Lecture Notes in Electrical Engineering 187

Rongjun Shen
Weiping Qian *Editors*

Proceedings of the 26th Conference of Spacecraft TT&C Technology in China

Shared and Flexible TT&C (Tracking, Telemetry and Command) Systems

TSINGHUA UNIVERSITY PRESS

Springer

Lecture Notes in Electrical Engineering

Volume 187

For further volumes:
http://www.springer.com/series/7818

Rongjun Shen · Weiping Qian
Editors

Proceedings of the 26th Conference of Spacecraft TT&C Technology in China

Shared and Flexible TT&C (Tracking, Telemetry and Command) Systems

TSINGHUA UNIVERSITY PRESS

Springer

Editors
Rongjun Shen
Chinese Academy of Engineering
Beijing
People's Republic of China

Weiping Qian
Beijing Institute of Tracking
and Telecommunications Technology
Beijing
People's Republic of China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
ISBN 978-3-642-33662-1 ISBN 978-3-642-33663-8 (eBook)
DOI 10.1007/978-3-642-33663-8
Springer Heidelberg New York Dordrecht London

Jointly published with Tsinghua University Press, Beijing
ISBN: 978-7-302-30159-2 Tsinghua University Press, Beijing

Library of Congress Control Number: 2012948182

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The start of the twenty-first century has seen rapid development of China's space endeavors. Breakthroughs have come one after another from manned spaceflight missions and lunar exploration missions. The first-generation Chinese tracking and data relay satellite system is already operational and a full constellation Compass/Beidou navigation satellite system will soon be in place. Very productive space applications have been developed for ocean environment and earth resource monitoring, communications, and meteorology. Even more exciting milestones are yet to come as the momentum builds in these and other rising fields in the near future. China's space station will soon be launched and will soon send a rover to land on the moon and return to the Earth. China will also send a probe to explore Mars.

The rapid development of China's space capabilities has included successful development of Tracking, Telemetry and Command (TT&C), and communication systems to meet unprecedented opportunities and challenges. As a result of the persevering efforts of generations of scientists, China already has TT&C and communication systems that encompass land-based ground stations, seagoing instrumentation ships, instrumentation aircraft and tracking, and data relay satellites. The system has the capability to provide a wide range of support to all types of spacecraft at all altitudes of earth's orbit. China's deep space TT&C network successfully provided support to its Chang'E-1 and Chang'E-2 lunar exploration missions and significant growth of the network is expected in the near future. However, there are still numerous issues needing in-depth studies in China's TT&C system. For example, interconnectivity between TT&C networks and the integrated utilization of resources needs further development, inter-department and inter-mission coordination is too complicated and more synergy is needed in the development of ground systems.

With "Shared and Flexible TT&C Systems" as its theme, the 26th Conference of Spacecraft TT&C Technology in China highlights the latest developments related to standardization, informatization, networking and intelligent technologies, and their applications in the field of aerospace TT&C. The objective is to promote the development of an interconnectivity-oriented, resource-sharing, responsive,

flexible, and efficient TT&C system with state-of-the-art technologies based on the specific needs in China. Only in this way is it possible for us to assure healthy long-term development of our TT&C system and to meet the future requirements of China's space program.

The conference, which is organized by the Spacecraft TT&C Committee of the Chinese Society of Astronautics, received more than 250 papers from experts nation-wide with 42 selected for publication. I firmly believe that publication of the proceedings of this important conference will promote international exchange and provide a new channel for sharing of the latest research achievements and engineering practices in the field of spacecraft TT&C.

October 2012 Rongjun Shen

Contents

Part I
Spacecraft TT&C System Design and Research

Chapter 1
Space-Based MA TT&C System and Technologies

Jianping Hu, Hongjun Yang and Maoge Xu

Abstract Multiple Access (MA) service provided by Tracking and Data Relay Satellite System (TDRSS) is a space based MA TT&C and communication service using satellite relay. Spaced-based MA TT&C system is specialized in the implementation and application. In this paper the performance and technology development of S-band MA service provided by TDRSS is analyzed. As a feature of TDRSS SMA services, the major difference from that of conventional communication satellite is that its MA assignment has to be relayed and managed by ground station. Key technologies related to space based MA TT&C and communication are determined and their implementation solutions are proposed.

Keywords Space-based MA TT&C and communication · Data relay satellite · MA access · Adaptive beam forming · Demand access · Calibration

1.1 Introduction

As a typical space based TT&C and communication system, data relay satellite system is able to provide user with tracking and data relay capabilities. Its Ku/Ka-band inter-satellite link is used for high rate data transmission, and its S-band single access (SSA) and S-band multiple access (SMA) services are used

An erratum to this chapter is available at 10.1007/978-3-642-33663-8_43.

J. Hu (✉) · H. Yang · M. Xu
Southwest China Institute of Electronic Technology, Chengdu 610036, China
e-mail: jphu63@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_1,

for command, telemetry, data transmission and ranging. With onboard phased array, SMA Service provides a communication path from customer platform via a geosynchronous data relay satellite to a customer control centre.

MA system of data relay satellite is a MA TT&C and communication system using space multi-beam and code division multiple target via satellite relay [1, 2]. An obvious difference from common MA communication satellite is that its MA service is relayed and managed by ground system, through which the communication and TT&C services between user platform and user centre can be realized. The space link involves on-board multi-beam and space-ground MA, resulted in special system implementation and application. The corresponding key technologies shall be studied and implemented in system design [3, 4].

1.2 Development of Space-Based MA TT&C and Communication Technology

Space Network (SN) concept is proposed firstly by NASA. Now its TDRSS system is established and consists of two generations of nine satellites and three ground stations. No Zone of Exclusion (ZOE) exists in the coverage provided by current TDRSS constellation, and 100 % orbit coverage is provided for user spacecrafts at 200 km altitude. The data relay satellites from ESA and NASDA provide only data relay capability without MA communication service. Space-based MA communication technology was used in the first and second generation of TDRSS [5]. In its third generation satellite system design, ground multi-beam forming, which was utilized in the first generation, is also used to enhance communication security.

In development of two generations of TDRSS satellite system, in addition to advanced technologies, its application range is expanded to include various users, including spacecrafts in LEO and MEO, South Pole science survey team, rocket, Long Duration Balloon Platform (LDBP) and so on.

The third generation satellite is designed based on second generation satellite with following differences: (1) SMA return beam is formed on ground, and (2) onboard command and telemetry link communication security (COMSEC) system is upgraded.

For user access, conventional pre-assignment mode was used in TDRSS initially. It can provide satisfied service for large spacecraft, but can not provide rapid and flexible automatic service for other spacecrafts such as small satellite because of high occupation cost on system equipment, prepare time and human resource. So in order to enhance customer MA access capability and increase customer number, Demand Access System (DAS) was used in ground station in 2004 to provide non-schedule MA return link service. With DAS, any number of beamformer can be incorporated in ground station, which is only limited by the number of available receiver/demodulator.

The Demand Access System will provide services by adding global system control, coordination functions, and data distribution capabilities into the ground systems. For the new services enabled by DAS, the DAS will [2, 6]:

1. Provide the capability for continuous, conflict-free, DAS MA return link services 24 h per day, 7 days per week upon demand from customers.
2. Provide an automated capability to transition DAS customer services between TDRSs/Space Ground Link Terminals (SGLT).
3. Provide the capability to support multiple, independent MA return links per TDRS/SGLT/ground station.
4. Meet or exceed the current communications performance and capabilities of the existing MA return link with the exceptions of the functions not possible due to the lack of tie-ins with the MA forward link.
5. Provide demodulation and data distribution capabilities for each DAS MA return data service.
6. Automate the operation of all DAS return link services.
7. Provide COTS data and control interfaces for DAS customers with the flexibility of accommodating non-standard/customer-unique telemetry interfaces.
8. Provide simple, low cost, modular expansion capabilities to facilitate the addition of DAS return link channels, as needs change.

With a new demand access capability, many customers will have low cost access to the SN. The new demand access capability is ideal for spacecraft flying in formation, which have relatively low data rate requirements or single spacecraft and other non-orbital vehicles that need a continuous communications link. Meanwhile, the adaptive ground-based multi-beamforming provides the system with flexible, stable, reliable, safety, expandable space access capability, and support DAS realization.

1.3 Key Techniques for Space-Based SMA TT&C and Communication

Space-based MA service is a special satellite MA communication system compared with common satellite communication system. Its MA access, customer management, FDM, and beamforming, calibration, and control of multi-beam antenna are precondition and guarantee to realize optimal communication performance. All key technical challenges shall be well resolved.

Refer to NASA's SMA designs and future plans from first to third generations of TDRSS, following development are observed:

1. Phased array antenna, which is used to implement Code Division + Space Division MA communication in space–space link.
2. FDM or Double DSSS, which is used in space-ground link for transmission of CDMA signals transmitted/received by elements in phased array.

3. Adaptive ground multi-beamforming, which can provide better adaptability to various customers and higher anti-jamming capability.
4. Multiple customer management and data relay, which is performed in ground station to enable DAS and provide better expandability.

This multiple channel MA communication is more complex than single channel communication and conventional MA satellite communication. Following technical challenges must be resolved.

1.3.1 Application and Management Mechanism for MA Communication

The application and management of MA communication is critical for ground operation control system, which is a key point for system normal operation. Space based SMA communication management using DAS is an effective method for application and improving of system capability.

Customers will realize many technical and logistical benefits from the implementation of DAS. Examples of benefits could include:

1. Immediate access to services in support of science and spacecraft safety.
2. Reasonable system resource configuration. A DAS return link could be configured to poll several different spacecraft automatically to reduce channel idleness and increase system availability.
3. Extended duration services without service interruption, which can improve operation flexibility and autonomy for customer vehicle.
4. Simplified scheduling and operation which can improve system automation and unattended monitoring levels.
5. Simplified expansion when service loading increases through the modular nature of DAS.

To establish DAS system, data flow and service flow with reasonable system configuration is core. Demand access system equipment configuration and data flow is shown in Fig. 1.1. Central management can be utilized for DAS equipment. IF signals from de-multiplex channels of several ground stations in same location are sent to given processor in expandable beamformer group via data distribution equipment for beamforming. The DAS equipment in ground station at different area can send data to central station DASCON via data transmission network to implement central control and management.

Schedule, initiating, and monitoring of DAS service are all performed automatically by DASCON, i.e. beamforming, demodulation, and return data distribution are all managed automatically based on DAS customer resource configuration requirements without operator intervention so as to real-time support space mission and ground emergency service.

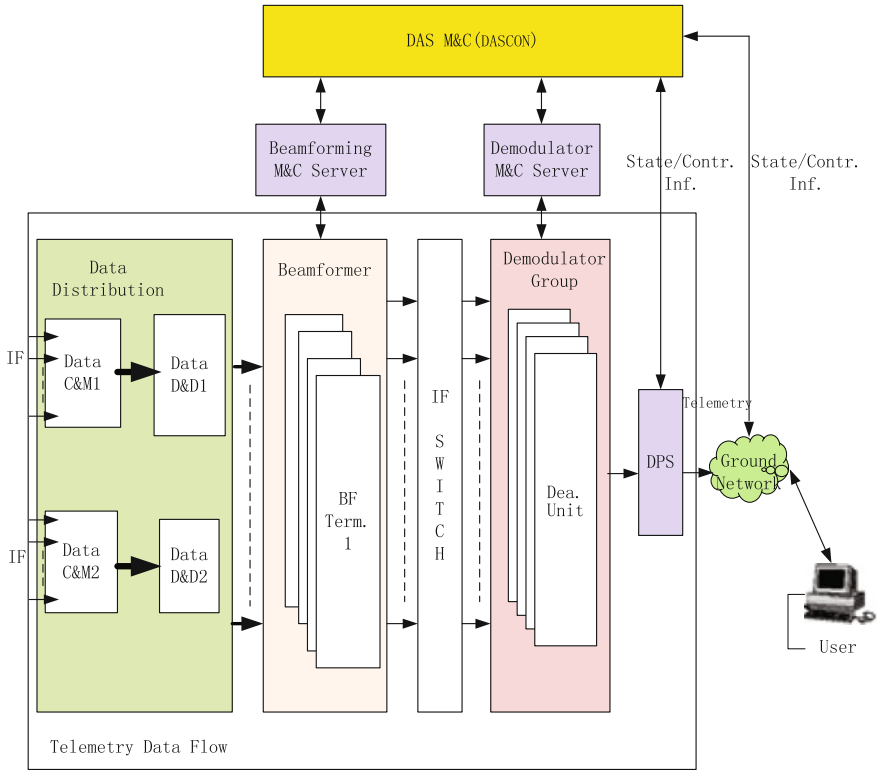

Fig. 1.1 Schematic diagram of space based MA communication DAS components and data flow. *M&C* Monitor and Control, *C&M* Copy and Multiplexing, *D&D* Demultiplexing and Distribution, *Term.* Terminal, *Demo.* Demodulator, *DPS* Data Processing Server

1.3.2 MA Access Methods

Time division multi-beam controlled by ground station is utilized in forward link meanwhile a forward link to customer is established.

Multi-beamforming is performed on ground at the same time for return link, which is key for multi-beamforming and customer access. SCDMA and CDMA are used to access multiple customers, which is a new technology for TT&C. So there is a challenge to select a proper MA access for special TT&C application. By comprehensive analysis, there are two options for SDMA and CDMA scheme of SMA return link [6, 7].

1.3.2.1 Multi-Beamforming After De-Spread

Due to all signals within 26° angle of view are received by each antenna element, each signal of 30 elements are de-spread at first to separate user signals received by each element, the signals from same user received by 30 elements are weighted and combined to obtain receive beam directing to corresponding user. Beam-forming shall begin at any initial weight value vector and adjust adaptively to optimal value.

No beam-forming is needed before de-spread in this solution. Therefore accuracy geometric relationship between TDRS and user vehicle is not necessary, especially with the cost of 15 dB combination gain reduced at de-spread.

1.3.2.2 Multi-Beamforming Before De-Spread

Signals from 30 elements are routed to six processors for complex number weighting to combine adaptively multiple beams corresponding to six users. The gain of combined beam is about 30 dB, providing good SNR for de-spread. Then signals from each user shall be further separated while all interferences shall be suppressed including same-frequency MA interferences from five users and other narrowband interferences.

This solution uses user ephemeris data, with known accuracy geometry relationship between TDRS and users as basis for open loop pointing. Adaptive beam-forming is performed as per maximum SNR or minimum mean square error rule.

When beam-forming is after de-spread, 15 dB combination gain is reduced as cost, so higher EIRP is required at user terminal, limiting possibility of part of small user vehicles accessing SMA. Consequently the access mode putting beam-forming before de-spread is necessary and is feasible technically.

1.3.3 MA Channel Features and Adaptability

In space–space and space-ground links, SSA/KSA digital communication features non-band-limit, single channel, single carrier, and limit power, so non-band-limit or low-band-limit data transmission model can be used. Because the signal is provided with constant envelope, non-linearity of channel has little effect on BER. Power amplifier can operate in saturation state, which is typical for power amplifier onboard spacecraft.

But SMA is different from SSA/SKA. Its space-ground return link operate in FDM multiple carriers mode, non-linearity of channel has more effect on signal characteristics. For single user spacecraft terminal, space–space link is low bandlimited with single channel and single carrier, so it has constant envelope. Low bandlimited and saturated power amplifier can be used in user terminal [8]. Space-ground return link from TDRS onboard transponder to ground terminal is

FDM multiplex with multiple carriers, and its envelope is fluctuant. Thus linear power amplifier shall be used in onboard transponder, and matched filter shall be added in ground receiver to reduce interference among codes.

Data transmission model of forward link is similar to that of return link, only except that ground station and transponder are at transmit end, and user terminal is at receive end. Space based MA forward link is a time-divided multi-beam link, establishing a forward link with one user each time. But its space-ground link is frequency-divided multiplex with KSA/SSA link, and there are multiple carriers in one channel. Linear power amplifier must be used in ground station to avoid effects on communication performance of forward MA and other SA services.

In addition, as an improvement on space based MA solution, constant envelope modulation can be used on channel modulation to improve spread spectrum code bit rate when occupied bandwidth channel allowed is constant.

1.3.4 Space Link Multi-Beamforming and Control Technology

Multi-beamforming is the key for MA communication. Based on NASA TDRSS system development, ground beamforming solution is accepted and verified. There are many methods for multi-beamforming on ground, such as main beam control (open loop), scan (open loop), and adaptive beamforming (close loop) [9]. Which kind of method shall be used decide the system space link performance and system application adaptability.

For main beam control, user spacecraft's ephemeris is used as time function, the TDRS location and attitude obtained from the tracking and telemetry data are also used as time function. From the two functions, ground station generates TDRS antenna pointing value and Doppler prediction value. Before user target is captured, open loop pointing software will predict the line-of-sight to user spacecraft and the phase and amplitude weight value of phased array using MA antenna coordinate as parameters. It is noted that the generated beam pointing is the beam pointing after 260 ms of space-ground transmission delay. For high speed spacecraft, the introduced point error is very high. Scan method is similar to this. The MA antenna beam shall scan in large scope as per some strategy. When ground system received signals from user target, antenna beam is controlled to approach to target to realize needed pointing.

Both the above methods are not flexible, especially for non-orbit users whose kinetic characteristic is more complex. Simpler beam open loop control mode perhaps causes antenna beam unreliable acquisition. For user targets with high dynamic characteristics, the target captured initially can also be lost at any time, so no reliable and stable tracking can be implemented.

The advantage of ground adaptive multi-beamforming (close loop) is that it can provide automatic target capture and auto-tracking capabilities, and it can make

the zero point of beam directional pattern aiming to interference when combined main beam pointing to the target. Therefore adaptive nulling antenna is formed to provide higher space filter and anti-jamming capability and reduce pressure on narrow-band strong interference caused by de-spread circuit of spread spectrum receivers.

In addition, ground multi-beamforming mode is precondition for using DAS service. Open loop control can not provide effective and flexible DAS control and expandable capabilities.

1.3.5 Beam Pointing Calibration Technology

During TDRS satellite operation, relative amplitude and phase of phased array can be partly changed because of some factors such as physical distortion of array caused by sunshine, sunshine and temperature effects on components within array and components aging. For onboard phased array antenna, on-orbit calibration shall be performed on its beam pointing performance.

In space based MA TT&C and communication system, beamforming is just a MA access means, and TT&C and communication effect achieving to maximum S/N is purpose. Its beam calibration requirement and means is different from that of radar system. Thus the key point of SMA system antenna beam calibration is integrated calibration on phase and amplitude in order to obtain optimum communication effect for all users in MA communication link.

In forward link, only a power amplifier at final stage is put after phase shifter in transmit channel. Therefore, the phase and amplitude inconsistency caused by them is small, so as to small S/N loss because of wide transmit beam width (about 7°). For TDRSS, the mean square root of phase inconsistency is 16° [2, 6]. When the amplitude inconsistency is 9 %, S/N loss will be 0.34 dB, so calibration is impossible in general. If the requirements are higher, the space-ground large loop calibration method similar to return link can be used. One of example is Vector Rotation Method.

For return link, the signals received by onboard phased array antenna shall be frequency division multiplexed, combined, and up-converted to Ka-band. After being routed through downlink channel and ionosphere, the signals is received, down-converted and de-multiplexed by ground equipment, then it can be combined. During signal processing and transmission, amplitude and phase inconsistency induced by different element signal frequency is main error source, shall be counted in ground multiple-beam processing and transmitted beam forming and processing. The effect of ionosphere periodic variation and thermal cycle of onboard equipment are also serious problem. In addition, satellite orbit shall be changed slightly with time variation, and then shall affect on antenna pointing.

For return link using open loop pointing mode, the beam shall be calibrated periodically by ground station [2, 6]. Using user signals simulated and transmitted by a set of fixed calibration equipment on ground station, calibration on channel

phase and amplitude is performed, during which user communication service perhaps will be interrupted, so system availability shall be affected. Because simulated user signal is routed through ionosphere, its characteristics is not consistence with actual spacecraft space-ground link, so will bring some calibration error.

For ground adaptive beamforming mode, fixed target calibration and adaptive calibration on ground can be incorporated effectively. Adaptive calibration can be performed without effect on system normal MA communication. Adaptive algorithm shall calibrate phase and amplitude of all channels automatically to obtain optimum communication effect and system availability. Fixed target calibration on ground can be performed during system initial establishment and maintenance, which is similar to calibration method for open loop pointing. Better adaptability can be provided with system by using adaptive ground multi-beamforming.

1.4 Conclusion

The development of MA TT&C service for NASA TDRSS system is introduced. SMA multi-beamforming for NASA TDRSS is developed from ground to onboard and then to ground, and DAS operation mode is used. The utilization of these technical measures can improve system application and expandable capability. Furthermore, these measures provide SMA system with better system functions and user number expandability, adaptability to various type users, and space link anti-jamming capability. Finally, key technologies and solutions are proposed for TDRS MA communication system particularity.

References

1. DuPree JE (1978) TDRSS multiple access receiving phased array system. ITC 337–344
2. Gitlin T, Nguyen DV (2000) Next generation TDRSS MA beam forming subsystem. ITC 841–848
3. Zillig D, McOmber R, Fox N (1998) TDRSS demand access service: application of advanced technologies to enhance user operations. In: SpaceOps 98 Symposium, Tokyo, pp. 1–5
4. Liu JX (1993) TDRSS overview. Telecommun Eng 39(3) (Supp.)
5. NASA (1996) 2nd TDRSS WORKSHOP
6. TRW (1983) TDRSS Design Report
7. Yi RT (1993) Discussion of ground-based multiple beam forming adaptive processing for TDRS onboard phased array antenna. Telecommun Eng 39(3) (Supp.)
8. Zhang JS (1979) Tracking and data relay satellite overview. J Spacecraft TT&C Technol
9. Hu JP (2004) Research of TDRSS user transponder technology. Telecommun Eng 44(3) (Supp.)

Chapter 2
The Investigation of a Novel Reentry Telemetry System

Xingwen Ding, Ming Chen, Xifu Huang and Ling Wu

Abstract With the development of the missile experimental research, the measurement bandwidth of super-fast-signal demands more and more (expected up to 100 MHz) in the reentry telemetry. However, the ability of the current PCM-PPK reentry telemetry system is limited. Also the PCM-PPK system is a non-universal system and impedes the improving of the telemetry technology. Based on the analysis of various modulation systems, the PCM-FM modulation system is proposed as a novel reentry telemetry system. The principle prototype of PCM-FM reentry telemetry baseband equipment is successfully developed with the code rate range from 10 to 100 Mbit/s.

Keywords Reentry telemetry · Novel modulation system · PCM-FM · High code rate

2.1 Introduction

As an important part of telemetry, the reentry telemetry is a kind of special telemetry technology which mainly measures various physical parameters of the aircraft (such as warhead) in the course of atmospheric reentry [1]. Its basic principle is the same as the telemetry technology of ballistic missiles during boost phase and the spacecrafts during operational segment, but its work environment

X. Ding (✉) · M. Chen
Beijing Research Institute of Telemetry, Beijing 100076, China
e-mail: dingdaming88@hotmail.com

X. Huang · L. Wu
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_2,

and measurement requirements are different from those of general telemetry. The reentry telemetry is more complicated and difficult than general telemetry.

In years' missile experiment, various receiving systems and methods are summarized according to the characteristics of reentry signal. Especially after a long-term study, the PCM-PPK system has solved those super-fast-signal telemetry problems. However, with the development of the missile experimental research, there appear more and more measurement requirements for super-fast-signal, and the measurement bandwidth is approximately up to 100 MHz. But the ability of PCM-PPK system is limited, so it is necessary to carry out relevant research on the telemetry technology of super-fast-signal in reentry telemetry system, in order to meet the demands of the developing experiment. The research work should also give consideration to both common signal telemetry and super-fast-signal telemetry, and seek ways to construct a unified telemetry system, so that it can reduce the amount of equipment and system complexity.

The passage first analyses the specificity and complexity of reentry telemetry, and then proposes the PCM-FM modulation system as a novel reentry telemetry system based on the analysis of various modulation systems. Finally, a principle prototype of PCM-FM reentry telemetry baseband equipment is developed with excellent performance.

2.2 Characteristics of Reentry Telemetry

When ballistic missile warhead is re-entering the dense layer of the atmosphere, it will encounter adverse mechanical, thermal and electromagnetic environment. The overload of warhead is up to tens of times of gravitational acceleration, and its vibration spectrum up to thousands of Hertz. The temperature of warhead shell can reach thousands of centigrade because of the friction between the high-speed reentry warhead and the surrounding air, so that the surrounding air conducts ionization and forms plasma sheath around the warhead, also called the "black-out", which can prevent the radio wave transmission. When hitting the ground, the overload of warhead will be up to tens of thousands of times of gravitational acceleration. Meanwhile, because of harsh electromagnetic environment in the warhead, the interference current on the ground wire can be up to tens of amps and it seriously interferes with the telemetry equipment.

Besides the poor mechanical, thermal and electromagnetic environment, the reentry telemetry still faces another serious problem: very short time for measurement. After the warhead flies out from the blackout area, the time for measurement is only about 3 s. During such a short time, it not only transmits real-time measurement parameters, but also need transmit the kept measurement parameters in the blackout area. Even more, the time for measurement of touch-down signals has only several hundred microseconds.

Obviously, the particularity of reentry telemetry mainly comes from three aspects: the particularity of the measurement parameters, the special internal

environment in warhead and the special external environment caused by warhead reentry to atmosphere. Due to the particularity of the reentry telemetry, it has the following characteristics [1, 2]:

1. Certain signals have a very wide frequency band.
2. The super-fast-signal has only one-time measuring chance and high-precision interval measuring requirement.
3. The height of warhead is low after flying out from the blackout area, so the time for transmitting signal and receiving signal is very short.
4. The layout of the telemetry components in the missile has strict requirements, so it is not easy to be destroyed by the detonation debris or stress wave.
5. The telemetry components in the missile should be small and light.
6. Because of the signal interruption caused by black-out area, the reentry telemetry system should have strong capability of the retransmission and large storage memory for retransmission.
7. The channel of reentry telemetry is a kind of variable parameter channel due to the movement and rotation of the warhead.
8. The reentry telemetry system has some features of high overload telemetry.
9. The reentry telemetry system should have sufficient power margin.

2.3 Investigation of the Novel Reentry Telemetry System

The novel reentry telemetry system must be greatly adapted to the characteristics of the reentry telemetry so as to ensure the completion of the reentry telemetry tasks. However, various modulation systems have different adaptability to reentry telemetry, so that it is necessary to carry out theoretical analysis and simulating study of various modulation schemes in order to select one (or more) modulation system which is suitable for the burst transmission of high code rate and also has high spectrum efficiency.

2.3.1 Phase Modulation System

In the reentry telemetry, the carrier modulation generally does not adopt phase modulation which is not suitable for time-varying channel. In addition, the PM demodulator needs a certain setting time to recover carrier, which is very unfavorable to reentry storage telemetry [1]. Phase modulation system generally uses coherent demodulation technique which should not be used in the reentry telemetry. Therefore, the phase modulation, such as PM, BPSK, QPSK, OQPSK, UQPSK, 8PSK, FQPSK, QAM, APSK, etc., (among them, QAM, APSK belong to both amplitude modulation and phase modulation) is not suitable as a novel reentry telemetry system.

2.3.2 Amplitude Modulation System

PCM-AM is not sensitive to carrier phase, but is rarely used in conventional telemetry. That is because its transmit power is proportional to the square of signal amplitude ($P = I^2R$). In order to ensure the transmission quality of the low amplitude telemetry signal, the transmission system needs additional power margin. Once the maximum-to-minimum ratio of the measured signal amplitude is large, its transmitter power will encounter big problems in design.

PCM-PPK, which is currently used in the reentry telemetry system, is just a kind of PCM-AM. Some data shows that PCM-FM transmitter power (for code rate 102.4 kbit/s) is required only less than 5 W; but PCM-PPK transmitter power (for code rate 51.2 kbit/s) requires no less than 300 W (pulse power) which is 60 times more than PCM-FM. Moreover, as the reentry telemetry further promotes the requirement of code rate (up to tens of Mbit/s and even hundreds of Mbit/s), it needs to further increase the PCM-PPK transmitter power, which leaves the system more difficult to be realized.

2.3.3 Spread Spectrum System

Spread spectrum system is a widely used TT&C system. It can complete different tasks such as ranging, velocity measurements, telemetry, security control, etc. It has the advantage of low power spectral density, anti-intercept, anti-jamming, security, multiple access communications, etc. However, spread spectrum system adopts coherent demodulation technique and needs three synchronization process: carrier synchronization, PN code synchronization and information code synchronization. Considering this factor, spread spectrum system is even less suitable for reentry telemetry system than the phase modulation system mentioned above.

2.3.4 OFDM System

The basic principle of OFDM system is to divide the whole channel into several orthogonal sub-channels, and then convert the high-speed serial data stream to multiple parallel transmission data streams which will be modulated on each sub-channel with low-speed.

Because of the orthogonality between each sub-channel, OFDM improves the band efficiency greatly. In addition to the high band efficiency, OFDM system also has some other advantages, such as resistance to multi-path delay spread, frequency selective fading, anti-inter-symbol interference, etc. But OFDM system

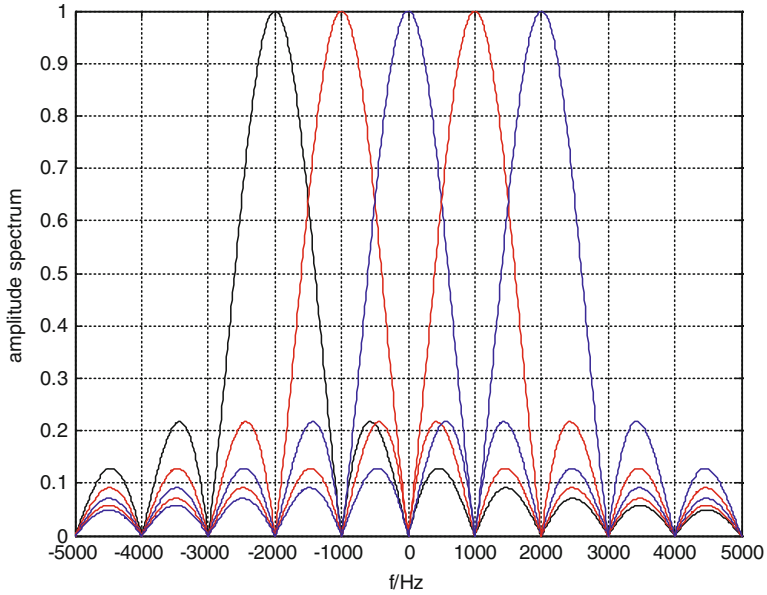

Fig. 2.1 The spectrum of OFDM signal

usually adopts coherent demodulation technique, and furthermore it has two serious shortcomings: sensitive to the frequency offset and high Peak to Average Power Ratio (PAPR). Therefore, OFDM system is not suitable for reentry telemetry either.

The spectrum of OFDM signal is shown in Fig. 2.1. Each sub-carrier spectral peak appears at other channels spectral null, so it requires the strict frequency synchronization. Once the system has frequency offset, the orthogonality among sub-carriers will be destroyed. That will cause serious "floor effect" on system performance. That means, no matter how to increase the transmitter power, it can not improve the system performance significantly.

Another disadvantage of OFDM system is that the PAPR is high. Figure 2.2 shows a diagram of OFDM signal in time domain. It shows that the envelope of OFDM signal has the fluctuation characteristic. That's because OFDM signal is a weighted sum of all the sub-carrier signals. When the phases of multiple sub-carriers are identical, the instantaneous signal power will be much greater than the average signal power, so its envelope dynamic range is large. That is, Peak to Average Power Ratio (PAPR) is high. High PAPR causes a high requirement of the linear range of the amplifier and D/A converter. If the linear range of the system can not meet the requirement, it will cause signal distortion and system performance degradation.

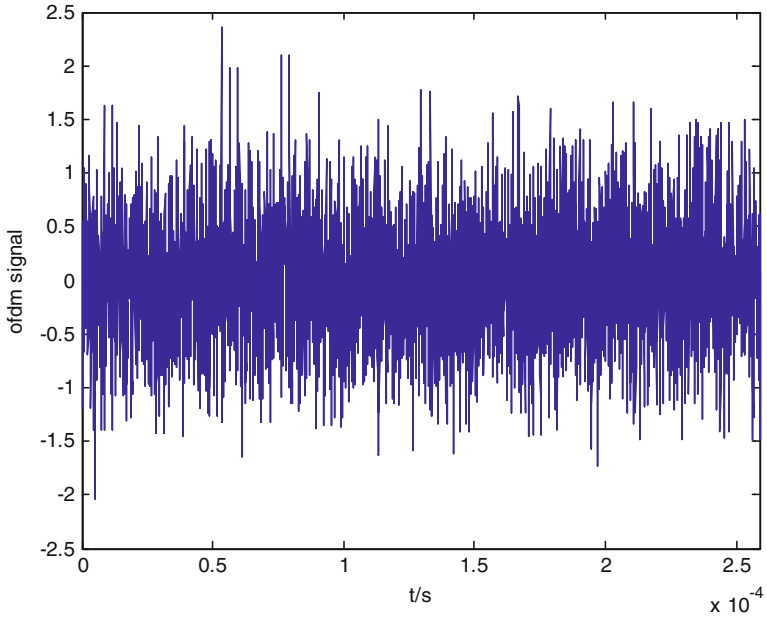

Fig. 2.2 OFDM signal in time domain

2.3.5 Continuous Phase Modulation System

Continuous Phase Modulation (CPM) system has constant envelope and excellent spectral characteristics. CPM is not sensitive to non-linear of the power amplifier, so the system can use non-linear power amplifier in practical applications. Furthermore, CPM signal can adopt non-coherent demodulation technique, which greatly simplifies the design complexity of telemetry system. Table 2.1 compares several common CPM systems: PCM-FM, MSK, GMSK and ARTM CPM. Among them, PCM-FM has the lowest complexity, the best demodulation performance and the lowest band efficiency, while ARTM CPM has the highest complexity, the worst demodulation performance and the highest bandwidth efficiency.

PCM-FM has been the main modulation system in the launch vehicle range telemetry system in china and abroad. That is because PCM-FM modulation system has the advantages of short acquisition time and strong capability of anti-flame, anti-fading, anti-jamming. Meanwhile, it can improve the bandwidth efficiency nearly 35 % using pulse shaping pre-modulation filter technology [3]. In addition, as shown in Fig. 2.3, the PCM-FM telemetry system with "MSD (Multi-symbol Detector) + TPC (Turbo Product Code)" technology can have nearly 9 dB channel gain at the bit error rate of 10^{-7}, compared to conventional Limiter Discriminator (LD) demodulation techniques [4]. Such a high channel gain can

Table 2.1 Performance of several common CPM systems

CPM system	99 % of Normalized bandwidth	Minimum euclidean distance	System complexity
PCM-FM	1.78R[a]	2.43	★[b]
MSK	1.18R	2	★★
GMSK (BT = 0.5)	1.04R	1.9	★★★
ARTM CPM	0.56R	1.39	★★★★★

[a] R represents signal code rate

[b] The more number of ★ represents the higher system complexity

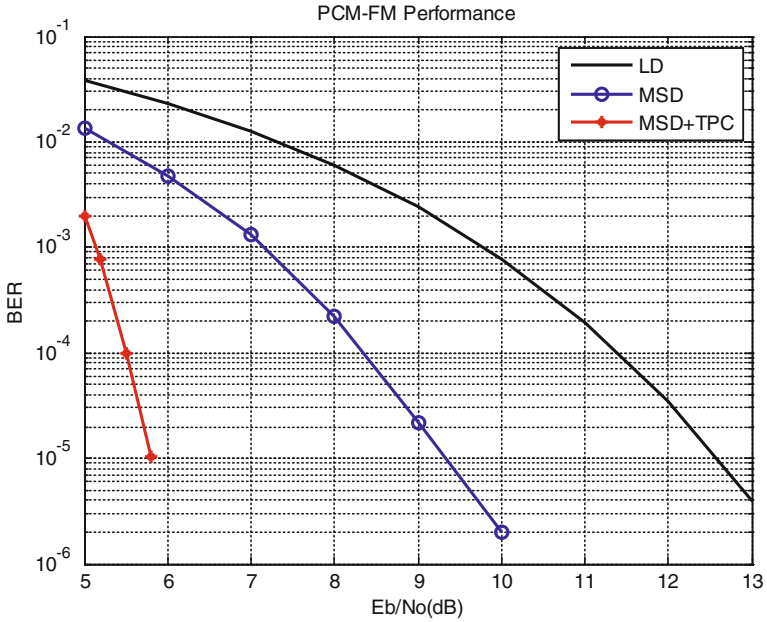

Fig. 2.3 PCM-FM performances with different demodulation technique (The parameters of PCM-FM modulation system are modulation index h = 0.7, the observation interval N = 5 and TPC (64, 57) × (64, 57))

significantly improve the transmission rate under the conditions of not increasing antenna aperture and transmitter power, and thus it can save cost for telemetry system.

Summarizing the modulation systems' adaptability to reentry telemetry above, PCM-FM can be chosen as a novel reentry telemetry system to unify modulation system, decrease equipment quantity and reduce system complexity.

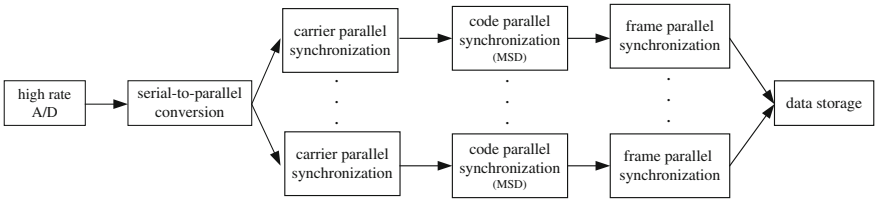

Fig. 2.4 The parallel demodulation process of the principle prototype

2.4 Principle Prototype of PCM-FM Reentry Telemetry Baseband Equipment

In order to meet the measurement requirements for the future reentry telemetry, it is estimated that the principle prototype of PCM-FM reentry telemetry baseband equipment needs to adapt the code rate up to 100 Mbit/s, which puts forward technical challenges on both hardware architecture and software architecture.

Firstly, the carrier center frequency of the traditional baseband equipment is 70 or 140 MHz, but the center frequency of the principle prototype needs to be higher in order to adapt to high code rate transmission. In the remote sensing system, the center frequency of baseband equipment with high code rate transmission (modulation system: QPSK, DQPSK, SQPSK, UQPSK, 8PSK, and total code rate: 10–600 Mbit/s) is 720 MHz. Therefore, the center frequency of the principle prototype for reentry telemetry can be also selected at 720 MHz.

Secondly, the traditional baseband equipment can complete demodulation of low code rate signal with serial processing, but the principle prototype for reentry telemetry with high code rate must adopt the parallel processing technique to reduce the processing rate, including serial-to-parallel conversion technique, carrier parallel synchronization technique, code parallel synchronization technique (parallel MSD technique), frame parallel synchronization technique, high code rate data storage technique, etc. The parallel demodulation process of the principle prototype is shown in Fig. 2.4.

With the use of advanced software radio architecture and parallel processing technology, the principle prototype of PCM-FM reentry telemetry baseband equipment has been developed:

1. Modulation system: PCM-FM (modulation index: $h = 0.7$);
2. Center frequency: 720 MHz;
3. Code rate: 10–100 Mbit/s, with step of 1 bit/s;
4. Demodulation method: non-coherent multi-symbol detection (MSD), the observed interval $N = 5$;
5. Carrier acquisition performance: Doppler range: ± 1 MHz, Doppler rate-of-change: ± 200 kHz/s, acquisition time: ≤ 1 ms;
6. Demodulation performance: the bit error rate $P_e \leq 1 \times 10^{-4}$ at $E_b/N_0 = 9$ dB.

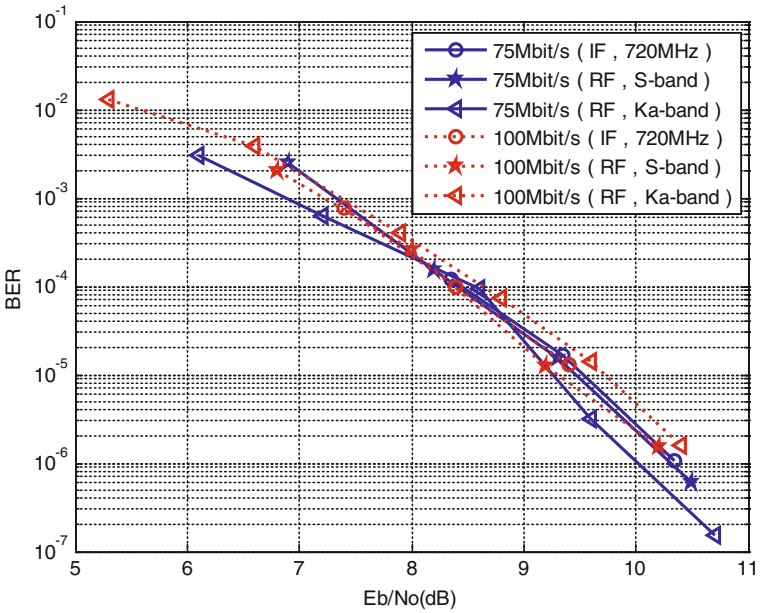

Fig. 2.5 The testing performance of the principle prototype

The principle prototype has been verified at IF (720 MHz) and RF (S-band and Ka-band) with wire connection in the laboratory, and its performance is shown in Fig. 2.5. It can be seen that the principle prototype of PCM-FM reentry telemetry baseband equipment has excellent performance and the demodulation loss is less than 0.5 dB compared with the theoretical performance ($P_e \leq 1 \times 10^{-4}$ at $E_b/N_0 = 8.4$ dB). Moreover, it can adapt to different radio frequency with very short acquisition time, so it is suitable for the burst transmission with high code rate in the reentry telemetry.

2.5 Conclusion

With the development of the missile experimental research, there appear more and more demands for the measurement bandwidth of super-fast-signal (expected up to 100 MHz) in the reentry telemetry. However, the ability of the current PCM-PPK reentry telemetry system is limited. According to the characteristics of super-fast-signal and reentry telemetry, the paper discusses various modulation systems' adaptability to reentry telemetry. On this basis, the PCM-FM modulation system is proposed as a novel reentry telemetry system. And the principle prototype of PCM-FM reentry telemetry baseband equipment has been successfully developed with excellent performance and the code rate range from 10 to 100 Mbit/s.

It achieves the anticipated goal of constructing a unified telemetry system to reduce the amount of equipment and system complexity, and also meeting the measurement requirement for the development of the missile experiment research.

References

1. Xie MX (1992) Reentry telemetry technology, vol 2. National defense industry press, Beijing
2. Huang Q (2006) Study on the characteristics of wait-receiving reentry-telemetry channel. Doctoral thesis of university of electronic science and technology, Chengdu
3. Wang KL, Liao XH, Wang F (2011) Research on high bit rate digital modulation technology. J Telem Track Command 32(1):22–27
4. Wang XB, Wu L, Xu SY (2007) Apply MSD and TPC in the PCM-FM telemetry system. J Telem Track Command 28(11):49–53

Chapter 3
Research on United Tracking Schemes Based on a New Information Frame Format

Haitao Nan, Zhiqiang Li, Wenming Zhu, Peng Jia and Feilong Li

Abstract In a Tracking, Telemetry and Command (TT&C) system for the upper-stage of a launch vehicle, a mass amount of instant and dynamic telemetry information need to be transmitted to the ground station from the launch vehicle. Meanwhile, ranging accuracy is seriously affected because the energy of telemetry information is more than the ranging information. To meet requirements for transmission of different information on the downlink of launch vehicle upper-stage, a new information frame format based on isochronous service and insert service is proposed. Moreover, three united tracking schemes including direct switch, free oscillation switch and accumulation branch assisted switch are proposed. The performance of the three schemes is simulated with MATLAB. The result shows that all three schemes can implement the switch between different loops, but there exists a phase jump at 20 ms both in direct switch and free oscillation switch. Only the accumulation branch assisted switch can switch smoothly. Therefore, it is the best scheme. The research findings are of high significance for practical applications.

Keywords Upper-stage · Frame format · Isochronous service · Insert service · Tracking scheme

H. Nan (✉) · Z. Li · W. Zhu · P. Jia · F. Li
Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China
e-mail: jack119wells@gmail.com

H. Nan
Xi'an Satellite Control Center, Xi'an 710043, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_3,

3.1 Introduction

The upper stage is the electronic system to pilot and control the launch vehicle. Because lots of instant dynamic telemetry information needs to be transmitted to the ground and the energy of telemetry information is more than the ranging information, the ranging accuracy shall drop significantly in the spread spectrum system. If the telemetry information and the ranging information can be transmitted in different time slots the delay is kept stable, the ranging accuracy will be improved significantly. The isochronous service and insert service regulated by the Advanced Orbiting System (AOS) of Consultative Committee for Space Data Systems (CCSDS) recommendations can be used for reference to design the frame format. Reference [1] shows the result that utilizing the insert service to implement pseudo-random code ranging is the best scheme under current condition. Furthermore, the feasibility of this scheme is validated through experiment. Reference [2] also validates this scheme. In this paper, an information frame format of a launch vehicle downlink signal is designed with reference to the idea of insert service. A tracking scheme is proposed and its performance is simulated.

3.2 Design of the Information Frame Based on Isochronous Service and Insert Service

S-band is used for transmit and receive and the modulation is BPSK. The downlink information transmission rate is 1.023 Mbit/s. Turbo code with 1/3 code rate is used for channel coding and the code block length is 1784 bits. The ranging frame length is 1000 bits and frame frequency is 1 frame/second. The synchronization header is 40 bits and the measurement information is 960 bits. To avoid influence of the combination course on the ranging time delay jitter, the frame header is not incorporated in channel coding and the rest information is incorporated in channel coding.

The transmission rate of the synchronization header should be reduced to ensure reliability because the synchronization header is not included in coding. The synchronization header rate is 1 kbit/s when the information rate is 1.023 Mbit/s. The Signal-to-Noise Ratio (SNR) is increased by about 30 dB compared to the rate not reduced. The synchronization header length can be arranged as 40 ms. For the purpose of isochronous insert, the frame frequency is 1 frame/second and the measurement information of every frame is 960 ms. The synchronization header adopts spread spectrum to match the downlink information rate. The numbers of Turbo blocks in each frame is $0.96 \times 1.023 \times 106/1784 = 550.5$. Therefore, 549 blocks can be arranged for measurement information and just 1 block for ranging information (Fig. 3.1).

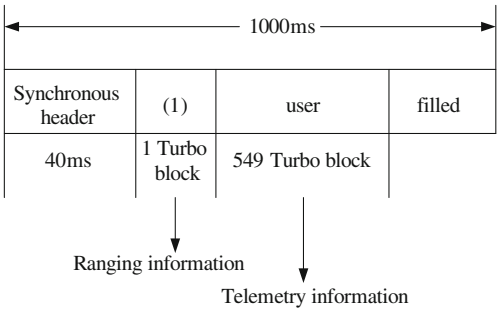

Fig. 3.1 Format of the downlink frame

3.3 Design of Tracking Program

The tracking loop can be designed as follows: the synchronization header is the data with the speed of 1 kHz, after spreading by C/A code with length of 3069 to the speed of 3.069 MHz. Here we will adopt Delay Locked Loop (DLL) to track the synchronization header. Behind the synchronization header is the telemetry information and ranging information with the speed of 3.069 MHz, which is coded by Turbo code and can directly adopt bit synchronization loop without the use of DLL. Thus, there exists the problem of switch between the two loops. In this paper, the direct switch, free oscillation switch and accumulation branch assisted switch are the main considerations.

3.3.1 Direct Switch

3.3.1.1 Scheme Design

As the simplest scheme, the principle of direct switch can be seen in Fig. 3.2. A, C is the place that DLL switches to information bit tracking loop; B, D is the place that information bit tracking loop switches to DLL. Under this scheme, the telemetry information and ranging information is time divided. Frequency difference and phase difference should be taken into account when the loop begins to work. But after each switch, the loop has to re-define the frequency difference and phase difference. Thus, this scheme is not the best choice although it is easy and convenient.

3.3.1.2 Simulation and Analysis

We set the parameters as follows: initial phase difference is $ph = \pi/3$. Initial frequency difference is $df = 2\,\text{Hz}$, and carrier-to-noise ratio is $C/N_0 = 75\,\text{dB} \cdot \text{Hz}$. Coherent integral time of DLL is $T = 0.1\,\text{ms}$. Tracking time of DLL is 20 ms.

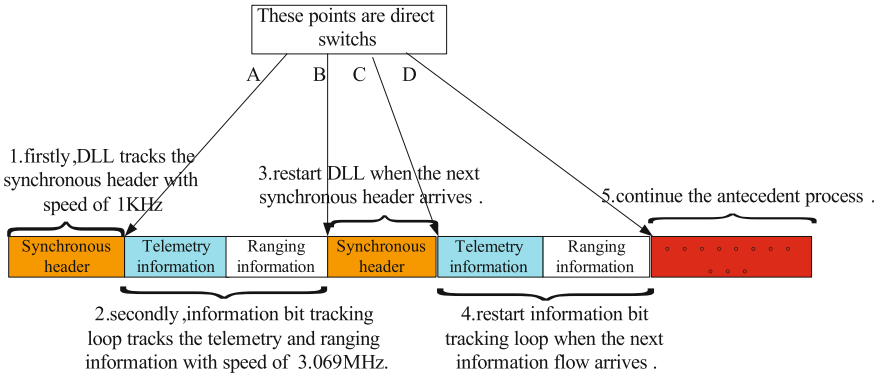

Fig. 3.2 The principle of direct switch

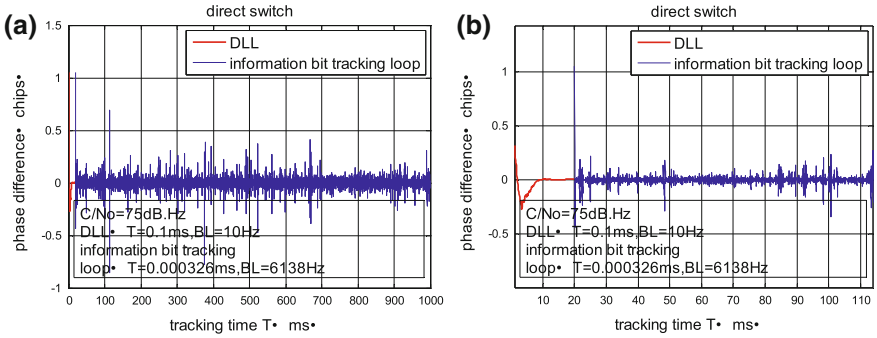

Fig. 3.3 **a** $C/N_0 = 75$ dB-Hz, DLL: T = 0.1 ms, BL = 10 Hz; information bit tracking loop: T = 0.000326 ms, BL = 6138 Hz. **b** $C/N_0 = 75$ dB-Hz, DLL: T = 0.1 ms, BL = 10 Hz; information bit tracking loop: T = 0.000326 ms, BL = 6138 Hz (0–100 ms amplified)

Bandwidth of the loop is $B_L = 10$ Hz; Coherent integral time of information bit tracking time is $T = 0.000326$ ms. Tracking time of this loop is 980 ms and bandwidth of the loop is $B_L = 6138$ Hz. Simulation result of the switch is shown in Fig. 3.3:

It is clearly indicated that the direct switch can realize the switching between DLL and information bit tracking loop. The phase jumped point in Fig. 3.3 is the switched point. The reason that the shaking performance of DLL and information bit tracking loop under $C/N_0 = 75\,\text{dB} \cdot \text{Hz}$ have big disparity is: the speed of the frame header of DLL is 1 kHz, $R_b = 30$ dB. The renewed period of loop is 0.1 ms, corresponding to 40 dB. Thus, $E_b/N_0 = 75 - 40 = 25$ dB. But for the information bit tracking loop, the speed of information flow is 3.069 Mbit/s, renewal period of loop is 0.000326 ms. Thus, $E_b/N_0 \approx 75 - 65 = 10$ dB. Therefore, the difference of the tracking results between these two loops is big. In the meanwhile, DLL converges at about 10 ms, after 100 times track, but information bit tracking loop converges at about 0.03 ms.

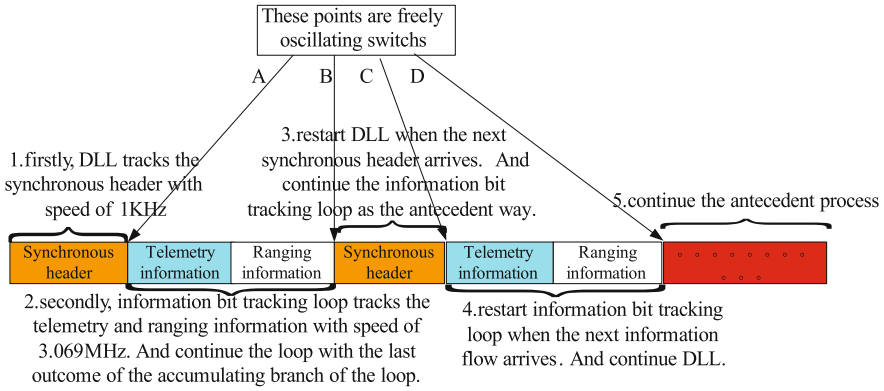

Fig. 3.4 The principle of free oscillation switch

3.3.2 Free Oscillation Switch

3.3.2.1 Scheme Design

Compared with stopping the DLL after tracking synchronization header in the direct switch, DLL can be designed to continue working until the next synchronization header arrives. In the meantime, DLL continues working with zero input, which amounts to predict the frequency difference and phase difference of the next synchronization header based on the last tracking result when the tracking finished. The principle of this scheme is shown in Fig. 3.4. A, C is the place that DLL switches to information bit tracking loop; B, D is the place that information bit tracking loop switches to DLL. Similarly, information bit tracking loop continues working with the zero input, which amounts to predict the frequency difference and phase difference of the next information flow based on the last tracking result when the track finished.

3.3.2.2 Simulation and Analysis

Under the same simulation condition as Sect. 3.3.2.1, the switch process of free oscillation switch is shown in Fig. 3.5.

Figure 3.5 is DLL free oscillation switch when the information bit tracking loop is working. In Fig. 3.5b, DLL does not stop working at 20 ms, but continues oscillation throughout 1000 ms. When the next synchronization header arrives, DLL begins to track again. There also exists a phase jump in free oscillation switch and that is because this scheme can't predict the frequency difference and phase difference of the next frame under high dynamic situation.

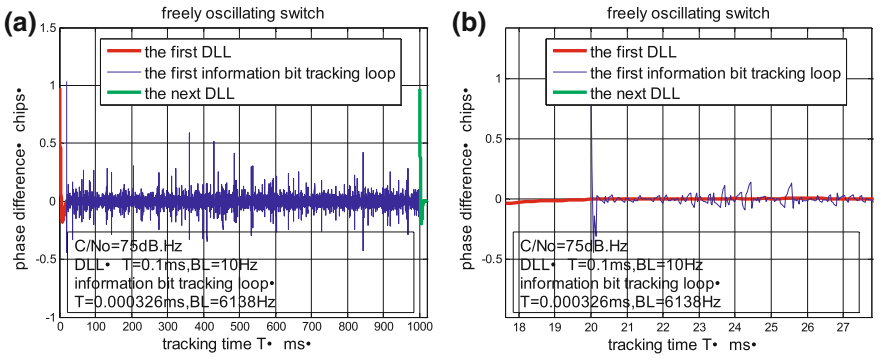

Fig. 3.5 **a** $C/N_0 = 75$ dB-Hz, DLL: T = 0.1 ms, BL = 10 Hz; information bit tracking loop: T = 0.000326 ms, BL = 6138 Hz. **b** $C/N_0 = 75$ dB-Hz, DLL: T = 0.1 ms, BL = 10 Hz; information bit tracking loop: T = 0.000326 ms, BL = 6138 Hz (18–27 ms amplified)

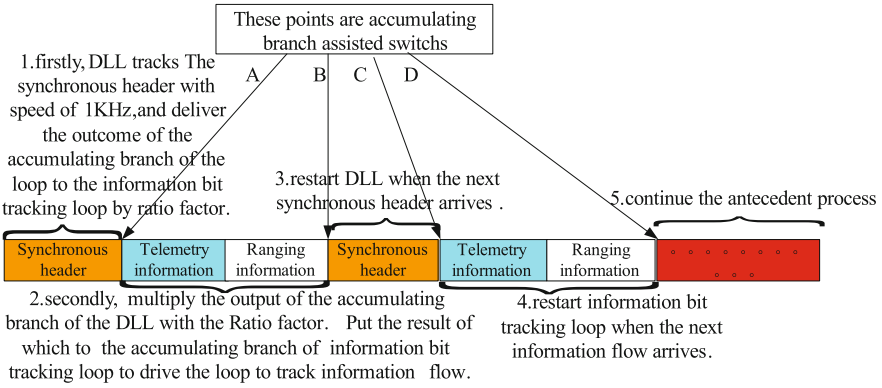

Fig. 3.6 The principle of accumulation branch aided switch

3.3.3 Accumulation Branch Assisted Switch

3.3.3.1 Scheme Design

If there exists an input of frequency steps to the loop, the output of the accumulation branch is a constant value when the loop is locked. This value is decided by parameters and the renewed period of the loop. If the ratio factor between the output of accumulation branch of DLL and information bit tracking loop can be calculated under the same input of frequency difference, this ratio factor can be arranged so that the output of accumulation branch of DLL is transformed to the accumulation branch of information bit tracking loop. Finally, the frequency difference can be precisely predicted and the loop can be switched smoothly. The

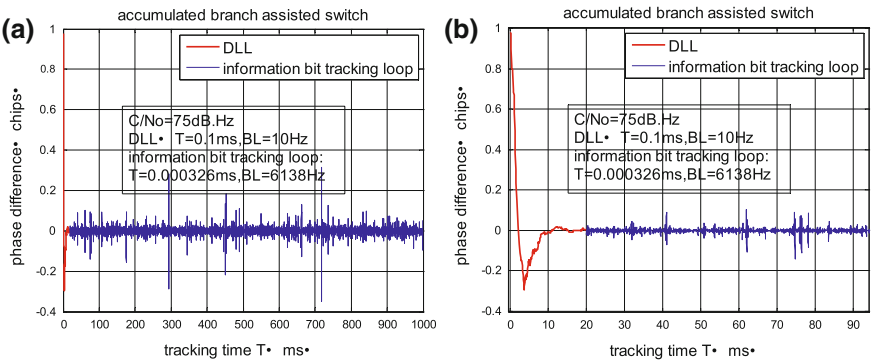

Fig. 3.7 **a** $C/N_0 = 75$ dB-Hz, DLL: T = 0.1 ms, BL = 10 Hz; information bit tracking loop: T = 0.000326 ms, BL = 6138 Hz. **b** $C/N_0 = 75$ dB-Hz, DLL: T = 0.1 ms, BL = 10 Hz; information bit tracking loop: T = 0.000326 ms, BL = 6138 Hz (0–100 ms amplified)

principle of this scheme is shown in Fig. 3.6. A, C is the place that DLL switches to information bit tracking loop; B, D is the place that information bit tracking loop switches to DLL.

3.3.3.2 Ratio Factor

The outputs of the accumulation branch of the DLL and information bit tracking loop under the same frequency steps input are [3, 4]:

$$\lim_{n\to\infty} H_{out1}(n) = \frac{K_{d1}}{K_{f1}T_{s1}}\Delta\omega \tag{3.1}$$

$$\lim_{n\to\infty} H_{out2}(n) = \frac{K_{d2}}{K_{f2}T_{s2}}\Delta\omega \tag{3.2}$$

Thus, the ratio factor is:

$$\eta = \frac{\lim\limits_{n\to\infty} H_{out1}(n)}{\lim\limits_{n\to\infty} H_{out2}(n)} = \frac{\frac{K_{d1}}{K_{f1}T_{s1}}\Delta\omega}{\frac{K_{d2}}{K_{f2}T_{s2}}\Delta\omega} = \frac{K_{d1}K_{f2}}{K_{d2}K_{f1}}\cdot\frac{T_{s2}}{T_{s1}} \tag{3.3}$$

It is evident that the discrimination gain K_d, filter gain K_f, loop renewed period T_s decide the value of ratio factor.

The renewed period can be calculated based on the simulation results of the two loops:

$$T_{s1} = 0.1 \times 10^{-3}\,\text{s}, \quad T_{s2} = 0.326 \times 10^{-6}\,\text{s} \tag{3.4}$$

Thus:

$$\eta = \frac{0.326 \times 10^{-6}}{0.1 \times 10^{-3}} K = 3.26 \times 10^{-3} K \tag{3.5}$$

where, $K = \frac{K_{d1} K_{f2}}{K_{d2} K_{f1}}$

Under the same condition as Sect. 3.3.2.1, the tracking results are illustrated in Fig. 3.7.

Simulation results indicate that the accumulation branch aided switch can precisely predict frequency difference and phase difference, and switching is smoothly done.

3.4 Conclusion

This paper proposes a design the downlink information frame of the upper stage of launch vehicles based on isochronous service and insert service specified by AOS. Furthermore, a design of tracking program is proposed based on the new frame format. Simulation indicates that the accumulation branch aided approach smoothly switches between DLL and the information bit tracking loop. Therefore, it is the best scheme and is of high significance for engineering applications.

References

1. Ding SQ, Yu ZJ, Fang HR (2002) Implementation of pseudo-random code ranging by using isochronous service of advanced orbit system. TV Technol (6):43–46
2. Ding SQ, Dong GL, Zhu KQ (2003) The experiment of verifying pseudo code ranging by insert service of AOS. Aerosp Shanghai (6):45–48
3. Gardner FM (2005) Phaselock techniques, 3rd edn. Wiley-interscience, New York
4. Kaplan ED, Hegarty CJ (2008) Understanding GPS principles and applications, 2nd edn. Publishing house of electronics industry, Beijing, pp 125–147

Chapter 4
A Research on the Architecture Design of Space Tracking, Telemetry and Control Networks

Xiangyang Lu, Lijuan Jia, Jin Hu, Jianguang Wu and Shiyong Du

Abstract The development of Chinese space tracking, telemetry and control (TT&C) networks are facing the challenge of informatization. In order to meet this challenge, it is necessary to research the architecture of the space TT&C networks according to the total development planning of our country's aerospace industry. This paper discusses the complex characters of space TT&C networks and then applies the Multi-Living Agents (MLA) theory to design and construct the future space TT&C networks. With this methodology, the objective of system design is concerned not only with the indicators of the system, but also the operation mechanism of the system. The significance of this theory in such architecture design is also indicated.

Keywords Multi-living agents theory · Space TT&C networks · Architecture design

4.1 Introduction

The TT&C network of our country has made remarkable progress in recent years. A huge system consisting of space TT&C station, task center, measurement and control center, application center and several other parts has been formed, which ensures the rapid development of our space industry [1, 2]. In the meantime, the

X. Lu · L. Jia · J. Hu (✉) · J. Wu · S. Du
School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
e-mail: bithj@bit.edu.cn

X. Lu
School of Electronics and Information, Zhongyuan University of Technology, Zhengzhou 450007, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_4,

rapid development leads to a higher demand for TT&C networks. From the perspective of external environment, a variety of factors have set higher requirement for the working efficiency and absolute stability of TT&C networks. These factors include the significant increase of aerospace crafts, major projects and tasks like new strategic arms, manned space flights, lunar exploration, and the internationalization demand of some tasks. Internally, the physical structure of TT&C network is getting increasingly complex, together with the presence of its space-based subsystem, the combination of various heterogeneous measurement and control equipment, and its rapid development and update, all of which have made the information exchange and task communication within the TT&C network confronted with major challenges. Due to historical reasons, the Chinese TT&C system has been in a development mode of "stovepipe system", which cannot meet current demand [3]. Specifically, resource allocation among the departments and tasks is not easy; the sharing and comprehensive utilization of data and process is difficult to realize; information resource waste is serious; and the effectiveness of system construction is low, etc. Therefore, in response to the rapid development of our country's aerospace industry, profound changes to the structure of TT&C networks are necessary. Informatization of the TT&C networks is an important approach of realizing the reformation. It reflects the national strategy, which is, fusion of information and industrialization [4].

The informatization of TT&C networks is a systematic project, in which the specific software and hardware system construction and implementation is just a small part. The IT strategic planning, professional work process classifying and settling, process optimization and management ascension prior to the system implementation all play a key role. All of these could be briefly summarized as system architecture design. At present, in the construction of information systems, there are a variety of concepts, technologies or standards that can be used to guide the system architecture design, such as cyber physics system (CPS) [5], service orient architecture (SOA) [6, 7] and cloud computing [8], etc. These traditional design technologies target at the indicators of systems. From the perspective of methodology, these are methods to describe the system from top to bottom using formalized words. All rules and relationships of the target system must be clarified, various exceptions should be foreseen. This task is a tremendous challenge in complex system design, and is usually accomplished with experience in real practice. As a result, although these technologies have been successful to some extent, there are still many problems in engineering application, for example, the scope and extent of its application is not clearly defined, thus making it difficult to form consistent and stable technical solution. Meanwhile, the implementation and continuous improvement is hard to be effectively guaranteed, especially for those complex systems that have high requirements for quality. Another problem is that, these techniques are generally bonded with products of different manufacturers and not well compatible with each other, due to the lack of underlying system theory guidance and uniform standards at all levels. Within the organizations that implement this technology usually form a kind of anti-patterns, which is, "vendor

lock-in" [9, 10], reducing the flexibility and sustainable development ability of the system.

From the viewpoint of system theory, the traditional design method focuses on the phenomenon (indicators) of the system, while neglecting the causes (mechanism) of the phenomena. In response to this problem, the Multi-Living Agents (MLA) theory developed in recent years [11, 12] has introduced a series of concepts into the architecture design of complex information systems, such as "living", "living agent" and "living self-organization mechanism". It is an effective methodology for the architecture design of large complex information systems. Based on this methodology, the hierarchical architecture design of TT&C information system can fully resolve the complexity, and avoid the risk of relying on experience, which enables the system architecture design method to be inherited stably within the development organization, and realize continuous development and innovation following consistent basic principles and a clear logic system, together with more robust implementation in practice.

This paper is arranged as follows. In Sect. 4.2, we analyze the complexity characteristics of our country's TT&C network. According to the information system with characteristics of the complexity, we derive the important methodology (MLA information system theory) in the system architecture design and construction in Sect. 4.3. In Sect. 4.4 the design method and steps of MLA are presented in the TT&C networks. In Sect. 4.5 we give conclusions.

4.2 The Characteristics of Chinese TT&C Networks

If we view our country's TT&C networks as a huge information system, the following important features are very striking. Firstly, the system is under strict constraint. This TT&C network plays an important role in national economy and people's life. Besides the demand for stronger functional ability, people put forward more stringent requirement than the traditional information system for a list of basic system qualities, including stability, reliability, security, and extendibility.

Secondly, what contradicts the strict constraint conditions is that, the TT&C networks are often in strong countermeasures environment, and these countermeasures are of general scope. First of all, the physical environment in which the system is operated has become much more challenging. The system not only runs on wide area network with complex topology, but also supports the open platform of some subnetwork in the internet. The nodes of the system are also deployed to general area, and expanded to no-man district, bad climate area, and space etc. The system users not only include cooperative users, but also non-friendly users and even hostile users. The core system also needs to consider safety in natural disasters, wars and the extreme conditions. With the complexity, changeability, and high uncertainty, the generalized countermeasures environment makes it difficult to realize strict constraint conditions of the system.

Thirdly, the system needs dynamic strategy coordination of a variety of functions. As the types and quantity of spacecraft increase, the diversity and

complexity of TT&C task are raised. The TT&C network system needs to provide comprehensive, long loop and wide range services, and implement the task in a fast response, flexible and efficient manner. The system needs to support a wide range and large quantity of business functions, most of which are not simply paralleled, or independent. Rather, they require resource sharing, coordination and collaboration among the subsystems, and even dispatched and balanced through complex business logic or artificial intelligence. From the physical point of view, the complexity of the system lies in the integration of various heterogeneous information subsystems, which makes the structure and components (such as system of TT&C software and hardware resources or equipment) of the system become complicated. There is no doubt that the complexity of the system brings further challenges to the strict constraint conditions.

Fourthly, the system needs to be considered as a subsystem embedded in a bigger system. This feature can be considered from two aspects. First, in the function level, although this system itself is already a large information system and its operations process has certain relative independence, however, it is still a subsystem of the whole space system, very much influenced by the development of the whole system. We need to consider this trait when designing the system architecture. Second, in the technical implementation level, the system is built on various heterogeneous technologies and many kinds of working mechanisms. Under the development of modern information technology, these techniques and mechanisms are greatly influenced by related ones. In order to provide sufficient and strong system performance, the realization technology of the system must be inspected by embedding it to a bigger context of technology and theory when we design and construct the system. Overall, this reflects the high demand for the scalability and flexibility of the system, in order to improve the system's business support ability, business development ability and business optimized combination ability, etc.

The above four characteristics constitute the system complexity of the TT&C networks. These put forward high requirements for the design of the system architecture. It is difficult to overcome these complexity problems if only to design the system following the traditional ways which focus on the indicators of the system and try to give a complete top-down description of the system in advance. Therefore, it is necessary to understand deeply the core working principles and basic mechanisms of complex systems, refine the innate characters of a good system, and design the system architecture with the guidance of scientific and effective methodology. The MLA complex information system theory is presented to this objective.

4.3 MLA Theory

MLA complex information system theory is proposed in recent years, aiming at the design of systems with the above four characteristics. From a dynamic evolution perspective, the most fundamental evolution theory and performance of

the artificial system is consistent with the physical system in principle. i.e., both systems are controlled by the basic principle of dissipative self-organization structure. The systems are of the open and dissipative structure, the fluctuation and far from the balance state, and will lead to the ordered system or evolution finally. Therefore, these basic principles of the system provide important guidance to the design of complex information system.

MLA complex information system theory sets up a theoretical framework, which is built on the essential dynamic characteristics of the system. This artificial system design methodology is derived from the existence, operation and evolution of the system's fundamental dynamic mechanisms. This method is suitable for complex system architecture design. The MLA theory shows that there are a great many variables affecting the dynamic rules of system and it is impossible to resolve all these factors, however in the design of artificial systems, we can start from the design objective, and focus on the adjustment and retention of the "living" at all levels of the living agent system, in order to better meet the needs for functionality and performance of the systems. The MLA's theoretical points are as follows. Firstly, a system constitutes with agents with various levels. The agents and their self-organization mechanism provide the functions of the system, together with the six-element dynamic model of the agent. Secondly, the agent itself and the sell-organization mechanism both have the "living" characters, which are the essence of system evolution. The theory presents the two-set model of the self-organization mechanism, and three-set model of the system agent structure. Thirdly, the retention and adjustment of the living at all levels is the key design objective. The theory gives the general structure and adjusting method of "living". In fact, many common information system architecture theory and system construction idea are all reflections of MLA theory from different viewpoints and at different levels, such as the previously mentioned cyber physical system, the service oriented architecture, cloud computing, and so on. Therefore, designing the system structure of TT&C information system based on MLA complex information system theory can provide more general, more profound methodology guidance, and has important practical implication to guarantee the design quality of the system.

4.4 The TT&C Networks System Architecture Design by MLA Theory

MLA information system theory converts the indicators based design into dynamic mechanism based design. The system is divided into different levels and profiles, which can be indicated with agents or sub-agents. Meanwhile, four levels living retention mechanism (LRM) is introduced, which are, the LRM of agent itself, LRM between the normal agents, LRM between the normal agents and management-control agents, LRM between management-control agents and the system

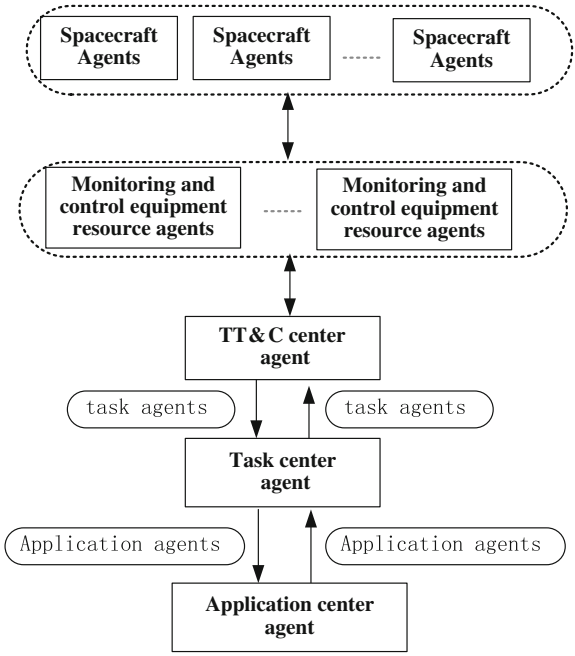

Fig. 4.1 Agents distribution of the system

administrators. The basic design goal of system architecture is to achieve better system "living", and realize retention and maintenance of the "living". "Living" is a concept that reflects the basic characteristics of system survival and evolution. It is the fundamental and uniform objective in the design of system architecture based on the framework of MLA, which can be quantified under certain circumstances. It can link the basic dynamic characteristics of the system evolution with the specific functions of the system, overcome some of the static, fixed design shortcomings in traditional design method, and unify the functionality with the quality factors such as stability, flexibility, thus to provide explicit direction and clear improvement strategy for the system architecture design.

According to this idea, we should first divide the system into different levels of agents, and design the agents based on the "living", to realize functions such as living retention and enhancement, etc. Limited by the length, this paper only discusses the agents division in large granularity. We design the application center, tasks center, and TT&C center in the TT&C networks as living agents respectively, and uses their own functions and the self-organization mechanism between each other to ensure the system functionality. Moreover, we consider the information transmitted between the physical agents as mobile living agents, in order to enhance the adaptability and flexibility of the system.

As shown in Fig. 4.1, application center agent produces application agents, which are in turn delivered to the task center agent. After the task center agent has

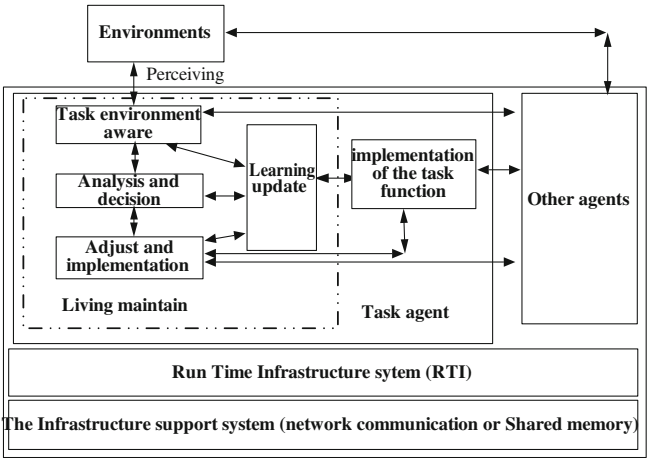

Fig. 4.2 Basic principle diagram of agents

processed the application agents, task agents are then formed and assigned to TT&C center agents, which dispatch measurement and control equipment resource agents to complete the task. In this process, the measurement and control equipment agents are transparent to the tasks center agents. The task center agent is responsible for processing logic levels of the task, such as task decomposition, sorting, etc. It is not concerned with the execution of task on the physical level. The TT&C center agent is responsible for this level.

Each agent is designed with LRM, which is introduced following the four levels mentioned above. The principle of agent structure is shown as Fig. 4.2.

Since the system is built on the living agents, different living measurement can be implemented with different living agents. Using the autonomy of agents, it is convenient to change different agents in order to realize different strategies of living, which improves the flexibility of system.

4.5 Conclusion

This paper analyzes the characteristics of Chinese TT&C networks system, such as the strong countermeasures, strict constraint of the system. It points out that the MLA theory is a powerful tool to design such complex information system. With this methodology, the objective of system design is concerned not only with the indicators of the system, i.e., the performance of the system, but also the operation mechanism of the system, i.e., the cause of the system performance. MLA theory uses "living" to describe the different levels capability in fulfilling the task during system operation. The design objective of system is reflected through the living agents. The capability of the entire TT&C system is described by the "living" of different level agents or subagents, as well as the system capability of overcoming

strict constraint, strong countermeasures conditions, obtaining high system qualities.

This paper presents preliminary analysis on the above subject, briefly explains the necessity and importance of the MLA theory in the architecture design of Chinese TT&C network system. The specific application result presented in this paper is just a preliminary overview of the rough architecture design, which aims at explaining main ideas of the new design methodology. Regarding the further detailed design, such as the division of more fine-grained living agents, and LRM of the static and dynamic characteristics in the different service behaviors of space TT&C network, will be explored in the following researches.

References

1. Qu W, Jia X (2010) The technical status and development of our country's aerospace TT&C networks. Sci Tech Inf 14:481–483
2. Yu ZJ (2008) System engineering of space tracking, telemetry and control networks. National defense industry press, Beijing
3. Jiang YQ, Feng YX, Liu ZG et al (2003) The model of network management for satellite information network. J Northeast Univ (Nat Sci Ed) 24(1):15–18
4. Jing JJ (2011) Current situation experience and trend analysis of the informationization and industrialization. China Inf Times 10:9–11
5. Lin F, Shu SL (2010) A review of cyber-physical systems. J Tongji Univ (Nat Sci Ed) 8:1243–1248
6. Wu Z, Li HQ (2008) The design and realization of distributed SOA architecture. Microcomput Inf 36:253–255
7. Votis K, Alexakos C, Vassiliadis B et al (2008) An ontologically principled service-oriented architecture for managing distributed e-government nodes. Netw Comput Appl 31:131–148
8. Chen K, Zheng WM (2009) Cloud computing: system example and research status. J Softw 20(5):1337–1348
9. Chen Y (2003) A new trend in complexity studies-agent based modeling and its implication. J Syst Dialect 1(11):45–50
10. William JB, Raphael CM, Hayds WM et al (1998) Anti-patterns refractor software, structure and project in crisis. Wiley, New York
11. Wang Y, Tao R, Li BZ (2008) Using the multi-living agent concept to investigate complex information systems. Sci China Ser F-Inf 52:1–17
12. Wang Y, Tao R, Zhang H (2010) Distributed intrusion detection system constructing analysis based on living agents. Sci China Ser F-Inf 40:613–623

Chapter 5
Discussion on Integration Management of TT&C Information

Yunsheng Hao and Linfeng Shang

Abstract To support more and more satellite and launch vehicle missions, tracking, telemetry and command (TT&C) system is challenged by requirements of more equipments operation, more complex system management and more intelligent decision-making. With the development of information technology (IT), promoting IT usage in TT&C system is the only way to meet the challenge. Based on the views of IT evolution stages, information integration issues, combined with TT&C system, this paper proposed a TT&C information integration management model and a corresponding application system, which both have three levels, a base-level with automation as subject, a middle-level with process management as subject and a high-level with intelligence as subject. Comprehensive discussion shows that the information integration stage in IT evolution is the key development stage for TT&C system. To improve information integration efficiency and ability, the three-layer model is an instructive model for practical application.

Keywords Integration management · Application system · Automation · Process management · TT&C

5.1 Introduction

Space exploration is one of the main themes in the twenty-first century. To support more and more satellite and launch vehicle missions, a comprehensive network of ground sensor stations are deployed to provide TT&C services, correspondingly, massive information is generated, including target kinematic measurements and sensors auto-detection data. With the rapid development of communication and

Y. Hao (✉) · L. Shang
Taiyuan Satellite Launch Center, Taiyuan 030027, China
e-mail: hys68@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_5,

computer techniques, isolated information could be transferred to an integration central station, stored, processed, managed there, and the central computer systems not only monitor sensors working status, but also make data analysis, output target trajectory estimation, and provide visualized information to support decision-making. TT&C system with all operators and commanders form a more sophisticated command, control, communication and intelligence (C^3I) system. Such a system is a key part in any launch mission, and is expected to be organized, and support high level commanders to make decisions more accurately, efficiently. This becomes information management's ultimate goal.

In fact, Information emerged with the emergence of human kind, IT is popularized with human progress and development, and IT evolution is becoming a mainstream trend. With regard to this mainstream, a point of view is worth noting: "From appearance, a mainstream generally manifests as result of some technical factors or ideological change, but fundamentally, the nature of problem is related to the structure changes of the society and economy, only that, nothing else." [1] IT evolution should focus on "evolution", "evolution" means a gradual process in which something changes into a different and usually more complex or better form, "IT evolution" refers to the process of integrating information and IT into production and life of contemporary human society.

As a process, evolution of information technology has staging characteristics. Influenced by Marx's social staging development theory, a six-stage model was developed by Richard L. Nolan during the 1970s [2] to account for the evolution of IT. Nolan's model proposed that IT evolved over time along six stages: initiation, contagion, control, integration, management, maturity. Over the years, the TT&C information management evolves exactly along such a path due to technological development, until recently it runs into a crux point. Mischel realized that in fact the integration stage and management stage could not be separated apart, and proposed another model, which had four stages, five indexes [3]. Both models are evolving along time dimension. They just tell us what stage we are in, and which direction we should go, but do not tell us how we should do just now. From the two models' perspectives and the level of IT application, the TT&C system for launch missions generally comes to a stage that requires information integration.

Information integration or fusion is a ubiquitous phenomenon naturally existing in the cognition process. Even animals can gradually enhance their information fusion abilities during their evolutions courses, and higher or lower animals are only different by degree in information fusion abilities. As a natural system, information integration is a creative fusion process. Useful information is selected and optimized on purpose, and be fused with others optimally for full play to enhance the overall functions of whole body or generate new functions. Similarly, TT&C information integration system as a manmade system fuses all the sensors information creatively, selectively, consciously, purposely, by optimal integration method to maximize decision accuracy.

How to look on TT&C information management level, how to operate such a complex system smoothly, what direction the information system should go to enhance the ability to accomplish any future launch mission, are the main topics discussed in this article.

5.2 TT&C Information Integration Management Application System

In this new era, TT&C system is challenged by new development requirements in TT&C operation and management, and ability promotion. Isolated "chimney" phenomenon in traditional TT&C system information management is increasingly unable to meet IT development requirements, and information system integration becomes the trend. Future TT&C system, the technical level's integration should be the foundation, process management level's integration should be the guarantee, and command service level of integration should be the guide.

Information integration from base technical layer to middle process layer and top service layer, shows that TT&C is not a pure technical system, but a complex human–computer interacting system. The efficiency of system depends on human–computer relationship and the similarity between model and real world. The TT&C system in reality is a hierarchical system. Top commanders are on the top layer, their staffs are located in the middle layer, and common operators are on the base layer. Accordingly, TT&C information integration management system should be a three layers "pyramid" like hierarchical information system and an integrated human–computer system which could fuse all the top, middle and bottom layers together, supply functions of management in longitudinal or transverse dimension. In the longitudinal direction, lower layers report to the upper layers the selected and verified information, and work according to the directives from the upper layers; in the transverse direction, different organization sections are responsible for different sub functions of TT&C, but share a common data management system, and operate coordinately with each other to form a whole system.

An application system of TT&C information integrated management has three layers as Fig. 5.1. Please note that the three layers are above an extra physician foundation layer. The foundation is the data exchange platform, which provides a database and information application interface for the top, middle, base layers. The base level mainly serves for TT&C equipment operators, in this layer TT&C equipment information system provides functions of the data acquisition, automation and control. The middle level is mainly for key staffs and senior technicians, and in this layer process management system brings together all kinds of information, and provides functions of organization management, task management, equipment management, training management and quality management. The top level is mainly for top commanders, and in this layer strategic management information system provides functions of decision support.

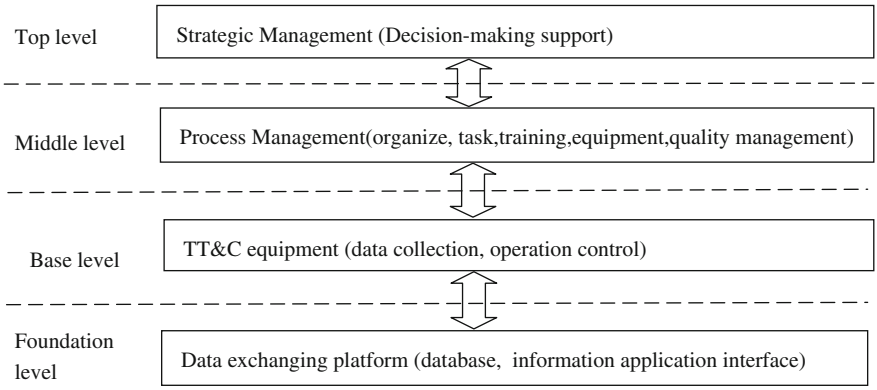

Fig. 5.1 TT&C information integration management application system

5.3 TT&C Information System

According to the theory of management decisions, the management decision problems could be divided into three types: structured and unstructured and semi-structured. The degree of support by information systems for decision-making varies with the types of decisions. In accordance with the actual situation of the TT&C system, structured decision-making is mostly present in the base level, unstructured decision-making is mostly present in the high-level and semi-structured decision-making is mostly present in the middle level. Structured decision-making problems are easier to describe and have more regularity, and information system is easier to program them into corresponding software modules, so base level information system is characterized by automation. The methods to deal with semi-structured decision problems need more flexibility to be adjusted to face application conditions change. Information systems can customize, optimize their processes, ensuring orderly operations under the dynamic changes, so middle information system is characterized with processing management. It is difficult to describe unstructured problems; they general have statistical characteristics, information system cannot be involved directly in the decision-making process, but can provide timely, accurate and relevant knowledge to the senior staff, so high level information system is characterized by intelligence. Of course, automation, process, intelligence not only exist in a certain level of information systems, but also exist in other levels, however, they might have different proportion.

Information system can be likened to a human body. Eyes, ears, nose and skin are all its sensors, brain is the computing center and neural system is the link between sensors and computing center. Many organs have the ability of local automatic reaction. If there occurs a slight cut in skin, skin rehabilitation process will be started immediately, the brain only gets a notification about the cut.

Decision making is more like using all the information coming from the eyes, ears, and nose. Analog to human body, TT&C system could be structured into a "local automatic", "orderly information exchange", "decision-making" model.

5.3.1 Base Automation Level

TT&C equipments are tools for primary operators. Today, information technologies in TT&C equipment are advancing rapidly. With the development of software, radio technology, optical imaging technology and signal processing technology, TT&C equipment develops to a greater extent in digitizing, integrating and embedded software, all kinds of knowledge embedded in the device, especially in data acquisition and operating, embarked on automation, including the automatic detection, automatic test, automatic calibration, equipment monitoring, and fault diagnosis. In auto detection, many event monitoring points (EMPs) are set up in different types of equipment. Those EMPs could be set into the minimal unit in every equipment subsystem, so the embedded software could automatically determine the system's health, report the important information to the data center. In automated test, the test system consists of test computer, test equipment, test software, test on key indicators according to the related specification. In automatic calibration, calibration equipment creates analog test signal or benchmark, check the system functionality and performance, and help to calibrate distance, angle parameters. In equipment monitoring, there are two general sub layers: system monitoring and subsystem controls. They both are responsible for system-wide status monitoring, parameter setting, task configuration and action scheduling. In fault diagnosis, various types of diagnostic knowledge are used to analyze the monitoring information, to make judgment and reasoning about device status. Once the device is found abnormal, operators can determine the cause of the malfunction and location, and give the interpretation of diagnostic reasoning process of the fault and its corresponding treatment.

5.3.2 Middle Process Management Level

In TT&C information integrated management system, from the vertical perspective, the middle layer is the link between top layer and base layer. It will interpret high-level decisions and directives into operating specifications, and entrust the base layer to implement, as well as to collect and process all kinds of information, and report to top layer. From the horizontal perspective, the middle layer is responsible for the entire main line of work; its scope of work covers the whole process of scientific research and experiments. So, it is obvious that the middle layer work is pivotal, and the operation management order in middle layer is extremely important.

Middle-level information systems can achieve process management by information integration, technology and operation management mechanism, embedding a variety of knowledge into the work process. First of all, the process management can effectively reduce the error probability and improve quality. The work of TT&C system is mostly a repetitive process; it is easy to implement standardization, to use software instead of hardware and to be divided into sub processes. Process management can promote each person's work in accordance with the required specifications, so error probability can be reduced, each process, behavior of each person and working equipments status can become visible, the problems are easy to be discovered and corrected, and the work quality can be improved. Secondly, the process management can promote horizontal integration between various functional departments to obtain efficiency. Process management can link up all different functions done by different people from different sections. For TT&C systems, requirements analysis, planning and construction, research and experiment, evaluation of results, integration, organization and management, task management, training management, equipment management, and quality management are all integrated through the strengthening of collaboration between functional departments to improve efficiency. Associated with it, just like the relationship between the productive forces and production relations, process management requires that the organizational structure in the middle level should have less sub levels and be more flatten. Third, the process management can adapt to various internal and external changes, and have more flexibility. The diversity of launch missions requires TT&C equipments have flexibility in deployment and network, and the "zero defect" demand of quality management requires TC equipments to keep technique improving. As a system with process management, TT&C can have a greater ability to keep order while adapting to internal and external changes by using of standardized interfaces and software realization.

Another characteristic of the middle-level information systems is the application of geographic information systems. Whether the flight of manned spacecraft or deployment of TT&C equipments, work related with TT&C has obvious space property, and has a very close relationship with geography. As a common language and accepted technology which is understood and used by everyone, geographic information systems (GIS) have the abilities of spatial information integration, analysis and visualization. In the past, it is a professional research tool to deal with academic issues, and now it is becoming a popular, easy-to-use information management tool. Based on GIS, TT&C system can clearly describe and show the target flight intuitively, help to make TT&C scheme for specific launch mission, and provide support to efficient decision-making. Further more, the customized applications of GIS could be developed on the basis of the general GIS platform, and adding process customizing, process operation, process optimization functions to original system become the main development direction, and using GIS in process management should be a pattern of TT&C information system development.

5.3.3 Top Intelligence Level

In the top layer of the TT&C information integrated management system, the strategic decision-making problems the senior staffs think about include long-term, medium-, short-term goals and the allocation of resources. The required information for decision-making cannot be clearly defined, decision-making practices vary from person to person, and the impact of decision-making cannot be accurately calculated, therefore, the high-level information systems cannot be a universal standard system, but only a decision support system. It must be suitable for the decision-makers' personal thinking habits, decision-making requirements, and provide relevant knowledge. Decision support system uses database, model base, method base, knowledge base, and human–computer interface and graphics components, to help the decision makers to solve semi-structured and unstructured decision-making problems. Its main functions consist of three aspects: First, performing information collection and management, including internal, external information and feedback information from actions of all decisions; Second, performing knowledge creation and management, including construction, storage of related models and methods, and making data analysis, making forecast based on the models and methods, and generating and managing knowledge based on the results; Third, establishing kindly human–computer interface, which provides convenient human–computer interaction and proper response speed.

It should be noted that IT evolution cannot be done overnight, it is a long process of development. It is necessary to grasp the characteristics of each development stage like decision support system, which is generally divided into four stages including data integration, data analysis, data decision-making, knowledge management. In the first stage, based on data integration, information management should perform the integration of workflow, and enable managers to grasp the system operation situation. In the second stage, through integrated data and process analysis, information management should obtain more useful references. In the third stage, information management should differentiate different types of decisions, and separate different functions of the decision-making, the structured decision-making problems can be solved by computer processing entirely, and semi-structured and unstructured decision-making problems can be solved by computer processing partially. In the fourth stage, information management should establish decision support system based on knowledge management, by data accumulation and data mining, build a knowledge base to provide decision support to higher level. To promote IT application, one must first find out what stage it is in now, then do research on what future stage looks like.

5.4 Conclusion

TT&C information integration management can be modeled by a three-layer pyramid, which has technical layer, process layer, service layer. This model is located between the fourth and fifth stages in time line of Nolan model. Application system shows that the above solid model could help us improve our management efficiency of TT&C information, and give us a more practical direction to future development.

References

1. Fan YS, Hu YG (2007) Strategic planning methods and practice of enterprise information. Publishing House of Electronics Industry, Beijing
2. Fu XL (2006) Strategic planning methods and practice of enterprise information. Beijing University of Posts and Telecommunications Press, Beijing
3. Xue HC (2012) Management information system. Tsinghua University Press, Beijing

Chapter 6
Informatization Maintenance of TT&C System Based on CBM

Meng Ren, Weijing Zhou, Jianhua Guo and Zhongkai Guo

Abstract In view of the complexity of TT&C equipment and the difficulty of maintenance, a CBM construction being applicable to TT&C system was established according to the technology of condition-based maintenance and the OSA-CBM standard. By the study of state information management, performance evaluation, fault diagnosis and other key technologies, the condition-based maintenance functions for TT&C system were implemented initially. In the field of space TT&C, a new maintenance mode was established. It can meet the needs of the developing space launch missions and the management of on-orbit satellites.

Keywords Condition-based maintenance · Performance evaluation and prediction · Fault diagnosis

6.1 Introduction

With constantly updated technologies and concepts, the way of TT&C system maintenance can be divided into two stages [1]:

STAGE 1: Breakdown Maintenance (BM). It is also called fault-based maintenance which refers to the maintenance after equipment failure or significant decline in equipment performance. This way of maintenance can reduce the unnecessary waste of manpower and maintenance costs, and avoid excess maintenance under the conditions of fewer TT&C tasks and simple equipment function in early period. However, due to the fault influence on the related equipment or the

M. Ren (✉) · W. Zhou · J. Guo · Z. Guo
Sanya TT&C Station, Xi'an Satellite Control Center, Sanya 572000, China
e-mail: renmeng041647@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_6,

whole system and even on the success of TT&C mission, breakdown maintenance can not guarantee trouble-free operation of the equipment and has been rarely used.

STAGE 2: Preventive Maintenance (PM). It is one to prevent interruption caused by the sudden failure or out-of-run of the equipment. The maintenance based on equipment's usage and service time is most widely used in preventive maintenance. The time interval between preventive maintenance depends on the size or life of the equipment, and can be set for half a year, a month or a week. But as to the actual work of TT&C stations, regular preventive maintenance has the following drawbacks:

1. Under present circumstances, the complexity and large scale of equipment in TT&C station makes it difficult to carry out the statistics of the life of components and systems, so maintenance period in a sense is blind, which always leads to over-maintenance or under-maintenance and can not achieve the desired effect.
2. As TT&C tasks become increasingly intensive, some of the equipment is in service all the year round, which affects the implementation of the regular maintenance.
3. Repeated removal and testing may lead to a decline in equipment performance, resulting in unnecessary losses.

Therefore, to meet the needs of regular space launch and satellites management, it is necessary to explore a maintenance mode to adapt to the current development trends in the field of space TT&C.

6.2 CBM: Condition-Based Maintenance

With the rapid development of technology of state monitoring, fault diagnosis, and decision-making, the technology of Condition-based Maintenance (CBM) on the basis of condition monitoring and fault diagnosis appeared [2]. Its working principle is to get condition information from the sensors implanted in the equipment or external test equipment when the system is running, and ultimately identify the needs of equipment maintenance by real-time or periodical evaluation of the condition information. Due to its ability to overcome the drawbacks of BM and PM, it attracts widespread attention.

To promote the development and versatility of CBM, ISO developed a series of standards on CBM. According to CBM Open System Architecture (OSA-CBM), CBM system consists of seven layers [3] (Fig. 6.1):

1. Data Acquisition Layer: to collect equipment running data, and provide services to the upper layer.
2. Data Manipulation Layer: to preprocess and calculate the data.

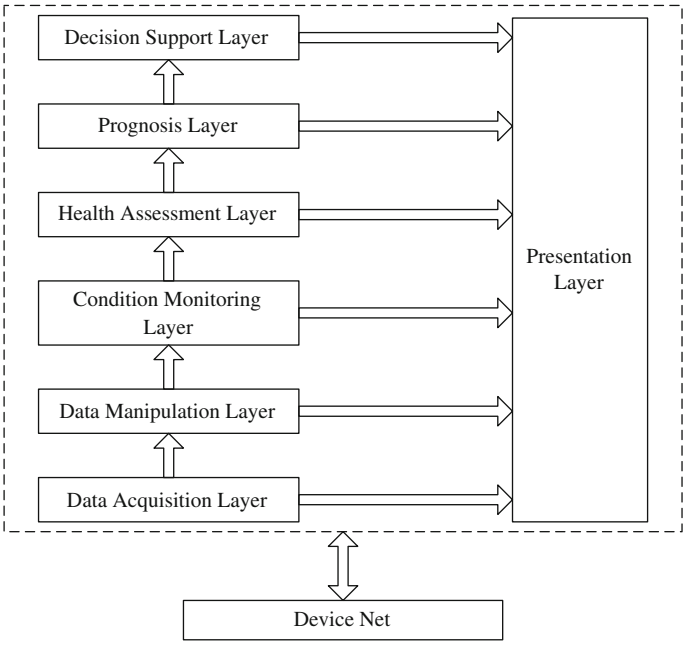

Fig. 6.1 Constitutional diagram of the OSA-CBM system

3. Condition Monitoring layer: to get the condition deviation by comparing the output data from the lower layer with its corresponding system standards, and realize alarm function if necessary.
4. Health Assessment layer: to assess the performance of system, subsystems or component, taking all factors into account.
5. Prognosis Layer: to predict the future performance of equipments, and infer its effective work time RUL.
6. Decision Support Layer: to provide maintenance strategy and the corresponding operation command. Considering the work of equipment, resources constraints, and other practical factors, The OSA-CBM has no clear requirement on this layer.
7. Presentation Layer: to display human–machine interaction interface and various information.

Based on current TT&C system architecture, this paper mainly analyzes the applicability of CBM in TT&C equipment maintenance, studies the corresponding technology CBM required and achieves a preliminary function in monitoring as well as evaluation on equipment condition, so as to provide decision support for current TT&C system maintenance.

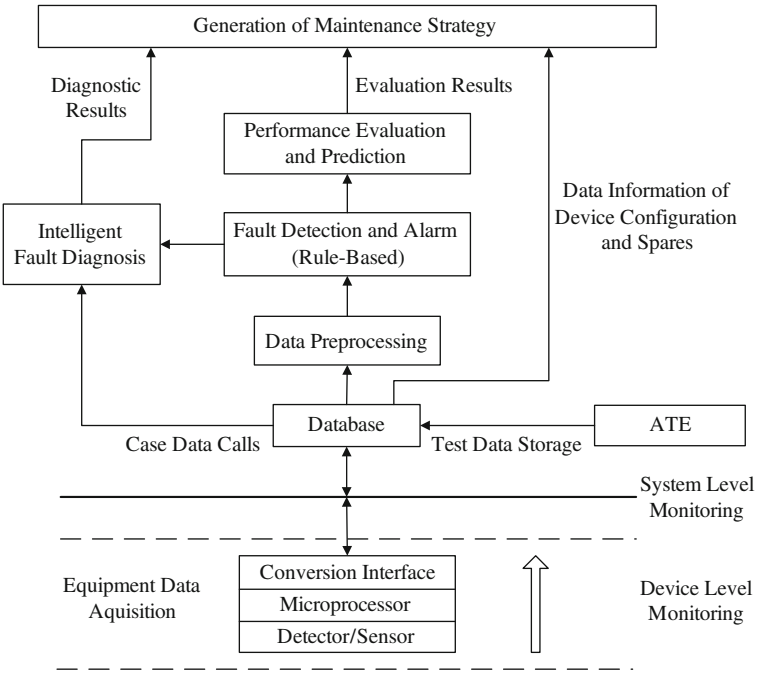

Fig. 6.2 Maintenance and decision theory of TT&C system based on CBM technology

6.3 Condition-Based Maintenance of TT&C System

According to the current TT&C system architecture, by the study of information management, performance evaluation, and fault diagnosis of the TT&C system, the CBM system suitable for TT&C equipment is designed according to the OSA-CBM. Its working principle is shown in Fig. 6.2.

6.3.1 Condition Information Management

Informatization is the general term of communication modernization, computerization and rationalization. That is to obtain sufficient information required of the application with the help of network and communication and make corresponding decisions. The implementation of CBM system in TT&C equipment relies on information technology and equipment to get adequate condition information, and the CBM system can achieve a sound judgment of system condition and fault information.

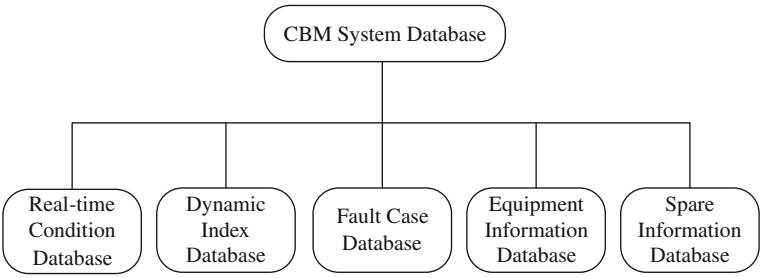

Fig. 6.3 CBM system database components

6.3.1.1 Condition Information Acquisition

Network-Based Monitoring of Equipment

Nowadays, network-based monitoring is widely used for TT&C system. Device level monitoring is by implanted detector/sensor to capture the condition of the equipment and make an initial processing by ARM microprocessor or other control devices; system level monitoring is by the monitoring network to make a unified management and achieve a centralized monitoring and processing of the condition. Non-network interface devices are converted by industrial interface converter to connect the monitoring network. Long-term operation of TT&C system suggests that the current monitoring network has a reliable capacity for data collection from the lower layer.

What's more, the data collection scope can be expanded by increasing the type and number of sensors. For example, we can use the temperature and vibration sensors to enhance the monitoring and analysis of the antenna motor operating condition.

Online Testing of TT&C System

The implanted detector/sensor are always unable to access some of the macro condition information, such as system performance indexes. To solve this problem, CBM allows the system to collect condition data by the external testing equipment. For TT&C system, Automatic test technology in developing can be applied to get the macro-performance data online by external testing instrument to ensure its normal operation and avoid unnecessary disassembly.

6.3.1.2 Database Design

According to CBM system, database design is composed as follows (Fig. 6.3):

Real-time Condition Database

Real-time condition database is set for storing the condition data getting from the lower layer when the equipment is in operation. Currently, TT&C system already has a proper and reasonable condition data base, and it can be considered to inherit and extend.

Dynamic Index Database

Dynamic index database is used for storing macro-performance data obtained through online testing. It should be as comprehensive as possible to fully characterize the macroscopic properties of the system condition. Taking the needs of assessment into account, each data table should include test conditions, the combination of equipment, test result, standard, and so on.

Fault Case Database

Fault case database is used for recording and storing the features extracted from faults when the equipment is in operation. In order to cite properly, the typical cases are stored in the form of vector in the database. Data structure of case vector is preliminary designed as follows: Fault Case = {Fault serial number, < fault symptoms1, weight > ··· < fault symptoms n, weight >, treatment results and so on}.

Equipment Information Database

The equipment information database is used to store the basic information of component type, technique condition, and quantity of all components relevant as well as a version of the software configuration. The database can be a reference for maintenance.

Spare Information Database

Spare information database is used to store the basic information of type, technique condition, and quantity of system spares. It provides the information basis for equipment maintenance and spare purchase.

6.3.2 System Fault Diagnosis

Fault diagnosis is to make judgments for system operation condition and exceptions so as to provide related strategy for the system recovery. For large and complex systems such as TT&C equipment, a single fault diagnosis method can not achieve good diagnostic results. With existing technology, the expert system based on two-level reasoning is employed to meet the needs of fault diagnosis in TT&C system.

6.3.2.1 Rule-Based Reasoning

TT&C system has the function of fault diagnosis and inspection to some extent. The monitoring and control subsystem checks the status of all equipments and gets fault location according to the preset fault model, and finally gives a fault alarm. Its main principle is based on RBR fault detection. In the anterior level of the diagnosis, rule-based diagnosis can be considered and the fault diagnosis capability of it can be further enhanced by improving the TT&C system rule base.

For the improvement of the rule base, both micro and macro level should be taken into account. On micro level, it is mainly to expand the diagnostic scope of the equipment on the basis of system monitoring. On the macro level, we should focus on making analysis and judgment of overall performance of TT&C system. With more and more application of automation technology, warning on automation process should be strengthened especially.

6.3.2.2 Case-Based Reasoning

Fault location of TT&C system is mainly completed by the monitoring and control subsystem in accordance with the corresponding fault model. This method is more accurate in device fault location but not in more complex situation. To solve this problem, the method of Case-Based Reasoning (CBR) can be considered. CBR uses the grey correlation theory to get the similarity between the fault feature vectors and case vector, and achieves assisted fault location for anterior diagnosis [4].

6.3.3 Performance Evaluation and Prediction

The purpose of performance evaluation and prediction is to make a scientific assessment of equipment performance and a certain degree prediction of the equipment condition. Regarding the complexity of TT&C system functions and many factors involved in assessment, the Analytic Hierarchy Process (AHP) can be adopted to ensure the comprehensiveness and practicality of the assessment. By

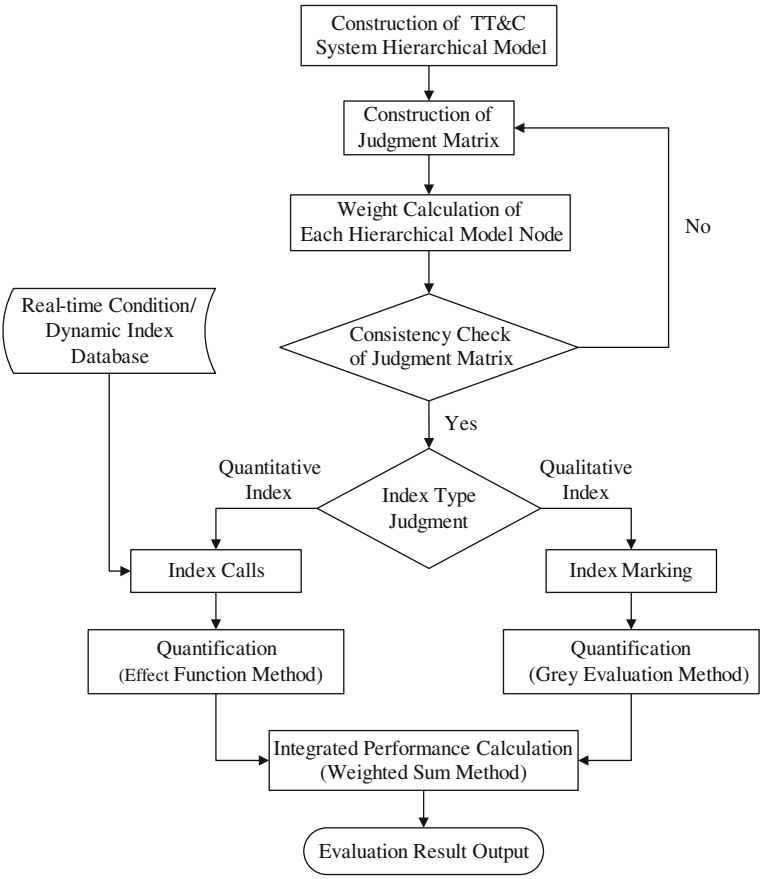

Fig. 6.4 Process of system performance evaluation

calling the corresponding data in the database, it makes integrated performance evaluation of system and components, and feeds back the condition information timely and accurately for the equipment managers.

Application of AHP for TT&C system performance evaluation can be divided into four steps. The evaluation process is shown in Fig. 6.4.

6.3.3.1 The Construction of TT&C System Hierarchy Model

To evaluate system performance with HAP, the evaluation hierarchy must be established firstly according to the logical relation of system functions. Hierarchical structure can be divided into three layers: target layer, criterion layer and index layer.

6.3.3.2 Weight Calculation of Each Hierarchical Structure Element

Weight is acquired by establishing judgment matrix and solving feature vector of it. To get a judgment matrix, the 1–9 scale method put forward by Satty can be used [5].

6.3.3.3 Index Quantification

The bottom layer indexes of TT&C system can be divided into two types: Quantitative and Qualitative index. They are quantified by effect function method and grey evaluation method separately.

6.3.3.4 Integrated Performance Calculation of TT&C System

According to quantitative results and corresponding weights of the indexes, the comprehensive performance of TTC&C system comes out of hierarchical calculation by linearity weighted sum method.

Based on reliable evaluation, we can make some predictions on the system performance. Taking the uncertain factors during the predictions into consideration, the gray prediction theory can be used [6].

6.4 Improvements and Application Prospect

At present, the functional design of basic database, system performance evaluation and fault diagnosis modules of CBM have been almost completed. Application effect demonstrates that the current technology based upon CBM can properly reflect the system condition and helps the maintenance staff to make decisions. To further enhance the reliability and practicability of the CBM system, the following aspects should be improved and perfected in the future:

1. Gradual completion of the condition information. For the imperfection of the monitoring points on equipment level and the irrationality of hardware and software interface design, there are still a lot of blind monitoring points for the whole TT&C system. The imperfection of condition information influences the reliability of the CBM system to some degree. The CBM system is an engineering system so that the condition information acquisition should be taken into consideration at the design stage.
2. The reasonability and practicability of assessment model, prediction model and diagnostic model deserve deep research.

3. Intelligent maintenance strategy. Based on the practicability, decision-making model that matching the TT&C system should be adopted to realize the optimization and intelligence of maintenance strategy.
4. The generality of the CBM system. For the universal existence of problems like various types and different functions of the current TT&C system, the generality of CBM is another question to be studied.
5. Long-distance CBM system based on network [7]. Considering the continuous extension of the space TT&C network scale and the constant development of remote control and automatic technology, the CBM system based on network will become the ultimate goal. The long-distance CBM system on TT&C center level can depend on the long-distance monitoring and control network. Long-distance CBM system will make it scientific and reasonable for the TT&C center to draw up work plan and maintenance schedule of the TT&C resources in the monitoring and control network.

6.5 Conclusion

For the maintenance problems of current TT&C system, the CBM technology is applied so as to enrich the monitoring and maintenance technology and make the performance condition of the system more transparent. Experiments on a type of TT&C system show that the application of the CBM technology can make the generation of maintenance strategy more scientific, flexible and efficient. It can offer a technological base for the TT&C system maintenance. With this advanced technical idea, the CBA system established above has an extensive reference value and promising applications.

References

1. Chang XY, Gao JJ (2008) Research on information and intelligence of equipment safety and maintenance. Guizhou Electr Power Technol (5):47–50
2. He JQ, Wang B (2009) Research on condition based maintenance of military equipment. Ship Electron Eng 29(12):42–44
3. Xu AD, Yu HB, Guo QJ (2005) Condition-based maintenance of equipment—research on CBM technology. Constr Mach Equip (6):9–13
4. Wang HY, Ni ZW, Yan J et al (2010) Research on application of gray-relational theory in CBR. Comput Technol Dev 20(5):96–99
5. Lu ZY, Fan JH (2011) Effective evaluation of tactical communications network based on AHP. Mod Electr Tech 34(1):57–60
6. He BG, Liu FT (2008) Gray prediction method of the reliability of electronic equipment in missile. Ship Electr Eng 28(3):160–161, 172
7. Hu JT, Xu AD, Guo QJ (2007) On the architecture of a network-based monitoring and maintenance system for key equipment. Inf Control 36(3):357–363, 370

Chapter 7
Preliminary Research on Management of Crosslinks of Navigation Constellations and Their Security System

Kunmei Cao, Taoming Chen and Bo Wang

Abstract Crosslinks of a navigation constellation are designed for inter-satellite ranging and communication and autonomous navigation so that the constellation can function continuously without the support of globally distributed ground stations. Crosslinks can also realize some other extended applications. As a significant part of a satellite platform, crosslinks can improve the function of a constellation. However, management problems related to crosslinks have to be solved. This paper discusses the tasks of a crosslink management system and proposes a design of its function and topology based on the characteristics of the crosslinks of the navigation constellation. Meanwhile, the crosslink management model and security protection issues are also discussed.

Keywords Navigation constellation · Crosslink · Management mission · Management mode · Security system

7.1 Introduction

It is a significant measure of strategic meaning and military value to set crosslinks in a global navigation satellite system. For example, the crosslinks of GPS [1] have already become important strategic assets of the US. As a type of military aerospace equipment, GPS can realize autonomous navigation for 60 days, greatly enhancing the survivability and availability of the satellite system during wartime through use of the crosslinks. Crosslinks can be used to modify the geometric relationship of observed orbit determination parameters, improve user positioning

K. Cao (✉) · T. Chen · B. Wang
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: mayan888@sina.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_7,

precision, update navigation information more quickly, and ensure the accuracy of star calendar forecast. Crosslinks can also be used to quickly obtain monitoring results of nuclear explosions from any position in the world to help the US master the global situation at any time. In the global navigation system of our country, crosslinks will be used to realize inter-satellite ranging and communication, and achieve autonomous constellation navigation. As a crosslink subsystem is an important part of a global navigation system, a scientific, optimized and efficient management system should be established to ensure that crosslinks operate normally during different stages of constellation networking, operation in orbit, network replenishment and so on to ensure that fault diagnosis and restoration of the crosslinks is timely and effective, to ensure that crosslink in-orbit testing is successful and the crosslink resources are allocated rationally.

Crosslink management system relates to the customers of constellation and crosslink resources, so a full-fledged security system should be established to avoid insecure elements brought into crosslink subsystem. The security system should be designed in an integral manner to make sure that the global navigation system could operate reliably. In this paper, we have a preliminary discussion on the crosslink management mode and its security-system so as to inspire more meaningful research.

7.2 Crosslink Mission Management

Crosslink management system integrates autonomous navigation, satellite-ground joint orbit determination, calendar uploading and platform control. Operation mode should ensure high reliability, excellent survivability and scalability. Consequently, it is necessary to design some auxiliary mechanism in order to meet various demands of equipment maintenance and in-orbit upgrade, from testing and validation, networking, in-orbit operation to network replenishment.

At the stage of verification test, means can be provided to verify autonomous navigation and satellite-ground joint orbit determination. In the networking process, it can offer some service functions when there are not enough satellites. During the in-orbit operation process, it is able to fully support autonomous navigation and satellite-ground joint orbit determination, calendar uploading, and emergency platform control requirements. Even when a new satellite joins the constellation or a satellite goes wrong and quits, the whole constellation will still run normally. In addition, the crosslink management system has functions for entire satellite network maintenance, testing, calibration and upgrade. The tasks of navigation constellation crosslink management can be summarized in the following five aspects:

1. Design of crosslink unit to be borne by individual satellites, in-orbit testing of the crosslinks, in-orbit testing of the crosslinks of repaired satellites and validation for access to the chain (if necessary).

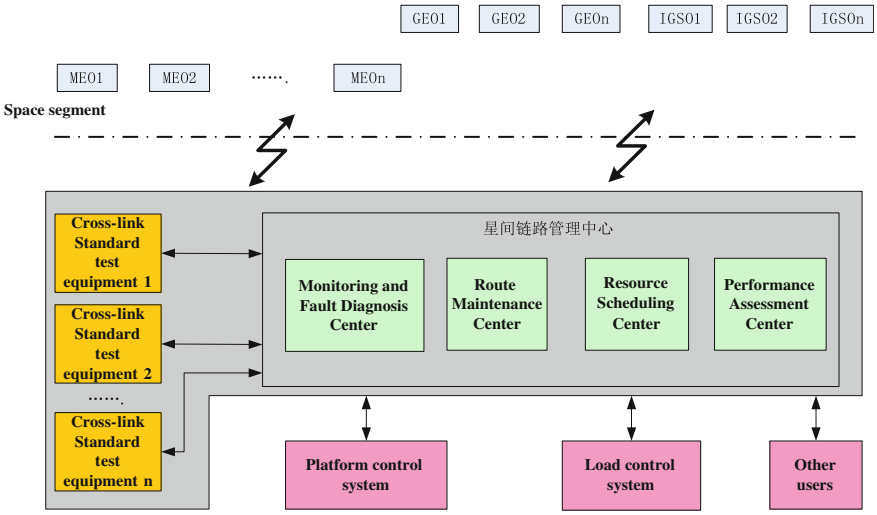

Fig. 7.1 Configuration of the crosslink management center

2. Maintaining the routing table of crosslinks; design, scheduling and calculation of crosslink routing table, uploading to satellites and broadcast to users based on the topological structures of crosslinks at different stages (including networking, operation, partial fault, network replenishment and other stages).
3. Monitoring the operation status of the crosslinks in orbit, and assisting platform control system in fault diagnosis and recovery in orbit.
4. Scheduling the crosslinks for multiple users, settling the conflicts among users, allocation of crosslink resource scientifically and rationally when necessary, monitoring of the usage of the crosslinks, and providing users with timely, effective information about the crosslinks;
5. Measurement of the performance of the crosslinks and optimization of the design.

7.3 Composition and Function

The crosslink management system is made up of standard test equipment and crosslink management center. The functional schematic is shown in Fig. 7.1.

7.3.1 Standard Crosslink Test Equipment

There could be many sets of standard crosslink test equipment (the standard test equipment is ground-based pseudo satellite crosslink equipment, and acts as platform and load controller). They are used for in-orbit testing of space-borne

crosslinks, constellation crosslinks and malfunctioning satellite. They could also be used to test malfunctioning satellite before joining the network again, to receive, demodulate, and to store telemetry information (including crosslink ranging and communication information) transmitted from the constellation to the crosslink management center for monitoring of the operation status of the crosslinks.

7.3.2 Crosslink Management Center

The crosslink management center should have the functions of monitoring and fault diagnosis, route maintenance, resource scheduling and performance evaluation.

The monitoring and fault diagnosis center performs data decryption, storage, processing of the crosslinks, monitoring of operation status of the constellation crosslinks in orbit, performs fault diagnosis of the crosslinks, performs monitoring management of the crosslinks standard test equipment, performs in-orbit testing and experimental data processing.

The route maintenance center accomplishes route design and calculation based on constellation's topological structure, uploads to satellite via the uplink channel of the platform control system or the standard test equipment of the crosslinks, and transmits the topology and routing information to users and enables them to keep synchronization.

If crosslink telecommunication resource is so strained that normal service can not be provided to multiple users, users must send crosslink application to the management center. Then, the resource scheduling center sets the principle of using crosslinks to solve conflicts between users based on user business priorities and the actual operation state of the crosslinks.

The performance assessment center collects crosslink operation status, user utilization and resource scheduling situation to evaluate the performance of the crosslinks and its management, and to optimize satellite-ground design based on actual usage.

7.4 Conception of Operation Mode of Crosslink Management

Crosslink management operation mode can involve monitoring and fault diagnosis, route maintenance, resource scheduling and so on.

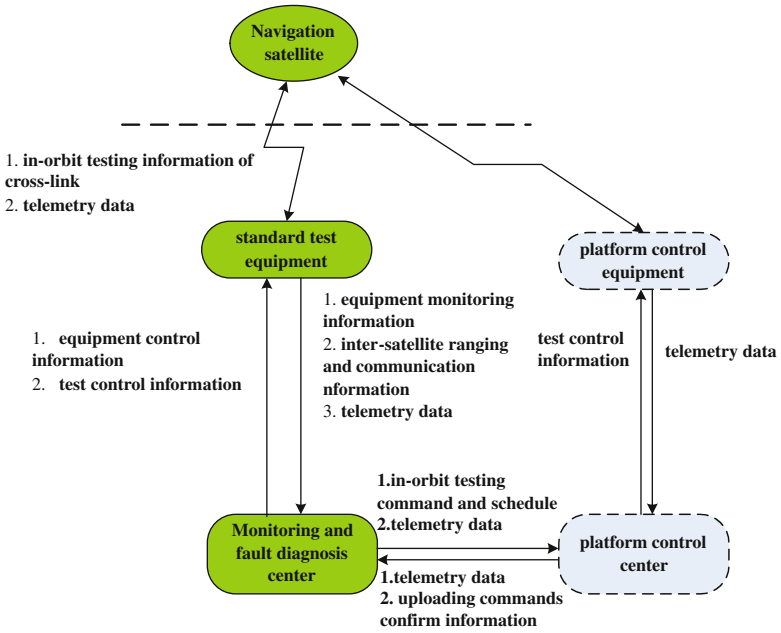

Fig. 7.2 Workflow of the crosslink system during in-orbit testing

7.4.1 Monitoring and Fault Diagnosis Operation Mode

7.4.1.1 In-Orbit Testing Operation Process

With cooperation of the platform control system, the crosslink management system is responsible for in-orbit testing of the crosslink system. The workflow schematic is shown in Fig. 7.2.

The monitoring and fault diagnosis center is in charge of data processing and overall scheduling of in-orbit test. It performs the in-orbit test in cooperation with the standard test equipment and platform control system.

The monitoring and fault diagnosis center develops an in-orbit testing program, sends satellite in-orbit testing instructions and data injection plan, which are transmitted to the platform control center. The platform control center formulates platform control programs, instructions, and data is uploaded to satellite via platform control equipment in order to set satellite state. Simultaneously, the platform control center sends the telemetry data, instruction received and data transmission condition to the monitoring and fault diagnosis center, and then the monitoring and fault diagnosis center begins in-obit testing as planned, and sends the telemetry data received by the standard test equipment to the platform control center to supplement telemetry data.

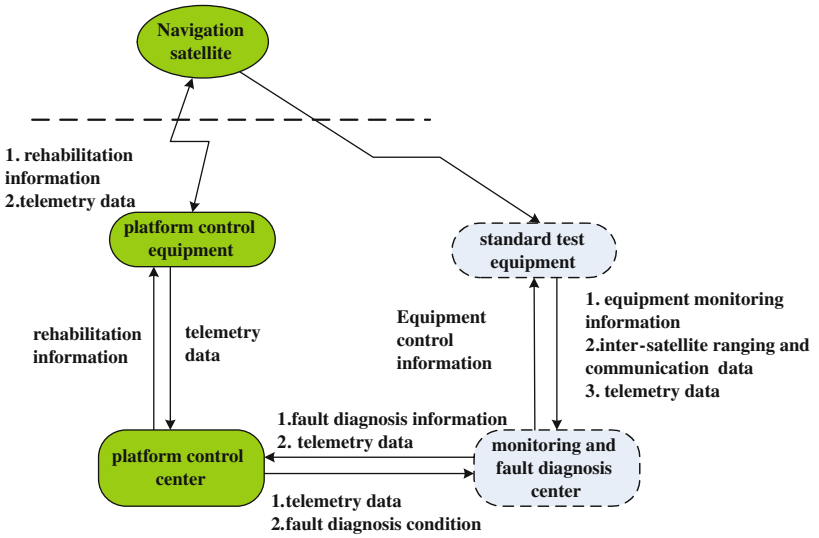

Fig. 7.3 Workflow diagram of monitoring and fault diagnosis

The monitoring and fault diagnosis center has remote monitoring on standard test equipment.

7.4.1.2 Operation Status Monitoring and Fault Diagnosis Operation Process

Monitoring of the operation status of the crosslink equipment of individual satellites and the constellation crosslinks is jointly fulfilled by the platform control system and the crosslink management system. The workflow is shown in Fig. 7.3.

The monitoring and fault diagnosis center receives satellite telemetry data of individual satellites and/or inter-satellite links, and sends the telemetry data received by the standard test equipment to platform control center. The two processes supplement each other. The monitoring and fault diagnosis center and the platform control center monitor the operation status of the crosslink equipment and constellation crosslinks. When something is wrong, the monitoring and fault diagnosis center, the platform control system and the satellite system shall locate the faulty satellite and evaluate the fault conditions based on the links between satellites and telemetry information of individual satellites, develop fault countermeasures, and rescue the satellite via platform control (if necessary, the standard test equipment shall help perform in-orbit testing). After elimination of the fault, the standard test equipment and the monitoring and fault diagnosis center shall perform an experiment and testing in orbit for the repaired satellite as part of workflow.

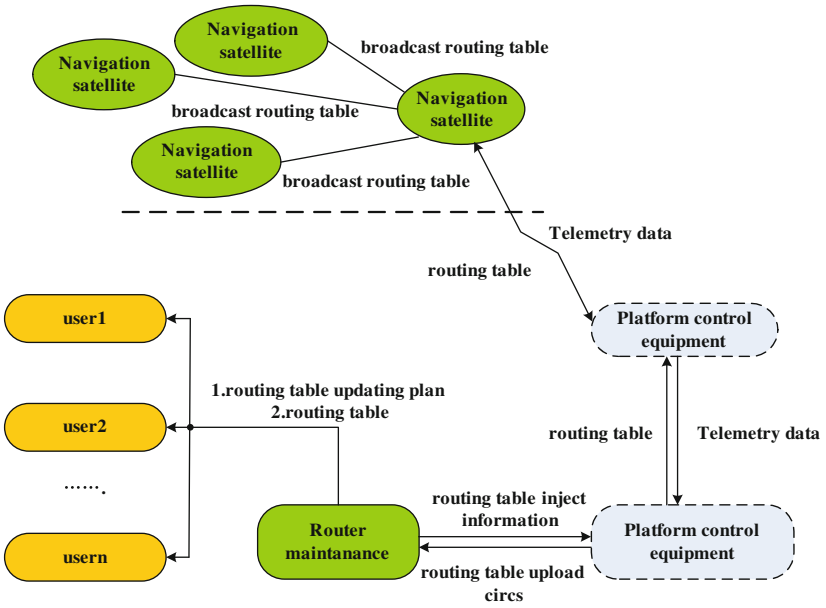

Fig. 7.4 Workflow diagram of route generation and uploading

7.4.2 Route Maintenance Operation Process

Based on crosslink topological structure and route design principle, the route maintenance center calculates the constellation routing table and then uploads to satellites via platform control. The workflow diagram is shown in Fig. 7.4.

Before the route is updated, the route maintenance center sends a route maintenance plan to users. Then, it calculates and generates a routing table based on the constellation topological structure and relevant principle. The routing table is sent to the platform control center for uploading to satellites and broadcasting to the whole network by satellites. From the downloaded telemetry data, route uploading results are assessed and a report is sent back to the route maintenance center. If the process is successful, the updated routing table will be broadcast to users by the route maintenance center and the routing table updating process is finished.

7.4.3 Resource Scheduling Operation Process

The resource scheduling center sets the principle for using crosslinks to solve conflicts between users according to user business priorities and actual operation of inter-satellite links. The workflow diagram is shown in Fig. 7.5.

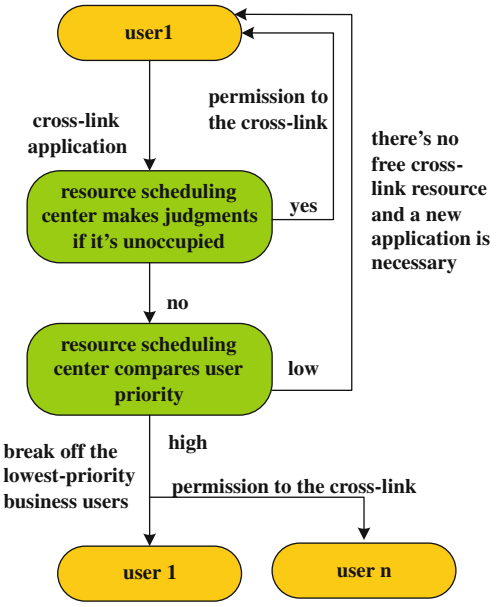

Fig. 7.5 Workflow diagram of crosslink resource

Users send applications to use crosslinks to the management center based on their own needs. Then, the resource scheduling center makes judgments whether the users' needs can be met based on the usage of crosslinks. If the answer is yes, the scheduling center will meet users' needs. If the answer is no, the management center shall continue to judge the priorities and then releases the users' resources from low-priority business for use by high-priority users. Low-priority users have to wait and make a new application for use of crosslink resources.

7.5 Security of Crosslink Management

Crosslink management security is related to satellite crosslink security, security of inter-satellite information transmission, orbit operation security of inter-satellite equipment platform, security of crosslinks and information between users and the management center, security of crosslinks and information between the platform control center and the management center, and security of crosslinks and information inside the management system.

Considering the security of inter-satellite crosslinks and information transmission, security design for the management system should be subordinated to the security design of satellite crosslinks. Considering the security of platform orbit operation, the technical means for satellite-ground information control of the management system should be subordinated to the platform control system.

In response to user requirements, appropriate level of channel and encryption and decryption mechanism defined by the users can be used between inter-satellite link users and inter-satellite link management system. The inter-satellite link management system shall cooperate. The channel between the platform control system and the inter-satellite link management system and information encryption and decryption mechanism shall be defined by the platform control and management system. The inter-satellite link management system shall cooperate.

The internal information security level within the inter-satellite management system and protection measures required shall be designed in detail in accordance with pertinent regulations.

7.6 Conclusion

The capability of a crosslink management system has a significant impact on normal operation and effective application of crosslinks of a satellite constellation. Taking into consideration of the design and topological structure of the crosslinks of the navigation constellation, an advanced, effective and cost-effective crosslink management system and security protection mechanism should be designed. Particular attention should be paid to establishment of an index system for quantitative measurement of the crosslinks and the performance of the management system.

Reference

1. Maine KP, Anderson P, Langer J (2003) Cross-links for the next-generation GPS. IEEE AC paper # LJ63, March 8–15

Chapter 8
Study on Threats to Security of Space TT&C Systems

Qi Wang, Bo Wang and Bin Wu

Abstract A space Tracking, Telemetry and Command (TT&C) system is a spaceflight support system and its security is of great significance. A framework of space TT&C system security threat assessment is proposed and it comprises of three aspects of factors: attack severity, value of asset and probability-to-succeed. An approach of qualitative gradation is introduced in the assessment. Based on the security threat assessment framework, threats to the security of space TT&C Systems during confrontation are analyzed. The analysis shows that threats of physical attacks on ground facility, RF jamming and High Power Microwave (HPM) weapon attack on satellites are outstanding and corresponding Protection measures are demanded.

Keywords Space TT&C system · Security · Threat assessment

8.1 Introduction

With the development of the spaceflight technology, the security of all parts of a spacecraft system is drawing more and more attention. A space Tracking, Telemetry and Command (TT&C) system is a support system for aerospace operation and its security is of critical impact on an entire aerospace system. Therefore, it is necessary to analyze the security threats and make appropriate protection plan to assure the security of space TT&C systems. Threats to space TT&C systems come from varied sources. First, this paper proposes a method for

Q. Wang (✉) · B. Wang · B. Wu
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: softfiona@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_8,

assessment of threats to the security of space TT&C systems. Then, threats to security of space TT&C systems during period of confrontation are analyzed because it is of fundamental importance to preparation of a protection plan.

8.2 Method of Space TT&C System Security Threat Assessment

Threats to security of space TT&C systems can be assessed from three aspects of factor: attack severity, value of asset and probability [1]. An approach of qualitative gradation is introduced in the assessment and a framework is constructed for assessment.

8.2.1 Attack Severity Assessment

Attack severity assessment adopts a hierarchical mechanism. Attacks are divided into different grades with a qualitative gradation method according to the after-effect of attacks. Threats to security of space TT&C systems can be divided into ten grades ranging from 0 to 9 according to severity of attack consequence. The qualitative gradation is illustrated in Table 8.1.

8.2.2 Value of Asset Assessment

The assets in a space TT&C system can be divided into three kinds as follows:

1. Information: Various TT&C data and system information existing in the form of the electron or signal, etc.
2. Service: TT&C service, etc.
3. Resources: TT&C equipment, environment, operators, rules and regulations, etc.

As each type of asset is of different value to the operation of the whole TT&C system, the value of each type of asset is different. The value of an asset can be considered in the assessment framework based on qualitative gradation of its importance to the system. That is to say, when the other conditions are the same, the attack of more important asset leads to more serious damage. The value of assets is classified into five grades, ranging from 1 to 5. The qualitative gradation is illustrated in Table 8.2.

Table 8.1 The qualitative gradation of attack severity

Grade	Grade definition	Type of attack
9	The space TT&C system is damaged and unable to be recovered in days	Ground facility physical attack, etc., which result in the space TT&C system damage
8	The TT&C function is lost	Software threats, ground facility physical attack, etc., which result in total loss of TT&C function
7	The TT&C function is lost partly, and only some space TT&C mission is possible	Jamming, software threats, ground facility physical attack, etc., which result in heavy loss of TT&C function
6	The TT&C function is lost partly, and only the basic TT&C mission is possible	Jamming, software threats, ground facility physical attack, etc., which result in partial loss of TT&C function and some damage to the mission
5	The TT&C function is lost partly, and the basic TT&C mission is still possible	Jamming, software threats, ground facility physical attack, etc., which result in partial loss of TT&C function but no damage to the mission
4	The TT&C function is lost in a little while, but there is no influence on the TT&C task	Jamming, software threats, ground facility physical attack, etc., which result in partial loss of TT&C function for a limited duration
3	The sensitive data are revealed	Interception, unauthorized access, etc., which result in sensitive data exposure
2	The TT&C information are revealed	Interception and stealing of data, etc.
1	Invalid threat	Invalid attack, which does not impact the space TT&C system
0	No threat	The space TT&C system operates normally

Table 8.2 Qualitative gradation of value of asset

Grade	Identification	Definition of value of asset
5	Very high	The significance of assets is very high, the damage of which leads to serious damage to the TT&C system
4	High	The significance of assets is relatively high, the damage of which leads to high impact on the TT&C system
3	Medium	The significance of assets is medium, the damage of which leads to medium impact on the TT&C system
2	Low	The significance of assets is relatively low, the damage of which leads to light impact on the TT&C system
1	Very low	The significance of assets is very low, the damage of which leads to negligible impact on the TT&C system

Table 8.3 Qualitative gradation of probability-to-succeed

Grade	Identification	Definition of probability-to-succeed
5	Very high	The attack is likely to happen and the Protection capacity is very low
4	High	The attack is probably to happen and the Protection capacity is relatively low
3	Medium	The attack is possibly to happen and the Protection capacity is medium
2	Low	The attack is of low probability to happen and the Protection capacity is high
1	Very low	The attack is likely to happen and the Protection capacity is very high

8.2.3 Probability to Succeed Assessment

The probability-to-succeed assessment is aimed to estimate the possibility of success in attack, which can be considered by both the feasibility of attack and the performance of corresponding Protection measures. The probability-to-succeed is divided into five grades, ranging from 1 to 5. The qualitative gradation is illustrated in Table 8.3.

8.3 Analysis of Threats to Space TT&C Systems

Taking a space TT&C system in confrontation for example, analysis of threats to the TT&C system is done. The threats mainly exist in the ground TT&C facilities, satellite-facility links, satellite TT&C equipment, constellation cross links, etc. Considering attack severity, value of asset and probability-to-succeed synthetically, analysis of threats to the space TT&C system in confrontation is given in Table 8.4.

As shown in analysis of threats in Table 8.4, the risks to security of TT&C systems in confrontation exist in the following parts:

Table 8.4 Assessment of threats to a space TT&C system in confrontation

Target of attack	Possible threat	Protection measures in use	Influence	Severity (0–9)	Value of asset (1–5)	Probability to succeed (1–5)	Risk (☆–☆☆☆)
Satellite	HPM weapon attack	The receiver has no reliability at high power	The satellite receiver burns	8	5	4	☆☆☆
Center	Physical attack on ground facility	No protection measures against the physical attack	The center is damaged and the TT&C function is lost	9	5	3	☆☆☆
Station	Physical attack on ground facility	No protection measures against the physical attack	The center is damaged or lost of control	7	4	3	☆☆
	Unauthorized access	Authentication of commands and access control	The data are stolen and the facilities are impaired	6	4	4	☆☆
Com. chain	Interception	Encryption limited	The key data are stolen	3	4	5	☆
Ground computer system	Software threat	Protection measures of computer system limited	The data are lost, and even the computer system is disabled partly	6	5	4	☆☆
Satellite-facility link	Interception	Protection of traffic via encryption and spread spectrum	The key data are stolen	3	4	5	☆
	Barrage jamming	Protection of traffic via spread spectrum	The TT&C function is lost	7	4	5	☆☆☆
	Deceptive jamming	Protection of traffic via encryption control and time marker	The satellite TT&C function is lost	8	5	1	☆
Constellation crosslink	Interception	No protection measures	The key data are stolen	3	4	5	☆

1. The satellite receiver may be attacked by HPM weapon. If it is not a forced amplitude-limiting digital receiver, the microwave weapon could destroy the receiver, and satellite TT&C function might be lost.
2. The center in ground may suffer physical attack. If the center facilities have no protection measures against the physical attack, the TT&C center might be damaged and the function couldn't be recovered in days.
3. The station in ground may suffer physical attack. If the station facilities have no protection measures against the physical attack, the TT&C station might be damaged and the facilities couldn't meet mission requirements.
4. Unauthorized access to the facilities and data of the station overseas may result in exposure of key data, disorder in system function or even leave the satellite to serious risks.
5. The data via the communication chain of commercial telecommunication satellites have the risk of being intercepted.
6. The ground computer system may be endangered by software threats such as backdoor attacking and viruses, which may lead to undesirable events and system damage.
7. If the information via satellite-facility links is encrypted, the risk of being intercepted will be reduced.
8. Barrage jamming will break down the satellite-facility links and the mission will be lost.
9. Deceptive jamming on the forward satellite-facility links may confuse the remote control information. Protection of traffic via encryption control and time marker is available to protect against the threat.
10. To transfer information without encryption via the constellation crosslinks has the risk of being intercepted.

It is obvious that ground facilities, satellites and links of a space TT&C system in confrontation may suffer various security threats. The analysis shows that the grade of attack severity, value of asset and probability-to-succeed of physical attack on ground facility, RF jamming and HPM weapon attack on satellites are relatively high and the corresponding risks are outstanding.

8.4 Conclusion

A framework for assessment of threats to the security of space TT&C systems is constructed and an approach of qualitative gradation is proposed. Threats comprise of three aspects of factors: attack severity, value of asset and probability-to-succeed. Analysis of threats to TT&C systems in confrontation is done based on the assessment framework. The analysis shows that the threats of physical attacks on ground facility, RF jamming and HPM weapon attack on satellites are outstanding and corresponding protection measures are required.

Reference

1. Lei J (2009) Research on the network security threat and situation assessment. Doctoral thesis of Huazhong, University of Science and Technology

Chapter 9
Construction Strategy Research on New Generation Central Computer System of Launch Centre

Shijie Song, Zhe Wang, Liping Zhang and Yongliang Yang

Abstract Measure and launch, TT&C (telemetry, track and command), communication, meteorology and duty subsystems of current launch centre are all independent in information acquisition, transmission, processing and application. Experiment and launch hall and command hall are separately constructed. As the frequency of launch missions increased and refined degree of system management improved, current launch centre faces the challenges of flat command to future launch missions, enhancing economic performance and refined management. To solve these problems, the next generation launch centre is designed following new construction strategies including development of uniform command and decision support software, construction of integrated information processing and application platform, unification of software development model and information communication standard. The construction strategies are described in detail in the paper and evaluated from aspects of mission organization, technical progress and users' requirements.

Keywords Experiment information system · Cooperative command · Decision support · Integrated display · Web services

9.1 Introduction

The experiment information system architecture of launch centre contains six sub-systems including information acquisition, information transmission, information application, operations management and security (Fig. 9.1). Information

S. Song (✉) · Z. Wang · L. Zhang · Y. Yang
Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: 1092089782@qq.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_9,

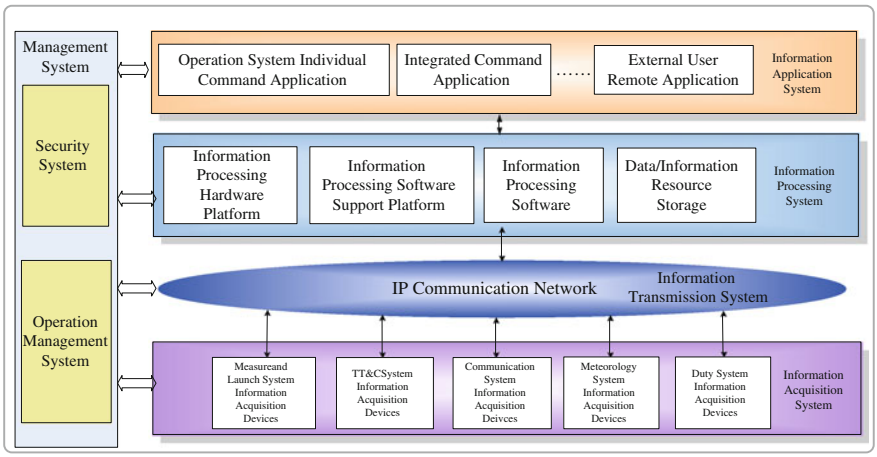

Fig. 9.1 Architecture of new generation launch centre experiment information system

acquisition system, information transmission system and information processing system are fundamental parts of the architecture, while operations management system and security system support efficient and reliable running of the experiment information system. Users use information provided by application system to make decision, control and command the system.

Information processing and information application function are mainly provided by central computer system. Information processing system consists of compute hardware platform, software support platform and information processing software platform. Its primary function is to process, store and distribute experiment information and provide information resource for launch control and measurement. Information application system consists of some individual information processing software such as command display, launch measurement, flight control and cooperative command software. Its main function is to form decision and control command through analyzing, indicating and selecting different kinds of scheme.

Measure and launch, TT&C (telemetry, track and command), communication, meteorology and duty subsystems of current launch centre are all independent in information acquisition, transmission, processing and application. Experiment and launch hall and command hall are separately constructed. In the respect of information processing and application, each operational system separately selects its processing platform and operation system, determines describing method, shared method, storage format and service method of information according to requirement of the subsystem and characteristics of experiment mission.

As the frequency of launch mission and the refined degree of management in-creased, the central computer system of current launch centre faces the following challenges.

1. The challenge of flat command for future experiment missions. Each operational system builds its own information system separated from each

other. It is complex to transmit information among systems so the commander on every level cannot get integrated experiment information from single information node, while the flat command cannot be truly realized.
2. The challenge to reduce system construction cost and improves efficiency. Infrastructure of each operational system such as communication platform, storage resource, information processing platform are often repeatedly built, in addition, the system hardly realizes load balance, so it is difficult to conform to the trend of lowering mission operation cost and improving efficiency.
3. The challenge for system integrated and refined management. Each operational system can only manage its own equipment, and information management networks are decentralized, so information cannot be integrated on top level, and equipment state and usage cannot be get at the same time.

Consequently, to develop experiment information system of launch centre, it is necessary to update the open degree and flexibility of the system, realize flexible construction of mission networks. An intelligent and integrative central computer centre is a key part of new generation launch centre to support command, control and management of experiment missions.

9.2 System Construction Strategies

In the process of central computer system construction and design, it is important to make clear universal requirement and individual requirement, data processing interface and application interface. The fundamental elements of the system are designed according to the requirement of universality and flexibility. Goals of the new system construction are to update command level and efficiency of launch mission, to adapt to diversity of launch missions, as well as to establish foundation for updating the ability of integrated experiment. System construction strategy contains the following items:

1. Intelligent cooperative command software and decision support software development. In the new generation launch centre, measure and launch hall and command control hall are constructed together. Every operational system develops and uses one universal cooperative command, decision support and integrated display software. The software provides workflow, current process state and decision support information to help commanders make accurate, scientific and efficient decisions in different mission stage.
2. Unified Information processing and application support platform construction and information resource construction. The purpose of unified information processing platform, application support platform and information resource construction is to make unified standard for selection and construction of hardware equipment such as information storage, public display and software architecture including operating system, database management system, development tools, application middleware and integrated architecture. Unified

platform provides technical support for development, testing, running, management, integration and extension of data processing software.
3. Uniform software development model and data/information interaction standard. It is necessary to uniform software development model and technical standard of each operational system, and build public services and components repository to improve reusability, maintainability and interoperability of software, furthermore, to increase development efficiency, multiple tasks support ability and extensibility of software system. Through unifying data and information collection, coding, storage and interaction format, information resource can be shared among different operational systems and users.
4. Uniform user interface. User operation interface, input interface and view display style is unified to provide users with convenient, logical and straightforward information display, to lower users' dependency on different terrain.

9.2.1 Functional Components

Central computer system of launch centre is the centre of experiment information distribution, storage, processing and application [1]; covers measure and launch, TT&C, communication, meteorology and duty operational system.

Information processing subsystem contains information sending and receiving, information storage, operational information processing and integrated information processing components. It is an efficient fundamental platform for operational system to integrate resource and share information, providing technical support for launch centre to integrated command and display.

Information application subsystem is divided into three functional components, cooperative command, decision support and integrated display [2]. Through cooperation among the three functional components, integrated command platform is constructed to serve launch centre and each operational system. The function of information application subsystem covers the whole process of launch centre mission which mainly contains measure and launch and TT&C. One main kind of process is launch measurement process of satellites and rockets on technical position and launch position, and the other is combined experiment process of TT&C system and real mission process. The workflow characteristics of communication, meteorology and duty subsystems are not obvious, so it only needs to display in integrated command platform.

Top level commander can handle mission workflow from macro aspect or micro aspect and control each mission process node based on the integrated command platform, furthermore, rapidly position problems, make corresponding decision, and evaluate integrated experiment or real mission effect. Commanders at each level can handle mission workflow and state of his subsystem and command sublevel commanders who also use the integrated command platform.

The main functional components of information processing and application system is shown as Fig. 9.2.

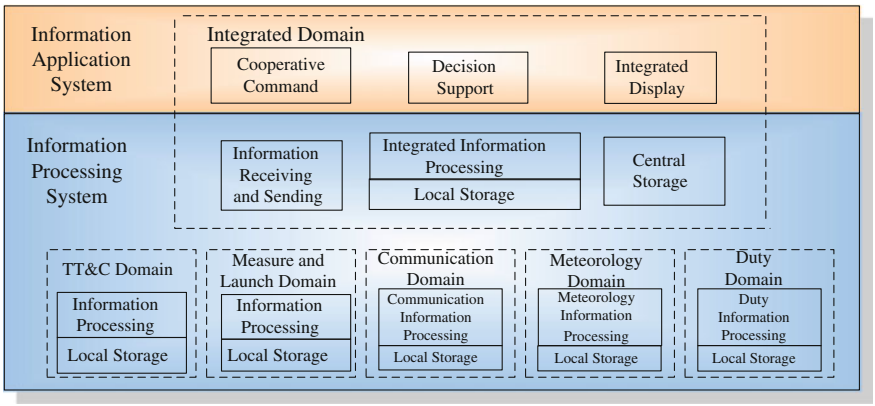

Fig. 9.2 Functional architecture of information processing and application system

Five operational systems are divided into measure and launch, TT&C, communication, meteorology, duty domain and integrated operational domain using virtual local area network technology. Each specialized domain is in charge of its own operational information processing function and local storage, while the integrated domain is in charge of receiving and sending of all operational information, integrated processing, local and central storage, as well as the functions related with integrated command such as cooperative command, decision support and integrated display. In addition to real time application, information application subsystems also include web services and unreal time digital data message services.

9.2.2 Real Time Information Receiving and Sending

Function of real time information receiving and sending is related with all data transmission of central computer system and experiment missions. The function enables subsystems be connected to mission IP network with real time information receiving and sending service, realizes regular mobility of mission information and interconnection of command nodes, ensures that original information of mission be stored real time.

Real time information receiving and sending component can adopt distributed data exchange scheme, deployed on outward data exchange nodes of central computer system as uniform outward interface of real time data. The exchanged data contains operational information of measure and launch, TT&C, communication, meteorology, duty domains and real time exchanged data with external isolated network. The components of the function are figured as Fig. 9.3.

TT&C operational domain exchanges real time data among launch centre, TT&C equipment, Beijing Centre and Xian Centre. Information receiving and

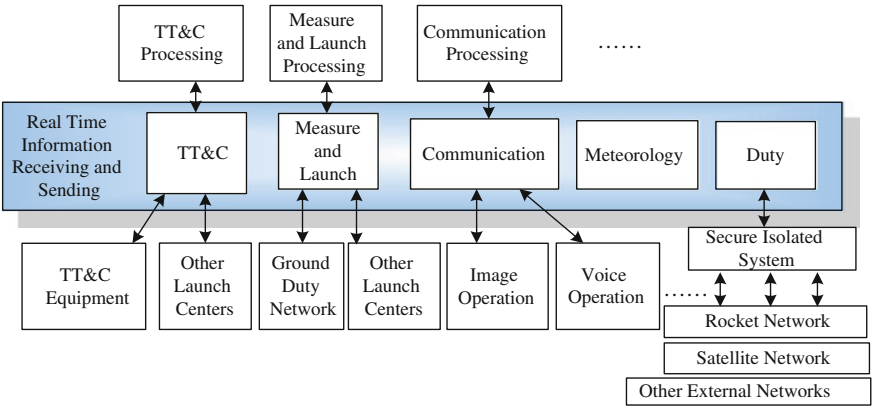

Fig. 9.3 Functional architecture of information processing and application system

sending components completes multicast in inner domain, solves the problem of switch between master and slave model, lowers coupling between centre equipment and outer equipment, and improves flexibility of centre equipment configuration.

Measure and launch domain exchanges real time data between launch centre and ground duty network, between launch centre and Beijing Centre through real time information receiving and sending components.

Real time information of outer networks such as rocket network and satellite network is exchanged with receiving and sending components of related operation domain via secure isolated system.

Integrated operation domain synthetically processes key information of TT&C and measure and launch. Real time data exchange between integrated domain and specialized domains in internal centre system is tightly coupled, so it does not need to go through real time receiving and sending components.

9.2.3 Information Storage

The main function of information storage component is to store and manage mission data including TT&C, measure and launch, communication, meteorology and duty data, as well as the command process information and equipment state information. The component meets the storage requirement of increasing data quantity and type and guarantees efficient management and application of all mission related information.

Data storage of central computer system of launch centre is divided into local storage of operational domains, local storage of integrated domain and two level concentrated storage of integrated domain. In one mission, local data storage of each operational domain is in charge of real time storage of original data, intermediate data and final data in its own domain. Local data storage of integrated domain is in charge of storing integrated processing data and displaying data. All

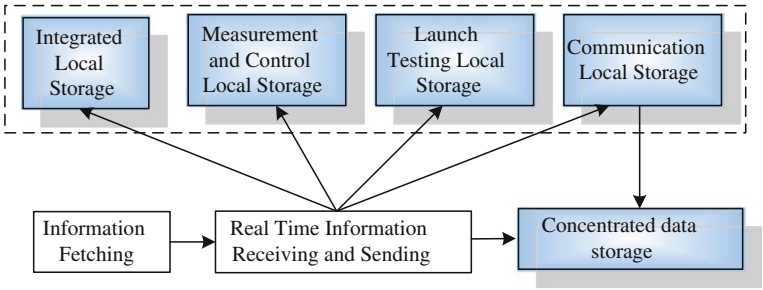

Fig. 9.4 Sketch map of information storage architecture

original data is stored while flowing into each operational domain through information receiving and sending components. After mission ends, intermediate data and final data stored in local storage is concentrated stored in unified format.

Concentrated data storage is mainly used for archiving past mission experiment data, and employs rocket state information subject, satellite state information subject, TT&C state subject, ground duty system subject, post mission data processing subject as database analysis model objects. The storage system provides upper layer with data support by efficiently organizing mission data in the way of subject analysis, meets mission and experiment data query and analysis requirement of system commander and related operation technicians, provides services across the whole process of integrated command and decision. The data storage architecture is shown as Fig. 9.4.

9.2.4 Operational Information Processing

Each operational processing system processes information of its own domain as fundamental display information [3]. For example, measure and launch system in-dependently executes unit test of technical position and subsystems; each subsystems of TT&C system independently executes module checking, telemetry full frame processing, remote measurement data fusion smooth processing, measurement and control data post processing. Operation information processing components send related information to integrated information processing components, and then integrated display component uses processing result to show operation details.

9.2.5 Integrated Information Processing

Integrated information processing components mainly process TT&C data and measure and launch data, while communication, meteorology, duty data will not be processed but only stored and shown in upper layer application.

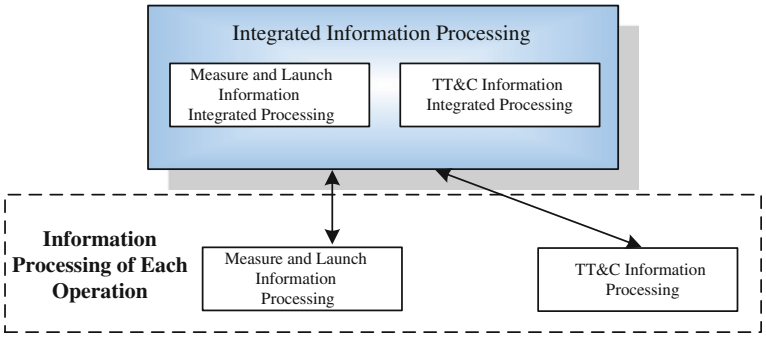

Fig. 9.5 Sketch map of data processing architecture

Based on data processing results of each operation, integrated information processing component provides data support to intelligent processing of integrated command progress state, intelligent analysis and decision, automatic transfer of flow nodes and automatic schedule display of integrated information. Architecture of data processing is shown as Fig. 9.5.

9.2.5.1 Measure and Launch Information Integrated Processing Function

Integrated information processing components make further process of specialized information processing result of base measure and launch system, and meet the requirement of integrated command by summarizing and analyzing inner relation of specialized data processing result. Through making multi-source information fusion to measure and launch data by progress, the components provide integrated display component and decision support component with data support by offering work progress and state of current test procedure.

9.2.5.2 TT&C Information Integrated Processing Function

Integrated information processing components further process specialized information processing result of base TT&C system. Under unified schedule of cooperative command component, the components automatically start joint test task and actual combat mission process. The components receive and execute project checking, equipment configuration command set by cooperative command component, and can check several different projects and equipment at the same time, thus meet the multitask requirement.

9.2.6 Cooperative Command

To meet the requirement of commanding own missions, systems of current launch centre including measure and launch, TT&C, communication, duty systems independently build up their own specialized command platform such as command and control central computer system, position C^3 system, meteorological integrated information system, position ground service network system, communication network management system.

Cooperative command component is the command platform for organization, operation and control in experiment mission. The platform integrates command in-formation of five systems and functions of integrated command, connects and analyzes procedures of each operational system. Depending on the differences among mission progress and focus of commanders' attention, the platform provides commanders with work flow, current procedure state and decision support information to guarantee accuracy, science, and effectiveness of each level commanders' decision.

Base commander makes use of the platform to fulfill integrated command, furthermore, specialized commanders of five subsystems respond to base commander and command his own subsystems, thus realizing automatic respond, execution and answer and convergence on a specified node. The platform replaces current command type which mainly depends on artificial voice dispatching with new command type mainly depending on information system and assisted by manual work. The new command type provides cooperative command ability of multiple commanders and realizes flat and networked cooperative command.

9.2.7 Decision Support

Decision support component makes comprehensive analysis to information of five subsystems according to operational requirement of commanders, and provides accurate, comprehensive, reliable decision suggestions making use of intelligent algorithm and logical reasoning to analyze useful data and information on the base of expert knowledge of aerospace field.

The component uses command information outputted by cooperative command support component as its input to drive decision rules repository in intelligent decision support system, analyzes, makes logical reasoning of processing result of fault diagnosis unit, mission evaluation unit and launch and test data processing components to provide decision support for commanders. At the same time, result of the component is the input information of integrated display component which helps commanders to handle decision support information.

On the other hand, logistics systems such as communication, meteorology and duty system have no close relationship with mission progress, so independent decision points can meet the requirement of these systems.

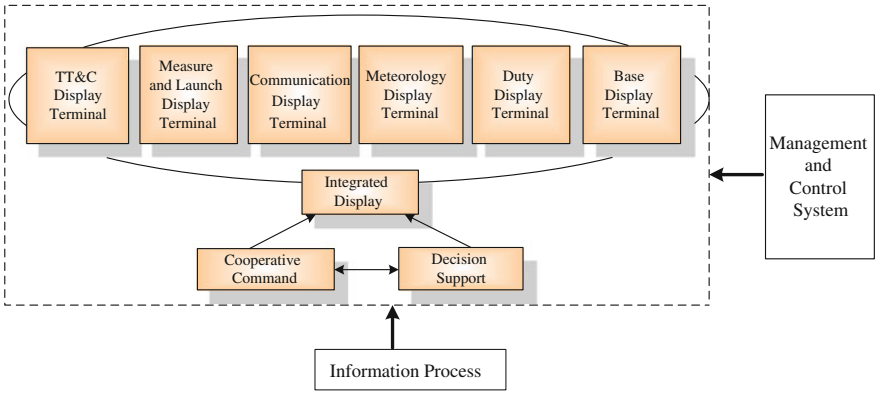

Fig. 9.6 Sketch map of integrated display system

9.2.8 Integrated Display

Integrated display component provides a convenient way to commanders to make decision through intuitive presentation of decision support information and cooperative command information.

Integrated display component is the top layer component of the system, and also a display platform of all data and information. Displayed information contains not only intelligent decision, information related with cooperative command, but also original information and parameters according to user's requirement so that the component can present overall state of mission to commanders.

The main content of integrated display includes information process, cooperative command, decision support, and control and management information. The component shows all stages of rocket test, launch, flight, star arrow separation and satellite orbited using technologies covering three-dimensional visualization, GIS system and virtual technology. The component provides individual operational interface and intelligence user management, meets the needs of asking for mission state information anytime, in addition, dynamic display related information according to different focus of users.

Integrated display software and that of each operational system are designed universally, thus every display terminal can show either integrated information or operational system information. Architecture of integrated display system is shown as Fig. 9.6.

9.2.9 Web Service

Web service is an effective way to raise information share ability. It contains two kinds of service.

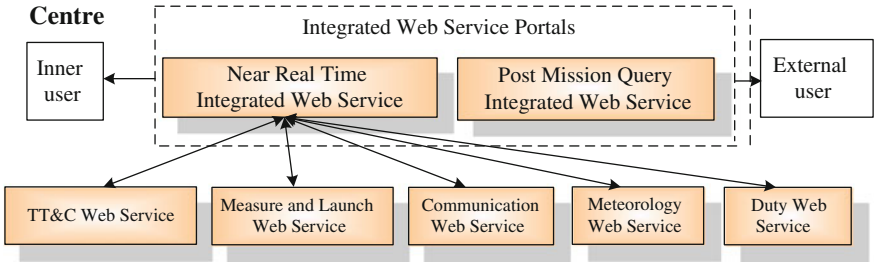

Fig. 9.7 Architecture of web service

One is to provide near real time services refreshed with mission progress to strict restricted inner and outer users. The service needs to configure web servers in each operational system, use local storage as after end database, while configure near real time web servers in integrated domain and use local storage of integrated domain as after end database. Web services of each operational system are linked to integrated web service portals in the way of hyperlink as uniform exit.

The other is to provide static history information query service of each operational domain to privileged users. The service needs to configure post mission query web servers in integrated domain and use concentrated storage as after end database.

Architecture of web service is shown as Fig. 9.7.

9.2.10 Non-real Time Message Service

Non-real time message mainly includes post mission measurement data, mission notification, and mission technological document, etc. Each internal operational system is configured message server to realize data transmission with assignment equipment and other message servers. The service employs FEP protocol to fulfill data exchange in file way. Non-real time message service is shown as Fig. 9.8.

9.3 Expected Effect Evaluation

9.3.1 Mission Organization

1. Through applying new technology, a universal and extensible technical architecture is designed for innovation of mission integrated command model and flexible configuration of mission command relation.

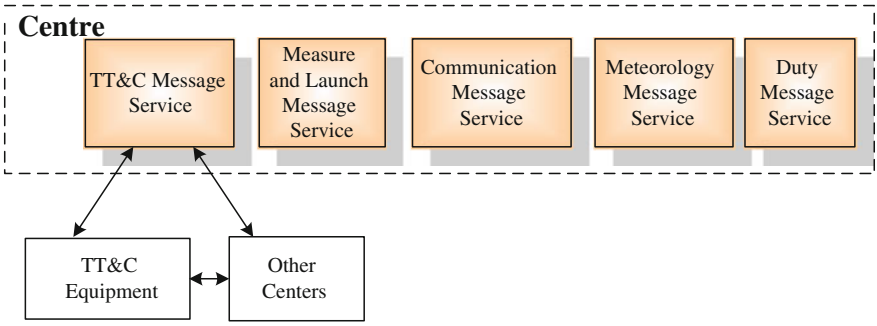

Fig. 9.8 Sketch map of non-real time message service

2. Information as bond, each operational system is close connected and can cooperate and interact in experiment mission.
3. The new system realizes digital information based mission organization and task management in the whole progress.
4. The new system realizes fine processes of each kind of mission information, extracts valuable integrated information for commanders and strengthens decision support degree of information system.
5. The method of organization and connection between inner subsystems and other operational systems is transferred from manual type to integrated command and organization type.

9.3.2 Technical Progress

1. The new design unifies software and hardware construction requirement of the whole system and technical standards, regulates information description format, shared information storage format and information access method.
2. Information share mechanism is build up that organization and command efficiency and standard of launch mission is raised, and it lays the foundation for updating integrated experiment ability.
3. Fault diagnosis mechanism and decision support mechanism are built up based on an abundant knowledge repository.

9.3.3 Meeting Requirement of Users

Information users of launch centre information experiment system can be divided into three types as base level decision user, system level command user and

terminal user. Construction of central computer system can meet the requirement of all the three kinds of users.

1. Requirement of base level decision users. Base level decision users mainly include base, headquarters, industrial department, leaders and experts of general unit. Their requirement mainly includes integrated command decision of experiment mission, effective data mining, and information diversified display based on information management. In addition, to query current and history operational information of specified domain, manage information, analyze and evaluate state of system or equipment, and make suggestions to mission progress and fault handling are also their requirements.
2. Requirement of system level command users. System level command users mainly refer to each specialized agencies, each operational system level commanders and so on. Their requirements mainly include real time analysis and mastery of experiment information, automatic management of experiment mission progress, hierarchical management and hierarchical command.
3. Requirement of operational level users. Operational level users are operators of computer equipment and each kind of experiment equipment. The main requirements include real time mastery of equipment information, analysis and report of equipment real time state, automatic and intelligent equipment operation, decision support in equipment failure model, and fine operability, maintainability and extensibility.

References

1. Meng WZ, Shi Y, Yuan G (2008) Construction of information network test system of snooting range. J Liaoning tech Univ (Natural Science) 27:187–189
2. Ju L, Zhang SG (2011) Research and design of test information integrated management system. Inform Technol 7:134–136
3. Xie XW, Gao SW (2011) The research on informatization of system testing for spaceflight equipments. Aerosp Control 27(2):61–64

Chapter 10
The Application of OFDM in UAV Telemetry

Hailong Zhao, Jian Zhang and Jie Zhou

Abstract OFDM will be a key technology in next mobile communications, and has potential applications in unmanned aerial vehicle (UAV) telemetry, so a UAV telemetry system based on OFDM is implemented on basis of analyzing its advantages in this paper. Firstly, the system scheme is presented. Then the baseband transmitter and receiver are designed in detail. All of OFDM baseband processing algorithms are accomplished in FPGA. The baseband test is developed, and the test result of PAPR is 12 dB, and the E_b/N_0 lose is about 1 dB compared with the theoretical value. Finally, the UAV telemetry system is built, and the flying experiments are done. The experiment results show that OFDM is available technology in UAV telemetry.

Keywords OFDM · UAV telemetry · Synchronization · PAPR

10.1 Introduction

At present, the development of the UAV is very fast, and the UAV will get more attention and more widely application in the future. As the link between the operator and UAV, the telemetry and telecontrol system is an important component part of the UAV. For telemetry link, the modulation mainly has FM, BPSK, QPSK and OQPSK [1].

In recent years, with the rapid development of communication technology, many advanced modulation appeared. Due to good anti-multipath performance and high frequency spectrum efficiency, OFDM is widely used in mobile communications, and it is studied in many famous universities and academic

H. Zhao (✉) · J. Zhang · J. Zhou
Institute of Electronic Engineering, CAEP, Mianyang 621900, China
e-mail: zhaohailong_029@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_10,

institutions [2]. In UAV telemetry, many scholars believe OFDM has potential application prospect [3–10], and is well worth to being studied.

The advantages of employing OFDM in UAV telemetry can be summarized as follows:

1. The more far the UAV flies, the lower the antenna elevation angle will be. Thus, multipath interference may be appeared owing to reflection from ground, mountain or sea surface. OFDM is more resilient against this multipath interference than single carrier systems, because OFDM divides the whole bandwidth into parallel narrow bands flat fading sub-channels.
2. OFDM makes efficient use of the spectrum by allowing overlap without loss of orthogonality. It can roughly double the spectrum efficiency.
3. OFDM is robust against narrowband interference and impulse noise, since such impairments only affect a small portion of total subcarriers. Those errors can be recovered by using adequate error correct codes together with interleaving.

OFDM, because of its advantages, will be a key technology in next generation wireless communications, and has potential advantages in UAV telemetry. Therefore, designing UAV telemetry system based on OFDM is necessary.

10.2 System Scheme

The hardware of the UAV telemetry system mainly consists of the high performance FPGA, DSP and high-speed ADC and DAC, and its diagram is shown in Fig. 10.1.

In Fig. 10.1a, the acquisition and coding card consists of video ADC, DSP and FPGA. It performs video compression coding and other telemetry data fusion, and sends the fusion data to FPGA1. FPGA1 mainly performs baseband transmitting, such as channel coding, data frame formation, inserting pilot tones, subcarriers allocation, constellation mapping, IFFT, PAPR reduction and digital up-conversion. Then, digital signal is converted to analog signal by DAC, and then this intermediate frequency (IF) signal is up-converted to radio frequency (RF) signal by up-convertor. Finally, the amplified RF signal is transmitted through antenna. In Fig. 10.1b, the telemetry RF signal is received through antenna, and is amplified by low noise amplifier (LNA), then is down-converted to IF signal by down-convertor, and IF signal is sent to FPGA2 through ADC. FPGA2 mainly performs baseband receiving, such as digital automatic gain control (AGC), synchronization, removing cyclic prefix (CP), FFT, channel estimation, decision and decoding. Finally, telemetry data is sent to computer through CPCI interface. The computer performs video decoding, instrument data separation, video and instrument data display.

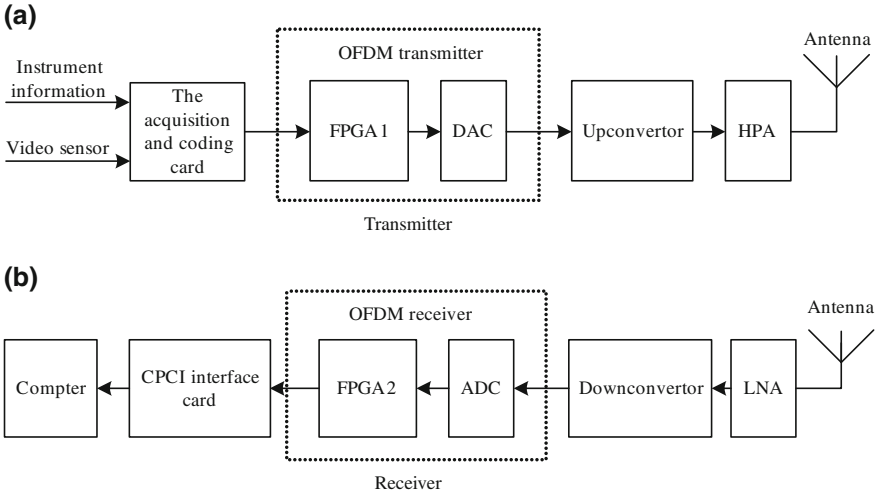

Fig. 10.1 The UAV telemetry system diagram

10.3 The Design of Baseband Transmitter and Receiver

All OFDM baseband processing algorithms of the UAV telemetry system are accomplished in FPGA, these algorithms can be divided into transmitting and receiving.

In transmitting part, the data inputted by the acquisition and coding card forms OFDM frame after scrambling, channel coding, inserting pilot tones and subcarriers allocation. In this design, considering the UAV telemetry channel is a slow fading channel, so pilot tones are inserted only in the frame head, and other OFDM symbols do not be inserted pilot tones. Then, we regroup the sequence into blocks by a serial-to-parallel converter. Each block forms in fact an OFDM symbol. Note that the original symbols can be obtained from a binary sequence by any traditional modulation methods. The symbol blocks are then processed by the inverse Fourier transform, which effectively makes the change from frequency domain to time domain. Finally, the time domain blocks are cyclicly prefixed, and PAPR is reduced and transmitted serially.

In receiving part, the received IF signal is sampled through ADC, and the sampling clock is provided by frequency synthesizer. The sampled signal is down-converted to baseband signal, and then synchronization including carrier frequency, symbol timing and sampling frequency synchronization is accomplished. OFDM signals are demodulated by removing the cyclic prefix from the received time domain samples, and performing FFT to find the frequency domain signals. Finally, the telemetry data is obtained by decision and decoding.

10.3.1 The Design of Baseband Transmitter

The baseband transmitter mainly includes eight modules, such as data scrambling, interweaving and channel coding, inserting pilot tones, subcarriers allocation, QPSK constellation mapping, IFFT, inserting CP, PAPR reduction, and digital up-converting. The detail designs are shown as follows.

1. Data scrambling. In order to facilitate debugging, a binary sequence is stored in internal ROM of FPGA. If debugging or Bit Error Rate (BER) testing, we use internal the binary sequence. If sending telemetry data, we use external data. Moreover, in order to avoid long even zeros or ones which influent synchronization algorithms, the input data is scrambled. Scrambling is accomplished by XOR the input data and PN sequence stored in internal ROM of FPGA.
2. Interweaving and channel coding. In UAV telemetry link, both interference and channel fading will lead to transmission errors on some subcarriers, so channel coding is necessary. Moreover, in order to avoid coding invalid, interweaving is also necessary. Interweaving is accomplished by operating double-port RAM. The coding type is convolution code, and its generating polynomial is (171, 133), and its coding efficiency is 1/2.
3. Inserting pilot tones and subcarriers allocation. Considering the UAV telemetry channel is a slow fading channel, so pilot tones are inserted only in the frame head, and other OFDM symbols do not be inserted pilot tones. Moreover, in order to avoid communication quality deterioration, we use the subcarrier switching technology. When the interference on some subcarriers is detected, these subcarriers will not send useful data in next transmitting. After inserting pilot tones and subcarrier allocation, the OFDM frame is formed according to a frame head following many symbols.
4. QPSK constellation mapping. QPSK constellation is mapped in gray way, because this mapping way may eliminate the possibility of a significant error. The constellation is represented by 8 bit complementary code.
5. IFFT. In order to restrain DC component and facilitate filter design, the number of used subcarriers is less than the points of IFFT. The first and middle input ports of IFFT are inputted zeros, and other ports are inputted useful data. IFFT is accomplished by using IP core in FPGA.
6. Inserting CP. Inserting CP actually improves the data rate, which requires system working clock improving. In order to avoid using two working clocks in FPGA, so we use the frame structure shown in Fig. 10.2, this structure can guarantee using a working clock. Inserting CP can be completed by controlling reading and writing of double-port RAM.
7. PAPR reduction. Considering the performance and complexity of the algorithm, we use Selective Mapping (SLM) algorithm to reduce PAPR [11, 12]. This algorithm will not lead to signal distortion, moreover, the more the number of channels is, the better the performance of PAPR reduction is.

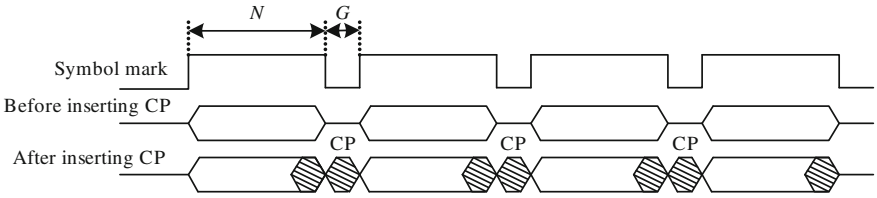

Fig. 10.2 OFDM frame structure

8. Digital up-converting. Up-sampling is necessary before digital up-converting, and this operation is accomplished by inserting zeros in middle of samples, and putting through low pass filter. The filter is used FIR IP core in FPGA. The up-sampled I and Q signals are produced Cosine and Sine signals respectively. The Cosine and Sine signals are generated by DDS IP core in FPGA. Finally, the IF signal is generated by adding produced I and Q signals.

10.3.2 The Design of Baseband Receiver

The baseband receiver mainly includes six modules, such as digital down-converting, symbol timing, carrier and sampling frequency synchronization, removing CP and FFT, channel estimation, residual frequency offset (RFO) tracking, constellation decision and decoding. The detail designs are shown as follows.

1. Digital down-converting. The ADC supports I and Q cross sampling, and each channel includes 1:2 serial-to-parallel converter, so there are four parallel signals sent to FPGA through ADC. Digital down-converting is accomplished in FPGA, and we first generate four Cosine and Sine signals by DDS IP core, then produce four parallel signals and Cosine and Sine signals respectively, filter high frequency component by FIR finally.
2. Symbol timing, carrier and sampling frequency synchronization. The symbol timing and carrier frequency synchronization are used to improve ML joint algorithm, and sampling frequency synchronization is used literature [13] algorithm. In joint algorithm, in order to reduce the amount of computation, solving phase angle is completed by look-up table, and solving correlation is completed by the way of sliding window.
3. Removing CP and FFT. Removing CP is actually recovering the symbol mark in Fig. 10.2 according to symbol timing estimation, and this processing does not need transform the working clock. After recovering the symbol marks, we do N points FFT starting with the recovered marks in every OFDM symbols. FFT is completed by IP core in FPGA.

4. Channel estimation. Because of only inserting pilot tones in frame head, the frame synchronization need to complete before the channel estimation. Channel estimation is accomplished by LS algorithm [14], and the estimation result of LS algorithm frequency response is

$$H_{LS}^{p}(k) = X_{p}^{-1}(k)R_{p}(k) \tag{10.1}$$

where, $X_p(k)$ is pilot tone in transmitter, and $R_p(k)$ is pilot tone in receiver. Since pilot tones are inserted dispersedly, the frequency response of every subcarrier is got by interpolation filter after the operation of (10.1). In LS algorithm, the frequency response estimation is related to symbol timing estimation. Assuming carrier and sampling synchronization are perfect, the channel frequency response of LS algorithm is

$$H_{LS}^{p}(k) = \frac{R^{p}(k)}{X^{p}(k)} = H^{P}(k)e^{j\frac{2\pi k\Delta d}{N}} \tag{10.2}$$

where, $H^P(k)$ is the actual frequency response, and $H_{LS}^{p}(k)$ is the frequency response estimation using LS algorithm. We can see that $H_{LS}^{p}(k)$ is related to symbol timing error Δd, so time estimation is updated only once in frame head. The RFO may lead to constellations rotation after channel estimation, so the RFO tracking is next work.

5. The RFO tracking. The RFO may lead to constellation rotation, and this rotation is related to the time. The longer the time is, the larger the rotation will be. The RFO tracking algorithm based on constellation characteristics proposed in literature [15] is used after channel estimation. In this algorithm, at first, constellations of every OFDM symbol are divided into four subsets, and then phase error is extracted by solving phase angle of every subset's average. Finally, the frequency offset compensation in the domain is achieved by using a two-tap loop filter. Solving phase angle is completed by look-up table.
6. Constellation decision and decoding. Since convolution code is used in transmitter, the Viterbi soft decision achieved by IP core is employed to get binary data. The descrambled data is sent to computer through CPCI interface. So far, the whole baseband design of transmitter and receiver are finished.

10.3.3 Test of Baseband Transmitter and Receiver

The test is developed after the baseband transmitter and receiver debugging is implemented. The envelope of OFDM is non-constant. If power amplifier works in nonlinear area, some high peak will be clipped. Thus, it will result in SNR loss. We test the PAPR through oscilloscope, and the result is 12 dB. Figure 10.3 shows BER curve, we can see that E_b/N_0 loses about 1 dB compared with the theoretical value.

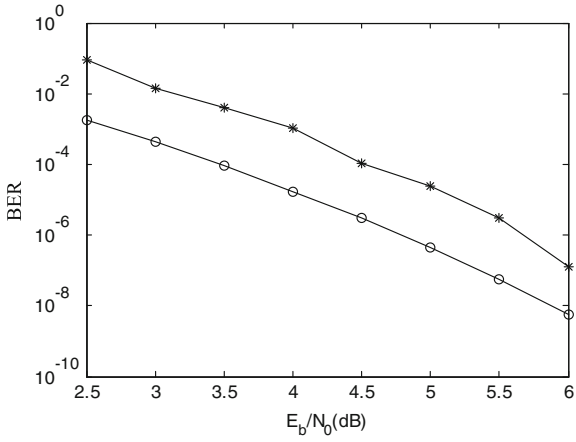

Fig. 10.3 Bit error rata curve

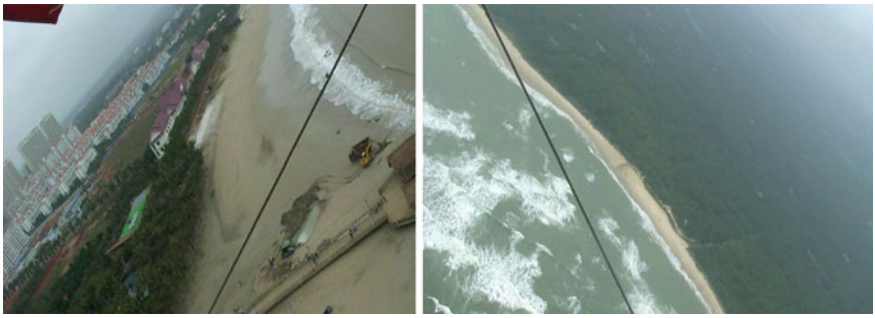

Fig. 10.4 Screenshots of a video in flight

10.4 System Experiments

After the design and debugging of baseband transmitter and receiver are completed, the UAV telemetry system is built. Then the flying experiments are done. The whole UAV system works well, and can transfer high definition video and telemetry in real time. Figure 10.4 is screenshots of a video in flight, and Fig. 10.5 shows telemetry information. Through the system experiments, it proves that OFDM is available technology in UAV telemetry.

10.5 Conclusion

This paper firstly analyzes the advantages of UAV telemetry based on OFDM, and then a system is designed. The system scheme is presented, and baseband

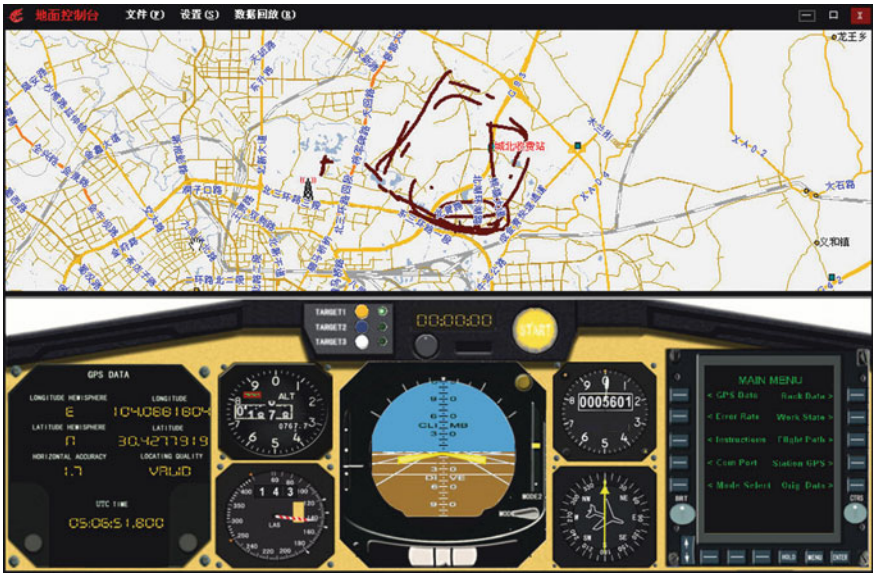

Fig. 10.5 Telemetry information display

transmitter and receiver are designed and tested in detail. Finally, the flying experiments are done. The experiment results show that OFDM is an available technology in UAV telemetry.

References

1. Wang YS (2005) The UAV communication technology. Cruise Missile 2:20–27
2. Pollet T, Van Bladel M, Moeneclaey M (1995) BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise. IEEE Trans Commun 43(234):191–193
3. You YH, Lee KT, Kang SJ (2008) Pilot-aided frequency offset tracking scheme for OFDM-based DVB-T. IEEE Trans Consum Electron 54(3):1053–1058
4. Durso CM (2003) A robust digital wireless link for tactical UAV's. In: 39th annual international telemetering conference
5. Zhang HJ, Wei C (2010) The theory of OFDM technology and the analysis of OFDM application at telemetry system. China Sci Technol Info 16:149–150
6. Ehichioya D, Golriz A, Cole-Rhodes A et al (2009) Performance comparison of OFDM and DSSS over an aeronautical channel. In: 45th annual international telemetering conference
7. Kamirah DK, Dean R (2009) OFDM performance on aeronautical channel. In: 45th annual international telemetering conference
8. Yao CF, Xiu CD, Qi DH (2010) Performance evaluation of OFDM synchronization technique of UAV data link. Commun Technol 43(1):17–19
9. Zheng JX (2011) Fading characteristics of UAV TT&C channel and anti-fading technology. Telecommun Eng 51(3):56–58
10. Zou CH, Liu GW (2007) Simulation and research of OFDM technique in UAV communication. J Syst Simul 19(10):2293–2327

11. Bauml RW, Fischer RH, Huber JB (1996) Reducing the peak-to-average power ratio of multicarrier modulation by selective mapping. Electron Lett 32(22):2056–2057
12. Lim DW, No JS, Lim CW et al (2005) A new SLM OFDM scheme with low complexity for PAPR reduction. IEEE Signal Process Lett 12(2):93–96
13. Hu DP, Zhang EY (2010) Non-data-aided sampling frequency synchronization algorithm for OFDM system. Signal Process 26(6):956–960
14. Coleri S, Ergen M, Puri A et al (2002) Channel estimation techniques based on pilot arrangement in OFDM systems. IEEE Trans Broadcast 48(3):223–229
15. Zhao HL, Zhang J, Zhou J (2012) A residual carrier frequency offset tracking algorithm based on constellation characteristics for OFDM systems. Signal Process 28(2):206–212

Chapter 11
TT&C System Design Based on Protocols and Master–Slave Structure

Feng Xu, Xiaofang Wang and Jianhong Zhao

Abstract The current TT&C System is characterized by monotonous in application layer protocol and simple in structure, which results problems in heavy workload in Space Mission preparation, complicated system joint test, and inflexible of facility adaptability. In this paper, a new TT&C System Designing strategy based on protocols and Master–Slave Structure is raised. According to the characters of the existing TT&C System data exchange protocol, the TT&C System could be upgraded by adding in supplementary protocols and data types to realize "join and play" function. Thus, the system joint test workload can be reduced, Space Mission preparation time can be shortened, and flexibility of TT&C System can be improved.

Keywords TT&C system · Protocol · Master–Slave structure

11.1 Introduction

TT&C System is consisted by TT&C facilities, communication networks, and central computer system. The current data exchange adopts PDXP Protocol based TCP/IP protocol application layer, which defined Mission Identification (MID), Source Identification (SID), Destination Identification (DID), Message Identification (BID), Length (L), and Data. Based on this protocol, TT&C facilities and

F. Xu (✉) · X. Wang · J. Zhao
Jiuquan Satellite Launch Centre, Lanzhou 732750, China
e-mail: xufengjianjian@sohu.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_11,

central computer system conduct independent data processing procedure according to different data types. TT&C System currently only assigns target functions for different facilities but does not clearly define its system structure.

Because of the monotonous application layer protocol and simple system structure, there are some prevailing problems in the TT&C System. First of all, there is heavy workload in software testing, different functional software must complete adaptability maintenance and test, covering from data channels test to data application test. Secondly, the system joint test is complicated, all related facilities must be coordinated with central computer system to complete testing works for functional software, and much manpower must be assigned to the test. Thirdly, facilities adjustment is not flexible, the center computer system is only assigned data processing function but no management function is assigned. Therefore, "join and play" between the center computer system and the facilities cannot be realized.

To solve the above prevailing problems in the TT&C System in practice, a new TT&C designing strategy based on protocols and Master–Slave structure is proposed in the paper. According to the strategy, management functions can be assigned to the central computer system by adding additional application layer protocols and most of the system testing and coordination works can be automatically completed by software according to TT&C System protocols. Thus, the workload in the mission preparing process can be reduced, adaptability of the TT&C System can be improved, and the "join and play" function can be realized.

11.2 Protocols

11.2.1 Protocols and Layers

Protocol is a set of rules followed by each involved parties. In computer technology, it mainly covers data type, data processing model, data processing sequence, electrical standard, and so on.

To reduce the complexity of data exchange system, communication networks are normally structured in different layers. Each layer's function is to deliver specific services to the next higher layer but shield the details of the process from its next higher layer.

Protocol and layer can abstract different applications to form common rules, adopting same data structure, logical sequence, processing model, and make the applications to be automatically completed by software and hardware. They reduce people's workload and are keys to an open style communication. Good layering structure and complete protocol are solid bases to realize broad and flexible applications.

11.2.2 The Existing Protocols in TT&C System

Currently, the TT&C System is using Private IP networks as its data communication networks between facilities and central computer system. The networks adopt TCP/IP protocol and define PDXP protocol in the application layer of TCP/IP. TCP/IP protocol is an open communication protocol with sound layering structure and complete communication protocols between different layers. The Private IP networks adopt all TCP/IP protocols in the layers below the application layer. Thus, the TT&C System has the basic condition for flexible applications.

Compared with the TCP/IP protocol based Internet, which is characterized by broad and flexible applications, TCP/IP protocol based TT&C System is characterized by inflexible and limited applications. The big difference in applications is the result of different protocols. In the application layer of TCP/IP, there are many protocols, including HTTP, SMTP, and FTP, supporting Internet applications, while there is only PDXP protocol supporting TT&C System and PDXP only defines data type without covering process model [1].

It is the deficiency in TT&C System application protocol in the TCP/IP application layer that results the problems, such as no automatic coordinating mechanism, heavy workload in software maintenance and testing, and inflexibility in application. Although many effective practices in modularity have been taken in different software in TT&C System, the practices cannot be extended to the whole system and the problems cannot be fundamentally resolved.

11.3 Master–Slave Structure

11.3.1 System Structure

Master–Slave structure means there is only one master controller (host) and multiple slave controllers (facilities) in the system. The host is responsible for managing all the resources in the system, including assigning of address, communication bandwidth, and processing time, etc., and coordinates the system works. The facilities apply system resources from the host and complete its works under host's general coordination [2, 3].

The Master–Slave structure's peer is the Equivalence structure, under which the facilities in the system have equal position and share system resources. In this structure, a set of contradictional settlement mechanism must exist to avoid system malfunctions resulted from different facilities occupying same system resources in same time.

No matter under the Master–Slave structure or the Equivalence structure, in the condition of there are complete protocols, different facilities in the system can realize "join and play" through software/hardware designing according to the

protocols. However, there is significant difference in work flow and requirement to protocols under different system structures [1].

Generally, the Master–Slave structure applies to central processing process, while the Equivalence structure applies to the distributed processing process.

11.3.2 The Structure of the Existing TT&C System

The existing TT&C System does not adopt a consistent structure and the system is characterized by inexistence of contradiction settlement mechanism among facilities, no shared resources, and inexistence of management function between the facilities and the central computer system. Thus, the current TT&C System is in fact constituted by some independent nodes connected for data communication only. The simple connection relationship among the facilities results that there is no applicable protocols in the system and the system is not flexible in practice.

The working procedure of the TT&C System is different facilities obtaining data from the tracked targets and sending to the central computer system for centralized processing; and the central computer system sends messages to controlling facilities to control the tracked target's operation according to the data processing result, or presents the processing result to people in a straightforward way. In the perspective of system workflow, the central computer system is in a dominant position.

In the current TT&C System, sometimes there are multiple centers working simultaneously in practice. In the time that the TT&C System has not adopted Private IP networks and the computer processing ability was weak, the multiple centers could relieve each specific center's workload. Nowadays the computer technology has experienced fast development and data communication bandwidth has dramatically increased. Multiple centers today are more of adoptability to existing administration structure of the TT&C System. The function of most of the centers is only data relay. Thus, the multiple centers in current TT&C System are not equal but in an obvious Master–Slave relationship.

Thus, it is appropriate for the TT&C System to adopt the Master–Slave structure and select the controlling center computer system according to different missions.

11.4 General Strategy for TT&C System Structuring

Based on the above analysis, the deficiency in the existing TT&C System results that intelligence and flexible application cannot be realized in the system. Thus, the strategy of structuring the TT&C System based on protocols and Master–Slave structure is proposed to solve the problems. The general idea of the strategy is that, on the basis of the existing TT&C System, a specific central computer system

could be defined as the master controller (host) through supplementary protocols. With the common protocols adaptable to all facilities in the system, the central computer system coordinates all the TT&C facilities' data communication, and completes all system coordination works excepting data processing result analysis. Thus, the "join and play" function between the facilities and the central computer system can be realized, flexibility of the system resources can be improved, and the workload of mission preparation can be reduced.

11.4.1 TT&C System Structure and Workflow

11.4.1.1 Defining Master Controller of the TT&C System

After adopting the Private IP networks, theoretically, all the TT&C facilities should have multidirectional data sending/receiving ability and can be assigned to different centers. Thus, intermediate nodes should be reduced as much as possible, and facilities should directly communicate with the master controller, instead of relaying through other centers. If multiple centers are really needed, a multiple layer structure could be adopted and a central computer system can be defined as the host and other centers can be assigned to the master controller as affiliated facilities. Meanwhile, the non-host computers are also host for facilities communicating with themselves.

11.4.1.2 Mission Preparation Workflow

The mission preparation procedure of the TT&C System mainly covers two parts, (i) checking the matching attributes of the system interfaces, validity of parameters including address, code, data type, etc. (ii) checking data processing model, the validity of the application results, such as ballistic curve and telemetry parameter. As the PDXP protocol in the current TT&C System does not define the rules of utilization of different types of data, too many items have to be jointly tested in practice.

In the Master–Slave framework, the coordination between the host and facilities is realized in the way of ask-response. The host and facilities can automatically communicate with each other through a "facility application, host allocation, and facility receiving" model. The TT&C System mission preparation workflow can be designed in this model.

1. First of all, facilities must join the central computer system.

The facilities joining the central computer system is the process of mutual recognition between the two parties. In this sense, the facility let the central computer system know "who I am, what mission I am going to join, what I can do in what condition, etc." The central computer system responds the facility

"I recognize you, what mission I am in charge, what I need you to do, what resource can be allocated to you, etc." The following testing procedures can only be carried out after completion of the mutual recognition process, which is the precondition of the dynamic allocation of facilities.

Under ideal conditions, all resources, such as IP address and communication bandwidth, of the facilities in the system should be allocated by the central computer system, which ensures efficiency and flexibility of the TT&C System. In the condition of highly reliable communication networks and regional binding of IP address, the recognition process can be simplified to below steps:

- The facility sends the basic information, including name, IP address, location, mission code, and data transmitting/receiving format, etc., in a certain frame frequency after startup until receiving the central computer system's response.
- Once receiving the basic info from the requesting facility, the central computer system responds the facility request by sending back the received information with a certain frame frequency until receiving the facilities' confirmation.
- The facility receives response to the central computer system. If the information included in the response is correct, the facility confirms the response. Otherwise, error message will be sent to the central computer and the facility joining process fails. Error notice will be presented to the operating people.
- If the central computer system receives affirmative response from the facility, it sends link monitoring information to the facility, representing successful completion of the joining process. Otherwise, the central computer system shows failing message.
- Once the facility receives link monitoring information from the central computer system, it responds link monitoring information to the central computer system, representing the facility is ready for next step of the works. The facility joining the process is completed and the facility will wait for instruction from the central computer system for the next step.

2. The next step is data format confirmation between the facilities and the central computer system.

Once facility joining process is completed, the facility can conduct data format checking under the central computer system's coordination. The checking procedure can be conducted either by the central computer system and a specific facility or the central computer system and all facilities connected to system. The checking procedure should follow the below steps:

- When the central computer system receives link monitoring information from facilities, it starts data format checking. The central computer system sends instruction message of "checking xx data transmission format" to the involved facilities. If multiple data format exists, different data format checking instruction should be sent successively.
- Once the facility receives the instruction message of "checking xx data transmission format" sent by the central computer system, it fills in requested data in

the data frame and sends back to the central computer system and then waits for the central computer system's judgment.

- Once the central computer system receives data sent by the facility, it checks the data format and makes judgment accordingly. If the data format is correct, it sends affirmative message to the facility, representing completion of the current data format confirmation process and the central computer system can enter into the next checking procedure. Otherwise, "data format error" message will be sent to the facility and facility joining process will be exited.
- If the facility received affirmative message from the central computer system, it will stop data sending and wait for next checking instruction. If a negative message was received from the central computer system, the facility will give error notice to operating people for checking and amendment.
- If all of the data formats sent by the facilities in the system were correct, the central computer system will coordinate the facilities to check receiving data format sent by the central computer system. The central computer system sends instruction message of "check xx data received format" to the facilities. If the facilities were expected to receive multiple data format, specific data format checking messages would be sent in sequence.
- Once the facility received "checking xx data receiving format" message, it will respond the center by sending "ready for checking" message and wait for data sent by the central computer system.
- Once the central computer system received the "ready for checking" message, it would fill in the required data in the data frame and send to the facility for checking.
- When the facility receives data sent by the central computer system, it conducts data format checking. If the format was correct, affirmative message will be sent to the central computer system. Otherwise, error message will be sent to the central computer system and error notice will be presented to the operating people for checking and amendment.
- If the central computer system received affirmative message from the facility, it will stop data sending and start to coordinate the next checking procedure. If error message was received, the facility joining process will be exited.
- After confirmation of all data formats, the central computer system sends message of "facility initialization success" to the facility. Once the equipment receives the message, it responds message of "joined the central computer system". Afterwards, the central computer and the facility keep the link monitoring information sending and wait for the followed application checking.

The facility joining processes and data transmission format checking processes can be completed by software. Through the two processes, the matching attribute checking procedure for the TT&C System interfaces is completed.

3. Thirdly, Data Application Checking

The procedure is to check the validity of the data processing algorithm and result. In the current technical conditions, such checking procedure would be very

costly, if the checking was completed by computer only. Thus, people must be involved in the procedure.

Once the data application checking procedure is completed, the TT&C System is ready for mission. The mission preparation process ends.

11.4.1.3 Facilities' Dynamic Exit

If a facility joining the central computer system stops sending link monitoring data, it represents the facility exits the system.

11.4.1.4 Data Processing Algorithm of the Central Computer System

The existing "fixed facilities and fixed data" processing algorithm of the central computer system must be improved to adapt to the Master–Slave structure based and facility dynamic joining allowed TT&C System.

Firstly, the algorithm must be capable of judging minimum system requirement.

If facility dynamic joining is allowed, number changes in the facilities involved in the mission must be allowed. However, the facilities number changes should not affect the mission completion. Thus, the central computer system must have the ability to judge if the facilities joining the central computer system are enough to form a minimum required system to carry out the mission. Whenever the facilities joining the central computer system are not enough to form a minimum required system to carry out the mission, the central computer system must be capable of sending error message to people for judgment.

Secondly, the algorithm must be capable of selecting most appropriate data from redundant facilities for processing and dynamic switching.

The TT&C System is normally redundant and allows dynamic joining of facility. Thus, the central computer system cannot conduct data processing with fixed facilities. It has make adjustment according to dynamic joining of the facilities. The central computer system must optimize its selection in the redundant data and get the best processing result. Meanwhile, it must have the dynamic data switching ability to ensure the data processing result whenever some facilities exit the system.

11.4.2 Supplementary Protocols of the TT&C System

According to the TT&C System workflow mentioned in Sect. 11.4.1, supplementary protocols must be added in the existing system.

11.4.2.1 Protocol for Facility Joining

The function of the protocols covers the content of "facility join the central computer system" described in Sect. 11.4.1.2. The protocols include central computer system protocols and facility protocols applying to the central computer system and the facilities respectively. Three types of data format should be added in the existing PDXP protocol.

1. Basic facility information sent by the facilities. The specific data definition should be added in the BID. The Data should be defined according to different facilities.
2. Responding message sent by the center computer. The specific data definition should also be added in the BID. The Data should be defined according to different receiving facilities.
3. Confirmation message sent by facilities. The message is an instruction and the data definition should be added in the BID. The Data should be defined together with other instruction messages.

11.4.2.2 Data Format Confirmation Protocol

The function of the protocols covers the content of "confirming the communication data format between facilities and the center" in Sect. 11.4.1.2. The protocols include central computer system protocols and facility protocols, applying to the central computer system and the facilities respectively.

The instruction messages in the protocols share same BID with that in protocols for facility joining, with same Data. The protocols do not need additional data format but Data should be defined based on existing format.

11.4.2.3 Data Processing Protocol Packet

The protocol packet is for data application checking. It cannot replace people's judgment but stipulate the existing data processing algorithm and simplify the joint test content. In the existing TT&C System, besides telemetry and controlling data related to aerospace products, other data also has fixed format and matured processing algorithm. The data can be regularized by protocols or standards and packed in software packet in same system. They could be installed in different centers and facilities easily and the operating people's workload could be lowered.

It is worth noting that the central computers not acting as host should install both above mentioned facility protocols and central computer system protocols.

11.4.3 Optimization of TT&C System Joint Test Items

If Sects. 11.4.1 and 11.4.2 mentioned mission preparation procedure and protocols being followed, the TT&C System joint test items with people involved could be significantly simplified. After the central computer system completes the initialization procedure for the involved facilities, the TT&C System will be ready for the TT&C applications, which is "join and play". Excepting in the checking items related to the spacecraft, such as voice checking in manned space flight, in other checking items, only one comprehensive test could complete mission preparation and make the TT&C System ready for mission.

11.5 Conclusion

In this paper, a new TT&C System structuring strategy was proposed based on the Master–Slave structure and protocols. The existing TT&C System could be upgraded by adding in supplementary data format and additional application layer protocols. With the upgrading of the existing TT&C System, "join and play" function of the existing central computer system and facilities could be realized, the automation of the coordination between the facilities and the central computer system could be improved, joint test items could be reduced, and the mission preparation workload could be reduced.

References

1. Tanenbaum AS (2004) Computer networks, 4th edn (trans: Pan Aimin). Tsingha University Press, Beijing
2. PCI Special Interest Group (1998) PCI local bus specification Rev 2.2. http://www.pcisig.com
3. USB Implementer's Forum (2000) USB specification Rev2.0. http://www.usb.org

Chapter 12
Study on the Application of LT Code Technology in Deep Space Communications

Tong Guo, Daheng Zhao and Xudong Li

Abstract Deep space communications have such characteristics as long delay, high code error rate and easily broken links, etc. Easily broken links may lead to a certain degree of random flash-break of TT&C signals, which in turn will result in the unrecoverable loss of data. Luby Transform (LT) code (Luby in Proceedings of the 43rd Annual IEEE Symposium Foundations of Computer Science (FOCS), pp. 71–280 (2002)) technology is an excellent special data recovery technology for scattered data. It can solve the issue of data frame loss in deep space communications, and is very suitable for the reliable transmission of large volume of data, but it cannot improve the code error performance of transmission. Therefore considering the issues such as high code error rate and easily lost data in deep space communications, the paper studied and implemented the concatenation of LT code and LDPC/TPC code, expected to provide more reliable solution for the acquisition of deep space communications data.

Keywords Deep space communications · LT code · LDPC/TPC code

T. Guo (✉)
Science and Technology on Information System Engineering Laboratory, National University of Defense Technology, Changsha 410073, China
e-mail: strongdragon@yahoo.cn

D. Zhao · X. Li
The 54th Research Institute, CETC, Shijiazhuang 050081, China
e-mail: dahengzhao@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_12,

12.1 Introduction

Deep space communications is faced with challenges such as large delay of data transmission, weak signals received, easily disrupted communication links and easily lost data, etc. It is necessary to use channel coding technology to recover information accurately under low SNR condition. So channel coding is one of the key technologies for aerospace TT&C and satellite communications.

A deep space communication channel can be modelled as an ideal additive white Gaussian noise (AWGN) channel. Its frequency band resource is relatively abundant, while its power resource is seriously restricted due to limited size of detection equipment and extremely long transmission distance. So the data transmission channel of deep space communications, considered as being power limited while bandwidth abundant, is a typical transmission channel that trades efficiency for reliability [1]. While as the deep space detection technology develops, the current requirement for transmission rate is increasingly higher, which greatly increases the implementation difficulty of coders and decoders.

Deep space communications have the following characteristics: large delay, limited storage capacity and processing capability of spacecraft; low efficiency of handshake process in communication; low efficiency of acknowledgement and retransmission; decreased throughput due to congestion control policy; asymmetry of information transmission and reception rate; waste of communication resources due to large amount of redundant data and feedback explosion in the case of multiple receivers [2].

Because communication and telemetry links have unstable link performance, flash-broken signals, and are not suitable to adopt retransmission-based error control, the use of LT code may improve the efficiency and reliability of transmission since it does not need any feedback channel but only a one-way link, so frequent processes of retransmission and acknowledgement are not required. It is expected to provide more reliable solution for the acquisition of deep space communications data by studying the concatenation technology of LT code and LDPC/TPC code.

12.2 Basic Concepts of Fountain Code

Digital Fountain technology for data distribution was first proposed by Luby et al. in 1998 [3]. A digital fountain has features similar to those of a water fountain: when you are filling a cup with water from a fountain, the only thing you care is to get enough water to satisfy your thirst, and you do not have to know which drops of water flow into your cup. Similarly, through digital fountain coding technology, a receiver receives coded packets from one or multiple transmitters, and it can reconstruct the original file as soon as enough coded packets are received. It is not important which specific coded packets among the coded packet sequence are

received. An ideal digital fountain should have the following characteristics: a source can generate a sequence of infinite coded packets using the original data; for a message segmented into k packets, the receiver may reconstruct it as soon as any k packets in the coded packet stream are received.

This reconstruction algorithm is expected to be very fast. Digital Fountain may be approximately implemented by loosening some requirements as follows: When coded by the transmitter, the number of coded packets can be finite; the coding and decoding algorithm can be slower; the number of coded packets the receiver has to receive to reconstruct the original message may be a little larger than the original number of packets k.

There are mainly three types of fountain codes currently proposed: Random Linear Fountain code, LT code and Raptor code.

12.2.1 Principle of LT Code

LT code, which is called universal erasure code, is the first one that has fully realized the concept of digital fountain. LT code is transmitted at unfixed rate, which completely conforms to the fountain coding concept. The data can be recovered as long as the number of received packets is a little larger than the minimum number of packets required to recover the source packets, which means LT code is approximately optimal for any erasure channels.

12.2.2 Degree Distribution of LT Code

LT code is a coding method of fountain codes, and degree distribution is critical to its performance. With a good degree distribution, the receiver may recover all the original packets with coded packets as few as possible. So, it is essential for LT code to find a good degree distribution.

Definition: degree of LT code d: each code word is obtained by performing bitwise exclusive OR operation between several original packets, and the number of original packets used to generate the code word is called degree. For example, in Fig. 12.1a, coding node T_1 is connected with S_1 and S_2, so the degree of T_1 is 2, and in a similar way, the degree of T_2 is 1, T_3 is 2, and T_4 is 2.

12.2.3 Coding and Decoding of LT Code

The coding process is as follows [4].

Assume that the original file is composed of several packets. Obtain several packets from the original file packets according to a certain probability distribution

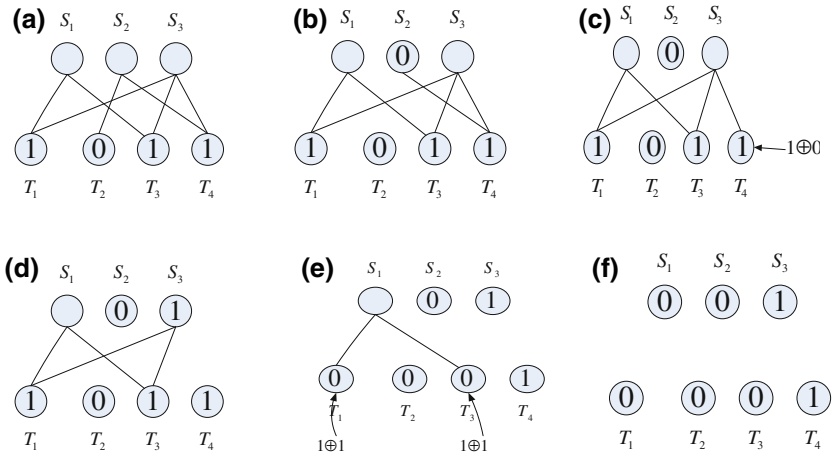

Fig. 12.1 Schematic diagram of LT decoding process. **a** Degree of T_2, $d_2 = 1$, $\therefore S_2 = T_2$, remove the connection between S_2 and T_2. **b** T_4 is connected with S_2, $\therefore T_4 = T_4 \oplus S_2$, remove the connection between S_2 and T_2. **c** Degree of T_4, $d_4 = 1$, $\therefore S_3 = T_4 \oplus S_2$, remove the connection between T_4 and S_3. **d** T_1 and T_3 are connected with S_3, $\therefore T_1 = T_1 \oplus S_3$, $T_3 = T_3 \oplus S_3$, remove the connection between them. **e** $S_1 = T_3 = 0$. **f** Decoding is finished

$\rho(d)$, then perform bitwise exclusive OR operation between them and get a new packet, and so on. The steps are:

1. Select a value of d randomly according to probability distribution $\rho(d)$. The design and analysis of a good order distribution $\rho(d)$ are critical to the performance of LT code.
2. Select $\hat{s}_1, \hat{s}_2, \ldots \hat{s}_d$ from the original packets $s_1, s_2, \ldots s_K$ as "neighbors".
3. Perform bitwise exclusive OR operation between the selected packets and get the new packet T_i. The value of d determines the number of connections of T_i.

Figure 12.1 is the schematic diagram of the LT decoding process. In each part of the figure, the three upper nodes are packet nodes, and the lower ones are four code words after coding. Code word T in the figure is 1,011, and the specific decoding process is as shown in Fig. 12.1.

1. Look for code word T_n with a degree of 1 in the received code words. The second output node in Fig. 12.1, i.e. code word T_2, is 0. If there is no such output node with a degree of 1, then the receiver continues to receive code words until it receives a code word with a degree of 1.
2. As we know from the coding principle, the code word with a node degree of 1 is equal to its "neighbor" S_k. In Fig. 12.1, $S_2 = T_2 = 0$.
3. For other output T_i ($i \neq n$) connecting with input node S_k, $T_i = S_k \oplus T_i$.
4. Delete all the connections of S_k.
5. Repeat steps 1–4. If each S_i is recovered, then the original file can be recovered successfully, otherwise the decoding fails.

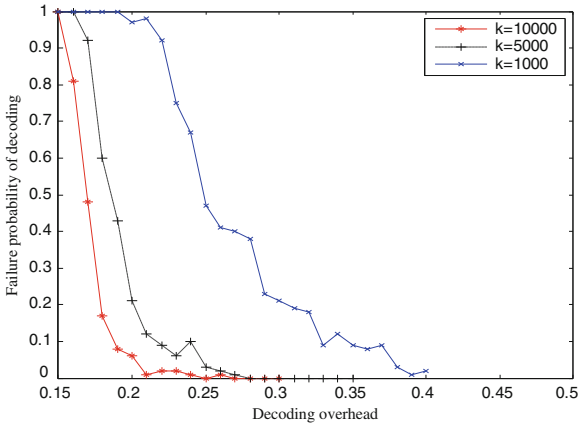

Fig. 12.2 Failure probability of decoding versus decoding overhead

Figure 12.2 shows the experiment result of using LT code for data recovery. In the experiment, the packet loss rate of the erasure channel is 10 %, and a robust solitary wave distribution is adopted for the degree distribution [5], $c = 0.03$, $\delta = 0.1$. The numbers of packets are $k = 10{,}000$, $k = 5{,}000$ and $k = 1{,}000$ respectively. The decoding overhead $\varepsilon = \frac{m-k}{m}$, in which m is the number of coded packets transmitted by the transmitter, and the number of coded packets actually received by the receiver is $m \times (1–10\ \%)$. It is assumed that, without loss of generality, the length of each packet is 1 bit. (Actually the packet consists of several bits. During coding, the bits are processed in parallel independently, and it is just the same during decoding. If each packet contains p bits, then running the decoding algorithm is actually using the same receiver generation matrix to perform the decoding algorithm of the p bits independently in parallel, and the decoding can only be considered as successful when all these p independent parallel decoding are successful.) In the experiment, the values of the overhead are taken within [0.15, 0.4] at an interval of 0.01. Run the coding and decoding algorithm 100 times at each value of decoding overhead, and get the failure probability of decoding. The relationship between the failure probability of decoding and the decoding overhead is as shown in Fig. 12.2.

As shown in Fig. 12.2, when $k = 10{,}000$ and the decoding overhead is above 0.27 (the receiver receives 11,430 coded packets), the failure probability of decoding in 100 times of coding and decoding experiments is 0, indicating that LT code is an excellent data recovery algorithm. We know from the degree distribution of LT code that the number of source packets k is also a critical factor affecting the success of decoding. Figure 12.2 shows the relationship between k and the decoding overhead. When $k = 1{,}000$, a decoding overhead of 0.4 is required to realize a decoding failure probability of 1 %, which is very large. However as k increases, the decoding overhead decreases very fast. When $k = 10{,}000$, a decoding overhead of 0.27 is sufficient for successful decoding. So LT code is very suitable for the reliable transmission of large volume of data.

12.3 Application of LT Code Technology in Deep Space Communications

Reliable downlink channel is necessary for deep space communications to download useful payload data and TT&C data to ground command and control center. While for most of the time, the deep space spacecraft is far away from the ground control station, the spatial attenuation of wireless carrier signal is very high, and the channel is affected by many interference factors such as blockage, cosmic rays and atmospheric attenuation, etc., so random and burst code errors are unavoidable, and the integrity of downloaded information cannot be guaranteed. In such case, the application of fountain code should be considered in deep space communications.

12.3.1 Concatenation Method of LT Code and TPC/LDPC Code

LT code, as a code with erasure channel as background, is currently applied in typical erasure channels in Internet communications. However, in non-erasure channels, this traditional LT code cannot be used. The remarkable feature of deep space communications is the weakness of signal, which leads to serious random code errors. Since LT code has no anti-interference or error resilience capability, if it is independently used in deep space communications for data recovery, the code error rate will be very high since the characteristics of LT decoding will spread the code errors. To realize the application of LT code in deep space communications for data recovery, a coding and decoding method is proposed in the paper by using TPC/LDPC code to realize LT-TPC/LDPC concatenated code.

The coding and decoding system diagram of this concatenated code is as shown in Fig. 12.3. With fountain code as external code and TPC/LDPC code as internal code, both the problems of packet loss and code errors can be solved. The TPC/LDPC coder, wireless channel and TPC/LDPC decoder can be considered as a virtual erasure channel. For each coded packet, if the TPC/LDPC decoding is successful, then it may be received; otherwise it will be lost. Through fountain decoding, the original packets are recovered.

12.3.2 Matlab Simulation Experiment for Concatenation of LT Code and TPC Code

Experiment conditions: the length of each data frame at the transmitter is 676. Perform inter-frame LT coding for 500 frames of data. The length of each frame is unchanged after LT coding, and the number of frames becomes 750. Then perform

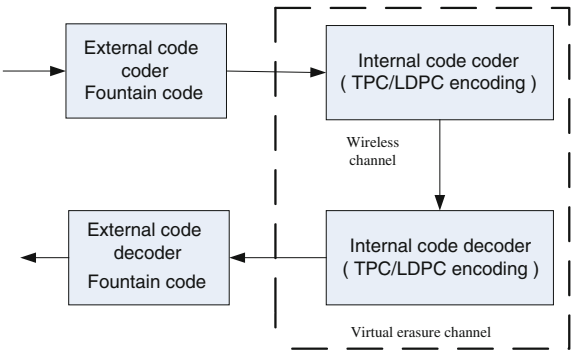

Fig. 12.3 Concatenated code system of TPC/LDPC code and fountain code

intra-frame TPC coding for these 750 frames of data. After TPC coding, the length of each frame becomes 1,024, and the number of frames is still 750. After transmission through channels with signal to noise ratio E_b/N_0 of 2.4, 2.8, 3, 3.2 and 3.4 dB respectively, the receiver performs inter-frame TPC decoding, then inter-frame LT decoding to recover data. Figure 12.4 shows the relationship between the code error rate and signal to noise ratio.

It is shown in Fig. 12.4 that when E_b/N_0 is 3.2 and 3.4 dB, the code error rate of LT-TPC concatenated code is 10^{-4} and the coding gain is about 5 dB, which indicates the good coding gain of the concatenated code.

12.4 Actual Performance of LT-TPC/LDPC Code in Engineering Implementation

In this section, the engineering implementation of LT-TPC/LDPC concatenated code is realized, and the actual performance of the concatenated code is studied. Using two industrial control computers, simulating the transmitter and the receiver respectively, two experiments are conducted: Performance test experiment for combined LT code and error correction coding; Channel on–off simulation, testing the LT code's data recovery performance.

12.4.1 Performance Test Experiment for Combined LT Code and Error Correction Coding

1. The block diagram of the experiment is as shown in Fig. 12.5.
2. Experiment conditions:

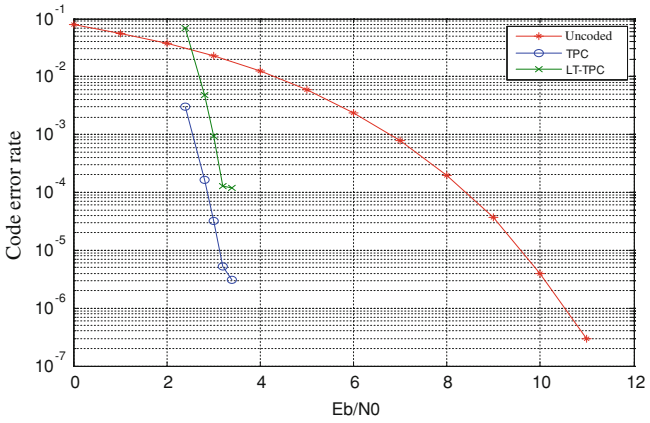

Fig. 12.4 Code error rate versus SNR

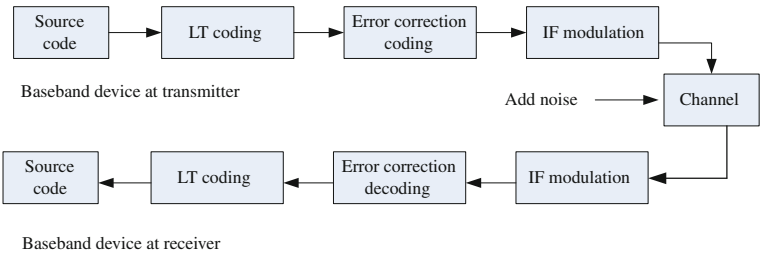

Fig. 12.5 Block diagram of performance test experiment for combined LT code and error correction coding

- Two industrial control computers, simulating the transmitter and the receiver respectively.
- Source code and packets: randomly generated by computer.
- Fountain code mode: LT coding.
- Code rate: 2 Mbit/s (when coded).
- Modulation: FM.
- Modulation IF; 70 MHz.
- Demodulation: FM + multi-symbol detection.
- Error correction coding: TPC code or LDPC code.

3. Experiment data:

- Data 1: the source code length is 488,000 bits (500*976), and through LT coding, it is divided into 900 packets of data, and the length of each is 976.
- Data 2: the source code length is 324,900 bits (100*3,249), and through LT coding, it is divided into 300 packets of data (300*3,249), and then, after TPC coding (300*4,128), the length of each packet is 4,128.

Table 12.1 Test results of code error rate

Internal code coding	E_b/N_0	External code coding: no LT code Code error rate	External code coding: concatenated with LT code Code error rate
No coding	12	1.0×10^{-7}	1.1×10^{-6}
	11	1.0×10^{-6}	1.4×10^{-4}
TPC code	5.8	2.2×10^{-7}	6.4×10^{-6}
	5.4	2.2×10^{-6}	1.6×10^{-5}
LDPC code	4.8	2.3×10^{-7}	5.3×10^{-6}
	4.6	3.1×10^{-5}	4.1×10^{-3}

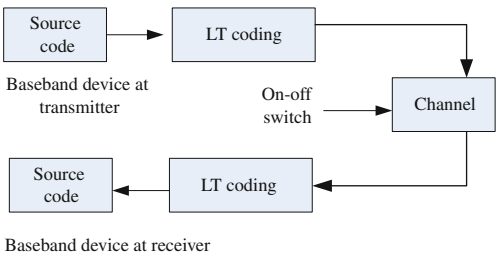

Fig. 12.6 Block diagram of channel on–off performance test experiment

– Data 3: the source code length is 102,400 bits (100*1,024), and through LT coding, it is divided into 300 packets of data (300*1,024), and then, after LDPC coding (100*2,080), the length of each packet is 2,080.

The test results are shown in Table 12.1.

It can be seen from Table 12.1 that the fountain code can solve the problem of data frame loss, while it does not have an error correction capability itself, on the contrary, it may even spread the code error. The best way to use fountain code is to perform error-free check of each data frame, and then use these error-free frames for fountain decoding.

12.4.2 Channel On–Off Performance Test

This experiment simulates the flash-break phenomena of communication link in deep space communications, and the block diagram is shown in Fig. 12.6. An on–off switch design is added to the channel, a cycle of 0.5 s is used, and the "on" duration is 0.2–0.5 s.

Data: the same as those of Data 1 in Sect. 12.4.1. The transmitter transmits a set of LT codes, with 900 frames as a cycle.

Table 12.2 Channel on–off performance test

On–off cycle (μs)	Duration of signal on (μs)	Code rate (M)	Number of signal fountain code packets	Success ratio of decoding (%)
460800	420800	2	900	100
460800	400800	2	900	100
460800	390800	2	900	67
460800	300800	2	900	0

The experiment results are shown in Table 12.2.

It can be seen from Table 12.2 that when the duration of signal on is 400,800 μs, which indicates that about 100 packets of data are actually lost due to the intermittent of signals, the decoding can be fully realized. When more packets are lost, the decoding will not be fully realized. There is a gap between this result and the theoretical value of 150 lost packets, which results from the effect of reacquisition of signal after intermittent.

12.5 Conclusion

In deep space communications, the concatenated coding and decoding method, which uses TPC/LDPC code as internal code and LT code as external code, effectively solves the data recovery problem caused by data loss due to flash-broken signals. However the code transmission efficiency is decreased and the code error rate is increased because of the use of LT code. How to increase the code transmission efficiency and further decrease the code error rate is the issue to be solved for the LT-TPC/LDPC concatenated code.

References

1. Zhai ZA, Luo L, Shi XH (2005) Research on channel encoding and decoding technologies for deep space communication. J Spacecr TT&C Technol 24(3):1–5
2. Yao WD et al (2009) Study on the fountain codes technology in deep space communications. Syst Eng Electron 31(1):40–44
3. Byers JW, Luby M, Mitzenmacher M et al (1998) A digital fountain approach to reliable distribution of bulk data. Proceeding of ACM SIGCOMM'98, Vancouver, pp 56–57
4. Luby M (2002) LT codes. In: Proceedings of the 43rd annual IEEE symposium foundations of computer science (FOCS), pp 71–280
5. Mu JJ, Jiao XP, Cao XZ (2009) A survey of digital fountain codes and its application. Chin J Electron 37(7):1571–1577

Chapter 13
Application of Multicarrier 2-Dimension Spread Spectrum in Aerospace TT&C

Jinbao Wang, Wenge Yang and Dong Liu

Abstract. Application of multicarrier 2-dimension spread spectrum (MC-2D-SS) is proposed in the field of aerospace TT&C systems to enhance their resistance to interference during mission operations. In principle, MC-2D-SS is the combination of direct sequence spread spectrum (DS-SS) and multicarrier CDMA (MC-CDMA). Therefore, it has both the characteristics of signal hiding of DS-SS and anti-inter-symbol interference (ISI) and anti-interference of MC-CDMA. A design of transmitter and receiver using MC-2D-SS is given for aerospace TT&C. Discussion is given to key technical issues of design, e.g. synchronization, ranging, anti interference and channel estimation, and a rough synchronization scheme is recommended. As a result, the TT&C system not only fully meets mission operation requirements, but also has high spread spectrum gain, strong anti-interference performance, ISI resistance capability and adaptive transmission characteristics.

Keywords: Multicarrier 2-dimension spread spectrum (MC-2D-SS) · Aerospace TT&C · Multicarrier CDMA (MC-CDMA) · Direct sequence spread spectrum (DS-SS) · Anti-interference · Synchronization

J. Wang (✉) · W. Yang
Optical and Electronic Equipment Department of the Academy of Equipment, Beijing 101416, China
e-mail: w22297202@126.com

D. Liu
Polytechnic Institute of Air Force Engineering University, Xi'an 710051, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_13,

13.1 Introduction

Along with the progress of Aerospace TT&C technology and the aggravation of space military countermeasure, a big problem which our aerospace TT&C system faced is how to command the spacecraft efficiently and ensure our space information system normal operation in a harsh environment. This condition will require the system accomplish the mission of telemetry, tracking, and commanding accurately for spacecraft in strong interference conditions. Although our country's present unified microwave TT&C system had met the need of aerospace, there are some disadvantages, such as subcarrier interference, low range accuracy, poor anti interference ability, nonsupport for multiple target control, equipment complexity and so on. On the other hand, Unified Spread Spectrum TT&C system has the advantages of large system capacity, support for multiple target tracking and anti interference ability, but it is a narrow band communication system and easy to suffer multiple access interference (MAI) and ISI. With the development of signal detection technology and the electronic interference, the system will not meet the requirements of aerospace TT&C for anti interference and anti interception. So the study of anti interference and anti interception aerospace TT&C system has been imminent.

Literature [1] proposed using chaotic sequence to construct spread spectrum TT&C system, which has good performance of security and anti interception. Literature [2, 3] suggested using the Frequency Hopping (FH) TT&C system, which has good capability of anti interception, and the United States Air Force Satellite Control Network (AFSCN) adopted Frequency Hopping mode to achieve the control tasks of the satellite [4]. Literature [5, 6] studied the Direct Sequence/Frequency Hopping(DS/FH) Hybrid Spread Spectrum TT&C system, which combines the advantages of DS-SS and FH spread spectrum and which is a more advanced anti interference system. In the three TT&C systems, it is difficult to achieve high spread spectrum gain for single spread spectrum system, and one weakness of DS/FH system is that it cannot withstand high strength tracking jamming. Based on MC-2D-SS technology, the aerospace TT&C system fuses the advantages of MC-CDMA and the DS-SS technique, which can obtain high spread spectrum gain, easily realize high rate data transmission, resist the time-frequency fading and have strong anti interference ability.

13.2 MC-2D-SS

13.2.1 Technical Principle of MC-2D-SS

Orthogonal frequency division multiplexing (OFDM) is one type of multicarrier, which appeared in the early 1950s, and until the 1970s it became much more practical along with the use of discrete Fourier transform to achieve multicarrier

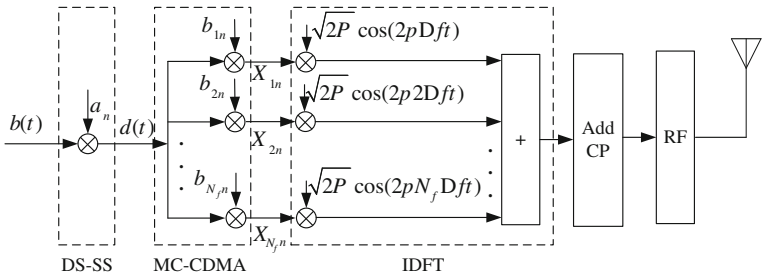

Fig. 13.1 The principle model of transmitter in multicarrier 2D spread spectrum

modulation and demodulation [7]. In 1993, three combined modes of OFDM and CDMA appeared [8–11]: multicarrier CDMA (MC-CDMA), which is also called OFDM-CDMA, multicarrier direct sequence CDAM (MC-DS-CDMA) and multiple tone CDMA (MT-CDMA). Of the three modes, MC-CDMA belongs to the frequency domain spread spectrum and MC-DS-CDMA and MT-CDMA belong to time domain spread spectrum. For MC-CDMA, CDMA was finished in the frequency domain and OFDM was finished in the time domain, therefore the final signal is OFDM signal and the characteristic (such as OFDM resist ISI) of OFDM and CDMA is reserved respectively. But for the MC-DS-CDMA and MD-CDMA, CDMA was finished in the time domain and the final signal equivalent to the summation of multiplex CDMA signals which modulate on each sub-carrier and no longer has the performance of OFDM. Especially for MT-CDMA, sub-carrier no longer maintains orthogonality and it becomes more difficult for synchronization and demodulation. Until the late 1990s, it proposed using DS-SS before MC-CDMA [12, 13], which combined with DS-SS (time domain spread spectrum) and multicarrier spread spectrum (frequency spectrum). For time-frequency two dimensional spread spectrum, CDMA was finished in the frequency domain, DS and OFDM was finished in the time domain, which could maintain the advantages of DS–SS and MC-CDMA and has time and frequency diversity effect.

The principle model of transmitter was shown in Fig. 13.1.

Assuming the rate of data bit sequence b(t) is $1/T_b$ and after DS spread the chip rate is 1/Tc, wherein $T_c = T_b/N_t$ and N_t is DS-SS code length. Let N_f as the MC-CDMA code length and α_n, β_{mn} as the DS and MC-CDMA code elements respectively, where n = 1, 2, …, N_t, m = 1, 2, …, N_f. We take single user and BPSK modulation to illustrate the MC-2D-SS idea. In Fig. 13.1, consider that a data of the data bit sequence b(t) is spread by DS-SS and the output of d(t) can be expressed as follows:

$$d(t) = \sum_{n=1}^{N_t} b(t)\alpha_n p(t - (j-1)T_c) \tag{13.1}$$

In MC-CDMA, copy the d(t) for N_f roads with the same width signals and each signal multiply with the frequency domain spread spectrum sequences β_{mn} respectively. The output:

$$X_{mn}(t) = d(t)\beta_{mn}p(t-(j-1)T_c), \quad n = 1,2,\ldots,N_t; \quad m = 1,2,\ldots,N_f \tag{13.2}$$

Assume that:

$$\boldsymbol{\beta_n} = (\boldsymbol{\beta}_{1n}, \boldsymbol{\beta}_{2n}, \ldots, \boldsymbol{\beta}_{N_f n})^{\mathrm{T}}, \quad n = 1,2,\ldots,N_t \tag{13.3}$$

$$\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots, \boldsymbol{\beta}_{N_t}) \tag{13.4}$$

$\boldsymbol{\beta}$ is the matrix of spread spectrum sequences in MC-CDMA and its column vectors are respectively corresponding to the different DS code.

Then send the two dimensional spread spectrum baseband signals into the IDFT module for processing. Let P be a sub carrier transmitting power and assuming that the sub carrier transmitting power is equal. The space of the any last two subcarrier spacing is $\Delta f = 1/T_c$ and the total subcarriers are N_f. Therefore the total band width is $(N_f + 1)\Delta f$. The outputs of IDFT are:

$$f(t) = \sum_{m=1}^{N_f} \sqrt{2p} X_{mn}(t) e^{j2\pi m\Delta ft} \tag{13.5}$$

Put formula (13.1), (13.2) into formula (13.5) and the outputs of IDFT can be expressed as:

$$f(t) = \sum_{n=1}^{N_t}\sum_{m=1}^{N_f} \sqrt{2p} b(t)\alpha_n\beta_{mn}p(t-(j-1)T_c)e^{j2\pi m\Delta ft} \tag{13.6}$$

The matrix expression of the f(t) is:

$$f(t) = \mathbf{s}\boldsymbol{\beta}\mathbf{U}\alpha\mathbf{p}\sqrt{2p}b(t) \tag{13.7}$$

where

$$\mathbf{s} = (e^{j2\pi\Delta ft}, e^{j2\pi 2\Delta ft}, \ldots, e^{j2\pi N_f\Delta ft}) \tag{13.8}$$

$$\mathbf{U} = \mathrm{diag}\{p(t), p(t-T_c), \ldots, p(t-(N_t-1)T_c\} \tag{13.9}$$

$$\boldsymbol{\alpha} = \boldsymbol{diag}(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha}_{N_t}) \tag{13.10}$$

$$\mathbf{p} = (p(t), p(t-T_c), \ldots, p(t-(N_t-1)T_c)^{\mathrm{T}} \tag{13.11}$$

Considering the formal (13.3), the two dimensions spread spectrum matrix can be expressed that

$$\mathbf{H} = \begin{bmatrix} \alpha_1\beta_{11} & \alpha_2\beta_{12} & \cdots & \alpha_{N_t}\beta_{1N_t}a_{11} \\ \alpha_1\beta_{21} & \alpha_2\beta_{22} & \cdots & \alpha_{N_t}\beta_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1\beta_{N_f1} & \alpha_2\beta_{N_f2} & \cdots & \alpha_{N_t}\beta_{N_fN_t} \end{bmatrix} \tag{13.12}$$

In MC-2D-SS system, if DS spreading code constant is 1, i.e. $\alpha_n = 1$, then the system will degenerate to the MC-CDMA, and if $N_f = 1$, which means the subcarrier number is 1, then the system will degenerate to DS-SS. At the same time, if $\boldsymbol{\beta}_i = \boldsymbol{\beta}_j$ which means the column vectors are the same as traditional MC-CDMA, then the any two column vectors of the two dimension spread spectrum matrix are related, which makes the signal detection methods simple and affects the characteristics of Low Probability of Intercept (LPI). So, it is necessary to enable the any two columns unrelated. On the other hand, if we regard the signals which go through DS as high data rate signal, then the MC-2D-SS is the traditional MC-CDMA, and if we regard the DS and multicarrier spread spectrum equivalent to a spread spectrum matrix $\mathbf{H}$, then the MC-2D-SS is the traditional MC-CDMA.

13.2.2 Advantages of MC-2D-SS

MC-2D-SS is the combination of DS and MC-CDMA, which has the advantages of two kinds of spread spectrum,

1. Easy to achieve high spread spectrum gain. MC-2D-SS gain is the summation of the time domain spread spectrum gain and frequency domain spread spectrum gain, so it is easy to generate high spread spectrum gain. Because of the severe link loss of TT&C, the signal power which the spacecraft transponder and the ground station received was very weak, so the high spread spectrum gain can significantly increase the signal-to-noise ratio and is conducive to improve the overall performance of the system. Another advantage of total gain which is the summation of DS gain and MC-CDMA gain is that it doesn't require high DS code rate and long frequency spread code, which is easy to realize.
2. Strong performance of anti interference. MC-2D-SS both have the ability of time diversity of DS signal and frequency diversity of MC-CDMA signal, thus it has good anti interception and anti interference ability. After DS, the information hides in the noise, which enhances the ability of LPI and anti interference. On the other hand frequency diversity has a good performance of resistance narrow-band interference (NBI).
3. The ability of resistance ISI and inter-channel interference (ICI). Only insert a cyclic prefix in the OFDM symbols (as long as the cyclic prefix length is greater than the maximum signal delay) and we can resist ISI and ICI effectively.

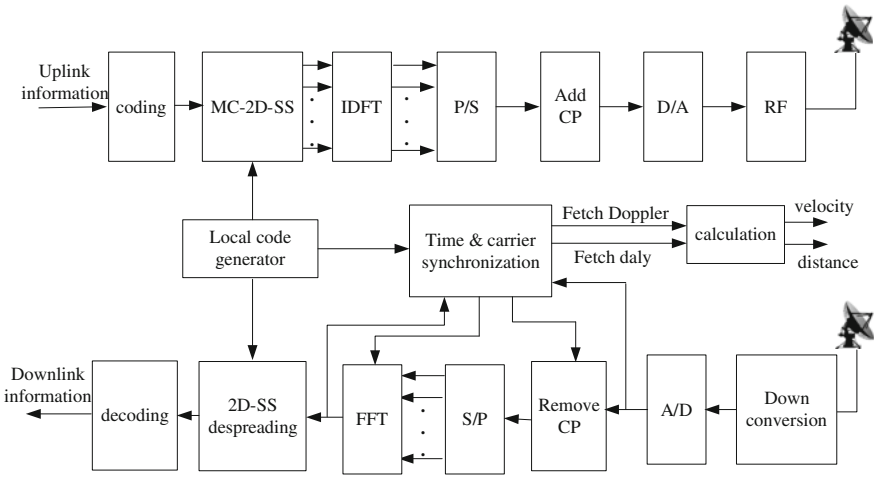

Fig. 13.2 Principle diagram of transmit and receive for aerospace TT&C based on MC-2D-SS

Compared with other methods (such as equalization and diversity technique), the complexity of system realization is low.

4. Adaptive transmission. In multicarrier spread spectrum system, according to the carrier signal to noise ratio or the interference degree of every subcarrier, we could adapt to change the modulation mode and data rate of every subcarrier, which greatly improve the transmission capacity of system and the anti interference performance.
5. High utilization efficiency of spectrum. Because the sub-carriers are orthogonal to each other, every sub-carrier spectrum has 1/2 overlap, rather than the traditional use of guard band separating the sub-channel, which greatly heighten the spectrum utilization efficiency.
6. Carry out modulation and demodulation function by FFT technology, which make the calculation and implementation easier.

Through the above analysis of MC-CDMA, we know that it has many merits and we could apply the technology in aerospace TT&C, which will greatly enhance the link transmission properties and improve the anti-interference ability of the system.

13.3 Application of MC-2D-SS for Aerospace TT&C

13.3.1 Design of Transmitting Terminal and Receiving Terminal

The schematic for application of multi-carrier 2-dimension spread spectrum in aerospace TT&C is given in Fig. 13.2.

At first, uplink information was encoded to increase signal error correction capability. Then the encoded signal was sent to two-dimensional spread spectrum, which could select different gain according to different space environment. When there is a better link environment, we could choose low gain and high efficient modulation to achieve high rate transmission. When there is a bad link environment or there exists interference, we can choose high gain and low modulation mode to meet the basic needs of TT&C. After spread spectrum, signal goes through the IDFT modulation and serial-to-parallel conversion and then inserts cyclic prefix (CP), which could eliminate multipath impact. At last the signal goes through the D/A conversion and modulates to the RF. After that it forms uplink signal.

Spacecraft transponder receives the signal that the ground station transmits and then takes coherent or incoherent processing and joins the telemetry signals, which form downlink signals and are transmitted to a ground station. The ground station receives downlink signals, which will first take conversion and A/D transformation. After that the signals will take timing and frequency synchronization, which correct the signal symbol timing deviation, Doppler frequency offset and clock error. After synchronization, signals will take the inverse transformation according to transmitting terminal and modulate the downlink information. At the same time we could fetch Doppler information and time delay information from the synchronization part, from which we could calculate the velocity and distance of spacecraft respectively.

13.3.2 Key Techniques

Applying MC-2D-SS in aerospace TT&C faced many difficulties in technology. On one hand, general technical problems need to solved, such as channel estimation, channel coding. On the other hand, special problems which were brought by application in aerospace TT&C must also be solved, such as synchronization in high dynamic, ranging.

13.3.2.1 Synchronization Technology

Signal synchronization is vital, because it directly affects whether subsequent signals could be processed normally and the acquisition of measurement information. The synchronization of MC-2D-SS includes the carrier synchronization, symbol timing synchronization and clock synchronization. Carrier synchronization is to provide a coherent carrier which is the same frequency and phase with received signal. Symbol timing synchronization is to determine the beginning and ending time of each OFDM symbol, which also mean determining the accurate FFT position of the window. Clock synchronization is to ensure that the receiving terminal and the sending terminal have the same sampling frequency. Among them, Carrier and symbol timing synchronization methods are generally divided into three types:

1. data aided algorithm: estimation based on the specific training information which was embedded in a transmitted signal [14].
2. un-data aided algorithm (or blind algorithm): The synchronization relies entirely on OFDM signal itself or its spectral characteristics [15].
3. algorithm based on the cyclic prefix [16]: estimation by using the signal's cyclic prefix.

In the three algorithms, data aided algorithm is fast, accurate and highly reliable, which could extend the estimation range by appropriate select sample number between training symbols, but the transmission of auxiliary symbol occupies system resources and reduces the utilization efficiency of spectrum resources. Blind synchronization only applies to some cases whose estimation performance is general and complexity is high. The merits of the algorithm based on cyclic prefix are that the calculation amount is small and the algorithm is simple, but the frequency estimation range is small and the time estimation is rougher. In addition, because the received signal is a spread spectrum signal and we know the received signal spread spectrum sequence, so we could adopt the related method to estimate the carrier frequency and symbol time.

In aerospace TT&C, because of the high relative velocity between spacecraft and ground station, there exists large Doppler frequency offset, which will destroy the orthogonality between sub-carriers of MC-2D-SS and lead to the mutual interference between the sub-channels. So how to realize the carrier synchronization of MC-2D-SS signal with large Doppler frequency offset is the priority among priorities.

In synchronization processing, we generally make the associated estimation of carrier and symbol time. At first, in time domain we estimate the partial carrier frequency offset (CFO, relative to the subcarrier spacing) and the rough symbol time, and then in the frequency domain we estimate the integer carrier frequency offset and the accurate symbol timing. Because the cyclic prefix estimation algorithm is simple but the precision is relatively coarse while without auxiliary information the spreading sequences Correlation Algorithm's estimation is accurate but the average acquisition time is longer, so the two estimation algorithms

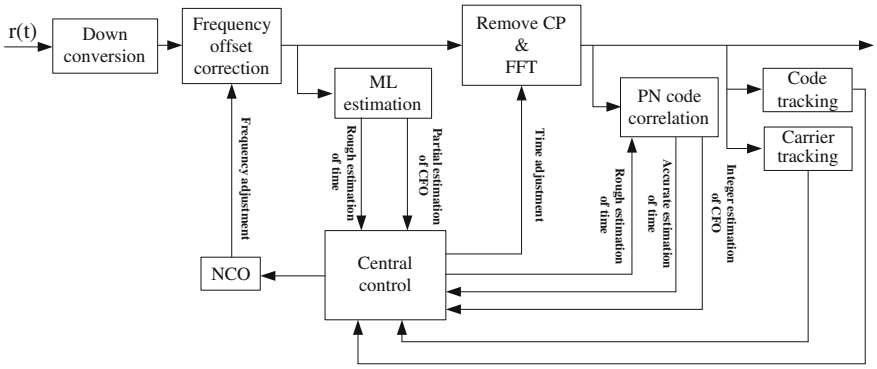

Fig. 13.3 Principle diagram of combination estimation of carrier and symbol time

can be combined. The processing was shown in Fig. 13.3. At first take estimation algorithm based on the cyclic prefix and then acquire the time rough estimation and partial carrier frequency offset estimation, which will be used to adjust the open time of FFT window, the local oscillator and the PN code phase. After that correlate the domain signals which go through FFT calculation and obtain the fine estimation of time and integer carrier estimation. At last send the fine time estimation to the code tracking loop and the carrier estimation which combine the partial and integer carrier estimation to the carrier tracking loop, which will realize the time and carrier accurate estimation respectively.

After the PN code and carrier tracking, we can obtain accurate delay information and the Doppler information, through which we could acquire the speed and distance measurement. On the time delay accurate estimation, in addition to the traditional delay locked loop, literature [17] proposed using frequency domain correlation method, which could get the phase difference between different subcarriers and then obtain accurate estimation of time delay. Under the same conditions, the ranging accuracy of the method at least increases by 4 times than PN code ranging method which uses a delay locked loop. literature [18] proposed iterated time delay estimation method based on the maximum likelihood function, which converges rate quickly and with one iteration the accuracy tends to convergence limit, and it is not affected by delay length and subcarrier frequency interval.

13.3.2.2 Anti-Interference Technology

One of the significant advantages of MC-2D-SS is that it has strong anti interference ability. Therefore it is necessary to study the anti interference technology. On one hand, we should study the anti interference performance of signal itself in different gain, different ratio of spread spectrum respectively at the same gain, different modulation modes, and different spread spectrum code, which could

provide the basis for design signal. Literature [19] points out that if there is the same transmitted bandwidth, there is the similar resistance NBI performance between the MC-2D-SS and the time domain spread spectrum or frequency domain spread spectrum. But MC-2D-SS could realize broad bandwidth easily, which has the advantage to resisting NBI. Literature [20] shows that when every multi-tone interference is just in the frequency center of any sub-carrier, if the row elements' summation of two dimensional spread spectrum matrix which corresponds to the interfered subcarrier is zero, the effect of multi-tone interferences could be eliminated completely. On the other hand, we should study the interference suppression methods. Because the MC-2D-SS is wideband signal, so we could focus on NBI suppression method.

13.3.2.3 Channel Estimation

The MC-2D-SS's channel can be equivalent to N independent parallel sub-channels, which is necessary for channel estimation when applied in aerospace TT&C. Channel estimation algorithm can be divided into blind estimation algorithm and estimation algorithm based on the pilot signal aided. Blind estimation means that by using the corresponding information processing technology it can obtain the channel estimation value whose characteristic is that it does not occupy the system bandwidth but its convergence rate is slow. As a result, it is unsuitable for aerospace TT&C application. Pilot symbol assisted method is to insert some known symbols and sequences in some fixed position of the transmitting signal and at the receiving terminal according to some algorithm to estimate the channels' performance by using the pilot symbols. Because aerospace TT&C is burst transmission mode, which needs short time for the channel estimation, thus the algorithm based on the pilot signal auxiliary is more suitable for it. Also because MC-2D-SS is a wide band signal for high gain, designing a pilot pattern which is good distributional in time domain and frequency domain is the key of channel estimation.

13.3.2.4 Coding Technology

In aerospace TT&C, the signal's loss is severe through the path of transmission because of long distance, which could be improved by coding and decoding technology to heighten the signal-to-noise ratio and error correcting capability. Currently, BCH code, Turbo code and LDPC code have been widely applied in wireless mobile communication, deep space communication and satellite communication. Introducing this technology in MC-2D-SS system for aerospace TT&C can effectively protect the transmission data, effectively overcome the defects of ISI and ICI and reduce the bit error rate.

13.4 Summary

Applying MC-2D-SS in aerospace TT&C cannot only complete telemetry, tracking, and commanding function, but can also finish multiple target management and control, overcome ISI and ICI, and the most important thing is that it has strong anti-interference performance. The solution of the key technologies such as synchronization and channel estimation, will promote the application of MC-2D-SS in aerospace TT&C, which will have great significance in meeting the needs of national space information confrontation.

References

1. Liu JX, Wen J (2009) Conception for a Ka-band spread spectrum TT&C system. Telecomm Eng 49(5):33–37
2. Henkle TG, Taggart DA (1995) Topics regarding 44/20 GHz frequency-hopping communication standards for telemetry, tracking, and commanding (TT&C). IEEE
3. Simone L, Salemo N, Mafei M (2006) Frequency-hopping techniques for secure satellite TT&C: system analysis & trade-offs. Alcatel Alenia Space-Italy. IEEE
4. Yang HY (2002) Air force satellite control network review. Telecommun Eng 42(3):138–144
5. Yang WG, Meng SY, Wang JB (2009) Analysis of a synchronization scheme for DS/FH TT&C signals. Telecommun Eng 49(6):5–9
6. Gou XT (2009) The study of hybrid spread spectrum system and its performance evaluation for spacecraft. Acad Trends 2:8–9
7. Yin CC, Luo T, Le GX (2004) Multi-carrier broadband wireless communications technology. Beijing University of Posts and Telecommunications Press, Beijing
8. Yee N, Linnartz JP, Fettweis G (1993) Multi-carrier CDMA in indoor wireless radio networks. In: Proceedings of IEEE PWLRC'93, Yokohama, Japan, pp 109–113
9. Fazel K, Papke L (1993) On the performance of convolutionally-coded CDMA/OFDM for mobile communication system. In: Proceedings of IEEE PIMRC'93, Yokohama, Japan, pp 468–472
10. Dasilva V, Sousa ES (1993) Performance of orthogonal CDMA codes for quasis synchronous communication system. In: IEEE ICUP'93, Ottawa, Canada, pp 995–999
11. Vandendorpe L (1993) Multitone direct sequence CDMA system in an indoor wireless environment. IEEE, 4.1.1–4.1.8
12. Xing H, Rinne J (1998) The performance analysis of a two dimensional CDMA system for frequency selective channels. IEEE 5:2537–2542
13. Tang YX, Li SQ (2004) The performance of generalized time-frequency domain two dimensional spread spectrum in AWGN channel. J Electron Inform Technol 26(2):248–253
14. Moose PH (1994) A technique for OFDM frequency offset correction. IEEE Trans Comm 42:2908–2914
15. Negi R, Cioffi JM (2002) Blind OFDM symbol synchronization in ISI channels. IEEE Trans Commun 50:525–1534
16. van de Beek JJ, Sandell M, Borjesson PO (1997) ML estimation of time and frequency offset in OFDM systems. IEEE Trans Signal Process 45(7):1800–1805
17. Ren G L, Sun C Q, Ni H et al (2009) OFDM based precise ranging technique in space. J Xidian Univ 36(2):203–206, 268
18. Ren GL, Bai Y, Ni H et al (2009) New precise space ranging algorithm for multicarrier systems. J Xidian Univ 36(6):957-961, 268

19. Qi J, Song LJ, Li SQ et al (2004) Anti narrowband interference performance analysis of two-dimensional spread spectrum communication. Mod Commun Technol 137(3):1–4
20. Zhou J, Tang YX, Xie SL et al (2006) Performance of time-frequency two dimensional spread spectrum system in multi-tone jamming environment. J Electron Inform Technol 28(11):2107–2110

Chapter 14
Parameter Estimation of Frequency Hopping Signals Based on Time Frequency Analysis

Wenge Yang, Meng Li, Libin Wang and Hao Zhang

Abstract Frequency hopping signal is analyzed in this paper using various time frequency analysis methods, including short time Fourier transform (STFT), Wigner-Ville distribution (WVD) and its improvements. By comparing the hopping patterns derived from these methods, spectrogram (SP) is chosen to estimate the hopping period, hopping timing and hopping frequency, as well as to obtain some valuable measurement results. Finally, a practical method is proposed to test the parameters of frequency hopping signals.

Keywords Short time Fourier transform · Frequency hopping graph · Frequency hopping time

14.1 Introduction

Frequency hopping technology has excellent ability in resisting interception and interfering, and has become an important system in modern military affairs. In complex electromagnetic environment, it is a key factor to acquire parameters of uncooperative frequency hopping signals to lead jammer to interfere the target communication system effectively, or to capture information of the target system. Parameter estimation of frequency hopping signals and its applications in problems

W. Yang · L. Wang · H. Zhang
Academy of Equipment, Beijing 101416, China
e-mail: yang_ttc@yahoo.cn

M. Li (✉)
Graduate School, Academy of Equipment, Beijing 101416, China
e-mail: limengcome2008@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_14,

such as detection, synchronization and direction have become a studying emphasis of researchers of different countries.

Frequency hopping signal is a typical non-stationary signal, and traditional Fourier transform cannot provide the information of hop period, hop timing and hop frequency simultaneously. Time Frequency Analysis, which analyzes signals by planar functions of time and frequency, has become a useful tool for studying unsteady signals such as frequency hopping signals. Literature [1] achieves the time representation of a frequency hopping signal by short time Fourier transform, get the edge information of the time–frequency analysis using wavelet transform, and then achieves the accurate estimation of frequency hopping rate by spectrogram analysis. This method avoids the influence of cross terms and is still available at a low *SNR*. Literature [2] puts forward a method of estimating hop period, hop timing and hop frequency using pseudo Wigner-Ville distribution, and this method is affected by cross terms and needs large calculation. Literature [3] brings forward that Gabor spectrogram can be used in the time–frequency representation and estimation of parameters of frequency hopping signals. Gabor spectrogram can restrain cross term interferences in WVD, and preserves advantages of WVD such as centralized energy and high time–frequency resolution.

This paper firstly compares and analyzes the effects of using different time–frequency analysis methods to process frequency hopping signals. Considering the results and calculation expense of different method, spectrogram is chosen to estimate the frequency hopping rate, hop timing and hop frequency, and some useful results is achieved.

14.2 The Model of Frequency Hopping Signal

It is assumed that signal s_n has N *FH* signals, and its hop period is T_h, and there are K complete frequency hopping signals, and the complete carrier frequency at $k(k = 1\ 2 \ldots K)$ is ω_{nk}. And the signal s_n can be figured as the formula below.

$$s_n(t) = a_n(t) \times \sum_{k=1}^{K} \exp[j(\omega_{nk}t + \phi_{nk})] rect\left(\frac{t - (k-1)T_{hn} - \Delta t_{n0}}{T_{hn}}\right) \tag{14.1}$$

where $a_n(t)$ is the baseband envelope of s_n, and ϕ_{nk} is the initial phase at k of s_n. The formula (14.1) can be simplified as the formula below.

$$s_n(t) = a_n(t)e^{j(\omega(t)t + \varphi(t))} \tag{14.2}$$

where $\omega(t)$ and $\varphi(t)$ express that the carrier frequency and the initial phase are time-varying. However, the actual measured frequency hopping signal is affected by the noise $n(t)$ from the environment, so the actual signal should be $s_n(t) + n(t)$.

The sampling signal f_s of the frequency hopping signal in the simulation is 10 MHz, and the middle frequency f_c is 3 MHz. Frequency hopping rate is

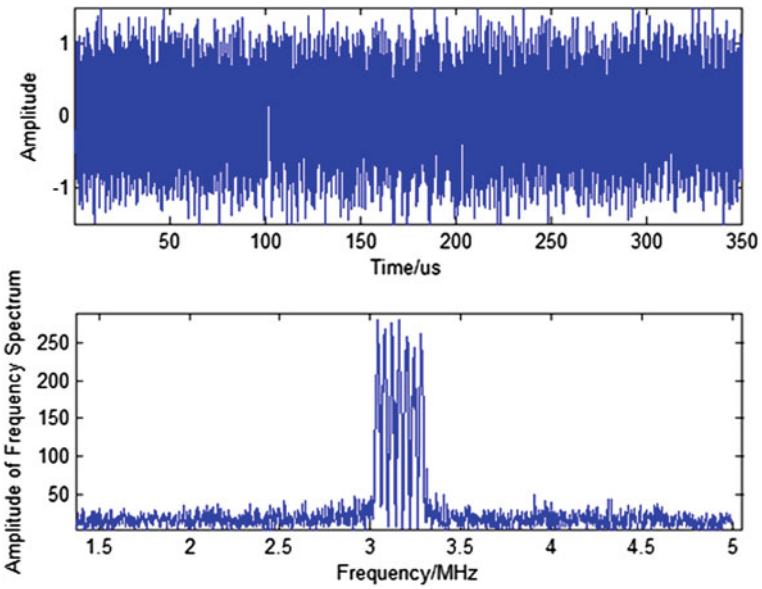

Fig. 14.1 The waveform and spectrum of a frequency hopping signal

20000 hop/s, and the hop period of a frequency hopping point T_h is 50 us, and the frequency hopping interval Δf is 40 kHz. In this instance, during a hop period, there are $N = f_s \times T_h = 500$ sampling points, and the frequency resolution is $f = {}^{f_s}/_{N} = {}^{f_s}/_{f_s * T_h} = {}^{1}/_{T_h} = 20$ kHz. In the test, eight frequency hopping points are chosen to simulate and analyze, and the corresponding hopping frequency F is [3.04 3.12 3.20 3.28 3.16 3.08 3.24 3.32] MHz. The frequency hopping signal generated from the test, conducted in the environment where Signal-to-Noise ratio is 10, is shown in Fig. 14.1.

14.3 Time Frequency Analysis

Time frequency analysis is a common method to process non-stationary signals. The frequency-time distribution figure can be achieved by processing a segment of data using time frequency analysis. The figure can reflect the frequency hopping intuitively, and estimation of some other parameters such as frequency hopping rate can be acquired. Time frequency analysis methods include short time Fourier transform, Gabor transform, WVD, PWVD, SPWVD and reassigning SPWVD. It is allowable to integrate applications to reach a good measuring result based on the characteristics of different methods. Time frequency analysis can be divided into linear time frequency representation, bilinear time–frequency distribution and bilinear time–frequency distribution by the reassignment method [4].

14.3.1 Linear Time Frequency Representation

Linear time frequency representation includes short time Fourier transform and Gabor transform. The meaning of short time Fourier transform can be expressed that it intercepts signals using window functions in time domain, and does Fourier transform with the part signal to get the Fourier transform of the segment of the signal at t. The Fourier transforms at different time can be achieved by moving the position of the center of the window function. The aggregation of these Fourier transforms is STFT(t, f), which is a planar function of time and frequency. But according to uncertainty principle, the time resolution and the frequency resolution of a short time Fourier transform are impossible to decrease at the same time and there is a lower limit of their time width and bandwidth. So the time–frequency centralization of the short time Fourier transform is limited.

The frequency hopping signal can be analyzed by simulation. For that the number of a sampling point is 500 fixedly, the frequency resolution of FFT is determinate. But it can enhance the calculating resolution to increase the number of frequency points of FFT, so the FFT of $N = 512$ points and $SNR = 10$ dB is conducted. The results of different window functions and window width are shown in Fig. 14.2.

It can be indicated that the time–frequency resolution of a short time Fourier transform is limited by the shape and width of the window function. The resolution of narrow window is higher but the frequency resolution will be low when there are few points. A wide window can intercept more points, so the time–frequency resolution is higher but the time resolution is low. The advantage of a rectangular window function is that the main lobe is centralized and can concentrate the power of a frequency point, while, its disadvantage is that there are high side lobes and negative side lobes causing high frequency interference and leakage or even negative spectrum. Hamming window has a wider and lower main lobe and side lobes decrease notably, so it is better than rectangular window in terms of reducing leakage. But the main lobe of Hamming window is widened, which means that the analyzing bandwidth is widened, causing that the resolution frequency decrease. A Gaussian window has no negative side lobes and can achieve minimum product of time and frequency width.

14.3.2 Bilinear Time–Frequency Distribution

Wigner-Ville distribution has the best time–frequency concentration, but it is unfeasible to do time–frequency analysis with multi-component frequency hopping signals. There will be masses of cross terms in time–frequency plane, and it is different to figure out auto terms. So it needs to be smoothed using window functions such as Pseudo Wigner-Ville distribution (PWVD) and Smoothed Pseudo Wigner-Ville distribution (SPWVD) [5].

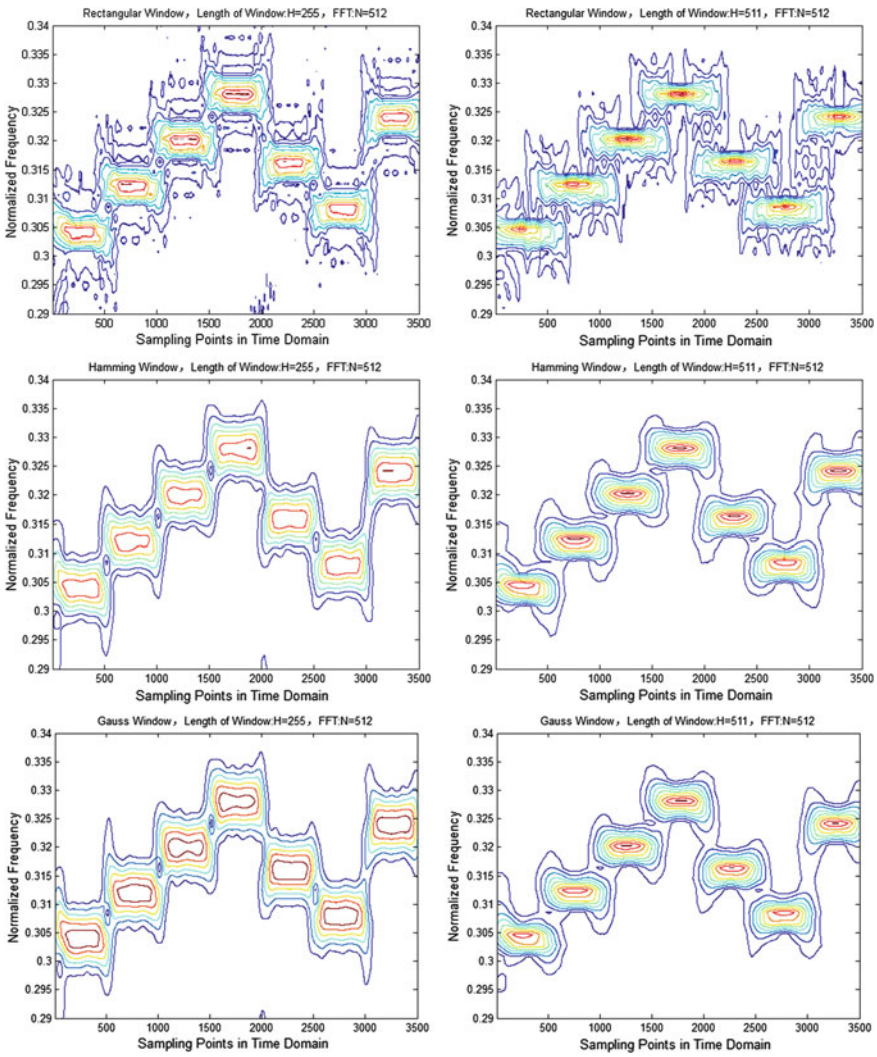

Fig. 14.2 The time–frequency contours of different window functions and window length of STFT

Spectrogram is defined to be the square of the module of short time Fourier transform, and the two dimensional convolution of the signal and the WVD of the window function.

The simulation results of the bilinear time–frequency distribution above are shown in Fig. 14.3.

From the figure, it can be seen that there are masses of cross term interference, especially at the center of the two signal frequencies. PWVD removes some interference along frequency axes, but there are still cross terms along time axes.

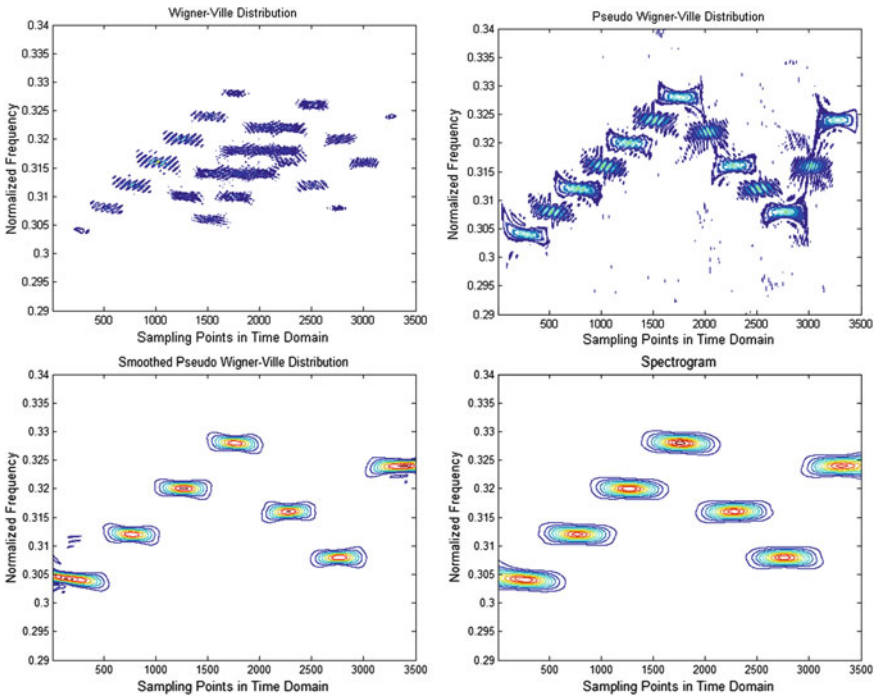

Fig. 14.3 The time–frequency contours of bilinear time–frequency distribution

SPWVD makes the signal smooth in both time and frequency domain, so the cross terms are reduced dramatically, but the time–frequency concentration decreases. The effect of spectrogram is very good, and the frequency hopping signal can be figured out clearly in time–frequency plane.

14.3.3 Bilinear Time–Frequency Distribution by the Reassignment Method

It can restrain cross terms interference greatly smoothing Wigner-Ville distribution, but there are still cross terms left for multi-component signals, and some new cross terms may be brought in. So frequency hopping signals can be analyzed using reassigning spectrogram and reassigning SPWVD, and the results are shown below.

As shown in Fig. 14.4, the frequency hopping graph generated by bilinear time–frequency distribution by the reassignment method has good performance in definition and concentration in both time and frequency domain, but it needs large calculation. And reassigning SPWVD is influenced by the environment noises notably.

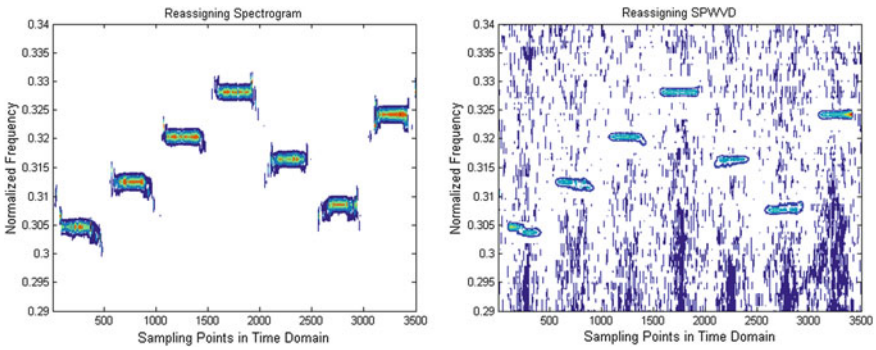

Fig. 14.4 The time–frequency contours of bilinear time–frequency distribution by the reassignment method

14.3.4 Summary of Time–Frequency Analysis

Short time Fourier transform has no cross term interference, but its concentration is low. Spectrogram based on short time Fourier transform has a good effect and the signal can be figure out clearly. SPWVD makes the signal smooth in time–frequency domain, and reduce both cross terms and concentration, getting a middle course between cross terms and concentration.

Cross terms and concentration is a couple of contradictions in all time–frequency distribution discussed above. It has been proved that a distribution that does not have any cross terms but has good concentration is impossible. But it can achieve a good result by integrating different time–frequency distributions and taking advantages of them. For example, we can get the approximate available area of the signal on time–frequency graph (setting a threshold according to SNR to intercept time–frequency graphs) by short time Fourier transform, and then multiply the intercepted time–frequency graph and WVD having a good concentration (calculating the Hadamard product of the two time–frequency matrixes) to get the integrating distribution.

14.4 Estimation of Frequency Hopping Signal Parameters by Time–Frequency Analysis

From the simulation results of frequency hopping graphs, spectrogram, SPWVD, reassigning spectrogram and reassigning SPWVD can generate good effects. But reassigning methods need large calculation, and computing speed is limited. So this paper uses spectrogram to process frequency hopping signals and tests the signal parameters further [6].

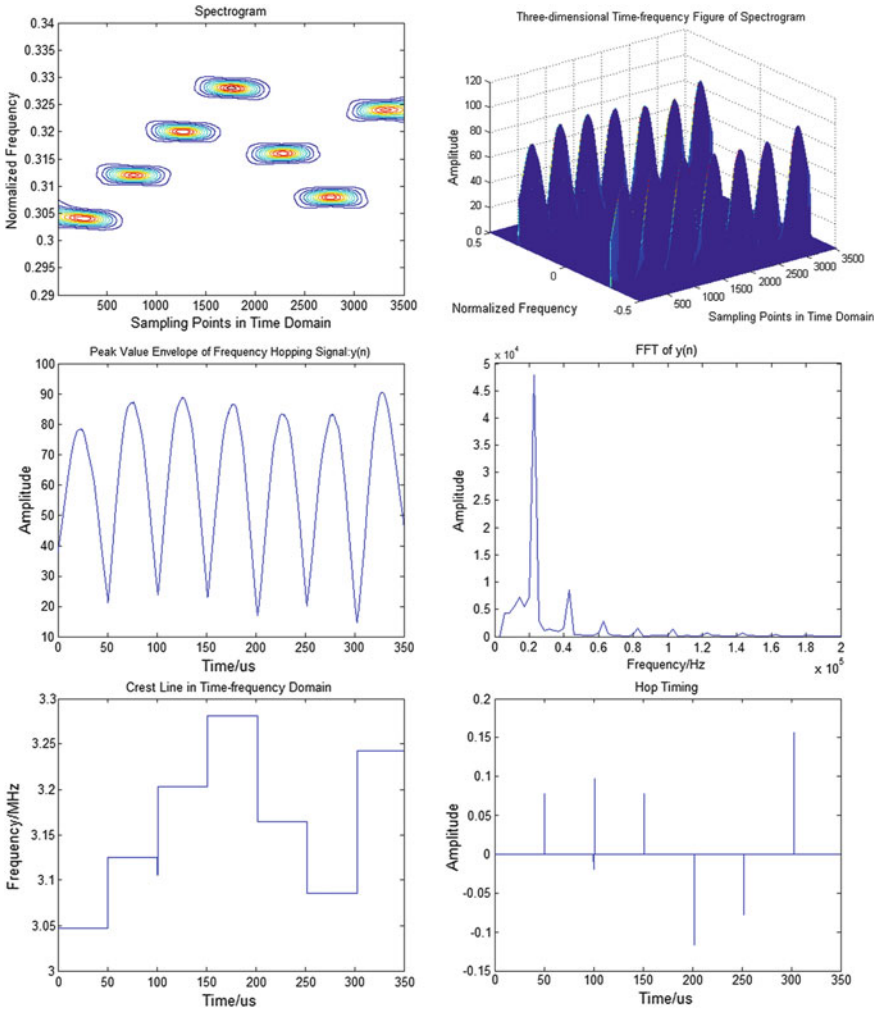

Fig. 14.5 The graph of spectrogram estimating parameters of a frequency hopping signal

The SNR is 10 in processing frequency hopping signals by spectrogram. And the time–frequency graphs are shown in Fig. 14.5.

The steps to estimate parameters of a frequency hopping signal by spectrogram are listed below.

1. Method 1:
   - To calculate DSP: SP(n,k) of the signal, and pick up its peak sequence, peak_seq(n)which is shown in the figure as the peak envelop of the signal.

- To do FFT with the peak envelop got in Step 1 to get the frequency hopping rate of the signal. And the reciprocal of frequency hopping rate is the resident time of a frequency hopping point.
- To take the maximum of the result of spectrogram as the frequency ridge. The hop period, hop frequency and hop timing of the frequency hopping signal can be seen in the figure.

2. Method 2:

- To analyze the pretreatment frequency hopping signal with spectrogram and get the time–frequency graph and the distribution function STFT(t,f).
- To take the barycenter of the time–frequency graph derived from Step 1 to get the hop frequency f_i and the time center of every hop frequency ti.
- To take the average of differences of time centers of hop frequencies to be the resident time of a frequency hopping point. It can be expressed in the formula, $T_h = \sum_{i=2}^{N_h} (t_i - t_{i-1})/(N_h - 1)$, in which Nh is the number of hopping frequency points.
- Tu sum the time center of every frequency hopping point, modulo the resident time, and divide by the number of hopping frequency points. Then the hop instant is achieved. Sum the hop instant and the resident time to be the hop timing. It can be expressed in the formula, $t_0 = \left[\sum_{i=1}^{N_h} (t_i - T_h/2)/T_h\right]/N_h$, in which t0 is the hop instant, ti is the time center, Th is the resident time, and Nh is the number of hopping frequency points. Then the hop timing is thop = t0 + Th.

It can measure the hop frequency accurately, f = [3.04 3.12 3.20 3.28 3.16 3.08 3.24] MHz. The resident time of a frequency hopping point can be measured, T_h = 50.971 us, and the error is *1.94 %*, so the frequency hopping rate is 19620 hop/s. The hop timing is T_hop = [51.4 101.5 151.4 201.9 251.6 302] us, and there are some errors between it and the actual value. The probable reason for the inaccurate result is that the signal is influenced by noises, or that the time–frequency analysis method, spectrogram, cannot meet the demand for precision.

14.5 Conclusion

Time–frequency analysis is the most common method to process frequency hopping signal. This paper analyzes and processes frequency hopping signals by time–frequency analysis methods such as short time Fourier transform, Wigner-Ville distribution and its extending distribution, and spectrogram. And it is validated that the effects of short time Fourier transform processing frequency hopping signals are influenced by the type and length of the window function. So it is not able to

make both the time resolution and the frequency resolution reach the best, and it can only balance the time resolution and the frequency resolution by choosing proper window functions and function length. Wigner-Ville distribution has good concentration but also serious cross term interferences. Pseudo Wigner-Ville distribution, Smoothed pseudo Wigner-Ville distribution and resigning smoothed pseudo Wigner-Ville distribution aim at reducing cross term interferences by changing Kernel functions. And cross terms are restrained but calculation increases notably. The concentration of spectrogram is not as good as that of Wigner-Ville distribution and its extending distribution, but it is much better than short time Fourier transform, it is free from cross terms and the calculation is less. This paper ultimately chooses spectrogram to estimate the parameters of frequency hopping signal and obtains desired results.

References

1. Zheng WX, Zhao GQ, Luo YJ (2006) Hop rate estimation for frequency hopping signals. Syst Eng Electron 28(10):1500–1502
2. Barbarossa S (1997) Parameter estimation of spread spectrum frequency hopping signals using time-frequency distributions. In: First IEEE signal processing workshop on signal processing advances in wireless communications, vol 4, pp 213–216
3. Zhang XD, Bao Z (1998) Fee-stationary signal analysis and processing. National Defence Industry Press, Beijing, pp 363–367
4. Liu XQ, Nicholas DS, Swami A (2005) Joint signal parameter estimation of wideband frequency hopped transmissions using 2-D antenna arrays. IEEE Trans Wirel Commun 4(6):3063–3074
5. Szmajda M, Gorecki K, Mroczka J (2010) Gabor transform, Gabor-Wigner transform and SPWVD as a time-frequency analysis of power quality. In: IEEE international conference on harmonics and quality of power, pp 1–8
6. Guo Y, Zhang EY, Shen RJ (2007) The time-frequency analysis and blind parameter estimation of frequency hopping signals. Signal Process 23(2):210–213

Part II
Instrumentation and Control Technology

Chapter 15
Flexible Hemispherical Simultaneous Multi-Beam TT&C Technology

Pengyi Wang, Yongfei Kong and Haizhou Wu

Abstract This paper first introduces the background and overseas research status of the hemispheric array antenna. An implementation structure, combining sub-array active phased antenna array with multi-plane digital beamforming, is proposed. Then, its design feature, development, expected performance and related key technologies are discussed, and finally a brief summary is given.

Keywords Multi-satellite TT&C · Large scale · Multiple beam · Hemispheric array antenna

15.1 Introduction

As the rapid development of scientific research satellite system, manned spacecraft system and navigation satellite system, the number of satellites becomes extremely large nowadays. However, the "chimney" state of traditional TT&C systems has been unable to meet the development needs of TT&C tasks. Today, the demand for earth-based TT&C systems is much higher. And it is a new and complex issue to manage so many satellites efficiently. How to efficiently manage plenty of satellites is the main challenge for the development of the earth-based TT&C systems.

An erratum to this chapter is available at 10.1007/978-3-642-33663-8_43.

P. Wang (✉) · Y. Kong · H. Wu
The 54th Research Institute, CETC, Shijiazhuang 050081, China
e-mail: wpybox1000@tom.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_15,

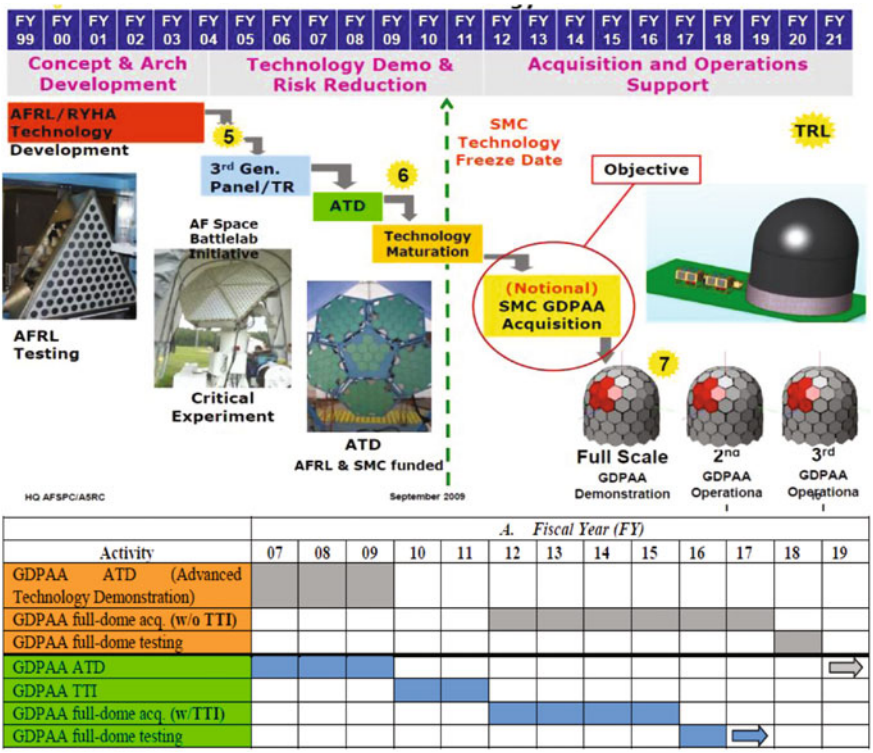

Fig. 15.1 Test procedure of American GDPAA

In order to address this challenge, two solutions have attracted extensive attention. One is the earth-based multi-target phased array TT&C system, which can be able to upgrade the ground TT&C station to the multi-target TT&C station. The other is the space-based TT&C system, which introduces relay satellites to realize simultaneous multi-target TT&C. Considering the earth-based multi-target TT&C management ability, the principal indicator of the performance of satellite TT&C, one-stop hemispheric simultaneous multi-target TT&C system, is emerging as a promising technique to deal with the sharply increasing number of satellite. Thus, this paper forces on the discussion of one-stop hemispheric simultaneous multi-target TT&C technology.

15.2 Developments of Analogous Technologies Abroad

In the area of earth-based multi-object TT&C, the United States Air Force is testing Geodesic Dome Phased Array Antenna (GDPAA) Advanced Technology Demonstration (ATD), demonstrating the advanced capabilities of communication

with satellites belong to the United States Department of Defense (DOD) [1, 2]. As shown in Fig. 15.1, the objective of the GDPAA-ATD program is to communicate with the Air Force Satellite Control Network (AFSCN). The phased array antenna is not only able to support the standard AFSCN implement, but only can provide the Air Force with whole airspace, more flexibility, timely response, and reliable telemetry, tracking and command, as well as the lower cost in the lifecycle. The experimental phased array antenna, consisting of 6 equivalent GDPAAs with an aperture of 10 m, has successfully communicated with Low Earth Orbit (LEO) satellites belong to the National Aeronautics and Space Administration (NASA) and the DOD. The preliminary test of GDPAA has been completed currently, in which four transmit (Tx) beams and four receive (Rx) beams are adopted, and the transmit frequency band range from 1.75 to 2.1 GHz and receive frequency band range from 2.2 to 2.3 GHz. Due to using the continuous-wave signal, the T/R module can be isolated with diplexer, and the maximum performance reach to 150 dB. The phase control of T/R modules uses the 4-bit phase shifter. The whole system test plans to be finished in 2017.

15.3 Recommendation of System Framework

15.3.1 System Characteristics

In general, phased array is the basic element for the phased array radar. However, according to the experience, there is a greatly different between the application of hemispheric TT&C phased array and the phased array radar, which are listed as follows:

1. Transmit-receive signal system: phased array radar and TT&C phased array adopt the transmit-receive time-sharing pulse system without modulated information and the transmit-receive continuous-wave signal system with modulated information, respectively.
2. Multi-target system: phased array radar is a time-sharing multi-target system, while the other is a simultaneous multi-target system. The later needs much more T/R modules and phase shifter modules with higher design complexity and cost.
3. The sizes of receiver and transmitter arrays: the correlated transmitted signal and received signal is used in the phase array radar, while the signals of TT&C phased array are independent. Thus, the design of the TT&C phased array is more flexible to achieve a better performance.
4. The design requirements of T/R modules: the integrated design of T/R modules in phased array radar allows the received and transmitted signals to be time-sharing. However, it is difficult to integrated design in TT&C phased array. Considering the transmitter–receiver isolation of continuous-wave, it is essential to apply transmit/receive filters. Thus, the design of T/R modules in

TT&C phased array is more difficult than that in phased array radar. In order to overcome this difficulty, transmit modules and receive modules should be designed separately.

5. The requirements of antenna polarization: in phased array radar, the received and transmitted signals usually operate with the single circular polarization. While in TT&C phased array, each beam can be switched between Right Hand Circular Polarization (RHCP) and Left Hand Circular Polarization (LHCP), which leads the simultaneous multi-polarization in received signals to be a problem. The multi-polarization makes it more difficult to design TT&C phased array antenna.
6. Multi-target tracking system: the phased array radar uses the time-division technology to achieve the multi-target tracking, which is with a simple waveform network and components. While multiple targets are tracked simultaneously in TT&C phased array, whose waveform network and components is complex.
7. Low-cost design requirement: the cost of phased array radar is higher than that of TT&C phased array. Take multi-beam TT&C station, with tele-control and tele-measure 16 targets, for example, the design cost is only half of that of 16 TT&C stations.
8. When using foreign articles about large-scale arrays, it is essential to be well aware of these above characteristics for the design.

15.3.2 System Framework

In order to fulfill hemispheric multi-target TT&C system, it is necessary to use electronic beam scanning on array antenna, which mainly includes passive phased array antenna, active phased array antenna and digital multi-beam array antenna. Considering the technological level and basic technology, using active phased array antenna is the best way. The great power network loss makes the passive phased array antenna uncomfortable for TT&C system. The cost of digital multi-beam array antenna can be much higher, because of its large cubage, complex routing and limited signal processing capability. According to the design of GAPAA [1, 2], here mainly discusses the implementation structure that combines the active phased array at sub-array level with multi-plane digital beamforming [3, 4].

Considering M targets, K planar array and N-element subarray, the implementation structure which combines the active phased array at subarray level with multi-plane digital beamforming can be shown as Fig. 15.2. The receiving subarray consists of N element channels, where “M targets’ amplitude and phase synthesis” is to form M targets’ signals through “M targets’ network”. Then the received signals are amplified when converted down to the low. These received M targets’ signals are delivered to “digital adaptive multi-beam-forming with multi-planar array”, and the output M baseband signals are delivered to M integrated basebands to deal with signal processing respectively. M output angular

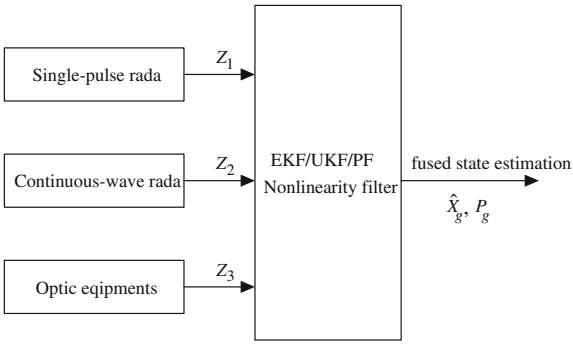

Fig. 15.2 Receiving framework of active phased array with N channels and M targets

errors are delivered to "M targets' beam controller" to complete the multi-beam-forming algorithm and angle tracking algorithm. The process of transmitting is much simpler that it only needs to transmit beamforming and antenna pointing.

15.3.3 Typical Array Arrangements and Simulations

15.3.3.1 Preliminary Simulation of Multiple Planar Arrays

Considering the ground Multi-satellite TT&C in S-Band, the aperture is set to be 10 m. At least 4 planar arrays are used to build up the hemispheric array antenna with 40000 elements. In order to realize the object of multi-target TT&C, the number of multiple simultaneous targets TT&C is chosen around 16.

The spatial gain characteristics of whole airspace are simulated and analyzed. The hemispheric array consists of 4 planar arrays with typical element antennas. Suppose the elevation of 4 planar arrays is 50°, namely, pointing to (0°, 50°), (90°, 50°), (180°, 50°) and (270°, 50°) respectively. The integrated pattern of whole airspace can be obtained through simulations, which is shown in Fig. 15.3.

The simulation results show that the gain of 4 planar arrays covers the dome's surface evenly and the normalized gain varies from −1 to 1 dB. The simulations also show that the whole airspace array has high grating lobes in some areas on the dome, which can be suppressed to be under −10 dB when using 8–16 planar arrays in synthesize pattern. Figure 15.4 illustrates the typical grating lobes pattern of the 4, 8 and 16 planar arrays, respectively. With the number of planar arrays increasing, the gain of grating lobes decreases from −5 to −19 dB.

15.3.3.2 Dome Phased Array

According to the framework of the GDPAA, the framework and beamforming technology of spherical conformal array are investigated [5]. Figure 15.5 illustrates the elements distribution of spherical conformal array. The two-dimensional

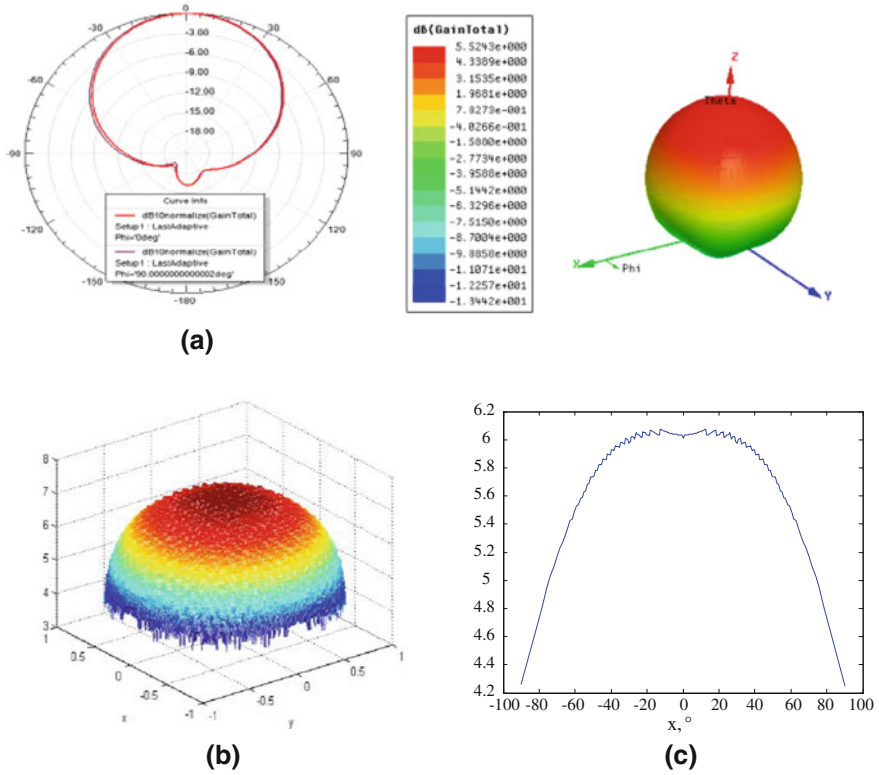

Fig. 15.3 Typical pattern of single antenna and spatial coverage characteristic of whole airspace array. **a** Pattern of single antenna. **b** Spatial coverage characteristic. **c** Slice pattern of spatial coverage

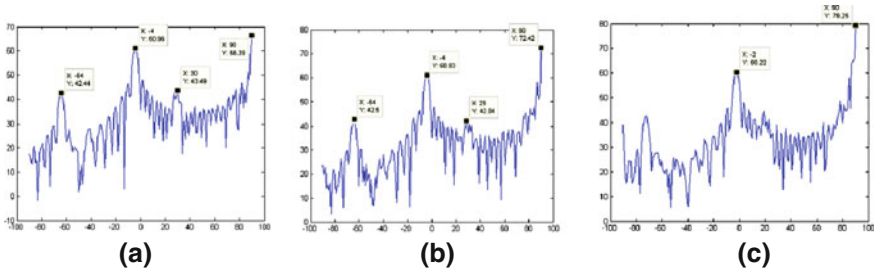

Fig. 15.4 Grating lobes with different numbers of arrays. **a** 4 planar arrays. **b** 8 planar arrays. **c** 16 planar arrays

slice pattern with v = 0 and three-dimensional beam pattern are both shown in Fig. 15.6. It is obvious that the main-lobe is narrowed and sidelobes are suppressed by compensation method. Grating lobes of spherical array combined by infinite planar arrays are non-existing in theories, but grating lobes of spherical array

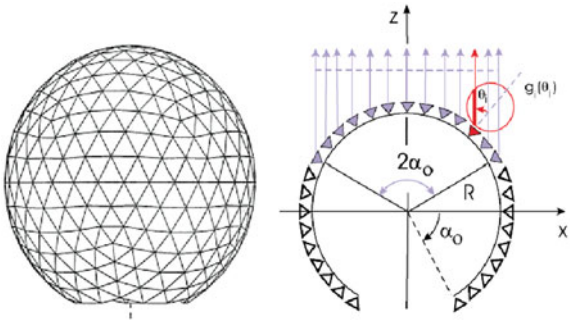

Fig. 15.5 Receiving framework of active phased array with N channels and M targets

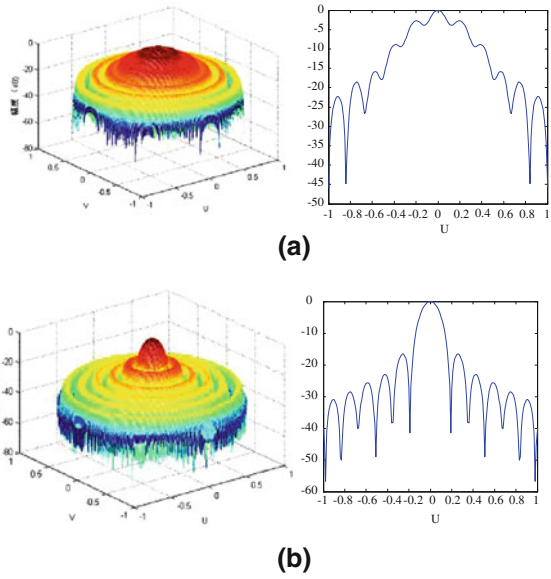

Fig. 15.6 Beam pattern of spherical conformal. **a** Beam pattern before compensated. **b** Beam pattern after compensated

combined by finite planar arrays are actually existent. However, the beam steering algorithm is complex and the structure of multi-planar spherical array is complicated.

15.3.3.3 Comparison of Array Arrangements

Approximate spherical array can cover whole hemispheric airspace and obtain better beam performance, but the electrical design, engineering and waveform network of the system structure are difficult and complex. There are some advantages to apply multi-planar arrays to cover whole hemispheric airspace in engineering.

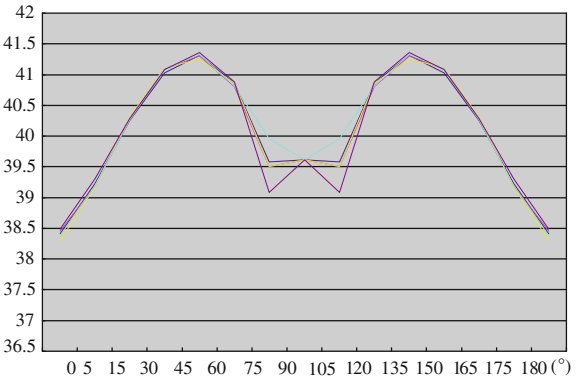

Fig. 15.7 Apple-shape pattern gain coverage with an equivalent aperture of 5 m

1. The simple design of the structure—combined by multiple planar arrays.
2. The simple processing and installation of the framework—combined by multiple planar arrays.
3. The simple beamforming algorithm—reconstructing multi-planar arrays.
4. The low risk of using multiple planar arrays technology to cover the whole hemispheric airspace—mature engineering.

Under certain grating lobes control, the decrease of the number of planar arrays is better to project implementation. The array combined by multi-planar arrays should be chosen first in engineering realization. The common problem of both spherical conformal array and the array combined by multi-planar arrays is that they need to face the beam transition from this planar array to another one when tracking flying targets in the whole hemispheric airspace.

15.3.3.4 Coverage of the Gain of 16 Planar Arrays with an Equivalent Aperture of 5 m

As shown in Fig. 15.7, when applying 16 planar arrays to synthesize, the slices in all directions have apple-shape coverage of the pattern gain. It requires low gain when the target locating above the array and high gain at low elevation. The gain reaches at least 41 dB at the range of elevation between 30° and 60°, and it also can reach larger than 39 dB at elevation 5° or 90°. The optimization design is obtained under the consideration of practical requirements of TT&C for the satellite.

Table 15.1 The performance contrast of proposed system and conventional multi-antenna system

Sequence number	Technical performance	Hemispheric airspace multi-beam phased array TT&C system	Conventional 16 antennas system
1	Antenna servo system	Generating 16 beams with one antenna	Generating single beam with one antenna
2		Tracking 16 targets with one antenna	Tracking single target with one antenna
3		Without servo system	16 servo drive systems
4		Without the difficulty of vertex tracking	Vertex tracking based on mechanical tilting
5		Without the influence of mechanical servo system time constant	The influence of mechanical servo system time constant on beam scanning
6	Antenna servo system	Without mechanical rotation component, high reliability	Many mechanical rotation components, easily abrasion, low reliability
7		Transmitting upstream signals to every satellite simultaneously	Transmitting upstream signals to every satellite simultaneously with 16 antennas respectively
8	Single station ability	Receiving every satellite's signal in hemispheric airspace simultaneously	Receiving every satellite's signal simultaneously with 16 antennas respectively
9		Eliminable interference among stations	Possible interference and shield among stations
10		One clock source, easy to process multi-target time information	Complex time information, difficult to coordinate
11	Multi-target management and cooperation	Multi-target centralized management	Multi-target decentralized management
12		Facile coordination	Difficult coordination
13		Efficient automatic management	Multiple stations, difficult to automatically manage
14	Construction cost	Low cost of equipment and construction	High cost of equipment and construction
15	Maintenance charge of whole life	One equipment, low maintenance charge	16 equipments, high maintenance charge
		Low construction cost of communication lines	High construction cost of communication lines
		Requiring small space	Requiring large space

15.3.4 Performance Contrast Between Proposed System and Conventional TT&C System

Take 16 targets for example, the paper gives a preliminary comparison between hemispheric airspace multi-beam TT&C phased array system and conventional multi-antenna system, which are summarized in Table 15.1.

Due to the hugeness, it is necessary to solve some key technologies to complete array antenna system of this scale.

15.3.5 Key Technologies Needed to be Broken

In order to fulfill the large-scale multi-beam array antenna system, it is expect to solve many key technologies as follows:

1. General technology of structure design for the large-scale array antenna system. The architecture is a significant factor in general technology of structure design. Since the large-scale electrical scanning array antenna system is different from part airspace array antenna system with small scale, choosing a proper architecture is exactly critical when considering the issues such as multi-target tracking, large-scale system, and complex-structure situation.
2. Large-scale array arrangement technology. Large-scale array arrangement technology is extremely complicate. Considering the whole airspace, multi-beam, electrical scanning and performance, a proper array arrangement scheme is a significant key to the implementation of this system.
3. Simultaneous multi-target precise angular tracking and hemispheric beams switching technology. Simultaneous multi-target precise angular tracking is the core technology of array antenna algorithms, which is solved by multi-target time-sharing angular tracking in previous phased array. However, in multi-target phased array TT&C system, it requires simultaneous multi-target precise angular tracking and measurement, seamless switching technology, which is more difficult.
4. The minimum design of beamforming network technology. The minimization of Beamforming Network is a difficult problem in implementation. Based on the level of electronic manufacture technique, the miniaturization of large-scale array antenna system is the technology bottleneck needs to be broken.
5. Precise calibration technology of large array. The precise calibration technology of array is the foundation of engineering implementation. In order to achieve expected function and performance, the multi-beam phased array algorithms must based on precise calibration.
6. Test technology of large array. The test of large array is different from any previous traditional antenna. How to complete the current test using existing technologies is the key factor to maintain the good performance.
7. General technology of structure design of the large-scale array antenna system. Because of multiple element channels and multiple target beams of large-scale array, the design cost is too high on the basis of previous technologies to realize the application in engineering. The foremost problem is solving the technology bottleneck with low design cost.

15.4 Conclusions

Compared with conventional phased array, the proposed system not only can be able to fulfill the performance of antenna with an equivalent aperture of 5, 10 m or even bigger, but also can realize hemispheric airspace simultaneous spacecraft TT&C in electrical scanning with a low cost. Because the number of simultaneous TT&C targets is no less than 16, the proposed system holds many advantages such as low average construction cost, low long-term maintenance charge, low infrastructure cost and great coordination ability.

References

1. Tomasic B, Turtle J, Liu S (2002) A geodesic sphere phased array antenna for satellite control and communication. In: International union of radio science, XXVIIth General Assembly, Maastricht
2. Liu S, Tomasic B, Turtle J (2005) The geodesic dome phased array antenna for satellite operations support. In: 18th international conference on applied electromagnetics and communications, Dubrovnik, Croatia
3. Mailloux RJ (1994) Phased array antenna handbook. Artech House Inc., Norwood
4. Zhang GY (1994) Phased array radar system. Defence Industry Press, Beijing
5. Boryssenko A, Prokhorenko V (2000) Phased-arrays effect in antennas with transient excitation. In: 2000 IEEE international conference on phased array systems and technology, pp 469–472

Chapter 16
Fast Identification and Modification of Angle Error Based on Prior Information of Velocity-Measurement Radar

Jianping Pan, Bo Qiang, Zongwei Liu, Yanan Hu and Shengxi Wang

Abstract A new method on the use of prior information is proposed to solve the problem that the accuracy of measured angles with a velocity-measurement radar is low. The method of amending the angle error of optical and pulse radar is applied to the angle data correction of velocity-measurement radar. So a method of amending error superposition of many complicated components is proposed, and the conventional method of amending angle error is improved. By eliminating the error item in measured angle data of the velocity-measurement radar, the accuracy of measured angle is improved and the measured angle data can be fully used in computing the trajectory. When the angles are amended, the steady and credible trajectory parameters can be computed. Thus, the trajectory measuring accuracy is improved.

Keywords Prior information · Angle error · Identification modification · Trajectory precision · Error model

16.1 Introduction

The Doppler radar provides not only the measurement parameter $\dot{S}$ in real-time task, but also the angle-information A and E. However, the angle measurement data is seldom effectively utilized in high precision measurement of ballistic missile. One of the reasons is the low precision of angle information measurement, the other reason is the uncorrected error terms in angle information. Therefore the

J. Pan (✉) · B. Qiang · Z. Liu · Y. Hu · S. Wang
Taiyuan Satellite Launch Center, Taiyuan 030027, China
e-mail: qiangbo23@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_16,

building of simple, rapid and effective models and methods on correcting the error terms of angle information is worthy of further study. The correcting methods and models is complete in data processing of impulse radar and optical-electronic theodolite, but the relative factors, fitting models and methods, quickly origin determination of error terms of angle information on Doppler radar are need to be verified by data processing results. This paper proposed a new method based on the error correcting methods on angle information of impulse radar and optical-electronic theodolite and the prior information of correcting models, which quickly determinate the error terms and the models applied on Doppler radar, and makes the angle information could be applied in actual combat.

16.2 Prior Information

The prior information is the information that proved to be reliable in a lot of experiments, and can be used as verification in other methods. It involves methods, math models, and experiment results. Mass data has been obtained from the application of impulse radar and optical electronic theodolite, especially in impulse radar, mature methods and models have been established and fully verified in actual combat on angle information correction. The typical angle error terms include dynamic delay error, axis error, zero error and radio wave refraction error.

16.3 Error Processing Methods on Measurement of Angle

During the process of error on measurement of angles, The error terms need to be identified firstly, and the corresponding error models should be established to correct the error value, thus the measurement data is satisfied for device precision index requirement [1].

16.3.1 Identification of Error Terms

The production of angle error is closely related to the equipments. According to the impulse radar and optical electronic theodolite, the error term of angle data mainly includes zero error, axis error (azimuth axis and pitching axis non vertical error, large cap non horizontal error, light axis and pitching axis non vertical error), light axis and electrical/mechanical axis mismatch error, gravity drop error, dynamic delay error, radio wave refraction error, time error, frequency error and velocity of light misfit error.

16.3.2 Establishment of Error Model

The error correction model of measured data is the mathematical expressions describing the variation of system errors, the key of establishment of the model is to obtain the structure and coefficients of every error term in the model. Different measurement device, tracking mode and measurement system need different error correction model [2].

16.3.2.1 Dynamic Delay Error

The dynamic delay means the electrical axis can not completely aim at the target, and have a small delay or leading angle to the motion of the target. It leads to measuring error of the angular displacement, velocity and acceleration on the relative motion between radar antenna and target [3]. The dynamic delay error is mainly caused by the variation of the angular velocity, namely the relative angular acceleration between target and radar, and they are in the direct ratio. In practical alignment calibration, the dynamic delay error of the corresponding point $\Delta\theta$ can be expressed by tracking angle error voltage ΔU and the directional sensitivity of the voltage μ, $\Delta\theta = \frac{\Delta U}{\mu}$.

The analysis shows that, different error origin and error term have different influence on the trajectory results, and the same error term produce different influence on target in different position and status. The errors which have the same magnitude may be caused by different origin or target in different position and dynamic characteristic, so in practice, we should make concrete analysis on concrete conditions.

16.3.2.2 Radio Wave Refraction Error

Radio wave refraction is the effect induced by radio wave propagation velocity variation in the atmosphere due to the variation of the atmospheric refraction index n according to different space (mainly according to height). The data measured by radar is not the real angle, distance and the variation of distance, the difference can be calculated according to the following formula [4], then the measured data could be compensated.

$$\Delta E_N = N_S^0 \cot E_c - \frac{N_S^0 (R + R_0 \sin E_c) \cos E_c}{CRR_0 \sin^3 E_c} (1 - e^{-20000c}) \tag{16.1}$$

The angle unit in the formula is radian, N_S^0 is the ground refractive index, C is the exponent of refractive index's variation to height, R_0 is the radius of the earth, R is the real distance of the target, E_c is the measured value of the pitch angle

export by the coder of antenna's axis angle. N_S^0, C is provided by the weather department.

According to the previous analysis, we can reach the following formula:

$$A, E_{measured\ value} = A, E_{real\ value} + A, E_{compensation\ value} \tag{16.2}$$

After the compensation of measured data, we can get the system error model of the measured angle data from the following formula:

$$\begin{aligned} A = {} & A_c + A_0 + \theta_m \sin(A_c + A_0 - A_m) \tan(E_c + E_0) \\ & + \delta \tan(E_c + E_0) + (K_z + K_g + k.\frac{\Delta U_a}{\mu_a}) \sec(E_c + E_0) \end{aligned} \tag{16.3}$$

$$\begin{aligned} E = {} & E_c + E_0 + \theta_m \cos(A_c + A_0 - A_m) + K_e \\ & + E_{mg} \cos(E_c + E_0) + k.\frac{\Delta U_e}{\mu_e} - \Delta E_N \end{aligned} \tag{16.4}$$

The symbols in the above formula are defined as following: A is the real value of the target's azimuth angle, E is the real value of the target's pitch angle, A_C is the measured value of azimuth angle provided by the coder of antenna's axis angle, E_C is the measured value of azimuth angle provided by the coder of antenna's axis angle, R_0 is the zero value of the antenna's distance, A_0 is the zero value of the azimuth angle; E_0 is the zero value of the pitch angle, θ_m is the non horizontal value of the big platform, A_m is the azimuth angle corresponding to the biggest non horizontal space of the big platform, δ is the degree of the non-orthogonality between azimuth axis and pitch axis; k_z is the transverse component coefficient of the mismatch between light and electrical axis. k_e is the vertical component coefficient of the mismatch between light and electrical a, k_g is the transverse component coefficient of the mismatch between light and mechanical axis, E_{mg} is the error coefficient of distortion caused by gravity, k is the constant 0.06, μ_a is the directional sensitivity coefficient of azimuth branch angle, μ_e is the directional sensitivity coefficient of pitch branch angle, ΔE_N is the compensation of radio wave refraction, ΔU_a is the azimuth error voltage, ΔU_e is the pitch error voltage.

16.4 Numerical Calculation Example

We use simulated data to evaluate the proposed method, firstly we use the trajectory based on GPS or other high precision trajectory as standard, then we calculate the angles of every Doppler radar from the trajectory, and make subtraction between the calculation results and the measured data, the comparison result is described in Fig. 16.1.

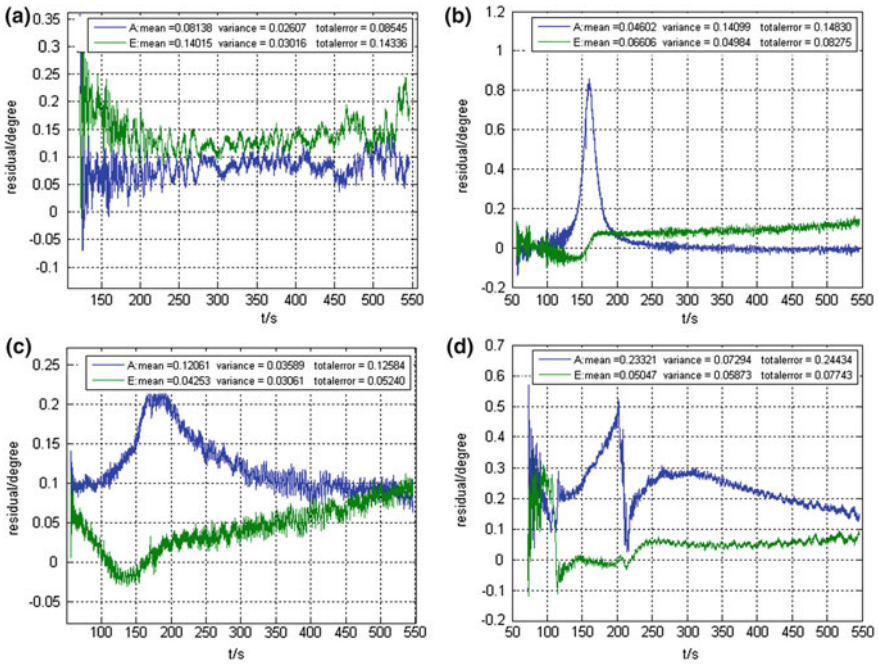

Fig. 16.1 Several typical angle error types

From Fig. 16.1, we can find that, after the data recovery of the measured angles of Doppler radar without any compensation, the errors are different. The typical 4 different error types are described in Fig. 16.1.

Type (a) error in Fig. 16.1 can be taken as the radio wave refraction error from the prior information of the impulse radar error terms. After the compensation of radio wave refraction, the data is stationary, and the precision satisfies the requirement, the comparison of the error between before and after the compensation is described in Fig. 16.2.

Type (b) error in Fig. 16.1 can be taken as the dynamic delay error from the prior information of the impulse radar error terms. After the compensation of dynamic delay, the data is stationary, and the precision satisfies the requirement, the comparison of the error between before and after the compensation is described in Fig. 16.3.

Type (c) error in Fig. 16.1 can be taken as the axis error from the prior information of the impulse radar error terms. After the compensation of axis error, the data is stationary, and the precision satisfies the requirement, the comparison of the error between before and after the compensation is described in Fig. 16.4.

Type (d) error in Fig. 16.1 can be taken as the cooperation of axis and dynamic delay error from the prior information of the impulse radar and optical electronic theodolite error terms. To solve the problem of several type of errors' superposition

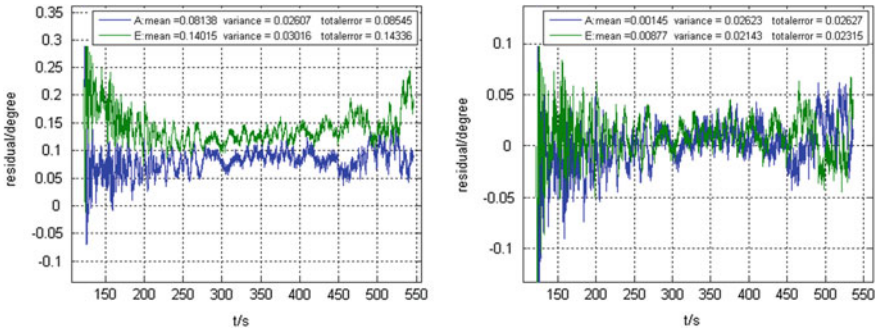

Fig. 16.2 The comparison between before and after the compensation of radio wave refraction

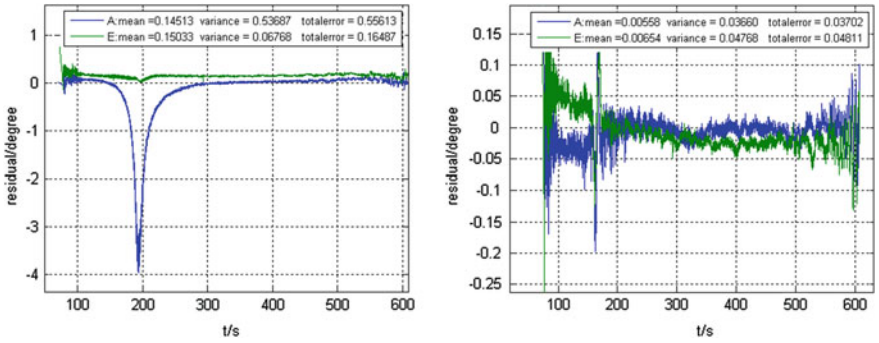

Fig. 16.3 The comparison between before and after the compensation of dynamic delay

that results in the error of angle data error, the original angle error correction model [5] is:

$$\begin{aligned} A = {} & A_c + A_0 + \theta_m \sin(A_c + A_0 - A_m) \tan(E_c + E_0) \\ & + \delta \tan(E_c + E_0) + \left(K_z + K_g + k. \frac{\Delta U_a}{\mu_a} \right) \sec(E_c + E_0) \end{aligned} \tag{16.5}$$

We improve the error model as:

$$\begin{aligned} A = {} & A_c + A_0 + \theta_m \sin(A_c + A_0 - A_m) \tan(E_c + E_0) \\ & + \delta \tan(E_c + E_0) + \left(K_z + K_g + k. \frac{\Delta U_a}{\mu_a} \right) \sec(E_c + E_0) \\ & + b \times \tan(E_c) \end{aligned} \tag{16.6}$$

In formula 16.2, we take $b \times \tan(E_c)$ as the correction term. It mainly references optical electronic theodolite t correction terms. The coefficient b can be adjusted to the concrete condition, it is mainly set for compensating axis and

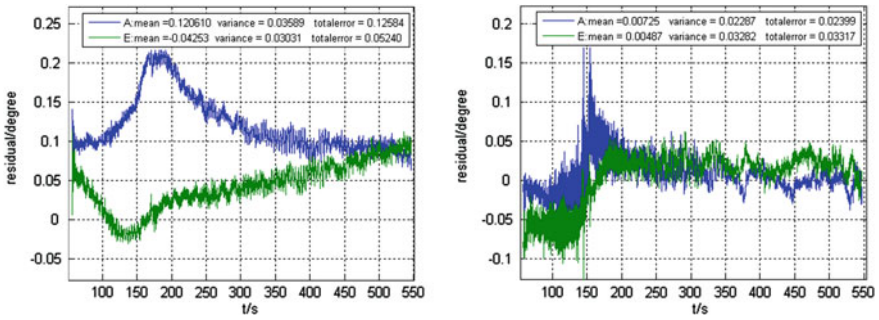

Fig. 16.4 The comparison between before and after the compensation of axis error

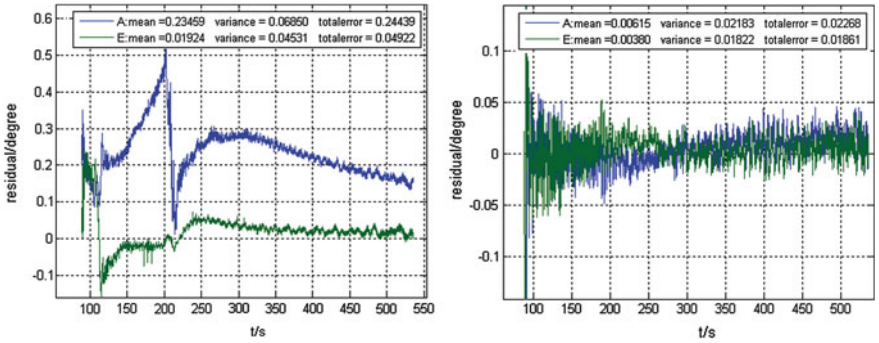

Fig. 16.5 The comparison between before and after the compensation of several type of error's superposition

dynamic delay errors. After the error compensation mentioned above, the data is stationary, and the precision is satisfied what the index requires, the compassion of the error between before and after the compensation is described in Fig. 16.5. It shows that the effect of the compensation method proposed is good.

According to the prior information correction on impulse radar and optical electronic theodolite, the 4 typical error terms in Fig. 16.1 are estimated and corrected. After compensation, the error shows white noise and no trend error, the precision is satisfied what the index requires. Therefore the effect of data correction is good and the corrected angle data can participate in the trajectory calculation, the calculation result is showed in Fig. 16.6.

Figure 16.7 describes the comparison of trajectories from GPS data and corrected $3\dot{S} + 3AE$ data of 3 Doppler radars. It shows that the trajectory could be acquired definitely via the velocity and angle data of 3 Doppler radars, and the precision is better than calculated result only based on the data on velocity.

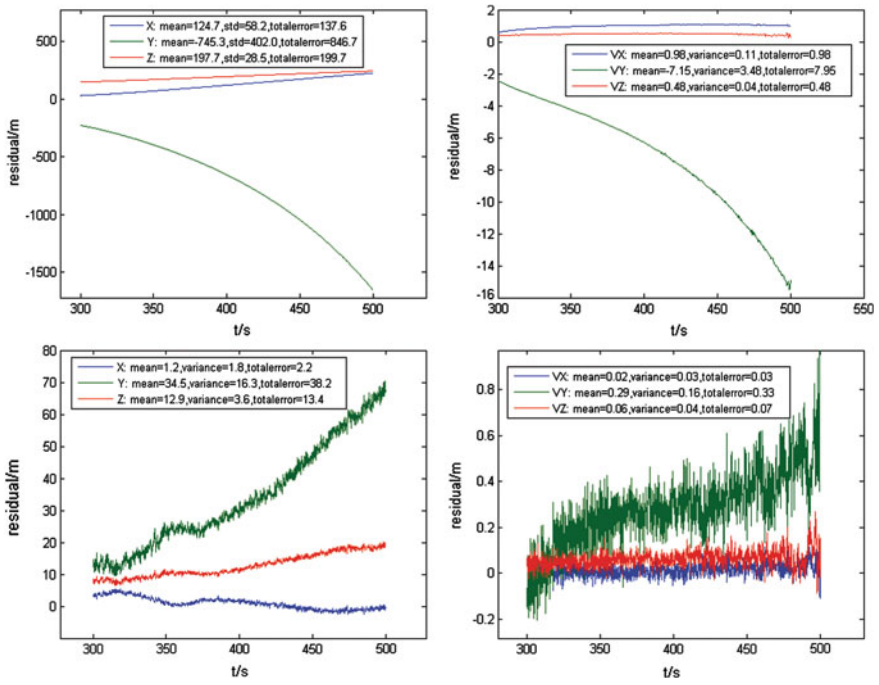

Fig. 16.6 Comparison of trajectories calculated by modified/unmodified angle data of Doppler radar. **a** Position. **b** Velocity

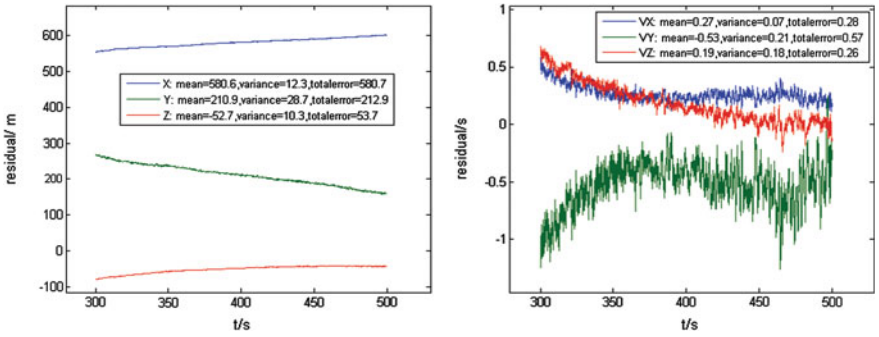

Fig. 16.7 Comparison of trajectories. **a** Position. **b** Velocity

16.5 Application Prospect

In this paper, we propose a new method to correct the angle data in real-time task for emergency, and improved the precision and reliability in real-time data process. The method can also be used on recent developed devices to quickly confirm

the error origins, and then correct the error to make new devices join real combat as soon as possible.

16.6 Conclusion

The errors of angle data on Doppler radar mainly include dynamic delay error, axis error, zero value error, radio wave refraction error. Here we proposed an angle modification method based on the prior information which can correct several complex error's superposition, improves the traditional angle modification method, and increase the precision of angle data measurement. In practice, measured angle data from 3 Doppler radars were utilized to calculate the trajectory, and the result's precision is stable and reliable, in that it can be used in real time task as introductory information. New real-time data process scheme is put forward, and the method is easy to realize. The method we propose in this paper can also be used in trajectory confirmation for the "One Principle and Three auxiliary" Doppler radar system, especially it is suit for experiments which have few measurement stations and low precision requirements in trajectory calculation, which require calculation of trajectory parameters of fall point.

References

1. Wang ZM, Zhou HY, Wu Y et al (1999) Calibration and estimation trajectory track data. National University of Defence Technology Press, Changsha, pp 452–465
2. Liu LS (2000) Post-flight data processing of trajectory measurement. National Defense Industry Press, Beijing, pp 265–273
3. Wang M, Shu LY, Hu SL (2008) Analysis and application of error influence on trajectory measured data. National Defense Industry Press, Beijing, pp 50–56
4. Hu SL, Wang M (2007) Data process technology on impulse radar measurement data. National Defense Industry Press, Beijing, pp 23–26
5. Zhao WC, Yao J, He MK (2004) Trajectory positioning and velocity determination based on angle measurement and height information fusion. J Spacecr TT&C Technol 26(4):54–57

Chapter 17
Optimized Simulation Analysis of Netted Ground-Based Radars for Near Space Vehicle

Daqing Chen, Dan Liu, Rongchun Wang and Zhe Zhang

Abstract Firstly the elements, functions and flight features of the Near-Space Vehicles (NSVs) are analyzed in this paper. Then the problems on tracking, telemetry and control (TT&C) of the NSVs and the requirements of the Ground-Based Radars for NSVs objects are analyzed. Based on the deployment scheme of the future Ground-Based Radars, the optimized usage of the TT&C resources of the ground navigation area is designed and simulated to fulfill the needed accuracy information of TT&C for the flight of the NSVs under limited resources condition. Several deployment schemes are simulated and optimized in this paper. The simulation results show that through deployment optimization to existing schemes, the performances, i.e., efficiency and resource consumption, are improved.

Keywords Radar · Near-space vehicle · Simulation · Netted ground-based radar

17.1 Introduction

Near Space is the region of Earth's atmosphere that lies between 20 and 100 km above sea level, and another definition is above which a commercial airliner flies but below the realm of an orbiting satellite, encompassing the stratosphere, mesosphere, and thermosphere. Due to the reasons of technology and knowledge about the Near Space, its strategic value has been received additional consideration until recently. Currently, several Near-Space vehicle design schemes are proposed

D. Chen (✉) · R. Wang · Z. Zhang
Taiyuan Satellite Launch Center, Taiyuan 030027, China
e-mail: dqchen64@sohu.com

D. Liu
No.23 Institute of the Second Academy, China Aerospace Science and Industry Corporation, Beijing 100854, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_17,

by many countries and the researches focus on stratosphere blimps (non-rigid airships), high altitude balloons and high-altitude long-endurance unmanned aerial vehicle (UAV) [1]. The whole flight course, including launching, flight, and landing, are monitored by the NSV TT&C system. However, the working range TT&C systems are built on the base of conventional orbits dynamics. Since the motion of NSVs in atmosphere is a kind of powered flight, which means it belongs to aviation aerodynamics and its orbit cannot be determined by conventional orbits dynamics [2]. This is the new challenge to current normal TT&C technology of missile range.

17.2 Technical Issues in TT&C of NSV

17.2.1 TT&C in Blackout Area

When a vehicle flies in Near Space with very high speed, the high-pressure and high-temperature shock wave will ionize the surrounding atmosphere, producing a plasma field and forming a plasma sheath, which will attenuate the radio wave and cause communication blackout [3]. This area is called blackout area. Black phenomenon has great impact on ground TT&C, thus the mitigation of blackout is a very important problem, although at the same time it is very difficult. The optional methods include increasing the operating frequency of RF devices and adding a static magnetic field outside the plasma sheath.

17.2.2 TT&C of Multi-Target

When the NSVs perform a regional mission, commonly they are formation flying [4].Consequently, the TT&C system must have the ability to deal with several targets. The methods of multi-target TT&C include: time division same-beam TT&C, code division same-beam TT&C, frequency division same-beam TT&C, time division multi-beam TT&C and ubiquitous multi-beams TT&C, etc.. These methods have their merits and drawbacks and among them the time division and the ubiquitous multi-beams technologies in multi-target TT&C are widely investigated recently.

17.2.3 Multipath Interference

For TT&C of the NSVs, the multipath interference should be considered since the flight altitude of the NSVs is lower than satellites, when the elevation angle is low the RF reflection by the ground non-target objects will affect the TT&C of the

targets. The multipath interference will cause the amplitudes and frequencies of the received signals fluctuating. If the receiver cannot adapt the high dynamic variation, it will be inoperable, i.e., losing lock. Thus increasing the dynamic range is the basic method for anti-multipath interference.

17.2.4 Whole Course TT&C

The NSV experiments need high percentage of coverage and whole course TT&C. Currently, the space-based and ground-based joint TT&C system is researched to perform whole course tracking of the NSVs. The method used in ground systems is multi-station relay, which is widely used in range experiments of China. Through reasonable optimization of deployment scheme the existing ground TT&C systems can basically meet the experiment phase for NSV design. In operation phase, the NSVs are in the low orbit, the space-based system based on data-relay satellite system should be developed. According to the conditions of China, the regional tracking and data relay satellite system need to be prior developed to realize the whole course TT&C. Therefore, making full use of the existing range TT&C resources and optimizing the deployment scheme of netting ground-based radars is a major problem for NSV TT&C.

17.3 Optimizing and Simulating of Ground-Based Radars Netting

A netted ground radar system which contains multiple spatially diverse monostatic radar components with a shared area of coverage performs searching, tracking, and identifying through observations from these netted radars. An optimized deployment scheme of netted ground radar can significantly improve the performance. Therefore, the deployment problem of netting radars is the key factor in increasing the performance of the netted ground radar for NSV TT&C.

17.3.1 Deployment Optimizing Simulation

Assumptions in simulation:

1. The system error can be eliminated through calibration, thus the system error can be ignored.
2. Generally the elevation angle range with sufficient accuracy is between 3° and 73°, therefore the radar positioning should consider the operational elevation angle range.

3. Dynamic delay can be eliminated by post processing so that the dynamic delay error can be ignored caused by approach point.

According to these assumptions, the approach point distance and the detection range are related to the flying altitude of the target.

When the detection equipment is a monostatic radar, different positioning scheme has different measurement accuracy. If the whole ballistic measurement accuracy is the key factor, the measurement quality of the whole ballistic is determined with the detection error range. If focusing on the measurement accuracy of some key point or some sector of the course, the detection error range of the point or sector is used. In a monostatic radar simulation, the detection range and the error analysis of the measurement data are mainly considered [5].

When two radars are used, an optimized data fusion algorithm is proposed and the data error related to different deployment scheme is simulated and compared. Finally the optimal deployment mode is obtained according to the simulation results.

17.3.2 Bistatic Data Fusion Algorithms

In order to improve the accuracy, a bistatic data fusion algorithm is proposed, which is summarized as follows [6]:

Data fusion is conducted according to the root mean square errors (RMSEs) of the two data groups under Cartesian coordinate. Assumed that the RMSE of the 1st radar for some point target is (DX1, DY1, DZ1) and (DX2, DY2, DZ2) for the 2nd radar. Fusion in three coordinate axis, i.e., X, Y, Z, respectively. Taking the fusion process of axis X for example, the RMSE need to reach its minimum for higher accuracy. According to the probability theory, the variance of the sum of two random variables equals to the sum of the two variances of these two random variables.

One observation of the two radars can be denoted as x1 and x2, which equal to the sum of their corresponding true value and random error respectively. Thus the fusion results of axis X is $kx1 + (1 - k)x2$, where k is the fusion coefficient. Its variance is $D(kx1 + (1 - k)x2) = D(kx1) + D((1 - k)x2)$, i.e., $k^2D(x1) + (1 - k^2)D(x2)$. If $D(x1) = m\,D(x2)$, the variance after data fusion can be simplified as $(k^2m + (1 - k^2))D(x2)$. When the derivation with respect to k equals to 0, the variance reaches its minimum, i.e., $2\,km + 2(k - 1) = 0$. Therefore, the accuracy reaches its highest point when $k = \frac{1}{m+1}$.

In the following, different deployment schemes in some scenario are simulated and the optimal station distribution can be obtained according to the data fusion errors.

17.4 Analysis of Experiment Results

The simulation scenario definition: the target is flying at a constant altitude with the altitude 8,000 m and the speed 300 m/s. The detection instrument is radar with the range error 5 m (1σ), the azimuth error and elevation error 0.5 mrad (1σ), respectively. The data rate is 20 Hz. The rectangle coordinate is northern east coordinate with positive x axis pointing northward (corresponding bearing angle 0°), y axis vertical upward, positive z axis pointing eastward (corresponding bearing angle 90°). The composite error can then be defined as $\Delta = \sqrt{\delta x^2 + \delta y^2 + \delta z^2}$ [7]. The composite error of one sector can be evaluated with the mean of several points, i.e., $\Delta = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\delta x_i^2 + \delta y_i^2 + \delta z_i^2}$.

17.4.1 Simulation of Monostatic Scenario

Assumed the radar is located on ground, the distance between the radar and the geocenter is re + hr (the radius of the earth plus the absolute altitude of the radar), and the distance between the target and the geocenter is re + ht (the radius of the earth plus the altitude of the target), the distance between the radar and the target is R. When the elevation angle e $\geq$ 3°, the maximum tracking range is $R_{\max}$. When the elevation angle e $\leq$ 73°, the minimum tracking range is $R_{\min}$. R, Re + ht and Re + hr meet the law of cosine, i.e.,

$$R^2 + (\mathrm{Re} + hr)^2 + 2R \cdot (\mathrm{Re} + hr)\sin(e) = (\mathrm{Re} + ht)^2 \tag{17.1}$$

If ht = 8,000 m, $R_{\max} = 128.28$ km and $R_{\min} = 8.365$ km, the operational range of the radar is 8.365–128.28 km.

If the approach point is zero and the target flies away from the radar, the composite error can be shown in Fig. 17.1.

The average error with the elevation angle from 73° to 3° is 37.4 m.

17.4.2 Simulation of Multistatic Scenario

The NSV TT&C mission is completed by netted ground radar through several radar relaying. Herein the optimal deployment scheme is obtained by comparing the fusion errors of different schemes.

1. If the two radars observe the same flight path from the same perspective, after data fusion, the random error will be $1/\sqrt{2}$of that of single radar. The calculated composite error is shown in Fig. 17.2.

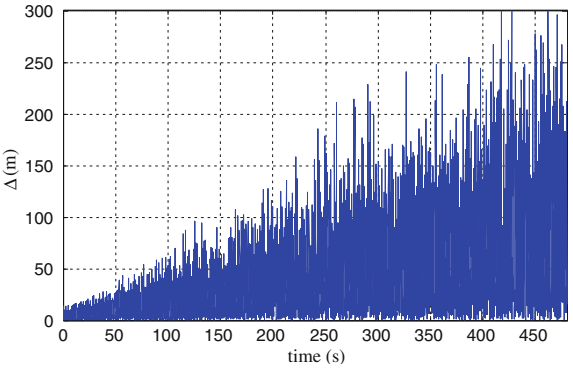

Fig. 17.1 The composite error of simulated radar data

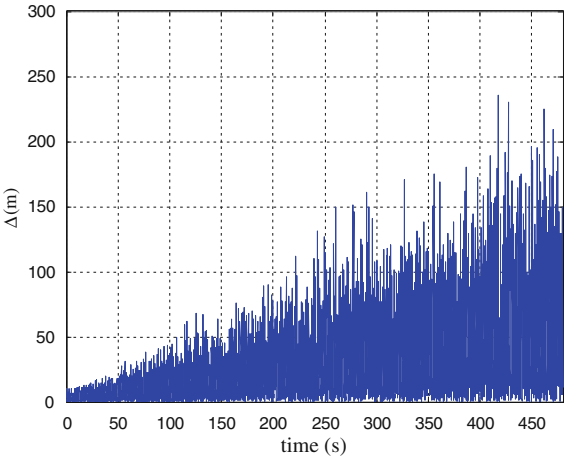

Fig. 17.2 Composite error of bistatic data fusion (from the same perspective)

2. If the two radars are deployed at the start position and end position of the flight path, the composite error is shown in Fig. 17.3.
3. If the 2nd radar is located at the middle of the flight path, the distance between the projections on the ground of two approach points is 2.5 km, the errors on axis X, Y and Z after data fusion are shown in Fig. 17.4.
4. If the two radars are deployed at 1/3 and 2/3 of the flight path and the distance between each radar and the approach point is 2.5 km, the composite error is shown in Fig. 17.5.

The average composite errors of these deployment schemes are listed in Table 17.1.

From Table 17.1 we can see that the average composite error reaches its minimum value when the two radars are deployed at 1/3 and 2/3 of the flight path, therefore this deployment scheme is optimal.

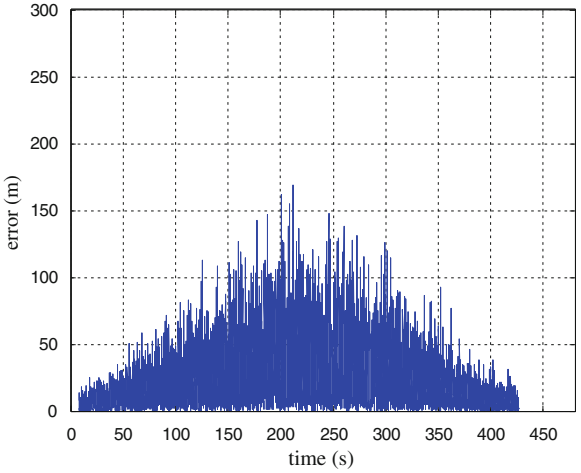

Fig. 17.3 Composite error of bistatic data fusion (Radars are deployed at the start position and end position of one segment of a flight path)

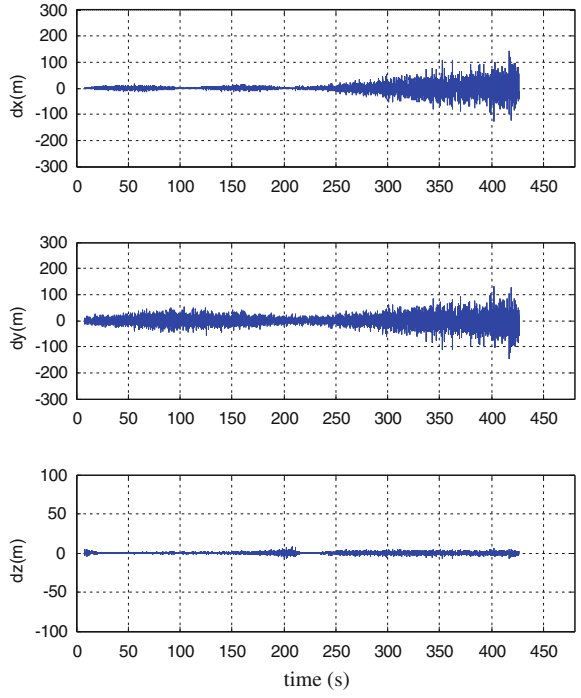

Fig. 17.4 Fusion errors on axis X, Y and Z (the 2nd radar is deployed at the middle of the flight path)

17.4.3 Conclusions from the Experiment Results

From the monostatic radar simulation, we can obtain the tracking error and the data error, thus the system indicator and the accuracy indicator can be determined and the appropriate TT&C facilities for NSV can be selected from the existing

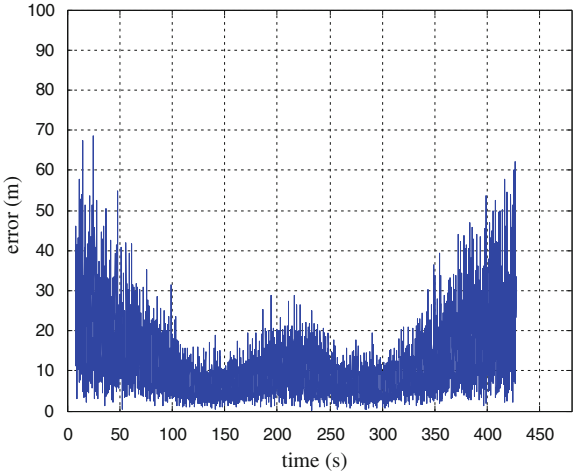

Fig. 17.5 Composite error of bistatic data fusion (Radars are deployed at 1/3 and 2/3 of the flight path)

Table 17.1 The average composite errors of different deployment scheme

	Deployment mode	The average composite errors ($3° \leq e \leq 73°$) (m)
1	The two radars observe the same flight path from the same perspective	26.4
2	The two radars are deployed at the start position and end position of the flight path	22.05
3	The 2nd radar is located at the middle of the flight path	18.5
4	The two radars are deployed at 1/3 and 2/3 of the flight path	11.95

TT&C resources. The optimal deployment scheme can be obtained from the simulation results of different bistatic deployment layout and this method can be extended to the optimization of multistatic radar scenario. Through the optimization of measurement and control facility resources, the requirement of NSV TT&C can be met.

17.5 Conclusions

The netting of ground radar for TT&C of NSV is different with the traditional netted radar and there are still many questions to be solved. In this paper the requirements to ground radar for TT&C of NSVs are analyzed. Considering the characteristics of the NSVs, a data fusion algorithm of the netted radar is proposed. The simulations focus on the optimization of the ground TT&C resources based on

the future ground radar deployment scheme, so that the needed accuracy information of TT&C for the flight of the NSVs under limited resources condition can be fulfilled. The simulation results show that through appropriate optimization of the former TT&C scheme, the efficiency and resource consumption of the netted radar can be improved. The optimal deployment scheme of netting ground radars for NSV TT&C is obtained, as a result, the subsequent vehicle experiments are guaranteed.

References

1. Mollanejad A, Khanli L M, Zeynali M (2010) Dynamic base station repositioning using genetic algorithm in wireless sensor network. In: Second international conference on computer engineering and applications (ICCEA), Bali Island, March 19–21, pp 521–525
2. Stereotyping KJ (2000) Improving particle swarm performance with cluster analysis. In: The congress on evolutionary computing, vol 2, La Jolla, Canada, July 16–19, pp 1507–1512
3. Zhou H, Chen WC, Yin XL (2002) Optimization of glide trajectory for a hypersonic vehicle. J Beijing Univ Aeronaut Astronaut 20(2):513–517
4. He YF (2007) The military applications of the near-space platform. Natl Defense Sci Technol 6:33–34
5. Liu CL, Ruan P, Xiong RS et al (2006) The synthetical analysis of error of intersection measuring multiple trajectory. J Xi'an Inst Technol 32(5):97–100
6. Wang ZM, Yi DY (1999) Calibration and evaluation of ballistic tracking data. National University of Defense Technology Press, Changsha
7. Zhao J, Zeng JC (2010) A virtual centripetal force-based coverage enhancing algorithm for wireless multimedia sensor networks. IEEE Sens J 10(8):1328–1334

Chapter 18
Techniques of High Efficiency and Linearity Transmitter

Tao Cao, Rong Zeng and Youjiang Liu

Abstract Linearity and power efficiency of transmitter are the two major concerns in designing modern wireless communication system. Consequently various competitive and complementary technologies have been investigated. The design techniques of high efficiency power amplifier and advanced linearity transmitter architecture are generally introduced in this paper, and the implementation of several efficiency enhanced power amplifiers, such as Doherty power amplifier, inverse class E power amplifier with series tunable parallel resonant tank, multi-band high efficiency power amplifier based on CRLH transmission lines and wideband (0.9–2.7 GHz) high efficiency power amplifier, are also presented. Furthermore, digital predistortion is used to increase the linearity of inverse class E power amplifier, and a high degree of spectral suppression for out-of-band is achieved. Moreover, the telemetry transmitter, which is suitable for 2 Mbps PCM signal, is developed based on high efficiency class E power amplifier. The results show the switching mode amplifiers have potential application in improving the power efficiency of telemetry transmitter.

Keywords High efficiency · Digital predistortion · Doherty · Class E power amplifier · Transmitter

18.1 Introduction

Due to the ever-increasing demands for communication, wireless connectivity, industrial power, and power conversion technology, the demand for high efficiency, radio frequency power amplifiers has never been greater. Wireless market

T. Cao (✉) · R. Zeng · Y. Liu
Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621900, China
e-mail: caotaog@gmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_18,

demands high efficiency amplifiers for portable units in order to extend battery lifetime. Wireless base stations and other high power RF transmitters require improved efficiency in order to increase reliability and lower the size and cost of heat-sinking. It is desirable to obtain high RF power amplifier efficiency in many practical applications.

On the other hand, wireless transmitter will need to adapt to handle the new requirements of the next generation of wireless system (4G). Wider bandwidths, higher bit rates, multiple antenna structures and new bandwidth efficient Orthogonal frequency Division Multiplexing (OFDM) modulation schemes are required [1]. As a result, these signals are with high peak-to-average power ratio (PAPR), and thus requiring linear power amplifiers to transmit them. Unfortunately, there is usually a trade-off between amplifier efficiency and amplifier linearity.

In order to meet different demands, the technologies of high efficiency and linearity transmitter must be investigated. In this paper, the design techniques of high efficiency power amplifier and advanced linearity transmitter architecture are generally introduced. The implementation of several efficiency enhanced power amplifiers and experiments of predistortion linearization are also presented. Furthermore, the potential of switching mode amplifiers in significantly improving the power efficiency of telemetry transmitter is demonstrated clearly.

18.2 High Efficiency RF Power Amplifier

The efficiency enhancement of traditional power amplifier depends on the decrease of conduction angle. The conduction angle of class AB, class B and class C amplifier reduces in turn, whereas efficiency of the amplifier rises. The efficiency cannot reach 100 % in reality because the output power cannot be zero. Therefore, designers must investigate into technologies of high efficiency power amplifier to achieve both lower dissipation and higher output power.

Doherty amplifiers (DPA), due to their properties of high efficiency and relatively better linearity at deep back-off region, have been widely adopted by communication equipment manufacturers [2]. The basic idea of the DPA is to properly combine two active devices, namely Main and Auxiliary, and exploit the active load modulation concept. The two amplifiers are combined at the output through an impedance inverter network. In the low power region, the Main amplifier only is active while the Auxiliary amplifier is kept off. In the medium power region both amplifiers are operating and the current provided by the Auxiliary device modulate the load seen by the Main device, in order to force such amplifier to operate in its maximum efficiency condition. Figure 18.1 shows Doherty amplifier test board of Freescale Inc. The amplifier circuits achieve 63 W output power with efficiency more than 45 % at 1.8–1.9 GHz. In order to improve linearity of Doherty amplifier, a baseband digital predistortion (DPD) is usually implemented.

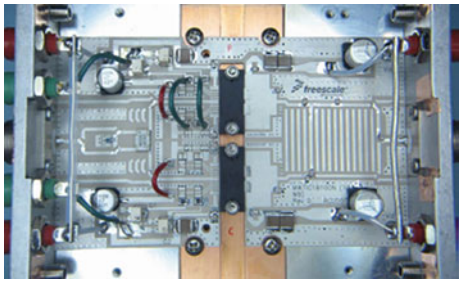

Fig. 18.1 Doherty amplifier test board of Freescale Inc

For power-amplifier and power-inverter applications, switching amplifiers such as class E and class F offer high efficiencies at high power densities. By operating the active device as a switch rather than a controlled current source, the voltage and current waveforms can, in principle, be made to have no overlap, reducing the theoretically achievable device dissipation to zero.

The class E amplifying approach has been developed mainly as a time-domain technique, with the active device treated as a nearly ideal switch [3]. Specifically, the switch is assumed to be effectively open circuit during the "off" period and a perfect short-circuit during the "on" period, and that the time required to switch between states is effectively zero. The drain voltage and current waveforms are designed with no overlap, which leads to 100 % drain efficiency. The class F approach has been developed in the frequency domain as a means of increasing the efficiency of class-AB and class-B amplifiers [4]. It is assumed that the amplifier is only partially compressed so that a relatively small number of harmonics have been generated at the drain. The ideal load network of class F amplifier short circuits even harmonics and open circuits odd harmonics, so the drain voltage waveform is square wave. This flattening eliminates the voltage across the transistor during the time it is conducting, thus increasing the efficiency. To date, the switching mode amplifiers can deliver more than 10 W output power with drain efficiency of 80 % at S band. Figure 18.2 shows high efficiency class E power amplifier module of Cree Inc. The amplifier obtains 10 W output power with efficiency of 83 % at 2.14 GHz.

18.3 Architectures of High Efficiency and Linearity Transmitter

In order to improve spectrum efficiency, non-constant envelope modulation formats such as QAM, WCDMA and OFDM are used. As a result, these signals are with high peak-to-average power ratio (PAPR), and thus are more sensitive to PAs nonlinearity. Unfortunately, there is often a trade-off between amplifier efficiency and amplifier linearity. To fulfill the stringent linearity requirements of wireless communication standards while achieving high efficiency, designers propose some advanced transmitter architectures, such as digital predistrotion (DPD), linear

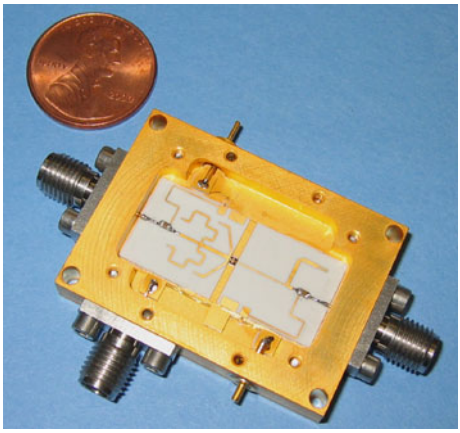

Fig. 18.2 The high efficiency class E power amplifier module of Cree Inc

amplification using nonlinear components (LINC), and polar transmitter. It is generally believed that switching-mode power amplifiers which can offer unity theoretical efficiency are the most promising solution in these transmitter architectures.

Digital predistortion (DPD) linearization has been used for many years, particularly in the microwave communication industries [5]. With linearity enhanced, PA can operate closer to its saturation region with higher power efficiency. For instance, the base station transmitter using Doherty PA with DPD technique can obtain work efficiency of 40 %, and meanwhile satisfy the linearity index of various microwave communication system. Moreover, in order to achieve better performance of efficiency, linearization for the high efficiency switching mode amplifier is also explored [6]. Figure 18.3 shows the digital adaptive predistorter block diagram. The predistorter is a device that precedes a nonlinear device such as an RF power amplifier. The magnitude of the predistorter gain increases when the magnitude of the power amplifier gain decreases, and the phase of the predistorter gain is the negative of the phase of the power amplifier gain. The result is that the magnitude and phase of the gain of the two devices in cascade becomes approximately a constant until the power amplifier reaches saturation. If the digital predistorter is designed to adjust itself automatically to cancel the nonlinearity of the power amplifier, a feedback path must be provided. This is accomplished by sampling the output of the power amplifier, downconverting the signal to baseband with the same local oscillator used for the upconversion, and converting the signal to digital with analog to digital converter. Anti-aliasing filters are necessary before the conversion to analog to prevent unwanted signals or noise from creating alias responses. DPD can provide good linearity without efficiency reduction using the exact behavioral model and memory compensation at the digital domain, but the algorithm is complicated and the solution is expensive.

The Chireix outphasing amplifier is another approach for the simultaneous realization of high efficiency and high linearity amplification, it is also known as linear amplification with nonlinear components (LINC), which was first introduced by

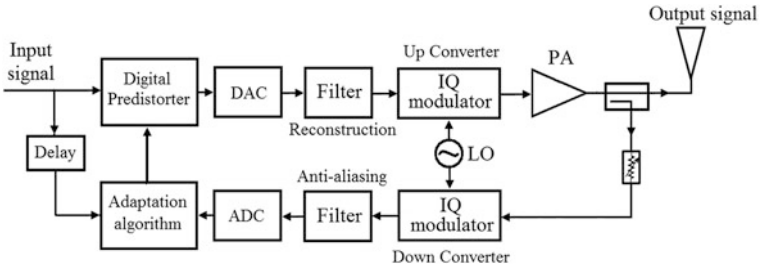

Fig. 18.3 The digital adaptive predistorter block diagram

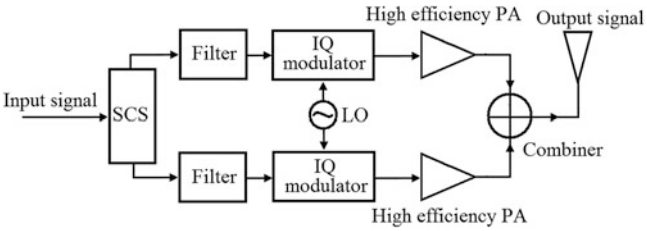

Fig. 18.4 The simplified LINC block diagram

Chireix [7] and further elaborated upon by Cox [8]. In the outphasing system, an amplitude and phase modulated input signal is decomposed into two constant envelope phase modulated signals. Highly efficient nonlinear PAs can then be used to amplify the constant envelope signals without traditional AM-AM or AM-PM distortion taking place in the individual branches. An amplified reconstruction of the original signal is obtained by summing the amplified branch signals in a passive power combiner. Figure 18.4 is the simplified LINC block diagram. Power combining is one of the major issues in outphasing systems. Conventional isolating Wilkinson combiners are not suitable since they lose most of the efficiency benefit offered by the nonlinear high efficiency PAs. Generally, the power-combining efficiency degrades rapidly as the crest factor of the original input signal grows. This problem can be avoided to a certain extent, although at the expense of linearity, by using a nonisolating power combiner structure. The resulting linearity deterioration can be compensated with careful calibration and predistortion algorithms.

Envelope Elimination and Restoration (EER) is a relatively old but attractive transmitter technique for high efficiency linear amplification of variable envelope RF signals. EER employs switching mode power amplifiers, which are extremely nonlinear but efficient, and the concept of supply voltage modulation to achieve linear performance. Polar transmitter is similar to the analog EER technique, but can be implemented digitally, as shown with the schematic in Fig. 18.5. In a polar transmitter, the phase and amplitude information of the RF input signal are calculated from its Cartesian coordinates and independently processed through the amplifier [9]. The RF modulated phase signal is typically processed through the input of a

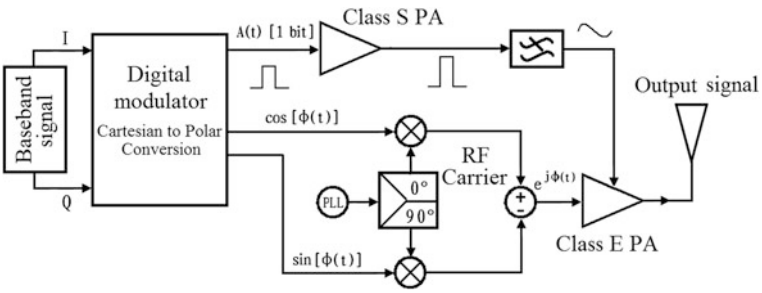

Fig. 18.5 Block diagram of the polar transmitter concept

nonlinear switching mode high efficiency PA, while the envelope information modulates the high efficiency PA's supply voltage. Polar modulation scheme is complex, and the total transmitter efficiency depends on both the PA and the supply modulating circuit efficiency. The linearity and efficiency limitations in the supply modulator will have a strong impact on performance of the whole transmitter system.

18.4 Experiment and Implementation

In order to promote the application of high efficiency and linearity techniques in the domain of telemetry engineering and improve the performance of telemetry transmitter, we devote our energies to explore possibilities for design methodology of high efficiency power amplifier and linearization technology. Our work can be summarized as follows:

1. The dual match technique and its implementation method are proposed to cope with the disadvantages of traditional design technique [10]. A practical Doherty power amplifier based on LDMOS devices with 55 dBm saturation output power is designed by this method. The photo of the experiment circuits is shown in Fig. 18.6. The improvement of efficiency with dual match is better than the one without dual match. Figure 18.7 depicts the experimental results of this Doherty power amplifier. The PAE of the Doherty amplifier is 15 % higher than that of the balance amplifier and more than 40 % within 8 dB backoff area. The test results indicate that the amplifier achieves 12 dB gain and 10 % higher PAE within a 6 dB output power backoff area. Experimental results show excellent agreement with simulations, which verifies the theoretical analysis.
2. Linearization for RF power amplifiers is indispensable in modern wireless communication systems. With enhanced linearity, PA can operate closer to its saturation region with higher power efficiency. A 200 W S band solid state power amplifier based on predistortion solution is developed. Figure 18.8 is the photo of this power amplifier. As the two-tone test is an important measure of linearity, we present a two-tone measurement for frequency spacing of 5 MHz.

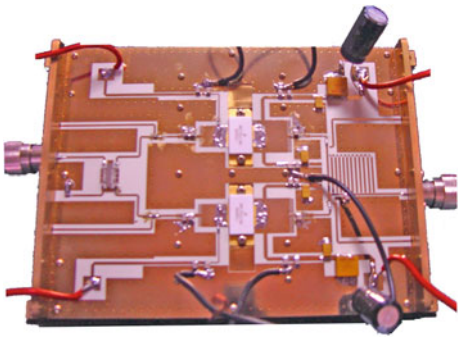

Fig. 18.6 The photo of the test board

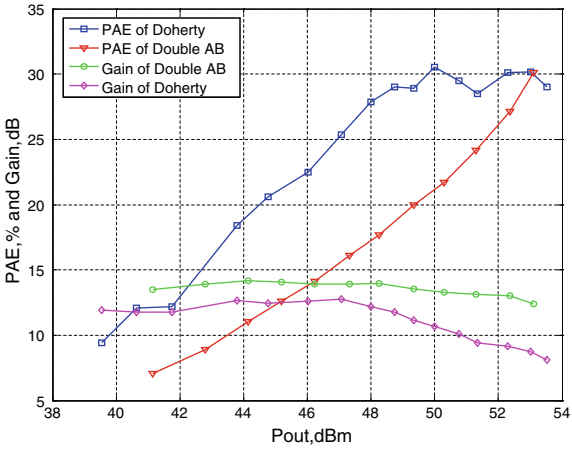

Fig. 18.7 The test results of dual match Doherty PA

Fig. 18.8 The photo of the 200 W S band solid state power amplifier

The measured IMD3 is −45 dBc when the output power is 100 W. And the IMD3 level decreases 25 dB within 7 dB backoff area.

3. Comprehensive theoretical analysis and design methodology of inverse Class E amplifiers with finite DC blocking capacitance is studied, and transmission-line amplifier using GaN transistor is built [11]. The practical PA circuit obtains output power 40.2 dBm at 2.3 GHz with drain efficiency of 76.1 % and power-

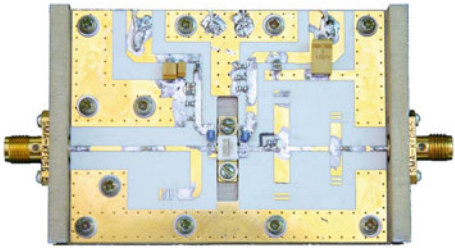

Fig. 18.9 The schematic and photo of the test board

added efficiency of 73.3 % when the supply voltage is 26 V. Furthermore, digital predistortion is used to increase the linearity of inverse class-E power amplifier, and a high degree of spectral suppression for out-of-band is achieved.

4. A modified inverse class E power amplifier with series tunable parallel resonant tank has been analyzed. The new configuration is similar to a hybrid combining features of Inverse Class E and Class F or Class1/F, and demonstrates varying degrees of trade-off between the peak drain voltage and current when the added parallel resonant is tuned to the different frequency. The theoretical principles and design equations required to calculate the optimum circuit component values were given in detail [12]. The proposed amplifier not only maintain 100 % efficiency theoretically, but also bring in some improvements such as higher output power, power output capability, and further relaxes design requirements. The practical amplifier implemented delivers 40.3 dBm output power at 2.3 GHz, and achieves PAE of 75.2 %, drain efficiency of 78.1 % and 13.3 dB power gain when operated from a 27 V supply voltage. The schematic and photo of the experiment circuits is shown in Fig. 18.9. Experimental results show excellent agreement with simulations, which verifies the theoretical analysis.
5. The design methodology of multiband high efficiency power amplifier based on composite right-left handed (CRLH) transmission lines is presented [13]. The optimal source/load impedance is found with source/load-pull technology. The CRLH transmission lines offer suitable source/load impedance and suppression of second harmonics simultaneously. The schematic diagram of the multiband high efficiency power amplifier is shown in Fig. 18.10. The proposed amplifier achieves 10 W output power, exhibits 13 dB power gain with drain efficiency over 60 % when operated at 0.9, 1.4 and 2.1 GHz respectively.
6. In order to linearize the inverse class-E PA with strong nonlinearity, we employ digital predistortion (DPD) for its linearization [6]. In this work, a GaN based inverse class E PA is designed with transmission line topological structure. There are four $\lambda/4$ shorting stub to construct the filtering network at its output network, which are used to suppress the 2nd, 3rd, 4th and 5th order harmonic components respectively. The measured results for the designed PA show that it has maximum power added efficiency (PAE) of 70 %. A novel linearization method with digital predistortion enhanced by up-converted dual-envelope injection (UCDEJ) method is proposed. And the linearization performance is

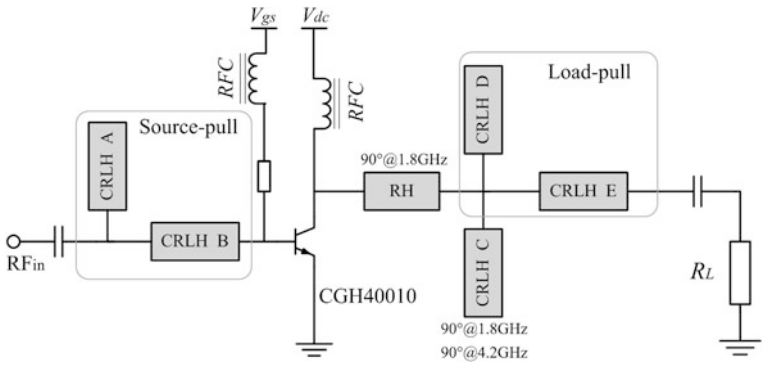

Fig. 18.10 The schematic diagram of the multiband high efficiency power amplifier

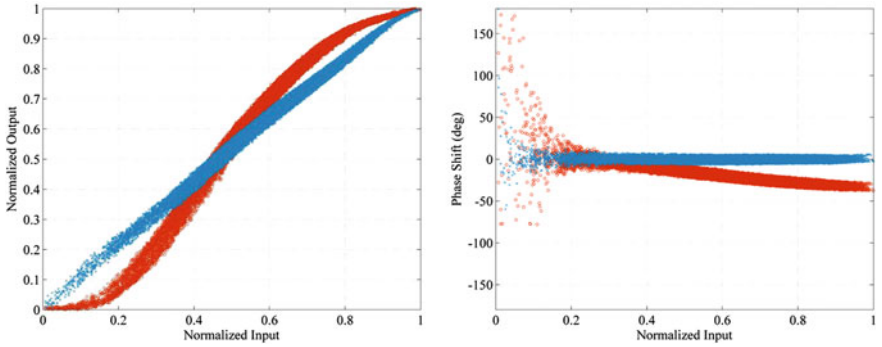

Fig. 18.11 Measured AM-AM and AM-PM for inverse class-E power amplifier linearization with digital predistortion enhanced by up-converted dual-envelope injection

validated with a one carrier WCDMA signal. The measured results reveal that, with performance enhanced by UCDEJ, it can get better improvement of spectral regrowth suppression. At this time, the inverse class-E PA is with PAE of about 36 %. Figure 18.11 gives the results of AM-AM and AM-PM of the inverse class-E PA before and after linearization. We can see that both amplitude and phase distortion are compensated very well.

7. In modern and future wireless communications systems, the increasing number of frequency bands and spectrum fragmentation require the development of circuits and subsystems having broadband capabilities. From the transmitter point of view, the power amplifier is the most critical component since its performance strongly influences the overall system features in terms of bandwidth, output power, efficiency, and operating temperature. This makes wideband PAs that cover many frequency bands while maintaining high efficiency an important research topic. Therefore we explore the designed methodology of high efficiency wideband power amplifier. Practical high efficiency wideband

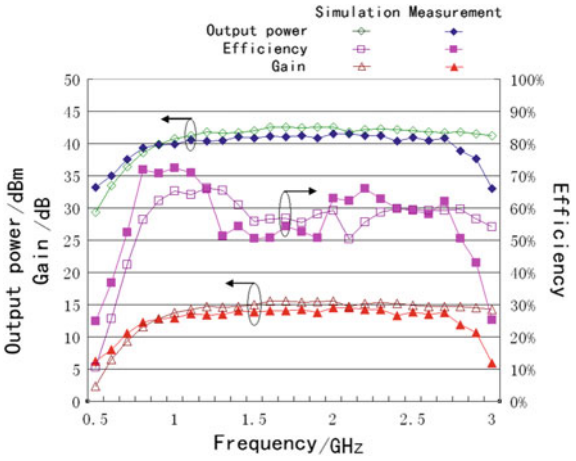

Fig. 18.12 The output power, gain and drain efficiency of the simulations and the measurements versus operating frequency

power amplifier circuit using GaN transistor is built. A method based on source-pull and load-pull simulation has been used to find optimum source and load impedances across the bandwidth and then used with a systematic approach to design wideband matching networks. Measurement results show that, across 0.9–2.7 GHz, the amplifier circuit is able to deliver 10 W output power with 13 dB power gain and obtain 51–72 % drain efficiency. The output power, gain and drain efficiency of the simulations and the measurements versus operating frequency are illustrated in Fig. 18.12. Comparing measured results with simulated results, the excellent agreement was obtained, which verifies the theoretical analysis. Furthermore, a simple predistortion solution is used to increase the linearity of power amplifier, and elaborate experimental results are achieved. The intermodulation product of power amplifier with predistortion as a function of input power is presented in Fig. 18.13. The measured IMD3 is below −40 dBc when the input power less than 23 dBm, and the IMD3 level decreases about 10 dB after predistortion. At this time, the drain efficiency of presented wideband PA is about 50 %, which means that both efficiency and linearity of the PA have desirable performance.

8. The adaptability of high efficiency class E power amplifier in telemetry system has been investigated. To confirm the flexibility of switching mode PA for constant envelop signal, PCM-FM system simulations have been carried out within Agilent Advanced Design Systems (ADS) suite. Simulation results illustrate that, for the frequency modulation signal, frequency spectrum and constellation have no distortion caused by switching mode PA. However, for the QPSK signal with non-constant envelop, linearization techniques should be implemented accurately to enhance the linearity of high efficiency class E power amplifier. And then, practical S-band transmission-line inverse class E power amplifier with series tunable parallel resonant tank using GaN transistor is built. The amplifier circuit is shown to be able to deliver 10 W output power and achieve drain efficiency of 78 %. By using this high efficiency power

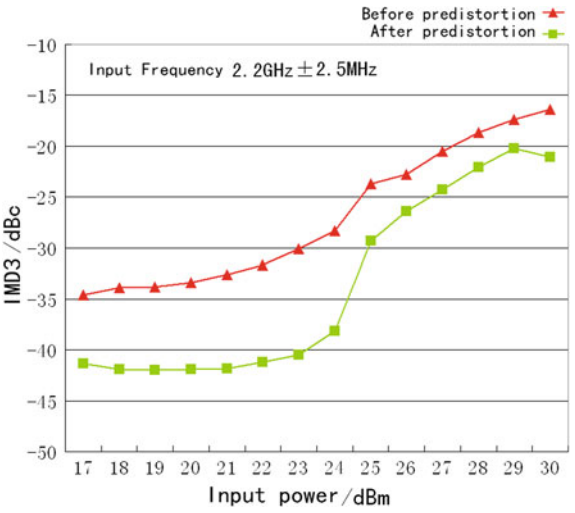

Fig. 18.13 The measured IMD3 of power amplifier with predistortion

Fig. 18.14 The photo of the high efficiency PCM telemetry transmitter

amplifier, the developed telemetry transmitter, which is suitable for 2 Mbps PCM signal, obtains an output power 9.8 W with efficiency of 57.6 %. Figure 18.14 shows the photo of the high efficiency PCM telemetry transmitter.

18.5 Conclusions

The continuous growth of wireless communications has been composed of the constant quest for higher data throughput, better signal quality, lower operating and deployment costs, and longer operating time. Consequently spectrum efficiency and power efficiency of transmitter are the two major concerns in designing modern wireless infrastructure for communication system. In general, the most power hungry device in a transmitter is a power amplifier which has nonlinear

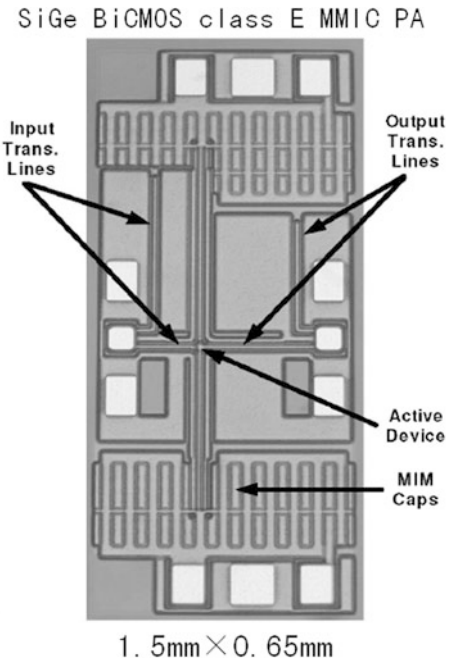

Fig. 18.15 The chip microphotograph of a millimeter-wave Class E PA

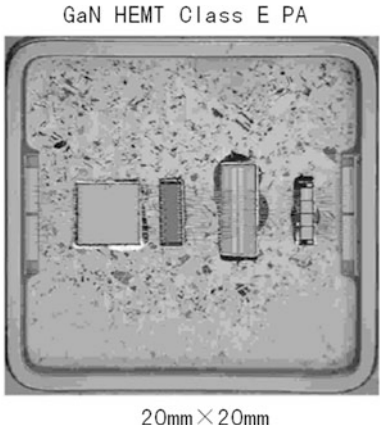

Fig. 18.16 The photo of GaN HEMT with Class E operation at S band

characteristics. In addition, modulation of non-constant envelope signals demands high linearity of a power amplifier. As a result, various competitive and complementary technologies have been investigated. This span from baseband signal processing techniques dealing with advanced modulation schemes and DPD linearization technologies to RF system technologies that extend from architecture (LINC, EER and Polar transmitter) to developments at the circuit (Doherty, switching mode PA) and device (MMIC) levels.

At the device level, emerging semiconductor technologies, such as gallium nitride (GaN) and silicone carbide (SiC) transistors are enabling switching mode PAs to achieve high performance. Figure 18.15 shows the chip microphotograph of a millimeter-wave Class E power amplifier [14]. The prototype IC is a single-ended single stage design that operates from a 1.2 V supply and employs an area of 0.98 mm^2. Measurement results show a saturated output power of 11.7 dBm with peak PAE of 20.9 % at 58 GHz. Figure 18.16 is the photo of GaN HEMT with Class E operation at S band [15]. The high output power of 100 W with high drain efficiency of 75 % and associated power gain of 12 dB was obtained at 2.14 GHz for CW operation with fairly low harmonics output. These achievements clearly demonstrated the potential of switching mode amplifiers in significantly improving the power efficiency of wireless communications transmitter. With the linearization technologies and innovatory RF architecture, the performance of future transmitter will be improved significantly.

References

1. Suzuki H (2008) Transmitter noise effect on the performance of a MIMO-OFDM hardware implementation achieving improved coverage. IEEE J Sel Areas Commun 26(6):867–876
2. Sungwook K (2009) Inverted-load network for high-power Doherty amplifier. IEEE Microw Mag 10(1):93–98
3. Lee YS, Jeong YH (2007) A high-efficiency Class-E GaN HEMT power amplifier for WCDMA applications. IEEE Microw Wirel Compon Lett 17(8):622–624
4. Negra R (2008) Concurrent dual-band class-F load coupling network for applications at 1.7 and 2.14 GHz. IEEE Trans Circ Syst II Express Briefs 55(3):259–263
5. Kim WJ, Stapleton SP, Kim JH et al (2005) Digital predistortion linearizes wireless power amplifiers. IEEE Microw Mag 6(3):54–61
6. Liu YJ, Cao T, Zhou BH et al (2011) Linearization for inverse class E RF power amplifier by memory polynomial digital predistortion. J Microw 27(3):79–82
7. Huttunen A, Kaunisto R (2007) A 20-W Chireix outphasing transmitter for WCDMA base stations. IEEE Trans Microw Theory Tech 55(12):2709–2718
8. Cox DC (1974) Linear amplification with nonlinear components. IEEE Trans Commun 22(12):1942–1945
9. Elliott MR (2004) A polar modulator transmitter for GSM/EDGE. IEEE J Solid-State Circ 39(12):2190–2199
10. Zeng R, Zhou J (2011) High efficiency and high power Doherty power amplifiers based on dual match. Semicond Technol 36(5):352–354
11. Cao T, Liu YJ, Zeng R (2011) S-band high efficiency GaN inverse-class E power amplifier. J Microw 27(4):49–52
12. Cao T, He SB, You F (2011) Performance study of inverse class E power amplifier with series tunable parallel resonant tank. Int J Microw Wirel Technol 3(4):405–413
13. Cao T, Liu YJ, Lv LM (2011) Multiband high efficiency power amplifier based on CRLH transmission lines. Microelectronics 27(4):49–52
14. Alberto VG, Scott R, Ullrich RP (2006) A 60 GHz class-E power amplifier in SiGe. IEEE Asian Solid-State Circ Conf 1(1):199–202
15. Norihiko U, Seigo S (2006) A 100 W class-E GaN HEMT with 75 % drain efficiency at 2 GHz. Eur Microw Integr Circ Confer 1(1):72–74

Chapter 19
Analysis of Feed Defocus's Effects on a Ka-Band Parabolic Antenna

Guolong He

Abstract The parabolic reflectors of deep space antennas are influenced by many facts, such as fabrication error, gravity and thermal deformation, which induce the feed deviating from the focal point of the paraboloid. The feed defocus would introduce phase errors over the paraboloidal dish, and degrade the antenna performance. In this paper, first the feed defocus is discussed from the view of reflector geometry, and the corresponding path length difference is derived. Then the effects on a Ka-band deep space parabolic antenna with 35-m diameter are analyzed as an example. From geometrical analysis and numerical calculations, the effects of feed defocus on some important antenna parameters such as gain, sidelobe level and pointing error are given and summarized. Finally, a method of feed defocus calibration is proposed.

Keywords: Parabolic antenna · Feed defocus · Antenna pattern · Gain · Sidelobe level

19.1 Introduction

The parabolic antennas are most widely used in the fields of deep space exploration, which have many unique characteristics and very good performance such as big dishes, high gain, very narrow beam and multi-frequency bands [1]. Its primary function is radiating the electromagnetic energy effectively delivered from the high power amplifier (HPA), and/or receiving the radiation power from a source. The parabolic antenna is a mechatronic system, which contains the accurate

G. He (✉)
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: qingchunruyu@qq.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_19,

paraboloidal dish, sensitive servo and control system, and the feed. The feed is used to convert the RF signal to electromagnetic wave, and illuminates the primary reflector with a particular energy distribution pattern. In order to acquire good performance, the feed should have the following characteristics [2]: (1) the phase center of the feed should locate on the focal point of the antenna reflector; (2) the phase of the feed does not vary with the observation angle; (3) the feed illuminates the primary reflector in a uniform manner; and (4) the feed should have a good directivity, which means the most of the energy is illuminated on the main reflector, while the spillover is reduced as low as possible.

In practice, the parabolic reflectors of deep space antennas are influenced by many factors, such as fabrication error, gravity and thermal deformation, which induce the feed deviating from the focus of the paraboloid. The feed defocus would give rise to phase errors over the paraboloidal dish, then degrade some antenna key parameters, such as higher gain loss, boarder beamwidth, lower antenna efficiency, higher sidelobe level (SLL) and cross-polarization. However, if these effects caused by feed defocus are carefully treated, some of them can be used in particular ways, such as near-field focusing measurement [3] and beam-squint conical scan [4].

This paper is organized as follows. In Sect. 19.2, the feed defocus from the view of paraboloid geometry is discussed, and the corresponding path length difference over the primary dish is derived. Then in Sect. 19.3 the effects on a Ka-band deep space parabolic antenna with 35-m diameter is analyzed as an example. From geometric analysis and numerical calculations, the effects of feed defocus on antenna key parameters such as gain, sidelobe level and pointing error are given and summarized. Finally, a method of feed defocus calibration is proposed in Sect. 19.4.

19.2 Geometric Analysis of Feed Defocus

19.2.1 Geometric Analysis of Axial Defocus

Deep space parabolic reflector antennas usually employ Cassegrain layout, which has a paraboloidal primary reflector and a hyperboloidal secondary subreflector. One foci of the subreflector coincides with the focus of the primary reflector, while the feed is on the other focal point. This Cassegrain layout can be equally considered as a single paraboloidal reflector for simplicity. The feed deviating from the focal point is decomposed in two normal directions: one along the reflector axis (axial defocus δ_a) and the other perpendicular to the axis (lateral defocus δ_l). As shown in Fig. 19.1, F is the focal point, and the feed is located on F′, δ_a denotes the length of axial defocus. By the cosine rule of the triangle PFF′, the following expression is obtained

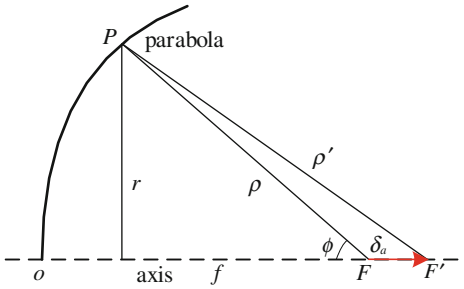

Fig. 19.1 Axial defocus

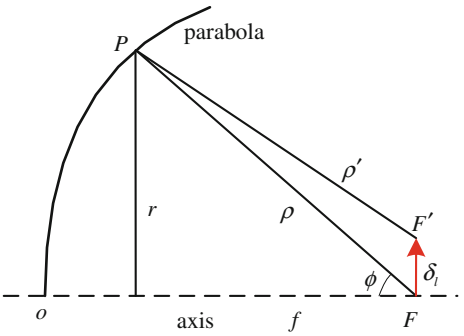

Fig. 19.2 Lateral defocus

$$\begin{aligned}\rho'^2 &= \rho^2 + \delta_a^2 - 2\rho\delta_a\cos(\pi - \phi)\\ &= \rho^2 + \delta_a^2 + 2\rho\delta_a\cos\phi\end{aligned} \tag{19.1}$$

Thus, the path length difference Δ_a caused by axial defocus is

$$\begin{aligned}\Delta_a &= \delta_a - (\rho' - \rho)\\ &= \delta_a - \frac{\delta^2 + 2\rho\delta_a\cos\phi}{\rho' - \rho}\\ &\simeq \delta_a(1 - \cos\phi)\end{aligned} \tag{19.2}$$

19.2.2 Geometric Analysis of Lateral Defocus

Lateral defocus is defined as the distance of the feed deviating from the reflector axis. Figure 19.2 shows the schematics, the feed is located on F′, and δ_l denotes the length of lateral defocus. Again, by the cosine rule of the triangle PFF′, the expression is obtained as follow

$$\rho'^2 = \rho^2 + \delta_l^2 + 2\rho\delta_l\cos(\pi/2 - \phi) \tag{19.3}$$

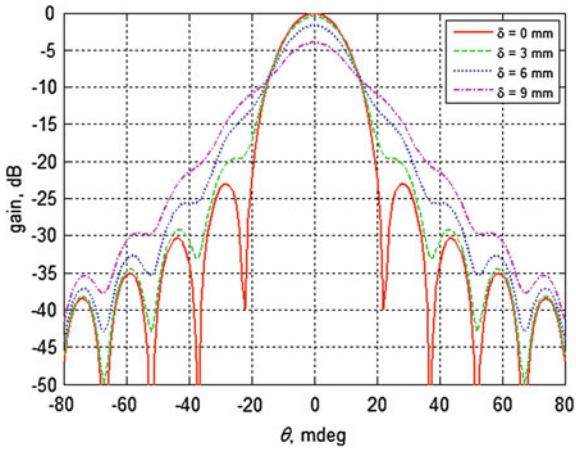

Fig. 19.3 The antenna power pattern with different axial defocuses

By the Taylor Series expansion,

$$\begin{aligned}\rho' &= \rho\sqrt{1+\left(\frac{\delta_l^2}{\rho^2}-2\frac{\delta_l}{\rho}\sin\phi\right)} \\ &\approx \rho+\frac{\rho}{2}\left(\frac{\delta_l^2}{\rho^2}-2\frac{\delta_l}{\rho}\sin\phi\right)+O(\delta^2) \\ &= \rho-\delta_l\sin\phi+\frac{\delta_l^2}{2\rho}+O(\delta_l^2)\end{aligned} \tag{19.4}$$

Thus, the path length difference Δ_l caused by lateral defocus is

$$\Delta_l = \rho' - \rho \approx \delta\sin\phi \tag{19.5}$$

19.3 Effects of Feed Defocus on Antenna Performance

19.3.1 Effects of Axial Defocus

As discussed in Eq. (19.2), the path length difference Δ_a caused by axial defocus is converted to phase error $\Delta\phi_a$ over the paraboloidal dish,

$$\Delta\phi_a = \frac{2\pi}{\lambda}\delta_a(1-\cos\psi_0)r^2 \tag{19.6}$$

From which the antenna field pattern $f(u)$ can be derived [1, 5]

Table 19.1 Effects of feed defocus on antenna performance

Defocus length (mm)	Effects of axial defocus on antenna gain loss (dB)	Effects of lateral defocus on sidelobe level (dB)
0	0[a]	−22.9404
3	−0.4093	−18.1778
6	−1.6739	−15.3177
9	−3.9140	−13.4233

[a] 0 dB means that the antenna gain is normalized as no axial defocus case

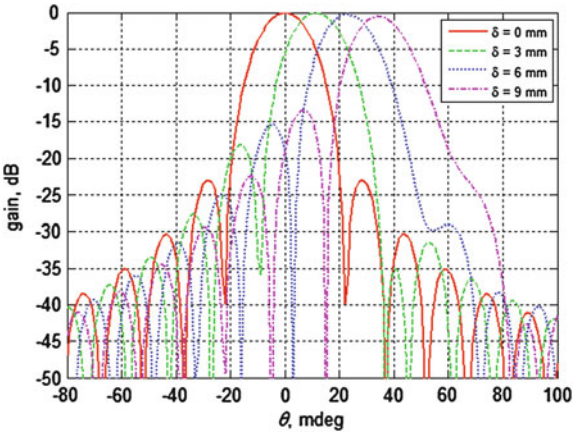

Fig. 19.4 The antenna power pattern with different lateral defocuses

$$f(u) = \int_0^1 F(r) J_0(ur) \exp(-j\Delta\phi_a) r dr \tag{19.7}$$

where

$$u = \frac{\pi d}{\lambda} \sin\theta \tag{19.8}$$

where θ is the angle from the axis, $J_0(.)$ is zero-order Bessel function, r is normalized radii, and $F(r)$ is the feed illumination function, which has a approximately quadratic functional form,

$$F(r) = 1 - (1 - \tau) r^2 \tag{19.9}$$

τ is called taper coefficient, and $0 < \tau < 1$. When $\tau = 0$, the illumination at the reflector edge is zero, and the feed has a uniform illumination when τ is 1. The antenna power pattern $g(u)$ is just the square of the antenna field pattern in Eq. (19.7),

$$g(u) = f^2(u) \tag{19.10}$$

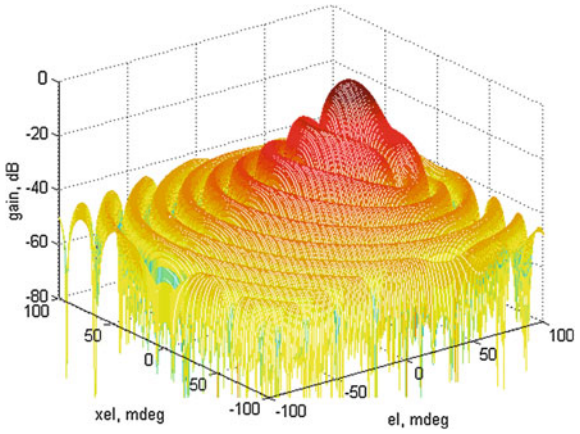

Fig. 19.5 3D plot of the antenna power pattern with lateral defocus $\delta_l = 9$ mm

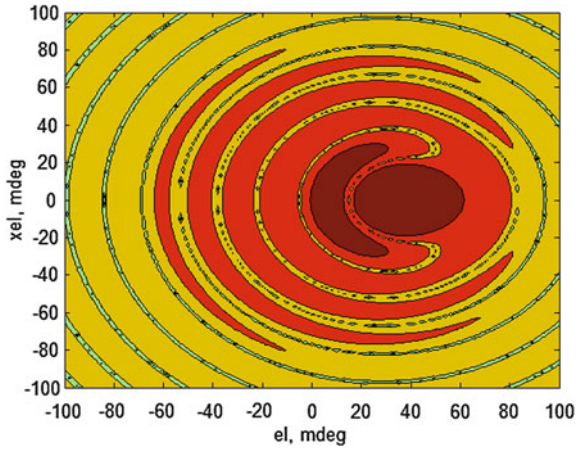

Fig. 19.6 Contour plot of the antenna power pattern with lateral defocus $\delta_l = 9$ mm

Here a Ka-band deep space parabolic antenna is used as example, the diameter of the primary dish $D = 35$ m, the working frequency $f = 32$ GHz, the corresponding wavelength $\lambda = 9.375$ mm, and the diameter-to-focus ratio $F/D = 0.325$, the aperture angle of the paraboloid $\Psi_0 = 74.88°$. Figure 19.3 illustrates the antenna power pattern with different axial defocuses ($\tau = 0.25$), and the numerical data is summarized in Table 19.1. Both of them show that, the antenna gain has a great loss when the axial defocus becomes large, especially when it's comparable with the wavelength. The gain loss is nearly 4 dB when the axial defocus $\delta_a = 9$ mm.

19.3.2 Effects of Lateral Defocus

Similar with Sect. 19.3.1, the phase error $\Delta\phi_l$ over the parabolic dish caused by lateral defocus can be expressed as

$$\Delta\phi_l = \frac{2\pi}{\lambda}\Delta_l = \frac{2\pi\delta}{\lambda}\sin\phi \tag{19.11}$$

In this case, the antenna field pattern $f(u)$ is [5, 6]

$$f(u) = 2\int_0^1 F(r)J_0\left[r\left(u - \frac{2k\delta}{4f/D1}1 + (D^2/8f)^2\right)\right]rdr \tag{19.12}$$

As shown in Eq. (19.10), the antenna power pattern g(u) is just the square of the antenna field pattern. Figure 19.4 shows the antenna power pattern with different lateral defocuses ($\tau = 0.25$), and the numerical data is also summarized in Table 19.1. Both of them show that, the antenna has a pointing error to the right when lateral defocus exists, and the sidelobe level on the left side is raised. The SLL is -22.9 dB with no lateral defocus, but raises to -13.4 dB when $\delta_l = 9$ mm. A 3D plot and a contour plot with lateral defocus $\delta_l = 9$ mm are given in Figs. 19.5 and 19.6, respectively, both of which show the asymmetrical sidelobe clearly.

19.4 Conclusions

From geometric analysis and numerical calculations, some conclusions about feed defocus's effects on antenna performance are derived in this paper: (1) axial defocus would cause antenna gain loss, and (2) lateral defocus induces antenna pointing error and merges higher sidelobe level.

In practical feed position calibration, one can first reduce lateral defocus by moving the feed position in the plane which is perpendicular to the paraboloid axis to the direction having the highest sidelobe, and make the sidelobe level near equal in every direction. After that, one can continue to move the feed to the different positions along the paraboloid axis, and measure the corresponding antenna gains, from which the best position along the axis can be fitted.

References

1. Imbriale WA (2006) Large antennas of the deep space network. Wiley Interscience, New Jersey
2. Kraus JD, Marhefka RJ (2004) Antennas for all applications, 3rd edn. McGraw-Hill Education, New York

3. Yaghjian AD (1986) An overview of near-field antenna measurement. IEEE Trans Antennas Propag 34(1):30–45
4. Fiebig D, Wohlleben R, Prata A et al (1991) Beam squint in axially symmetric reflector antennas with laterally displaced feeds. IEEE Trans Antennas Propag 39(6):774–779
5. Baars JWM (2007) The paraboloidal reflector antenna in radio astronomy and communication. Springer, New York
6. Ruze J (1965) Lateral-feed displacement in a paraboloid. IEEE Trans Antennas Propag 13(5):660–665

Chapter 20
Study on ISAR Imaging of Stepped-Frequency Chirp Signal

Haotian Yuan, Shuliang Wen and Zhen Cheng

Abstract The application of the stepped-frequency Chirp signal on the wideband phased-array radar was analyzed in this paper. The window function was used to design the radar key parameters to minimize the ghost peak and the sampling loss. Then a velocity compensation method was presented to solve the Doppler Effect. Then the effects of the aperture fill time on the antenna direction pattern were also analyzed and a grouping phase matching rule was given. At last was the imaging result of the real data, which shows the validly of the proposed methods.

Keywords: Wideband phased-array radar · Stepped-frequency chirp signal · Aperture fill time · Ghost peak

20.1 Introduction

In order to solve the contradiction between the transmitted energy and the data rate of stepped-frequency signal, some scholars propose to use Chirp sub-pulse to replace the rectangle sub-pulse in the stepped-frequency signal and the new signal of the sub-pulse whose centre frequency is well-distributed and stepped, which is the signal of Chirp Sub-pulse Stepped-Frequency. The stepped-frequency chirp signal is also called stepped frequency chirp, frequency-jumped burst or synthetic bandwidth. It is Tradex radar [1] at S wave band of Air Force Base in 1974 that first realized this wave form. Afterwards, Aegis SPY-1 radar, AN/MPQ-53 radar and PSTER all made use of this wave form [2].

H. Yuan (✉) · S. Wen · Z. Cheng
Beijing Institute of Radio Measurement, Beijing 100143, China
e-mail: joe007@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_20,

Although the imaging principle of stepped-frequency chirp signal is simple and is easy for engineering realization, it will come out with grating lobe which this paper names as "ghost peak". Besides, the stepped-frequency chirp signal is Doppler sensitive signal, the movement of radar target will lead the synthetic range profile to range migration and waveform distortion, and therefore, compensating the velocity of target shall be done before the coherent and synthetic treatment of pulse. What's more, wideband phased array radar may generate the "dispersion" effect because of the fill time of the aperture, leading two important problems [3] one is the instant bandwidth of the signal is restricted by the aperture fill time; the other is when the antenna scanning beam deviates from the normal direction, the beam pointing changes as the frequency of emitted signal changes. Firstly, this paper puts forward the parameter design rule specific to the ghost peak; secondly, a velocity compensation method is given specific to the movement of target. Lastly, specific to the effect of the aperture fill time, a grouping phase matching rule, a digital compensation method and the imaging result of the real data are given.

20.2 Parameter Design

The key parameters of modulated stepped-frequency chirp radar signal includes stepped-frequency $\triangle f$, number of stepped-frequency N, sub-pulse bandwidth B_m, sampling frequency f_s, time of pulse repetition T_r, and sub-pulse width T_p. The largest difference between parameter selection of modulated stepped-frequency signal and that of stepped-frequency chirp signal lies in the selection of relation between stepped-frequency $\triangle f$, sub-pulse bandwidth B_m, and sampling frequency f_s. This paper focuses on the selection principle of the above three parameters.

Weighting is often used in modulated stepped-frequency signal sub-pulse compressing in engineering to reduce side lobe and broaden the main lobe. The broadening multiple (n) is determined by the window function.

Let the window function of sub-pulse compressing be $w(t/Tp) = v(t/Tp) \cdot rect(t/Tp)$, where $w(t)$ is the window function of time normalization and $v(t)$ is periodic extension of $w(t)$. When the ratio of sub-pulse compression $D = Bm \cdot Tp$ is very large and the linear modulated frequency sub-pulse signal after pulse compressing is approximately a constant (C), the output signal of sub-pulse after windowing in the main lobe position can be expressed as the following formula:

$$\begin{aligned} y_{sub}(t) &= IFFT\{S(f) \cdot S^*(f) \cdot W(f/B_m)\} \\ &= IFFT\{C \cdot W(f/B_m)\} = C \cdot w(B_m \cdot t), \quad t \ll Tp \end{aligned} \tag{20.1}$$

In formula (20.1), $W(f/B_m)$is the Fourier transform of $W(B_m \cdot t)$. The above equation shows that the output main lobe of Chirp sub-pulse signal after windowing can be approximately represented as the Fourier transform of window function $W(f/B_m)$ with its frequency domain as B_m.

20.2.1 Sub-Pulse Bandwidth

Corresponding with the width of envelop τ after pulse compressing, the modulated stepped-frequency Chirp signal sub-pulse bandwidth can be regarded as the stepped-frequency Chirp signal with an output pulse width of τ and envelop of $\sin c(\cdot)$. Therefore, the selecting of stepped-frequency Chirp signal with an output pulse width of τ can be referred when selecting the modulated stepped-frequency signal sub-pulse bandwidth. Generally, $\tau\Delta f \leq 1$ (tight constraint conditions) is used to prevent the range profile from aliasing. The selection of sub-pulse bandwidth B_m is based on how to define τ.

If define $\tau = \frac{2n}{B_m}$ (n refers to the broadening multiple of main lobe after windowing and pulse compressing), the aliasing of range profile can be completely eliminated theoretically. According to the tight constraint conditions, the corresponding sub-pulse bandwidth should be:

$$B_m \geq 2n\Delta f \tag{20.2}$$

Since the ghost peak of side lobe can not be completely eliminated, certain ghost peak of main lobe is tolerable in engineering. Let $\tau = \frac{2m}{B_m}$ and the tolerable maximum ghost peak range be $A_{\min}$, m is the broadening multiple of main lobe ghost peak $A_{\min}$ after windowing and pulse compressing. Since m is determined by $A_{\min}$, the relation between sub-pulse bandwidth and modulated stepped-frequency is obtained as the following formula.

$$20\lg(w(m)) = A_{\min} - 20\lg(w(0)) \tag{20.3}$$

Corresponding sub-pulse bandwidth

$$B_m \geq 2m\Delta f \tag{20.4}$$

20.2.2 Sampling Frequency

As for modulated stepped-frequency Chirp signal, the envelope of linear modulated frequency sub-pulse after pulse compressing is $\sin c(\cdot)$. When there are multiple scattering points within one sample distance unit, the sampling output is affected by $\sin c(\cdot)$ instead of the result of merely adding equal weighting of scattering point echo. The highest SNR (Signal to Noise Ratio) within one sample distance unit is abstained only after the pulse compressing of scattering point echo, while there are losses in varying degrees for SNR of other scattering point echo. To minimize the affect of $\sin c(\cdot)$ on IDFT, sampling stitching algorithm is needed to improve the sampling frequency. However, there still are some losses in the scattering points of range profile after stitching. The loss range becomes less as the sampling frequency becomes higher.

For imaging radar, the target range profile should reflect the target scattering property as real as possible, so the sampling frequency should be as high as possible. However, if the sampling frequency is too high, the heavy computation will seriously affect the processing speed of signal. Therefore, the optimal sampling frequency should be adopted to meet allowable loss requirement with less computation.

So when designing sampling frequency f_s, the influence of window function on envelop after pulse compressing should be taken into account.

It is obvious that the relation between sampling frequency and sub-pulse bandwidth is completely determined by maximum sampling loss A_{los} and window function property $w(t)$. Let sampling period and sub-pulse meet $T_s = \frac{2s}{B_m}$, where s is determined by A_{los}. The formula is expressed as follows:

$$20\lg(w(s)) = A_{los} - 20\lg(w(0)) \tag{20.5}$$

Corresponding sampling frequency

$$f_s \geq \frac{B_m}{2s} \tag{20.6}$$

Therefore, the relation between stepped-frequency Δf, sub-pulse bandwidth B_m, and sampling frequency f_s can be represented as follows:

$$f_s \geq \frac{B_m}{2s} \geq \frac{2m\Delta f}{2s} = \frac{m}{s}\Delta f \tag{20.7}$$

20.3 Velocity Estimation and Compensation

20.3.1 Compensation Accuracy Requirement

The main effect of target motion on synthetic range profile is the linear phase error and quadratic phase error. The linear phase error causes the coupling time shift of IDFT output and the quadratic phase error causes waveform divergence of IDFT output. The requirement of compensation accuracy is listed as follows:

Linear phase error compensation accuracy:

$$|\Delta V|_{lin} < \frac{C}{4Nf_oT_r} \tag{20.8}$$

Quadratic phase error compensation accuracy:

$$|\Delta V|_{qua} < \frac{C}{4N^2\Delta fT_r} \tag{20.9}$$

It can be concluded that in order to get optimal range profile, motion compensation is required and the compensation accuracy should be in line with the

above two formulas. Because of the synthetic bandwidth is less than carrier frequency ($N\Delta \ll ff_0$), the compensation accuracy of quadratic phase error $|\Delta V|_{qua}$ should be less than that of linear phase error$|\Delta V|_{lin}$. In typical parameters, the compensation accuracy of quadratic phase error is 10 m/s order of magnitude and that of linear phase error is 1 m/s order of magnitude.

20.3.2 Velocity Estimation and Compensation

The velocity compensation of synthetic range profile is conducted after accurate velocity estimation is obtained. Pulse train is often used in the velocity estimation in radar system. The velocity compensation after sub-pulse compressing can be directly multiplied by corresponding phase factor as follows:

$$\exp(j2\pi \frac{2\hat{v}}{C}(f_0 + i\Delta f)iT_r) \tag{20.10}$$

20.4 Phase Matching Grouping and Digital Compensation

According to the stepped-frequency chirp signal, pulse–pulse correlation can be applied in the wideband phased array radar under the ideal case. However, due to the stepped-frequency chirp signal is very sensitive to Doppler so a pulse train (in one frame) with minor changes in target motion feature can be used in radar design. Hence a high pulse repetition frequency or medium pulse repetition frequency with very short pulse repetition interval is always being required. Meanwhile transferring on the side of the phase shifting is expected to be limited within an extremely ultimate short period when changing the pulse to pulse phase shifter, which requires the phase shifter to be outstanding. On considering the existing problems of spending a relative long period in transferring the antenna phase shifter and comparatively heavy load in phase matching calculation and transmission, a solution of pulse group phase matching might be adopted in this project. That is by dividing the stepped-frequency chirp signal into several groups, antenna phase shifter in each group can be achieved by one matching base on one frequency. Through this way the loss of offset antenna beam under the same frequency can be limited in the allowable range while the frequency change becomes smaller and the size (mainly depends on various groups of effective wideband) of each group can be selected appropriately. At this point the problems of how to set a way of grouping to ensure the loss of offset antenna beam is limited in a rational range and figure out the impact of composing the range profile resulted from different grouping have to be solved in this project.

20.4.1 The Aperture Effect of Grouping Phase Matching

On neglecting the mutual coupling impact, take a linear array made by M array elements with d spacing, set the 0 array element on one side of the linear array as the reference point, which has a spacing R from the target, and the corresponding delay is τ, then the echo of the array element m in this linear antenna is:

$$s_{m,n}(t) = A_{m,n} rect\left(\frac{t - nT_r - \tau - m\Delta\tau}{T_p}\right) e^{j\pi K(t-nT_r-\tau-m\Delta\tau)^2} \cdot e^{j[2\pi(f_0+n\Delta f)(t-\tau-m\Delta\tau)+\theta_n]},\ m = 0, 1, \ldots, M-1 \quad (20.11)$$

In the above formula, $A_{m,n}$ is the amplitude of the n times' back wave in the linear m, and $\Delta\tau$ is the time differential between the adjacent array elements when the scanning beam reached the θ angle and $\Delta\tau = \frac{d\sin\theta}{C}$, θ_n is then initial phase item of sub pulse emission.

Due to the uniform distribution of m array elements, the echo amplitude of each array element can be approximately equal, that is $A_{m,n} \approx A_n$.G_s is the quantity of each group of sub pulse on grouping the phase matching; K is the frequency-modulation-slope of the stepped-frequency chirp signal; f_0 is the initiation carrier frequency. Mod() is the complementation operation; floor() is the integer arithmetic operation; ceil() is the upward integer arithmetic operation. In case of separating the whole array element by grouping grouped phase matching in one time domain. When the phase matching proceeded by G group composed by N sub pulse, the corresponding sub pulse quantity of each group is G_s ($G_s = ceil(\frac{N}{G})$, the quantity of the final group of sub pulse is $N - (G-1)\cdot G_s$) and the mid-frequency of each group sub pulse is f_g(mid-frequency f_g differs from various groups), each group's phase matching depends on its mid-frequency f_g, afterwards, the echo signal [4] of the echo signal (back wave) n in the whole linear array can be realized:

$$s_n(t) = \frac{1}{M}\sum_{m=0}^{M-1} A_{m,n} rect\left(\frac{t - nT_r - \tau - m\Delta\tau}{T_p}\right) e^{j\pi K(t-nT_r-\tau-m\Delta\tau)^2} \cdot e^{j(2\pi(f_0+n\Delta f)(t-\tau-m\Delta\tau)+\theta_n)} e^{j\left(2\pi f_g m\Delta\tau\right)} \approx A_n x(t-\tau)\cdot \frac{1}{M}\sum_{m=0}^{M-1} e^{-j(2\pi f\_eff(t,n)\cdot m\Delta\tau)} e^{j\pi K(m\Delta\tau)^2} \quad (20.12)$$

In the above formula, $x(t-\tau)$ is the formula after the corresponding delay τ of stepped-frequency chirp signal; $f\_eff(t,n)$ is the frequency difference of the corresponding groups at the n sub pulse in the time of t after grouping, which is regarded as the "Equivalent instantaneous wideband" at the time of grouping the phase matching in the phased array radar for the linear stepped-frequency chirp signal, and

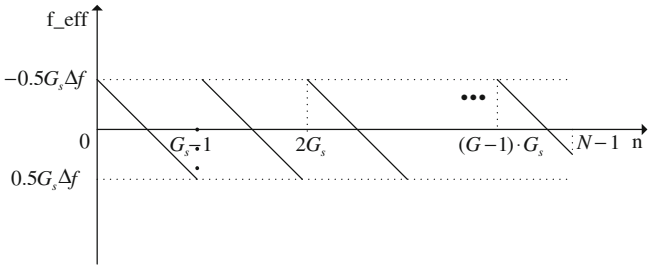

Fig. 20.1 Diagram of the equivalent instantaneous wideband changes over the sub pulse variation when grouping the phase matching

$$f\_eff(t,n) = f_0 + n\Delta f + K \cdot rect\left(\frac{t - nT_r - \tau}{T_p}\right) \cdot (t - nT_r - \tau) - f_g$$
$$= K \cdot rect\left(\frac{t - nT_r - \tau}{T_p}\right)(t - nT_r - \tau) + (\mathrm{mod}(n, G_s) - 0.5G_s))\Delta f \quad (20.13)$$

$$f_g = f_0 + (floor(\frac{n}{G_s}) + 0.5)G_s \cdot \Delta f \quad (20.14)$$

$$K \cdot rect\left(\frac{t - nT_r - \tau}{T_p}\right)(t - nT_r - \tau) \le B_m \ll G_s\Delta f \quad (20.15)$$

Since the pulse quantity G_s of each group is comparatively larger (usually greater than 10) and the sub pulse wideband of the stepped-frequency chirp signal is relatively smaller than all equivalent wideband, thus

Ignore the impact from the sub pulse, take the mid-frequency t_n into consideration, where $t_n = nT_r + \tau$, then

$$f\_eff(t,n) \approx f\_eff(t_n, n) = (\mathrm{mod}(n, G_s) - 0.5G_s)\Delta f \quad (20.16)$$

Figure 20.1 is the information of equivalent instantaneous wideband changes over frequency when grouping the phase matching.

For the general phased array antenna, usually $\Delta\tau$ is smaller (in nanosecond grade) and the frequency-modulation-slope K of the stepped-frequency chirp signal is comparatively lower (normally lower than 1013), so the impact of $e^{j\,\pi K(m\Delta\tau)^2}$ can be ignored. Formula (20.12) can be simplified as

$$s_n(t) = A_n x(t - \tau) \cdot \frac{1}{M}\sum_{m=0}^{M-1} e^{-j(2\pi f\_eff(t)\cdot m\Delta\tau)} \quad (20.17)$$

Where the local oscillation of the stepped-frequency chirp signal is

$$s_{ref}(t) = rect\left(\frac{t - nT_r}{T_p}\right)e^{j2\pi((f_0 + n\Delta f)t + \theta_n)} \quad (20.18)$$

After the frequency mixing processing to formula (20.12), the output signal afterwards is

$$\begin{aligned} r_n(t) &= s_n(t) \cdot s_{ref}^*(t) \\ &= A_n rect\left(\frac{t - nT_r}{T_p}\right) e^{-j2\pi(f_0 + n\Delta f)\tau} e^{j\pi K(t - nT_r - \tau)^2} \\ &\quad \times h_1(f\_eff(t,n)) \cdot h_2(f\_eff(t,n)) \end{aligned} \tag{20.19}$$

In the above formula

$$h_1(f\_eff(t,n)) = \frac{\sin(\pi M \Delta\tau f\_eff(t,n))}{M \cdot \sin(\pi \Delta\tau f\_eff(t,n))} \tag{20.20}$$

$$h_2(f\_eff(t,n)) = e^{-j(\pi(M-1)\Delta\tau f\_eff(t,n))} \approx e^{-j(\pi M \Delta\tau f\_eff(t,n))} \tag{20.21}$$

Concluded from formula (20.19), we can see its single-dimensional range profile, for the stepped-frequency chirp signal phased array radar, is influenced by both the amplitude weighting function $h_1(f\_eff(t,n))$ and phase weighting function $h_2(f\_eff(t,n))$.

20.4.1.1 The Influence of the Phase Weighting Function to the Single-dimensional Range Profile

Because of the equivalent wideband effect of the grouping phase matching, the required actual phase position is different from the disposing-phase. The phase difference after phase matching should be the periodic function since the equivalent instantaneous wideband $f\_eff(t,n)$ is also a periodical function. The effect of the phase matching error on the ideal range profile is similar to the phase quantization error on the ideal antenna directional pattern; the difference between them is that the influence of the former to the single-dimensional range phase is in time domain while the latter to the antenna directional pattern is in spatial domain. Obviously, the influence of the phase weighting function $h_2(f\_eff(t,n))$ to the single-dimensional range phase equals to the convolution sums [5] of the range phase and Fourier series. The influence can be analyzed in the way of analyzing the relevant quantization error [6], which is analyzing the error functions by adopting the Fourier series. Where

$$h_2(f\_eff(t,n)) \approx h_2(f\_eff(t_n,n)) = e^{-j(\pi M \Delta\tau(\mathrm{mod}(n,G_s) - 0.5G_s)\Delta f)} \tag{20.22}$$

It shows, base on formula (20.21) and Fig. 20.2, that the periodic function is the phase weighting function, the sequencing time period is G_s, and the discrete fundamental wave angular frequency is $\omega_0 = \frac{2\pi}{G_s}$. According to the influence from quantification error to directional pattern, it shows that the influence of h_2 generates at the time ghost peaks (the Ghost Peak generated hereof is similar to the quantization side lobes caused by quantization error of the phase shifter. Only

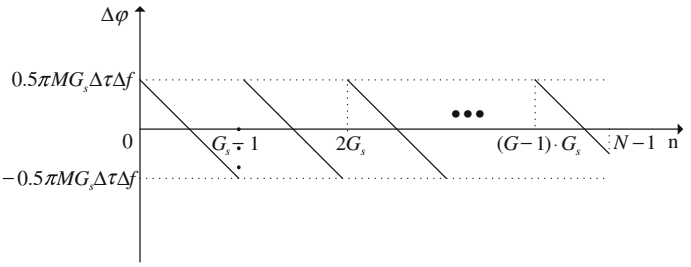

Fig. 20.2 Diagram of the phase item of the phase weighting function changes over the sub pulse variation

the First ghost peak shall be given consideration in reality) forming at both sides of the range phase. Besides, phase weighting function will also cause amplitude attenuation in the single dimensional ranged phase with an attenuation magnitude of b_0.

Based on the discrete fundamental wave frequency ω_0, the position of the first ghost peak is:

$$\Delta R = \frac{\omega_0}{2\pi/N}\Delta r = \frac{2\pi/G_s}{2\pi/N}\cdot\frac{C}{2N\Delta f} = \frac{C}{2G_s\Delta f} \tag{20.23}$$

With generality in consideration and set G_s as an odd, then

$$w_2(n) = \begin{cases} h_2(f\_eff(t_n, n + \frac{G_s-1}{2})), & -\frac{G_s-1}{2} \le n \le \frac{G_s-1}{2} \\ 0, & \text{others} \end{cases} \tag{20.24}$$

Then

$$w_2(n) = \exp(-j\pi M\Delta\tau \cdot n \cdot \Delta f),\ -\frac{G_s-1}{2} \le n \le \frac{G_s-1}{2} \tag{20.25}$$

Knowing the Fourier series, then

$$b_k = \frac{1}{G_s}\sum_{n=-\frac{G_s-1}{2}}^{\frac{G_s-1}{2}} w_2(n)\exp(-j\frac{2\pi nk}{G_s}) = \frac{\sin\pi((\frac{M\Delta\tau\cdot\Delta f}{2}+\frac{k}{G_s})G_s)}{G_s\cdot\sin\pi(\frac{M\Delta\tau\cdot\Delta f}{2}+\frac{k}{G_s})} \tag{20.26}$$

Make $k = 0$, then the attenuation of the single dimensional range profile caused by phase mismatching of:

$$b_0 = \frac{\sin\pi(\frac{M\Delta\tau\cdot\Delta f}{2}G_s)}{G_s\cdot\sin\pi(\frac{M\Delta\tau\cdot\Delta f}{2})} \tag{20.27}$$

Make $k = -1$, we have the First ghost peak on the right side of the single-dimensional range profile of:

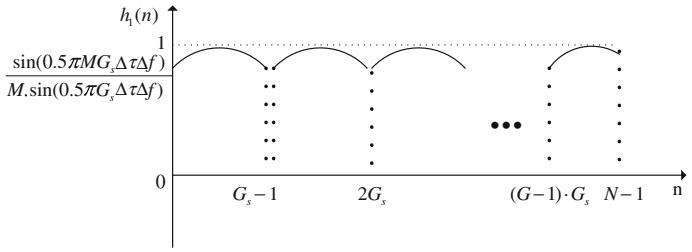

Fig. 20.3 Diagram of the phase item of the amplitude weighting function changes over the sub pulse variation

$$b_{-1} = \frac{\sin\pi((\frac{M\Delta\tau\cdot\Delta f}{2} - \frac{1}{G_s})G_s)}{G_s \cdot \sin\pi(\frac{M\Delta\tau\cdot\Delta f}{2} - \frac{1}{G_s})} \tag{20.28}$$

Make $k = 1$, then the First ghost peak on its left side as:

$$b_1 = \frac{\sin\pi((\frac{M\Delta\tau\cdot\Delta f}{2} + \frac{1}{G_s})G_s)}{G_s \cdot \sin\pi(\frac{M\Delta\tau\cdot\Delta f}{2} + \frac{1}{G_s})} \tag{20.29}$$

20.4.1.2 Influence of Amplitude Weighting Function to the Single-Dimensional Range Profile

It is similar to the phase weighting function. The amplitude weighting function can also be analyzed in the same way of analyzing the Fourier series.

$$\begin{aligned} &h_1(f\_eff(t,n)) \approx h_1(f\_eff(t_n,n)) \\ &= \frac{\sin(\pi M\Delta\tau(\mathrm{mod}(n,G_s) - (0.5(G_s-1))))\Delta f}{M\cdot\sin(\pi\Delta\tau(\mathrm{mod}(n,G_s) - (0.5(G_s-1))))\Delta f} \end{aligned} \tag{20.30}$$

According to the formula of $f\_eff(t,n)$, the function distribution of h_1 is as the following Fig. 20.3, $\frac{\sin(0.5\pi MG_s\Delta\tau\Delta f)}{M\cdot\sin(0.5\pi G_s\Delta\tau\Delta f)}$ hereof is the minimum value of function h_1.

It shows, base on formula (20.30) and Fig. 20.2, that h_1 is also a discrete periodical function and the G_s is the discrete period, and the corresponding angular frequency of the discrete fundamental wave is $\omega_0 = \frac{2\pi}{G_s}$.

If make

$$w_1(n) = \begin{cases} h_1(f\_eff(t_n), n + \dfrac{G_s-1}{2}), & -\dfrac{G_s-1}{2} \le n \le \dfrac{G_s-1}{2} \\ 0, & \text{others} \end{cases} \tag{20.31}$$

Then

$$w_1(n) = \frac{\sin(\frac{\pi M d \sin\theta}{C} n \cdot \Delta f)}{\sin(\frac{\pi d \sin\theta}{C} n \cdot \Delta f)}, -\frac{G_s - 1}{2} \le n \le \frac{G_s - 1}{2} \tag{20.32}$$

The Fourier series of periodical function $w_1(n)$ can also be obtained through:

$$a_k = \frac{1}{G_s} \sum_{n=-\frac{G_s-1}{2}}^{\frac{G_s-1}{2}} w_1(n) \exp(-j\frac{2\pi nk}{G_s}) \quad k = -\frac{G_s - 1}{2}, -\frac{G_s - 3}{2} \ldots, \frac{G_s - 1}{2} \tag{20.33}$$

As $w_1(n)$ is the even function in regard to n, hence

$$\begin{aligned} a_0 &= \frac{1}{G_s} \sum_{n=-\frac{G_s-1}{2}}^{\frac{G_s-1}{2}} w_1(n) \\ a_k &= a_{-k} \end{aligned} \tag{20.34}$$

20.4.1.3 Digital Compensation Method

According to formula (20.19), we know that during matching the phase, the echo signal will be affected by the amplitude and the phase weighting function, so the amplitude and phase weighting function need to be compensated at the same time. The compensation formula is as follows:

$$\begin{aligned} H_1(n) \cdot H_2(n) &= \frac{1}{h_1(f\_eff(t_n, n))} \cdot \frac{1}{h_2(f\_eff(t_n, n))} \\ &= \frac{\sin(\pi \Delta\tau(\mathrm{mod}(n, G_s) - 0.5G_s)\Delta f)}{\sin(\pi M \Delta\tau(\mathrm{mod}(n, G_s) - 0.5G_s)\Delta f)} \\ & e^{j(\pi M \Delta\tau(\mathrm{mod}(n, G_s) - 0.5G_s)\Delta f)} \end{aligned} \tag{20.35}$$

After being compensated, the signal becomes

$$r_n^{'}(t) = A_n rect\left(\frac{t - nT_r}{T_p}\right) e^{-j2\pi(f_0 + n\Delta f)\tau} e^{j\pi K(t - nT_r - \tau)^2} \tag{20.36}$$

From formula (20.36), we know that after the amplitude and phase weighting function were compensated at the same time, the aperture effect of the phased array antenna can be solved effectively.

In addition, to eliminate the influence of the phase weighting function to range profile, phase center matching method can be selected, which means selecting the center element of the array other than Array Element 0 as the reference point.

20.4.2 Grouping Phase Matching Rule

From Ref. [3], the amplitude uniform distribution phased array antenna pattern is

$$|F(\theta)| = \left| \frac{\sin(\pi M \frac{d}{c}[(f_0 + \Delta f)\sin\theta - f_0 \sin\theta_0])}{\sin(\pi \frac{d}{c}[(f_0 + \Delta f)\sin\theta - f_0 \sin\theta_0])} \right| \tag{20.37}$$

The direction of the wideband signal beam changes along with the frequency of the signal.

For stepped-frequency Chirp signal, because grouping phase matching method can be used, so after grouping phase matching, the corresponding pattern becomes

$$|F(\theta)| = \left| \frac{\sin(\pi M \frac{d}{c}[(f_g + f\_eff(t,n))\sin\theta - f_g \sin\theta_0])}{\sin(\pi \frac{d}{c}[(f_g + f\_eff(t,n))\sin\theta - f_g \sin\theta_0])} \right| \tag{20.38}$$

Now the deviation of the beam produced by the antenna becomes

$$\Delta\theta = -\frac{f\_eff(t,n)}{f_g}\tan\theta \tag{20.39}$$

If the maximum scan angle of the antenna is $\theta_{\max}$, the maximum deviation of the beam produced by the antenna is

$$\Delta\theta_{\max} = \frac{(G_s - 1)\Delta f + B_m}{2f_0}\tan\theta_{\max} \approx \frac{G_s \Delta f}{2f_0}\tan\theta_{\max} \tag{20.40}$$

Among which, B_m is the impulse bandwidth and Δf is the stepped value of the frequency.

During grouping matching, in addition to the influence of weighting function to the range profile, the deviation of the beam shall also be considered. As the influence of weighting function to the range profile can be compensated digitally, the number of groups mainly depends on the allowed beam deviation loss. If uniform weighting is adopted for the antenna, at the maximum scan angle $\theta_{\max}$, the 1 dB beam width of the antenna is

$$\theta_{1dB} = \frac{0.523\lambda}{L\cos\theta_{\max}} \tag{20.41}$$

Among which, L is the length of the antenna.

At the maximum scan angle $\theta_{\max}$, the 3 dB beam width of the antenna is

$$\theta_{3dB} = \frac{0.886\lambda}{L\cos\theta_{\max}} \tag{20.42}$$

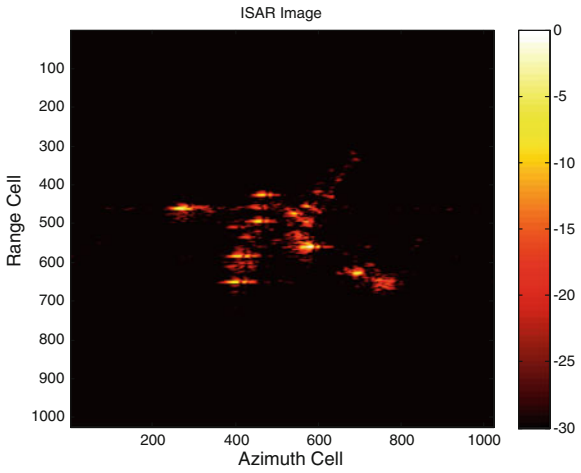

Fig. 20.4 Imaging result of the aircraft passing-by

If the allowed beam deviation is within 1 dB, it is required that

$$\Delta\theta_{\max} \leq \frac{\theta_{1dB}}{2} \tag{20.43}$$

After the parameters was introduced, the following result can be obtained

$$G_s \leq \frac{0.523C}{\Delta fL \sin\theta_{\max}} \tag{20.44}$$

If the allowed beam deviation is within 3 dB, it is required that

$$\Delta\theta_{\max} \leq \frac{\theta_{3dB}}{2} \tag{20.45}$$

After the parameters were introduced, the following result can be obtained:

$$G_s \leq \frac{0.886C}{\Delta fL \sin\theta_{\max}} \tag{20.46}$$

From the above description, we can see that under different requirements on beam deviation, the frequency modulation stepped phased radar grouping rule is also different.

20.5 Imaging Result

According to the parameter selection principles and actual phase matching method described in the paper, we had conducted test and verifications with phased array radar. Figure 20.4 is the processing result of the measured data of civil aircraft.

From the figure, it can be seen that the ghost peak of the aircraft is less than −30 dB and the two-dimensional ISAR imaging result of the aircraft is well.

20.6 Conclusions

The paper has investigated the specific application of frequency modulation stepped signal in phased array radar. Window function based parameter design principle is adopted for the ghost peak and the sampling loss of the stepped-frequency Chirp signal when picking up the range profile and the mathematical relationship between the stepped frequency increment, sub-pulse bandwidth and sampling frequency is given. For Doppler sensitive step frequency signal, velocity compensation method is given. The aperture effect of the wideband phased array is analyzed. Through the analysis, it is derived that the weighting function caused by the phase matching error produced during grouping matching is the convolution of the ideal range profile and the Fourier series of the weighting function, and proved that the effect of the phase matching error on the ideal range profile is similar as the phase quantization error on the ideal antenna directional pattern and the digital compensation method is promoted. Finally, the influence of the phase matching error produced during grouping matching to the directional pattern of the antenna is analyzed, the grouping phase matching rule is presented and the imaging result is given finally. It is proved that through the method described in the paper, the frequency modulated stepped signal can be used in wideband phased array radar successfully.

References

1. Hong S (1974) TRDEX frequency jump burst waveform. MIT Lincoln laboratory project report RMP47, 28, Lexington, June 1974
2. Rabideau DJ (2002) Nonlinear synthetic wideband waveforms. MIT Lincoln laboratory, Lexington, 02420-9108
3. Zhang GY, Zhao YJ (2006) Phased array radar technology. Publishing House of Electronics Industry, Beijing
4. Yuan HT (2009) Investigation of high velocity target frequency modulation stepped radar imaging technology. Doctoral thesis of 2nd institute of China aerospace science and technology corporation
5. Qin ZY (1991) Analysis of phased array antenna with linear system theory. Syst Eng Electron Technol 5:7–13
6. Barton DK, Ward HR (1969) Handbook of radar measurement. Prentice-Hall, New Jersey

Chapter 21
A Carrier Acquisition and Tracking Algorithm for High-Dynamic Weak Signal

Ruifeng Duan, Rongke Liu, You Zhou, Qingping Song and Zhiqiang Li

Abstract This paper studies acquisition and tracking of carriers by the Tracking, Telemetry and Command (TT&C) systems in low Signal-to-Noise Ratio (SNR) and high dynamic conditions for deep space applications. An associated carrier acquisition and tracking algorithm is proposed and it features FFT frequency-domain shift and accumulation algorithm in combination with variable bandwidth FLL-PLL algorithm with Doppler rate pre-compensation. The algorithm not only captures signal with high dynamics and low SNR but also achieves high tracking accuracy. Simulation results show that the proposed associated frequency estimation algorithm is capable of carrier frequency estimation with an error less than 0.5 Hz and it tolerates high Doppler dynamics under E_b/n_0 as low as 3 dB.

Keywords Deep space exploration · High dynamics · Frequency-domain shifting · Variable bandwidth · Doppler rate pre-compensation

21.1 Introduction

As a result of the long communication distance and high speed of the communication target in a deep space mission, the signals received by TT&C system is usually of low Signal-to-Noise Ratio (SNR) and high Doppler dynamic [1]. Therefore, acquisition and tracking of high-dynamic weak signal from a deep space probe is a great challenge for a deep space TT&C system. The carrier capture is a crude estimate of the Doppler frequency and Doppler rate, which is

R. Duan · R. Liu (✉) · Y. Zhou · Q. Song
School of Electronic Information Engineering, Beihang University, Beijing 100191, China
e-mail: rongke_liu@buaa.edu.cn

Z. Li
Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_21,

implemented through the local oscillator pre-compensation in segment FFT algorithm on the received signal to do two-dimensional search [2]. The algorithm computational complexity increases dramatically with broadening of Doppler frequency and Doppler rate range. Carrier tracking relies on refined estimation of frequency and phase. Mainly related to the algorithms like Maximum Likelihood Estimation (MLE) [3], Cross-Product Automatic Frequency Control (CPAFC) frequency Locking Loop assisted Phase Locking Loop (FLL-PLL) [4], Extended Kalman Filter (EKF) and improved algorithm [5–7] and so on [2–4]. MLE estimation algorithm achieves a fine estimate of the frequency and phase at the price of high computational complexity and poor real-time capability. The CPAFC algorithm is a negative feedback system based on the frequency discriminator. It is obviously limited by the loop bandwidth and is difficult to take into account both the dynamic property and the tracking accuracy in low-SNR. EKF linearly processes nonlinear parametric equation, and iteratively estimates current state by making use of input signal with high estimation accuracy. However, highly complicated implementation makes high speed computation difficult. Furthermore, EKF estimates the phase of the signal and has rather slow convergence rate and narrow capture range. FLL-PLL, a mature technology, not only exploits FLL accommodation ability for high dynamics signals but also takes advantage of high tracking accuracy. In particular, switching mode is used in low SNR. However, switching mode decreases the robustness of tracking when Doppler rate exists in receiver signal.

Following analysis of the existing carrier acquisition and tracking algorithms for weak and high dynamic signals in deep space, this paper proposes FFT frequency-domain shifting accumulation algorithm that simplifies coarse carrier acquisition process. Simultaneously, variable bandwidth FLL-PLL structure is used for accurate tracking of carrier with remaining Doppler frequency and Doppler rate. Furthermore, pre-compensation measure of Doppler rate is also added to FLL-PLL structure to complete refined tracking for carrier in order to overcome sensitivity of the switch operating mode for Doppler rate.

The rest of the paper is organized as follows. Section 21.2 describes signal model and provides a design of the overall acquisition and tracking structure based on the signal model. Section 21.3 proposes frequency-domain shifting accumulation algorithm and analyzes its acquisition accuracy and computational complexity. Section 21.4 proposes variable bandwidth FLL-PLL with pre-compensation of Doppler rate to improve traditional tracking algorithms. Section 21.5 gives the simulation result, which shows that the proposed joint frequency estimation algorithm achieves a precise estimate of the carrier frequency in extremely low SNR and high dynamic conditions. Section 21.6 concludes this paper.

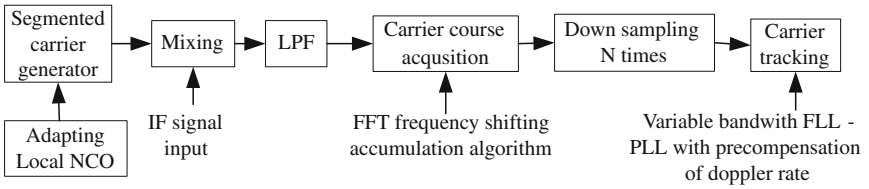

Fig. 21.1 Overall acquisition and tracking block diagram

21.2 Design of the Overall Acquisition and Tracking Structure of the Joint Frequency Estimation Algorithm

The signal frequency in this paper uses the constant acceleration model based on the high dynamic motion model of Jet Propulsion Laboratory (JPL). To maintain general applicability, we set signal IF to zero, and the received signal is expressed as

$$y(t) = Ae^{j(2\pi(f_d + at/2)t)} + n(t) \tag{21.1}$$

where A is the data bit with the value of 1 or -1, f_d is Doppler frequency, a is Doppler rate, $b(t)$ is the data bit stream with the value of 1 or -1, $n(t)$ is a white Gaussian noise sequence with $n(t) \sim N(0, \sigma^2)$. The task of carrier acquisition and tracking is accurate estimation of f_d and a.

A hierarchical estimation algorithm is used in this paper. After completion of the coarse carrier acquisition, the residual Doppler frequency and Doppler rate respectively satisfy $f_{\text{rem}} \leq f_{\text{req}}, a_{\text{rem}} \leq a_{\text{req}}$. The frequency error is less than 0.5 Hz after carrier tracking is completed. The overall acquisition and tracking block diagram is shown in Fig. 21.1.

We assume that the carrier Doppler frequency range is $[f_{d\min}, f_{d\max}]$ and Doppler rate range is $[a_{\min}, a_{\max}]$.When the carrier Doppler frequency covers a wide range, sampling frequency must be high enough. Furthermore, the SNR of IF signal is also very low and this brings more difficulty to subsequent processing.

The proposed algorithm in this paper searches Doppler frequency and Doppler rate simultaneously in two dimensions among many segments. In Fig. 21.1, the bandwidth of the Low Pass Filer (LPF) is $(f_{d\max} - f_{d\min})/L$.

The local carrier coming from segment carrier generator module l can be written as follows:

$$y\_loc_l(t) = \mathrm{e}^{-j(2\pi f_l t)}, \quad l = 1, 2, \ldots L \tag{21.2}$$

Mixed with input signal, we can obtain:

$$y\_cha_l(t) = y(t) \times y\_loc_l(t) = Ae^{j(2\pi(\Delta f_{dl} + at/2)t)} + n(t)e^{-j2\pi f_l t} = s_l(t) + n_l(t) \tag{21.3}$$

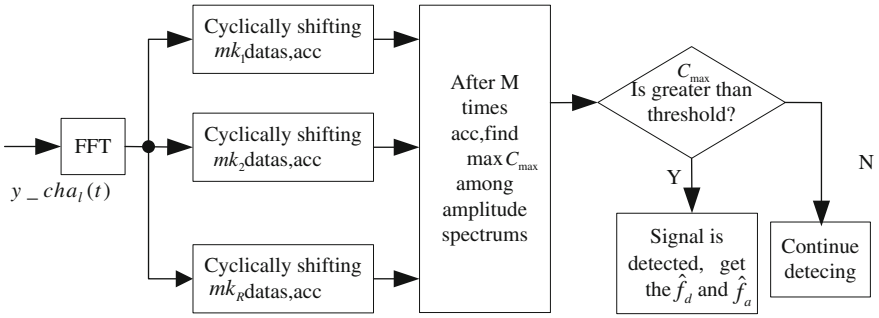

Fig. 21.2 Schematic of FFT frequency shifting accumulation algorithm

where $\Delta f_{dl} = f_d - f_l$, if and only if the specific local frequency f_l approaches f_d to the greatest extent, Δf_{dl} reaches the minimum value. After mixing, only this signal $y\_cha_l(t)$ in the special segment can pass the LPF, while other signals in other segment cannot pass. This special signal can be expressed as

$$y\_cha_m(t) = Ae^{j(2\pi(\Delta f_{dm} + at/2)t)} + n(t)e^{-j2\pi f_m t} \tag{21.4}$$

21.3 FFT Frequency Shifting Accumulation Algorithm for Coarse Carrier Acquisition

The carrier of the above mentioned beat frequency signal in every segment can be coarsely captured by FFT frequency shifting accumulation algorithm. Its schematic is shown in Fig. 21.2.

In Fig. 21.2, after T seconds, $y\_cha_l(t)$ can be expressed as follows:

$$\begin{aligned} y\_cha_l(t+T) &= Ae^{j2\pi(\Delta f_{dl} + a/2\cdot(t+T))\cdot(t+T)} + n_l(t+T) \\ &= C_T \cdot s_l(t)e^{j2\pi a\cdot T\cdot t} + n_l(t+T) \end{aligned} \tag{21.5}$$

where $C_T = e^{j2\pi(\Delta f_{dl}\cdot T + aT^2/2)}$, a constant independent of t. The FFT corresponding to $y\_cha_l(t+T)$ can be expressed:

$$y\_cha_{lT}(f) = C_T \cdot S_l(f - a\cdot T) + N_T(f) \tag{21.6}$$

When the Doppler rate differs, the signal FFT in the same time span can be obtained by shifting FFT of signal in initial time span at a different rate aT along frequency axis. Consequently, compensation to signal Doppler rate can be completed by shifting the FFT of signal in reverse direction at a different rate.

In Fig. 21.2, $m = 1, 2, \ldots M$, the cyclically shifting number $k_1, k_2, \ldots k_R$ in every matching branch corresponds to Doppler rate estimation $a_1, a_2, \ldots a_R$ for the local

carrier. R is the total number of matching branches. Assuming sampling frequency f_s, sampling point's number N at each time, we have the following relationship:

$$k_r = a_r \times N/f_s/(f_s/N) = a_r \times N^2/f_s^2, \quad r = 1, 2, \ldots R \tag{21.7}$$

The Fourier transform spectrum of mth segment data can be expressed as:

$$R(f, m) \approx AT \cdot \mathrm{sinc}\left(\frac{2\pi(f - (\Delta f_{dl} + maT))T}{2}\right) e^{-j2\pi(f-(\Delta f_{dl}+maT))\frac{T}{2}} + N(f) \tag{21.8}$$

The equation is established when $|a| \cdot T < 2/T$. In rth branch, we can shift single DFT result at a step size of k_r. After M times of accumulation, the average periodogram power spectrum can be written as:

$$\begin{aligned} |\bar{R}_r(K, m)| &= \frac{1}{MN}\sum_{m=0}^{M-1} |R(K + mk_r, m)|^2 \\ &\approx \frac{1}{M}\sum_{m=0}^{M-1}\left\{A^2N \cdot \left|\mathrm{sinc}\left(\frac{(K - (k_d + m(a - k_r/N^2)N)N}{2}\right)\right|^2 + |N(K)|^2\right\} \end{aligned} \tag{21.9}$$

The Sinc function in the spectrum shifts at a different degree in the reverse direction in every accumulation according to difference of Doppler rate estimation, $k_r \cdot f_s^2/N^2$ (related to a_r). When $k_r \cdot f_s^2/N^2$ in some branch approaches the real Doppler rate a to the greatest extent, the Sinc function shifting rate reaches minimum $|\Delta a_r|_{\min} \cdot T$ after compensation by shifting in the reverse direction. After the same times of accumulation, the corresponding power spectrum reaches maximum nearby the real carrier frequency. As a result, we can get respectively the Doppler frequency and Doppler rate frequency estimation $\hat{f}_d$ and $\hat{f}_a$.

In practice, the algorithm will encounter the situation where k_r is not an integer and rounding operation will affect acquisition performance. To solve the problem, we often pad zeroes to sampling data for d times before computing the FFT. Compared to the algorithm proposed in Ref. [8], the algorithm proposed in this paper can reach a commensurate acquisition accuracy and acquisition probability but the computational complexity reduces by R/d times. Assuming N is the number of FFT points, frequency estimation satisfies $f_{\mathrm{res}} \leq f_s/N$ and Doppler rate estimation satisfies $a_{\mathrm{res}} \leq (a_{\max} - a_{\min})/R/2$ after completion of the coarse acquisition. The acquisition probability is 90 %. The output signal of the local NCO after completion of acquisition is

$$y\_loc(t) = e^{-j(2\pi(f_{res}+a_{\mathrm{res}}t/2)t+\theta)} \tag{21.10}$$

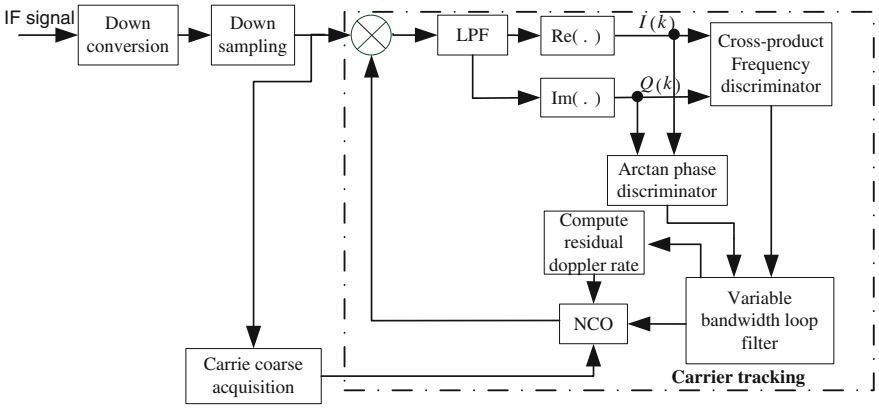

Fig. 21.3 Variable bandwidth carrier tracking with pre-compensation of Doppler rate

21.4 Variable Bandwidth FLL-PLL Carrier Tracking Loop with Doppler Rate Pre-compensation

After completion of coarse carrier acquisition, the carrier tracking loop is turned on. The schematic block diagram is shown in Fig. 21.3.

FLL can reach lower tracking threshold compared with PLL. It features better resistance against dynamics and interference. But it has big frequency errors. On the other hand, PLL can reach high tracking accuracy. Consequently, we can turn on PLL to track carrier after FLL tracking is locked. This not only improves tracking robustness but also improves tracking accuracy. When forward coarse carrier acquisition is completed, the beat frequency signal still has residual Doppler rate. The second-order FLL assisted third-order PLL is used in carrier tracking in this paper to make sure there shall be no steady-state errors.

After completion of carrier acquisition, the baseband signal entering carrier tracking module can be expressed as:

$$y\_cap(t) = Ae^{j(2\pi(f_{\text{rem}}+a_{\text{rem}}t/2)t+\theta)} + n_{\text{rem}}(t) \tag{21.11}$$

where $f_{\text{rem}} = f_d - f_{\text{res}}$ is the residual Doppler frequency, and $a_{rem} = a - a_{res}$ is the residual Doppler rate. $n_{\text{rem}}(t)$ is the noise output signal, which can be obtained when white Gaussian noise passes the coarse carrier acquisition module and can be deemed as white noise within signal bandwidth range. The aim of carrier tracking is completion of accurate estimation of f_{rem} and a_{rem}, and assurance of the demodulation performance. $I(k)$ and $Q(k)$ can be expressed as:

$$I(k) = A\cos(2\pi(f_{\text{rem}} + a_{\text{rem}}kT_s'/2)kT_s') \tag{21.12}$$

$$Q(k) = A\sin(2\pi(f_{\text{rem}} + a_{\text{rem}}kT_s'/2)kT_s') \tag{21.13}$$

Table 21.1 Acquisition results based on FFT frequency shifting accumulation algorithm

Item	Value
Symbol rate/(bps)	100
E_b/N_0/(dB)	3
Segment number	8
FFT points number	8192
Accumulation times	120
Doppler rate/(Hz/s)	−800 to 800
Doppler frequency/(MHz)	−2.4 to 2.4
Doppler rate estimation accuracy/(Hz/s)	±25
Doppler frequency estimation accuracy/(Hz)	±48
Acquisition probability	93 %

where T_s' is the inverse of the sampling frequency after down-sampling.

Cross-product Frequency Discriminator (FD) output can be written as:

$$\begin{aligned} disf &= \sin(2\pi(f_{\text{rem}}(t_1 - t_2) + a_{\text{rem}}(t_1^2 - t_2^2)/2)/(t_1 - t_2) + n_{disf} \\ &\approx 2\pi(f_{\text{rem}} + a_{\text{rem}}(t_1 + t_2)/2) + n_{disf} \end{aligned} \tag{21.14}$$

Arctan Phase Discriminated (PD) result is $2\pi(f_{\text{rem}} + a_{\text{rem}}kT_s'/2)kT' + n_{dispha}$

Obviously, when residual Doppler rate $a_{\text{rem}} \neq 0$, FD result is the linear function interfered by noise with the slope a_{rem}, PD result is the second parabolic function interfered by noise.

The loop bandwidth B_L is determined by following formula:

$$\rho = P \times S_L/(N_0 B_L) \tag{21.15}$$

where ρ is loop SNR, P is input signal power and S_L is the loop squaring loss. N_0 is the noise single power spectral density of Single Side Band. To get high tracking accuracy, B_L must be set to a very small value. However, only small carrier Doppler rate can be tolerated by the loop as a result. When Doppler rate is larger than the capture range, the normal tracking of the carrier will not be guaranteed.

This paper proposes variable width FLL-PLL tracking strategy with Doppler rate pre-compensation. At the beginning period of tracking, FLL-PLL switching mode is used. When FLL tows the residual carrier Doppler frequency and the Doppler rate to a small range, in which PLL can track reliably, FLL is turned off and PLL is turned on. When PLL is working alone, the residual Doppler rate estimation a_{adp} can be computed by making use of a series of adjacent discrete frequency adapting values $f_1, f_2, \ldots f_n, f_{n+1}, \ldots f_{n+p}, f_{n+p+1}, \ldots f_{n+2p}$. The specific equation is shown as follows:

$$a_{\text{adp}} = [(f_{2p} - f_p) + (f_{2p-1} - f_{p-1}) + \cdots (f_{p+1} - f_1)]/(pT_{adp})/p \tag{21.16}$$

where T_{adp} is the time interval between adjacent two frequency adapting values.

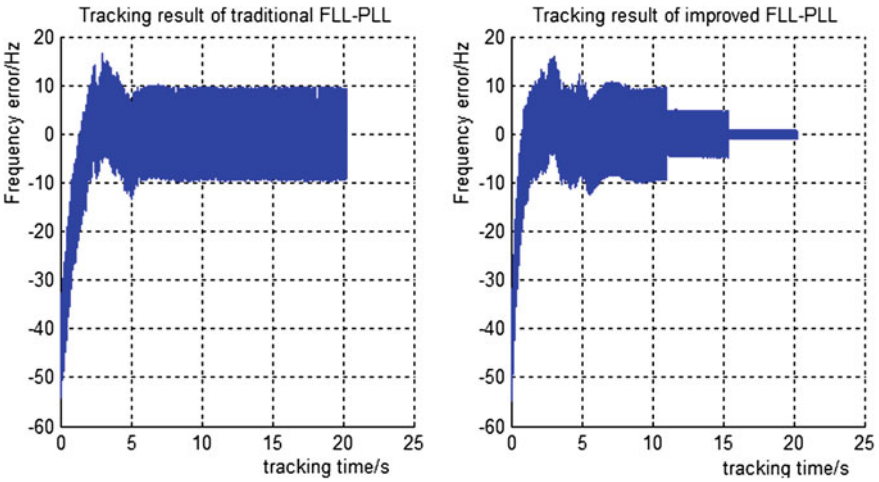

Fig. 21.4 Variation of the final instantaneous frequency error with time

Compensating a_{adp} to the local carrier NCO and the Doppler rate of LPF signal in carrier tracking module will be almost zero. The loop bandwidth can be set to a very narrow value, and tracking accuracy will be improved further.

Taking as an example the echo signal of a deep-space probe, the Doppler frequency varies from −2.4 to 2.4 MHz, the Doppler rate varies from −800 to 800 Hz/s, the symbol rate is 100 bit/s, and E_b/n_0 is 3 dB.

21.5 Simulation Result and Analysis

In the course carrier acquisition phase, Doppler frequency is divided into 8 segments for parallel capturing. The step of Doppler rate is 50 Hz/s in every matching branch, totally 32 branches. The times of zero padding interpolation d is 7. After down-sampling, the signal sampling frequency of each frequency segment is 800 kHz. The signal model for simulation is established as Eq. (21.1). The Monte Carlo simulation of 1000 times is executed based on the coarse carrier acquisition algorithm of Fig. 21.2. The acquisition result is shown in Table 21.1.

The computational complexity of the proposed algorithm decreases by 70 % compared with that in Ref. [8].

In carrier tracking phase, $|f_{\text{rem}}| \leq 48$ Hz, $|a_{\text{rem}}| \leq 25$ Hz/s. The traditional FLL-PLL loop and proposed tracking loop in Fig. 21.2 is respectively used to simulate carrier tracking. The final instantaneous frequency error with time varying relations is shown in Fig. 21.4.The proposed algorithm changes its bandwidth twice. The first variation happens when FLL is turned off and PLL is turned on, and the noise bandwidth is changed from $B_L = 5$ to $B'_L = 2.5$. For the second time, the noise bandwidth is decreased to 0.3 Hz after completion of residual Doppler rate compensation.

As shown in Fig. 21.4, the bandwidth of the traditional FLL-PLL is no less than 5 Hz to track the LPF signal with a residual track Doppler rate of 25 Hz/s. After completion of carrier tracking, the instantaneous frequency error reaches ±9 Hz, and this does not meet the requirements of subsequent demodulation. After use of the propose algorithm, the noise bandwidth is decreased to 0.3 Hz, and the final instantaneous frequency error is within ±0.5 Hz, meeting requirements for high demodulation performance.

21.6 Conclusions

An associated carrier acquisition and tracking algorithm for weak high-dynamic signal is proposed for deep-space applications. Firstly, coarse carrier acquisition is completed by the FFT frequency shifting accumulation algorithm at a low computational complexity. On this basis, we use variable bandwidth FLL-PLL with pre-compensation of Doppler rate to realize carrier tracking in extremely low SNR and high dynamic conditions. When E_b/N_0 is 3 dB, the Doppler frequency varies from −2.4 to 2.4 MHz, the Doppler rate varies from −800 to 800 Hz/s and the symbol rate is 100 bps. Simulation shows that the proposed algorithm provides a high frequency estimation accuracy of less than 0.5 Hz and lowers computational complexity dramatically. Therefore, the algorithm shows a valuable prospect for real-time carrier acquisition and tracking application in low-SNR and high dynamic conditions.

References

1. Fountain GH, Kusnierkiewicz DY, Hersman CB et al (2008) The new horizons spacecraft. Space Sci Rev 140(1–4):23–47
2. Han MF, Wang YQ, Wu SL et al (2009) A fast algorithm on parameter estimation of LFM signals under low SNR. Trans Beijing Inst Technol 29(2):147–151
3. Yan L, Hua F, Kam PY (2009) Improved, approximate, time-domain ML estimators of chirp signal parameters and their performance analysis. IEEE Trans Signal Process 57(4):1260–1272
4. Lu HB, Wang WW, Gu QC (2010) Carrier synchronization technique for low SNR and high dynamic condition. Inf Control 39(4):451–454
5. Li WB, Liu SJ, Zhou CH et al (2007) High dynamic carrier tracking using Kalman filter aided phase-lock loop. In: International conference on wireless communications, networking and mobile computing, pp 673–676
6. Miao JF, Sun YR, Liu JY et al (2009) A Kalman filter based tracking loop in weak GPS signal processing. In: International conference on fuzzy systems and knowledge discovery, pp 438–442
7. Zhu YL, Yang DK, Liu ZK (2009) Filte ring algorithm used for high dynamic GPS frequency estimation. J Beijing Univ Aeronaut Astronaut 35(1):24–27
8. Satorius E, Estabrook P, Wilson J et al (2003) Direct-to-earth communications and signal processing for Mars exploration rover entry, descent, and landing. IPN Progress Report 42-153, pp 1–35

Chapter 22
An Acquisition Algorithm for DS/FH TT&C Signal Using Subband-Accumulation Method

Xiao Chen, Zhiqiang Li, Wenming Zhu and Dekan Lou

Abstract The problem of fast acquisition of Direct Sequence/Frequency Hopping (DS/FH) spread spectrum signal is studied. With the existence of ionospheric dispersion, the carrier phase continuity of a phase-coherent FH system is disrupted and the complexity of acquisition is drastically increased. By searching the Total Electron Count (TEC), i.e. the electron density, along the path length, the receiver is capable of compensating for the phase pattern of local replica FH signal and performing quasi-coherent acquisition. A new acquisition strategy named Band Division Accumulation is proposed. It is proved that the strategy can broaden the TEC search period and thus reduce search complexity. The connection between the number of sub-bands and energy loss is analyzed. The accumulation loss of acquisition can be adjusted by altering the number of sub-bands, providing flexibility to design of receivers.

Keywords TT&C · DS/FH signal · Ionospheric dispersion · Phase estimation · Band division · Incoherent accumulation

22.1 Introduction

Jamming and anti-jamming of a telemetry link is an important issue in designing a spacecraft electronic system. Combining the excellence of direct sequence spread-spectrum and frequency hopping spread-spectrum signal, a hybrid DS/FH [1]

X. Chen (✉) · Z. Li · W. Zhu · D. Lou
Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China
e-mail: ty_chenxiao@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_22,

telemetry signal features a low spectrum density which is difficult to be spied and a frequency hopping pattern that dodges the hostile interference. These qualities has endowed it a promising future in anti-jamming telemetry applications [2, 3]. However, this introduction of the DS/FH signals in the field of aerospace telemetry has also brought about a bunch of technical problems. The first problem is that the capturing complexity increases and acquisition time is much longer. The second problem is that signal tracking, precise ranging and velocity measuring becomes more difficult [4, 5]. Studies have shown that ionospheric dispersion disrupts the linear transmission of broadband signals, causing a significant delay difference between frequency hops as the instantaneous carrier frequency changes from hop to hop. This paper aims at performing fast acquisition of DS/FH signals by canceling the ionospheric dispersion. In the following discussion, a band-division search and accumulating method is proposed based on quantitative analysis of the dispersion effect. The acquisition time and complexity are analyzed and simulated.

The remainder of this paper is organized as follows. In Sect. 22.2, we describe the ionospheric dispersion's effect in terms of phase leap and time delay and compare the ionosphere-disrupted signal with a vacuum propagation situation. In Sect. 22.3, the acquisition model of C-DS/FH signals is briefly discussed and the effect of ionospheric dispersion on the acquisition of signal is analyzed. Then, a search method named Band Division is proposed. The performance of Band Division method is evaluated in Sect. 22.4. Section 22.5 gives the CAD simulation results obtained by using Matlab 7. Finally, conclusions are drawn in Sect. 22.6.

22.2 Ionospheric Dispersion

Due to its multiple search ranges, low spectrum density and broad spectrum, fast acquisition of a DS/FH system is much more complicated than that of a DSSS system. In order to enhance anti-jamming capability and obtain a superior tracking precision, a coherent DS/FH(C-DS/FH) system, which is sensitive to the presence of nonlinearity in the phase characteristics of the channel transfer function, is presented and analyzed.

In absence of ionospheric effect, the received C-DS/FH signals can be written as

$$s(t) = \sqrt{2S}\sum_{i=0}^{\infty} a_i g(t - iT_h; T_h)\cos(\omega_i t + \phi_0) \tag{22.1}$$

where

S is the average power of the signal,

T_h is the frequency hopping cycle,

a_i, i = 0, ± 1, ± 2 ± 3, …, is a sequence of binary independent identically distributed symbols,

$g(t, T)$ is a gate function and ω_i is the hopping carrier frequency,

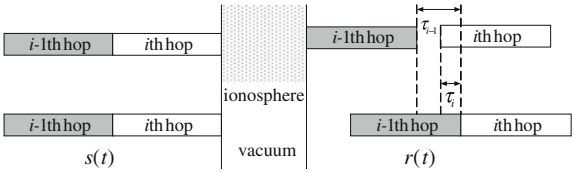

Fig. 22.1 Frequency hopping signal through ionosphere

ω_i is properly set to get integral numbers of carrier cycles in each frequency hopping period.

TT&C link of a spacecraft transmits radio signals through ionosphere and suffers from ionospheric dispersion. The nonlinearity of channel transfer function and group delay distortion of telemetry signal introduced a transient leap of the carrier phase between consecutive hops which has to be compensated so as to fulfill the coherent demodulation.

Ionospheric dispersion's effect on GPS signal acquisition is discussed in detail in ref. [6]. Figure 22.1 shows the time delay of FH signal travelling through ionosphere.

Comparing to the signal transmitting through vacuum, a C-FHSS signal disrupted by dispersion is regarded as below:

$$r(t) = \sqrt{2S}\sum_{i=0}^{\infty} g(t - iT_h - \tau_i)\cos(\omega_i t + \phi_i) + n(t) \tag{22.2}$$

where

$$\phi_i = \phi_0 - \omega_i \tau_i \tag{22.3}$$

ϕ_i is the phase leap regarding the vacuum traveling path, τ_i is the time delay regarding the vacuum traveling path, both can be expressed in terms of TEC as follows:

$$\tau_i = -\frac{40.3 \cdot \mathrm{TEC}}{f_i^2 \cdot c}(s) \tag{22.4}$$

$$\phi_i = 360^\circ \times f\tau_i = -\frac{4836 \cdot \mathrm{TEC}}{f} \times 10^{-8}(\mathrm{deg}) \tag{22.5}$$

Total Electron Count (TEC) is the electron density along the path length in the ionosphere. The value of TEC is within the range of 10^{16}–10^{19} and varies in less than a TECU (10^{16}) during acquisition time. There exist other factors, e.g. nonlinearity in trans-receiver, ionosphere storm and scintillation influencing, that affect the channel characteristic of C-FHSS signals. But analysis of this paper is merely based on the connection between TEC and dispersion.

22.3 Acquisition Model

The acquisition of hybrid DS/FH signals is a four-dimensional search (search of carrier frequency, PN code phase, Doppler and phase leap between consecutive hops). By applying the connection between TEC and ionospheric dispersion, the delay and phase leap can be compensated to perform coherent demodulation. Generally the receiver on the spacecraft needs to search the value of TEC in a certain range to obtain the delay and phase leap.

Thus in the acquisition process of DS/FH signals, the receiver performs a four-dimension search including the hopping frequency, PN code phase, Doppler and TEC. During the search process, the receiver searches the value of TEC and adjusts the phase of the local replica carrier to match the received carrier phase leap pattern. As long as the local replica signal reaches the correct phase pattern, the receiver will be able to capture the signal. In comparison with that of a DSSS telemetry system, the acquisition of a C-DS/FH system is bound with the burden of searching phase leap, and this multiplies design complexity. A search technique is proposed in this paper to solve time complexity in acquisition.

22.3.1 Compensation of Phase Leap

Assuming that the FH pattern and PN code phase are acquired and the carrier Doppler is compensated, we have the carrier phase leap between f_i, f_j as:

$$\varphi = \phi_i - \phi_j = \lambda_{\mathrm{TEC}} \cdot \mathrm{TEC} \cdot \frac{f_j - f_i}{f_i f_j} \tag{22.6}$$

where λ_{TEC} is defined as the *phase leap constant* and

$$\lambda_{\mathrm{TEC}} = 4836 \times 10^{-8} (\mathrm{deg}.Hz) \tag{22.7}$$

To compensate for the local replica phase leap, we choose the carrier phase of a FH hop with frequency f_0 as the reference phase, thus the relative phase leap in the ith interval can be written as:

$$\varphi_i = \phi_i - \phi_0 = \lambda_{\mathrm{TEC}} \cdot \mathrm{TEC} \cdot \frac{f_i - f_0}{f_i f_0} \tag{22.8}$$

And to match the received phase leap pattern, the local replica is designed as:

$$v(t) = \sqrt{2S} \sum_{i=0}^{\infty} g(t - iT_h) \cos(\omega_i t + \hat{\varphi}_i) + n(t) \tag{22.9}$$

The carrier phase in each hop of local replica FH interval is adjusted to match the received phase leap pattern through search of TEC. Consequently, the receiver is capable of performing synchronization.

22.3.2 Analysis of TEC Search Step

Let TEC search step be Δ, the phase estimation error of frequency point *fi* is expressed as:

$$\varphi_{error} = \varphi_i - \hat{\varphi}_i = \lambda_{\mathrm{TEC}} \cdot \frac{f_i - f_0}{f_i f_0} \cdot \frac{\Delta}{2} \tag{22.10}$$

In frequency hopping band B, we estimate the relative carrier phase leap of frequency f_i compared to that of central frequency of the FH band f_0. Obviously, the carrier on the edge of the FH band gets the largest estimation error:

$$\varphi_{error,\max} = \varphi_i - \hat{\varphi}_i = \lambda_{\mathrm{TEC}} \cdot \frac{B/2}{(f_0 + B/2) f_0} \cdot \frac{\Delta}{2} \tag{22.11}$$

Constraining the phase leap estimation error within the maximum value θ, we set the TEC search step as:

$$\Delta = \frac{2\theta \cdot (2f_0 + B) \cdot f_0}{\lambda_{\mathrm{TEC}} B} \tag{22.12}$$

TEC search complexity reduces when the frequency hopping band increases. This means that the acquisition time can be cut down significantly by dividing the FH band into smaller sub-bands.

22.3.3 Band Division Accumulation

The main purpose of Band Division is to shrink the TEC search complexity by a factor of 1/K by dividing the FH band into K sub-bands. The acquisition module for Band Division method is shown in Fig. 22.2. The incoming signal is downconverted to digital IF and multiplied with a locally generated FH carrier and PN sequence. Then, the output is sampled and processed in the sub-band accumulation module.

The frequency hopping band is divided into K sub-bands and the FH signals are assumed to be uniformly distributed in the original FH band. In each sub-band, there exist N hops. The local replica carrier phase leap in each sub-band is compensated by searching through the TEC range with a large step. Then, the signal power in each sub-band can be coherently integrated as

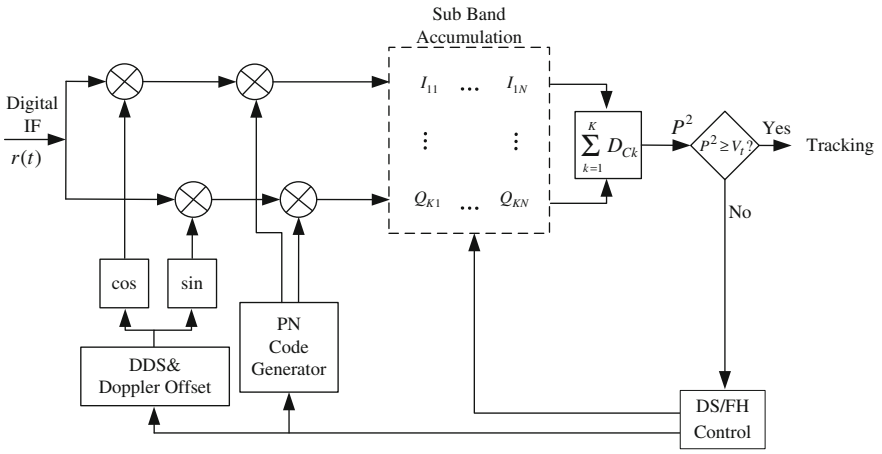

Fig. 22.2 Band-division accumulation method

$$D_{Ck} = (\sum_{n=1}^{N} I_{kn})^2 + (\sum_{n=1}^{N} Q_{kn})^2 \tag{22.13}$$

And the energy in K sub-bands is incoherently accumulated to get

$$P = \sum_{k=1}^{K} D_{Ck} \tag{22.14}$$

If $P > V_{th}$, the initial acquisition can be claimed. The overall accumulation time is

$$\tau_{sum} = K \times N \times T_h \tag{22.15}$$

This acquisition module works simultaneously with the FH pattern generator and search the real value of TEC at the speed of $f_{search} = 1/\tau$. The band division method reduces the TEC search step and shortens the mean acquisition time to a great extent, whereas the sampled digits have to be restored and processed separately according to sub-bands and this shall consume extra hardware resources.

22.4 Performance of Band Division Method

22.4.1 Search Time of TEC

When performing quasi-coherent accumulation in a sub-band, the TEC search step increases and the bandwidth of a sub-band decreases comparing to the case of the original FH bandwidth.

Table 22.1 Search times of the band-division method

No.	K	B/MHz	Δ(TECU)	M
1	1	100	7.47	134
2	5	20	36.65	28
3	10	10	73.13	14

TEC Range (10^{16}–10^{19})

Assuming the original bandwidth to be 100 MHz and maximum estimation error to be $\theta = 20°$, the search times M corresponding to different number of sub-bands K obtained is as shown in Table 22.1.

It can be seen that TEC search time falls when the number of sub-bands increases.

22.4.2 Energy Accumulation

When performing demodulation of DS/FH signals, the frequency hopping band is divided into several sub-bands within which the FH signals are coherently integrated and the integration results in each band are incoherently accumulated, provided that we have obtained an accurate estimation of TEC by searching through the TEC range. In each sub-band a quasi-coherent accumulation with an integration interval $N \times T_h$ is performed to obtain the incoherent accumulation factor D_c. Then, the output of K bands incoherent accumulation is judged to check for acquisition. The error in searching TEC causes a phase estimation deviation in each FH interval and this leads to energy loss in accumulation as follows:

$$L_i = \cos^2 \Delta\varphi_i \tag{22.16}$$

Usually we confine the phase error as

$$\Delta\varphi_B \leq 20° \tag{22.17}$$

And the maximum energy loss in a sub-band is

$$L_B = \cos^2 \theta \tag{22.18}$$

The accumulation gain in a sub-band is

$$G_B = N \cos^2 \theta \tag{22.19}$$

The incoherence among K sub-bands brings in square loss. Square loss [7] is related to the number of sub-bands K and the coherent integration result D_c in each sub-band:

$$L_{ACC} = \left[1 + \sqrt{1 + \frac{9.2K}{D_c}}\right] \Big/ \left[1 + \sqrt{1 + \frac{9.2}{D_c}}\right] \tag{22.20}$$

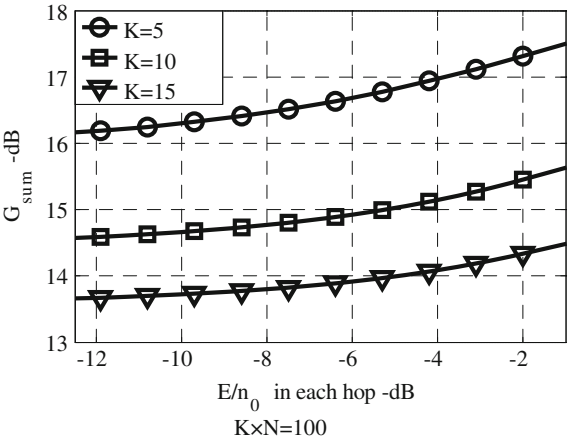

Fig. 22.3 Accumulation gain of the band-division method

The total accumulation gain of the Band Division method during time interval τ_{sum} is

$$G_{SUM} = KN\cos^2\theta\left[1+\sqrt{1+\frac{9.2}{D_c}}\right]\Big/\left[1+\sqrt{1+\frac{9.2K}{D_c}}\right] \tag{22.21}$$

22.5 Simulation Results

Simulation is performed with Matlab 7. S band is used as the radio frequency. The FH bandwidth is 100 MHz and FH rate is 10Khops/s and $C/N_J = -50$ dB. The maximum phase leap estimation error is $\theta = 20^\circ$.

Figures 22.3 and 22.4 show G_{sum} given by (4.6) respectively for $K \times N = 100$ and $K \times N = 200$, forming a comparison of accumulation gain versus number of sub-bands.

Figure 22.5 shows TEC search step given by (3.7).

Figure 22.6 shows TEC search time M while assuming TEC $\in [10^{16}, 10^{19}]$.

It is noted that the Band Division Method shortens the acquisition time by reducing TEC search complexity. However, it increases accumulation loss by introducing incoherent accumulation among sub-bands.

22.6 Conclusions

Ionospheric effects on acquisition of DS/FH signals are analyzed and a solution is presented. First, a C-DS/FH system is described where the frequency hopping carriers are phase-continuous. Then, the effect of dispersion on carrier phase

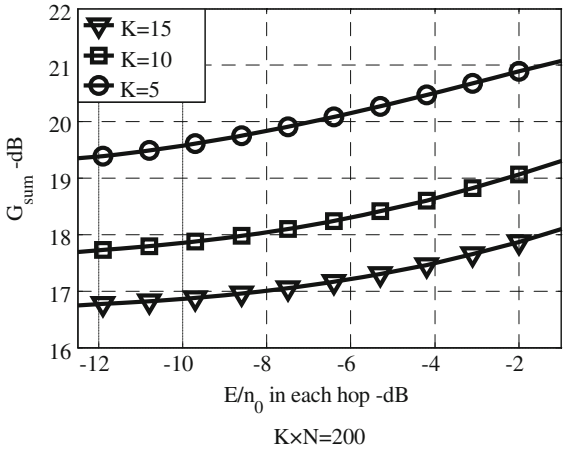

Fig. 22.4 Accumulation gain of the band-division method

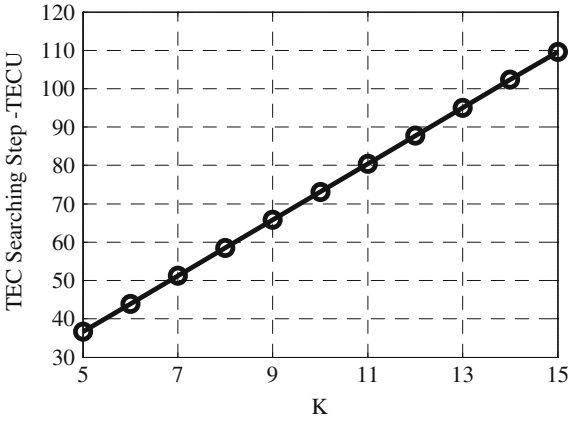

Fig. 22.5 TEC searching period of the band-division method

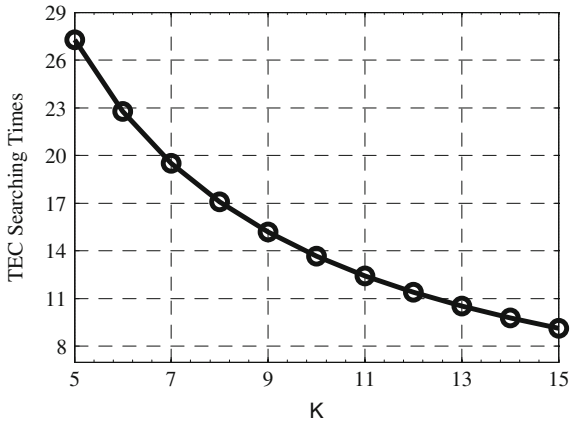

Fig. 22.6 Search complexity of the band-division method

continuity is presented and a search technique is considered to settle the problem. Since the dispersion has a connection with bandwidth, it is possible to reduce search complexity by dividing a FH band into several sub-bands.

The above assumption is justified by the fact that, after searching TEC in a narrower band, the mean acquisition time is reduced at the cost of deteriorating the accumulation performance. In particular, it is found that the receiver can make a compromise between acquisition time and anti-jamming capability by altering the number of sub-bands. This provides more flexibility for design of receivers.

References

1. Cherubini G, Milstein LB (1989) Performance analysis of both hybrid and frequency-hopped phase-coherent spread-spectrum systems-part I: a hybrid DS/FH system. IEEE Trans Commun 37(6):600–611
2. Simone L, Salerno N, Maffei M (2006) Frequency-hopping techniques for secure satellite TT&C: system analysis & trade-offs. Alcatel Alenia Space-Italy, IEEE
3. Zhiqiang Li (2008) Modern satellite telemetry technology. Institute of Comm. PLA University of Science and Technology, Nanjing
4. Zhang B, Shao DR, Li SJ (2005) Study on fast synchronization of hybrid DS/FH system. J Beijing Univ Aeronaut Astronaut 31(11):1226–1231
5. Zhao GB, Liu YF, Guo SL (2011) Research on DS/FH hybrid spread spectrum range rate measurement technology. J Spacecraft TT&C Technol 30(4):20–27
6. Kaplan ED, Hegarty CJ (2006) Understanding GPS-principles and applications second edition. ARTECH HOUSE, INC.685 Canton Street, Norwood
7. Marcum FI (1947) A statistical theory of target detection by pulsed radar, rand research memo. RM-754, December 1947, with Appendix, RM-753, July 1948

Chapter 23
Anti-Fading Analysis of Diversity-Synthesized Technology

Mingxin Kou and Jun Yan

Abstract This paper analyzed the anti-fading performance of diversity-synthesized technology in FM telemetry engineering practice based on FM Baseband device principles and algorithms, designed experiment to simulate the situation that one direction rotational signal flash in engineering practice and tested the telemetry bit error rate. Through the discussion of device principle, experimental validation and engineering experiences, this paper analyzed the impact of differential-mode loop's frequent loss of lock in FM telemetry engineering practice. It is normal in FM telemetry engineering practice that the fast fading of FM telemetry one direction rotational signal in a short time causes differential-mode loop loss of lock frequently, the technology of diversity-synthesized can be used as anti-fading effectively in FM telemetry engineering.

Keywords FM modulation baseband · Differential-mode loop · Common-mode loop · Diversity-synthesized · Fading

23.1 Introduction

In a spacecraft FM telemetry engineering practice, it happened that a few seconds of telemetry one direction rotational signal flashed within rocket attitude adjustment phase, with the poor signal quality and frequently loss of lock of differential-mode loop. Although common-mode loop lock, code synchronization of telemetry data, frame synchronization and sub-frame synchronization are normal in practice,

M. Kou (✉) · J. Yan
China Satellite Maritime Tracking and Control Department, Jiangyin 214431, China
e-mail: kasimkou@gmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_23,

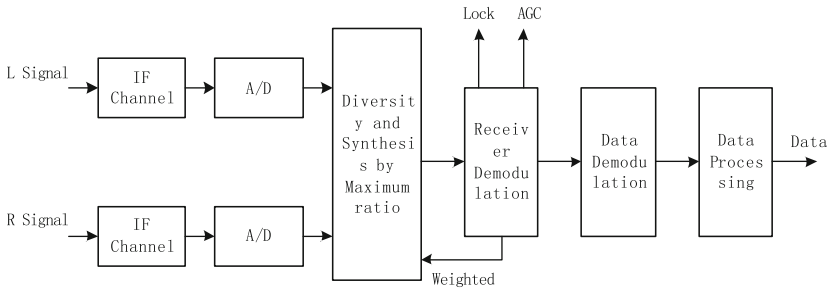

Fig. 23.1 The workflow of FM telemetry-baseband

frequently loss of lock of differential-mode loop may cause a degree of telemetry bit error, so this paper was wrote to analyse it.

23.2 Profile of FM Baseband Workflow

FM telemetry baseband transforms left and right rotational signal, which is sent by IF switch network, filtered through IF analog channel and amplified by AGC, from the analog domain to digital domain by A/D converter. Digital IF sampling signal is sent into FPGA and DSP, we complete different kinds of digital signals processing using FM telemetry demodulator software algorithms in the digital domain, thus, achieving the function of FM signal's diversity-synthesized, carrier wave demodulation, code synchronization, time code generation/demodulation, and so on. Figure 23.1 shows the workflow of FM telemetry-baseband [1].

23.3 Factors that Affect Downlink Frequency Modulated Left and Right Rotational Signal

The polarization of the FM signal is the linear polarization. Generally, there is no obvious difference in strength between the received left rotational signal and right rotational signal. The reasons for the undulating of one direction rotational signal includes rockets attitude, rockets plume, Doppler shift and the shake of attitude adjustment phase signal. This paper analyzed the impact of the one direction rotational signal flash. Detailed discussion of its causes was not included.

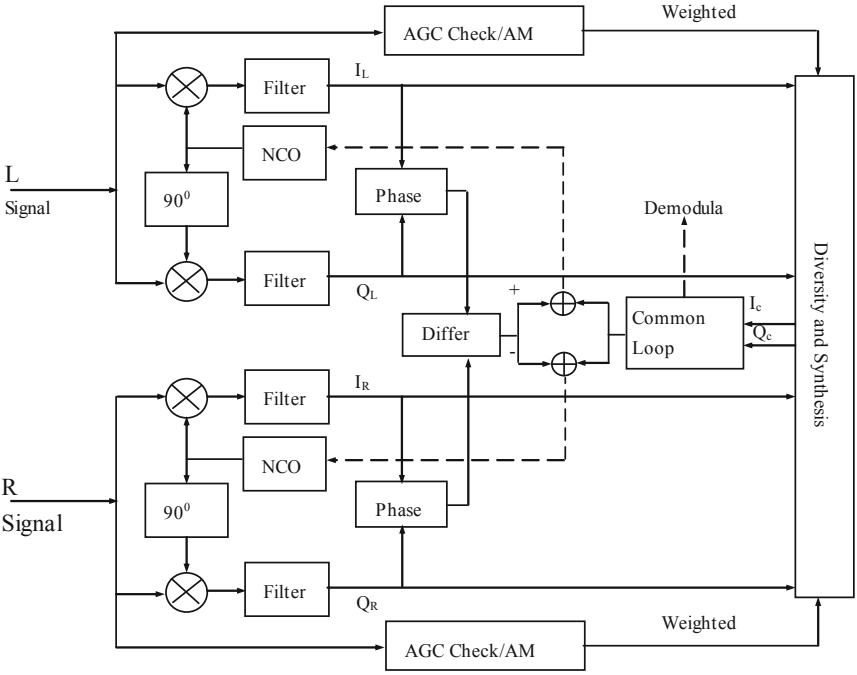

Fig. 23.2 The algorithm processes of FM diversity reception demodulation module

23.4 Impact Analysis of FM Baseband Differential-Mode Loop Lose Lock

Figure 23.2 shows the algorithm processes of FM diversity reception demodulation module, which is the key of the impact analysis of FM baseband differential-mode loop loss of lock.

23.4.1 The Analysis of Diversity-Synthesized Module

FM telemetry receiver uses inspection before polarization synthesis techniques. Diversity-synthesized module determines the weighting factor by AGC/AM weighted approach, to achieve the left and right rotational signal's maximal-ratio synthesis.

Table 23.1 The AGC and SNR records when telemetry signal was stable after the task target was found in telemetry engineering practice

	Baseband A			Baseband B		
	L Signal	R Signal	Synthesized	L Signal	R Signal	Synthesized
AGC(V)	2.85	3.22	–	3.49	3.16	–
S/φ(dB)	22.7	23.2	25.4	22.8	22.8	25.8
Doppler(kHz)	72 kHz			72 kHz		

23.4.1.1 The Analysis of Diversity-Synthesized Algorithm

In theory, diversity-synthesized algorithm is that there would be about 3 dB SNR gain after left and right rotational signal are processed by maximal-ratio diversity-synthesized on condition that the signal are normal, the weighting factor of one direction rotational signal becomes larger when it becomes stronger, the weighting factor would be 1 when there is only one direction rotational signal (Table 23.1 shows the AGC and SNR records when telemetry signal was stable after the task target was found). In the telemetry engineering practice, left rotational signal flash off caused the adjustment of its weighting factor, which may have an effect on telemetry bit error rate. But without the comparison of rocket raw data and telemetry demodulation data, it is hard to determine the degree of the influence.

23.4.1.2 The Analysis of AGC Detection Circuits

AGC detection circuits get out digital control output of the AGC voltage after amplitude detecting, digital filtering the signal and AGC time constant controlling. The AGC digital control output completes AGC closed-loop operation by exporting analog control voltage using D/A converter and sending it to the IF analog channel's AGC amplifier, in order to keep IF signal level which is A/D collected in the dynamic range of the input signal stable.

23.4.1.3 The Method of Diversity-Synthesized Weighting Factor Calculation

In the task, the AGC time constant's setting of the receiver is 50 ms, so when the RF signal fading rate are up to 200 Hz or above, weighting coefficient would change lag behind the signal changes, if we only use the receiver AGC voltage on the maximal-ratio synthesizer to weight, thus, synthesized signal is distorted, which affects the receiver demodulation performance. For this reason, when designing FM baseband, we detect the AM envelope which is not tracked by AGC in IF signal, add it with the AGC voltage after processed, get maximal-ratio weighting factor and send it to diversity-synthesized. When the signal is fading fast, maximal-ratio diversity-synthesized will react and calculate the weighted coefficient quickly.

For the design of diversity-synthesized algorithm, if there is a stable hand signal, we can prevent the flash of synthesized signal when one direction rotational signal flashes using the method of combine AGC and AM envelope to calculate the weighted coefficient. It is a timely and effective way to ensure the stability of the synthesized signal.

23.4.2 The Analysis of FM Baseband Symmetric Diversity of Phase-Locked Loop

FM baseband symmetric diversity of phase-locked loop consists of two key parts: differential-mode loop and common-mode loop. Common-mode loop is to track the input signal common frequency and phase changes, differential-mode loop is to track the frequency and phase change between the two input signals and symmetry control of the two input signals corresponding to the local oscillator NCO frequency and phase.

We can get the local carrier wave and the input signal common-mode error after two synthetic orthogonal baseband signals are digital frequency discriminated inside FPGA and loop filtered. Left and right rotational I and Q quadrature signal together are sent into differential-mode loop, to get differential-mode error. Common-mode error is sent to the left and right rotational signal two NCO and pushed to a common frequency in the same direction to ensure that the receiver IF signal fall to the centre of the filter bandwidth. Differential-mode error is divided into two and added to the digitally controlled oscillator in the opposite polarity, so that two-way local carriers are pushed to the opposite direction on a public frequency. Thus, through the sum of the differential-mode signal and common-mode error signal, we control the output frequency of the digitally controlled oscillator, which makes the two signals input to the sub-collection of synthesizers always maintain the same frequency and phase and fall on the design of filter passband centre frequency. Carrier wave sends out data directly after demodulation to back-end code synchronization and frame synchronization module.

The design of symmetric diversity of phase-locked loop can guarantee the continuity of tracking and avoid data loss caused by signal deep fading.

23.4.3 The Analysis of Carrier Wave Demodulation Algorithm

Time domain expression of left and right rotational signal received in FM demodulation module [2] is:

$$s_{FM}(t) = A_0 \cdot \cos[\omega_0 t - \int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau] \tag{23.1}$$

In the formula above, ω_0 is the carrier wave angular frequency, A_m is the modulation signal amplitude, $m(t)$ is the modulation signal, k_f is the modulation index. The received signal and the digital local oscillator signals are to pass the orthogonal mixer and low pass filter, which can get left and right rotational orthogonal baseband signal IL(t), QL(t) and IR (t), QR(t).

$$I_L(t) = A_L \cos(\int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_L) \tag{23.2}$$

$$Q_L(t) = A_L \sin(\int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_L) \tag{23.3}$$

$$I_R(t) = A_R \cos(\int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_R) \tag{23.4}$$

$$Q_R(t) = A_R \sin(\int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_R) \tag{23.5}$$

After maximal-ratio diversity-synthesized, left and right rotational orthogonal baseband signal will become synthesized orthogonal baseband signal Ic(t), Qc(t).

$$I_C(t) = A_C \cos(\int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_C) \tag{23.6}$$

$$Q_C(t) = A_C \sin(\int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_C) \tag{23.7}$$

Orthogonal baseband signal contains all the information of the input signal, which is fully able to express the amplitude and phase characteristics of the input signal. After synthesized orthogonal signal phase detecting, we can get instantaneous phase of modulation signal $\theta_m(t)$,

$$\theta_m(t) = \operatorname{arctg} \frac{\mathrm{Ic}(t)}{\mathrm{Qc}(t)} = \int_{-\infty}^{t} k_f \cdot A_m \cdot m(\tau) d\tau + \theta_c \tag{23.8}$$

After differential processing the instantaneous phase, we can get the original modulation signal $f(t) = A_m m(t)$, completing the FM demodulation.

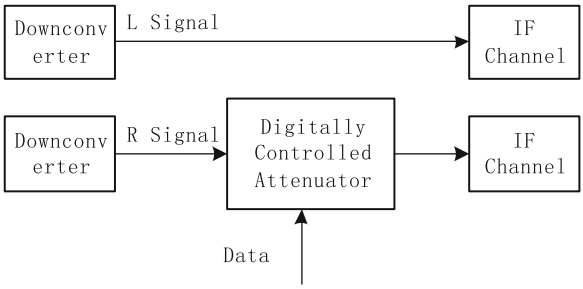

Fig. 23.3 The procedure of experimental design

Orthogonal baseband signal is diversity-synthesized by left and right rotational orthogonal baseband signal in maximal-ratio, we can see from the expression that synthesis effect has something to do with single rotational signal's frequency and phase; it has nothing to do with amplitude. Stability of the amplitude of the signal does not affect the synthesis results, but in fact, the degree of influence of the magnitude stability is closely connected with the achieve method of devices, no specifically discussed here.

23.5 Experiment Design and Test Analysis

Because it is difficult to simulate differential-mode loop frequent loss of lock in the real FM telemetry, here we design a simple experiment to start quantitative analysis.

23.5.1 Experimental Design and Recording

Fade the left and right rotational signal using the foremast bit error rate testing, fade the left rotational signal to the signal strength of properly lock, fade the right rotational signal to AGC threshold, then control the right rotational signal's strength by digitally controlled attenuator, to make differential-mode loop loss of lock frequently (you can control the differential-mode loop loss of lock by adjusting the attenuation and frequency), observe and record error rate. Figure 23.3 shows the experimental procedure.

Table 23.2 shows the test results.

Table 23.2 Test record sheet of bit error rate

Attenuation frequency (Hz)	AGC and S/N			
	2.5 V/20 dB	1.5 V/13 dB	1.5 V/12.8	0.3 V/12.8
2	0	3.22e-7	7.07e-5	7.43e-5
0.5	0	3.23e-7	5.73e-5	5.94e-5
0	0	3.01e-7	4.52e-5	4.88e-5

23.5.2 Difference Between Experimental Environment and Real Telemetry Environment

In the experimental environment, telemetry channel is stable, left-hand signal is always stable. While in real telemetry environment, left rotational signal and right rotational signal will fade or increase at same time, or one fade, the other increase.

In the experimental environment, signal attenuation adjustment happened in the frequency converter export, attenuator fades both the signal and noise, SNR is unchanged. In real telemetry environment, many factors will cause fading, man-made fading cannot reproduce the real telemetry environment.

23.5.3 The Analysis of Result

Seen from the results, when AGC lock is normal and SNR is much larger than the threshold, a single rotational signal flash will only lead to differential-mode loop loss of lock frequently, and it will not cause the telemetry bit error. When SNR is close to the threshold, a single rotational signal flash will cause differential- mode loop loss of lock frequently, and it will have a certain impact on the telemetry bit error, but the effect is not obvious. We also get that the technology of maximal-ratio diversity synthesized algorithm and diversity-synthesized PLL does a poor job in improving SNR, it cannot reach the ideal situation of 3 dB gain, the key is to ensure that common route signal can normally received and demodulated in the case of single rotational signal fast fading.

23.6 Conclusions

Seen from the principle of FM baseband receiver demodulation of telemetry data and the results of experiment design, differential-mode loop loss of lock frequently will not affect the telemetry data if a single rotational signal's AGC and SNR are normal. Seen from feedback on use of data, tasks arc telemetry data is complete

and effective. It is normal that in FM telemetry engineering practice the fast fading of FM telemetry single rotational signal in a short time causes differential-mode loop loss of lock frequently, the technology of diversity-synthesized can play the role of anti-fading effectively in test tasks.

References

1. Beijing research institute of telemetry (2007) TW-113 multi-function digital baseband design report
2. Beijing research institute of telemetry (2006) Telemetry technology

Chapter 24
Analysis on the Application of Feed-Forward Technology for Space Tracking, Telemetering and Control Ships

Dingxin Yang and Ting Yuan

Abstract Shipborne radar devices are influenced by ship-swaying in terms of their tracking accuracy, especially under bad sea conditions, which making it very important to employ a technique which not only maintains tracking stability but also isolates ship-swaying effects. A scheme which compensates antenna azimuth and elevation angles for feed-forward controls is proposed. First the radar ship-swaying feed-forward compensation technique is introduced, followed by feed-forward equations deduced by compensating tracking errors caused by ship-swaying and target movements. Then, the deduction results are analyzed and concluded that ship-swaying feed-forward data can be validly produced without target distance values provided.

Keywords Ship-swaying · Feed-forward compensation · Azimuth · Elevation

24.1 Introduction

In order for a radar to correctly track a target, when the target deviates from the zero point of the antenna difference beam, the tracking receiver has error voltage signals which is transferred into the driven antenna of the servo system to eliminate the error voltage. If the extent where the target deviates from the zero point of the antenna difference beam is within the extent where the radar tracking converges, the target can be normally tracked. However, because of ship swaying and target movements, shipborne radars inevitably produce greater tracking errors.

D. Yang (✉) · T. Yuan
China Satellite Maritime Tracking and Control Department, Jiangyin 214431, China
e-mail: ydxhit@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_24,

When the velocity and the acceleration of either the target or the swaying ship are much greater, there are more target-missing risks. In order to overcome ship-swaying effects, it is requested that the velocity and the acceleration of the radar should be greater than those of the target. However, only with the inherent radar tracking loop circuit and gyroscope stabilization will realizing expected ship-swaying isolation accuracy be a challenge. Employing feed-forward stabilization technique will substantially improve system performance. As for shipborne radars, installing a gyroscope onto the antenna pedestal is commonly used to overcome ship-swaying influences. It has been proved that, as for wide-beam, low-accuracy radars, gyroscopes have the advantages of feasibility and simplicity. With the development of the spacecraft technology and the advance of the target dynamic performance, stricter demands are imposed on the radar tracking performance. Therefore, as for the narrow-beam, high-accuracy radars, feed-forward compensation techniques are essential apart from gyroscopes.

As for space tracking, telemetering and control ships, radar feed-forward data is computed with the target's position, velocity and acceleration as well as ship-swaying angles, ship-swaying velocities and ship-swaying accelerations. The above values are forecast by the shipborne central computer system using accumulating data collected by various measuring devices.

24.2 Mathematical Feed-Forward Model

Mathematically, ship-swaying feed-forward is formulated by deducing prospective target movement values to compute antenna azimuth and elevation velocities and accelerations produced by the target movements and ship swaying, according to measuring data.

24.2.1 Relationship Between Radar Angular Velocities and Surveying Coordinate Velocities

Under a surveying coordinate system, rectangular coordinates can be described by polar coordinates as follows [1]:

$$\begin{cases} X_c = R\cos E\cos A \\ Y_c = R\sin E \\ Z_c = R\cos E\sin A \end{cases} \tag{24.1}$$

We take derivatives of Eq. 24.1 with respect to time which is omitted in the equations:

$$\begin{cases} \dot{X}_c = \dot{R}\cos E\cos A - \dot{E}R\sin E\cos A - \dot{A}R\cos E\sin A \\ \dot{Y}_c = \dot{R}\sin E + \dot{E}R\cos E \\ \dot{Z}_c = \dot{R}\cos E\sin A - \dot{E}R\sin E\sin A + \dot{A}R\cos E\cos A \end{cases} \tag{24.2}$$

We write Eq. 24.2 using the matrix form:

$$\begin{bmatrix} \cos E\cos A & -R\sin E\cos A & -R\cos E\sin A \\ \sin E & R\cos E & 0 \\ \cos E\sin A & -R\sin E\sin A & R\cos E\cos A \end{bmatrix} \begin{bmatrix} \dot{R} \\ \dot{E} \\ \dot{A} \end{bmatrix} = \begin{bmatrix} \dot{X}_c \\ \dot{Y}_c \\ \dot{Z}_c \end{bmatrix} \tag{24.3}$$

We can have the determinant value of the leftside 3 × 3 matrix:

$$\Delta = R^2\cos E \tag{24.4}$$

Let:

$$\Delta_A = \det\begin{bmatrix} \cos E\cos A & -R\sin E\cos A & \dot{X}_c \\ \sin E & R\cos E & \dot{Y}_c \\ \cos E\sin A & -R\sin E\sin A & \dot{Z}_c \end{bmatrix}$$

$$\Delta_E = \det\begin{bmatrix} \cos E\cos A & \dot{X}_c & -R\cos E\sin A \\ \sin E & \dot{Y}_c & 0 \\ \cos E\sin A & \dot{Z}_c & R\cos E\cos A \end{bmatrix}$$

That is:

$$\Delta_A = -\dot{X}_c R\sin A + \dot{Z}_c R\cos A \tag{24.5}$$

$$\Delta_E = -\dot{X}_c R\sin E\cos E\cos A + \dot{Y}_c R\cos^2 E - \dot{Z}_c R\sin E\cos E\sin A \tag{24.6}$$

When E≠ 90°, Δ ≠ 0, according to the Cramer's rule, we have:

$$\dot{A}_c = \frac{\Delta_A}{\Delta} = \frac{-\dot{X}_c\sin A + \dot{Z}_c\cos A}{R\cos E} \tag{24.7}$$

$$\dot{E}_c = \frac{\Delta_E}{\Delta} = \frac{-\dot{X}_c\sin E\cos A + \dot{Y}_c\cos E - \dot{Z}_c\sin E\sin A}{R} \tag{24.8}$$

With Eqs. 24.7 and 24.8, we have that the antenna azimuth angular velocity and elevation angular velocity are related to surveying coordinate velocities, and the latter is also related to ship-swaying attitude angles and parameters of dynamic target characteristics.

Similarly, we can also have the relationship between antenna angular velocities and coordinate parameters under the horizontal coordinate system:

$$\dot{A}_L = \frac{\Delta_A}{\Delta} = \frac{-\dot{X}_L \sin A + \dot{Z}_L \cos A}{R \cos E} \tag{24.9}$$

$$\dot{E}_L = \frac{\Delta_E}{\Delta} = \frac{-\dot{X}_L \sin E \cos A + \dot{Y}_L \cos E - \dot{Z}_L \sin E \sin A}{R} \tag{24.10}$$

The radar tracking error is a function of the velocity error and the velocity error is proportional to the angular velocity. As for land-based radars, the velocity error is only related to target movements. As for shipborne radars, the velocity error is related to both target movements and ship-swaying parameters.

24.2.2 Deduction of Feed-Forward Equations

We can take coordinate transformations between surveying coordinate systems and horizontal coordinate systems. With the help of orthogonal transformations, we can have a new orthogonal coordinate system from the old one [1]. We firstly transform the surveying coordinate system to the corresponding horizontal one:

$$\begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} = M_K M_\varphi M_\theta \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{24.11}$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = M_\theta^T M_\varphi^T M_K^T \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} \tag{24.12}$$

In the above equations, M_K, M_φ, M_θ represent yawing, pitching and rolling parameter matrices, respectively. The superscript T means transposition. We also have:

$$M_K = \begin{bmatrix} \cos K & 0 & -\sin K \\ 0 & 1 & 0 \\ \sin K & 0 & \cos K \end{bmatrix}$$

$$M_\varphi = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

We take derivatives of Eq. 24.12 with respect to time and then take Eq. 24.11 into the result:

$$\begin{bmatrix} \dot{X}_c \\ \dot{Y}_c \\ \dot{Z}_c \end{bmatrix} = \left[\dot{M}_\theta^T M_\theta + M_\theta^T \dot{M}_\varphi^T M_\varphi M_\theta + M_\theta^T M_\varphi^T \dot{M}_K^T M_K M_\varphi M_\theta\right] \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} + M_\theta^T M_\varphi^T M_K^T \begin{bmatrix} \dot{X}_L \\ \dot{Y}_L \\ \dot{Z}_L \end{bmatrix} \tag{24.13}$$

From Eq. 24.13, we have that surveying coordinate velocities are related to ship attitude parameters, surveying coordinate parameters and horizontal coordinate velocities. In other words, surveying coordinate velocities are related to ship attitude parameters and dynamic target parameters. We then unfold the matrices:

$$\dot{M}_\theta^T = \dot{\theta} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\sin\theta & \cos\theta \\ 0 & -\cos\theta & -\sin\theta \end{bmatrix}$$

$$\dot{M}_\varphi^T = \dot{\varphi} \begin{bmatrix} -\sin\varphi & -\cos\varphi & 0 \\ \cos\varphi & -\sin\varphi & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\dot{M}_K^T = \dot{K} \begin{bmatrix} -\sin K & 0 & \cos K \\ 0 & 0 & 0 \\ -\cos K & 0 & -\sin K \end{bmatrix}$$

$$\dot{M}_\theta^T M_\theta = \theta \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$M_\theta^T M_\varphi^T \dot{M}_K^T M_K M_\varphi M_\theta = \dot{K} \begin{bmatrix} 0 & \sin\theta\cos\varphi & \cos\theta\cos\varphi \\ -\cos\theta\sin\varphi & 0 & -\sin\varphi \\ -\cos\varphi\cos\theta & \sin\varphi & 0 \end{bmatrix}$$

$$M_\theta^T \dot{M}_\varphi^T M_\varphi M_\theta = \dot{\varphi} \begin{bmatrix} 0 & \cos\theta & -\sin\theta \\ -\cos\theta & 0 & 0 \\ \sin\theta & 0 & 0 \end{bmatrix}$$

We take the above six equations into Eq. 24.13 and we then have:

$$\dot{M}_\theta^T M_\theta + M_\theta^T \dot{M}_\varphi^T M_\varphi M_\theta + M_\theta^T M_\varphi^T \dot{M}_K^T M_K M_\varphi M_\theta$$
$$= \begin{bmatrix} 0 & \dot{\varphi}\cos\theta + \dot{K}\sin\theta\cos\varphi & -\dot{\varphi}\sin\theta + \dot{K}\cos\varphi\cos\theta \\ -\dot{\varphi}\cos\theta - \dot{K}\sin\theta\cos\varphi & 0 & \dot{\theta} - \dot{K}\sin\varphi \\ \dot{\varphi}\sin\theta - \dot{K}\cos\theta\cos\varphi & -\dot{\theta} + \dot{K}\sin\varphi & 0 \end{bmatrix} \tag{24.14}$$

We also have:

$$M_\theta^T M_\varphi^T M_K^T$$
$$= \begin{bmatrix} \cos\varphi\cos K & \sin\varphi & \cos\varphi\sin K \\ -\cos K\cos\theta\sin\varphi - \sin\theta\sin K & \cos\theta\cos\varphi & \sin\theta\cos K - \cos\theta\sin\varphi\sin K \\ \sin\theta\sin\varphi\cos K - \sin K\cos\theta & -\sin\theta\cos\varphi & \sin\theta\sin\varphi\sin K + \cos\theta\cos K \end{bmatrix} \tag{24.15}$$

We take Eqs. 24.14 and 24.15 into 13:

$$\begin{aligned} \dot{X}_c = {} & (\dot{\varphi}\cos\theta + \dot{K}\sin\theta\cos\varphi)Y_c + (\dot{K}\cos\theta\cos\varphi - \dot{\varphi}\sin\theta)Z_c \\ & + \cos\varphi\cos K\dot{X}_L + \sin\varphi\dot{Y}_L + \cos\varphi\sin K\dot{Z}_L \end{aligned} \tag{24.16}$$

$$\begin{aligned} \dot{Y}_c = {} & (\dot{\varphi}\cos\theta - \dot{K}\sin\theta\cos\varphi)X_c + (\dot{\theta} + \dot{K}\sin\varphi)Z_c \\ & + (-\cos\theta\sin\varphi\cos K - \sin\theta\sin K)\dot{X}_L + (\cos\theta\cos\varphi)\dot{Y}_L \\ & + (\sin\theta\cos K - \cos\theta\sin\varphi\sin K)\dot{Z}_L \end{aligned} \tag{24.17}$$

$$\begin{aligned} \dot{Z}_c = {} & (\dot{\varphi}\cos\theta - \dot{K}\cos\theta\cos\varphi)X_c + (\dot{K}\sin\varphi - \dot{\theta})Y_c \\ & + (\sin\theta\sin\varphi\cos K - \cos\theta\sin K)\dot{X}_L + (-\sin\theta\cos\varphi)\dot{Y}_L \\ & + (\sin\theta\cos\varphi\sin K + \cos\theta\cos K)\dot{Z}_L \end{aligned} \tag{24.18}$$

At last, we take Eqs. 24.1, 24.16, 24.17 and 24.18 into 24.7 and 24.8, and then we can have the angular velocities under the surveying coordinate system. The acceleration can be obtained by take derivatives of the velocity equations with respect to time.

24.3 Analysis on Application of Feed-Forward Technique

As the center of shipborne device data processing, the central computer system provides radar devices with feed-forward data including that caused by target movements and that caused by ship swaying. Target feed-forward data is described as $\dot{A}_1$ (azimuth angular velocity), $\dot{E}_1$ (elevation angular velocity) and $\ddot{A}_1$ (azimuth

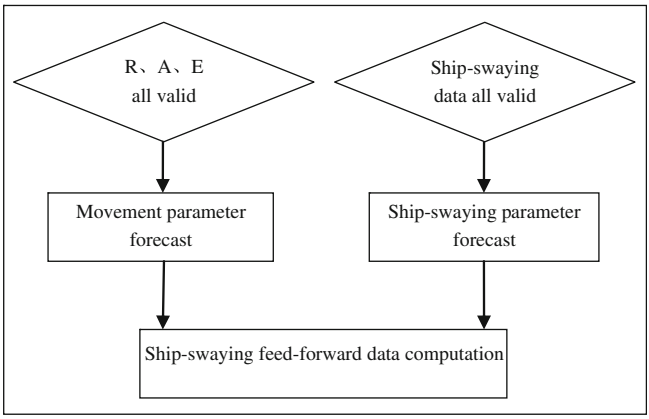

Fig. 24.1 Procedure of ship-swaying feed-forward computation

angular acceleration), $\ddot{E}_1$ (elevation angular acceleration). Ship-swaying feed-forward data is described as $\dot{A}_2$ (azimuth angular velocity), $\dot{E}_2$ (elevation angular velocity) and $\ddot{A}_2$ (azimuth angular acceleration), $\ddot{E}_2$ (elevation angular acceleration). All the above data is computed by the deduction introduced in the previous section and is implemented by computer algorithms [1–4], which is also described in Fig. 24.1.

From the computation procedure, we find that it is necessary for ship-swaying feed-forward computation to have valid azimuth, elevation and distance values. Combined with applications, we have two categories of targets, one category includes geostationary satellites or beacon towers, the other category includes moving targets such as rockets, spaceships, non-geostationary satellites and beacon balloons. We next analyze the application of feed-forward techniques according to the above two categories.

24.3.1 Analysis on Application for Targets of the First Category

From analyzing target movements of the first category, we find that, as these targets are geostationary, the velocity and the acceleration values are both zero under the horizontal coordinate system when ignoring the ship velocity. With Eqs. 24.9 and 24.10, we have $\dot{A}_L = 0$ and $\dot{E}_L = 0$, that is, the target feed-forward value is zero. We then take this result into Eqs. 24.16, 24.17 and 24.18:

$$\dot{X}_c = (\dot{\varphi}\cos\theta + \dot{K}\sin\theta\cos\varphi)Y_c + (\dot{K}\cos\theta\cos\varphi - \dot{\varphi}\sin\theta)Z_c \tag{24.19}$$

$$\dot{Y}_c = (\dot{\varphi}\cos\theta - \dot{K}\sin\theta\cos\varphi)X_c + (\dot{\theta} + \dot{K}\sin\varphi)Z_c \tag{24.20}$$

$$\dot{Z}_c = (\dot{\varphi}\cos\theta - \dot{K}\cos\theta\cos\varphi)X_c + (\dot{K}\sin\varphi - \dot{\theta})Y_c \tag{24.21}$$

We then take Eqs. 24.1, 24.19, 24.20 and 24.21 into 7 and 8:

$$\begin{aligned}\dot{A}_c = {} & \dot{\varphi}\sin\theta - \dot{K}\cos\varphi\cos\theta - \tan E_c[(\dot{\theta} - \dot{K}\sin\varphi)\cos A_c \\ & + (\dot{\varphi}\cos\theta + \dot{K}\cos\varphi\sin\theta)A_c]\end{aligned} \tag{24.22}$$

$$\dot{E}_c = -(\dot{\varphi}\cos\theta + \dot{K}\cos\varphi\sin\theta)\cos A_c + (\dot{\theta} - \dot{K}\sin\varphi)\sin A_c \tag{24.23}$$

We also have the ship-swaying acceleration: $\ddot{A}_c = (\dot{A}_c)'$, $\ddot{E}_c = (\dot{E}_c)'$.

From the final result, we find that there are no distance variables in the equations and the equations are only related to ship-swaying conditions. Therefore, if a target of the first category is being tracked, feed-forward data can be computed without target distance values. All that is needed is an angular transformation from under the horizontal coordinate system to under the surveying coordinate system. We can conclude that feed-forward of geostationary targets only originates from ship swaying.

24.3.2 Analysis on Application for Targets of the Second Category

When tracking moving targets, the corresponding feed-forward data can be computed using Eqs. 24.9 and 24.10. In these situations, it is essential to have target distance values in order to obtain feed-forward data, as target velocities and accelerations all result from target distance values and the latter imposes great influence on results of the two equations mentioned before.

Although feed-forward data cannot be obtained without distance values, self-tracking can help the antenna steadily and accurately focus on the target without ship swaying, as the antenna's self-tracking itself has the isolation effect with valid angular and distance values collected during self-tracking periods. Therefore, we can regard that the antenna's self-tracking has isolated the final feed-forward result from target movements and all we need here is the feed-forward from ship swaying.

When we have only valid angular data in hand, we can use the equations deduced previously to compute ship-swaying feed-forward data for the servo system, helping to improve tracking accuracy.

24.4 Conclusions

In this paper, we firstly introduce the antenna feed-forward technology, deduce the mathematical feed-forward model in detail and then analyze the application based on the software implementation of the shipborne central computer system. Finally,

we conclude that ship-swaying feed-forward data can be obtained without target distance values.

Feed-forward technology plays an important role in improving antennas' tracking performance, eliminating ship-swaying influence on target tracking conditions. However, there are also insufficiencies with respect to the technology. The software implementation, for instance, has a substantial effect on feed-forward accuracy which is partly determined by the accuracy of forecast algorithms. As the advance of real-time processing capability and computation accuracy, the feed-forward performance will surely be optimized in the future.

References

1. Jian SL (2009) Introduction on maritime tracking and control technology for space tracking, telemetering and control ships. National defense industry press, Beijing
2. Liu LS (2000) Post-flight data processing of trajectory measurement. National defense industry press, Beijing
3. Qu YX (2009) Ship-swaying stabilization technology of tracking, control and communication equipment for space tracking, telemetering and control ships. National defense industry Press, Beijing
4. Zhang ZH (2009) Methodology of ship attitude data processing for space tracking, telemetering and control ships. National defense industry press, Beijing

Chapter 25
A Telemetry Data Fusion Method Based on Optimal Weighted

Ping Jiang, Yangwei Dong and Xuemei Zou

Abstract To solve the difficult problem of multi-sources telemetry data use in spaceflight TT&C, data pretreatment procedure and data fusion method are studied in this paper. On basis of the existing telemetry data processing, a new data pretreatment procedure is designed, realizing the function of data screening, extracting and selecting. A new data fusion method based on optimal weighted is put forward as well. The algorithm and distribution principle for optimal weighted values are discussed and presented. In the end, after the data pretreatment and fusion stages, a more accurate and reliable data result is provided. Theoretical analysis and real mission data implication show that data quality after pretreatment and fusion is improved obviously. It enhances the precision of data processing, and reduces risks caused by error data. The method has great value in engineering application, and can make reference for the further research of multi-sources data processing.

Keywords Multi-sources telemetry · Data screening · Data selecting · Data confusion · Optimal weighted

25.1 Introduction

Telemetry is a science of using automatic equipment to make scientific measurements and transmit them by radio to a receiving station. As for spaceflight mission, the working and environment state parameters of spacecraft subsystems,

P. Jiang (✉) · Y. Dong · X. Zou
Beijing Spaceflight Control Center, Beijing 100094, China
e-mail: jiangpinghere@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_25,

as well as the physiology state of astronauts in manned spacecraft are collected to be telemetry data, and is transferred to ground tracking stations, and finally delivered to the flight control center. Flight control center receive and process the telemetry data, and finally translate them into all kinds of parameters, according to which the flight control personnel monitor the mission state and make control decisions.

With the rapid development of Chinese aerospace in recent years, the scale of TT&C network is steadily amplified, leaving multi-sources telemetry data transmitted into flight control center at the same time. Considering the different performance of every channel affected by space electromagnetic interference and equipment, it means a lot to find a method to make better use of the multi-sources telemetry data and to improve the accuracy of data processing results.

25.2 Multi-Sources Data Fusion

Data fusion [1] is a technique of data colligating and processing, which is actually an integration and application of lots of traditional subjects and new technology. In a broad sense, it includes communication, pattern recognition, decision making theory, uncertainty theory, signal processing, estimation theory, optimal technique, computer science, artificial intelligence and neural network, etc. Mangolini defined it as that data fusion is a technique of multi-sources data processing by using a series of technology, tools and methods which improves the quality of data [2].

Considering the characteristics of telemetry data of spaceflight, we make use of the concept of data fusion. With multi-sources telemetry data fusion, we combine telemetry data transmitted from different channels by using some proper methods, and provide an optimal evaluation of flight vehicle state.

After using the multi-sources telemetry data fusion method, the influence of channels and receiving equipments can be dispelled, and the frame loss rate and BER can be reduced. A unique succession and reliable telemetry data is made.

25.3 Multi-Sources Telemetry Data Processing

Telemetry data processing [3] is a procedure of extracting, analyzing and synthesizing. One of the basic missions of flight control center is to process all kinds of telemetry data frames that received in real time, and supply users with result frames. The flow of work is shown in Fig. 25.1.

At present, flight control centre processes and uses telemetry data mostly in the way as follows: All the telemetry data transmitted from all stations are processed and displayed at the same time, and users selected what they need separately. Technicians make comprehensive judgment of the state of flight vehicle by all the results. Telecontrol and other software select one result on basis of respective

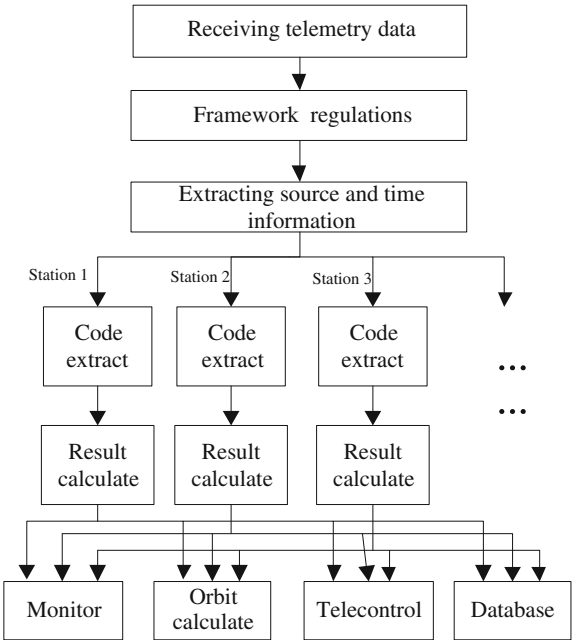

Fig. 25.1 Work flow of telemetry data processing

principles. In this work mode, risks are made because of the different use of multi-sources data. If we realize the data fusion at the data processing stage, the result delivered to the users will be unique and more reliable.

25.4 Multi-Sources Telemetry Fusion Method

The final aim of multi-sources telemetry data fusion is to make a result with low BER covering the whole course from all tracking stations. To make this, in off line telemetry processing, we usually select data of low frame error rate from the multi-sources data in sections. But this method doesn't work in real-time processing, because it needs a period of time to make statistics of frame loss rate and frame error rate. And it still has the problem of error bits.

Considering the problem of frame loss and bit error, in our data processing system, we design a new pretreatment function. It makes code selection and fusion initially before processing, which guarantees the data quality and promotes efficiency of system.

The pretreatment of telemetry data includes four stages as data screening, code extracting, data selecting and data fusion. Flow of work with the new pretreatment function is shown in Fig. 25.2.

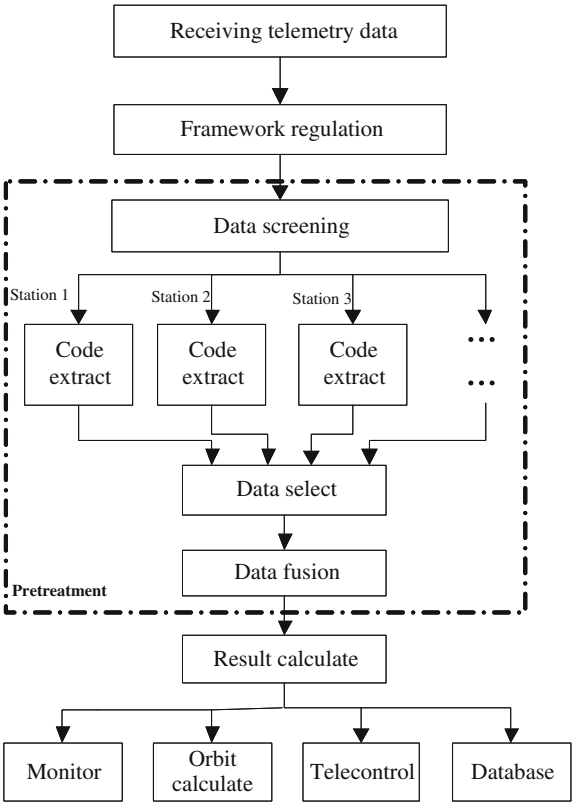

Fig. 25.2 New work flow of telemetry data processing

25.4.1 Data Screening

Data screening basically reviews the rationality and continuity of data, as well as the validity of data. We confirm the rationality of data by checking whether it is received at the right time and from the right station. If the system receives data out of plan, then abandons it.

We confirm the continuity of data by checking whether the data is continuous while the station is tracking. If the data breaks off, it is abandoned.

We confirm the validity of data by checking whether the telemetry data framework conforms to the regulations. It is abandoned when the frame synchronous code or the frame frequency is uncorrected.

25.4.2 Data Selecting

Mostly, after the pretreatment of telemetry data, there are still several sources of data, which all have right framework. We pick one from them by data selecting, which is done after the code extracting. The procedure of data selecting is as follows:

Assume a telemetry parameter is transmitted by K sources, make P_i the code of source i. Make P_i a set (P_1, P_2, P_3, … P_i), record the number of times P_i appears as N_i, and we get:

$$N_1 + N_2 + N_3 + \cdots + N_i = K$$

Note $N_x = \text{Max}\ (N_1, N_2, N_3, \cdots, N_i)$.

If there is only one $N_i = N_x$, then P_x is the data selecting result.

If the number of $N_i = N_x$ is bigger than one or N_x=1, we can not get the sole selecting result. In this circumstance, we make use of the data fusion algorithm to P_i.

25.4.3 Data Fusion

Make b_{ik} the ith bit of code P from source k, and it can be 0 or 1.

If

$$0 \le (b_{i1} + b_{i2} + \cdots b_{ik})/k \le 0.5$$

b_i of the fusion result equals 0.

If

$$0 \le (b_{i1} + b_{i2} + \cdots b_{ik})/k = 0.5$$

abandon this code because no result can be made.

If

$$0.5 < (b_{i1} + b_{i2} + \cdots b_{ik})/k \le 1$$

b_i of the fusion result equals 1.

In this way, we get every bit of code P and make it the result of data fusion.

25.4.4 Data Fusion Based on Optimal Weighted

Data fusion method talked in Sect. 4.3 is based on that the BER of stations are in average. In real mission, we consider it different when talking about the reliability of different channels. For the TT&C network existing, there are ground stations, instrumentation ship and tracking and data relay satellites. Because of the different factors affecting transference, data quality received and delivered by different channels should be considered separately.

For this consideration, we improve the data fusion method by using of optimal weighted. We supply every kind of channel with different weighted value, and the new algorithm is adjusted as follows:

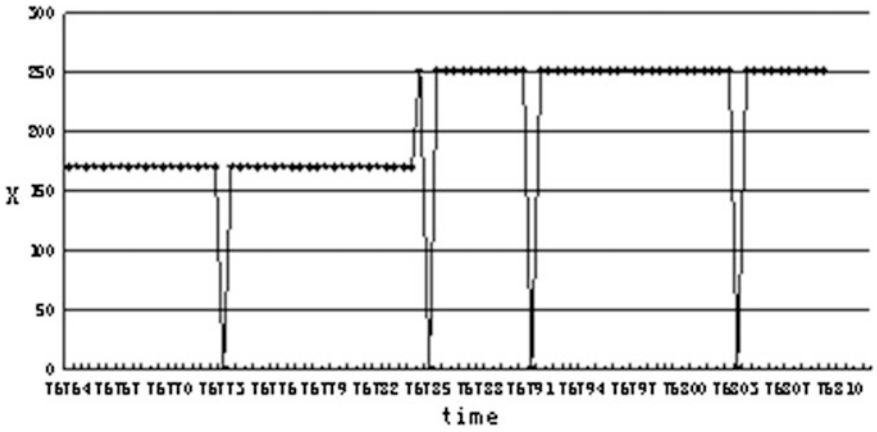

Fig. 25.3 Parameter X from ground station 1

If

$$0 \le \frac{\sum_{n=1}^{k} b_{jn} \times W_n}{\sum_{n=1}^{k} W_n} \le 0.5$$

b_j of the fusion result equals 0.

If

$$0 \le \frac{\sum_{n=1}^{k} b_{jn} \times W_n}{\sum_{n=1}^{k} W_n} = 0.5$$

abandon this code.

If

$$0.5 < \frac{\sum_{n=1}^{k} b_{jn} \times W_n}{\sum_{n=1}^{k} W_n} \le 1$$

b_j of the fusion result equals 1.

The weighted value can be set according to the real tracking condition. For example, we can set them in such principles as:

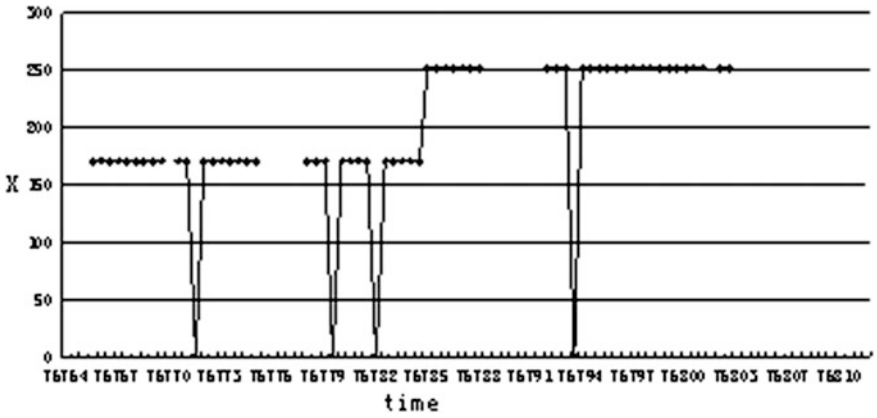

Fig. 25. 4 Parameter X from ground station 2

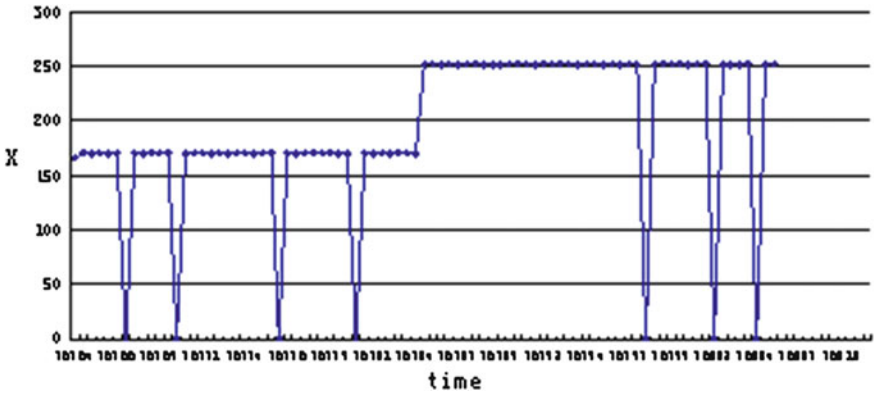

Fig. 25.5 Parameter X from ship

1. In any case, the minority is subordinate to the majority.
2. Ground stations take priority over instrumentation ships, and instrumentation ships take priority over tracking and data relay satellites.

25.5 Application

Choose a period of telemetry data from real spaceflight mission. There are two ground stations, one instrumentation ship and one tracking and data relay satellite, of which the weighted value are indicated as x, y, z.

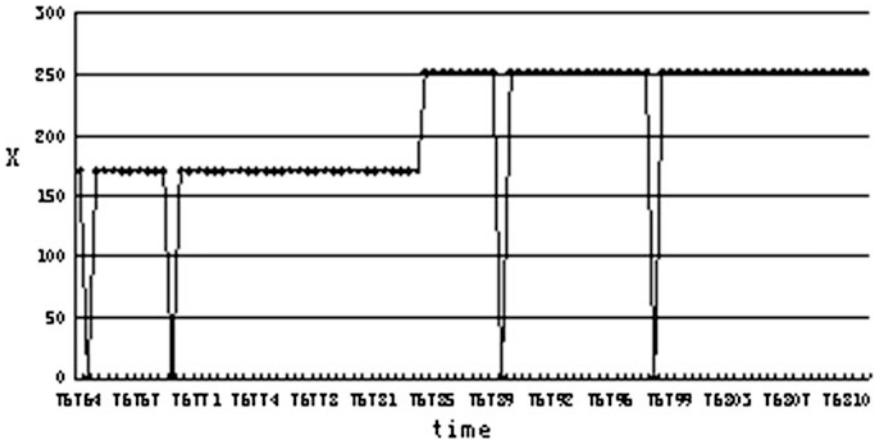

Fig. 25.6 Parameter X from delay satellite

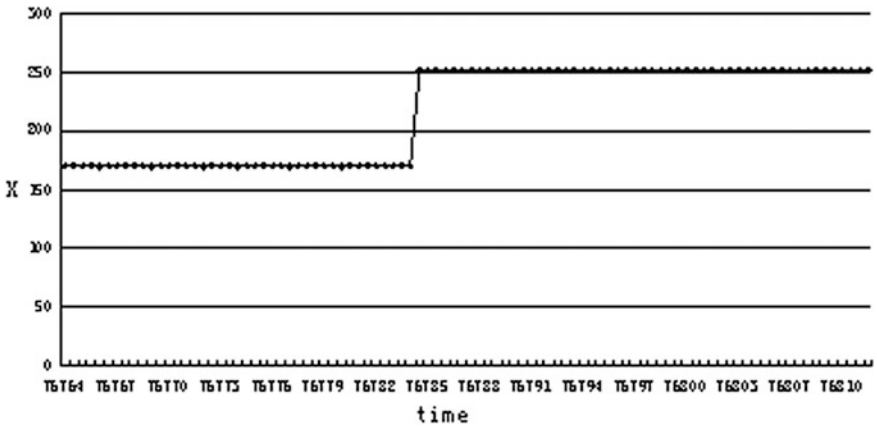

Fig. 25.7 Parameter X after data fusion

According to the principles stated above, we can get

$$\begin{cases} z + \mathrm{y} > x \\ \mathrm{x} > y > z \end{cases}$$

Make

$$\mathrm{x} = y + 1 = z + 2$$

we can distribute the weighted value as

$$\begin{cases} x = 4 \\ y = 3 \\ z = 2 \end{cases}$$

The values of parameter X from the four sources are shown in Figs. 25.3, 25.4, 25.5, 25.6. It can be seen from the above figures that the start–end time of four sources are different, and they all have error bits. We process the data with data fusion method based on optimal weighted, and get the parameter value curve in Fig. 25.7.

As indicated in Fig. 25.7, after data fusion, the error bits are cast off, and a whole and correct result is given.

In addition, we choose 1 h data with more stations to verify the reliability and efficiency of the processing. The result shows that the BER is zero and it can meet the real-time data frequency requirement.

25.6 Conclusions

In conclusion, the data fusion method based on optimal weighted presented in this article can solve the problem of code error from multi-sources and give an integrated and more reliable result. Theoretical analysis and experiment data show that the data quality after fusion is improved obviously. It reduces risks caused by error data, and lowers the difficulty of follow-up work. Time used for data processing is shortened, and the efficiency is promoted. The method has great value in engineering application, and can make reference for the further research of multi-sources data processing.

References

1. Waltz E, Llinas J (1990) Multisensor data fusion. Artech house, Boston
2. David LH (1992) Mathematical techniques in multisensor data fusion. Artech house, New York
3. Liu YC, Fang HR, Zhang F (2001) Telemetry data processing. National defense industry press, Beijing

Chapter 26
A Data Fusion Method of Multi-Sources Measurement Data Based on Federated UKF Filter

Hong Chen, Jian Jiang and Lin Wang

Abstract To ensure the tracking reliability and accuracy, and to increase the utilizing efficiency of measurement data in space flight tracking, a data fusion method in real-time orbit tracking based on federated UKF filter is put forward. The key technology of the multi-source measurement data fusion scheme is also analyzed in this paper. The computer simulation shows that the reliability of the real-time orbit tracking can be increased remarkably and the tracking accuracy can be increased by thirty percent. In this method, the fused ballistic curve is continuous and the convergence can meet the requirement of real-time guidance, which has great meanings on the overall level of spaceflight commending and controlling system.

Keywords Multi-source data fusion · Federated UKF filter · Orbit tracking

26.1 Introduction

To ensure the reliable tracking of the target in space flight tasks, different kinds of tracking equipments such as optic equipments, radars and GPS are comprehensively used at the same time. In order to take advantage of the measurement data of different kinds of tracking equipments, the commonly method is to reversely figure out the orbit of the target according to the measurement data and choose one as the final orbit to output according to the quality of the measurement data [1]. This method has the shortages of low utilization rate of measurement data and bad

H. Chen (✉) · J. Jiang · L. Wang
Jiuquan Satellite Launch Center, Lanzhou 732750, China
e-mail: chenhong7409@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_26,

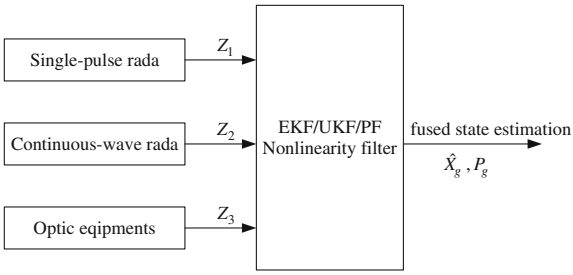

Fig. 26.1 The central data fusion method

tracking reliability. The multi-source data fusion method based on federated UKF filter presented in this paper can take full advantage of the potentiality of the tracking systems and can also increase the orbit tracking accuracy and reliability.

26.2 Previously Data Processing Before Data Fusion

In center computer of the Spaceflight Commanding & Controlling Center, some sampling data called outliers whose values are much higher than the natural values always exist in the received data of every measurement equipment. Further more, some data are missed because of interference. So, the outliers must be found and the missed data must be renewed before data fusion. On the other hand, because every measurement equipment in the tracking system doesn't work at the same time and the data sampling periods of measurement equipments are not always uniform, arrival time difference of received measurement data exists in the center computer and the measurement data must be aligned. Interpolation algorithm such as the Lagrangian three-point interpolation algorithm is the classical method to solve the asynchronous time problem caused by the different data sampling period [2, 3]. Commonly, the sampling time of one measurement equipment in the tracking system can be chosen to be the reference time and the sampling time of other measurement equipments should be unified to the reference time.

26.3 Traditional Data Fusion Methods in Target Tracking

In space flight tracking, the state equation used to describe the flight character of the target and the measurement equations of the measurement equipments such as radars and optic equipments are strongly nonlinear. The methods of nonlinear filter are always adopted to estimate the flight state of the target. The structures of data fusion based on nonlinear filter can be classified into the centrality structure and the distributed structure. The centrality structure of data fusion in target tracking is shown in Fig. 26.1.

In the centrality structure, all measurement data of different kinds of tracking equipments constitute the measurement vector of the EKF/UKF/PF nonlinear filter [4–6]. Theoretically speaking, the centrality structure has the optimization state estimation performance and is easy to implement. But there are many serious problems in the central data fusion system. Firstly, the computation complexity will be N^3 times (N is the dimension of the measurement vector). For example, assuming the tracking system are composed of one single-pulse radar, one continuous-wave radar and three optic equipments, the measurement vector dimension of the central filter will be 17. In that case the real-time performance of the orbit tracking can not be satisfied because of the remarkably increased computation complexity. Secondly, the central nonlinear filter has low reliability. The error will be expanded to the whole filter when the local measurement data is wrong and the orbit tracking will be completely failed. So, the central data fusion system is seldom adopted in practical application.

The simply method to fuse the measurement data in a distributed data fusion system is to set the weighted coefficient corresponding to the target state estimation of different measurement equipments. Commonly, we can determine the weighted coefficients according to the measurement precision because different measurement equipments have different precision. For example, assuming that the state estimation result of the ith measurement equipment is $[x_i, y_i, z_i]^T$ and the variance is $\left[\sigma_{x_i}^2, \sigma_{y_i}^2, \sigma_{z_i}^2\right]^T$ (where $i = 1, 2, \ldots M$), the fused estimation value of the orbit can be given by

$$\begin{aligned} \hat{x}_0 &= \left(1 \Big/ \sum_{i=1}^{M} \frac{1}{\sigma_{x_i}^2}\right) \sum_{i=1}^{M} \frac{x_i}{\sigma_{x_i}^2}, \; \hat{y}_0 = \left(1 \Big/ \sum_{i=1}^{M} \frac{1}{\sigma_{y_i}^2}\right) \sum_{i=1}^{M} \frac{y_i}{\sigma_{y_i}^2}, \\ \hat{z}_0 &= \left(1 \Big/ \sum_{i=1}^{M} \frac{1}{\sigma_{z_i}^2}\right) \sum_{i=1}^{M} \frac{z_i}{\sigma_{z_i}^2} \end{aligned} \tag{26.1}$$

The distributed data fusion method based on weighted coefficient of precision has ideal effect when all of the measurement equipments work properly. But the measurement precision of a equipment is only a probability statistic, the measurement data can be completely wrong when the equipment of the tracking system doesn't work properly or the target is missing. In that case, the data fusion system has poor tracking performance.

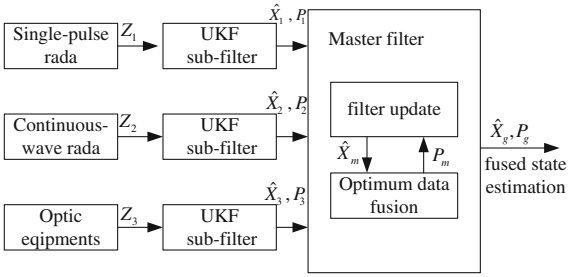

Fig.26. 2 The distributed data fusion structure of multi-source measurement data

26.4 Distributed Data Fusion Method Based on Federated UKF Filter

26.4.1 System Model

To avoid the shortages in the central data fusion and the traditional distribute data fusion based on weighted coefficient, a distributed data fusion model based on federated UKF [7] is adopted to improve the performance of the space flight tracking system. The system model can be shown in Fig. 26.2. The federated filter is composed of one master filter and a certain number of sub-filters. In Fig. 26.2, the state estimation of one sub-filter is independent of the other sub-filters. $\hat{X}_g$ and P_g represent the fused state and it's variance. $\hat{X}_i$ and P_i represent the state estimation and it's variance of the ith sub-filter. $\hat{X}_m$ and P_m represent the state estimation and it's variance of the master filter.

In Fig. 26.2, different kind of measurement equipment estimates the state of target with its own UKF sub-filter. The orbit of the target is fused in the master filter according to the state estimation results of every UKF sub-filter. The model of Fig. 26.2 has much lower computation complexity and better system fault tolerance. The fault of local measurement equipment can hardly influence the final precision of the fused orbit. In space flight tracking, the cases that the measurement equipments malfunction or track the wrong target are common. So the distributed data fusion model showed in Fig. 26.2 is more suitable to be used in space flight tracking.

26.4.2 UKF Sub-filter

UKF filter chooses a set of sigma sample points to compute the mean and covariance of the state distribution. Assuming that the state estimation vector of the UKF filter is X, a discrete nonlinear system can be described as follows:

$$X(k+1) = \Phi(k+1,k)X(k) + W(k) = F[X(k), W(k)] \tag{26.2}$$

$$Z_i(k) = H_i(k)X(k) + v_i(k) \qquad i = 1, 2, \ldots, n \tag{26.3}$$

where $\phi(k+1,k)$, $W(k)$ and $H_i(k)$ represent the state transformation matrix, noise matrix and the measurement matrix of the ith sensor respectively, $E[w(k)w^T(j)] = Q(k)\delta(k,j)$,$E[v(k)v^T(j)] = R(k)\delta(k,j)$

UKF consists of the following steps:

Step 1: Set the initial values for state mean $\bar{X}_0$ and error covariance P_0.

Step 2: Compute the $2n_x + 1$ sigma sampling points χ_i and their weighted coefficient W_i.

$$\begin{gathered}
\chi_0 = \overline{x}, \quad W_0^m = \kappa/(n_x+\kappa), W_0^c = \kappa/(n_x+\kappa) + (1-\alpha^2+\beta), i = 0 \\
\chi_i = \overline{x} + (\sqrt{(n_x+\kappa)P_x})_i, \quad W_i^m = W_i^c = 1/2(n_x+\kappa), i = 1, \ldots, n_x \\
\chi_i = \overline{x} - (\sqrt{(n_x+\kappa)P_x})_i, \quad W_i^m = W_i^c = 1/2(n_x+\kappa), i = n_x+1, \ldots, 2n_x
\end{gathered} \tag{26.4}$$

Step 3: Compute the predicted state mean and the predicted covariance as:

$$\chi_i^x(k+1/k) = F(\chi_i), i = 0, \ldots, 2n_x \tag{26.5}$$

$$x(k+1/k) = \sum_{i=0}^{2n_x} W_i^m \chi_i^x(k+1/k) = \sum_{i=0}^{2n_x} W_i^m F\left[\chi_i^x(k/k)\right] \tag{26.6}$$

$$P(k+1/k) = \sum_{i=0}^{2n_x} W_i^c \left[\chi_i^x(k+1/k) - \overline{x}(k+1/k)\right]\left[\chi_i^x(k+1/k) - \overline{x}(k+1/k)\right]^T + Q(k) \tag{26.7}$$

$$Z_i(k+1/k) = H(\chi_i), i = 0, \ldots, 2n_x \tag{26.8}$$

$$\bar{Z} = \sum_{i=0}^{2n_x} W_i^m Z_i(k+1/k) \tag{26.9}$$

Step 4: Update the state mean and covariance as

$$P_{ZZ} = \sum_{i=0}^{2n_x} W_i^c \left[Z_i(k+1/k) - \overline{Z}\right]\left[Z_i(k+1/k) - \overline{Z}\right]^T + R(k) \tag{26.10}$$

$$P_{XZ} = \sum_{i=0}^{2n_x} W_i^c \left[\chi_i^x(k+1/k) - \overline{x}(k+1/k)\right]\left[Z_i(k+1/k) - \overline{Z}(k+1/k)\right]^T \tag{26.11}$$

$$K_k = P_{XZ}P_{ZZ}^{-1} \tag{26.12}$$

$$x(k+1/k+1) = x(k+1/k) + K_k\left[Z_k - \overline{Z}(k+1/k)\right] \tag{26.13}$$

$$P(k+1/k+1) = P(k+1/k) - K_k P_{ZZ}(k+1/k) K_k^T \tag{26.14}$$

26.4.3 Data Fusion in the Master Filter

In the federated UKF filter, the process information $Q^{-1}(k)$ and $P^{-1}(k)$ are distributed between the sub-filters and the master filter as follow

$$Q_i(k) = \beta_i^{-1} Q(k) \tag{26.15}$$

$$P_i(k) = \beta_g^{-1} P(k) \tag{26.16}$$

where β_i is the information distribution coefficient and satisfies [8]

$$\sum_{i=1}^{n} \beta_i + \beta_m = 1 \tag{26.17}$$

Finally, all of the state estimations in every UKF sub-filter are fused in the master filter. The fused state estimation and the covariance matrix can be expressed as

$$P_{\mathrm{g}}(k) = \left[\sum_{i=1}^{n,m} P_i^{-1}(k) \right]^{-1} \tag{26.18}$$

$$\hat{X}_g = P_g \sum_{i=1}^{m} P_i^{-1} \hat{X}_i \tag{26.19}$$

26.5 Analysis of the Key Techniques

26.5.1 Modeling of the Flight Target

In the target tracking system based on nonlinear filter, the modeling veracity of the state equation used to describe the flight character of the target has great influence on the tracking performance. The models of target movement consist of single model(SM), static multiple model (MM) and interacting multiple model (IMM). There are many kinds of single model such as CA model, CV model, Singer model [9] and CS model [10]. Different movement model is used to estimate the target state in different period in MM and satisfies the flight character better than SM. But only one model filter is used at the same time in MM and the static model switch will lead to tracking delay. IMM has ideal tracking performance through adopting

multiple models at the same time and has been a research hotspot in target tracking. But the computation complexity of the IMM model will be remarkably increased when there are many models in IMM. In addition, the unnecessary competition among models will reduce the tracking performance along with the increase of models. So, model set in IMM should be optimized during the process of data fusion.

26.5.2 Covariance Matrix Processing

In space flight target tracking, every measurement equipment has the specifically tracking area because of the earth curvature. For example, a radar can not track the target from the launching of rocket to 20 s. The output data of the radar will be a changeless value and the state estimation covariance matrix of the sub-UKF filter will be very small. From 20 to 400 s, the radar can track the target normally and the state estimation covariance matrix of the sub-UKF filter becomes larger and larger. After 400 s, the target is missing and the covariance matrix falls sharply. So, the covariance matrix of the sub-UKF filter keeps small though the output measurement data is invalid. As we have known that the weighted coefficient of data fusion is inversely proportional to the covariance matrix. So the covariance matrix of the sub-UKF filter should be processed properly to avoid decreasing the tracking performance of the whole data fusion system.

26.5.3 Reliability Techniques of the Data Fusion System

Reliability is the most important factor in space flight tracking and the measurement data fusion system must work properly in any case of abnormal condition. On the one hand, the movement model adopted in the data fusion system must satisfy the character of the maneuvering target. Otherwise, the data fusion filter will be divergent. On the other hand, square root operation is needed in UKF filter during the process of UKF iteration and the covariance matrix of the UKF filter must be positive. When the covariance matrix becomes non-positive, software error will occur. Valid steps should be taken to ensure the performance of system fault-tolerance.

26.6 Computer Simulations

In this section, the performance of the data fusion based on federated UKF filter in space flight tracking is investigated and the current statistical (CS) model is adopted in this paper. There are one single-pulse radar, one set of continuous-wave

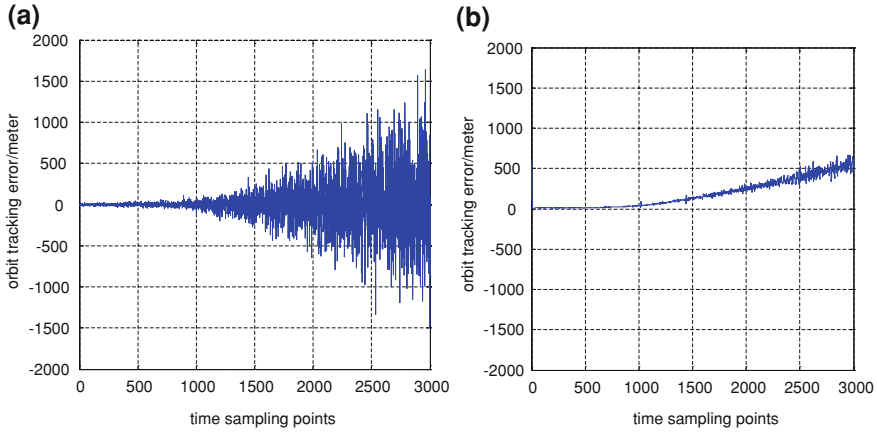

Fig. 26.3 Orbit tracking performances of different methods

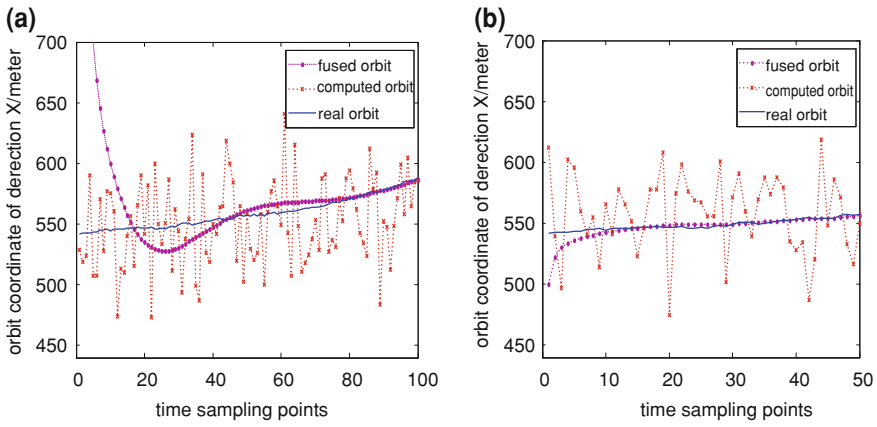

Fig. 26.4 The convergence performance of federated UKF filter

radar (composed of one master station and four assistant stations) and three optic equipments in the space flight tracking system, and the sampling period of the measurement data is 50 ms. The orbit tracking performance based on federated UKF data fusion is shown in Fig. 26.3. Computer simulation shows that the precision of orbit tracking based on federated UKF will be increased more than thirty percents compared with the orbit tracking result which directly computed according to the measurement data of the single-pulse radar.

The convergence rate of the data fusion based on federated UKF filter is very important in space flight real-time orbit tracking. The convergence performance of

the federated UKF filter is shown in Fig. 26.4. The convergence of the federated UKF filter will be faster when the value of the state set initially approaching the real value. In that case, the federated UKF filter can reach convergence after several times of filtering irritations.

26.7 Conclusions

A data fusion method in real-time orbit tracking based on federated UKF filter to make maximal use of the space flight tracking system is presented in this paper. Computer simulation shows that the precision of the orbit tracking will be increased remarkably and the convergence rate can meet the requirements of the real-time tracking in space flight. Further more, the reliability of the tracking system will be effectively improved because of the ideal fault-tolerance performance.

References

1. Liu LS (2000) Post processing of measurement data. National defense industry press, Beijing
2. Blair WD, Rice TR (1991) Asynchronous data fusion for target tracking with a multitasking radar and option sensor. SPIE 1482:234–245
3. Peng Y, Xu Y, Jin HB (2005) Analysis of time asynchronization algorithms in multisensor data fusion systems. Radar countermeas 2:16–19
4. Fang JC, Shen GX, Wang DJ (1999) Adaptive extended Kalman filter model for integrated land vehicle navigation system. J Beijing Univ Aeronaut astronautieds 25(2):235–239
5. Juliter S, Uhlmann JA (2000) New method for the nonlinear transformation of means and covariance in filters and estimators. IEEE trans autom control 45(3):477–482
6. Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proc radar signal proc 140(2):107–113
7. Carlson NA (1988) Federated Kalman for fault-tolerant integrated navigation. Processing of IEEE position location and navigation symposium, pp 110–119
8. Gu QT, Wang S (2003) Optimized algorithm for information sharing coefficients of federated filter. J Chin Inert Technol 11(3):1–6
9. Singer RA (1970) Estimating optimal tracking filter performance for manned maneuvering targets. IEEE Trans AES 6(4):473–483
10. Zou HR, Jin ZL (2000) Maneuvering target tracking. National defense industry press, Beijing

Chapter 27
Dim and Weak Target Detection Technology Based on Multi-Characteristic Fusion

Jia Tang, Xin Gao and Gang Jin

Abstract As the signal of small weak target is weak, its characteristic information is prone to weakening and loss and is difficult to maintain stability. In this paper, a self-learning mechanism is built using continuous signal correlative characteristics. The method is based on K-L transform fusion technology and D-S theory of decision-making. Meanwhile, the fusion technology of "fusion after the first decision-making" and "decision-making after first fusion" organic integration is developed. To verify the detection capability of this technology for small dim targets, a new type of tracking processor is developed. It processes more than one hundred groups of raw image data divided into two groups. The image data is taken by a space target surveillance telescope. Experimental results show that dim targets can be detected and tracked stably and reliably using this technology. Also provided with a complex background suppression and anti-jamming capability, a target can be detected stably and reliably when its SNR is about 1.2.

Keywords Multi-characteristic · Fusion · Dim and weak target · Detection · K-L transform · D-S theory of decision-making

J. Tang (✉) · X. Gao
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: ttj.j@163.com

G. Jin
China Aerodynamics Research and Development Center, Mianyang 621000, China

G. Jin
Chinese Academy of Sciences, Institute of Optics and Electronics, Chengdu 610209, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_27,

27.1 Introduction

Range optical equipment is required to detect and track specific objects as far as possible. At far distance between the optical equipment and the object, the dimension of target image taken by the photoelectric system is very small and the intensity is weak. As the weak target signal lacks shape and texture characteristics information, most conventional means fail to detect the target. Therefore, detection and tracking of dim and weak targets has long been a challenge and a hot subject of research in the field of signal processing.

Data fusion technology is an integrative process of a variety of sources of information processing, control and decision-making. The method of multi-feature fusion processes a variety of signal characteristics in the feature domain to get a new "feature" followed by target detection and identification based on this new feature. There are many fusion approaches, e.g. weighted average method, voting method, DS evidence theory, linear classifiers and artificial neural network simulation [1–5]. Generally speaking, these methods fuse the results of various signal features and then obtain the final result through decision-making.

The signals of small weak target are weak, so the characteristic information is prone to weakening and loss and it is difficult to maintain stability. Therefore, it calls for an effective self-learning mechanism and multi-feature fusion method for stable extraction of small weak target signals and re-capture of the target as early as possible after target signal lost.

In this paper, a self-learning mechanism is developed using continuous signal correlative characteristics, and the fusion technology of "fusion after the first decision-making" and "decision-making after first fusion" organic integration is developed.

27.2 Technical Principle

27.2.1 Discrete Karhunen–Loeve Transform Fusion Technology

Discrete Karhunen–Loeve Expansion transform (K-L transform) is a mathematical method for pattern recognition. It is an orthogonal linear transformation similar to the Fourier transform and Walsh transform. K-L transform performs compression of the extracted feature information to determine the largest value of the transformed "features" vector. To a certain sense, it can also be considered an adaptive information fusion technology.

Discrete K-L transform expands orthogonally according to statistical nature as (27.1) below:

$$g = A\left(f - m_f\right) \tag{27.1}$$

where, f is data in random field scale as N, whose statistical properties characterized with the mean vector m_f and covariance matrix C_f. A is orthogonal transformation matrix. Its line is composed by the eigenvectors of C_f.

It can be seen from the nature of the K-L transform covariance matrix C_g in the transform domain g is a diagonal matrix. Its main diagonal elements are the eigenvalues of covariance matrix C_f the elements outside the main diagonal are all zero. The main diagonal elements satisfy $\lambda_i \geq \lambda_{i+1} (i = 1, 2, \ldots, N-1)$. At the same time, λ_i is also the eigenvalue of C_f. We can see that C_f and C_g have the same eigenvalues and eigenvectors. This means the main diagonal element is the eigenvalue of the original image and of the transform domain covariance matrix at the same time. Covariance matrix transformed can be substituted by the matrix before transformation.

Because each element in the diagonal of the covariance matrix reflects the variance of each component in the transform domain, it provides a measure of correlation between the elements. The characteristic value can be expressed by the "track" of eigenvalue matrix, as shown in (27.2) below:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_N = C_{11} + C_{22} + \cdots + C_{NN} \sin^{-1} \theta \tag{27.2}$$

Equation (27.2) reflects a new distribution of the original data feature information in the transform domain. The eigenvalue component $(\lambda_1, \lambda_2, \ldots, \lambda_n)$and the first characteristic value λ_1 can be taken to form a new characteristic parameter. The new characteristic parameters describe the largest energy distribution and shape distribution of the data.

Due to the different choice of feature extraction methods and the dimension values, there is a big difference between the characteristics of the combinations of the same model samples. First, the original feature should be standardized to eliminate the adverse effects caused by the unbalanced nature of the combined feature values, as shown in (27.3) below:

$$Y = (X - \mu)/\sigma \tag{27.3}$$

where, μ is the mean vector of training samples and σ is the mean value of the standard deviation vector in each component.

When characteristic dimensions involved in the combinations are unequal, the characteristics of a higher dimension in the combinations still have a clear advantage after standardization because the characteristics of a higher dimension value occupies a larger proportion in the scatter matrixes generated in the combination. To overcome this effect, take a weighted combination of standardized characteristics is taken when the dimensions of feature vector are unequal.

K-L transform is optimal orthogonal transformation of a mean square error criteria. The transform matrix not only contains the information of the original image, but also naturally reflects the characteristics of the transform domain. The K-L transform is different from the other orthogonal transform. K-L transform realizes characteristic information compression while performing a fusion of characteristic information at the same time.

27.2.2 Dempster-Shafer Theory of Decision-Making

Dempster-Shafer theory of evidence (D-S theory of evidence) as a decision-making theory makes decisions through combining multiple evidence. Compared with probability decision theory, it is not only able to handle uncertainty because the knowledge is not accurate, but also able to handle the uncertainties because the knowledge does not know.

During target detection, the evidence can be assumed on the domain {target, non-target, uncertainty}. $U = \{u_1, u_2, \ldots, u_n\}$ is a collection of factors affecting the judgment. Weight coefficient for the judgment of the corresponding factors in the decision process is $\tilde{A} = \{a_1, a_2, \ldots, a_n\}$. a_k is the weighting coefficient corresponding to the k-th factor and $\sum_{k=1}^{n} a_k = 1$ and it can be assumed that $a_1 \geq a_2 \geq \cdots \geq a_n$.

The evidence involved in the fusion comes from the degree of membership of fuzzy comprehensive decision-making. The collection composed by the membership degree corresponding to the factors is shown below:

$$R = \{r_1, r_2, \cdots, r_n\} \tag{27.4}$$

During the process of tectonic evidence, $r_{k1} = r_k$ is used to represent the credibility of detecting as the target when the k-th factor is considered. $r_{k2} = 1 - r_k$ is used to represent the unreliability of detection of a target. τ_k is the degree of adjustment. δ_k is the degree of expansion corresponding to u_k. Their values are determined by simulation.

Evidence matrix $M = (m_{ki})_{n\times 3}$ is defined. $m_{k1} = a_{k1} r_{k1}$ is used to represent the level of trust of the target when the k-th factor is considered and $m_{k3} = 1 - a_k$ is used to represent the level of uncertainty of the target. Each $M_k = (m_{k1}, m_{k2}, m_{k3})$ is an evidence. All the items of evidence are fused by D-S evidence theory fusion rules and the final comprehensive judgment vector is got as shown in Eq. (27.5):

$$M = \{m_1, m_2, m_3\} \tag{27.5}$$

m_1, m_2 and m_3 respectively represent the level of trust of the target, the level of trust of non-target and the level of trust of target uncertainty.

27.3 Experiment

In order to verify the detection capability of this technology for small dim targets, a large scale semi-physical simulation experiment is carried out. First, a new type of tracking processor using the new tracking and dealing algorithm is developed. Then, the space target image data taken by other optical equipments is sent into the new track processor. The module's minimum detectable magnitude and SNR are detected. The image source for the experiment is the space target image data taken

by space target surveillance equipment, whose advantage is to facilitate quantitative assessment.

27.3.1 SNR Calculation Method

The lowest Signal-to-Noise Ratio (SNR) for the target is an important indicator to judge weak target signal extraction capability. There are many ways to calculate the SNR. Equation (27.6) is used here.

$$SNR = \frac{\sum_{i}^{N_s} abs(S_i - \bar{B})}{N_s} \Bigg/ \sqrt{\frac{\sum_{i}^{N_B} (S_B - \bar{B})^2}{N_B}} \tag{27.6}$$

where, S_i is signal value corresponding to each signal pixel. S_B is signal value corresponding to each background pixel. $\bar{B}$ is the mean of the background signal, defined as $\bar{B} = \sum_{i}^{N_B} S_B \Big/ N_B$. N_B is total number of pixel processing area. Taking the target as the center and taking the 101 × 101 image area minus the 5 × 5 of the center of the wave gate, there are a total of 10176 pixels. N_S is pixel number occupied by a given target. For point targets, the area can be fixed to take 3 × 3 pixels in the wave door center for convenience.

The SNR is approximately linearly proportional to the target brightness Q. According to the relationship between the target brightness and magnitude, Eq. (27.7) can be given:

$$\frac{SNR_{n-k}}{SNR_n} \propto \sqrt{\frac{Q_{n-k}}{Q_n}} = 2.512^{k/2} \tag{27.7}$$

Where, $n - k$ is magnitude for a reference star. n is magnitude for the observed star. k is the magnitude difference between the reference star and the observed star.

27.3.2 Experiment Results

To test the proposed technology, the new type of tracking processor has processed more than 100 groups of raw image data divided into two groups. The image data is taken by a space target surveillance telescope. Each part of the test system is working properly. Target detection and tracking state is reliable. Processing speed is about 110 frames per second.

In the first batch of data, the target SNR of most images is even <1. The new processor is stable to detect the target. Lost target because of lens shaking can be

Table 27.1 Part of the typical image data file processing results

File name	Minimum SNR	Highest SNR	Average SNR	Detection rate (%)	Notes
20090519_105388_5	0.38	32.02	8.26	100	Double Star intersection
20090506_105412_2	0.49	19.21	7.29	100	
20090506_106006_8	0.73	23.62	7.24	93.91	Target into the cloud
20090615_834780_1	0.69	15.49	6.13	100	
20090611_107302_1	0.14	51.62	6.12	100	
20090519_105412_1	0.58	14.86	6.06	100	
20090506_105412_1	0.86	13.57	5.81	100	Clouds
20090506_513681_5	0.83	10.11	3.08	100	
20090506_107306_2	0.31	4.6	2.15	100	

Table 27.2 Selected stellar processing results

No.	Object location on the image		Mean SNR for each exposure time				Magnitude
	X	Y	154 ms	62 ms	26 ms	10 ms	
1	427	352	5.79	2.8	2.11	1.22	11.84
2	514	934	4.90	3.02	2.15	1.49	11.9
3	81	484	4.14	2.67	1.88	1.47	11.9
4	98	742	3.05	1.97	1.44	1.25	11.9
5	130	986	5.05	2.82	1.87	1.21	12
6	213	52	2.54	1.77	1.38	X	11.9
7	248	796	2.13	1.51	1.31	X	12.2
8	23	682	2.53	1.77	1.27	X	12.4

quickly re-captured automatically. Table 27.1 lists a part of SNR results and detection rate for some typical image data, including the higher SNR, lower SNR, and the clouds and multi-satellite intersection. Processing results of this batch show that the new technology can stably and reliably detect and track low SNR space targets, and has good resistance to the intersection of interference and cloud interference.

The second batch of image data for different exposure time is obtained from the same airspace in order to facilitate assessment of weak signal detection capability of the new image processing algorithm. Image exposure time includes 154, 62, 26, and 10 ms. The difference between adjacent exposure time is about 2.5 times, equivalent to the brightness change around magnitude 1 for the same stars. Table 27.2 lists the part of processing results of the star. The objectives set forth in the table are failures of existing equipment capability to detect. Using the proposed technology, the object can still be detected accurately even if in case of 10 ms integral conditions. The table gives the location coordinates of the target in the image and the average SNR of the detected target under different exposure time (X-target signals are too weak to be detected with the proposed technology).

Data processing results from the batch of the actual detection images show that the new technology can stably and reliably detect and track the low SNR space objects when the target SNR is only 1.2. Detectable magnitude is no less than magnitude 12, higher than the design value of 11.

27.4 Conclusions

Experiment results show that the technology proposed in this paper is capable of stable and reliable detection and tracking of dim and weak targets. Low SNR space objects, the target SNR of which is only 1.2, can be successfully detected and tracked. Moreover, it is provided with a complex background suppression and anti-jamming performance. The processing speed is faster than 100 frames per second.

References

1. Reed IS, Gagliardi RM, Stotts LB (1998) Optical moving target detection with 32D matched filtering. IEEE Trans Aerosp Electron Syst 24(4):327–336
2. Porat B, Friedlander B (1990) A frequency domain algorithm for multi-frame detection and estimation of dim targets. IEEE Trans Pattern Anal Mach Intell 12(4):398–401
3. Succcary R, Cohen A, Yaractzi P et al (2001) A dynamic programming algorithm for point target detection: practical parameters for DPA. Proc SPIE 4437:96–100
4. Johnston LA, Vikram K, Leigh A (2002) Performance analysis of a dynamic programming track before detect algorithm. IEEE Trans Aerosp Electr Syst 38(1):228–242
5. Kirubarajan T, Bar-Shalom Y (2004) Probabilistic data association techniques for target tracking in clutter. Proc IEEE 92(3):536–557

Chapter 28
Distortion Correction for Optical Measurement Systems in a Test Range

Rujie Wang, Liangliang Wang, Lei Zhang and Jia Tang

Abstract A flexible method for timely correction of distortion of optical measurement system in a test range environment is proposed to eliminate the negative impact of distortion on accuracy. The fitting relationship between the measured miss distance and real miss distance is solved because theoretically the compound angle is unchanged. Operation of the method is simple and quick and there is no need of special-purpose instruments. The method is validated in an experiment of distortion correction. The result of data analysis shows that the measurement accuracy is increased from 100 to 20 arc s. Therefore, it has a high value for extended application.

Keywords Distortion correction · Optical system · Miss distance

28.1 Introduction

At present, most optical measurement systems in test ranges (such as theodolites, high-speed TV and ballistic cameras), are angle measurement systems using optical imaging devices. The front-end of the system mainly consists of lens and photo detector. Because of the limits of technological level, errors are produced in processing and assembly of the lens; or the lens imaging are not strictly consistent with the ideal pinhole model; or the detector arrays are not completely regular. All these may lead to the generation of distortion. With the increase of the field of view, distortion will grow rapidly, resulting in performance image distortion and

R. Wang (✉) · L. Wang · L. Zhang · J. Tang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: Rujiewang@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_28,

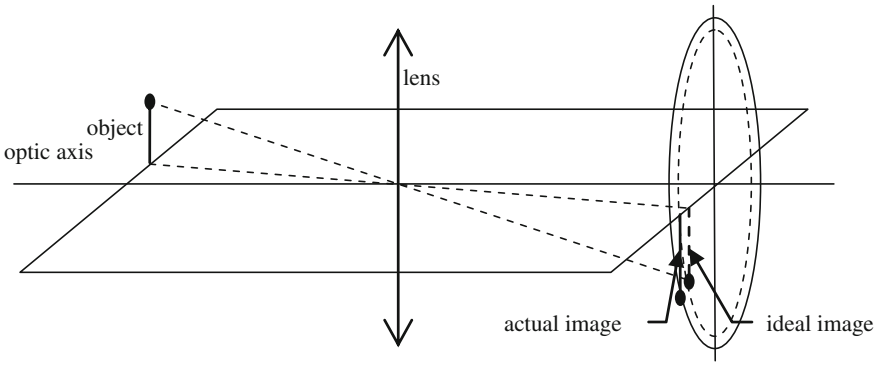

Fig. 28.1 Nature of distortion

size disproportion. Although the distortion does not affect the clarity of the image, it shall change the geometric position of the imaging and this will produce measurement errors [1]. To an angle measuring system with 15° field of view, if the distortion rate of edge is 1 %, then the angle measurement error can reach 4.45 arc min. Under normal circumstances, the angle measurement accuracy of the optical system requires <1 arc min. So such huge distortion error is absolutely not allowed and it must be calibrated to ensure measurement accuracy [2].

Distortion correction is usually done for equipment in laboratory. However, long-distance transport may produce structural deformation or position offset when the system is delivered to a test range. Part of equipment slowly releases stress. The equipment may also be subjected to long time of cold temperature and exposure. All these will cause distortion changes and the original calibration parameters will be no longer applicable. Therefore, distortion correction should be done again when precision drops dramatically.

28.2 Distortion Model

Ideally, the object and its image strictly form a triangular relationship. But in reality, angle measuring systems hardly satisfy such a condition. There is a departure between the actual image and the ideal image and this is the essence of distortion (Fig. 28.1).

In terms of the sources of error, main distortions can be divided into three types [3]: radial distortion, eccentric aberration and the thin lens distortion. Among them, radial distortion is ax-symmetrical aberration, while the latter two distortions are non-symmetrical aberrations, which include both radial and tangential distortion (Fig. 28.2).

Radial distortion: radial distortion is mainly caused by the defects of the lens (lens group). Magnification in the abaxial region is different from the near-axis

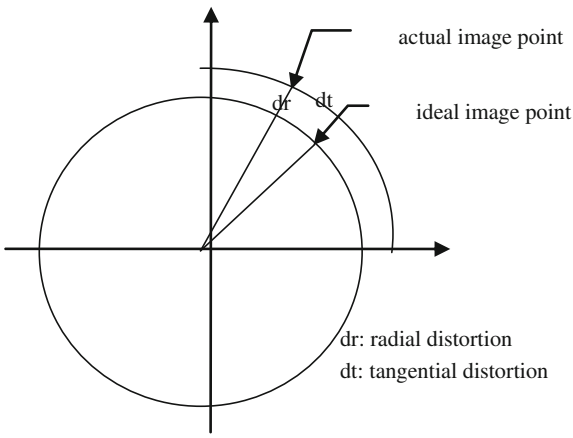

Fig. 28.2 Categories of distortion category

region. As a result, the image points inward (abaxial district magnification is bigger than the paraxial district) or outward (the magnification of the abaxial is smaller than region of the paraxial region) and both cases cause deviation from the optical axis center. According the distorted imaging shape of the square target, the two distortions are named pincushion (saddle-shaped) distortion and barrel (drum-shaped) distortion (Fig. 28.3).

The mathematical model of radial distortion is as follows:

$$\delta_x = k_1 x(x^2 + y^2) + k_2 x(x^2 + y^2)^2$$
$$\delta_y = k_1 y(x^2 + y^2) + k_2 y(x^2 + y^2)^2$$

In this formula, the k_1 and k_2 represent the radial distortion coefficient; (x, y) represent the actual imaging coordinate; δx, δy represent distortion value.

Centrifugal distortion: the optical axis of each lens in the lens group is not strictly collinear. This is the basic cause of centrifugal distortion. The mathematical model is as follows:

$$\delta_x = p_1(3x^2 + y^2) + 2p_2 xy$$
$$\delta_y = 2p_1 xy + p_2(x^2 + 3y^2)$$

p_1, p_2 represent the centrifugal distortion coefficient.

Thin lens distortion: thin lens distortion is a kind of image deformation due to lens design and manufacture imperfection or CCD array processing errors. This deformation consists of the radial component and tangential component. Its mathematical model is as follows:

$$\delta_x = s_1(x^2 + y^2)$$
$$\delta_y = s_2(x^2 + y^2)$$

s_1, s_2 represent the wave of the distortion coefficient.

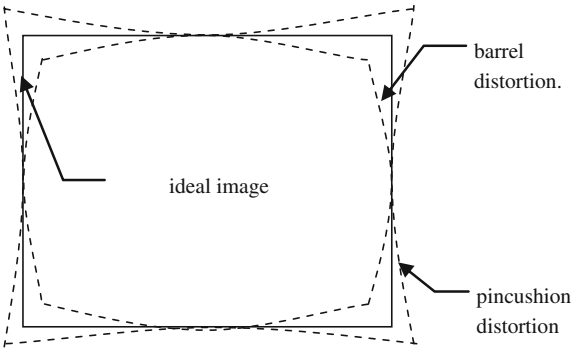

Fig. 28.3 Radial distortion

Because of distortion, for the object point, the relationship between the ideal image plane coordinates (u, v) and the actual coordinates is:

$$u = x + k_1 x(x^2 + y^2) + k_2 x(x^2 + y^2)^2 + p_1(3x^2 + y^2) + 2p_2 xy + s_1(x^2 + y^2)$$
$$v = y + k_1 y(x^2 + y^2) + k_2 y(x^2 + y^2)^2 + 2p_1 xy + p_2(x^2 + 3y^2) + s_2(x^2 + y^2)$$

28.3 Simple Method of Distortion Correction in a Test Range

Indoor optical measurement system distortion correction is usually performed with a collimator and a special grid plate target [4, 5]. Interpretation is carried out after stationary shooting and finally data processing is performed to solve the distortion parameters. However, distortion correction is a big challenge in a test range because there are no collimator and grid target. To perform distortion correction quickly and accurately in a test range, this paper proposes a flexible calibration method. The operation is relatively simple and practical application has achieved satisfactory results. The steps are as follows:

1. Select an infinity of fixed-point target, make its image in the center of the field of view, shoot multiple images for interpretation to calculate an average. Get the true position of the target (A_0, E_0);
2. Make the point target imaging at different locations in the field of view, shoot image for interpretation, getting a group of encoder values (A and E) and measuring miss distance (x, y);
3. According to the principle of synthetic angle keeping unchanged, from the actual location of the target (A_0, E_0), you can get a group of theoretical miss distance data (u, v) corresponding to the encoder values (A and E);
4. Select the appropriate fitting method, solve the fitting parameters from the measuring miss distance (x, y) to the theoretical miss distance (u, v).

Table 28.1 Comparison between two distortion correction methods

Method name	Grid plate method	Handy correction method
Scope	Indoor	Indoor and test range
Required tools	Collimator, grid plate	Infinity fixed-point object
Sample size	Number of grid points	Infinitude
Correction effect	Very good	Good

The following should be noted when using the above method:

1. The selected points should be as many as possible and spread evenly over the whole calibration field of view.
2. After interpretation some odd point data should be eliminated.

In terms of the guiding idea, the handy method and the grid plate calibration method have no difference. For both method, the way is to find the transformation between the actual image coordinates and the theoretical image coordinate. The difference lies in that the grid plate method shoots multiple grid points in a stationary setup while the simple method turns to shoot a single fixed-point object. The benefits of doing so is that we can obtain a large number sample data of field of view anywhere, instead of only grid points. Therefore, the solved distortion transformation features more general implication. The disadvantage is the equipment rotation brings about more errors. However, the new errors are very small or negligible compared with the measurement errors caused by distortion. Comparison between two distortion correction methods is shown in Table 28.1.

28.4 Data Analysis

In experiments, the detected target is a far-away light at night. Operate equipment with a sine guide and shoot. Then, select 80 points through interpretation and use a variety of fitting methods in data processing.

Cubic polynomial fitting:

$$u = a_3x^3 + a_2x^2 + a_1x + a_0$$
$$v = b_3y^3 + b_2y^2 + b_1y + b_0$$

Radial fitting:

$$u = x + a_1x(x^2 + y^2) + a_2x(x^2 + y^2)^2$$
$$v = y + b_1y(x^2 + y^2) + b_2y(x^2 + y^2)^2$$

Table 28.2 Distortion correction processing results

Fitting methods	A-miss distance measurement accuracy	E-miss distance measurement accuracy
Original data	98.72″	117.38″
Cubic polynomial fitting	35.54″	33.22″
Radial fitting	54.95″	47.53″
Amendable radial fitting	26.99″	22.15″
Binary cubic polynomial fitting	20.65″	16.05″

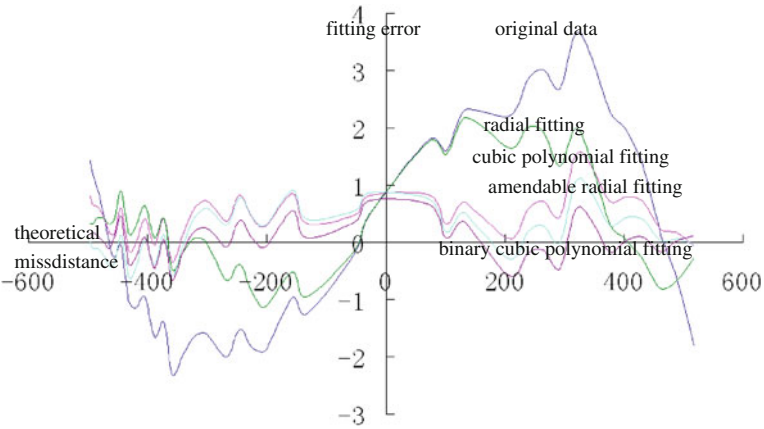

Fig. 28.4 A-miss distance data of different fitting methods

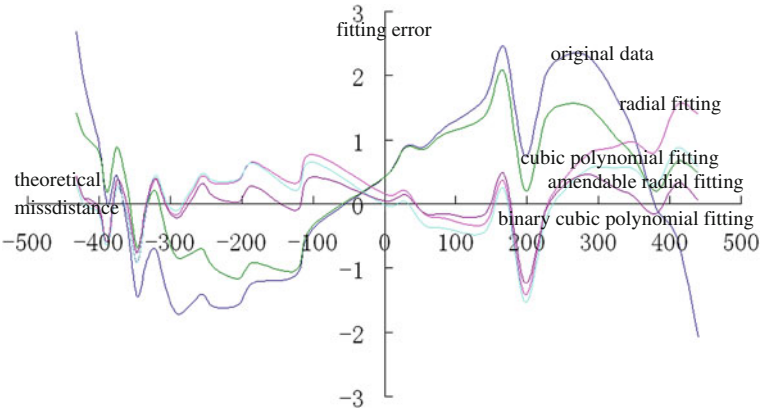

Fig. 28.5 E-miss distance data of different fitting methods

Amendable radial fitting:

$$u = a_0 x + a_1 x(x^2 + y^2) + a_2$$
$$v = b_0 y + b_1 y(x^2 + y^2) + b_2$$

Binary cubic polynomial fitting:

$$u = a_{30}x^3 + a_{20}x^2 + a_{10}x + a_{00} + a_{03}y^3 + a_{02}y^2 + a_{01}y + a_{12}xy^2 + a_{21}x^2y + a_{11}xy$$
$$v = b_{30}x^3 + b_{20}x^2 + b_{10}x + b_{00} + b_{03}y^3 + b_{02}y^2 + b_{01}y + b_{12}xy^2 + b_{21}x^2y + b_{11}xy$$

It can be seen from Figs. 28.4, 28.5 and Table 28.2 that binary cubic polynomial fitting method gets high processing precision with low fitting error (difference between the fitting miss distance and the theoretical miss distance). However, 20 parameters have to be solved and calculation is more complicated. For radial fitting, the accuracy is poor. This is probably because the focal plane of camera and the lens have assembly errors (the optical axis does not pass through the center of the focal plane, or the optical axis is not strictly vertical with the focal plane). The amendable radial fitting method features much less calculation load and higher accuracy and meets our requirements well.

28.5 Conclusions

The method proposed in this paper can effectively solve the problem of distortion correction for optical measurement systems in test range environment. The method is simple and convenient and there is no need of special-purpose devices or instruments. By selecting the appropriate fitting method, it can reduce computation and ensure measurement accuracy. Therefore, it has a high practical value.

References

1. He ZC, Hu BA (2002) Optical measurement system. National Defense Industry and Press, Beijing
2. Dong SF (1992) GJB1381.1-92 The accuracy evaluation of trajectory measuring equipments for missile and spacecraft test Photoelectric theodolites. National Defense Science and Technology Industry Council
3. Zhu ZT, Li SF (2005) Lens distortion and proofreading technology. Opt Tech 31(1):136–138
4. Xing ML, Liu JP, Lin JM et al (2003) Distortion measurement of CCD camera with a large-field short focal length lens. Opt Tech 29(2):377–379
5. Guo Y, Yang H, Yang ZJ et al (2008) Distortion measurement of lens in CCD camera system. J Appl Opt 29(2):377–379

Part III
Information Transfer and Processing

Chapter 29
Research on Multi-Path QoS Routing Strategy for the Satellite Network

Guanghua Song, Mengyuan Chao, Bowei Yang, Hua Zhong and Yao Zheng

Abstract The satellite network, together with the traditional ground network, constitutes the Space–Ground Interconnection Network (SGIN), which can provide global coverage for communications and become an important developing trend of the next generation network. Currently, there exist lots of routing algorithms that are based on the characteristics of satellite networks. However, most of them focus on a single transmission performance index while ignoring others. A multi-path parallel routing strategy for the satellite network based on Quality of Service (QoS) requirements and information of inter satellite links (ISLs) is proposed. Firstly, the strategy utilizes the historical and near-real-time information of the ISLs to conduct off-line routing computing on the ground. After that, real-time information of the ISLs and QoS requirements for data communications will be reconsidered on board to adjust the pre-computed routing scheme. Finally, data transmission will be allowed to take place on several routes between the source and the destination in parallel. This strategy not only ensures high arrival rate and low latency of data transmissions, but also fully utilizes link resources of the satellite network and balances the workloads among ISLs efficiently.

Keywords Satellite networks · QoS routing · Multi-path routing

G. Song · M. Chao · B. Yang (✉) · H. Zhong · Y. Zheng
School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China
e-mail: boweiy@zju.edu.cn

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_29,

29.1 Introduction

The satellite network can provide remote data transmission services with high bandwidth and high flexibility for data traffics such as voice, images and videos. It has advantages such as wide coverage, flexible networking and easy utilization. Since satellite communication is less affected by geographical environments and climate conditions, it is especially suitable for communications in mountain and sea areas where ground network is always difficult to deploy. As a result, the satellite network, together with the traditional ground network, constitutes the Space-Ground Interconnection Network (SGIN), which can provide global coverage of Tracking, Telemetry and Command (TT&C), navigation and communication and has become an important developing trend of the next generation network. In 2006, Shen Rongjun [1], academician of Chinese academy of engineering, proposed the idea of constructing the SGIN of China. In 2008, Zhang Jun [2] described the advantages of SGIN and pointed out that, the information transmission ability of SGIN will drive the formation and development of the burgeoning space industry.

To successfully construct SGIN, an efficient routing algorithm for satellite networks with Inter-Satellite Links (ISLs) is necessary. Different from traditional ground networks, the satellite network, which is made up of satellites and ISLs, has characters such as high dynamics, frequent link switching, unbalanced load, easy interruption and so on. Furthermore, the processing ability of computers on board is limited. Therefore, common routing protocols such as RIP, IGRP and OSPF in the ground are not well applicable to satellite networks. Many experts and scholars have conducted research in special routing technologies for satellite networks. Among the existing routing technologies, some focus on the topology of satellite networks [3–5], hoping to solve the problem of frequent changing topology with the minimum cost; some focus on the switching between ISLs [6], hoping to find a transmitting path which contains most stable ISLs; some focus on the transmitting delay [7], hoping to carry out the data transmitting task with low end-to-end delay; others focus on load balance of the whole network [8], the QoS requirements of different traffics [9], and the interconnection efficiency with the ground IP network [10].

After studying the existing satellite routing algorithms, we find that these algorithms solve the routing problem to some extent, however, there are still shortages in some ways: (1) Some algorithms create the routing table by conducting off-line computing [3, 4]. If some satellite nodes or ISLs break down, these algorithms could not well adapt to the changes, resulting in rapid performance degradation; (2) Some algorithms put forward high requirement to the on-board storage capacity [4], which is hard to meet in practice; (3) Some algorithms could well interconnect with the IP network [10], but they ignore the scalability of the routing table; (4) Some algorithms could ensure low delay of data transmitting [7], but they ignore the load balance of the whole network; (5) Some algorithms do not consider the generality [9], and only apply to the polar LEO satellite networks.

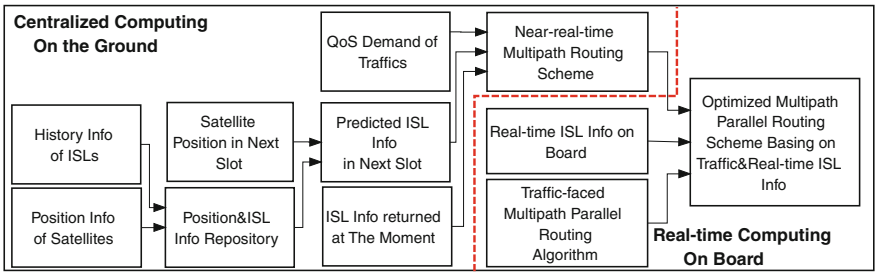

Fig. 29.1 Main working flow of multi-path QoS routing strategy

As far as we know, no algorithms have been reported that combine the global optimization and local adjustment, and improve the transmission quality by adopting multiple-path parallel transmission.

In this paper, considering the characteristics of the satellite network with ISL, we propose a multi-path parallel routing scheme based on QoS requirements of traffics and information of ISLs. The scheme utilizes satellite orbit parameters and the historical and near-real-time feedback information of ISLs to conduct off-line global routing computing on the ground. Then, it adjusts the pre-computed routing scheme according to the real-time ISL information and QoS requirements of traffics, making the traffic data travel along multiple proper paths in parallel, to ensure that the data can be conveyed as fast as possible. At the same time, it can make full use of the ISL bandwidth and help achieve load balance of the whole network.

29.2 Main Working Flow

As shown in Fig. 29.1, the multi-path QoS routing strategy is divided into two parts: the pre-computation stage on the ground and the real-time calculation stage on board. The pre-computation stage mainly includes the following steps: (1) To establish a "Position & ISL Information Knowledge Base" according to the regularly gathered ISL information and satellite orbital information; (2) To calculate the satellite position and speed information in the next slot and inquire the "Position & ISL Information Knowledge Base" to predict the ISL information in the next slot, which includes parameters such as the largest available bandwidth, delay, signal strength, etc.; (3) To combine the near-real-time feedback ISL information, the forecast global ISL information got from the above procedures and the possible QoS demands of different traffics to calculate the near-real-time multi-path routing scheme in the next slot. In the real-time calculation stage on board, LEO satellites will combine the uploaded "near-real-time multi-path routing scheme" with the QoS demands of traffics and the real-time ISL information to calculate the optimized multi-path parallel routing scheme. In Sect. 29.3 and Sect. 29.4, the pre-computation stage on the ground and the real-time calculation stage on board will be further discussed respectively.

29.3 Multi-Path Routing Pre-computation

29.3.1 Dynamic Topology of Satellite Networks

LEO satellites constantly rotate around the earth, and when they get into the polar circle, inter-orbit ISLs of them will be turned off, so the topology of the satellite network is dynamic. The dynamic characteristic of the entire network brings about some difficulties to the routing computation. By analyzing the changing topology of the satellite network, we can design a reasonable time partitioning strategy, which will transfer the dynamic network topology into a set of static network topologies. These static network topologies provide the necessary structural basis for subsequent global multi-path routing computation.

In order to get accurate orbit parameters of the LEO satellite, we can utilize professional tools like STK to conduct the satellite network simulation, record the satellite position information and ISL ON/OFF information, analyze the constellation characteristics, and determine a reasonable slot to divide the dynamic topology.

29.3.2 Position & ISL Information Knowledge Base

The cyclical change of physical parameters like position, angle, speed and the satellite network topology determines that the available bandwidth and delay of each ISL will change periodically. Therefore, through studying of the relationship among all these factors, we can set up a "Position & ISL Information Knowledge Base", which will offer a great help for forecasting the global network link status in the future and will provide an important premise for the global multi-path routing pre-computation.

The process of establishing "Position & ISL Information Knowledge Base" is complex, which needs lots of periodic statistical work about the information of each ISL in the satellite network. The information includes statistical parameters like the peak value, the mean value and the variance of the bandwidth and the delay of each ISL respectively, and will be further used to determine the maximum available bandwidth, the minimum delay and some other important parameters of each ISL. The statistical information finally constitutes "Position & ISL Information Knowledge Base".

29.3.3 Multi-Path Routing Pre-computation

By inquiring the "Position & ISL Information Knowledge Base" and combining the near-real-time ISL information, we can roughly predict each ISL's maximum

available bandwidth and propagation delay in the next slot. On this basis, multi-path routing pre-computation can be conducted according to various QoS demands beforehand. For example, to meet the delay requirement of the TT&C data, it needs to calculate the shortest path between two nodes according to the shortest path algorithm. To meet the bandwidth requirement of the real-time video data, it needs to calculate the path with the maximum available bandwidth between two nodes. Through some improvements on existing graph algorithms, we can find out alternative paths to meet the QoS demands of different traffics from the global optimization view, which provide foundations for further calculation on board.

29.4 Multi-Path Parallel Routing

29.4.1 Link Bandwidth Allocation

The multi-path routing pre-computation process provides each LEO satellite with several paths for different QoS demands in the next slot. Therefore, when different traffics are to be transmitted from one LEO satellite to others along several paths, the following two questions come up: (1) how to assign the data of a traffic flow to each ISL; (2) how to allocate the available bandwidth of an ISL to each traffic flow. The router on board needs to weigh in the multiple alternative paths, and to eventually select one or more reasonable transmission paths for each traffic flow, according to the priority of current traffic and real-time ISL information. The rationality of the link bandwidth allocation strategy not only determines whether the current traffic can arrive at the destination node successfully, but also determines the utilization of link resources of the satellite network during a period of time in the future.

The allocation of the available link bandwidth is determined by both the features of each traffic flow and the real-time ISL information, so the final allocation scheme is a balanced result of various factors. When designing the link bandwidth allocation strategy, game theory and related resource allocation theory can be exploited to balance the QoS requirements of the traffics, the actually available link bandwidth, the current delay and other factors. Meanwhile, combing with software simulation methods, we can evaluate the proposed bandwidth allocation algorithm. By tuning the parameters, we can finally implement an algorithm that achieves the best effects.

Figure 29.2 illustrates the transmission process of a traffic flow from S34 to S11. At node S12, S13, S22, S23, the former traffic flow is divided into several different flows for the following parallel transmission, according to the real-time ISL information and QoS demands. Through a certain identification strategy, we can establish the corresponding relationships between the former data flow and the split data flows, which make it easy to conduct the assembly process of all the data fragments at the destination node. In Fig. 29.2, a simple “length/offset” strategy is

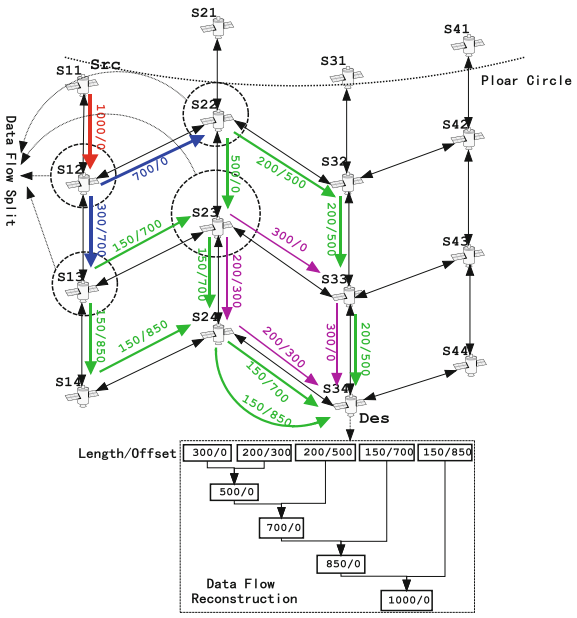

Fig. 29.2 An example of multi-path parallel routing

depicted: when all the data fragments arrive at the destination node, we first sort them according to their "offset". Then, if the offset of fragment1 + the length of fragment1 = the offset of fragment2, they can be immediately merged. As a result, the length of the newly merged fragment1 is (the length of fragment1 + the length of fragment2), while the offset remains the same.

29.4.2 Link Bandwidth Pre-emption

Different from the traffics in the ground network, traffics in the satellite network such as high resolution images and videos put forward high requirements for link bandwidth. When traffics with high priority need to be sent out, the available bandwidth may have been occupied by traffics with low priority. Therefore, the router on board should let the traffic with higher priority preempt the available link bandwidth to meet its requirements. Pre-emption of the link bandwidth not only concerns whether the traffic data with high priority could arrive successfully at the destination, but also takes into account whether the traffic data with low priority can be still successfully transmitted to the destination after the bandwidth is preempted. Therefore, preemption of link bandwidth is a key point to realize overall optimization of the network.

As shown in Fig. 29.3, the original traffic with a bandwidth of 3Mbit/s was routed along S31-S32-S22-S23-S24, via the link S22-S23. Assume that a bandwidth of 3Mbit/s of the link S22-S23 has been occupied, and a bandwidth of

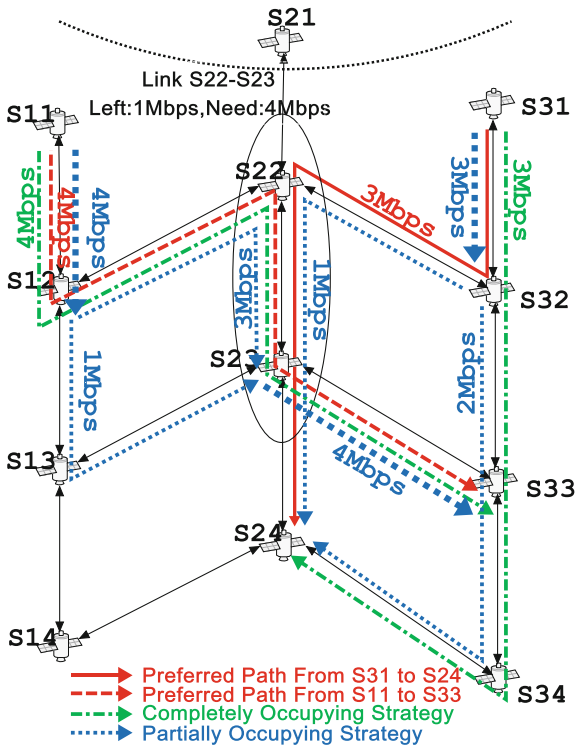

Fig. 29.3 An example of link bandwidth pre-emption

1Mbit/s is free. At this moment, if a traffic with higher priority that requires a bandwidth of 4Mbit/s is to be transmitted along S11-S12-S22-S23-S33, which should go through the link S22-S23. The available bandwidth of the link S22-S23 could not satisfy the traffic. To deal with this issue, pre-emption strategy based on traffic priority is applied to allow the traffic with higher priority to completely or partly preempt the bandwidth of the link. To successfully transmit the data, the traffic whose bandwidth is preempted will assign part of its data to other paths.

29.4.3 Two Typical Applications of Multi-Path Parallel Transmission

To concretely illustrate the advantages of multi-path parallel transmission in the satellite network, two typical applications will be introduced below.

1. Multi-Path Parallel Transmission of High Bandwidth Video Streams.

As shown in Fig. 29.4, satellite node *A* needs to transmit a video stream with a bandwidth of 800kbit/s to node *G*, but the remaining link bandwidth in the satellite network could not support the transmission with a single path. However, if we

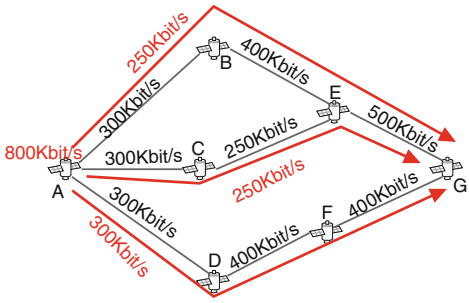

Fig. 29.4 Transmission of high bandwidth video streams

divide this video stream into three parts: 250kbit/s, 250kbit/s and 300kbit/s, and separately transmit them along the path *A-B-E–G, A-C-E–G, A-D-F-G*, and merge them at node *G*, the effective bandwidth of the transmission is 800kbit/s. Therefore, by dividing traffic stream into several parts and transmitting them along multiple paths, the effective bandwidth of a traffic flow can be highly improved.

2. Multi-Path Parallel Transmission of Low Delay TT&C Data.

As shown in Fig. 29.5, the satellite node *A* needs to transmit TT&C data with high priority to the node *G*, and the data can tolerate a maximum delay of 120 ms. Through querying node *A*'s routing table, we find that there are three paths *A-B-E–G* (120 ms), *A-C-E–G* (105 ms) and *A-D-F-G* (110 ms) that can satisfy the transmission requirements. However, compared with the ground network, the failure probability of links in the satellite network is much higher. In order to improve the arrival rate of the TT&C data, we can let the path *A-C-E–G* with the shortest delay transmit the data, and let path *A-B-E-G* and path *A-D-F-G* transmit the replicas of the data respectively as well. In this way, even if there is a link malfunction in one path, other paths can ensure that the data arrives in time. Therefore, by transmitting replicas of data along several paths, the usability of the network link can be effectively improved.

29.4.4 Multi-Path Parallel Transmission Network Protocol

To support Multi-path Parallel Routing on board, we propose a draft of eUDP, an extension of the UDP protocol, as shown in Fig. 29.6. In the draft, if a session is setup between two nodes, a unique "Association_ID" marking the session will be allocated. The offset of a fragment is recorded by the "Start_Pos_at_Initial" field, and the length of the fragment is recorded by the Length field. In this way, eUDP can well support any segmentation and reconstruction of traffic streams and multi-path parallel transmission of a traffic flow. The Route field records the planned path from the source node to the destination node. If the effect of real-time link status is not considered, the packet will be transmitted along this path. In addition, since this draft is based on the UDP/IP protocol, it is compatible with the existing protocol stack.

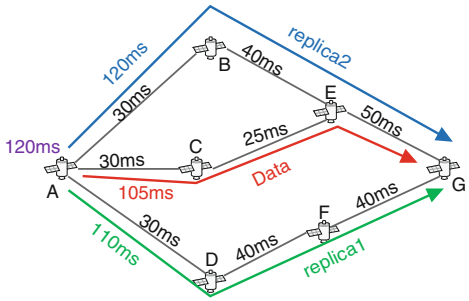

Fig. 29.5 Transmission of low delay TT&C data

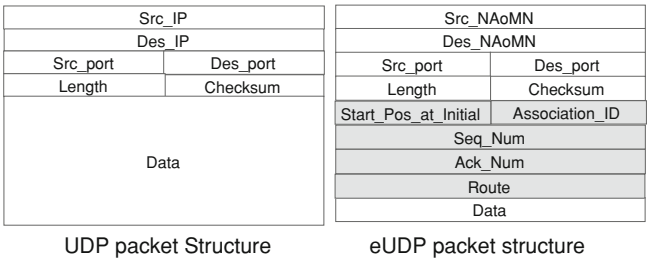

Fig. 29.6 Multi-Path parallel transmission network protocol

29.5 Conclusions

To meet the demands of the development of satellite applications and to make up the insufficiency of existing routing algorithms, this paper presents a multi-path parallel routing scheme that is based on the demands of traffics and the real-time ISL information. By using satellite orbit parameters, historical ISL information, and periodic feedback near-real-time ISL information to conduct offline routing pre-computation, the scheme effectively overcomes the shortcomings of the routing algorithms in traditional ground networks. In addition, the scheme combines the real-time ISL information and the QoS demands of traffics to adjust the pre-computed routing scheme, making the traffic data to be transmitted in parallel along multiple paths. Theoretical analysis shows that the presented scheme ensures high accessibility of the traffics, full utilization of link resources and load balance of the satellite network.

References

1. Sheng RJ (2006) The conception of constructing China's space-ground interconnection network. Chinese Eng Sci 8(10):19–30
2. Zhang J (2008) Space-ground interconnection network for the future. Int Aviat 9:36–39
3. Werner M, Delucchi C, Vogel HJ et al (1997) ATM-based routing in LEO/MEO satellite networks with inter satellite links. IEEE J Sel Areas Commun 15(1):69–82
4. Wang JL, Yan J, Cao ZG (2007) Optimization for routing algorithm based on snapshot sequence in LEO satellite network. J Tsinghua Univ (Natural Science Edition) 47(4):559–563
5. Wang JF, Li L, Zhou MT (2007) Topological dynamics characterization for LEO satellite networks. Comput Netw 10(3):43–53
6. Li H, Gu XM (2005) Research on routing algorithm and switching performance in LEO/MEO satellite communication systems. Syst Eng Electron 27(7):1145–1153
7. Korçak Ö, Alagöz F, Jamalipour A (2007) Priority-based adaptive shortest path routing for IP over satellite networks. Int J Commun Syst 20(3):313–333
8. Rao Y, Wang RC (2010) Agent-based load balancing routing for LEO satellite networks. Comput Netw 54(17):3187–3195
9. Rao Y, Wang RC (2011) Performance of QoS routing using genetic algorithm for polar-orbit LEO satellite networks. Int J Electron Commun 65(6):530–538
10. Hashimoto Y (1998) Design of IP-based routing in a LEO satellite network. In: Proceeding of the 3rd international workshop on satellite-based information services, Dallas, pp. 81–88

Chapter 30
Constant Modulus Blind Equalization Analysis for High Speed Implementation

Dalong Yang, Dahai Chen and Wen Kuang

Abstract Constant modulus equalization algorithm (CMA) is widely used for mitigating inter-symbol interference (ISI) in various communication systems, for its robustness and low computational complexity. In practice, signed error CMA may be adopted to simplify the implementation. However, the simplification severely degrades its steady state performance. In this paper, a detailed analysis is taken to compare the mentioned two algorithms in Field Programmable Gate Arrays (FPGA) implementation. Furthermore, a novel error function is proposed for the high speed implementation consideration. Numerical simulation demonstrates that the proposed algorithm behaves much better than the reference with similar computational complexity, which makes the proposed algorithm more suitable for the area of high speed applications.

Keywords Blind equalization · High speed application · Constant modulus algorithm · FPGA

30.1 Introduction

With the rapid development of telemetry technology, the telemetry environment becomes more and more complicated, which increases the performance requirements to the telemetry systems. In order to ensure reliable communication, channel equalizers are usually adopted in the receiver to compensate the inter-symbol

D. Yang (✉) · D. Chen · W. Kuang
Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621900, China
e-mail: yangdleo@gmail.com

D. Yang
Department of Engineering Physics, Tsinghua University, Beijing 100084, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_30,

interference (ISI), which mainly caused by multipath propagation and the band limitation effect. Consequently, channel equalization techniques have been studied for decades and various algorithms were proposed. Blind equalization attracted many attentions, considering that no training sequence is needed for start-up. Considering its robustness and low computational complexity, Godard's constant modulus algorithm (CMA) [1] became one of the most popular blind equalization algorithms. Numerous modifications on CMA were made to improve its convergence rate, steady state performance, phase ambiguity resolution, high-order modulation adaptability, and so on. The frequently used algorithms include the variable step-size CMA [2], multi-modulus algorithm (MMA) [3], and fractionally spaced CMA [4]. These algorithms all increase the computational complexity of the original one. For transmitting MPSK modulation signals, the traditional CMA behaves well enough for its nearly Wiener solution performance, and suitable for high speed implementation. In this paper the CMA in MPSK modulation scheme is selected to be analyzed for high speed implementation.

30.2 CMA and Its Simplification

Assuming a multipath channel scenario with additive white Gaussian noise, the signal $\{a(k)\}$ is transmitted through the channel $\{h(k)\}$, and then we get the signal $\{x(k)\}$ at the receiver, indicated by

$$x(k) = \sum_{i=0}^{M-1} h(i)a(k-i)e^{j\varphi(k)} + n(k) \tag{30.1}$$

where $\varphi(k)$ denotes phase offset and $n(k)$ denotes additive complex noise process independent of $\{a_n\}$. Filtered by the L taps adaptive equalizer vector $\boldsymbol{w}(k) = [w_0(k)\ w_1(k)\ \cdots\ w_{L-1}(k)]^{\mathrm{T}}$, we get the equalized signal sequence $\{z(k)\}$:

$$z(k) = \boldsymbol{x}^T(k)\boldsymbol{w}(k) \tag{30.2}$$

where $\boldsymbol{x}(k) = [x(k)\ x(k–1)\ \cdots\ \mathrm{x}(k–L+1)]^{\mathrm{T}}$.

In the baud-spaced condition, CMA tries to minimize the cost function $D^{(p)}$

$$D^{(p)} = E\left(|z_n|^p - R_p\right)^2, R_p = E|a_n|^{2p} \Big/ E|a_n|^p \tag{30.3}$$

where p is the order of the algorithm.

Without special explanation in this paper, CMA means the algorithm for $p = 2$. And the CMA update function is given by

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \mu z(n)(|z(n)|^2 - R_2)\boldsymbol{x}^*(n) \tag{30.4}$$

where μ is the step-size.

For implementation simplification with less quantization bit width of the tap coefficients, the signed error update function is usually used

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu z(n) \cdot sign(|z(n)|^2 - R_2)\boldsymbol{x}^*(n) \tag{30.5}$$

where the function sign(·) means the sign function,

$$sign(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \tag{30.6}$$

The equalizer performance is largely due to its steady state error. The fluctuation of the tap updates directly influence the equalizer's output error. We use the variance to characterize the degree of the fluctuation.

$$\begin{aligned} Var_{Godard} &= \mu^2 Var\left[z(n)(|z(n)|^2 - R_2)\boldsymbol{x}^*(n)\right] \\ Var_{Sign} &= \mu^2 Var\left[z(n) \cdot sign(|z(n)|^2 - R_2)\boldsymbol{x}^*(n)\right] \end{aligned} \tag{30.7}$$

Where the subscript Godard means the standard Godard CMA, and the Sign mean the signed one. Assuming the equalizer has already converged, the fluctuation expectation should be zero. For MPSK modulation signal, the ideal amplitude of $z(n)$ should be a constant, indicated by A.

According to the mathematical theorem:

$$E\left(X^2 \cdot Y^2\right) \le E\left(X^2\right) \cdot E\left(Y^2\right) \tag{30.8}$$

We get:

$$\begin{aligned} Var_{Godard} &= \mu^2 E\left|z(n)(|z(n)|^2 - R_2)x^*(n)\right|^2 \\ &\le E\left|(A^2 - R_2)\right|^2 \cdot \mu^2 A^2 \cdot E|x(n)|^2 \\ Var_{Sign} &= \mu^2 Var\left[z(n) \cdot sign(|z(n)|^2 - R_2)x^*(n)\right] \\ &\approx \mu^2 A^2 \cdot E|x(n)|^2 \end{aligned} \tag{30.9}$$

According to Eqs. (30.9), when the step-sizes for the two algorithms are the same, the variance of the signed CMA is proportional to the square of the equalizer's input and output signal. But for the standard CMA, the variance is much less, because the value of $(A^2 - R_2)$ is very small when the equalizer has converged. Therefore the standard CMA is not so sensitivity to the amplitude fluctuation of the input signal. For convergence rate, the signed CMA will be faster for its bigger update value. For hardware requirements, the standard CMA needs a wider data quantization bit width for its small update value, in order to ensure the stability of the convergence (Table 30.1).

Table 30.1 The comparison of the standard CMA and signed CMA

	Standard CMA	Signed CMA
Error function	$e_n = z_n(\|z_n\|^2 - R_2)$	$e_n = z_n \cdot sign(\|z_n\|^2 - R_2)$
Equalized signal	Fluctuation Small and constellation compact	Fluctuation big and constellation loose
Quantization	requirement	Wider bit width
Normal		
Overall evaluation	High performance, slow convergence rate, more hardware requirement	Low performance, fast convergence rate, less hardware requirement

30.3 New Error Function

From Sect. 30.2, the two algorithms both have their own advantages and disadvantages. With a comprehensive consideration of the two algorithms, we propose a new kind of error function:

$$e_n = z_n f(v_n), v_n \triangleq (|z_n|^2 - R_2) \tag{30.10}$$

where the specified form of function $f(\cdot)$ determines the performance and complexity of the error function. Considering the numerical implementation, the function $f(\cdot)$ could be given by

$$f(v) = \begin{cases} v & ,|v| > a \\ a \cdot sign(v) & ,|v| < a \end{cases} \tag{30.11}$$

where the parameter a is a constant value, for example the integer power of 2, and its value determine the quantization bit width of the error function.

Using this error function, we can not only ensure a smaller steady-state variance than that of the signed CMA, but also decrease the quantization bit width and get a faster convergence rate than the standard CMA. We use mixed CMA to identify the new algorithm. This kind of adjustment is equivalent to adopt different update step-size and switch between the standard CMA and the signed one at the same time. At the beginning, use the standard CMA with a big step-size, and when the equalizer error is small enough, switching to the signed CMA with a $1/a$ reduced step-size. Overall, the new algorithm is an algorithm between fixed step-size algorithm and variable step-size one, and much computational simpler than the variable step-size algorithm. In Ref. [2], the high order error function was used to adjust the step-size:

$$\mu(k) = \mu + [M_m(k+1) - M_m(k)] \tag{30.12}$$

$$M_m(k) = E[e^m(k)] \approx \frac{1}{L+1} \sum_{i=0}^{L} e_i^m(k) \tag{30.13}$$

where $M_m(k)$ is the m order origin moment, which is estimated by time average of (L + 1) samples of error function. The computational complexity is not so cheap for estimating high order statistics, especially for complex data, and this limits this variable step-size algorithm's application in practical systems.

30.4 Numerical Simulation

For the three kind of algorithms, we use 8PSK modulation signals to simulate the equalizer's performance. The impulse response of the channel which is used in [5] is characterized by

$$\begin{aligned} H(z) = & (-.005 - j.004) + (.009 + j.030)z \\ & + (-.24 - j.104)z^2 + (.854 + j.520)z^3 \\ & + (-.218 + j.273)z^4 + (.049 - j.074)z^5 \\ & + (.16 + j.020)z^6 \end{aligned} \tag{30.14}$$

All three algorithms have 17 taps using center tap initialization method. The center tap is initialized to 1.2. The amplitude expectation of the transmitted signal $\{a(k)\}$ is equal to 1, so the CMA constant parameter is set to 1. The step-size is set to 1.5×10^{-3}. The constant value a for mixed CMA is set to 0.1. We use truncate quantization method to give the fixed point simulate results.

For evaluating the equalizer's performance we use the residual inter-symbol interference (RISI) which is defined by

$$RISI(n) = \frac{\sum_i |h(i) * w_i(n)|^2 - |h(i) * w_i(n)|^2_{\max}}{|h(i) * w_i(n)|^2_{\max}} \tag{30.15}$$

where the symbol * means the convolution operator.

30.4.1 The Effect of Different Amplitudes of Input Signal

The performance of the standard CMA, signed CMA and the mixed new one is simulated when the amplitude of the equalizer input signal varies from 0.2 to 2.3. The mean value and variance of RISI which is calculated in steady states are shown in Figs. 30.1, 30.2 respectively.

It can be seen from the figures above, the Signed CMA is very sensitive to the amplitude of the equalizer input, which is also indicated by analysis in Sect. 30.2. And the RISI mean value for mixed CMA is less than 0.1 when the amplitude

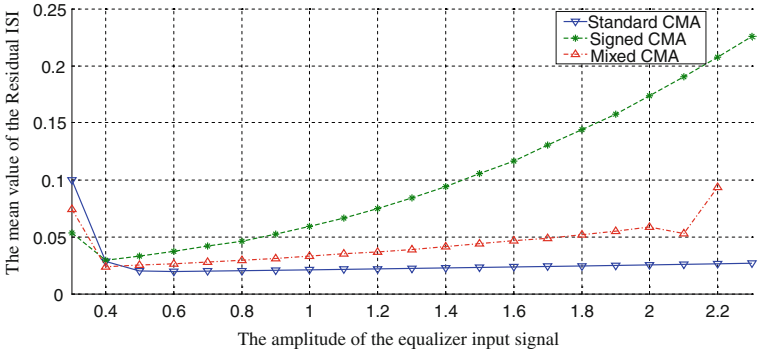

Fig. 30.1 The mean value of RISI for different equalizer input amplitude

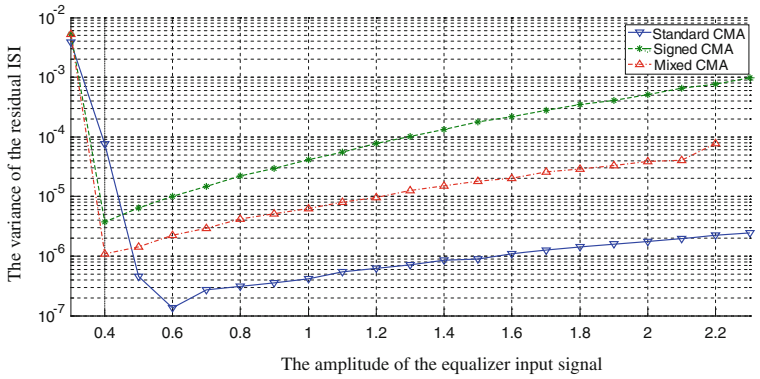

Fig. 30.2 The variance of RISI for different equalizer input amplitude

varies from 0.3 to 2.2 and the variance is also much less than the signed CMA, indicating the better adaptation ability and performance than signed CMA. The mixed CMA is a kind of sub-optimal form of the standard CMA for hardware implementation simplification.

30.4.2 The Convergence Rate

The convergence rate is evaluated by using the same input signal with the amplitude of 1.0 and equalizer parameters for each algorithm, the RISI curves are shown in Fig. 30.3.

As indicated by Fig. 30.3, the convergence rate of signed CMA is the fastest one but with the poorest performance, and the standard CMA is the slowest one but with the best performance. The mixed CMA's convergence rate is slightly faster than the standard CMA algorithm, and the performance is nearly the same with it.

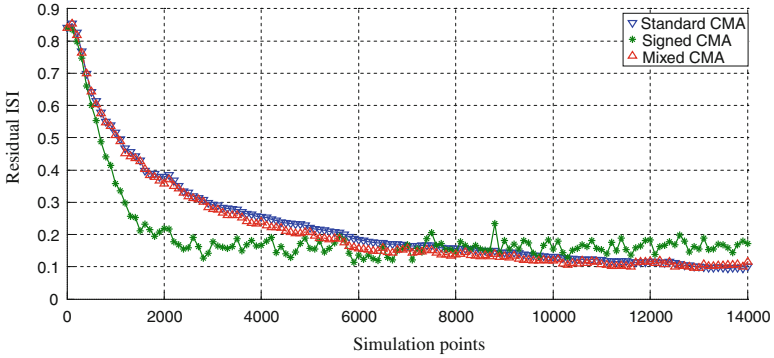

Fig. 30.3 The convergence curve for each algorithm

Table 30.2 The convergence performance of each algorithm for different quantization bit width

	Signed CMA		Mixed CMA		Standard CMA	
Bit width	Mean of RISI	Variance of RISI	Mean of RISI	Variance of RISI	Mean of RISI	Variance of RISI
30	0.1120	2.00×10^{-4}	0.0332	6.61×10^{-6}	0.0220	1.13×10^{-6}
20	0.1120	2.00×10^{-4}	0.0332	6.40×10^{-6}	0.0245	2.29×10^{-6}
19	0.1120	1.99×10^{-4}	0.0335	6.37×10^{-6}	0.0293	4.01×10^{-6}
18	0.1120	1.99×10^{-4}	0.0342	6.69×10^{-6}	0.0413	7.62×10^{-6}
16	0.1117	2.04×10^{-4}	0.0492	2.78×10^{-5}	0.1284	9.88×10^{-5}
15	0.1133	2.00×10^{-4}	0.2268	4.23×10^{-4}	0.2582	4.33×10^{-4}
14	0.1162	2.14×10^{-4}	0.5295	1.90×10^{-3}	–	–
13	0.1350	3.72×10^{-4}	–	–	–	–
12	9.8162	28.5894	–	–	–	–

30.4.3 The Effect of Quantization Bit Width

The above two simulations are done with sufficient quantization bit width, and in this section we analyze the influence of quantization bit width to the equalizers' performance. The amplitude of the equalizers' input signals is fixed at 1.0, and the input signals are quantized to 9-bit signed number. When using different quantization bit width to indicate the equalizer tap coefficients, the convergence performance of each algorithm is shown in Table 30.2.

From Table 30.2, we can see that the impact of quantization error on the equalizer's performance is very great, and varies for different algorithms. The standard CMA, although theoretically able to achieve the best performance, is most sensitive to the quantization bit width. When the quantization bit width is less than 20, the deterioration becomes obvious. When the bit width is less than 16, it becomes the worst one even. The mixed CMA could get the best performance in median quantization bit width of 16–18, which is suitable for practical use.

30.5 Conclusions

In this paper, we have analyzed the standard CMA and the signed CMA for the theoretical performance and the impact of quantization bit width. Considering the two algorithms cannot satisfy the convergence performance and the hardware consumption at the same time, we have proposed a new kind of error function, and gave a specified realization for simple hardware implementation which is less computational complexity than the standard CMA algorithm. Numerical simulations have shown that the proposed new algorithm has a good performance close to the standard CMA algorithm and better convergence rate. The mixed CMA is a good compromise of the standard CMA and signed CMA in simpler hardware implementation and convergence performance. It is a tradeoff between the hardware consumption and convergence performance and is the most suitable form for practical implementation.

References

1. Godard DN (1980) Self-recovering equalization and carrier tracking in two-dimensional data communication systems. IEEE trans commun COM-28:1867–1875
2. Ying Xiao, Zhenxing Li, Yuhua Dong (2008) Constant modulus algorithms using variable step-size. Commun Technol 41(5):47–49
3. Yuan JT, Tsai KD (2005) Analysis of the multimodulus blind equalization algorithm in QAM communication system. IEEE Trans Commun 53:1427–1431
4. Johnson CR Jr, Schniter P, Endres TJ et al (1998) Blind equalization using the constant modulus criterion: a review. Proc IEEE 86:1927–1950
5. Picchi G, Prati G (1987) Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm. IEEE Trans Commun COM-35:877–887

Chapter 31
Study on Space Mission IP Network QoS Technologies

Yunjun Chen, Yan Liu and Shengwang Xu

Abstract This paper addresses the QoS technologies and strategies in space mission IP network. It is analyzed of the primary factors, including that causes congestion and influences the QoS performance. This paper aims to provide a top-level design on the QoS strategy and commits suggestions on the development of the QoS technologies in next generation space mission IP networks. With respect to the three potential congestion causes, i.e., excessive traffic access, unbalanced resource, and transmission mechanism, corresponding countermeasures are proposed. The QoS strategy of the space mission IP network is a systematic solution whose first role is to avoid congestion and guarantee the transmission with high priority when congestion occurs. The architecture of the QoS strategy of the space mission IP network can be summarized as a congestion avoidance mechanism based on traffic management, traffic output control, as well as additional service mechanisms. The congestion avoidance mechanism includes a set of strategies such as traffic policing, traffic shaping, and port rate limitation. The differential service mechanism includes a set of strategies such as service classification, priority marking, priority trust, and priority queue dispatch. With these strategies implemented, the end-to-end congestion management and QoS guarantee is achieved in the space mission IP network.

Keywords Space mission · IP network · QoS

Y. Chen (✉) · Y. Liu · S. Xu
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: chenyunjun@bittt.cn

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_31,

31.1 Introduction

Quality of Service, or QoS [1], is the service requirement that should be satisfied when data is transmitted in a network, which can be quantified with bandwidth, time delay, delay jitter, packet loss rate, and throughput. The radical goal of the implementation of QoS strategy is to avoid congestion beforehand as well as to minimize the loss when congestion occurs.

There are two types of QoS models, i.e. integrated services model, or IntServ [1], and differentiated services model, or DiffServ [2]. As a stream-based QoS technology, IntServ model claims its most significant advantage of end-to-end QoS guarantee, and disadvantage of requiring much more resources and complicated management. As a classification-based QoS technology, DiffServ model is much easier to implement with better flexibility.

During the process of space flights, all kinds of real-time data such as command, telemetry, voice, and so on, claims very strict requirement with respect to packet loss, delay, jitter, and availability. To meet that requirement, a light loading strategy is implemented in the space mission IP network which provides good transmission performance with the cost of bandwidth. Moreover, a QoS strategy should be implemented, which tries to avoid congestion beforehand and guarantee the reliable transmission of important information when abnormal circumstances occur.

31.2 Analysis on Factors Influencing the QoS of Space Mission IP Network

A radical precondition to guarantee the QoS is that the network does not interrupt. That is to say, only when the reliability of a network is guaranteed that it is possible to discuss how to guarantee the QoS such as packet loss, delay, and jitter. To guarantee the reliability of the network, a thorough architecture is deployed in the space mission network which can be summarized as *double network plains organization and double routes data transmission*. It guarantees the reliability of the network with redundancy of equipments, circuits, and information. With that precondition, congestion is the only factor that influences the QoS. Therefore, it is the radical way to implement effective management on congestion to guarantee the QoS.

31.2.1 Architecture of the Space Mission IP Network

As is illustrated in Fig. 31.1, the space mission IP network consists of MANs and LANs which are connected by a WAN. The MANs or LANs are organized in hierarchical architecture. The MANs are comprised of a core layer, an aggregation

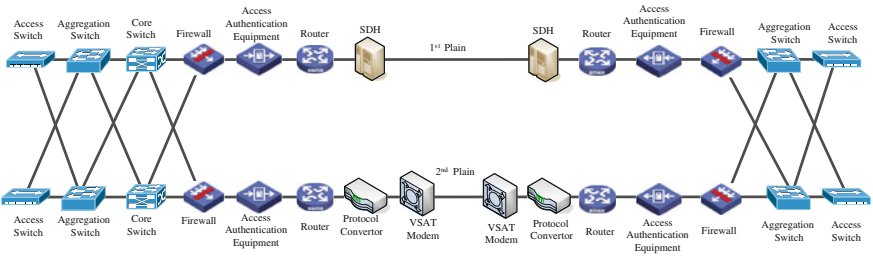

Fig. 31.1 Architecture of space mission IP network

layer, and an access layer while the LANs are comprised of an aggregation layer and an access layer. Switches are used for the MANs or LANs while routers for the WAN. Fiber circuits are used for the first route while satellite circuits for the second route. Firewalls are used between the MANs or LANs and the WAN.

31.2.2 Analysis and Countermeasure on Factors Causing Congestion

The primary factors that could cause congestion in the space mission IP network include excessive traffic access, unbalanced resource, and transmission mechanism, as is illustrated in Fig. 31.2.

31.2.2.1 Analysis on and Countermeasure against Excessive Traffic Problem

Analysis on Excessive Traffic Problem

The bandwidth of a network link, especially the WAN link, is deployed in terms of the data transmission requirement of all users. However, congestion will occur at the bottleneck if the actual access rate of one or several users exceeds their requested rate.

There could be two types of excessive traffic access problem. The first one is average excessive traffic problem. If telemetry equipment has requested a bandwidth of 64 kbit/s, and the telecommunication department sets a 64 kbit/s circuit for the WAN transmission link. However, when the actual access data rate exceeds 64 kbit/s and is as high as 100 kbit/s, it will exceed the throughput of the WAN and cause congestion and packet loss at the entrance of the WAN. The second type is interval excessive traffic problem. We continue to take the telemetry equipment as the example. The actual access rate of that equipment does not exceed 64 kbit/s. If it does not keep a well-distributed data transmission style but accumulate for a

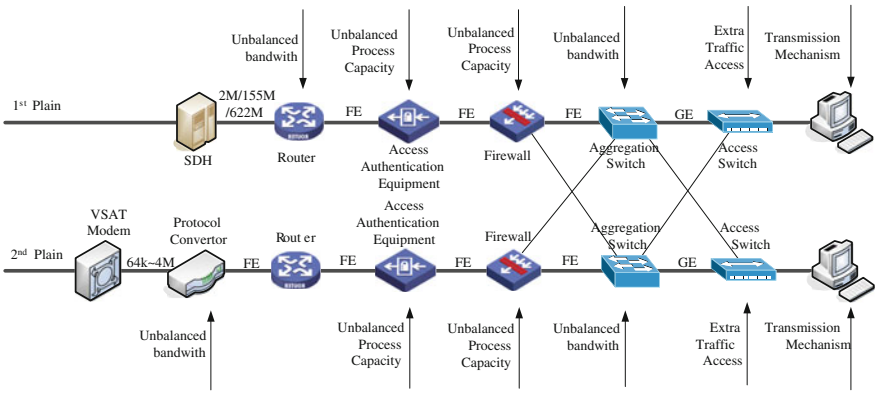

Fig. 31.2 Locations of possible factors causing congestion

while the data to be sent and burst transmit them intermittently, then during some time the sending data rate would exceed the throughput of the network and cause congestion and packet loss. Even if all the equipments keep a well-distributed data transmission style, one packet per second, the excessive traffic problem still could happen. It occurs when all the terminal equipments send data at the same time.

Countermeasure against Excessive Traffic Problem

Coordination between the terminal equipments and the network should be carried out and an agreement should be achieved that all the terminal equipments are supposed to send data at a stable rate and in a well-distributed style. Besides, a strategy must be implemented at the entry of the network to monitor and control the access traffic of users and discard those data that exceeds the agreed rate.

Traffic Conditioning [2] includes metering, policing, shaping, and packet marking. Traffic policing is the process of controlling packets. It monitors the access packet rate and penalizes the excessive traffic to enforce a reasonable traffic profile. The following should be noticed when implementing traffic policing strategy:

1. As for token bucket mechanism used in traffic policing, the token rate determines the packet rate and the bucket volume determines the burst rate. Consequently, it is really critical for these two parameters.
2. As to the DiffServ model, it is much more critical to implement the traffic policing strategy to the traffic with higher priority. The reason is that the higher the traffic priority is, the more influence it imposes on other traffic. As to the traffic with lower priority, it will be discarded by the network equipment if congestion occurs and will not influence the traffic with higher priority even when its access rate exceeds the agreed rate.

31.2.2.2 Analysis on and Countermeasure against Unbalanced Resource Problem

Analysis on Unbalanced Resource Problem

Unbalanced resource problem includes unbalanced bandwidth and unbalanced process capacity.

Unbalanced Bandwidth

Unbalanced bandwidth means the inconsistency between the entry port and exit port at a network node. When the entry port rate exceeds the exit port rate, the packets coming from the entry port will be suffered at the exit port. If the suffer volume is overflowed, the congestion occurs and packet loss happens.

In the space mission IP network, the nodes with unbalanced bandwidth ports are those following:

1. In both the first and the second network plain, the aggregation switch is connected with the access switch with a GE port and the firewall with a FE port.
2. The router of the first network plain uses a FE port as the LAN interface while a 2 M, 155 M or 622 M SDH port as the WAN interface.
3. The router of the second network plain uses a protocol convertor to connect with the VSAT modem and the protocol convertor is connected with the router with a FE port while the VSAT modem with a synchronous port from 64 kbit/s to 4 Mbit/s.

Unbalanced Process Capacity

Unbalanced process capacity means the process rate of some nodes in the network cannot satisfy the requirement of line speed transmission and causes a throughput bottleneck. If the incoming packet rate exceeds the process capacity of that node or equipment, there will be congestion happening because that node or equipment cannot process and transmit the packets as fast as enough.

All the routers and switches in the space mission IP network owns line speed process capacity and the bottleneck of the throughput occurs primarily at the security equipments such as the firewall and the access authentication equipment.

Countermeasure against Unbalanced Resource Problem

Traffic in the WAN is less than that in the MANs and LANs. The resources of long distance transmission circuit are less than local resources. The bandwidth of the WAN link is less than that in the MANs or LANs. The process capacity of the security equipment is less than that of the switches and routers due to their different function. That is to say, the unbalanced resource is the real fact in the network, which should be faced and considered carefully when the QoS strategy is designed. Concrete measures include the following items.

1. Setting appropriate port suffer at the nodes with unbalanced resource problem.
2. Due to the fact that QoS strategy cannot be implemented on some equipment such as firewall, access authentication equipment, protocol convertor, and the VSAT modem, it is necessary to find a way to transfer the throughput bottleneck there to their upstream neighbor node where the QoS strategy can be implemented. To achieve this goal, traffic control strategy should be implemented at the upstream node which includes traffic rate control and traffic profile control. The traffic rate control strategy is implemented to avoid the traffic rate exceed the process capacity of the downstream neighbor node. The traffic profile control strategy is implemented to force the traffic pass the downstream neighbor node smoothly and well-distributed.

31.2.2.3 Analysis on and Countermeasure against Transmission Mechanism Problem

Analysis on Transmission Mechanism Problem

IP information transmission mechanism can be connection-oriented or connectionless.

As a connection-oriented reliable transmission protocol, TCP sets up a connection before transmitting data. The reliability of TCP protocol is guaranteed by CRC, timer, data sequence number as well as ACK. The sender assigns a sequence number for each packet to ensure the right sequence and eliminate the duplicated data. The receiver sends an ACK to the sender to ensure reliable transmission. If some packets are discarded due to channel code error or congestion, the sender will retransmit the packets which will make the congestion worse and even cause congestion collapse. The source-based TCP congestion control mechanism can effectively avoid congestion collapse [3]. However, when many TCP connections enable slow start [4] and congestion avoidance [4] or TCP global synchronization at the same time, they decrease traffic at the same time and that make the total traffic fluctuates and ineffective use of the bandwidth.

Countermeasure against Transmission Mechanism Problem

As a connectionless protocol, UDP provides connectionless datagram service and do not retransmits the packets lost. It will not make congestion worse from the congestion avoidance point of view although it cannot guarantee the arrival of the packets.

The space mission IP network is primarily deployed over VSAT and fiber circuits and the stability and performance of the circuits can meet the requirement

of the space mission. That is to say, the connection-oriented TCP reliability is not the only choice to guarantee the information reliability any more. Thus, in summary, UDP protocol should be used to transmit data in the space mission IP network.

31.2.3 Analysis Conclusions

All kinds of factors mentioned above could cause congestion which will influence the QoS. To avoid congestion effectively, systematic technical strategy should be implemented. As it should be, a differential service strategy should be implemented in the network to make sure that important information can be transmitted reliably with priority and that the QoS is guaranteed even when congestion occurs.

Therefore, the QoS strategy of the space mission IP network should be a systematic solution that first tries to avoid congestion and guarantee the transmission of the information with high priority when congestion occurs.

31.3 Top Level Design of the QoS Strategy

31.3.1 Architecture of the QoS Strategy

As illustrated in Fig. 31.3, the architecture of the space mission IP network QoS strategy can be summarized as a congestion avoidance mechanism based on access traffic management and output traffic control plus a differential service mechanism based on traffic classification and priority queue dispatch. The congestion avoidance mechanism includes a set of strategies such as traffic policing, traffic shaping, port rate limit, and so on. The differential service mechanism includes a set of strategies such as service classification, priority marking, priority trust, and priority queue dispatch.

31.3.2 Classification of Service Priority

According to different requirements, the service traffic is classified into six priority classes with the TOS [5] value of 5, 4, 3, 2, 1, and 0 respectively. Class 5 has the highest priority and class 0 the lowest. Class 1 is reserved. The DSCP [6–8] values are assigned correspondingly.

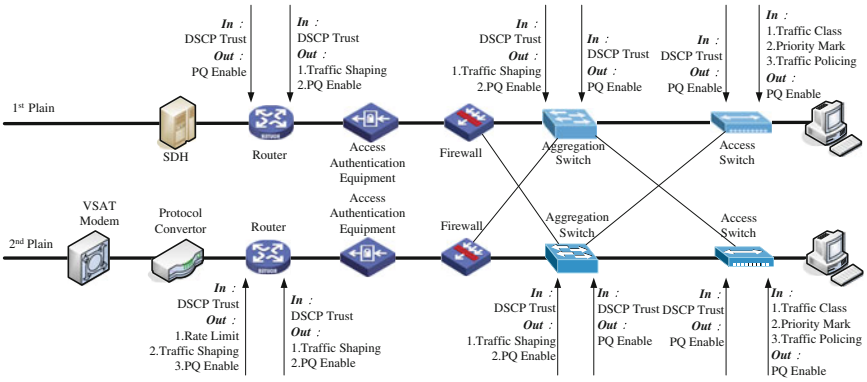

Fig. 31.3 Architecture of QoS strategy in space mission IP network

31.3.3 Disposition of QoS Strategy

The concrete disposition of the QoS strategy in the space mission IP network is left out due to the volume limit of the paper.

31.3.4 Testing Method of QoS Strategy

It is necessary to carry out a test to check the validity and effectiveness of the QoS strategy before the space mission.

31.3.4.1 Effectiveness Test of Access Traffic Policing

A network tester or testing terminal installed testing software is used at an access switch to send normal service traffic and excessive service traffic up to 120 % of the policed traffic. At the receiving end a network tester or testing terminal installed testing software is used to check the received traffic. The packet loss rate of the normal service traffic should be less than 0.1 % and that of the excessive traffic should be more than 0.1%.

31.3.4.2 Effectiveness Test of Priority Class Marking

A testing traffic is send by the user terminal or a testing terminal and a packet capturing software is used to check whether the marking result of the priority class is coordinate with the strategy.

31.3.4.3 Effectiveness Test of Priority Queue Dispatch

A network tester is used to send background traffic with the priority 0 and a user terminal or testing terminal installed testing software is used to send normal service traffic with higher priority. The total traffic exceeds the throughput of the WAN bandwidth. The receiving user terminal or testing terminal checks the normal service traffic with higher priority. The packet loss of the normal service traffic should be less than 0.1 %.

31.3.4.4 Effectiveness Test of Port Rate Limit

This test is implemented together with that of priority queue dispatch simultaneously. If the packet loss of the normal service traffic with higher priority is less than 0.1 %, then the strategy of port rate limit is working effectively. Otherwise, if the packet loss of the normal service traffic with higher priority equals that of the background traffic with the priority 0, then the strategy of port rate limit is not working effectively.

31.4 Conclusions

It is a complex systematic engineering task to design the QoS strategy of the space mission IP network. All kinds of factors should be taken into account such as the capacity of all the equipments from end to end, the complexity of implementation, and difficulty of management. As a complete and specific strategy, the QoS strategy discussed in this paper has been verified in many space missions including Tiangong-1, Shenzhou VIII, and Shenzhou IX missions. On the other hand, due to the performance of the equipment, there still remain some flaws in the actual effectiveness of the QoS strategy. For example, it is hard to implement an exact traffic policing due to large granularity of the access traffic policing in the switches. It is hard to achieve exact matching between the port rate limit and the VSAT circuit rate due to the difference between the Ethernet frame and HDLC frame. These questions mentioned above make it hard to setup the WAN circuit exactly and efficiency of bandwidth is decreased because quite large bandwidth margin has to be reserved.

It is a process of improvement and perfection to develop the QoS technology and the implement of the QoS strategy in the space mission IP network. The radical goal of the perfection is to increase the efficiency of the bandwidth when guarantee the QoS requirement such as packet loss rate, delay, jitter. The following aspects should be taken into consideration on the next generation space mission IP network:

1. To increase the process capacity of the security equipments to increase the thorough throughput of the IP network;
2. Addition IP supporting mechanism to the VSAT modem and implement some QoS function at the Ethernet port;
3. To develop the test and management technology of the QoS strategy and achieve the goal of central management of the QoS strategy;
4. To study the next generation QoS technology and achieve knowability and controllability in the next generation space mission IP network.

References

1. Braden R, Clark D, Shenker S (1994) Integrated services in the Internet architecture. IETF RFC 1633
2. Blake S, Black D, Carison M et al (1998) An architecture for differentiated services. IETF RFC 2475
3. Floyd S (2000) Congestion control principles. IETF RFC 2914
4. Allman M (1999) TCP congestion control. IETF RFC 2581
5. Postel J (1981) Internet Protocol. IETF RFC 791
6. Nichols K, Blake S, Baker F et al (1998) Definition of the differentiated services field (DS Field) in the IPv4 and IPv6 headers. IETF RFC 2474
7. Jacobson V, Nichols K, Poduri K (1999) An expedited forwarding PHB. IETF RFC 2598
8. Heinanen J, Baker F, Weiss W et al (1999) Assured forwarding PHB group. IETF RFC 2597

Chapter 32
System Level Design of Address Allocation for a Private IP Network

Yalin Huang, Zongyin Zhao, Yan Liu and Xu Yao

Abstract IP address allocation is very important to data routing, inter-networking and OAM of large IP networks. In this paper, the high level architecture of a private IP network is presented firstly. Then its uniqueness is summarized as large scale multilayer network architectures, independent dual transferring routing of WAN, dual transferring planes of MAN/LAN, support of special requirement such as multicast and external networking. Various address allocation requirements from the above mentioned features are analyzed. Consequently, address spaces are divided according to the network type, routing, devices, and special application requirements. After that, each address space and its future evolution path is presented in details. Finally high level design and analysis of each allocation scheme is presented in sections according to the sequence of MAN, LAN, WAN, multicast and external networking. Field deployment result in the past 2 years has proved the success of the proposed address allocation plan.

Keywords IP network IP address allocation Multicast External networking

Y. Huang (✉) · Z. Zhao · Y. Liu · X. Yao
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: yalin0223@sina.com

X. Yao
School of Computer and Communications Engineering,
University of Science and Technology Beijing, Beijing 100083, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_32,

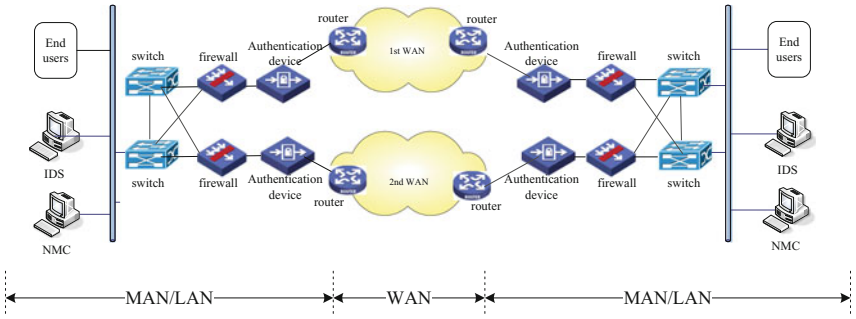

Fig. 32.1 High level diagram of PIP

32.1 Introduction

IP address is used to annotate host in TCP/IP network. A sound IP address allocation scheme is very important to uniformed addressing, network accessibility as well as enhanced maintenance and network management capability.

A large private IP network (PIP in short hereafter) has been established to support multiple services in parallel. It consists of WAN/MAN/LAN with many network nodes and has many unique features compared with public Internet, such as wide geographic coverage, large number of network nodes, two independent WAN routes, support of multicast and networking with external IP network. Address allocation for it needs to resolve several technical obstacles.

In the following sections an address allocation scheme is presented based on analysis of PIP architecture and its uniqueness.

32.2 Architecture of PIP and Its Features

32.2.1 High Level Architecture

As shown is Fig. 32.1, a PIP consists of WAN, MAN and LAN, security devices (authentication device, firewall, intrusion detection system), network management system.

WAN consists of two physically separated networks. One is a WAN mainly includes "routers + 1st routing links", namely 1st WAN. The other consists of "routers + 2nd routing links", namely 2nd WAN.

Generally MAN is deployed in large organization, whereas LAN is deployed in subordinate units far from central office. Three-layer (core-convergence-access) architecture is deployed for MAN while only two layers or one layer architecture is used for LAN.

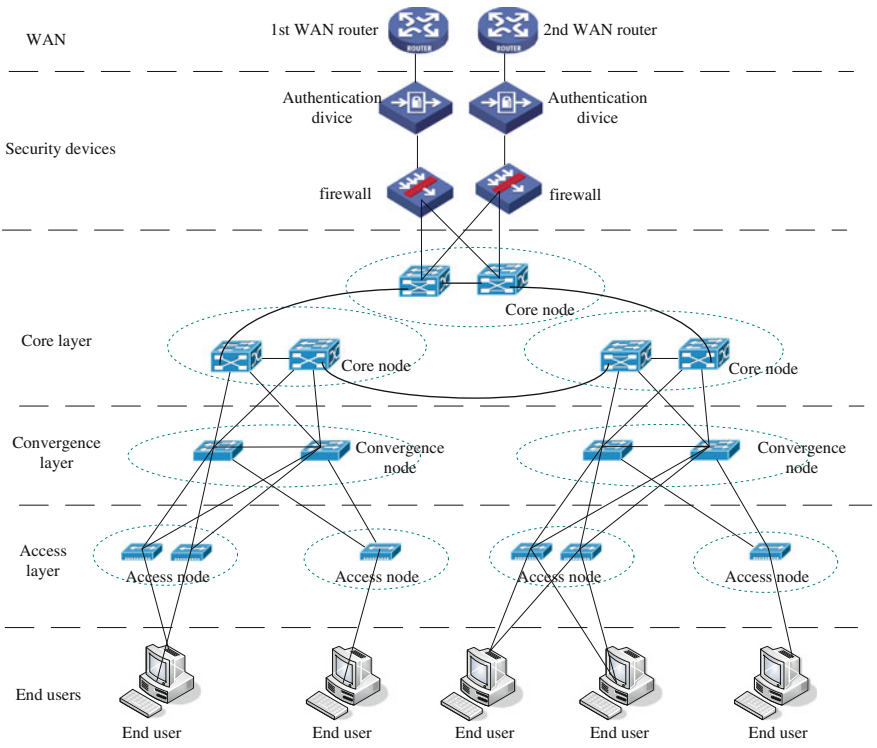

Fig. 32.2 WAN/MAN/LAN architecture

Normally, each MAN/LAN has networking links with 1st WAN and 2nd WAN separately Fig. 32.2.

32.2.2 The Features of PIP

32.2.2.1 Wide Network Coverage and High Scalability

The numbers of LAN, users, network type are quite large in PIP. This number grows consistently. Therefore, high network scalability is essential for IP address allocation plan.

32.2.2.2 Independent Dual Routing of WAN

To meet data transmission reliability requirement, important user data packets must be routed simultaneously via 1st and 2nd WAN. Consequently, important

users should have 2 network interfaces. Each of them has a different IP address belonging to a separate IP subnet. In unicast case, when sending data to different destination addresses, data packets can be routed in different WANs. In multicast case, sending data from different source addresses will force multicast data packets be routed via different WANs. Therefore, 2 user address spaces are needed, corresponding to the 1st and the 2nd WAN respectively.

32.2.2.3 Dual Transferring Planes

In MAN and LAN, dual-switch is deployed at all core/convergence nodes and at most access nodes. 2 network interfaces for each important user can be linked with 2 switches in one node separately. Normally, 2 switches in one node will relay data according to user IP address and back up each other in case of device failure. Thus, a MAN (LAN) can be conceived as 2 MANs (LANs) connecting with 1st and 2nd WANs separately, and network switches in convergence or above layers shall be capable of data relaying for 2 user address spaces simultaneously.

32.2.2.4 Layer 3 Routing Between Switches

Given wide coverage and various user types, network switches resource may be wasted to handle unknown traffic without reasonable VLAN scope definition [1]. To reduce broadcast domain, network switches in PIP are connected at layer 3 in routing mode and VLAN is designated at access switches. Consequently, IP address must be allocated for internetworking interface of switches, and it must not belong to VLAN for user address.

32.2.2.5 Security Devices Between WAN and MAN/LAN

Between WAN and MAN/LAN, there are security devices like authentication device and firewall, which work at layer 2 and connect with a WAN router and dual transferring planes of MAN (LAN). Device management address needs to be allocated for security devices and it must belong to same VLAN that covers router LAN interface and up-link port of network switch.

32.2.2.6 Network Management Requirement

PIP management center authorizes management terminals for network management of corresponding management areas. Network management terminal is responsible for configuration, maintenance, performance and failure management for all network devices within its management area. So device management address must be allocated for all networking devices.

32.2.2.7 Special Application Level Requirements

PIP needs to support some special application level requirement, like multicast and external networking. Multicast is implemented mainly by Source Specific Multicast (SSM for short) while in special case, Any Source Multicast (ASM for short) is also needed. As a result, group addresses must be allocated in advance. External networking has many scenarios, such as level I domestic networking, level II domestic networking, and international networking. Consequently user address, network device address and interface address must be allocated accordingly.

32.3 High Level Allocation of Address Space

According to PIP architecture and its characteristics, it can be treated as two IP connected networks. Therefore, 2 non-overlapping independent address spaces, 1st address space and 2nd address space, are needed for these 2 networks. Data sending to user at 1st space will be routed via 1st WAN and data sending to user at 2nd space will be routed via 2nd WAN.

The 1st and 2nd address space can be further divided into WAN address space, MAN address space and LAN address space. These 3 address spaces are not overlapped also. Based on current requirement and future upgrade requirement in planned time frame, planned MAN number is around 30 and planned LAN number is around 200.

Moreover, group address must use fixed Class-D address (First 4 bits of first octet is 1110) [2], so WAN/MAN/LAN address must not use same address space allocated for multicast.

To meet above mentioned requirement, 8 bits of first octet can not distinguish address types needed. So, first octet of address is used to identify network type and its corresponding sub-space (WAN, MAN, LAN, multicast, external networking).

First octet of IP address is defined as below:

[000]: 1st WAN
[010]: 2nd WAN
[100]: 1st sub-space for MAN/LAN. Each MAN is designated one unique value; a fixed value of first octet is reserved for all LANs. Second octet is used to further identify a specific LAN
[110]: 2nd sub-space for MAN/LAN. Each MAN is designated one unique value; a fixed value of first octet is reserved for all LANs. Second octet is used to further identify a specific LAN
[1110]: Multicast.

32.4 Address Allocation for MAN

The number of MAN is around 30, so last 5 bits of first octet can be used to label different MAN. Following octets can be used to further label user address, network device address and interface address. Because connections between switching nodes are all layer 3 connections, these 3 types of address are closed coupled with network topology. Moreover, all network switching nodes need to be distinguished by unique associated number.

Taking architecture of MAN as example, generally, within a MAN, there are less than 10 core switching nodes. Beneath one core switching node, there are less than 10 convergence nodes. Beneath one core (or convergence) node, there are less than 30 access nodes. Thus within a MAN, there are less than 10 core nodes, 100 convergence node and 3300 access nodes. Obviously, within a MAN, one octet cannot distinguish all switching nodes. Hence core nodes and convergence nodes can be numbered uniformly within a MAN, and access nodes need to be numbered under corresponding convergence node.

32.4.1 MAN User Address

To specify a user address, user location must be specified by what core/convergence/access node a user belongs to. So, meaning of each octet must be decided according to network layered architecture. To effectively indicate user position, the 1st octet is used to label associated MAN, the 2nd octet is used to label associated core or convergence node and 3rd octet is used to label associated access node. 4th octet is host id. To reduce broadcast domain at layer 2, multiple VLANs can be further defined according to service and usage type [3].

32.4.2 MAN Network Device Address

To specify address for a network device, its relation to a specific network node needs to be specified. Given this, similar to user address allocation case, meaning of each octet needs to be defined according to network architecture.

The 1st octet is used to label MAN a network device belongs to. The 2nd octet is used to distinguish from end user address space and also to ease network maintenance. Therefore, to select all network devices in a MAN, a 16 bit mask can be used to separate network devices and users. The 3rd octet is used to label associated core or convergence node. The 4th octet is used to label a specific network device. As each network node can consist of 2 devices backing up for each other, 2 address spaces must be reserved for each node.

32.4.3 MAN Interface Address

To specify interface address for a network device, link used must be specified. Because there are many links within a MAN, if using switch nodes ID at 2 ends to identify a link, it is impossible to find a suitable address space within a MAN address space. Therefore, links between switches are numbered together within a MAN. Numbering scheme is: Link group ID + link ID. Link group ID represents various types of links between network devices, while link ID labels a specific link within a link group. Therefore, it is enough to use two octets to label link group ID and link ID.

32.5 Address Allocation of LAN

The 1st octet of a LAN address is a fixed value. The 2nd octet is used to label LAN ID. To specify a MAN/LAN, all LANs are numbered after MAN, range is 32–255. If number of LANs is larger than 255, the 8th bit of the 1st octet will carried 1. Following same approach of MAN, user address, LAN network device address as well as LAN interface address can be allocated similarly.

Similar to MAN user address, LAN user address is specified according to its associated switching node. As switching node id will not be larger than 255, so one octet is used to number all networking nodes. Multiple VLANs can be defined to reduce broadcast domain.

After the 1st and the 2nd octets are specified, the 3rd octet is to distinguish network devices from user. The 4th octet is to label a specific network switch.

Similar to MAN interface address allocation, LAN interface address is also defined as: Link group ID + Link ID.

32.6 WAN Address Allocation

There are 3 scenarios of WAN connection: 1 is dedicated point to point link, 2 is connected via layer 2 network, 3 is virtual tunnel connection.

32.6.1 Point to Point Link Case

When WAN routers are connected via point to point link, address allocated must specify associated 2 WAN sites. As only one octet can distinguish all MAN/LAN in use, the 2nd and the 3rd octet can be used to label 2 sites connected by point to point link respectively. The 4th octet is used to label interface at two ends of link.

When LAN id is larger than 255, the 7th and the 8th bits of the 1st octet can be used as carrying bit for the 2nd and the 3rd octet, which means the 7th and the 8th bit of the 1st octet can be combined with 8 bits of the 2nd or the 3rd octet to form a 9 bit number to identify LAN ID.

32.6.2 Connected via Layer 2 Network Case

When WAN routers are connected via layer 2 network, to distinguish from point to point case, the 2nd octet is a fixed value. The 3rd octet is layer network id. To avoid unexpected broadcast packet due to unknown layer connection, layer 2 network within WAN needs to be numbered together within range of 1–255. The 4th octet is router id and is used to label router within layer 2 network.

32.6.3 Tunnel Connection Case

If WAN routers are connected via tunnel across public network or other networks, two types of point to point address must be allocated. One type is tunnel address, i.e., virtual point to point link address. Its allocation can follow approached depicted in 6.1. The other type is connection address from WAN router to public network router, i.e., actual working address with local meaning only as long as it will not conflict with local router address.

32.7 Group Address Allocation

Multicast includes 2 cases, ASM and SSM. Group address is used to group multiple hosts that need to receive same information and is not coupled with network routing, which is closely coupled with application systems. Therefore, group addresses can be assigned by application systems as long as they are not conflicted.

32.7.1 SSM Address

The 1st octet of SSM address is fixed as 232 [4]. As multicast depends on application system heavily, the 2nd octet is used to distinguish different specific multicast applications. Considering it is impossible to control multicast data stream, to implement multicast control policy on security devices, address space is

allocated by several application modes, such as multicast within sub-network, cross sub-network multicast within field, inter-field multicast.

32.7.2 ASM Address

Similar to specified source group SSM address allocation, any source group address is allocated according to multicast source function and multicast group requirement.

In this case, 225 is used for cross-field multicast and 226,227 are used for cross sub-network multicast within a field. 228,229 are used for multicast within a sub-network [5]. The 2nd and the 3rd octet represent application ID and field ID respectively. The 4th octet is defined by each application system.

32.8 Inter-Networking Address

PIP is required to be capable of inter-networking with external network. External peer network can be domestic I or II, or international network.

1st octet of interface address is a fixed value. So in remaining 3 octets, information of networking type, address type (WAN, LAN), routing type (1st route or 2nd route), field ID(local field ID and peer field ID) must be included. Obviously, it is impossible to allocate one octet for each information type. So one octet is assigned as inter-networking field ID and remaining 2 octets are used to label networking, address, routing type and host ID.

32.9 Conclusion

IP address of PIP has been allocated following the above mentioned paradigm. In 2011 and 2012, multiple projects have been successfully supported by PIP. Field deployment result has proved this address allocation paradigm, based on Various Length Subnet Mask techniques, meets various requirements of address allocation, and has reduced the size of routing table significantly. Moreover, it also facilitates easy maintenance, configuration and ensures clarity and simplicity of address allocation.

References

1. Webb K, Froom R et al (2003) Building Cisco multilayer switched network, 4th edn. Pearson Education
2. Postel J (1981) Internet protocol. RFC 791
3. Pummill T, Manning B (1995) Variable length subnet table for IPv4. RFC 1878
4. Holbrook H, Cain B (2006) Source-specific multicast for IP. RFC 4607
5. Deering SE (1989) Host extensions for IP multicasting. RFC 1112

Chapter 33
Research on the QoS Guaranteed Mechanism for the Private IP Network

Lihua Liu, Tun Wu, Zongyin Zhao and Qian Zhang

Abstract In the private network, end-to-end QoS plays very important role for the real-time trial data transmission. Considering the features of the network and the respective QoS requirements, the general principle of QoS deployment is studied in this paper. Following the principles, the model of end-to-end QoS used in the private network is established firstly. Then it is analyzed and demonstrated that how to deploy the QoS strategy in the kernel layer, cluster layer, and access layer of the private network, respectively. Furthermore, the solution is also designed to accomplish the QoS for the burst data and in other special conditions. Finally, the all-around scheme of end-to-end QoS is proposed for the private network, which provided the theoretical basis and guidance for bringing the QoS into effect. The proposed mechanism could be a valuable reference of significant practical meaning for other private IP networks.

Keywords The private IP network · End-to-end QoS · Service model · Strategy

33.1 Introduction

Following the development of the trial private network and IP services, the trial information such as the real-time data, dispatch, image, timing signal and other trial data has been carried with the private network based on the IP protocol. However, IP networks used to be connectionless, information is transmitted in the

L. Liu (✉) · T. Wu · Z. Zhao · Q. Zhang
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: ericabao@163.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_33,

way called best-effort. It is essential to design a scheme of QoS for reliable links of real-time services. Besides, different kind of services contained in the private network, such as real-time data, voice, dispatch, image and other services and also trial documents, meet different real-time and reliability level. Above all, it is necessary to analyze and discuss the QoS scheme of the IP private networks to guarantee the reliability of the links of end-to-end services.

33.2 QoS Scheme for End-to-End Services

The QoS scheme relies on a specific network structure, especially for the private IP network. Consequently, an appropriate end-to-end QoS model is needed to characterize. Considering the differences of kinds of services and devices of each layer of the private network, various QoS methods should be applied in each layer.

When the QoS scheme is applied for the private IP network, the following aspects is needed to be considered:

1. QoS scheme for end-to-end services, including deploying strategy in the kernel layer, cluster layer and access layer, respectively. There is correlation between the strategies in different layer.
2. Different QoS mechanism should be configured for different user and service.
3. Strategies to improve the redundancy of the private network should be applied, such as backup storages for some network nodes.
4. The scheme could not affect the extending of the private network and the services.

33.3 QoS Models for the Private Network

A QoS model is an integrated scheme of many kinds of QoS methods. Recently, the most familiar QoS models include the Best-Effort model, the Integrated model (called IntServ for short), and the Differentiated model (called DiffServ for short) [1].

The private network discussed in this paper could provide the services of real-time data, dispatch, image, timing signal, telephone, network management, documents and so on. As the specialty of each service, different level of reliability for each kind of service should be considered, when an appropriate QoS model is selected.

The QoS scheme of basic IP links is characterized as the Best-Effort model. Whether or not the IP packages have reached the end and the time it takes are not guaranteed. So the private network could not be characterized by the Best-Effort model.

The IntServ model takes the methods that the networks ensure resources required by the services, so the quality of some specific services could be reached. The IntServ model could be used to model the private network and it could reach

the QoS demands. However, the services of the IntServ model are configured by using RSVP signaling, and the resources required should be reserved by every network device and node. For the networks contain large amount of data volume, it is difficult to ensure the resources for each kind of services: the increment of the QoS would increase the load of the networks; if the resources of the networks are limited, the configuration would never be accomplished [2]. As to the private network considered in this paper, resources required by some kinds of services could be estimated or calculated, the methods of resource reservation could be used.

The strategies which the DiffServ model applies are that different QoS method is used for different kinds of services. For the devices in the networks, it only needs to identify the QoS level of each IP package, but recording the information of the links of the services. The classes of QoS level should be instituted to ensure the different QoS demands, and the packages which mark as high QoS level would have the priority of transmission, especially for the situation when the network congestion happens. However, in the private network, QoS should be guaranteed for every kind of services, especially for the real-time service. And the different classes of QoS level should be applied only the situations which the networks have bad performances. As to the private network, the resources of the networks and the requirements of actual services should be estimated, because the bandwidth of some links (especial for the satellite communication links) is limited.

The QoS model of the private network could be designed as follows.

1. Resource Reservation. As the description of the IntServ model, in the private network, the resources required by the services could be estimated, such as the bandwidth of each transmission channel. Performances of the private network, such as the transmission efficiency, could be improved by reducing the load of the networks. Generally, 50 % of the total bandwidth is designed for the metropolitan area network and the wide area network would take the number of 70 %.
2. Differentiated Services. In the private network, the classes of QoS level could be instituted to differentiate different kinds of services. For example, classify and marking the priority for the service on the access switch, then take the technology of CAR, buffer mechanism and queue schedule to deal with the priority data on the following network equipment to achieve the goal of the differentiated QoS. The application of the QoS methods in the private network will be discussed in Sect. 33.4.

33.4 Implementations of the QoS Mechanism for the Private Network

As mentioned in Sect. 33.2, in order to differentiate different kinds of services of the private network, first of all, the classes of QoS level should be instituted to identify the IP packages of real-time data, dispatch, image, timing signal,

telephone, documents and so on. High QoS level should be configured for the services needed to be transmitted in the first place. Methods could be applied in every device of the private network to achieve the goal of the differentiated QoS, such as classification by priority field, flow CAR, queue schedule.

As a manner of multi-layers of the private network, QoS strategies should be configured in each layer of the networks.

33.4.1 Access Layer

The access layer of the private network provides the accesses of a mount of services, data, voice, images and so on. The access type of each service could be different from others. The methods which classify kinds of services and institute the classes of different QoS level of services should be applied in this layer, and the institution could be configured in every node of the networks. The load of the networks could be alleviated, too. Generally, the services could be divided into five QoS levels, namely, EF, AF41, AF21 and BE, marked as 5, 4, 3, 2 and 0, respectively. The maker 1 is a reserved QoS level.

Besides, in order to avoid the network congestion, the data rate of every user of the private network should be at a configured level, and the user should not be allowed to have useless operations.

33.4.2 Cluster Layer

The cluster layer commonly locates at the area where kinds of the services should be provided. In this layer, services would be integrated and configured for transmission, and the resources of the networks would be managed in this layer. The cluster layer could use queue schedule technology according the QoS level configured in the access layer, namely, the services of high QoS level and the packages of special users have the priority of transmission.

Two most familiar queue schedule algorithms are the Strict-Priority algorithms (SP) and the Weighted Round Robin algorithms (WRR). As to the private network discussed in this paper, real-time services which need reliable transmission should be configured as high QoS level and have the priority of transmission. In WRR algorithm, resources of bandwidth would be reserved for each service, the services with high QoS levels could not be transmitted firstly, if the network congestion happens, the services would be affected. Therefore, in the private network, the SP algorithm should be applied to ensure the priority of the services of high QoS level, i.e. the services with higher QoS level would have the priority of transmission and those with lower QoS level would be treated later.

Besides, as to the users and terminals straight connect to the cluster layer, the services should be classified and configured the QoS levels.

33.4.3 Kernel Layer

The scalar of the kernel layer is much smaller, but the throughput of the kernel layer is high and the performance of the kernel layer plays an important role of the whole private networks. The QoS level of the services should be used to guarantee the QoS. However, the firewall and encryption equipment which connect the layer could not be configured the QoS mechanism for the private network. Limitation of output data rate should be applied in this layer to avoid the congestions in the firewall and encryption equipment. As to the users and terminals straight connect to the kernel layer, the services should be classified and configured the QoS levels.

33.4.4 Routers

Routers in the private network connect subnets of different area. For the QoS mechanism, the services should be classified and configured the QoS levels in the routers. Besides, considering the resources of the private network, output data rate of the routers should be configured to reduce the consumption of the resources, for example, bandwidth of the channels.

33.4.5 Implementations of the QoS Mechanism for Some Special Situations

33.4.5.1 QoS Mechanisms for the Burst Data

In the private network discussed in this paper, some measure equipments could generate the IP packages of burst data. In this situation, limitation of data rate of the burst data would increase the loss rate of the IP packages, and the storage of the burst data could cause delay and jitter in the private network. Considering the local subnets of the private network used to have the bandwidth of kbit/s, the speed limitation of the burst data should not be configured to access the kernel layer, but it is necessary to configure in the output of the switch of the kernel layer in order to match the speed of the firewall and encryption equipment. The output of the routers in the private network should math the speed of the wide area network to guarantee the demands of the QoS.

33.4.5.2 Limit Rate

In the private IP network, the wide area network is consisted of satellite communication subnets and ground fiber subnets [3]. Data rate of different

transmission link could be different and limited, especially to the satellite communication links. Besides, configurations of the QoS mechanism for the private network would be different for each device. So, the limitation of the total data rate of the networks should be configured when network congestion happens, and the loss rate of IP packages of high QoS level could be minimized. The limitation of the data rate of the following transmission links and devices should be considered.

1. The transform equipment of protocol used in the satellite communication links, which is applied for exchange the protocols from the Ethernet interface to the RS530 or V.35.
2. The optical transceiver and the HDSL device used for Ethernet extending.
3. The output interface of the kernel switch which connects to the router should be configured the limitation of the data rate, considering the performances of firewall and encryption equipment in the links.

33.4.5.3 Security Device

In the private network, security and encrypted information are used to transmit kinds of service. So, firewall and encryption equipment are implemented in the links between the switches of the kernel layer and the routers. The firewall could have the authority to access the private network, and prevent useless services to improve the performance of the QoS mechanism. However, QoS methods could not be applied in most encryption equipments.

Above all, the implementations of the QoS mechanism for the private network are depicted in Fig. 33.1.

1. The services should be classified in the access switches according the five elements of each service: source IP address, destination IP address, source port number, destination port number, and transmission protocol.
2. According to the demands of the QoS mechanism, the services could be divided into five QoS levels, namely, EF, AF41, AF21, and BE, marked as 5, 4, 3, 2 and 0, respectively.
3. The CAR is applied for the services marked by 5, 4, 3 and 2, flow control is not used for the services marked by 0.
4. The classes of the QoS level should be instituted and configured in every interface of the devices (switches and routers).
5. The PQ schedule should be configured in every interface of the devices.
6. Limitations of data rate should be estimated and configured in every interface of the switches and routers, considering the resources of each transmission link.

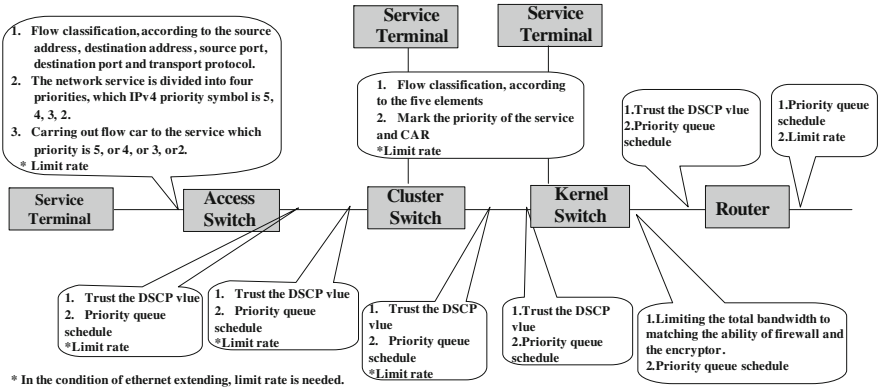

Fig. 33.1 The scheme of end-to-end QoS for the private network

33.5 Other Suggestions

As the complexity of the private IP network, there are many other factors not mentioned in this paper could also have influences on the performances of the QoS mechanism. So, the following suggestions should also be considered.

1. To the implementations of the QoS mechanism, QoS methods should be applied to all devices in the private network, such as firewall and encryption equipment. Besides, researches of the QoS methods used for the networks which have limited resources, especially for the satellite transmission links of WANs.
2. To the network management, the services and the flows should be monitored, and the data rate of every user of the private network should be at a configured level, and the user should not be allowed to have useless operations, such as Ping.

33.6 Conclusions

The QoS mechanism designed and discussed in this paper could fulfill the demands of the end-to-end QoS for the private IP network, and the implementations of the QoS mechanism in each layer, are also presented. The mechanism could be an example of significant practical meaning for other private IP networks.

References

1. Firoiu V, Boudec L, Towsley D (2002) Theories and models for internet quality of service. In: Proceedings of the IEEE, Minnesota
2. Li W (2002) Differentiated service based on IntServ and DiffServ. Telecom World
3. Zhao ZY (2008) Research on the framework and critical technology for the development of the trial private network. Beijing institute of tracking and telecommunication technology, China

Chapter 34
Distributed Data Service Platform Based on Narrowband Network Environment

Xu Yan, Guoping Hu and Dahai Zhai

Abstract A distributed data service platform is established mainly to solve the problem of data sharing among several data centers in narrowband WAN environment, provide continuously reliable and position-independent data services for the users and realize the objective of transmitting right information to right customer in right way (3R). The multi-center data service platform optimizes data consistency algorithm against WAN features (small bandwidth and long delay) based on network technology, distributed technology and parallel computing technology, using peer-to-peer cluster-based network system framework, combined with "Cloud storage" framework features, and provides such characteristic functions as strategy-based oriented sharing, priority control and network flow control as well as capabilities of synchronously sharing file data and database data facing overall situation and subject to unified management. The global distributed data redundancy storage also provides the users with new data backup for disasters; intelligent troubleshooting and inter-node high self-fusion features can provide efficient data fault-tolerant and data service recovery capabilities in case of fault of network or other hardware as well as continuous and non-stop data services for the users. System joint debugging in actual narrowband WAN operation environment shows that the system can fully meet the engineering demands in synchronous data sharing, system stability etc., and has good promoting value.

Keywords Distributed · Peer-to-peer network · Data sharing · Data storage · Data service

X. Yan (✉) · G. Hu
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: yx68909@163.com

D. Zhai
The Fifth Research Institute of Telecommunication Technology, Chengdu 610062, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_34,

34.1 Introduction

The current society has entered information age and the computation and storage capabilities of traditional central computer systems cannot meet the people's demands. The requirements of future information systems for computing speed, system reliability and system cost effectiveness make it necessary to develop distributed systems, of which the demands and push factors lie in:

High cost-effectiveness	The parallelism of the distributed system reduces the processing bottlenecks. Compared with large computers with the same processing efficiency, the distributed computer system consisting of several general computers is much cheaper;
Inherent application	The system itself is of distributed type and various applications on it require mutual coordination and data sharing;
High reliability	Redundancy is necessary for biological evolution, so is for information technology. The distributed system has potential to run continuously in case of any fault;
Scalability	The distributed system is easy to be expanded to include more resources;
Data sharing	The distributed system can effectively support users from different places to share information and resources (software and hardware).

Since 1990s, the distributed systems with processing of file data as the core have been developed greatly, but the distributed systems processing database data are mostly realized based on LAN, and the WAN-based distributed systems with database as the core are rarely applied in reality. The reason for that mainly lies in uncertainty of WAN connection and time delay, which makes it difficult to satisfy the strict requirements for transactional consistency and safety of the database system.

A distributed system - multi-center data service platform based on narrowband WAN environment, with distributed data storage and distributed data sharing services as the core is introduced in this paper.

34.2 Introduction to the Distributed System

The distributed system is an integration of several independent computers which are single systems for the users [1]. The main objective is to facilitate the users to access or share with other users the resources, and meanwhile the system resources must be within the controlled scope. In term of design realization, the distributed

system mainly covers the distributed system management, the distributed data management and the distributed service management.

The distributed system management mainly includes network topology, communication protocol, route algorithm, etc. The network topology is the basic framework of the distributed system and it determines the system availability and scale. Since the features of WAN (long time delay and instability) make it more difficult to design the network topology, communication protocol, etc. of the WAN-based distributed system, for example, all video systems mainly using P2P technology adopt loose coupling (or discrete) framework strategy, but the strategy also causes such disadvantages as no manageability of the system.

The distributed data management mainly includes data redundancy management, data consistency protocol, fault-tolerant algorithm, etc., aiming at providing a systematic method to execute concurrent data transaction in serial mode. The data consistency protocol is the hottest topic till now. Traditional method has concurrent control protocols based on lock, time stamp, etc. These protocols can guarantee data consistency among several copies in theory, but they raise extremely high requirements for the network environment, and the network time delay, network stability, etc. have a direct bearing on the execution efficiency of above concurrent control protocols.

The distributed service management covers several sub-items such as intelligent scheduling and load balancing. The distributed system provides huge processing capacity. However, good resource scheduling method is required for the utilization of such capacity. The scheduling model and load balancing are closely related to each other and shall be designed according to specific application demand. Generally, no high load balancing is required in WAN environment. Nearby service principle is adopted for the data within the service capacity scope and the service requests exceeding the service scope will be scheduled remotely.

34.3 Realization Principles

34.3.1 Distributed Storage System Framework

The distributed data storage system is the core component of the multi-center data service platform, using peer-to-peer cluster-based network system framework, as shown in Fig. 34.1. Clustering means grouping the servers according to regions, the servers within the same region form a service node via LAN and the service nodes are connected via WAN. Peer-to-peer means that the service nodes have equal status without any control center or resource center and they form a whole via networking to provide services for the service applicants in a unified manner.

Generally, each node consists of several servers responsible for synchronous sharing service of data. Data are stored in the system in the form of copy and the quantity of copies may be set flexibly, subject to various copy redundancy rules

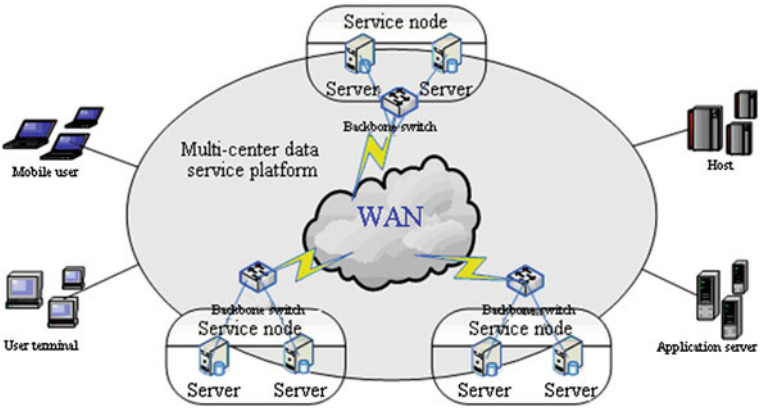

Fig. 34.1 Peer-to-peer cluster-based network system structure

including partial redundancy and full redundancy. Data completeness and consistency shall be kept among all copies to realize synchronous data sharing.

34.3.2 Network Topology Management

In LAN environment, the network fault rate is very low and the detection results may be obtained at ms level, so the network topology management of the distributed system based on the network environment is very sensitive and meanwhile the data consistency is completely applicable to the realization of such data consistency protocols as traditional 2PL (2-phase lock protocol) and 2PC (2-phase commit protocol). However, it is completely different in narrowband WAN environment, and the features of small bandwidth, long time delay and instability increase the difficulties of network topology management and data consistency guarantee.

For the features of narrowband WAN, slow processing is conducted on the network topology management to reduce the network detection frequency and increase the tolerance for network data loss and no response to peer-to-peer network. The results of such method are the reduction of timeliness of network fault detection, but very stable operation of the system in poor network environment, no frequent network troubleshooting and reduction of additional expenses of the system.

Ring test is adopted as the network fault detection algorithm: The servers in the node form a logic ring, adjacent servers in the ring conduct detection, in case of any network fault or any other type of fault, other servers in the node will be informed and troubleshooting will be conducted; the nodes also form a logic ring, and the inter-node detection method is the same as server detection in node. Ring test can reduce network communication expenses and save system resources.

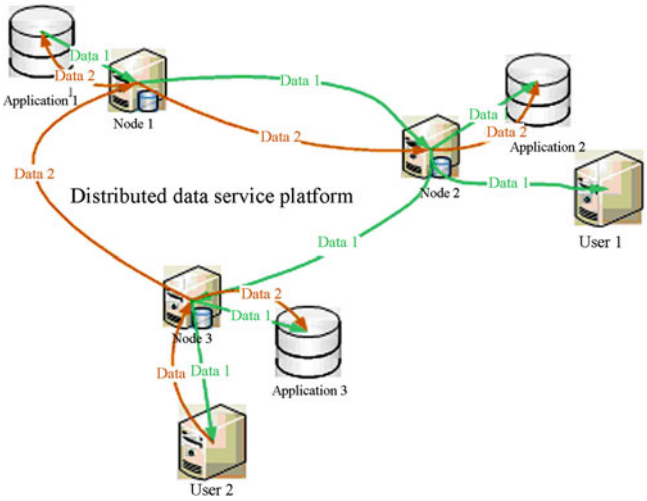

Fig. 34.2 Abstract data flow diagram of timely consistency protocol

34.3.3 Data Consistency Protocol

In order to guarantee atomicity and consistency of transaction, the distributed database system must guarantee that all participated distributed transactions are committed or no one is committed. However, such guarantee is realized by the commit protocol of distributed transactions [1]. Generally, basic 2-phase lock protocol (2PL) and basic 2-phase commit protocol (2PC) are used to guarantee data consistency.

For features of WAN (small bandwidth, long time delay and instability), the multi-center data service platform provides two data consistency protocols: real-time consistency protocol and timely consistency protocol, and different protocols may be selected for different businesses according to their demands to realize data sharing. Real-time consistency protocol refers to that any operation of sharing data must meet atomicity feature of transactions and any modification will be executed at the same time on all data copies by means of using basic 2-phase lock protocol and basic 2-phase commit protocol during transaction, and will be mainly used for the businesses with very high requirements for real-time and completeness. For the applications without high requirements for real-time data, timely consistency protocol will be used. Under timely consistency protocol, the user's modification to the system data will not be immediately kept consistent among all data copies and finally the data of all copies will be kept consistent after a "gradual diffusing" process, as shown in Fig. 34.2.

The timely consistency protocol algorithm is realized by means of integrating all data copies into a logic ring, when certain datum is modified, such operation is recorded and orderly transmitted to other data copies via the ring structure for data modification. Generally, each service node designates a master data copy

Table 34.1 Testing records of execution efficiency of data consistency protocols

Protocol	Time delay (ms)	Number of data copies	Number of data	Data size (byte)	Execution time (ms)
Real-time consistency protocol	4	10	1000	1000	20938
Timely consistency protocol	4	10	1000	1000	2326
Real-time consistency protocol	40	10	1000	1000	190726
Timely consistency protocol	40	10	1000	1000	2518

responsible for transmission of shared data among service nodes and several slave data copies. Data updating is firstly executed on the main data copy, then the main data copy server records the data operation log upon success, and then the content of the log file is synchronized to the slave data copy server of this node and the master data copy server of other nodes via special log synchronizing thread. Since the timely consistency protocol reduces adverse impact of network time delay on system transaction performance and eliminates the impact of the number of servers participating in the transaction on the transaction execution time by means of pipeline data transmission type, the data synchronization performance has been improved greatly, which has been proven in the experiment.

In the lab environment, the testing environment simulating 5 service nodes is established to prove the execution efficiencies of the real-time consistency protocol and the timely consistency protocol, where each service node consists of two servers and corresponds to one user, and the space for 10 data copies is established on 10 servers by full redundant storage mode of data. WAN simulation environment is formed among nodes via 100 M router (network time delay equipment is added among router links to simulate WAN time delay environment). The test records are as shown in Table 34.1. Seen from the test data, the timely consistency protocol can efficiently reduce the impact of network time delay on transaction execution efficiency.

34.3.4 Realization of Key Strategies

For actual narrowband WAN environment, in order to realize inter-node complete data sharing and symmetrical management application mode, the key strategies used for the multi-center data service platform include the following aspects:

1. Narrowband WAN: General distributed systems raise high requirements for network environment, while the multi-center data service platform can work smoothly in WAN's 2 M to 100 M high time delay network environment, with

good network adaptability. For special network conditions, the data consistency algorithm is optimized: Based on the basic principle of serial execution of conflicting transactions, the transaction execution process is changed from traditional active calibration of current transaction to passive calibration of subsequent transaction, so as to eliminate the impact of network time delay on the algorithm efficiency. However, the data transmission routing mechanism is also changed from traditional simultaneous data transmission mechanism of multiple copies to orderly transmission among multiple copies, which greatly eliminates the impact of network bandwidth on data peak and prevents network congestion.

2. Intelligent task scheduling: It has distributed intelligent dynamic task scheduling function to prevent central scheduling mode from causing system bottleneck and key fault point, guarantee system load balancing and enable effective utilization of system and network resources.
3. High reliability of data service: The data center and the data collection point can be connected with the server in any distributed data storage system, so the system has good fault-tolerant capability and it can satisfy the data service request of the data center and the data collection point in the distributed data storage system even in case of fault. The data of the data collection point or the data center stored into the distributed data storage system are subject to copy backup to guarantee high reliability of data.
4. Real-time processing of data: The external trigger function of the distributed data storage system may be utilized: When the data sent by the data collection point reaches each node, the server of the distributed data storage system can real-time send triggering message to the application program on the server, so as to commit the data to each data center. For the data from the data center to any other data center or the data collection point, the objective of real-time data processing can also be realized via the trigger function.

34.4 System Structure

The multi-center data service platform externally provides position-independent data sharing service via unified access interface and hierarchical system structure is adopted, including the network layer, the infrastructure layer, the core layer and the application layer, as shown in Fig. 34.3. The network layer provides network communication function; the infrastructure layer provides software and hardware operation environment and storage medium; the core layer is responsible for realizing data distribution management, synchronization control and dynamic scheduling of tasks among multiple centers, establishing unified shared information images and providing the application programs with position-independent transparent file access and database access interfaces; and the application layer provides the users with various customized applications via application interfaces.

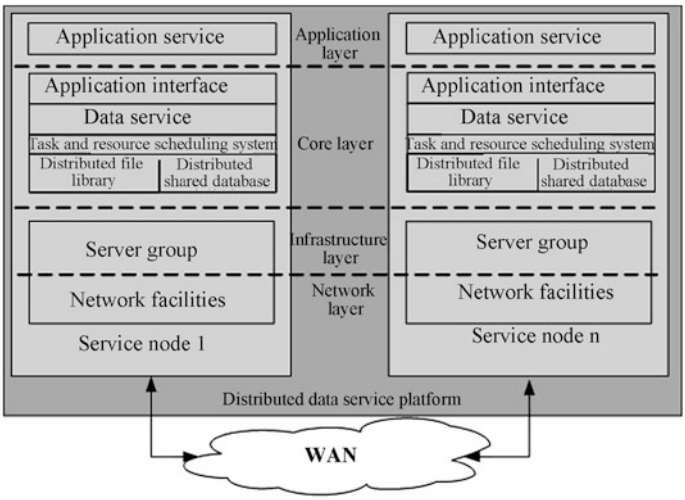

Fig. 34.3 Structural diagram of system

The main parts of the system structure are described as follows:

1. Application interface. Application interfaces include user access interfaces and Application Programming Interfaces (API): The user access interfaces provide the users with methods and tools for direct access to shared data; while the Application Programming Interfaces (API) provide software programming interfaces for third party software developers.
2. Data service. It provides shared information customization, inter-node authentication, user authorization control, configuration of operation parameters, etc. and provides the users with customizable safe and efficient shared information services.
3. Task and resource scheduling system. The task and resource scheduling system timely tracks the resource conditions of adjacent nodes by means of event triggering and searches for corresponding service resources for the application service request according to the service node loading conditions, the service distance, etc., so as to realize load balancing of the whole network and fault isolation. It provides management function for resources of the whole network and determines the resource and resource metadata placement, searching and positioning services.
4. Distributed file library. The distributed file library stores and manages unstructured file data, associates the file data with the file contents index, access authorization, sharing scope, etc. and externally provides unified position-independent file access interfaces. The shared file library realizes management of file copies distributed at various operation management nodes to guarantee consistency updating among multiple copies, provide delayed updating capability and guarantee the visitors to obtain the latest data. The shared file library has access control capability to prevent access of unauthorized users and prevent misoperation of authorized users.

5. Distributed database. The distributed database stores and manages structured data and stores shared data in the form of base table of database. The data service platform integrates the databases distributed at various service nodes to realize unified management on the data copies on them, externally provides unified transparent standard SQL access interfaces, provides data consistency and completeness guarantee mechanism and provides transaction control mechanism. The distributed database system is also the task and resource scheduling system and the storage system for file management information of the file sharing library, providing global data recording, storage, management and sharing capabilities.
6. Application service. The application service provides the users with various customized applications via application interfaces, mainly including data access interfaces, system management interfaces and data backup application for disasters, etc. provided for users.

34.5 System Features and Key Technologies

34.5.1 Peer-to-Peer Cluster-Based Network System Structure

The peer-to-peer cluster-based network is network system framework for distributed utilization and sharing of resources, using the system framework similar to "Cloud storage" which is dominant in current network. Cloud storage is extended and developed from cloud computing concept and it refers to the function of integrating and coordinating a large amount of different types of computing equipment via application software by clustering application, network technology and distributed technology, so as to jointly provide data storage and business access function [2] externally.

The similarity between the peer-to-peer cluster-based system framework and the cloud storage system framework lies in application of distributed technology, p2p technology, data storage technology, virtualization technology, etc. The difference between the peer-to-peer cluster-based system framework and the cloud storage system framework lies in that the cloud storage also adopts such central framework technology as clustering technology and grid technology and it is integrated and established by several grid frameworks, meaning that there is central point for cloud storage within local scope, while for the peer-to-peer cluster-based network, only distributed parallel technology is adopted, each subnet group is divided clearly, each service node in the network also plays the role of server and client, the nodes are completely equal, i.e. each service node also has such functions as receiving, storing, transmitting and integrating data, with clearer and more concise system structure model; cloud storage network lays more emphasis on central management and transparent service, while the

peer-to-peer cluster-based system is more inclined to distributed management and distributed service.

In addition to conventional distributed technical features such as load balancing and fault-tolerant mechanism, the multi-center data service platform also provides data priority control, network flow monitoring, directional data sharing, etc.

The peer-to-peer cluster-based network system is realized using the protocol of full redundant routing among servers. In such protocol, each server autonomously maintains a global routing table, creating, extending, exception handling, etc. of nodes in the system are detected and maintained via the routing table among servers, and all routing information is dynamically consistent. Among the service nodes and among servers in the system, the type of combining "reliable heartbeat" (TCP) and "unreliable heartbeat" (UDP) is adopted for system state monitoring. "Reliable heartbeat" guarantees monitoring accuracy, while "unreliable heartbeat" reduces the communication loads among system servers. By means of monitoring, faulty servers of the system and dynamically extended servers can be found out in time, and the network topology structure can be reconfigured intelligently.

Compared with current cloud network structure, the peer-to-peer cluster-based network has more powerful manageability and service controllability, with flattening, multi-center, anti-destruction and non-stop network features.

34.5.2 Realizing Sharing of Database Data and File Data

File data sharing is mainly used for large file data storage processing, with file data storage engine as the core. The file data storage engine is modified and designed from the INNODB storage engine of MYSQL database. The file storage path field is newly extended special field (the field type is file) defined by the newly extended syntax, and the system will automatically fill and manage the content. The methods for the other fields and conventional datasheet fields are the same. The data storage area of the file data storage engine is divided into the following two parts: One is the file data storage area (file system) and the other is the data record storage area (database base table) which stores the storage path of conventional data and file data in the file data storage area (file type field).

Database data sharing is direct data sharing among database base tables. Text data and conventional non-text data are included in database base tables and shared synchronously via synchronous SQL commands among database base tables.

Unified data storage management and standard SQL access interface are provided externally, and the data distribution details are masked. Based on copy management, synchronization and recovery of incremental data is realized, and data access is transferred smoothly among service nodes, so as to effectively increase the reliability and access efficiency of the database system. In order to support the development of application programs, it provides standard C-API and JDBC interfaces.

34.5.3 Realizing High Reliable Task Scheduling Based on Load Balancing and Fault-Tolerant Mechanism

Due to distributivity and dynamics of network resources as well as randomness of service requests, the loads of service nodes in the distributed system are unbalanced and therefore the system availability is reduced. According to the operation state and the position of the resource content, the multi-center data service platform uses intelligent scheduling and load balancing algorithm to realize load-balanced resource scheduling, so as to provide services nearby.

Load balancing algorithm is conducted based on nearby service and resource balancing principles. Each service node properly collects the load information of each server in the adjacent service node via additional information of "heartbeat information" monitoring data packet among servers. The task scheduling system selects the server with the minimum load in adjacent service node as the service provider according to the scheduled task type, required resource and network structure. Different monitoring technologies are adopted automatically according to the actual network conditions among service nodes, so as to efficiently complete monitoring of abnormal nodes. Based on fault detection, the faulty nodes can be isolated via dynamic reconfiguration of data and system, so as to guarantee continuous and healthy operation of the system.

Since each server in the system has the capability of completing task scheduling independently and can participate in task scheduling, the fault of any server will not affect operation of the whole system, so as to avoid system bottleneck caused by central scheduling.

34.5.4 Selection of Intelligent Shared Path

The network topology information is obtained by means of collecting "heartbeat information" monitoring data among servers and stored in the form of routing table. The structure includes all known server node information within the system scope as well as logic proximity relation among server nodes which may be modified dynamically according to the "heartbeat information" monitoring results and can reflect the physical link change state within certain time range from time to time.

Based on network topology information, the system conducts network quality statistics from the following three parameters: fault rate of network connection state, network bandwidth and network congestion and selects the optimal route for data transmission.

The weighted quantization model for network quality is defined as follows: $Y = (\Delta 1\beta + \Delta 3\gamma + \Delta 4\varepsilon)/\Delta 2\alpha$; where Y represents the weighted value describing network quality, β represents the network fault rate, α represents the network bandwidth, ε represents the network time delay, γ represents the network packet loss rate (the network packet loss rate is used to comprehensively describe the

network congestion and the whole load of the peer-to-peer server), and Δ1, Δ2 and Δ3 respectively represent the weighted value of the former three items. In the formula, the higher the weighted value of network quality Y, the lower the network quality, and vice versa. Seen from the formula, Y is directly proportional to the network fault rate, the network time delay and the network packet loss rate and inversely proportional to the network bandwidth.

Test in actual project environment shows that if there are several physical network channels among several centers of the system, the method of selecting the data transmission path based on the network quality can effectively prevent faulty link from affecting the system service quality.

34.5.5 Data Sharing Mode Based on Strategy

Customizable data sharing strategies are provided on database and data recording levels. The database sharing strategies are realized depending on database copy management function to control the database copy storage scope, while the data record sharing strategies depend on extended SQL syntax to set data sharing scope in SQL to realize local sharing of data.

The system supports coexistence of several shared strategies and provides single or bulk data sharing capability, and the data sharing content and scope may be customized according to the strategies. The shared data provider releases the data and shared strategies on the platform. The users subscribe their required data according to the shared strategies. The users beyond the strategy scope are unable to obtain the shared data.

34.5.6 Network Flow Control and Priority Mechanism

Network bandwidth is always the most important performance bottleneck in any distributed system: Limited network resources are shared by several applications in the system. In the network using preemptive mode, the unlimited usage of bandwidth will necessarily cause network congestion, which results in failure to obtain good service quality for each application, and effective network flow control is necessary. Therefore, the network flow control method adopted by the data access platform is to firstly set the total network bandwidth occupancy proportion, then establish the network flow control table according to the network topology information, and real-time record the message volume among servers within each time section. For all transaction threads to be communicated among servers, the flow shall be applied for from the network flow control table firstly and then the messages shall be sent. Each server in the service node exchanges the network flow value of the previous cycle on a regular basis and then the network

bandwidth to be dominated by the server in the next cycle is calculated, so as to strictly control the network flow.

In order to guarantee the service quality requirements of different businesses when the resources conflict with each other, the data service platform also provides priority-based service mechanism. For the requirements for different service levels, each data sharing object (file, database, etc.) is set with individual priority, and the businesses with higher priority have more network resources and service resources than those with lower priority. When the resources conflict with each other, the businesses with higher priority take over those with lower priority to guarantee that the businesses with higher priority always obtain better service quality guarantee in multi-business environment.

34.5.7 Realizing Mass Large File Data Storage and Synchronous Sharing

64-bit Linux operation system and ext4 file system are selected to realize mass large file data storage support. Large file data are independently stored in the file system, while relevant attribute information of the large files is stored in datasheets in database recording mode. Large file data as well as relevant attribute data and base table data are shared synchronously among several shared nodes to guarantee completeness of data transmission.

Technical means of fragment transmission and breakpoint transmission are used for large file data to increase the large file data storage and synchronous sharing efficiency.

Intelligent compression method is used during file data transmission process. Compression anticipation is conducted according to the file type and data size. Only the file data satisfying the compression conditions will be subject to compressed transmission to increase the network utilization rate and synchronous data sharing efficiency.

34.5.8 Continuous Service and Data Backup for Disaster

The system has self organization capability and self reconfiguration capability, e.g. when the service nodes are added or separated dynamically, the system can automatically reorganize the nodes to form new service platform topology structure; when the servers in the nodes are added or separated dynamically, the system can automatically reconfigure the shared node. The reconfigured system has the function and operation mode unchanged, and the reconfiguration process is transparent to the users.

The function and status of each service node in the multi-center data service platform are equal and there is no control center or resource center or key fault node. When the proximal service node fails, the platform may select other nodes alternatively; when certain server in the node fails, other local servers may be connected and used. Therefore, the platform will greatly increase the sustainability of data services.

Meanwhile, since the shared data redundancy is stored at the server of each data service node, which provides the users with data backup and recovery means, i.e. when any fault of the information center system causes data loss, the nearby service node can be used for recovery; when the lost data exceed the storage scope of the service platform, other center may be used to obtain data prior to data recovery.

34.6 Conclusions and Outlook

The multi-center data service platform optimizes data consistency algorithm against WAN features (small bandwidth and long delay) based on network technology, distributed technology and parallel computing technology, using peer-to-peer cluster-based network system framework, combined with "Cloud storage" framework features, and provides such characteristic functions as strategy-based oriented sharing, priority control and network flow control as well as capabilities of synchronously sharing file data and database data facing overall situation and subject to unified management. The global distributed data redundancy storage also provides the users with new data backup for disasters; intelligent troubleshooting and inter-node high self-fusion features can provide efficient data fault-tolerant and data service recovery capabilities in case of fault of network or other hardware as well as continuous and non-stop data services for the users.

Currently, the system has realized engineering and has been put into formal operation. The multi-center distributed data service platform including 4 service nodes has been established. Joint debugging in actual operation environment shows that the system can meet the engineering demands in synchronous data sharing performance, system stability etc., and has good promoting value.

In field performance test of the system, for the two 10 MB data samples compressed in typical 2Mbit/s WAN environment, the sharing time among nodes is less than 180 s; for the compressed 2000 10 KB data samples, the sharing time among nodes is less than 210 s; in the system stability test, if 1 of 2 servers is closed, the current data sharing task will not be affected distinctly, and the data synchronization performance will be kept unchanged; when any new server is added, the topology reconfiguration time is less than 100 s, i.e. when not subject to data recovery operation, the server newly added within 100 s can provide data access.

To sum up, the distributed system has been developed continuously from concept to realization, and the distributed data storage system has realized the capability of providing the users with reliable and efficient distributed data storage and sharing services. In future, the system fault-tolerant capability will be further improved, optimized scheduling algorithm will be designed, the resource sharing capability will be strengthened, and researches and practices on the system expansion and management, the system safety and identity authentication etc. will be strengthened.

References

1. Yang J F et al (trans) (2004) Distributed systems principles and paradigms. Tsinghua University Press, Beijing
2. Sa SX (2001) Introduction to database systems, 3rd edn. High Education Press, Beijing
3. Liu K, Dong LJ (2011) Cloud storage and management. Comput Syst Appl 20(6):

Chapter 35
Impact Analysis of the Leap Second to the Computer System in Beijing Aerospace Control Center

Tonghua Li, Yuqiu Liang and Xia Wang

Abstract It analyzes the impact of the leap second at 07:59:60 July 1, 2012 to the timing system and applications in Beijing Aerospace Control Center (BACC), and proposes a solution for the leap second adjustment. First it introduces the source of the leap second. Next it articulates the timing system device, timing system clients, timing system topology, as well as their working mechanism. Then it describes the experiments on the leap second adjustment, presents the test contents, procedures, and results under the automatic and manual adjustment modes. It also presents the test results of the leap second adjustment under the NTP protocol. The test results indicate that the timing system devices work well under both automatic and manual adjustment modes, and the clients can synchronize with the timing system device in 4 s. When the adjustment is executed on the timing system device, the clients can synchronize with the timing system device after a pause of one second at the synchronization point. As for the NTP conditions, the clients can not synchronize within 5 min, and can synchronize immediately if the clients are rebooted. The leap second will make a big impact on the time-sensitive systems such as the TT&C devices and applications. It suggest that the leap second adjustment and synchronization need to be executed outside the critical command and control procedures, and a thorough implementation plan needs to be made. The most important point is that the synchronization must be after a delay of 4 s in the leap second adjustment.

Keywords The leap second · Timing system · The standard time · The time of compute system

T. Li (✉) · Y. Liang · X. Wang
Beijing Spaceflight Control Center, Beijing 100094, China
e-mail: lith215@126.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_35,

35.1 Introduction

Time is an important parameter in people's aerospace activity. Timing system is the necessary condition to guarantee the validity and accuracy for aerospace TT&C. The timing system of computer system in TT&C system is achieved by the nodes in computer system that receive standard-time signal from the timing system devices under the definite time synchronization mode, and calibrate time of the local computer system. From this, it can keep a unified time scale in the whole TT&C system [1].

According to the arrangement of China's national GPS clock, the leap second was added after the Beijing time 7:59:59 on July 1, 2012. The next second was going to be 7:59:60. And then the time would get into 8:00:00. In this paper, the principle of the leap second generation and adjustment is analyzed, the technology state of timing system device and working mechanism are studied, the results of our leap second tests are provided. Through these researches, the time variation law of timing system devices and computer system's devices under manual and automatic leap second are obtained. The influence of the timing system and application software in BACC is analyzed, and the leap second adjustment solution of computer system is given.

35.2 The Source of Leap Second

The purpose of leap second is to guarantee the differences between UTC time based on atomic clocks and the UT time based on the rotation of the earth within one second. The extra second was mandated by the International Earth Rotation and Reference Systems Service (IERS) in Paris. Normally, it is the last second on 31 December or 30 June.

The Beijing time is based on the UTC time. Because of the slowing rotation of the Earth, the UT time will lag behind the UTC time. It is adopted by positive leap second, which means the UTC time will be set one second behind to keep the differences between the UT and UTC within one second. At the leap second moment, the UTC time will take place one second step; it will stop one second in contrast with the normal second sequence.

35.3 The Technology States of Timing System

35.3.1 The Timing System Devices

The timing system devices generate the B-code time as the standard time signal source in Spaceflight Control Center, including B (DC) code and B (AC) code,

according to the GJB2991A-2008 B code time interface terminal general specification.

The timing system devices can support the leap second adjustment by two ways: one is the automatic leap second, which means that the leap second can be set positively or negatively by default, and the device can perform automatically. The other one is manual leap second, which means that one second can be added or minus on the device by manual control.

35.3.2 The Timing System Device for the Client

The timing system devices for the clients in the computer system of BACC include two types:

1. Physical clients: They are the clients in traditional sense, including B code terminals, timing system cards and sound cards.
2. Logical clients: This paper defines the clients which adjust time by TCP/IP protocols as the logical clients, including NTP timing system clients, UDP multicast timing system clients and all other clients adjusting time by network protocol.

B code terminals receive the standard time code signal from the timing system devices. They provide the needed time signals by decoding the time signal and synchronize with standard time. They output the signals including B (DC) codes and time interrupt signals. The B (DC) codes contain day, hour, minute, second and micro second. The interrupt signals contain 1 and 20 Hz signals. The B code terminals are used for branching on the standard time to clients.

Timing system card is installed in a computer which needed timing system signal. It uses the timing system device to synchronize its clock, and uses the high precisely constant temperature crystal to realize punctuality. The timing system card provides accurate time signals including hour, minute, second and microsecond.

The sound card is one standardized component of a PC. It can achieve 16-bit dual-channel hi-fi data acquisition. The top sampling frequency is 44.1 k/s, the lowest is 8 k/s. If we modulated the B (DC) code with 1 Hz sine, we can obtain the B (AC) code. As the result, the sound card can sample the B (AC) code precisely, and turn it into the digital signal which PCs can use. The PC will correct its local system time according to the digital signal.

Network Time Protocol (NTP) is the accepted timing system protocol on Internet. It uses hierarchical time distribution model, and runs over UDP to transmit time packets. It can choose the closest and most authentic offset for local clock by the method of statistics and probability theory. Then the PC can correct the time of its system by local clock adjustment model.

UDP multicast: It synchronized the servers based on IP multicast. Under the normal circumstances, the time-server provides the standard time, and the other servers must keep in step with it.

35.3.3 The Connection Modes of Timing System Devices

There are three connection modes of the timing system devices and the clients: direct mode, indirect mode and network connection mode.

1. Direct mode. In the direct mode, the timing system devices connect the clients by cable. The direct-mode clients include B code terminal, timing system cards and sound cards etc. The B code terminals and timing system cards use B(DC) code, and the sound cards use B (AC) code.
2. Indirect mode. In the indirect mode, the timing system devices connect the clients by B code terminals. The indirect-mode clients are timing system cards, which use B(DC) code.
3. Network connection mode. This mode means that the timing system devices connect the clients by TCP/IP network. The signal of timing system is transmitted through the network, not the special transmission media. NTP and UDP multicast all use this mode—transfer the standard signal of timing system to the clients by network protocol.

35.3.4 The Timing System Mechanism of Center Computer System

The goal of the timing system is to synchronize the system time of servers with the standard time. According to the different timing system device and connection modes, there are several timing system mechanisms:

1. The time-server with timing system cards: The time-server attached the timing system device directly to receive the B (DC) code signal. It gets the timing system card's signal by the preset period, multicasts the time packets to the LAN's clients.
2. The function-server with timing system cards: The server gets the B (DC) code signal indirectly. When it receives the time synchronization signal from the time-server, it will correct the time of its system according to the local timing system cards.
3. UDP multicast clients: These servers receive the UDP multicast packets through the time-service client application. The UDP multicast packets transferred by the time-server contain standard-time and time-synchronization signal. When the server receives the packets, it will correct its local time according to the standard time.

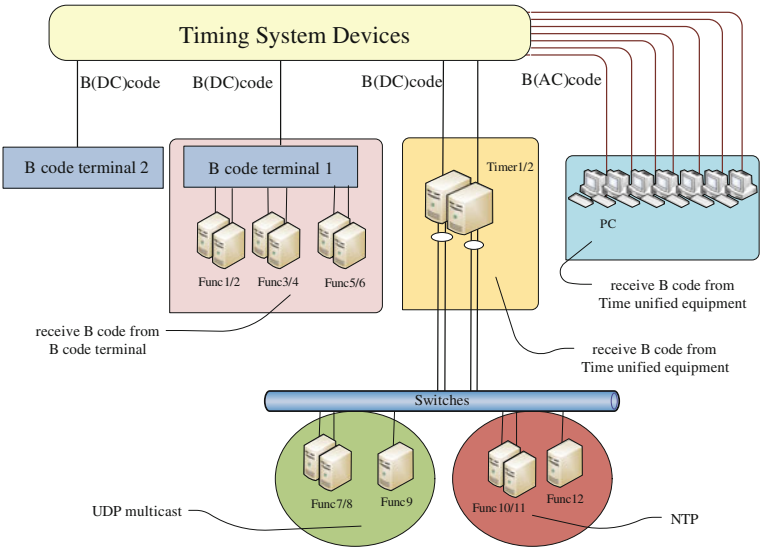

Fig. 35.1 The timing system topology in the computer system

4. NTP clients: The servers are configured as the NTP clients, using NTP protocol to adjusting local time. The time-servers are designated as NTP servers. The synchronization period is configured by the client.
5. The sound cards clients: Some servers use the sound cards installed on main board to obtain B (AC) code signal adjusting local time. The synchronization period is configured by clients.

Overall, the two essential factors of timing system are standard time source and synchronization period. For timing system cards and network cards clients, the time-server undertakes the essential core role whichever kind of time adjusting mechanisms. For timing system cards clients, standards time is caused by local timing system cards, the time-servers only provides the synchronization period. As for UDP multicast clients, the time-servers provide both standard time and synchronization period. That is, the time-servers are responsible for the timing system clients and UDP multicast clients in whole system. The sound cards clients are individual clients, the standard-time and synchronization period are produced by themselves.The timing system diagram is shown in Fig. 35.1.

Table 35.1 "Automatic mode" test results

Seq	Device name	Timing sequence	System clock sequence
1	Timing device	59 **60** 00 01 02 03	–
2	B-code terminal1	59 **00 60** 01 02 03	–
3	B-code terminal 2	59 **00 60** 01 02 03	–
4	timer1	59 **00 00** 01 02 03	59 00 01 02 03 04
5	timer2	59 **00 00** 01 02 03	59 00 01 02 03 04
6	func1	59 00 **01 01** 02 03	59 00 01 02 03 04
7	func2	59 00 **01 01** 02 03	59 00 01 02 03 04
8	func3	59 00 **01 01** 02 03	59 00 01 02 03 04
9	func4	59 00 **01 01** 02 03	59 00 01 02 03 04
10	func5	59 00 01 **02 02** 03	59 00 01 02 03 04
11	func6	59 00 01 **02 02** 03	59 00 01 02 03 04
12	PC 1	59 00 01 02 **03 03**	59 00 01 02 03 04
13	PC 2	59 00 01 02 **03 03**	59 00 01 02 03 04

35.4 Testing of Leap-Second

35.4.1 Testing of Leap-Second

In order to control the changing condition of the standard time and the time of computer system for both the timing system devices and timing system clients in the whole system, the testing of leap second is performed.

35.4.1.1 The Test Content

The test devices include timing system devices and the timing system clients:

1. The timing system devices: the standard timing system devices and direct–connection cable.
2. The timing system clients: two B code terminals, two time-servers (Timer1/2) with timing system cards, six main servers (func1–6) with timing system cards.

The timing system runs under the "automatic mode" and "manual mode" separately, observing time variety of the faceplate of timing system device, the timing system clients and the computer system. During the tests period, the automatic synchronization configuration in time-servers and the sound cards clients was disabled; the configuration is enabled after the tests.

The tests under "automatic mode" and "manual mode" about the leap second were both completed separately. The results are consistent. Tables 35.1 and 35.2 show the results of leap second test.

Table 35.2 "Manual mode" test results

Seq	Device name	Timing sequence	System clock sequence
1	Timing device	59 **59** 00 01 02 03	–
2	B-code terminal1	59 **00 00** 01 02 03	–
3	B-code terminal 2	59 **00 00** 01 02 03	–
4	timer1	59 **00 00** 01 02 03	59 00 01 02 03 04
5	timer2	59 **00 00** 01 02 03	59 00 01 02 03 04
6	func1	59 00 **01 01** 02 03	59 00 01 02 03 04
7	func2	59 00 **01 01** 02 03	59 00 01 02 03 04
8	func3	59 00 **01 01** 02 03	59 00 01 02 03 04
9	func4	59 00 **01 01** 02 03	59 00 01 02 03 04
10	func5	59 00 **01 01** 02 03	59 00 01 02 03 04
11	func6	59 00 **01 01** 02 03	59 00 01 02 03 04
12	PC 1	59 00 01 02 **03 03**	59 00 01 02 03 04
13	PC 2	59 00 01 02 **03 03**	59 00 01 02 03 04

35.4.1.2 The Analysis of Test Results

From the results above, in the automatic leap second mode, the faceplate of timing system device runs as the same as IERS described. The devices including main servers, time-servers, B code timing devices, and the sound cards clients run inconsistently from the beginning of "07:59:60" until "08:00:03". The servers' system time keeps the step of the standard time sequence of "59,00,01,02,03……".

Under the manual leap second mode, the faceplate of timing system device appears "07:59:59" twice, the devices including main servers, time-servers, B code timing devices, and the sound cards clients run inconsistently from the second "07:59:59" until "08:00:02". The servers' system-time keeps the step of the standard time sequence of "59,00,01,02,03……".

The tests results indicate that under both automatic and manual leap second modes the timing system devices work well. The timing system clients run consistently with timing system devices at most 4 s later. When the time synchronization period is cancelled, the servers system time will keep the step of the standard time sequence of "59,00,01,02,03……". That is, after 4 s it will be 1 s faster than timing system. When leap second is performed, the timing system clients, UDP multicast clients and sound cards clients will be synchronize with the timing system after enabling the time synchronization period of time-servers and sound cards. It will make a pause of 1 s while synchronizing.

35.4.2 NTP Tests

In order to master the changing condition, we put up a timing synchronization environment [2], and then conduct the NTP clients leap second tests.

Table 35.3 NTP Clients' devices test record

Seq	Cycle configuration	NTP Server System clock	NTP Client System clock	Restart NTP-service in Client
1	Default configuration: the initial value of synchronization period is 64 s, then changed to 1,024 s	Adjusted about 1.7 s fast	Observing 5 min, time is not synchronized	The client time is synchronized with NTP Server time
2	Default configuration: the initial value of synchronization period is 64 s, then changed to 1,024 s	Adjusted about 1.9 s slow	Observing 5 min, time is not synchronized	The client time is synchronized with NTP Server time
3	The synchronization period is 16 s	Adjusted about 1.8 s fast	Observing 5 min, time is not synchronized	The client time is synchronized with NTP Server time
4	The synchronization period is 16 s	Adjusted about 1.9 s slow	Observing 5 min, time is not synchronized	The client time is synchronized with NTP Server time

35.4.2.1 The Tests Content

Under the test circumstance, a PC server is configured as a NTP server, a SUN server is configured as NTP client. During the default configuration and enabled the timing synchronization period, the administrator adjusted the NTP server system clock one second fast or slow, then observed the time changing condition in NTP client.

35.4.2.2 The Analysis of Test Results

The results (Table 35.3) show that, The NTP clients do not synchronize shortly during the NTP servers occurring two-second jump. If the NTP clients need to synchronize immediately, the NTP service should be restarted in the NTP clients.

35.5 The Impact and Suggestion

The timing system clients that follow the GJB2991-1997 standard do not support leap second protocol, which means that the devices could not output 07:59:60 normally. They can synchronize after the timing system runs leap second in four seconds, and it will make a second pause during that period. The servers in the computer system of BACC will make 1 s pause also during synchronization. So

the leap second will make a great impact on time sensitivity system such as measure and control system as well as measure and control application.

The automatic mode is suitable for no key service during the leap second, otherwise, the servers should synchronize timing system avoiding the key service. The thorough plan should be made to guarantee the application software works well while the system time has 1 s pause.

Taking the two essential factors into account, during the leap-second, the time synchronization period in time-server should be stopped in order not to synchronizing automatically. In addition, in order to keep the timing source stable, the server system clock should be synchronized 4 s after the timing system runs the leap second.

References

1. Huang XD, Cheng QQ (1996) Missile measurement and control manual. National Defense Industry Press, Beijing, pp 197–198
2. Sun Microsystems, Inc. (2006) System administration guide: network services, May 2006, pp 59–62

Chapter 36
The Exploration and Practice of Itinerant Testing for TT&C Device Software at the Launch Range

Peng Fu, Wei Li, Liang Zhao, Wei Zhang and Jing Zhang

Abstract In view of the latent software faults and unsteadiness of telemetric tracking and command (TT&C) device software at the Launch Range, Software Itinerant Testing (SIT) is brought forward. Software faults are attributed to software manufacturer's lack of necessary ability and shortage of process specification. Comparison of SIT and software project testing shows that SIT is a convenient activity that can be taken. The first time that SIT worked at launch range took about 8 weeks and found 52 faults. The result shows that TT&C device software is under unreliable state and with low robustness. SIT is an effective measure to solve TT&C Device software faults and improve the robustness and reliability of TT&C device software.

Keywords TT&C device · Software fault · Software update · Software itinerant testing

36.1 Introduction

The TT&C devices at the launch range are composed of hardware and software. With the development of fundamental industry technology, the trend of miniaturization and integration, the reliability of electronic equipments are greatly improved. Nowadays, more functions are implemented by software. Software has become the key factor of TT&C devices function. At the same time, the consummation of software development process and the non-conformance of software

P. Fu (✉) · W. Li · W. Zhang · J. Zhang
Taiyuan Satellite Launch Centre, Taiyuan 030027, China
e-mail: wallffpp@hotmail.com

L. Zhao
Beijing System Engineering Institute, Beijing, 100101 China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_36,

update activity result in more latent faults in software. This increases the risk of launching task. For example, after the update of optical device software, the third party software regression testing has found more than one fault. So to improve the reliability of TT&C devices, regular SIT is put forward.

36.2 The Comprehensive Analysis of TT&C Device Software

TT&C devices can be divided into four classes as optical, radar, telemetric and remote control device [1]. Devices in each class can be further divided into more types. Some devices have been put into use from the 1970s, and have experienced many times of hardware and software updates. Some devices have just been deployed and their software still not thoroughly explored. According to the statistical information, the proportion of the software malfunction has increased from 10 to 30 % in the recent 20 years at the launch range. The main reason that causes the increase tendency of device software faults can be analyzed as follows:

1. The increase of the ratio of software functions in TT&C devices. More functions that used to be implemented by hardware are now done by software, and the number of software configuration items in device software system has increased.
2. The management of software development process is not enough. According to the collected information, the software capability maturity level varied greatly from different device manufacturers. The manufacturers usually put more emphasis on the development of hardware than on software. Generally they pay little attention to software project tracking, quality assurance, and configuration management. Especially, for legacy devices whose software had been updated for many times, no software engineering work was done by the developing team [2].
3. Software requirement analysis is not thorough. Generally device software requirement analysis is not thorough, which causes an obvious gap between the functions that were realized by the device software and the real requirements of the launch range. At the same time, with the change of launch task, the real software requirements are continuously changing. These differences and changes require the update of device software. Update may bring new faults into the software.
4. The software update management is not rigorous. The update of device software is frequent. Investigation of the situation of TT&C device update shows that the proportion of the device types whose software had been updated is 68.8 %; and the proportion of the number of devices that had undergone software updates is 81.4 %.

The update activity management is not rigorous. According to the software engineering best practice and the launch range's rules for software update, a complete software update process should include original software fault report, thoroughly discussed update scheme, application for software update, rigorous configuration management, regression testing and professional technical review. But in reality, these rules are not followed and most of time one or more practices are leaped, especially for legacy device software update.

36.3 What Does SIT Do

SIT is to run periodic configuration item testing or system testing to TT&C device software at the launch range. Before the testing activity deployment, some fundamental characteristics should be fully understood about SIT which includes testing precondition and method of SIT.

36.3.1 The Characteristics of SIT

SIT is different from common configuration item testing or system testing. It is also different from common software regression testing. Its characteristics include:

1. SIT is not the necessary condition for software commission. The testing in software development and maintenance phase is a necessary task for software commission, which should be followed by a formal technical review. Software that is not tested and reviewed can not be qualified for commissioning [3]. The object of SIT is the software that has been commissioned. It is a supplement to regular testing. Even software faults that were found in SIT have not been corrected; the software can still be used.
2. No precondition for SIT. Usually there are clear preconditions for software testing. For example, complete documents, finished unit testing and software be checked out from controlled library are all necessary for configuration item testing. Regular regression testing requires complete documents and software should be under the control of configuration management. For SIT, its object may be software that has been updated, or software that runs unsteadily. The department of device management, the user of device and the department of software testing can decide when to start SIT.
3. Different strategies of SIT. Testing strategies in software development phase should originate from software task document or development document, software maintenance testing strategies should originate from fault report and software update scheme. For SIT, although there is no complete document and no testing before commissioning, the fundamental fault has been solved and the functions of device software were well known by the manipulator of the device

after long software running period. Software testers can use some methods to get the necessary information of the software by discussing with users and operating the software personally. Then testing strategies can be worked out from the knowledge we learned from. Usually testing scope is the updated part or the influenced part of the software, including the key part and the weakness of the software.

36.3.2 Selection of Software to Run SIT

Two kinds of device software are suitable for doing SIT:

1. Device software that has been running for a period of time. After the TT&C device has been transmitted, the device software comes into maintenance stage. Generally the software may not run steadily. The device users hope to know and improve the quality of software, so they ask for SIT.
2. Device software that has been updated. After deployment of the TT&C device set in use, users continually bring forward new functional and performance requirements. Device manufacturers want to update the device software on site. Usually there is no regression testing for this kind of update [4]. This kind of software is also suitable for doing SIT.

36.3.3 Methods of SIT

SIT is different from common testing; there are some special methods to enhance the effect.

1. Planning of SIT. SIT is a massive work, which involves many devices, as well as their users and manufactures. It will take a long time to finish a complete SIT. A complete SIT includes testing planning, testing implementation, software updating, and regression testing and testing summary. SIT should be treated as a project. Before start, the testing activities should be adequately planned to make clear testing strategy, key points of testing, test time, clear testing goal, adjustment of testers, uniformity of testing documents, etc. A detailed implementation plan can help to make sure the testing activities proceed smoothly, and the work done by different personnel in different stages be unified.
2. Qualified Tester for SIT. For SIT, qualified software tester is necessary. SIT is usually arranged in the interval of launch missions, so the time to do SIT is not abundant as common testing. It has to test lots of device software in a short period of time. At the same time, there are many difficulties when doing SIT, such as lack of software documents, difficulty to learn software requirements comprehensively, etc. So the test team should consist of personnel that have

strong sense of responsibility, are familiar with function and performance of device software, and are familiar with software testing specification [5]. In this way, more test cases with rational coverage will be designed, more software faults will be found, and testing efficiency will be increased. Even device software has been tested before running, SIT is also meaningful. Different testers can found different faults in the same software. In addition, the testing ability of testers is constantly improved. So the same tester test the same software again may also find new faults that were not found in an early testing.

3. Reuse of software test cases. There are many similar functions between the same types of device software. The test cases should be reused as much as possible when doing SIT. In this way, the test resources can be saved; testers will have more time to design more targeted test cases according to the characteristics of tested device software. If the test case that was not passed before is reused to test the same type device software, it will be easier to detect errors in software.
4. SIT can be divided into several phases. Usually SIT can not test all device software in one time. First, the TT&C devices at the launch range are located in different places and are geographically far away from each other. The long journey will make testers tired and reduce its efficiency. Secondly, the number of devices is very big which need long test time but the interval period of missions is short. So the testing task should be divided into several phases according to the device layout and the type of device. Different personnel join in the test team in different phases. In this way, different test methods can be brought into testing. This will help find more faults.

36.4 The Practice of SIT

After discussion of the characteristics, conditions, testing methods and having carefully planned, the test team can begin to carry out SIT. The first time SIT at launch range took about 8 weeks and tested about 15 device software configuration items.

36.4.1 Design of Software Test Case

According to the testing plan, 172 test cases are designed to test TT&C devices. All test cases are fully executed. The distribution of test cases is presented in Table 36.1.

Table 36.1 Distribution of test case

Device type	Number of test cases
Optical device	14
Radar device	146
Safe control device	12
Total	172

Table 36.2 The number of software faults and its serious level

Device type	Level 2	Level 3	Level 4	Level 5	Total number of each level
Optical device	0	3	3	0	6
Radar device	1	3	36	3	43
Safe control device	0	0	3	0	3

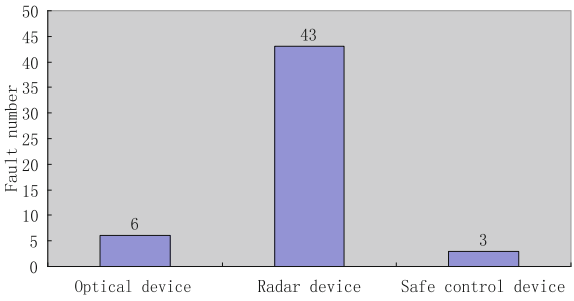

Fig. 36.1 The number of software faults for each device type

36.4.2 The Distribution of Software Faults

1. The number of software faults and its serious level. The distribution of software faults according to its serious level is presented in Table 36.2 and Fig. 36.1.
2. The proportion of software faults among different device types. The proportion of software faults among different device types is presented in Table 36.3 and Fig. 36.2.
3. The distribution of software faults of different serious level. Among a total of 52 software faults, the number of 2-level faults is 1, which account for 1.9 %. The number of 3-level faults is 6, which accounts for 11.6 %. The number of 4-level faults is 42, which accounts for 80.7 %. The number of 5-level software fault is 3, which accounts for 5.8 %. The detail distribution is shown in Table 36.4 and Fig. 36.3.

Most of the test cases executed in SIT were newly designed. Reuse of these test cases in testing old device also found some new software faults.

Table 36.3 The proportion of software faults among different device types

Device type	Fault number	Proportion of each type (%)
Optical device	6	11.6
Radar device	43	82.6
Safe control device	3	5.8

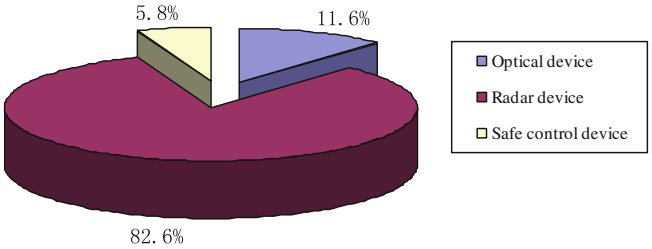

Fig. 36.2 The number of software faults for each device type

Table 36.4 The distribution of software faults of different serious level

Fault level	Fault number	Proportion of each level (%)
Level 2	1	1.9
Level 3	6	11.6
Level 4	42	80.7
Level 5	3	5.8

36.4.3 Analysis of SIT Results

36.4.3.1 Faults Arise from Poor Process Quality

Software quality consists of software process quality and software product quality. Software process includes management process, engineering process and supporting process. Software product consists of document, code and data. Software process quality is positively interrelated with software product quality. The larger the scale of software or the higher the critical level of the software, the stronger the correlation between software process quality and software product quality. Most of TT&C device software at the launch range is large scale software. Usually, if the software process quality is poorer, more faults will be latent in software product. This phenomenon is confirmed by investigation of software manufacturer whose device software is faulty.

Software faults that found by SIT were discussed and negotiated among device management department, device manufacturer and device user. Some faults were selected to be modified according to the principle of getting optimal return of investment.

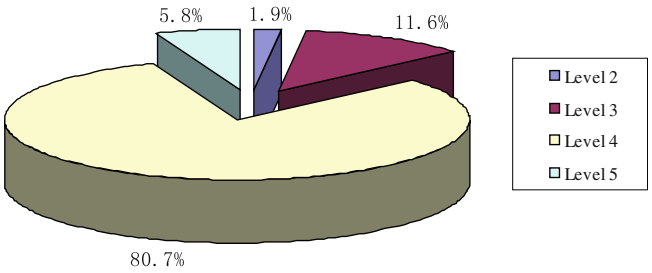

Fig. 36.3 The distribution of software faults of different serious level

36.4.3.2 Evaluation of Software Quality

Most of the serious software faults were found by extreme operating, exceptional operating, exceptional data input or exceptional data form input. According to the statistics of SIT and analysis of faults, we can conclude that: The TT&C device software can steadily run under limited condition. If the environment and condition go beyond limit, such as receiving exceptional data or exceptional data form, the software will fall into exception. Because the device software at the launch range is high level real-time software, high reliability is required. So we consider that quite a few TT&C device software at the launch range are in the status that its reliability is not high enough and its robustness is not strong enough.

36.4.3.3 Evaluation of SIT Effect

SIT is an effective measure to find TT&C device software faults. It can be confirmed that most of the reasons that cause device software faults will still exist in the near future. So SIT should go on. The interval between two SIT activities may depend on the running situation of new device software and update situation of old device software. Three to five years may be a reasonable choice.

36.5 Conclusions

The exploration and practice of TT&C device SIT at the launch range show that this activity has great significance. It is a new method to understand and enhance the quality of TT&C device software. After completion of the first time SIT, the failure rate of TT&C device software is obviously decreased and the reliability and robustness are increased, which greatly improved TT&C device's capability to complete launch mission.

References

1. Liu YC, Zhang JS (1996) Missile and satellite monitoring and control systems engineering. National Defense Industry Press, Beijing
2. Li W, Cheng XF (2010) Improvement of TT&C device software life cycle quality management. Missile Test Technol 1:32–34
3. Zhang HF (2008) An introduction to software engineering. China Machine Press, Beijing
4. Sun YT, Lan RE, Peng J (2011) Discussion on the reliability of military software. Technol Found Natl Def 1:31–34
5. Fan SX, Nie WS, Tao QZ (2001) Discussion on the military software testing technology. Sci Technol Innov Her 25:23

Chapter 37
Research on Quality Assurance Method Based on Software Defect Analysis

Qianran Si and Guoying Yan

Abstract Software defect information represents the characteristics of software processes and it can provide feedback to software quality assurance activities. The basic principle of software defect baseline is discussed, and a software quality assurance framework based on defect baseline feedback information is proposed. The relation between software activities and defect information collected through the process is studied. The statistical characteristics of defect information and it development trends along with the software process advance can reflect the effectiveness and efficiency of different quality assurance activities as feedback. Combined with practical projects, software defect metrics based on Orthogonal Defect Classification framework is built. The effectiveness and efficiency evaluation of quality assurance activities using the result of defects analysis is carried out with this quality assurance model to determine the direction of the QA improvements and help the software process integral improvement.

Keywords Software defects · Defect analysis · Orthogonal defect classification · Software quality assurance

37.1 Introduction

A software defect is commonly called a bug, which is the software's departure from its expected attributes, meaning the dissatisfy substances in software products. The software defect is the software's immanent attribute. Software failure

Q. Si (✉) · G. Yan
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: sqrblule@hotmail.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_37,

mechanism shows that a software failure is the dynamic characteristic in software runtime under specific internal and external conditions, caused by software defects [1]. Software defects are correlative with software quality, and they reflect the validity and completeness of software.

In software development process, software QA activities are aiming at decrease defects injected, raise the efficiency of defects detected and removed, and minimize the defect omission risk after product delivery etc., so as to improve software quality. Research shows that the statistical characteristics of software defect information reflect the features of software QA activities [2]. The relationship between the software defects and the software QA activities is analyzed, and QA method based on software defect analysis is promoted.

37.2 Relationships Between Software Defect Information and QA Activities

Making use of software defect information collected through software development process can help plan and control software quality activities effectively, and provide support for the improvement of software quality. It will be a good choice to establish software defect collection infrastructure and defect classification framework, optimize QA strategy according to software defect analysis result, and establish software QA improvement model based on software defect baseline. In one software project, software defect information from different phases can be used to evaluate the effectiveness and validity of software QA activities. Software defect analysis result provides support for the QA activities improvement, and the software defects metrics help identify the right directions of QA improvement.

Figure 37.1 shows the correlation between software defect analysis activities and QA activities in classical waterfall development model. With a reasonable defect classification framework, software problems collected can be abstracted into different types of software defects, and stored in the project database. According to the classification framework, serious of software defect metrics can be established which form the basis for defect analysis. QA and developers analyze the statistical distribution characteristics of currently collected defect information, with the help of reference information from historical projects and experiential knowledge of domain experts. Through this procedure, information about the validation, effectiveness and efficiency elimination of QA activities and the development process characteristics can be gained, which help the adjustment of QA policies, so as to improve the whole QA process and the software quality continuously.

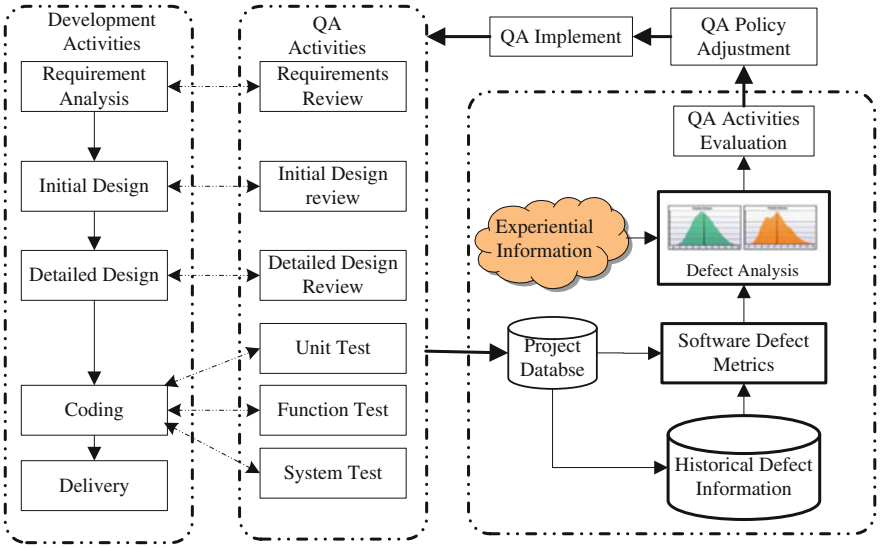

Fig. 37.1 Correlation between defect analysis and QA activities

37.3 Defect Classification

As we need the feedback information from software defect analysis, the correlation between the defect metrics and software process need to be established. The research leaded by Ph.D. Ram Chillarege at IBM Research showed that the cause-effect relationship between the software reliability growth curve and the variation characteristic of a specific defect type can be identified if the software defect was classified according how it was fixed [2]. This research forms the basis of Orthogonal Defect Classification [3], invented by IBM Research. Though ODC, phase based statistical defect analysis can be done, and quick process feedback with defect data can also be achieved.

In ODC, one defect has eight core attributes, and when the defects are collected and analyzed in-process during an ongoing software development, information about defects is available at two specific points in time. When a defect is opened, the Activity, Trigger and Impact attributes are captured, and when a defect is fixed, the Type, Qualifier, Source, Severity and Age attributes are captured. Each defect detected in software lifecycle is unambiguously assigned eight orthogonal attributes, the value set of which is in a restrictive set, related to software process features [4].

Table 37.1 Correlation between software defect type and software activities

	Design	Design review	Coding	Code review	Unit test	Function test	System test
Assignment/ Initialization			★	★	★		
Checking			★	★	★		
Algorithm/Method	★	★	★	★	★	★	
Function/Class	★	★				★	
Interface/Messages	★					★	★
Timing/Serialization	★	★					★
Relationship	★	★					★

37.3.1 Definition of Defect Classification Framework

The value set of each attribute should be customized according to the characteristic of specific software organization. To the development model shown in Fig. 37.1, the value set of Activity (meaning defect detection activities in ODC) contains requirement review, initial design review, detailed design review, unit test, function test and system test. Obviously, the Activity attribute value set belongs to QA activities.

Trigger means the conditions under which the defect is exposed. Correlation exists between the activities and triggers [2]. Defect Type attribute is determined just by the defect fix operation done by software developers. Type value set contains Assignment/Initialization, Function/Class, Algorithm/Method, Checking, Interface/Messages, Timing/Serialization, and Relationship. Other attributes' value set can be determined according to the organization's process characteristics.

37.3.2 The Correlation Between Defect Information and QA Activities

To build the software defect baseline, the correlation between software activities and defect type information need to be identified. It is necessary to determine which kind of software activities tends to inject specific types of software defects, and which kind of tends to detect specific types of software defects. There is a particular correlation between defect type information and software QA activities. The statistical characteristics of software defect type information reflect the validity of software QA activities.

According to Table 37.1, the rough variation tendency in statistics of different types of software defects can be obtained. Then the development tendency of different types of software defects in theory can be determined.

To evaluate the software QA activities with defect information, necessary supportive information is needed. This information comes from the experiential

knowledge of field experts and the statistical characteristics information of historical software projects. This information is stored in software defect baseline database. It's really hard to get an absolute value of certain defect baseline item, but an acceptable float scope can be obtained, and we can do evaluation combined with other types of defect metrics.

To defect injection activities (design and coding), marking with ★ means current activity tends to inject specific types of defect, and not marking with ★ means current activity inject specific types of defect with low possibly. To defect detection activities (Design Review, Code Review, Unit Test, Function Test and System Test), marking with ★ means current activity can detect specific software defect more effectively, while not marking with ★ means it will detect specific software defect less effectively.

37.4 QA Based on the Feedback of Software Defect Analysis

With software defect information collected in software, firstly calculates the statistical distribution characteristics of software defects in different dimensionality of ODC framework, compares with the referential values of corresponding defect baseline items, and then analyzes the diversity between them. The analysis result can be used to evaluate the efficiency and effectiveness of software QA activities, so as to identify the potential weak links of them.

37.4.1 QA Activities Evaluation and Control

There is a potential baseline in ODC framework, the injection and detection probabilities of different types of software defects are different from one software phase to another, so the statistical characteristics of software defects with the software process carrying on reflect the character characteristic of the software development processes [5]. This means different QA activities' capacities to detect specific types of software defects vary from one to another. On the other hand, by checking the proportion of different types of defects detected by different QA activities, it's easy to deduce whether or not the QA activities are involved in the right time, and whether they are effective or efficient enough. This'll be helpful for the determination of potential weak points in QA activities, and determining the right directions of QA improvement.

Figure 37.2 shows the statistical distribution of different types of software defects detected in unit test, function test and system test phases. Normally in unit test phase, the proportion of "Assignment/Initialize" defects and that of "Checking" defects will be bigger relatively; in function test phase more "Function/Class" defects and "Algorithm/Method" defects will be detected, and while system test tends to detect defects of "Timing/Serialization" and

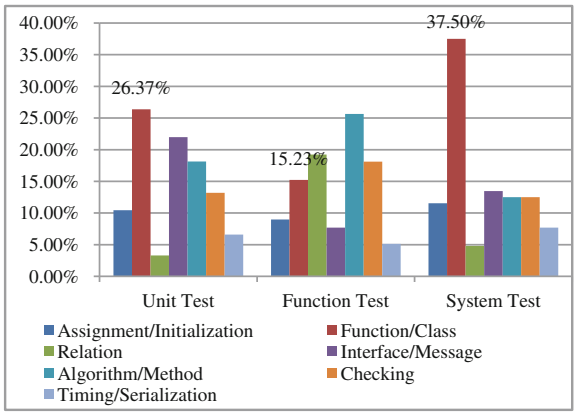

Fig. 37.2 The statistical distribution of different defect type

"Relationship". If the statistical distribution of different defect types in current project deviates too much from the desired defect baseline, some potential problems needs to be considered.

In Fig. 37.2 the proportion of software defect of "Function/Class" detected in function is less than that in system test, which should not have found so much "Function/Class" defect. This means the function test is not effective enough, causing that too much software defects of "Function/Class" type are omitted and leaked to the system test phase, so there are extraordinarily much defects of "Function/Class" type in this phase. But the emphasis of system test is not on the "Function/Class" defects, so there's high risk that "Function/Class" defects might omitted to the later phases.

The reason that causes the problem above is probably that the function test of the software is too insufficient to stage into the system test period, and the system test involves in too early. In practice, if the proportion of certain type of defects varies too much from the normal value in defect baseline, some adjustment to current project's certain QA activities should be done timely according to the feedback information from the defect analysis, so as to carry out QA activities more effectively. On the other hand, the result and experience gained from the current project's defect analysis will be supportive to the future projects, and helpful for the software QA activities improvement.

37.4.2 Defects Detection Efficiency Analysis

Researches about the efficiency of QA activities and cost-benefit shows that 80 % of software defects come from 20 % of the modules and 80 % of those are detected with 20 % of the QA work [6]. So if the QA activities are better targeted, the efficiency will be improved. To certain QA activity, the detection efficiency of different Triggers (from ODC) varies from one to another, and each trigger has its

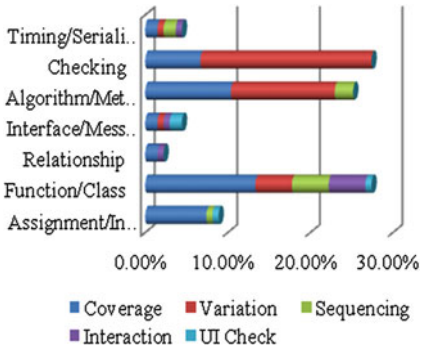

Fig. 37.3 The statistical distribution of defect triggers and types

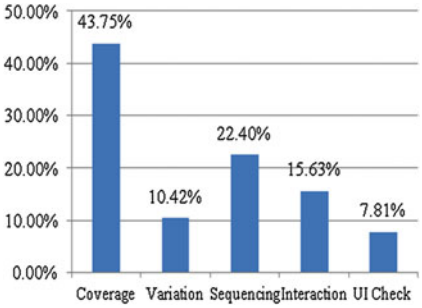

Fig. 37.4 The statistical distribution of test cases related to triggers

better targeted defect types. Figure 37.3 shows the statistical distribution of software defects detected with different triggers in the function test phase of one large-scale distributed data process software project. Figure 37.4 shows the statistical distribution of the test cases related to different triggers in this phase of the project.

From Figs. 37.3 and 37.4, it's easy to find the difference of each trigger's defect detection efficiency. Obviously the defect detection efficiency of "Variation" is 5–6 times of that of "Coverage". We need to ensure that in function test design phase a considerable proportion of test cases are "variation" related, so as to improve the efficiency of the test's defect detection. In reviewing the function test design, the statistical proportion information about the test cases related to different Triggers need to be considered. The proportions of test case related to different triggers can be compared with the values in defect baseline items, to determine whether or not the test design is sufficient enough and how to do some adjustment about the test design according to their deviation. The related reference value in the defect baselines can be obtained by combining experts' experience knowledge and the statistical features from historical projects.

37.5 Conclusions and Outlook

To establish software baseline, firstly build serious of software defect metrics using defect information collected form software development process based on Orthogonal Defect Classification, and then provide the result of software defect analysis as feedback information to software QA activities. The defect analysis is done according to the correlation between defect attribute information and software process in ODC framework. The evaluation of software QA activities' effectiveness and efficiency can be done based on software defect baseline, the result of which helps the control and improvement of QA activities. The feedback of defect analysis favors the identification of weak points in QA activities, serves as the basis of QA policy adjustment, and helps determine the direction of QA activities' improvement. This method is supportive for the improvement of the whole software process.

Supportive information stored in software defect database is needed to carry out effective QA activities based on defect baseline. So it's an important task to make full use of accumulated project data and experiences of domain experts to build more effective and trustworthy defect baseline database in the future. It is an iteratively forwarding process to do QA based on defect baseline, and the accumulation of historical information will be helpful for the accurate evaluation of current QA activities.

References

1. Wang DS, Gong YZ (2008) On software defects. Appl Res Comput 25(12):3531–3533
2. Chillarege R, Bhandari IS (1992) Orthogonal defect classification-a concept for in-process measurements. IEEE Trans Softw Eng 18:943–945
3. Chillarege R, Kao W L, Condit R G (1991) Defect type and its impact on the growth curve. In: Proceedings of the 13th International Conference on Software Engineering, pp 246–255
4. IBM Research Center for Software Engineering (2009) Details of ODC v5.11. http://www.research.ibm.com/softeng/ODC/DETODC.HTM
5. Freimut B, Denger C, Ketterer M (2005) An industrial case study of implementing and validating defect classification for process improvement and quality management. In: 11th IEEE international software metrics symposium (METRICS 2005), pp 10–19
6. Eman KE, Wieczorek I (1998) The repeatability of code defect classification. In: Proceedings of the international symposium on software reliability engineering, pp 322–333

Part IV
Trajectory, Orbits and Navigation

Chapter 38
Single-Station Orbit Determination with Astrometric Positioning and SLR Techniques

Guoping Chen, Xiaogong Hu, Yong Huang, Yong Yu, Zhenghong Tang, Zhongping Zhang and Yezhi Song

Abstract The principle of single-station orbit determination using astrometric positioning and SLR techniques is to add a laser ranging unit onto the optoelectronic telescope system, so as to conduct laser ranging and astrometric positioning angle measurement at the same time. By melting and processing all the data together, the new technique will achieve single-station orbit determination of the objects. In this paper, the actual measurement data of satellite Ajisai is analyzed for its single-station single-lap and multi-lap orbit determination. However, due to the limited amount of actual measurement data, objects of three types of orbit altitude have to be simulated and the single-station orbit determination analysis is just based on them in order to make the conclusion more universal. After studying the actual and simulated data, it turns out that for single-station single-lap orbit determination, the ranging measurement data (measurement duration longer than 4 min) will improve the orbit determination accuracy from several kilometers to tens of meters and the 24-h prediction accuracy from thousands of kilometers to several kilometers.

Keywords Single-station orbit determination · Precision orbit determination · Laser ranging measurement · Astrometric positioning

G. Chen (✉)
Graduate School of Chinese Academy of Sciences, Beijing 100049, China
e-mail: cgp365@126.com

G. Chen · X. Hu · Y. Huang · Y. Yu
Z. Tang · Z. Zhang · Y. Song
Shanghai Observatory of Chinese Academy of Sciences, Shanghai 200030, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_38,

38.1 Introduction

With the rapid development of space exploration, the increasing number of space debris is causing greater damage to the space vehicles working and to be launched and will pose a potential threat to the manned space flight. To face the threat, we should make full use of different techniques to closely monitor the space debris and based upon which, to undertake collision warning, avoidance and clearance of the space debris. Small objects on the high or medium orbits are usually less bright. According to some calculation, the brightness of the space objects with a diameter of 10 cm on the geostationary orbit is only 20 magnitudes [1] and they are only visible through large caliber optical telescopes. Laser radar has a high accuracy of ranging measurement, but to achieve diffused reflection laser ranging measurement of the low orbit objects, the radar must have a high power laser in coordination with a large caliber reception telescope.

Shanghai Observatory developed a new system which has a high power laser ranging unit added onto a large caliber optoelectronic telescope. The optoelectronic telescope can get the angle measurement data [2] with astrometric positioning while the laser can get the ranging measurement data [3]. This paper focuses on the impacts of high accuracy laser ranging measurement data on the accuracy of single-station single-lap or multi-lap orbit determination and prediction.

38.2 Single-Station Orbit Determination with Actual Measurement Data

In this paper, we used the actual laser ranging measurement data and the astrometric positioning angle measurement data of satellite Ajisai. Its obit altitude is 1485 km with the inclination of 50° and cycle time of 116 min. The actual laser ranging measurement data come from Sheshan laser observation station of Shanghai Observatory while the actual astrometric positioning angle measurement data come from two devices: one is the ranging measurement telescope with an optical lens in Sheshan, which measures the angle and ranging alternatively, the other is the rotating CCD drift scan optoelectronic equipment in Tianhuangping, Anji, Zhejiang. After some analysis, we find that there are five groups of angle and ranging measurement data, each of which can be catalogued into the same lap. See Table 38.1.[1]

In the above data, the longest duration of ranging measurement is 5.6 min, while the shortest is 0.4 min; the longest duration of angle measurement is 10.2 min, while the shortest is 5.1 min. The data of lap 1 and 2 come from Anji,

[1] In the following tables, we will replace all the angle measurement with AM and ranging measurement with RM.

Table 38.1 Actual ranging and angle measurement data of satellite Ajisai

Lap	Beginning time of RM (UTC)	Duration /(min)	Beginning time of AM (UTC)	Duration /(min)
1	2011.03.07 11:46:50	1.2	11:40:57	6.8
2	2011.03.30 11:16:51	4.8	11:12:29	7.5
3	2011.11.15 19:03:23	4.3	19:08:59	5.1
4	2011.11.15 21:03:25	5.6	21:09:09	10.2
5	2011.11.21 11:36:16	0.4	11:29:29	5.5

Table 38.2 Accuracy of actual angle measurement of satellite Ajisai

Lap	Azimuth accuracy Arc s	Elevation accuracy Arc s
1	5.3	5.5
2	4.7	5.7
3	16.7	13.7
4	13.6	9.5
5	9.7	8.1

Zhejiang, while that of lap 3, 4, 5 come from Sheshan. As the distance between Anji and Sheshan is relatively short, we will consider the angle and ranging measurement data come from the same station in the following analysis.

38.2.1 Accuracy of the Actual Measurement Data

Before using the above data, we must analyze their accuracy so as to set the corresponding weight when determine the orbit. According to website ILRS, from March 1st, 2011 to February 29th, 2012, for satellite Ajisai, the largest ranging measurement error of Sheshan station is 43.92 mm while the average error is 27.36 ± 6.35 mm.

By comparing the actual angle measurement data with the precise ephemeris of satellite Ajisai, we got Table 38.2. Precise ephemeris is achieved by precise orbit determination using global SLR data, and the ephemeris accuracy [4] of satellite Ajisai is about 1 m. The angle measurement data accuracy of lap 1 and 2 is 4.7–5.7 arc s, more precise than that of lap 3, 4 and 5, whose accuracy is 8.1–16.7 arc s. The main reason for this is that the former two groups of data is collected by the rotating drift scan equipment, which has a large diameter of 30 cm, and which locates in a better place (at night the sky is darker than magnitude 19). Besides, in the rotating drift scan mode, the star images of the reference star and the target can both keep round, so all of the above help to get a more precise astrometric positioning result.

Table 38.3 Actual measurement data single-station single-lap orbit determination of satellite Ajisai

Lap	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
		Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1	AM	356.0	400.4	250.0	443.4	3646.4	2.15E + 5	13844.0	2.15E + 5
	AM + RM	147.0	64.8	112.0	195.9	405.8	12850.0	1035.5	12872.0
2	AM	1855.7	1168.6	418.9	1970.1	69081.0	1.05E + 6	66727.0	1.06E + 6
	AM + RM	16.5	15.6	33.6	35.4	58.3	3363.5	173.8	3364.5
3	AM	2543.7	2547.0	1002.9	3693.1	3.37E + 5	2.15E + 6	134060.0	2.17E + 6
	AM + RM	10.4	20.9	44.8	50.5	32.6	2067.2	93.2	2068.5
4	AM	75.7	140.2	48.2	166.1	1191.7	71999.0	4571.2	72052.0
	AM + RM	20.0	37.5	68.9	73.2	27.3	1321.1	175.9	1323.4
5	AM	1126.3	1090.9	517.7	1651.3	94832.0	1243500.0	79964.0	1248500.0
	AM + RM	136.9	108.9	85.7	194.8	2142.2	120050.0	7860.5	120290.0

38.2.2 Single-Station Single-Lap Orbit Determination with Actual Measurement Data

The single-station single-lap orbit determination of satellite Ajisai with actual measurement data uses the numerical method [5]: gravity field uses JGM-3 model, 20 × 20 order; Earth tide perturbation; Atmospheric drag perturbation, atmospheric drag coefficient Cd: 2.2, mass: 685 kg, area: 3.63 m^2, atmospheric density mode: Jacchia-77; Light pressure perturbation, coefficient Cr: 1.0; Lunisolar gravitational perturbation; Elevation mask: 10° without calculating the drag coefficient of atmosphere and light pressure [6]. By comparing the results with the precise ephemeris of satellite Ajisai, we got Table 38.3.

From Table 38.3 we find that when determine the orbit with only angle measurement data, the accuracy is from 166.1 to 3693.1 m. According to the duration in Table 38.1, lap 4 has the longest angle measurement duration of 10.2 min and the highest orbit determination accuracy of 166.1 m, while lap 3 has the shortest angle measurement duration of 5.1 min and the lowest accuracy of 3693.1 m. It's the same with 24-h prediction accuracy: lap 4 has the highest accuracy of 72 km while lap 3 has the lowest accuracy of 2175 km. So the duration of angle measurement is closely related to the accuracy.

When we combine the angle and ranging measurement data together, both orbit determination accuracy and 24-h prediction accuracy are greatly improved. The orbit determination accuracy of lap 2, 3, 4 is higher than 80 m, while the 24-h prediction accuracy higher than 3.5 km. However, in lap 1 and 5, the orbit determination accuracy is higher than 200 m, but the 24-h prediction error of lap 1 is nearly 13 km while that of lap 5 is as large as 120 km. By studying the data in Table 38.1, it turns out that the ranging measurement for lap 1 lasts only 1.2 min and even worse for lap 5, which only lasts 0.4 min. On the contrary, lap 2, 3 and 4 all last longer than 4.3 min. So we can conclude that the highly accurate ranging

Table 38.4 Actual measurement data of single-station multi-lap orbit determination of satellite Ajisai

Lap	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
		Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
3 + 4	AM	31.1	49.5	13.9	50.8	44.4	348.9	24.1	349.0
	AM + RM	2.8	2.9	2.9	5.0	4.8	18.1	4.0	18.1

measurement data improved the accuracy of orbit determination and the longer we measure the ranging, the better the result will be.

No matter using the angle measurement data only or the combined data, the 24-h prediction errors appear mainly in direction T, namely the along track direction.

38.2.3 Single-Station Multi-Lap Orbit Determination with Actual Measurement Data

In Table 38.1, the data of lap 3 and 4 belongs to the same day and can be used to determine the orbits, while others are not suitable for multi-lap orbit determination as they belong to different days. The settings are the same as those in Sect. 38.2.2. Usually, in multi-lap orbit determination we should calculate atmospheric and light pressure, but in this case we skipped that because the time span between the two laps is short, only 2 h. Besides, according to our tests, if we calculate the atmospheric and light pressure here, it will lead to greater errors. Whatever, this is just a special case. Generally it's better to calculate the atmospheric and light pressure when we can measure the data for a relatively long time. By comparing the results with the precise ephemeris of satellite Ajisai, we got Table 38.4.

According to Table 38.4, if we determine the orbit using the angle measurement data from two laps, the orbit determination accuracy is about 50 m and the 24-h prediction accuracy is higher than 350 m. By some comparison, we find that the 24-h prediction accuracy in Table 38.4 is much higher than that in Table 38.3, namely, angle measurement data from two laps is more precise than the combined angle and ranging measurement data from a single lap. The reason for this is that data from two laps can better describe satellite orbits.

In Table 38.4, if we combine angle and ranging measurement data together, the orbit determination accuracy is higher than 5 m and the 24-h prediction accuracy is higher than 20 m. There are two reasons for this. On the one hand, the precise ranging measurement data improved the accuracy; on the other hand, the altitude of satellite Ajisai is 1485 km and is less influenced by the atmosphere, so the error is smaller.

No matter using the angle measurement data only or the combined data, the 24-h prediction errors appear mainly in direction T.

38.3 Single-Station Orbit Determination with Simulated Data

As we can only get limited amount of actual measurement data, we simulated objects of three different altitudes to make the analysis more universal: 1,500, 800 and 400 km, all the inclination is 80°.

38.3.1 Simulation Conditions

Station: Sheshan, Shanghai; Elevation noise: 2 arc s; Elevation system error: 1 arc s; Azimuth noise: 2 arc s; Azimuth system error: 1 arc s; Ranging measurement noise: 1.0 m; Ranging measurement system error: 0.4 m; Elevation mask: 10°; Observation conditions: do not observe earth shadow, do not observe in daytime.

Dynamic model: gravity field uses JGM-3 model, 20 × 20 order; Earth tide perturbation; Atmospheric drag perturbation, atmospheric drag coefficient Cd: 2.2, mass: 1000 kg, area: 20 m^2, atmospheric density mode: Jacchia-77; Light pressure perturbation, coefficient Cr: 1.0; Lunisolar gravitational perturbation.

There are two types of orbit determination using simulated data: single-station single-lap and single-station multi-lap. In each type, there are also two ways: angle measurement data determination and determination combining angle and ranging measurement data. Both ways use numerical method. Settings for single-station single-lap orbit determination are as follows. Gravity field uses JGM-3 model, 20 × 20 order; Earth tide perturbation; Atmospheric drag perturbation, atmospheric drag coefficient Cd: 2.2, mass: 1000 kg, area: 20 m^2, atmospheric density mode: Jacchia-77; Light pressure perturbation, coefficient Cr: 1.0; Lunisolar gravitational perturbation; Elevation mask: 10° without calculating the drag coefficient of atmosphere and light pressure. The settings for single-station multi-lap orbit determination are almost the same except that the drag coefficient of atmosphere and light pressure are counted.

Under the simulated conditions, the object whose altitude is 1500 km can be observed 16 laps in 4 days; the 800 km object can be observed 9 laps in 7 days and the 400 km object, without observing the earth shadow, can be observed 4 laps in 4 days. Considering the factors such as the duration of observation and the lifting and dropping section, for each simulated object, we analyzed its data from 4 laps.

38.3.2 Single-Station Single-Lap Orbit Determination with Simulated Data

The orbit cycle of 1500 km object is 116 min and the orbit determination result is shown in Table 38.5.

Table 38.5 Single-station single-lap orbit determination for the simulated 1500 km object

Lap	Duration (min)	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
			Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1	16.0	AM	49.1	37.4	11.2	59.1	579.0	30994.0	2438.4	31084.0
		AM + RM	5.5	3.2	5.8	8.5	14.3	383.8	27.1	384.6
2	13.3	AM	57.9	44.9	49.9	84.5	726.0	43585.0	3381.0	43711.0
		AM + RM	4.1	5.7	2.4	7.2	10.0	122.6	13.6	123.4
3	9.0	AM	252.7	148.7	149.7	326.4	2328.9	1.40E + 5	11366.0	1.41E + 5
		AM + RM	1.3	2.7	1.4	3.2	20.5	1099.1	85.6	1102.5
4	11.7	AM	45.5	23.4	59.7	76.7	474.0	27528.0	2092.2	27533.0
		AM + RM	9.0	6.5	3.0	11.5	26.6	445.9	34.9	447.0

Table 38.6 Single-station single-lap orbit determination for the simulated 800 km object

Lap	Duration (min)	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
			Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1	5.0	AM	126.9	83.3	191.0	244.0	4753.3	1.81E + 5	12626.0	1.82E + 5
		AM + RM	3.5	0.6	2.6	4.4	17.1	1032.0	74.8	1033.0
2	10.3	AM	42.4	47.7	16.8	64.4	730.4	49656.0	3440.0	49679.0
		AM + RM	2.9	0.7	7.8	8.3	12.7	515.9	49.1	515.9
3	9.3	AM	26.7	31.0	13.1	36.5	484.7	29914.0	2073.9	29942.0
		AM + RM	5.7	4.9	7.0	7.9	16.0	779.8	63.6	781.5
4	6.0	AM	29.0	21.9	43.7	56.8	590.7	39159.0	2813.4	39200.0
		AM + RM	4.0	2.7	0.9	4.9	15.0	273.0	19.4	273.6

In Table 38.5, the orbit determination accuracy of single-station single-lap angle measurement data ranges from 59.1 to 326.4 m. The duration for lap 1 is 16.0 min and orbit determination accuracy is 59.1 m while 24-h prediction accuracy is 31 km. The duration for lap 3 is 9.0 min and orbit determination accuracy is 326.4 m while 24-h prediction accuracy is 141 km. So we can conclude that the accuracy is related to the duration: the longer we observe the higher accuracy we will get.

When we combine the angle and ranging measurement data together, the orbit determination accuracy ranges from 3.2 to 11.5 m while 24-h prediction accuracy ranges from 123.4 to 1102.5 m. The main prediction errors appear in direction T. The ranging measurement data greatly improved the accuracy of orbit determination and prediction.

The orbit cycle of 800 km object is 101 min and the orbit determination result is shown in Table 38.6.

In Table 38.6, the orbit determination accuracy of single-station single-lap angle measurement data ranges from 36.5 to 244.0 m while the 24-h prediction accuracy ranges from 29.9 to 181.8 km. When ranging measurement data is combined, the orbit determination accuracy improved from 4.4 to 8.3 m and the 24-h prediction accuracy improved from 273.6 to 1033.0 m. The main prediction errors appear in direction T.

Table 38.7 Single-station single-lap orbit determination for the simulated 400 km object

Lap	Duration (min)	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
			Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1	5.0	AM	45.6	82.8	41.0	101.8	3742.3	1.54E + 5	9644.2	1.55E + 5
		AM + RM	5.1	1.6	7.3	9.1	54.1	3000.6	178.8	3006.0
2	5.0	AM	6.0	18.5	16.4	25.4	394.8	29214.0	1783.5	29270.0
		AM + RM	4.1	0.5	3.2	5.2	21.2	174.0	18.9	174.1
3	5.0	AM	25.9	33.7	20.9	47.4	918.4	68145.0	4256.6	68280.0
		AM + RM	5.0	1.2	5.9	7.8	35.8	1691.1	98.9	1694.0
4	5.3	AM	25.8	38.0	21.8	50.9	1012.2	59291.0	3623.4	59410.0
		AM + RM	0.9	0.4	3.2	3.4	6.3	435.6	25.7	436.4

Table 38.8 Single-station multi-lap orbit determination for the simulated 1500 km object

Lap	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
		Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1 + 2 + 3	AM	6.5	3.3	4.1	8.3	19.9	72.3	10.8	72.7
	AM + RM	2.8	3.0	1.1	4.0	5.3	21.2	2.6	21.2
2 + 3 + 4	AM	5.5	4.0	3.4	6.3	22.1	98.1	14.2	98.4
	AM + RM	4.3	4.7	0.7	6.3	2.5	16.7	1.7	16.7

The orbit cycle of 400 km object is 93 min and the orbit determination result is shown in Table 38.7.

In Table 38.7, the orbit determination accuracy of single-station single-lap angle measurement data ranges from 25.4 to 101.8 m while the 24-h prediction accuracy ranges from 29.2 to 155.0 km. When ranging measurement data is combined, the orbit determination accuracy improved from 3.4 to 9.1 m and the 24-h prediction accuracy improved from 174.1 to 3006.0 m. The main prediction errors appear in direction T. There is one thing to point out, when the object altitude is lower, the real atmospheric condition becomes more complicated. We cannot fully simulate the atmospheric environment. Besides, we use the same atmosphere model when we simulated the observation data and determine the orbits. So the real 24-h prediction accuracy will be worse than the simulated one.

38.3.3 Single-Station Multi-Lap Orbit Determination with Simulated Data

Determine the orbit using the single-station multi-lap data of objects 1500, 800 and 400 km, with the atmospheric and light pressure calculated, we got the following results.

In Table 38.8, the orbit determination accuracy of single-station multi-lap angle measurement data ranges from 6.3 to 8.3 m while the 24-h prediction accuracy

Table 38.9 Single-station multi-lap orbit determination for the simulated 800 km object

Lap	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
		Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1 + 2 + 3	AM	6.1	2.3	4.7	8.0	5.2	23.7	7.1	23.9
	AM + RM	1.6	0.9	0.9	1.9	1.1	8.7	1.2	8.8
2 + 3 + 4	AM	11.3	6.4	5.4	11.8	20.9	59.5	7.5	59.5
	AM + RM	2.2	1.6	1.3	2.6	3.8	20.4	2.2	20.4

Table 38.10 Single-station multi-lap orbit determination for the simulated 400 km object

Lap	Orbit determination type	Orbit determination accuracy/m				24-h prediction accuracy/m			
		Direction R	Direction T	Direction N	3D position	Direction R	Direction T	Direction N	3D position
1 + 2 + 3	AM	5.2	2.6	11.9	13.0	12.0	41.9	32.1	44.2
	AM + RM	1.1	0.6	0.3	1.1	3.6	14.9	1.6	15.0
2 + 3 + 4	AM	5.3	3.2	1.3	6.0	16.7	70.4	4.5	70.6
	AM + RM	2.1	1.0	1.5	2.3	3.9	15.4	3.7	15.4

ranges from 72.7 to 98.4 m. When ranging measurement data is combined, the orbit determination accuracy improved to 4.0 to 6.3 m and the 24-h prediction accuracy improved to 16.7 to 21.2 m. The main prediction errors appear in direction T. Compared to the results in Table 38.5, the accuracy of multi-lap orbit determination and 24-h prediction is usually higher than the single-lap one.

In Table 38.9, the orbit determination accuracy of single-station multi-lap angle measurement data ranges from 8.0 to 11.8 m while the 24-h prediction accuracy ranges from 23.9 to 59.5 m. When ranging measurement data is combined, the orbit determination accuracy improved to 1.9 to 2.6 m and the 24-h prediction accuracy improved to 8.8 to 20.4 m. The main prediction errors appear in direction T. Compared with the results in Table 38.6, the accuracy of multi-lap orbit determination and 24-h prediction is higher than the single-lap one.

In Table 38.10, the orbit determination accuracy of single-station multi-lap angle measurement data ranges from 6.0 to 13.0 m while the 24-h prediction accuracy ranges from 44.2 to 70.6 m. When ranging measurement data is combined, the orbit determination accuracy improved to 1.1 to 2.3 m and the 24-h prediction accuracy improved to 15.0 to 15.4 m. The main prediction errors appear in direction T. Compared with the results in Table 38.7, the accuracy of multi-lap orbit determination and 24-h prediction is higher than the single-lap one. Due to the same reason in Sect. 38.3.2, the real prediction accuracy will be worse than the simulated one.

38.4 Conclusions

The single-station orbit determination of satellite Ajisai with actual measurement data shows: (i) For single-station single-lap orbit determination, the ranging measurement data (measurement duration longer than 4 min), will improve the orbit determination accuracy from several kilometers to tens of meters and the 24-h prediction accuracy from thousands of kilometers to several kilometers. (ii) The short-time ranging measurement data (measurement duration shorter than 1 min), will improve the orbit determination accuracy to within 200 m. (iii) For single-station two-lap orbit determination, the ranging measurement data will improve the orbit determination accuracy from 50 to 5 m and the 24-h prediction accuracy from 350 to 18 m. (iv) The accuracy of single-station multi-lap orbit determination and prediction is higher than that of single-station single-lap. (v) The main prediction errors appear in direction T.

The simulated data of objects 1500, 800 and 400 km shows: (i) For single-station single-lap orbit determination, the ranging measurement data will improve the orbit determination accuracy from hundreds of meters to within 12 m and the 24-h prediction accuracy from more than 100 km to within 3 km. (ii) For single-station multi-lap orbit determination, the ranging measurement data will improve the orbit determination accuracy from tens of meters to within 7 m and the 24-h prediction accuracy from nearly 100 m to within 22 m. (iii) The accuracy of single-station multi-lap orbit determination and prediction is higher than that of single-station single-lap. (iv) The main prediction errors appear in direction T.

References

1. Wu LD (2011) Orbit and detection of man-made satellites and space debris. China Science and Technology Press, Beijing
2. Zhao M (2006) Introduction to astrometry. China Science and Technology Press, Beijing
3. Yu Y, Mao YD, Li Y et al (2010) Analysis of the Shanghai observatory 30 cm rotating CCD drift scan telescope astrometric accuracy. Chinese Academy of Sciences Shanghai Observatory Annuals 31
4. Chen GP, He B, Zhang ZB et al (2010) Analysis of the CPF ephemeris accuracy. Chinese Academy of Science Shanghai Observatory Annuals 31
5. Ye SH, Huang C (2000) Astrogeodynamics. Shandong Science and Technology Press, Jinan
6. Li JS (1995) Precise orbit determination of man-made satellite. PLA Publishing House, Beijing

Chapter 39
On Nominal Formation Flying Orbit with a Small Solar System Body

Yuhui Zhao, Shoucun Hu, Xiyun Hou and Lin Liu

Abstract It is very difficult for an explorer to orbit a low-mass and irregularly shaped small body in the solar system. In this case, formation flying with the small body is a workable solution. This paper discusses two strategies of formation flying based on two different dynamic models. The orbit resulted from the C–W equation and the halo orbit around libration point L1 in the Circular Restricted Three-body Problem (CRTBP) are considered as nominal orbit respectively. Numerical simulation indicates the effect of the magnitude of μ on the stability and other features of the halo orbits, where μ is a parameter weighing the gravity of the small body. The result shows that the CRTBP is more fuel-saving and therefore a more appropriate dynamic model for solving the formation flying problem. This paper also works out a dynamic model involving solar radiation pressure. Simulation result shows that, in this condition, the C–W equation has no significant advantage.

Keywords Formation flying · Nominal orbit · Halo orbit · C–W equation · Solar radiation pressure

39.1 Introduction

In 2010, the small body explorer Hayabusa of JAXA completed its journey and returned to the Earth, and its success drew increasing worldwide attention to small body exploration, which requires long period observations to obtain detailed

Y. Zhao (✉) · X. Hou · L. Liu
Astronomy Department, Nanjing University, Nanjing 210093, China
e-mail: zhaoyuhui@mail.nju.edu.cn

Y. Zhao · S. Hu
Purple Mountain Observatory, The Chinese Academy of Sciences, Nanjing 210093, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_39,

information. However, it is very difficult for an explorer to orbit a low-mass and irregularly shaped small body. In this case, fly-by and formation flying are two alternatives to explore small bodies.

This paper studies the dynamics and technology of formation flying with small bodies in the solar system. For the motions of an explorer in the solar system, the two most frequently used dynamic models are the perturbed two-body problem and the perturbed circular restricted three-body problem. The first one is usually used for circling orbits in the gravity field of a central body such as the sun or other planets while the second is more applicable for interplanetary cruise in deep space explorations.

Generally speaking, most small bodies in the solar system are small and light weighted. To solve the problem of formation flying with this kind of celestial bodies, different strategies should be adopted under different conditions: if the gravity of the small body can be neglected, C–W equation used in formation flying is applicable; if its gravity cannot be ignored, the problem can be solved based on the dynamics of motions around libration points in the CRTBP.

This paper mainly studies the dynamics and station keeping strategies of collinear libration point L1 of CRTBP with small μ, which is applicable in the case of formation flying with small bodies, and figures out the features of the halo orbits. Orbit control for station keeping of the spacecraft is also discussed and the relationship between energy consumption and the parameter μ is presented. Our work shows that formation flying with a small body on the basis of halo orbit around libration point is feasible in small body explorations. This strategy, in some cases, consumes less energy for orbit control than that of the strategy based on C–W equations. However, if the influence of solar radiation pressure on the C–W equation type of formation flying is taken into consideration, we will find its great influence on the orbit configuration of C–W equation and in this case the C–W equation type of formation flying does not have any advantage.

39.2 Dynamic Model of CRTBP

In the dynamic model of circular restricted three-body problem of the sun, a small body and a spacecraft, assuming that the mass of two primaries are m_1 and m_2 ($m_1 > m_2$), the following normalized unit can be used:

$$[M] = m_1 + m_2, \quad [L] = d, \quad [T] = \left(d^3/\mu\right)^{\frac{1}{2}} = \frac{1}{n} \tag{39.1}$$

where d is the distance between the sun and the small body, n is the angular velocity of the relative motion of the two primaries, $\mu = \frac{m_2}{m_1+m_2}$. In the rotating frame, the two major bodies lie on two points on the x-axis at $(-\mu, 0)$ and $(1-\mu, 0)$ respectively. Assuming that the mass of the spacecraft is m and its position is $\vec{r} = (x, y, z)^T$, the equation of motion can be written as [1–3]:

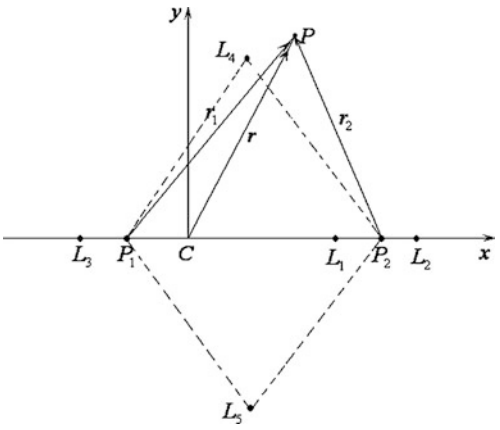

Fig. 39.1 Rotating frame and five libration points

$$\begin{cases} \ddot{\vec{r}} + 2(-\dot{y}, \dot{x}, 0)^T = (\partial\Omega/\partial\vec{r})^T \\ \Omega(x, y, z) = (\mu(1-\mu) + x^2 + y^2)/2 + (1-\mu)/r_1 + \mu/r_2 \end{cases} \tag{39.2}$$

where $\mu = m_2/(m_1 + m_2)$, r_1, r_2 are respectively the distances between the spacecraft and the two primaries. There are five stationary positions in this dynamic system, namely, libration points L_i including three collinear points on the x-axis and two triangular points in the x–y plan, all of which are depicted in Fig. 39.1.

Taking collinear libration point L_1 as an example, the position of L_1 can be denoted as $\vec{r}_1 = (1 - \mu + \gamma_1, 0, 0)^T$, and the relative position of the small body is $\vec{r} = (\xi, \eta, \zeta)^T$. From the conversion $\vec{R} = \vec{r}_1 + \vec{r}$, the linear equation of motion obtained from Eq. (39.2) is:

$$\begin{cases} \ddot{\xi} - 2\dot{\eta} - (1 + 2c_2)\xi = 0 \\ \ddot{\eta} + 2\dot{\xi} - (1 - c_2)\eta = 0 \\ \ddot{\zeta} + c_2\zeta = 0 \end{cases} \tag{39.3}$$

Where $c_2 = \mu/r_{11}^3 + (1-\mu)/r_{12}^3$,$r_{11}, r_{12}$ are respectively the distances between L_1 and the two primaries. In the sense of linear approximation, the equation of motion can be depicted as:

$$\begin{cases} \xi = C_1 e^{d_1 t} + C_2 e^{-d_1 t} + C_3 \cos d_2 t + C_4 \sin d_2 t \\ \eta = \alpha_1 C_1 e^{d_1 t} - \alpha_1 C_2 e^{-d_1 t} - \alpha_2 C_3 \sin d_2 t + \alpha_2 C_4 \cos d_2 t \\ \zeta = C_5 \cos d_3 t + C_6 \sin d_3 t \end{cases} \tag{39.4}$$

The solution to this equation is unstable through an exponential type defocusing while it's conditionally stable when $C_1 = C_2 = 0$. Therefore, the initial conditions for the conditionally stable solution are [1–3]:

$$\dot{\xi}_0 = \left(\frac{d_2^2}{\alpha_2 d_2}\right)\eta_0\,, \quad \dot{\eta}_0 = -(\alpha_2 d_2)\xi_0 \tag{39.5}$$

Then the solution can be written as:

$$\begin{cases} \xi = \xi_0 \cos d_2 t + \eta_0 \sin d_2 t/\alpha_2 \\ \eta = -\alpha_2 \xi_0 \sin d_2 t + \eta_0 \cos d_2 t \\ \zeta = C_5 \cos d_3 t + C_6 \sin d_3 t \end{cases} \tag{39.6}$$

In the real dynamic model there exist various kinds of perturbations, which have a significant effect on the orbit. Meanwhile, the neglected higher order terms of Eq. (39.4) also weaken the stability of the periodic orbits. In this case, a more stable nominal orbit under a more accurate dynamic model is required.

Richardson worked out the third-order approximation solutions to the kinetic equations for halo orbits, providing a more accurate orbit in the real dynamic model. An accurate quasi-halo orbit of CRTBP can also be worked out using numerical methods based on multiple shooting algorithm.

39.3 C–W Equation

For most small bodies in the solar system, the parameter μ of the three-body problem is very small. Under this circumstance, the dynamic model of the restricted two-body problem ($\mu = 0$) may be a better choice. In the latter model, the origin of the rotating frame in Fig. 39.1 moves to the center of the Sun, and the positions of L_1 and L_2 coincide. Therefore, the position of L_1 becomes $\vec{r}_1 = (1, 0, 0)^T$, and the linear equation of the motion can be written as:

$$\begin{cases} \ddot{\xi} - 2\dot{\eta} = 3\xi \\ \ddot{\eta} + 2\dot{\xi} = 0 \\ \ddot{\zeta} + \zeta = 0 \end{cases} \tag{39.7}$$

This equation is called C–W equation. The solution to this equation is [4]:

$$\begin{cases} \xi = -\frac{3}{2}C_2 + \frac{1}{2}C_3 \sin t - \frac{1}{2}C_4 \cos t \\ \dot{\xi} = \frac{1}{2}C_3 \cos t + \frac{1}{2}C_4 \sin t \\ \eta = C_1 + C_2 t + C_3 \cos t + C_4 \sin t \\ \dot{\eta} = C_2 - C_3 \sin t + C_4 \cos t \end{cases} \tag{39.8}$$

The initial conditions for the periodic solution are:

$$\dot{\xi}_0 = \eta_0/2\,, \quad \dot{\eta}_0 = -2\xi_0 \tag{39.9}$$

This is the mechanism of formation flying of satellites of the earth, and it may also be used in formation flying with very small bodies in the solar system. This is

the mechanism of formation flying of satellites of the earth, and it may also be used in formation flying with very small bodies in the solar system.

39.4 Numerical Simulation and Results

In this part, we adopt the theta-D control method [5–7] to solve the problem of station keeping in formation flying with the small body based on respectively the two different dynamic models.

39.4.1 Results Based on Halo Orbit Control

For near-earth small bodies, assuming that the semi major axis of their revolution 1AU, the energy consumption for station keeping can be worked out using the control strategies mentioned above(taking the halo orbit around L_1 point of the CRTBP composed of the sun, the small body and the spacecraft as an example). Assuming that the amplitude of halo orbit in z-direction is $0.155\gamma_1$, differential correction method and numerical simulations are used to calculate the nominal orbits for different μ see Fig. 39.2. The variation of amplitude in x-direction due to μ is shown in Fig. 39.3a. The approximate 10 years' energy consumption for station keeping in these orbits is shown in Fig. 39.3b.

As μ decreases, both the amplitude of halo orbits and the energy consumption for orbit control will decrease. If the ratio of the amplitude in x-direction to γ_1 keeps unchanged (=0.155), the periods of halo orbits are almost the same (about 190 days). As a result of the stability of CRTBP and the dynamics of halo orbit formation, nominal halo orbits do not exist if the ratio is too large or too small. Therefore, different amplitude should be considered for different μ in nominal orbit design.

39.4.2 Comparison of the Results of Different Strategies

Taking the near-earth small body 2005TF49 as an example, the formation flying orbits for an explorer with this small body are designed using respectively halo orbit calculated in numerical ways and C–W equation. This small body has a semi major axis of 155866014.57599 km and a gravity of about 10^{-19} to the sun. The amplitude in x-direction of the halo orbit around L_1 is about 7.78 km, while the distance between L_1 and the small body is about 50.16 km, which is also taken as the amplitude on x and y direction in orbit design with C–W equation. The amplitudes in z direction of both the halo orbit and the nominal orbit with

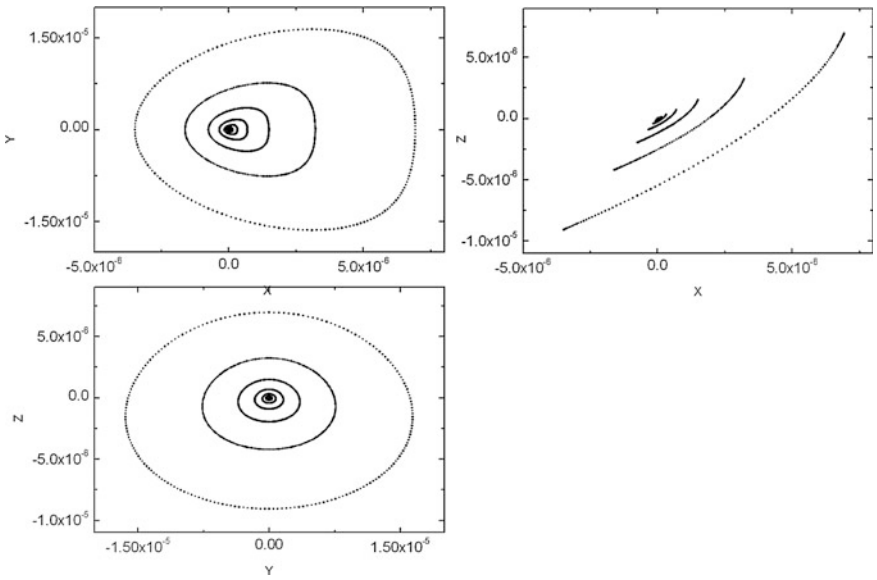

Fig. 39.2 Projection of halo orbits

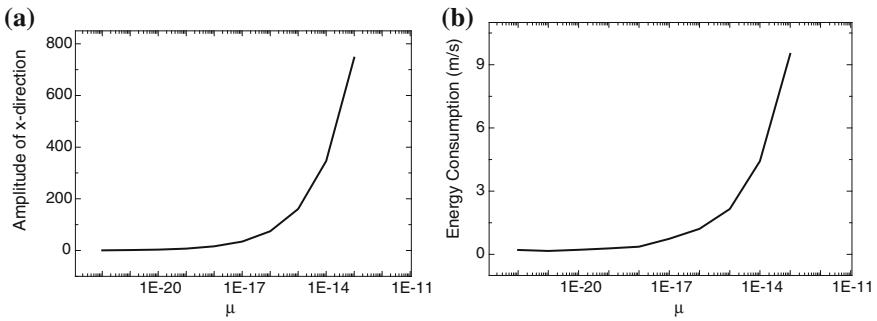

Fig. 39.3 **a** Variation of amplitude in X-direction of halo orbits due to μ; **b** variation of energy increment due to μ

C–W equation are all set at about 14 km. Figure 39.4 shows the nominal orbits resulted respectively from the two dynamic models. The variation of orbit control due to time is depicted in Fig. 39.5.

Given an initial error of 1×10^{-6} in the computation, the total cost in the former model is 0.277 m/s while it is more than 6 m/s using C–W equation. Comparison shows that it costs less for station keeping when the nominal orbit is computed in CRTBP.

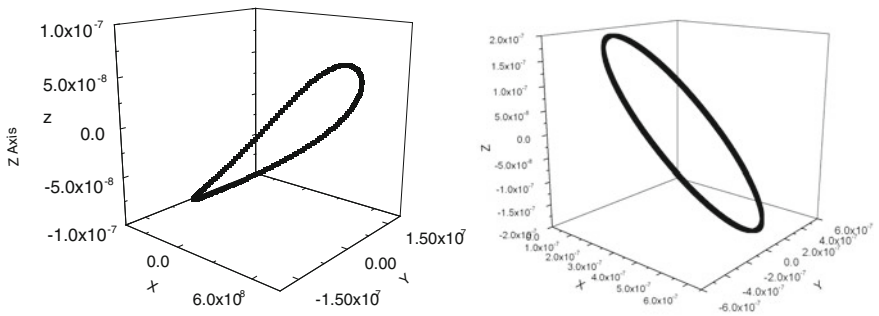

Fig. 39.4 **a** Nominal halo orbit; **b** nominal orbit resulted from C–W equation

39.5 The Effect of Solar Radiation Pressure

Simulation results above have not taken into consideration the effect of solar radiation pressure perturbation. The aim of the omission is to show the effect of the small body's gravity on the dynamics. However, solar radiation pressure perturbation is not negligible when the explorer formation flies with a small body in a real dynamic model. Solar radiation pressure is a kind of surface force. Since the area-to-mass ratio of the small solar system body is usually far less than that of the spacecraft, the effect of solar radiation pressure perturbation on the small body can be ignored while it has to be considered for that of the spacecraft. It is this differentiation that leads to the failure of the C–W equation configuration. This situation is quite different from that of the formation flying of two earth satellites, which have similar area-to-mass ratio and thus have similar effects by solar radiation pressure. The discussion above explains the different situations when applying C–W equation to formation flying of two earth satellites and to that of a spacecraft in formation flight with a small solar system body.

Considering the solar radiation pressure perturbation, the equation of motion of the explorer in the rotating frame can be depicted as:

$$\begin{cases} \ddot{\vec{r}} + 2(-\dot{y}, \dot{x}, 0)^T = (\partial\Omega/\partial\vec{r})^T + \varepsilon_{lt}\boldsymbol{\alpha} \\ \Omega(x, y, z) = (\mu(1-\mu) + x^2 + y^2)/2 + (1-\mu)/r_1 + \mu/r_2 \end{cases} \tag{39.10}$$

where $\boldsymbol{\alpha}$ is the nominal direction of solar radiation pressure perturbation, $\varepsilon_{lt} = \frac{(1+\eta)S}{m}\frac{\rho_{AU}\Delta_S^2}{r^2}$, η is the reflection factor of the explorer and $0 \leq \eta \leq 1$, S/m is the area to mass ratio. $\rho_{AU} = 4.5605 \times 10^{-6}$ N/m^2 is the solar radiation pressure at 1 AU. In this case, the form of Eq. (39.7) can be written as:

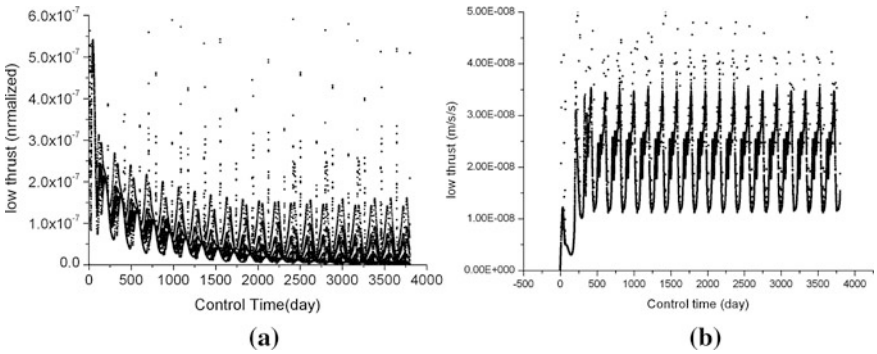

Fig. 39.5 **a** Variation of orbit control for nominal halo orbit due to time; **b** variation of orbit control for nominal orbit computed by C–W equation

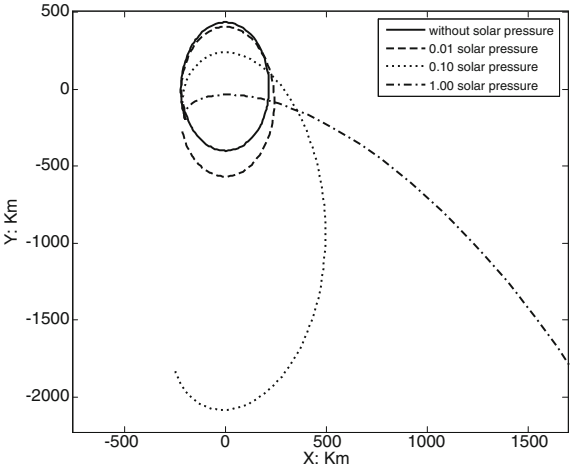

Fig. 39.6 The effect of solar radiation pressure on nominal orbit based on C–W equation

$$\begin{cases} \ddot{\xi} - 2\dot{\eta} = 3\xi + \varepsilon_{lt}\alpha_x \\ \ddot{\eta} + 2\dot{\xi} = \varepsilon_{lt}\alpha_y \\ \ddot{\zeta} + \zeta = \varepsilon_{lt}\alpha_z \end{cases} \tag{39.11}$$

Assuming that the area and the mass of the explorer are respectively 18 m^2 and 1500 kg, $\varepsilon_{lt} \sim 10^{-5}$ can be obtained when the normalized units in (39.1) are used. The condition to apply C–W equation is $\varepsilon_{lt} \ll \sqrt{\xi^2 + \eta^2 + \zeta^2} \ll 1$, that is, $\sqrt{\xi^2 + \eta^2 + \zeta^2} \gg 10^3$ km. However, for formation flying, the distance between the explorer and the target body is generally not allowed to be large for scientific purposes. A distance of 10–100 km is acceptable, but the effect of solar radiation pressure, in this case, will destroy the configuration of C–W equation (as illustrated in Fig. 39.6). Therefore, the application of C–W equation in formation flying problems is not preferable when solar radiation pressure is taken into consideration.

39.6 Conclusions

This paper studies formation flying with small bodies in the solar system. Both CRTBP and perturbed two-body problem are considered. Halo orbits around L_1 and orbits based on C–W equation are respectively applied to form nominal orbits. As μ decreases, the amplitude of halo orbits energy consumption for orbit control decreases. Because of the higher accuracy of its dynamic models, energy consumption for station keeping for halo orbits is less than that of the formation flying orbits using C–W equation. A dynamic model considering the solar radiation pressure is also studied while simulation results show that it has no significant advantage to use C–W equation in this case.

Acknowledgments This work was supported by national Natural Science Foundation of China (NSFC 10903002, 11033009).

References

1. Hou XY (2008) Dynamics and their applications of libration points. Doctoral thesis of Nanjing University, Nanjing
2. Koon WS, Lo MW, Marsden JE et al (2006) Dynamic systems, the three-body problem and space mission design. World Scientific, Berlin. ISBN 978061524095-4
3. Gomez G, Libre J, Martinez R et al (2001) Dynamics and mission design near libration points. World Scientific, Singapore
4. Liu L, Hu SJ et al (2005) An introduction of astrodynamics. Nanjing University Press, Nanjing
5. Ming X, Balakrishnan SN (2002) A new method for suboptimal control of a class of nonlinear systems. In: Proceedings of IEEE conference on decision and control, Las Vegas, Nevada, Dec 10–13
6. Ming X, Dancer MW, Balakrishnan SN et al (2004) Station keeping of an L2 libration point satellite with θ-D technique. In: Proceeding of the 2004 American control conference Boston, Massachusetts, June 30–July 2
7. Ming X, Balakrishnan SN, Stansbery DT et al (2004) Nonlinear missile autopilot design with theta-D technique. AIAA J Guid Control Dyn 27:406—417

Chapter 40
On Orbit Control Utilizing Solar Sails Around Asteroids

Shengxian Yu, Xiyun Hou and Lin Liu

Abstract The interest of exploring asteroids has been growing over these years due to their primordial origins and possible economic benefits. The gravities of the asteroids are mostly very small and irregular, so it is necessary to study the orbits around them with special treatment. In this paper, two kinds of nominal orbits for probes are firstly proposed. One kind is around the collinear libration points of the restricted three-body problem formed with the Sun, the asteroid and the probe. The other kind is the formation flying orbits around the asteroid. Due to the inevitable existence of errors in practice, station keeping of these two kinds of nominal orbits is necessary. Solar sail propulsion is used in this paper. Two control techniques including varying the solar sail area and varying the solar sail pitch and yaw angles are considered. The linear optimal feedback control law is used. Numerical simulations are made to both kinds of nominal orbits. The results show that the controlled orbits converge to the vicinity of the nominal orbits rapidly. The technique of varying the solar sail pitch and yaw angles is better than varying the solar sail area.

Keywords Asteroid · Solar sail · Collinear libration point · Station keeping

S. Yu (✉) · X. Hou · L. Liu
School of Astronomy and Space Science, Nanjing University,
Nanjing 210093, China
e-mail: yushx@nju.edu.cn

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_40,

40.1 Introduction

Much attention has been paid to the exploration of the asteroids due to their primordial origins and possible economic benefits. Till now, several missions have visited them [4, 20, 21]. The exploration ways include flying by, circling around or landing on the asteroids. In the paper, two possible nominal orbits for probes to explore the asteroids are proposed. One is the quasi-periodic orbits around the collinear libration points of the restricted three-body problem formed with the Sun, the asteroid and the probe. The other is the formation flying orbits around the asteroid. For the first kind of orbits, the collinear libration point L_1 is very close to the asteroids. Since the asteroids are usually geometrically irregular, the perturbation of the non-spherical gravitational terms cannot be simply neglected. Nevertheless, the distance between the point L_1 and the asteroid which increases once the solar pressure perturbation is taken into account. Therefore, in order to reduce the effect of the non-spherical gravitation perturbation, halo orbits around the point L_1 are considered in our work. Analytical representations of the halo orbits are firstly given in the simple circular restricted three-body problem and then numerically refined to obtain the true halo orbits in the real force model. For the second kind of nominal orbits, the distance between the probe and the asteroid is large enough so that the mass of the asteroid can be neglected. Therefore, the model of restricted two-body problem is used as the first approximation. This kind of exploration is similar to satellite formation flying. The deputy is the probe and the chief is the asteroid. Analytic solutions in the restricted two-body problem can be constructed and then numerically refined to obtain the true orbits in the real force model.

Both kinds of nominal orbits are unstable, especially for the first kind. Farquhar [5] suggested some station-keeping strategies for quasi-periodic orbits around the collinear libration points. Breakwell et al. [3] proposed a station-keeping method for spacecrafts moving around the halo orbits in the vicinity of the Earth–Moon libration point L_2. Howell and Pernicka [15] used the target point strategy for libration point trajectories. Gómez et al. [7] used the Floquet strategy which cancels the unstable component of the nominal orbit at the nodal points. Besides the ones mentioned, there are also some others [1, 6, 12, 17]. In our work, station-keeping with solar sails is considered. The linear optimal feedback strategy is used. It is similar to the one used by Breakwell et al. [3]. The difference is that the problem in this paper is of finite time. The station keeping strategy is applied to both kinds of nominal orbits and some numerical simulations are made.

In the paper, when we refer to the real force model, we mean the gravitational model of the nine planets (including Pluto), the Moon, the Sun, the asteroid and the solar pressure. The major bodies' motions are given by the ephemeris DE405. The asteroid's orbit is given by numerical integration by neglecting its mass. Denote the set of these bodies as $S = (\mu_1, \ldots, \mu_{11}, \mu_{12})$, where $\mu_1 - \mu_9$ indicate the reduced masses of the nine planets and $\mu_{10}, \mu_{11}, \mu_{12}$ indicate the reduced masses of the Moon, the Sun and asteroid. The mass unit is the sum of the masses of the Sun and the asteroid. The length unit is the mean distance between the Sun and the asteroid.

40.2 Halo Orbits Around the Collinear Libration Point

For asteroids with small masses such as Apophis, the collinear libration point L_1 under the influence of the solar pressure perturbation has a proper distance from the asteroid. A possible way for the probe to explore the asteroid is to locate around L_1. Similar to the Sun-Earth system and the Earth–Moon system, halo orbits around L_1 can be used as nominal orbits. In this section, analytical expressions of halo orbits in the restricted three-body problem will be introduced first and then improved numerically in the real force model. The asteroid Apophis is taken as an example. The non-spherical gravitational terms of the asteroid are not considered. Their effects will be discussed in the conclusion section.

40.2.1 The Positions of the Collinear Libration Points with Solar Pressure

The perturbation acceleration of the solar pressure is

$$\mathbf{F}_s = \beta \frac{1}{\Delta^2} \left(\hat{\Delta} \cdot \mathbf{n} \right)^2 \mathbf{n} \tag{40.1}$$

where $\hat{\Delta} = \boldsymbol{\Delta}/\Delta$, $\boldsymbol{\Delta} = \mathbf{r} - \mathbf{r}_s$, $\boldsymbol{r}$ and $\boldsymbol{r_s}$ are the position vectors of the probe and the Sun respectively. $\mathbf{n}$ is the normal direction vector of the solar sail. $\beta = (\kappa S/m)\rho_s \Delta_s^2$, $\Delta_s = 1$ AU, $\kappa = 1 + \eta$, η is the reflection coefficient of the solar sail, $\eta = 0.4$ in the paper. S/m is the ratio of the area to the mass of the solar sail, S/m=100 m^2/1 t is adopted, corresponding to $\beta = 1 \times 10^{-4}$ in dimensionless units. $\rho_s = 4.5605 \times 10^{-6}$ N/m^2 is the radiation pressure of the Sun at the distance of $\Delta_s = 1$ AU.

In the paper, the normal direction vector of the solar sail $\mathbf{n}$ is always paralleled to $\hat{\Delta}$ at the nominal orbits. The acceleration $\mathbf{F}_s$ is an inverse square force, just as the central gravitational force of the Sun, so the position of the collinear libration points L_1 and L_2 can be obtained by solving the following equation

$$\xi + q\frac{(1-\mu)}{(\xi-\mu)^2} \mp \frac{\mu}{(\xi+1-\mu)^2} = 0 \tag{40.2}$$

where $\mp$ indicate L_1 and L_2 respectively. ξ is the distance between the collinear libration point L_1 (L_2) and the asteroid. $q = 1 - \beta/GM_s$, $\beta = 1 \times 10^{-4}$ corresponds to $q = 0.999892333$.

The positions of L_1 and L_2 under the influence of the solar pressure (called new L_1 and new L_2 hereinafter) are obtained according to Eq. (40.2). They are listed in Table 40.1. Table 40.1 shows that the distance between L_1 (L_2) and the asteroid is very close to Apophis without solar pressure. Once the solar pressure is considered

Table 40.1 The distances of L_1 and L_2 from Apophis

	Without solar pressure ($q = 1$)	With solar pressure
L_1	$1.653975024 \times 10^{-7}$	$3.589004976 \times 10^{-5}$
L_2	$1.653975206 \times 10^{-7}$	$1.122655102 \times 10^{-8}$

(S/m=100 m^2/1t is adopted), the distance between L_1 and Apophis is two orders lager than before while L_2 is even closer to Apophis. Therefore in order to reduce the effect of Apophis' non-spherical gravitation perturbation, the nominal orbits around the new L_1 are considered.

40.2.2 Restricted Three-Body Problem

In a synodic coordinate centered at the center of masses, the equations of motion for the restricted three-body problem are [8]

$$\begin{cases} \ddot{X} - 2\dot{Y} = \Omega_X \\ \ddot{Y} + 2\dot{X} = \Omega_Y \\ \ddot{Z} + Z = \Omega_Z \end{cases} \tag{40.3}$$

where

$$\begin{aligned} \Omega &= \frac{1}{2}\left(X^2 + Y^2\right) + (1-\mu)r_1^{-1} + \mu r_2^{-1}, \\ r_1^2 &= (X-\mu)^2 + Y^2 + Z^2, \\ r_2^2 &= (X-\mu+1)^2 + Y^2 + Z^2. \end{aligned} \tag{40.4}$$

Moving the origin of the coordinate to the collinear libration point L_1, we have

$$x = -\frac{1}{\gamma}(X - \mu + 1 - \gamma), \quad y = -\frac{1}{\gamma}Y, \quad z = \frac{1}{\gamma}Z \tag{40.5}$$

where γ is the distance between the new L_1 and the asteroid. The analytical solutions of the motion equations in this new coordinate can be written in the following form [16]. Although the solar pressure perturbation is considered in the paper, the expansion form of the analytical solutions is the same.

$$\begin{cases} x(t) = \sum\limits_{N\geq 1,i,j\geq 0}^{\infty}\left(\sum\limits_{|k|\leq i,|m|\leq j} x_{ijkm}\cos(k\theta_1+m\theta_2)\right)\alpha^i\beta^j \\ y(t) = \sum\limits_{N\geq 1,i,j\geq 0}^{\infty}\left(\sum\limits_{|k|\leq i,|m|\leq j} y_{ijkm}\sin(k\theta_1+m\theta_2)\right)\alpha^i\beta^j \\ z(t) = \sum\limits_{N\geq 1,i,j\geq 0}^{\infty}\left(\sum\limits_{|k|\leq i,|m|\leq j} z_{ijkm}\cos(k\theta_1+m\theta_2)\right)\alpha^i\beta^j \end{cases} \tag{40.6}$$

where α is the amplitude in plane and β is the amplitude out of plane. $\theta_1 = \omega t + \phi_1$ and $\theta_2 = vt + \phi_2$. ϕ_1, ϕ_2 are arbitrarily chosen integration constants. $N = i + j$ is the order of the analytical solutions. The two frequencies ω and v are of the form

$$\omega = \sum_{i,j=0}^{\infty}\omega_{ij}\alpha^i\beta^j, \quad v = \sum_{i,j=0}^{\infty}v_{ij}\alpha^i\beta^j \tag{40.7}$$

Usually, ω and v are not rational with each other. In this case, we have the so called Lissajous orbits. But for large amplitude orbits, it's possible for $\omega = v$. In this case, we have the so called halo orbit, expressed in the following form.

$$\begin{cases} x(t) = \sum\limits_{N\geq 1,i,j\geq 0}^{\infty}\left(\sum\limits_{|k|\leq i} x_{ijk}\cos(k\theta)\right)\alpha^i\beta^j \\ y(t) = \sum\limits_{N\geq 1,i,j\geq 0}^{\infty}\left(\sum\limits_{|k|\leq i} y_{ijk}\sin(k\theta)\right)\alpha^i\beta^j \\ z(t) = \sum\limits_{N\geq 1,i,j\geq 0}^{\infty}\left(\sum\limits_{|k|\leq i} z_{ijk}\cos(k\theta)\right)\alpha^i\beta^j \end{cases} \tag{40.8}$$

where $\theta{=}\omega t + \phi$. ϕ is arbitrarily chosen integration constant. Analytical formulae of these two kinds of orbits can be obtained with the Lindstedt–Poincaré method [16]. In this paper, the third-order analytical solutions of halo orbits are considered.

For the asteroid Apophis, its mass is about $M_a = 2.7 \times 10^{10}$ kg. The dimensionless units of the restricted three-body problem consisting of it and the Sun are

$$\begin{cases} [L] = R_0 = 0.9223002432 \text{ AU}, \\ [M] = M_{sun} + M_a, \\ [T] = \left[R_0^3/G(M_{sun}+M_a)\right]^{1/2} \approx 51.490492199842 \text{ d} \end{cases} \tag{40.9}$$

40.2.3 The Real Force Model

For the real force model, various perturbations exist. Two prominent ones are gravitational perturbations from the orbit eccentricity of Apophis and other major bodies besides the Sun. In a synodic coordinate centered at one of the collinear libration points, the probe follows ([11, 13])

$$\ddot{\rho} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_s \tag{40.10}$$

where

$$\begin{cases} \mathbf{F}_1 = -2\mathbf{C}^T\dot{\mathbf{C}}\dot{\rho} - \mathbf{C}^T\ddot{\mathbf{C}}\rho - \mu_{11}\mathbf{r}/r^3 + \mu_{11}\mathbf{r}_0/r_0^3 \\ \qquad - \sum_{i\in S, i\neq 11} \mu_i\left(\boldsymbol{\delta}_i/\delta_i^3 - \boldsymbol{\delta}_i^0/\left(\delta_i^0\right)^3\right) \\ \mathbf{F}_2 = -\mu_{11}\mathbf{r}_0/r_0^3 - \sum_{i\in S, i\neq 11} \mu_i\left(\boldsymbol{\delta}_i^0/\left(\delta_i^0\right)^3 + \mathbf{r}_i/r_i^3\right) \\ \qquad - \ddot{\mathbf{r}}_0 - 2\mathbf{C}^T\dot{\mathbf{C}}\dot{\mathbf{r}}_0 - \mathbf{C}^T\ddot{\mathbf{C}}\mathbf{r}_0. \end{cases} \tag{40.11}$$

In the above equations, $\mathbf{r}, \mathbf{r}_i, \mathbf{r}_0$ are the vectors in the Sun-centered synodic coordinate. $\mathbf{r}$ is the position of the probe. $\mathbf{r}_i$ is the position of μ_i. $\mathbf{r}_0$ is the position of the collinear libration point. $\boldsymbol{\rho} = \mathbf{r} - \mathbf{r}_0$ is the position of the spacecraft from the collinear libration point. $\boldsymbol{\delta}_0^i = \mathbf{r}_0 - \mathbf{r}_i$ is the position of the collinear libration point from μ_i. $\mathbf{C}$ is the transformation matrix between the Sun-centered synodic coordinate and the Sun-centered sidereal coordinate.

Due to various perturbations, the analytical solutions expressed by Eq. (40.8) are only approximate orbits in the real system. In order to obtain the orbits in the real system, numerical refinement is used. The analytical solutions in the restricted three-body problem can be used as initial seeds. There are several methods of numerical refinement [9, 14, 19]. In the paper, the multiple shooting method mentioned in Ref. [9] is used.

The initial time is 0.0 h, Mar 14th, 2012, corresponding to JD = 2456000.5. The motion of Apophis is given by integration in the real system. The initial orbit elements of Apophis at time JD = 2455800.5 are

$$\begin{cases} a = 0.9223002432\ \text{AU}, & e = 0.1910762290, \\ i = 3.3319600435^\circ, & \Omega = 204.4304100445^\circ, \\ \omega = 126.4244766663^\circ, & M = 287.5823055950^\circ \end{cases} \tag{40.12}$$

Figure 40.1 shows the halo orbit in the real force model. The duration of this orbit is about 8.458 years.

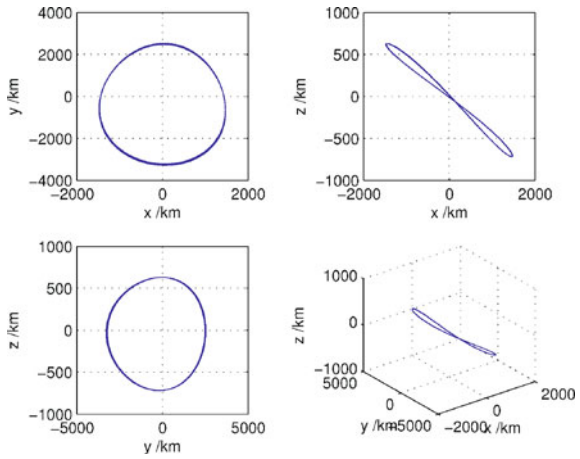

Fig. 40.1 The halo orbit around the new L_1 in the real force model for $\alpha = 1.08 \times 10^{-5} R_0$

40.3 Formation Flying Orbits Around the Asteroid

For most asteroids, their masses are very small compared with the Sun and thus can be neglected. In this case, the restricted three-body problem can be reduced to the restricted two-body problem. A possible way to explore the asteroid is to fly directly around it, like the satellite formation flying. The asteroid can be taken as the chief and the probe can be taken as the deputy.

40.3.1 Restricted Two-Body Problem

For the restricted two-body problem, the origin of the synodic coordinate is the asteroid. In this coordinate, the motion equations are

$$\begin{cases} \ddot{x} - 2\dot{y} = x + 1 - \dfrac{x+1}{r^3} \\ \ddot{y} + 2\dot{x} = y - \dfrac{y}{r^3} \\ \ddot{z} + z = -\dfrac{z}{r^3} \end{cases} \tag{40.13}$$

Expanding Eq. (40.13) around the asteroid and ignoring the nonlinear terms, we have

$$\ddot{x} - 2\dot{y} = 3x,\ \ddot{y} + 2\dot{x} = 0,\ \ddot{z} + z = 0 \tag{40.14}$$

Usually Eq. (40.14) is called C–W equation. If the initial values $x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0$ at t_0 satisfy

$$x_0,\ \ y_0,\ \ \dot{x}_0 = y_0/2,\ \ \dot{y}_0 = -2x_0,\ \ z_0,\ \ \dot{z}_0 = \pm(y_0/2x_0)z_0 \tag{40.15}$$

The solution of Eq. (40.14) is of the following form [18]

$$\begin{cases} x = x_0 \cos t + (y_0/2)\sin t, & \dot{x} = -x_0 \sin t + (y_0/2)\cos t \\ y = -2x_0 \sin t + y_0 \cos t, & \dot{y} = -2x_0 \cos t - y_0 \sin t \\ z = z_0 \cos t + \dot{z}_0 \sin t, & \dot{z} = -z_0 \sin t + \dot{z}_0 \cos t \end{cases} \tag{40.16}$$

Of course, different flying formation can be obtained by changing the initial values $t_0, x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0$. In the paper, $x_0 = y_0 = z_0 = 1 \times 10^{-4} R_0$ is used.

40.3.2 The Real Force Model

Again, the real force model is considered. The perturbations of the nine planets (including Pluto), the Moon and the solar pressure are considered. The asteroid's orbit is given by numerical integration in the real force model. In a synodic coordinate centered at an asteroid in the real system, the probe follows

$$\ddot{\rho} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_s \tag{40.17}$$

where

$$\begin{cases} \mathbf{F}_1 = -2\mathbf{C}^T\dot{\mathbf{C}}\dot{\rho} - \mathbf{C}^T\ddot{\mathbf{C}}\rho - \mu_{11}\mathbf{r}/r^3 + \mu_{11}\mathbf{r}_{12}/r_{12}^3 \\ \qquad - \sum_{i\in S, i\neq 11,12} \mu_i\left(\boldsymbol{\delta}_i/\delta_i^3 - \boldsymbol{\delta}_i^0/\left(\delta_i^0\right)^3\right) \\ \mathbf{F}_2 = -\mu_{11}\mathbf{r}_{12}/r_{12}^3 - \sum_{i\in S, i\neq 11,12} \mu_i\left(\boldsymbol{\delta}_i^0/\left(\delta_i^0\right)^3 + \mathbf{r}_i/r_i^3\right) \\ \qquad - \ddot{\mathbf{r}}_{12} - 2\mathbf{C}^T\dot{\mathbf{C}}\dot{\mathbf{r}}_{12} - \mathbf{C}^T\ddot{\mathbf{C}}\mathbf{r}_{12}. \end{cases} \tag{40.18}$$

In the above equations, $\mathbf{r}, \mathbf{r}_i$ are the vectors in the Sun-centered synodic coordinate. $\mathbf{r}$ is the position of the probe. $\mathbf{r}_i$ is the position of μ_i. $\boldsymbol{\rho} = \mathbf{r} - \mathbf{r}_{12}$ is the position of the probe from the asteroid. $\boldsymbol{\delta}_0^i = \mathbf{r}_{12} - \mathbf{r}_i$ is the position of the asteroid from μ_i. $\mathbf{C}$ is the transformation matrix between the Sun-centered synodic coordinate and the Sun-centered sidereal coordinate.

Due to the perturbations, the linear analytical orbits deviate from the real orbits. Using the analytical orbits as initial seeds, the real orbits in the real force model can be found with multiple shooting method. Figure 40.2 is the refined result. The duration of this orbit is about 8.458 years.

For the above orbit, we ignore the mass of Apophis. To see the effects of this approximation, we integrate the orbit by considering the mass of Apophis. The same initial values as the nominal orbit are used. The deviation from the nominal orbit is shown in Fig. 40.3. It can be concluded from Fig. 40.3 that the position error is at most $O(40.10)$ m in about 8.458 years for $O(10^4)$ km high orbits by ignoring the mass of Apophis. Therefore, for asteroids with small masses, the restricted two-body model can be used if the probe is not very close to an asteroid.

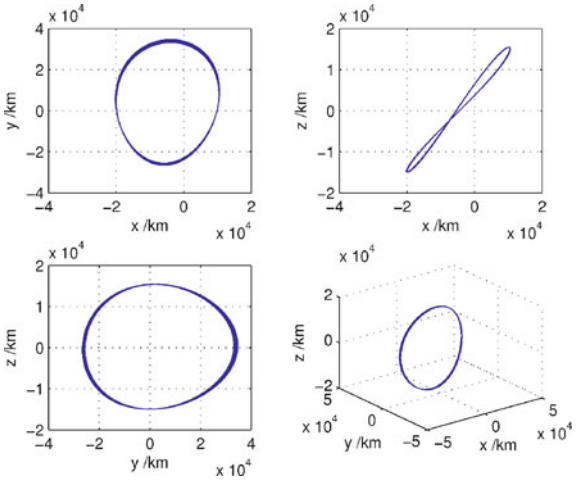

Fig. 40.2 The formation flying orbit around Apophis in the real system for $x_0 = y_0 = z_0 = 1 \times 10^{-4} R_0$

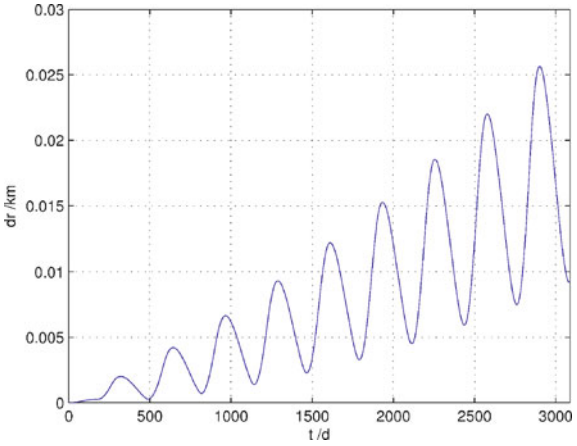

Fig. 40.3 The deviation of the orbit considering the mass of Apophis from the nominal orbit

40.4 Station Keeping

Both kinds of nominal orbits are unstable, especially for the orbits around the collinear libration point L_1. Due to the inevitable existence of errors, station keeping of the probe is necessary. In our work, the linear optimal feedback strategy is used and the solar sail propulsion is considered. Two control techniques are used. One is to vary solar sail area, and the other is to vary the solar sail pitch and yaw angles.

40.4.1 Control Strategy

Equations (40.10) and (40.17) are all nonlinear systems, so they should be linearized first as the following form

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{u}(t) + \mathbf{E}(t) \tag{40.19}$$

where $\mathbf{x} = \left(\Delta\boldsymbol{\rho}^T, \Delta\dot{\rho}^T\right)^T$ is the deviation of the state vector from the nominal orbit. $\mathbf{E}(t)$ is the error of the force model. Since the nominal orbit is constructed in the real force model, $\mathbf{B}(t) = 0$. $\mathbf{F}(t)$ is of the following form

$$\begin{cases} \mathbf{F}(t) = \begin{pmatrix} \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \\ -\mathbf{C}^T\ddot{\mathbf{C}} + \mathbf{D} & -2\mathbf{C}^T\dot{\mathbf{C}} \end{pmatrix} \\ \mathbf{D} = -\mu_{11}\left(\dfrac{\mathbf{I}_{3\times3}}{r^3} - 3\dfrac{\mathbf{r}^T\mathbf{r}}{r^5}\right) - \displaystyle\sum_{i\in S, i\neq11} \mu_i\left(\dfrac{\mathbf{I}_{3\times3}}{\delta_i^3} - 3\dfrac{\boldsymbol{\delta}_i^T\boldsymbol{\delta}_i}{\delta_i^5}\right) \end{cases} \tag{40.20}$$

$\mathbf{G}(t)\mathbf{u}(t)$ is the linearized result of $\mathbf{F}_s$. Denote the yaw angle and the pitch angle of the position direction vector $\hat{\Delta}$ of the probe from the Sun as (θ, ϕ). $(\theta + \alpha,\ \phi + \delta)$ denote the yaw angle and the pitch angle of the sail normal direction vector $\mathbf{n}$. $\alpha = 0$ and $\delta = 0$ at the nominal orbits. For the first control strategy of varying the solar sail area, $u = \beta$. $\mathbf{G}(t)$ is expressed as ([2, 10])

$$\mathbf{G}(t) = \begin{pmatrix} 0 & 0 & 0 & \dfrac{1}{\Delta^2}\cos\phi\cos\theta & \dfrac{1}{\Delta^2}\cos\phi\sin\theta & \dfrac{1}{\Delta^2}\sin\phi \end{pmatrix}^T \tag{40.21}$$

For the second control strategy, $u = (\alpha, \delta)^T$. $\mathbf{G}(t)$ is

$$\mathbf{G}(t) = \begin{pmatrix} 0 & 0 & 0 & -\frac{\beta}{\Delta^2}\cos\phi\sin\theta & \frac{\beta}{\Delta^2}\cos\phi\cos\theta & 0 \\ 0 & 0 & 0 & -\frac{\beta}{\Delta^2}\sin\phi\cos\theta & -\frac{\beta}{\Delta^2}\sin\phi\sin\theta & \frac{\beta}{\Delta^2}\cos\phi \end{pmatrix}^T \tag{40.22}$$

In this problem, for any initial state $\mathbf{x_0}$ at time t_0, we wish to minimize the cost function

$$J = \int_{t_0}^{t_f} \left(\mathbf{x^T A x} + \mathbf{u^T u}\right) dt \tag{40.23}$$

where $\mathbf{A} = \begin{pmatrix} k_1\mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & k_2\mathbf{I}_{3\times3} \end{pmatrix}$. The parameters k_1 and k_2 are a measure of the 'tightness' of the desired control. The optimal control which makes the minimum cost function can be expressed as [3]

$$\mathbf{u}^*(t) = -\mathbf{B}^{-1}(t)\mathbf{G}^{\mathbf{T}}(t)\mathbf{P}(t)\mathbf{x}(t) \tag{40.24}$$

where $\mathbf{P}(t)$ is a 6×6 matrix which satisfies the matrix-Riccati equation

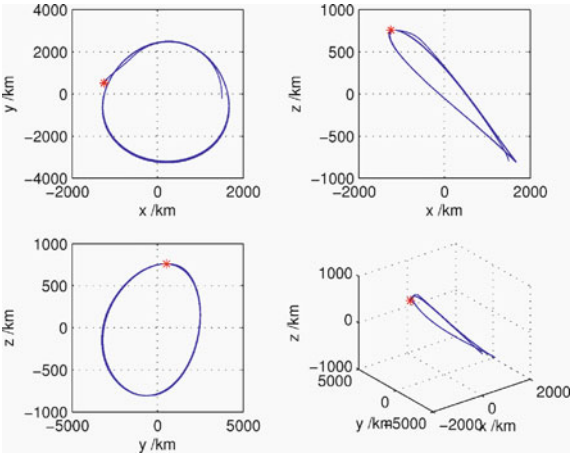

Fig. 40.4 The controlled halo orbit with solar sail area variable

$$-\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{F}(t) + \mathbf{F}(t)^T\mathbf{P}(t) - \mathbf{P}(t)\mathbf{G}(t)\mathbf{B}(t)^{-1}\mathbf{G}(t)^T\mathbf{P}(t) + \mathbf{A}(t) \tag{40.25}$$

The boundary condition at the final time t_f is $\mathbf{P}(t_f) = \mathbf{0}$. The optimal index is $J^* = \mathbf{x}^{\mathbf{T}}(t_0)\mathbf{P}(t_0)\mathbf{x}(t_0)$. The optimal orbit is the solution of the differential equation

$$\dot{\mathbf{x}}(t) = \left[\mathbf{F}(t) - \mathbf{G}(t)\mathbf{B}^{-\mathbf{1}}(t)\mathbf{G}^{\mathbf{T}}(t)\mathbf{P}(t)\right]\mathbf{x}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{40.26}$$

Since the matrix-Riccati equation is nonlinear, it is hard to be solved analytically. We solve it by numerical integrations. After obtaining the solution of $\mathbf{P}(t)$, the optimal control and the optimal orbit can be obtained by solving Eqs. (40.24) and (40.26).

40.4.2 Sail Area Variation

In the paper, the initial orbit insertion error $\mathbf{x} = \left(\Delta\boldsymbol{\rho}^T, \Delta\dot{\rho}^T\right)^T$ is set to be

$$\Delta\rho_i{=}1 \times 10^{-6} \approx 137974 \text{ m}, \quad \Delta\dot{\rho}_i{=}1 \times 10^{-6} \approx 0.031 \text{ m/s}, \quad i = 1, 2, 3 \tag{40.27}$$

If the position control parameter k_1 is too small, the position deviation of the controlled orbit from the nominal orbit may not converge, so $k_1 = 1000$ and $k_2 = 0$ are adopted.

According to the station-keeping strategy given above and using the halo orbit in the real force model shown in Fig. 40.1 as nominal orbit, the controlled orbit is shown in Fig. 40.4. The asterisk "*" in the figure denotes the initial position of the probe. Figure 40.5 is the position deviation curve of the controlled orbit from the nominal orbit. Judging from Fig. 40.5, the deviation decreases to a certain value and then oscillates at the same frequency as the nominal orbit. Figure 40.6 is

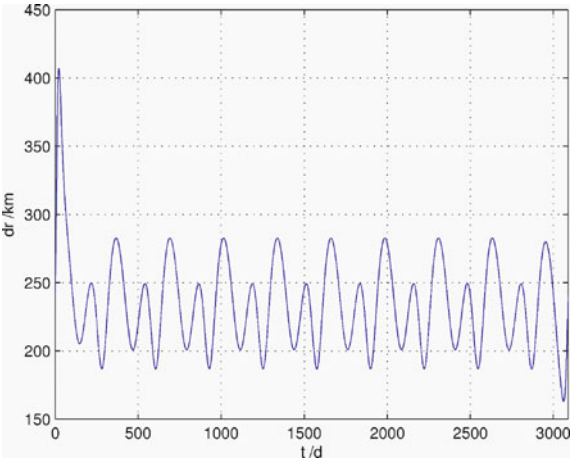

Fig. 40.5 The position deviation of the controlled halo orbit from the nominal orbit

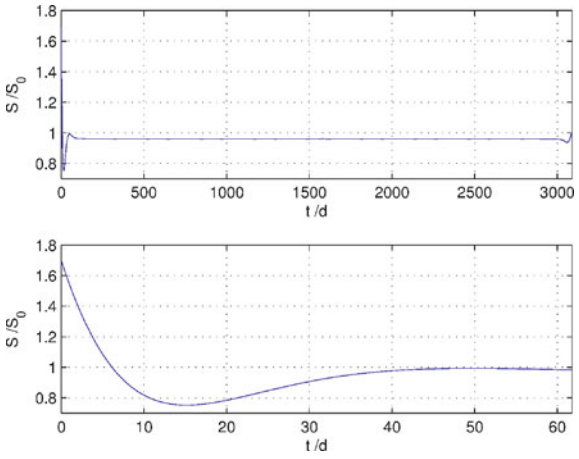

Fig. 40.6 The variation of the sail area relative to the initial area

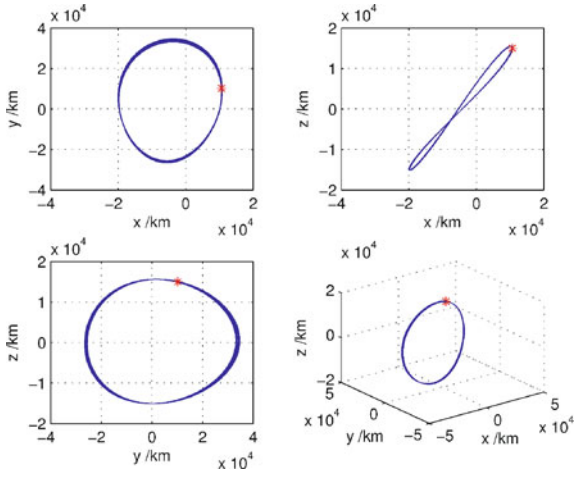

Fig. 40.7 The controlled formation flying orbit around Apophis

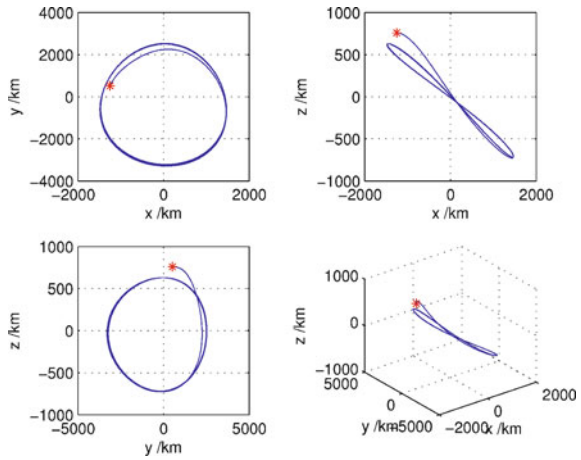

Fig. 40.8 The controlled halo orbit with pitch and yaw angle variable

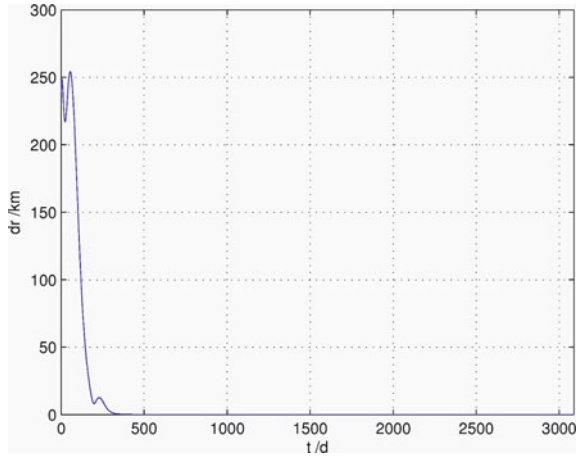

Fig. 40.9 The position deviation of the controlled halo orbit from the nominal orbit

the variation curve of the sail area relative to the initial sail area. The sail area "S" changes greatly at first and then converges to the vicinity of the initial area "S_0".

Using the formation flying orbit in the real force model as nominal orbit shown in Fig. 40.2, the controlled orbit is shown in Fig. 40.7. Because the position control parameter k_1 is very large and the solution $\mathbf{P}(t)$ of Eq. (40.26) is mainly decided by $\mathbf{A}(t)$, the matrix $\mathbf{P}(t)$ is almost the same as the result for the first kind of nominal orbit. As a result, the deviation of the controlled formation flying orbit from the nominal orbit and the variation curve of the sail area are almost the same as the results shown in Figs. 40.5 and 40.6. To save space, the details will not be presented here.

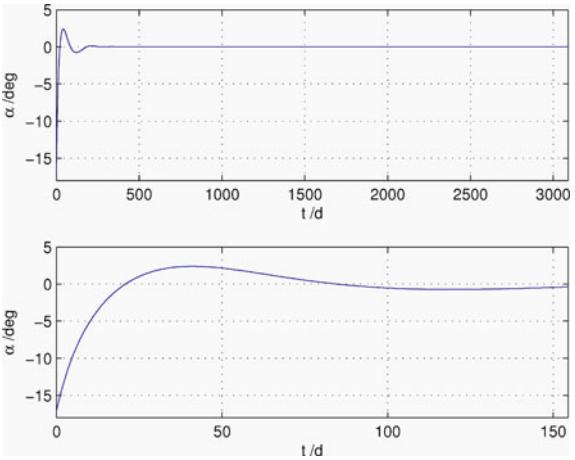

Fig. 40.10 The variation of the sail yaw angle

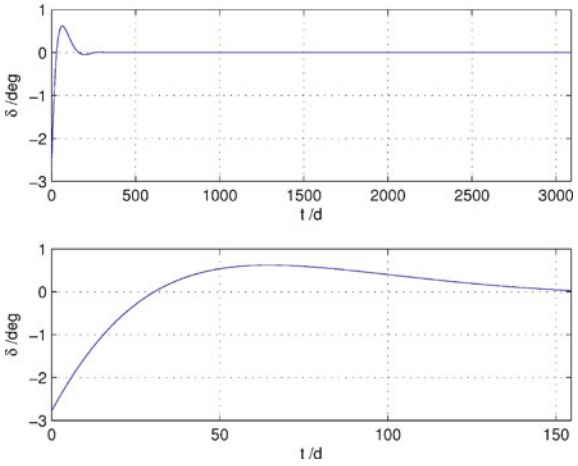

Fig. 40.11 The variation of the sail pitch angle

40.4.3 Sail Pitch and Yaw Angle Variation

The initial orbit insertion error $x = \left(\Delta\boldsymbol{\rho}^T, \Delta\dot{\rho}^T\right)^T$ is also the same as Eq. (40.27). $k_1 = 10^9$ and $k_2 = 0$ are adopted. The controlled results of the halo orbit in Fig. 40.1 is shown in Figs. 40.8, 40.9, 40.10, 40.11.

Different from the results of the first control strategy, Fig. 40.9 shows that the position deviation of the controlled halo orbit from the nominal orbit decreases rapidly and converges to 0 very soon. The sail pitch angle and the yaw angle also change greatly in 150 days when the position deviation is large. Using the same initial orbit insertion error, the controlled formation flying orbit is shown as Fig. 40.12. Also the position deviation of the controlled formation flying orbit from the nominal orbit and the variation curve of the yaw and pitch angles of the

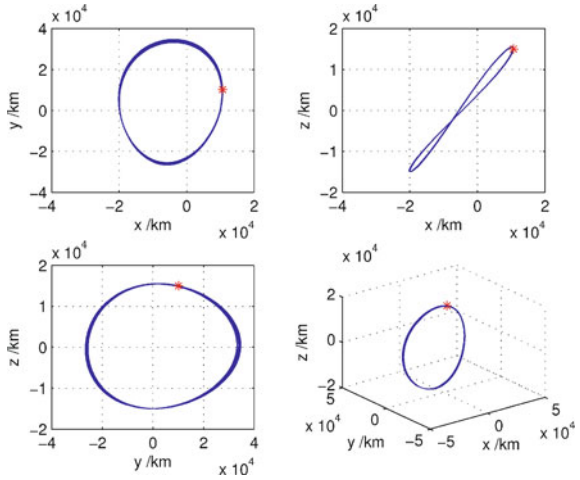

Fig. 40.12 The controlled formation flying orbit around Apophis

sail are almost the same as the results shown in Figs. 40.9, 40.10, 40.11. Form the results above, it can be concluded that the control strategy of changing solar sail pitch and yaw angles is better than the first one for the two kinds of nominal orbits in the paper.

40.5 Discussions and Conclusions

In this work, two kinds of nominal orbits were proposed for the exploration of asteroids. Solar sail propulsion was used to provide station-keeping around the nominal orbits. Two control techniques including varying the solar sail area and varying the solar sail pitch and yaw angles were considered. The numerical results showed that both kinds of nominal orbits work well and the linear optimal feedback strategy is feasible. For the control strategy of changing the sail pitch and yaw angles, the controlled orbits can converge to the nominal orbits rapidly. But by changing the sail area, the control results show that the position deviation from the nominal orbits is an oscillation function around a certain deviation with the same frequency as the nominal orbits. Therefore, for the two kinds of nominal orbits, the control strategy of changing pitch and yaw angles is better than changing the sail area.

The perturbation of the asteroid's non-spherical gravitation is not considered in the paper. For the two kinds of nominal orbits, they are not very close to the asteroids, so the influence is small. Especially for the second kind ones, the gravitational perturbation of the asteroid as a particle is already neglected, so the perturbation of its non-spherical gravitation is even smaller and can also be neglected.

Acknowledgments This work was supported by national Natural Science Foundation of China (NSFC 10903002, 11033009).

References

1. Anderson RL, Lo MW, Born GH (2003) Application of local Lyapunov exponents to maneuver design and navigation in the three-body problem. In: AAS 03-569, AIAA/AAS astrodynamics specialist conference, Big Sky, Montana, 3–7 Aug
2. Bookless J, McInnes CR (2008) Control of Lagrange point orbits using solar sail propulsion. Acta Astronaut 62:159–176
3. Breakwell JV, Kamel AA, Ratner MJ (1974) Station-keeping for a translunar communications station. Celest Mech 10(3):357–373
4. Cheng AF (2002) Near earth asteroid rendezvous: mission summary. In: Bottke WF Jr, Cellino A et al. (eds) Asteroids III. University of Arizona press, Tucson, pp 351–366
5. Farquhar RW (1970) The control and use of libration point satellites. Technical report TR R346, NASA
6. Folta D, Vaughn F (2004) A survey of earth–moon libration orbits: station-keeping strategies and intra-orbit transfers. In: AIAA 2004-4741, AIAA/AAS astrodynamics specialist conference, Providence, Rhode Island, 16–19 Aug
7. Gómez G, Howell K, Masdemont J et al (1998) Station-keeping strategies for translunar libration point orbits. Adv Astronaut Sci 99:949–967
8. Gómez G, Jorba Á, Masdemont J et al. (2001) Dynamics and mission design near libration point orbits. Vol. III, Advanced methods for collinear points. World Scientific, Singapore
9. Gómez G, Masdemont J, Simó C (1998) Quasi-Halo orbits associated with libration points. J Astronaut Sci 42(2):135–176
10. Hou XY, Liu L (2009) On orbit control of spacecrafts with solar sail around the earth–moon collinear libration point. J Astronaut 30(6):2249–2257
11. Hou XY, Liu L (2010) On quasi-periodic motions around the triangular libration points of the real earth–moon system. Celest Mech Dyn Astron 108:301–313
12. Hou XY, Liu L (2011) Station-keeping of small amplitude motions around the collinear libration point in the real earth–moon system. Adv Space Res 47:1127–1134
13. Hou XY, Liu L (2011) On quasi-periodic motions around the collinear libration points in the real earth–moon system. Celest Mech Dyn Astron 110:71–98
14. Howell KC, Pernicka HJ (1987) Numerical determination of Lissajous trajectories in the restricted three-body problem. Celest Mech Dyn Astron 41:107–124
15. Howell KC, Pernicka HJ (1993) Station-keeping method for libration point trajectories. J Guid Control Dynam 16(1):151–159
16. Jorba A, Masdemont J (1999) Dynamics in the centre manifold of the collinear points of the restricted three body problem. Phys D 132:189–213
17. Kulkarni JE, Campbell ME, Dullerud GE (2006) Stabilization of spacecraft flight in halo orbits: an H∞ approach. IEEE Trans Control Syst Technol 14(3):572–578
18. Liu L, Wang H, Ma J (2004) On the formation flying of satellite constellations. Chin Astro Astrophy 28(2):188–199
19. Morrison DD, Riley JD, Zancanzro JF (1962) Multiple shooting method for two-point boundary value problem. Commun ACM 5(12):613–614
20. Thomas VC, Makowski JM, Brown GM et al (2011) The dawn spacecraft. Space Sci Rev 163(1–4):175–249
21. Yoshikawa M, Fujiwara A, Kawaguchi J (2007) Hayabusa and its adventure around the tiny asteroid Itokawa. Highlights Astron 14:323–324

Chapter 41
Orbit Determination of Lunar Probe Brake Course Based on Compensation to Dynamic Parameters

Shijie Chen, Lan Du, Zhongkai Zhang, Quying Danzeng, Ruopu Wang, He Wang and Qifu Zhang

Abstract During orbital maneuver process, the probe relays on propulsion system to generate thrust acceleration, which tunes attitude and orbit. The complicated course of propulsion makes thrust hard to be modeled. The difficulty of precisely modeling thrust acceleration mainly includes two aspects: firstly, the begin-and-final epochs of attitude control and orbital control are not easy to determine; secondly, the acceleration of attitude control and orbital control are not easy to model accurately. It focuses on modeling and building nearly real-time filter to estimate dynamic parameters to compensate the uncertainty of the force model during the maneuver process. In the first section, the background of modeling thrust acceleration was introduced. In Sect. 41.2, the linear attitude control acceleration and the average orbital control acceleration model were used to describe the accelerating motion. In Sect. 41.3, an Extended Kalman Filter (EKF) was developed for the orbit determination with thrust involved, also the variation equation about state vector with dynamic parameters were put forward for the linearization of the non-linear dynamic system. In Sect. 41.4, the third lunar orbital brake of Chang'E-1 was processed to ascertain that the algorithm developed here can estimate the acceleration precisely during the continuous thrust maneuver process. EKF Results show that compared with the precise post-results, the error is 216 m in semi-major axis and 0.001 in eccentricity.

Keywords Lunar probe · Dynamic model compensation · Orbital maneuver · Dynamic parameters estimation

S. Chen (✉) · L. Du · Z. Zhang · Q. Danzeng ·
R. Wang · H. Wang · Q. Zhang
Information Engineering University, Zhengzhou 450052, China
e-mail: tianyaniao123@tom.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_41,

41.1 Introduction

In recent years, the orbit determination of probe has been focused on dealing with maneuver. At present, there are mainly three methods, namely, point positioning method, dynamic method and kinematical method. Due to the difficulty in dynamic modeling for the braking lunar probe, Li Jin-ling put forward the instantaneous state vector (ISV) algorithm, which realizes orbit determination geometrically epoch by epoch. The deficiency is that the positioning needs more observations and cannot be predicted orbit [1, 2]. Li Peijia discussed the short-arc kinematical orbit determination method, which adopts higher order polynomial functions to describe the short-arc trajectory including the maneuver. However, the relatively more unknowns need to accumulate long-arc observations to fit better, which affect orbit determination in real-time [3]. Both the ISV algorithm and the kinematical method avoid accurate dynamic modeling, but will not satisfy the requirements in efficiency. In this paper we introduce the orbit determination filter with the real-time compensation of dynamic parameters during orbital maneuver.

The difficulty of precisely modeling thrust acceleration mainly includes two aspects [4]. Firstly, the begin-and-final epochs of attitude control and orbital control are not easy to demarcate because the received remote sensing data rate is low, the direct judgment from remote sensing data has great error. Secondly, the attitude control force and the orbital control force are not easy to model accurately because the designed value of engine's parameters are not consistent with the actual condition, the accumulative error of dynamics leads to a rapid decline in the precision of orbit determination.

The maneuvering parameters will be estimated in this paper accurately in two steps.

Firstly, the begin-and-final epochs of attitude control and orbital control can be determined with the tracking observables almost in real-time, which is the foundation of accurate modeling for orbital maneuver course. It aims to precisely separate the attitude control course and the orbital control course from the cruise course. If the observables own high sampling rate and precision of remote sensing data or orbit sensing data, the begin-and-final epochs of thrust can be demarcated through three algorithms. (i) Differential algorithm, the variation rate of velocity-related observations are constructed to monitor the jumping points, which change the trend of curve [5, 6]. (ii) Statistical test algorithm, predict residuals are applied to construct test statistic to diagnose dynamics model fault. (iii) Acceleration data fitting algorithm [7]. Without acceleration data, herein, the differential algorithm and the statistical test algorithm have been adopted to demarcate the epochs of the respective control courses, which will be presented in other articles.

Secondly, it is the on-line compensation to dynamic parameters, which is essential to realize precise orbit determination for orbital maneuver course, and is also the focus of the paper. On the basis of precise demarcation for the begin-and-final epochs, the average thrust is modeled to describe approximately the accelerating motion and estimated precisely along with position and velocity vectors.

Taking the third lunar orbital brake process of lunar capture of Chang'E-1 probe as an example, the proposed method is validated with Chinese VLBI network observations. The results show that it can be used to the rapid trace process.

41.2 Dynamic Model

During orbital maneuver course, the probe relays on propulsion system to tune attitude and brake. The probe's propulsion system generates thrust acceleration, whose magnitude is far beyond the other perturbation forces [8]. The ignition posture is to be established before orbital maneuver [9, 10], meanwhile, perturbation acceleration in the direction of centroid is also produced during attitude adjustment course [11, 12]. Therefore, the precise modeling of attitude control force and orbital control force is the key to the accurate dynamic modeling of orbital maneuver course.

In the Lunar Centered J2000.0 Inertial coordinate system, whose basic plane is the earth J2000.0 mean equator plane, the dynamic equation of the lunar probe during lunar capture is

$$\ddot{\boldsymbol{r}} = a_0 + \boldsymbol{a}_{\varepsilon} + \mu_1 \boldsymbol{a}_{tac} + \mu_2 \boldsymbol{a}_{thrust} \tag{41.1}$$

where $\ddot{\boldsymbol{r}}$ is the acceleration vector, $\boldsymbol{a}_0$ is the gravitational acceleration vector of lunar point mass, $\boldsymbol{a}_{\varepsilon}$ is the perturbed acceleration vectors which is under the influence of all the natural perturbations, $\boldsymbol{a}_{thrust}$ is the acceleration vector exerted by orbital control force, $\boldsymbol{a}_{tac}$ is the acceleration vector of attitude control force. Name $t_{tacbegin}$ and $t_{tacfinal}$ are the beginning and final epoch of attitude control course, $t_{brakebegin}$ and $t_{brakefinal}$ are the beginning and final epoch of orbital control course, and μ_1, μ_2 are the switch functions of attitude control force and orbital control force respectively. For example when $t_{tacbegin} \le t \le t_{tacfinal}$, $\mu_1 = 1$ means that the attitude control force is acting. Usually the attitude control course and the orbital control course are independent, namely, after the attitude control course ends, the ignition posture is established, and then orbital control is exerted on the probe. The probe relays on the jet of high pressure cylinder to adjust both magnitude and orientation of attitude and thrust, and the mass and velocity of the probe will continue change with time.

41.2.1 Attitude Control Force Model

It is convenient to establish the attitude control force model in RTN coordinate system. Because the jet process is not even, linear model can be adopted to approximate the perturbation acceleration, and

$$\boldsymbol{a}_{tacRTN} = \begin{bmatrix} \boldsymbol{a}_{tacR} \\ \boldsymbol{a}_{tacT} \\ \boldsymbol{a}_{tacN} \end{bmatrix} = \boldsymbol{c}_0 + \boldsymbol{c}_1 \Delta t \tag{41.2}$$

where $\boldsymbol{a}_{tacR}$ $\boldsymbol{a}_{tacT}$ $\boldsymbol{a}_{tacN}$ are the radical (R), transverse (T), normal (N) component of $\boldsymbol{a}_{tacRTN}$, $\boldsymbol{c}_0$ and $\boldsymbol{c}_1$ are the constant and slope vectors of $\boldsymbol{a}_{tacRTN}$ in RTN coordinate system, Δt is the time interval between the current epoch and the beginning epoch. The vector $\boldsymbol{a}_{tacRTN}$ should be transformed into J2000.0 coordinate system.

$$\boldsymbol{a}_{tac} = \boldsymbol{M}\boldsymbol{a}_{tacRTN} \tag{41.3}$$

where $\boldsymbol{M}$ is the rotation matrix from RTN coordinate system to J2000.0 coordinate system. Actually, only $\boldsymbol{c}_1$ is estimated in order to improve the systematic observability of estimation.

41.2.2 Orbital Control Force Model

Thrust acceleration varies with time during orbital maneuver course along with the decrease of the mass of the fuel. In Lunar Centered J2000.0 inertial coordinate system, assume that the direction of orbital control force is known, the thrust acceleration induced by the propellant is

$$\boldsymbol{a}_{thrust} = a\boldsymbol{p} \tag{41.4}$$

where a is the magnitude of thrust acceleration, $\boldsymbol{p}$ is the thrust direction, which can be determined by attitude sensors in the practical engineering application [13].

Suppose that the jet process of propulsion system is uniform, it is reasonable to establish the average thrust model [14, 15]. Assume that the mass of probe at the brake beginning epoch t_0 is m_0, when the propellant burns on, the mass flow rate is $\dot{m}$, the specific impulse is I_{sp}, and then the variable acceleration induced by the propellant is

$$a = \frac{\dot{m} I_{sp}}{m - \dot{m}(t - t_0)} \tag{41.5}$$

where t is the current epoch of orbital control course. Equation (41.5) indicates that the acceleration varies not only with the duration of the propellant burn, but also relates to the initial mass of probe, the mass flow rate and specific impulse of propellant. For the sake of improving the observability of dynamic system, it only takes the mass flow rate $\dot{m}$ and the impulse I_{sp} as unknown parameters to be estimated.

Extend the state of the probe as $\boldsymbol{X} = (\boldsymbol{r}, \dot{\boldsymbol{r}}, \boldsymbol{q})^{\mathrm{T}}$ and the differential equation can be expressed as

$$\dot{\boldsymbol{X}} = \begin{bmatrix} \dot{\boldsymbol{r}} \\ \ddot{\boldsymbol{r}} \\ \dot{\boldsymbol{q}} \end{bmatrix} = \boldsymbol{f}(\boldsymbol{X}) = \begin{bmatrix} \dot{\boldsymbol{r}} \\ \boldsymbol{a}_0 + \boldsymbol{a}_\varepsilon + \mu_1 \boldsymbol{a}_{tac} + \mu_2 \boldsymbol{a}_{thrust} \\ \dot{\boldsymbol{q}} \end{bmatrix} \tag{41.6}$$

where $\boldsymbol{x} = (\boldsymbol{r}, \dot{\boldsymbol{r}})^{\mathrm{T}}$ is the position and velocity vector, $\boldsymbol{q}$ is the dynamic parameter vector, which is expressed in different ways in attitude control course and orbital control course,

$$\boldsymbol{q} = \begin{cases} \boldsymbol{c}_1 & \mu_1 = 1 \quad \text{the attitude control force model} \\ (\dot{m}, I_{SP})^{\mathrm{T}} & \mu_2 = 1 \quad \text{the orbital control force model} \end{cases} \tag{41.7}$$

41.3 EKF with Dynamic Model Compensation

41.3.1 Partial Derivative

Compared with the general EKF [16], the complicated part is the partial derivative $\boldsymbol{H}_k$ of augmented states

$$\boldsymbol{H_k} = \frac{\partial z}{\partial \boldsymbol{X}} | (\boldsymbol{X} = \boldsymbol{X}_k^-) = (\frac{\partial z_k}{\partial \boldsymbol{r}_k^-}, \frac{\partial z_k}{\partial \dot{\boldsymbol{r}}_k^-}, \frac{\partial z_k}{\partial (\boldsymbol{r}_k^-, \dot{\boldsymbol{r}}_k^-)^{\mathrm{T}}} \cdot \frac{\partial (\boldsymbol{r}_k^-, \dot{\boldsymbol{r}}_k^-)^{\mathrm{T}}}{\partial q}) \tag{41.8}$$

where z is the observable, $\frac{\partial z_k}{\partial \boldsymbol{r}_k^-}, \frac{\partial z_k}{\partial \dot{\boldsymbol{r}}_k^-}$ is the partials to the position-velocity vector from z, $\frac{\partial (\boldsymbol{r}_k^-, \dot{\boldsymbol{r}}_k^-)}{\partial q}$ is the partials to the dynamic parameters from position to velocity vector, which is computed through the integration of variational equation below.

41.3.2 Variational Equation

Compared with cruise course, the extended state transition matrix (t, t_0) is more complex and can be obtained by the integration of variation equation as follows,

$$\dot{\Phi}(t, t_0) = \begin{bmatrix} \dot{\Phi}^{\boldsymbol{r},\boldsymbol{v}}(t, t_0) & \dot{\boldsymbol{S}}(t) \\ 0 & \dot{\Phi}^{\boldsymbol{q}}(t, t_0) \end{bmatrix} = \boldsymbol{A}_t \Phi(t, t_0) \tag{41.9}$$

$$\boldsymbol{A}_t = \frac{\partial \dot{\boldsymbol{X}}_t}{\partial \boldsymbol{X}_t}, \ \Phi^{\boldsymbol{r},\boldsymbol{v}}(t, t_0) = \begin{bmatrix} \frac{\partial \boldsymbol{r}_t}{\partial \boldsymbol{r}_{t_0}} & \frac{\partial \boldsymbol{r}_t}{\partial \dot{\boldsymbol{r}}_{t_0}} \\ \frac{\partial \dot{\boldsymbol{r}}_t}{\partial \boldsymbol{r}_{t_0}} & \frac{\partial \dot{\boldsymbol{r}}_t}{\partial \dot{\boldsymbol{r}}_{t_0}} \end{bmatrix}, \boldsymbol{S}(t) = \begin{bmatrix} \frac{\partial \boldsymbol{r}_t}{\partial q} \\ \frac{\partial \dot{\boldsymbol{r}}_t}{\partial q} \end{bmatrix}, \ \Phi^{\boldsymbol{q}}(t, t_0) = \frac{\partial q}{\partial q_0}$$

where $\Phi^{\boldsymbol{r},\boldsymbol{v}}(t, t_0)$ and $\Phi^{\boldsymbol{q}}(t, t_0)$ are the state transition matrix of position-velocity vector and dynamic parameters respectively, $\boldsymbol{S}(t)$ is the sensitivity matrix of dynamic parameters., the same is the computation of $\Phi^{\boldsymbol{r},\boldsymbol{v}}(t, t_0)$, the difference is

Table 41.1 The flying control program of the third brake of CE-1 (2007.11.7)

		Epoch	Engine type	Thrust
Begin epoch of cruise section		0:00:0.0	–	–
Maneuver section	Begin epoch of attitude control section	0:20:23.580	Attitude control engine	10 N
	Begin epoch of orbital control section	0:24:28.818	Orbital control engine	490 N
	Final epoch of orbital control section	0:33:23.580	Orbital control engine	490 N
Final epoch of cruise section		0:36:0.580	–	–

the computation of sensitivity matrix $\boldsymbol{S}(t)$ and dynamic parameters transition matrix $\boldsymbol{\Phi}^{\boldsymbol{q}}(t, t_0)$.

The differential equation of the position-velocity transition matrix can be expressed as [17],

$$\dot{\boldsymbol{\Phi}}^{\boldsymbol{r},\boldsymbol{v}}(t, t_0) = \begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} \\ \boldsymbol{G}_{3\times3} & \boldsymbol{0}_{3\times3} \end{bmatrix} \cdot \boldsymbol{\Phi}^{\boldsymbol{r},\boldsymbol{v}}(t) \tag{41.10}$$

where $\boldsymbol{G}$ is the gradient matrix of gravitational field.

The differential equation of sensitivity matrix can be expressed as,

$$\dot{\boldsymbol{S}}(t) = \begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} \\ \boldsymbol{G}_{3\times3} & \boldsymbol{0}_{3\times3} \end{bmatrix} \cdot \boldsymbol{S}(t) + \begin{bmatrix} 0 \\ \frac{\partial}{\partial q}(\mu_1 \boldsymbol{a}_{tac} + \mu_2 \boldsymbol{a}_{thrust}) \end{bmatrix} \tag{41.11}$$

The partial derivative about maneuver force with dynamic parameters can be expressed as,

$$\frac{\partial}{\partial q}(\mu_1 \boldsymbol{a}_{tac} + \mu_2 \boldsymbol{a}_{thrust}) = \begin{cases} \Delta t \cdot \boldsymbol{M} & \mu_1 = 1 \\ (\frac{\partial a}{\partial m}, \frac{\partial a}{\partial I_{sp}}) \cdot \boldsymbol{p} + a \cdot \frac{\partial \boldsymbol{p}}{\partial (\dot{m}, I_{sp})^{\mathrm{T}}} & \mu_2 = 1 \end{cases} \tag{41.12}$$

The differential equation of dynamic parameters transition matrix can be expressed as,

$$\frac{\partial \dot{\boldsymbol{q}}}{\partial \boldsymbol{q}} = \begin{cases} 0_{3\times1} & \mu_1 = 1 \quad \text{the attitude control force model} \\ 0_{2\times1} & \mu_2 = 1 \quad \text{the orbital control force model} \end{cases} \tag{41.13}$$

41.4 Application Example Analysis

The CE-1 lunar probe has been manipulated to accomplish three orbital brakes near the perilune before becoming the polar satellite with the orbital altitude of 200 km. Taking the third lunar orbital brake process of CE-1 as an example, the proposed algorithm is validated with domestic VLBI network observations. The flying control

Table 41.2 The flying control program of the third brake of CE-1 (2007.11.7)

Item	Model
Coordinate system	Lunar centre J2000.0 inertial coordinate system
Lunar gravity field model	LP165p 20 × 20
Third body position	JPL DE403 (earth, sun)
Solar radiation pressure model	Fixed ratio about area with mass
Initial orbit and epoch	The solution of combined system with USB and VLBI (2007.11.7.0)
Earth rotation parameters	IERS Bulletin B (2007.11.7)
Integral method	RKF78
Estimation method	EKF
Observation	VLBI net data after preprocess
Solution parameters	Position velocity + dynamic parameters

program of the third brake of CE-1 is scheduled as Table 41.1, the orbital determination strategy is summarized as Table 41.2.

It is reasonable to stress that:

1. The domestic VLBI nets data are adopted alone to determine orbit herein. However, the initial orbit is provided by the solution of combined USB and VLBI data in order to have a higher precision in radical direction [18].
2. The thrust direction is simplified in this article. Unfortunately, the thrust direction can not be described accurately because there is no installation parameter of nozzle and attitude data. Therefore, we use the precise initial orbit to predict orbit to obtain the velocity direction of perilune, and then the opposite direction is set as the thrust direction. The modeling error can be absorbed partly through the thrust acceleration value.
3. The dynamical noise of state propagation equation is also simplified in this article, which is set as one in a thousand of the initial value.

41.4.1 Estimation of Dynamic Parameters

The precise on-line estimation of dynamic parameters is very important, which is essential to the realization to the precise orbit determination of maneuver course. Figure 41.1 shows the curve of the maneuvering acceleration.

Figure 41.1 indicates that attitude control section is from 20.5 to 24.5 min lasting about 4 min; while orbital control section is from 24.5 to 33.5 min, which lasts about 9 min. Besides, the magnitudes of attitude and the orbital control acceleration are about 0.03 and 0.35 m/s^2, respectively.

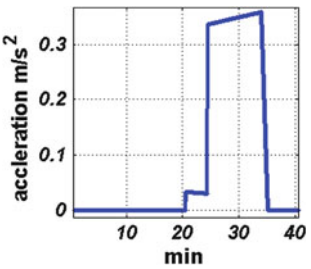

Fig. 41.1 The magnitude of the maneuver acceleration

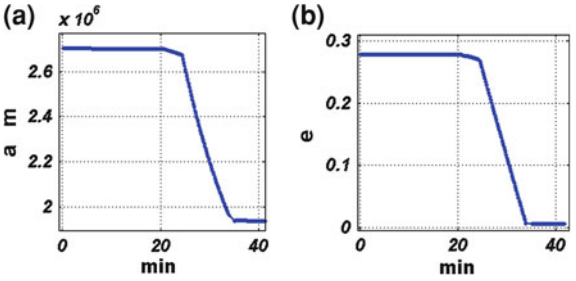

Fig. 41.2 Variation in orbital elements of the third lunar orbital brake process of CE-1. **a** Semi-major axis. **b** Eccentricity

41.4.2 Evaluation of Brake Effect

The Kepler elements from the estimated states are adopted to be compared with the Post-results of precise orbit to evaluate the brake effect. Figure 41.2 demonstrates the curve of semi-major axis and eccentricity, respectively. The comparisons between the post-results and real-time ones are summarized in Table 41.3.

Table 41.3 indicates that there is certain deviation between the post-result and the on-line filter result, the error is 216 m in semi-major axis and 0.001 in eccentricity.

Residual curve is another important index evaluating the orbit determination. Taking the Kunming-Urumqi baseline for example, Fig. 41.3 shows the residuals of time delay and time delay rate, respectively.

It indicates evidently that the dynamic model is accurate, but there are still small systematic errors in model.

1. The dynamic model compensation restrains the divergent trend of residuals to a great extent. In maneuver section, the time delay rate residuals are up to 3, and 100 ps/s with- and without-dynamic model compensation, respectively. Therefore, about 97 % residuals are absorbed by dynamic parameters.
2. Due to the simplify of thrust direction, there are still small model system errors, the magnitude of time delay rate residuals range from 1 to 3 ps/s during maneuver process.

Table 41.3 Orbit elements of CE-1 after braking (0:35:0.0)

	Semi-major axis/(km)	Eccentricity
Post-result (from USB and VLBI jointly)	1938.630	0.006
Real-time EKF with dynamic parameters compensation	1938.846	0.007

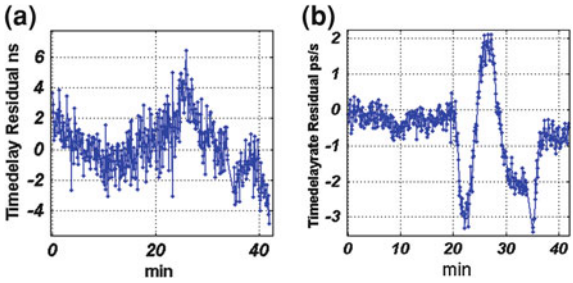

Fig. 41.3 Kunming-Urumqi VLBI residual of the third lunar orbital brake process of CE-1. **a** Time delay. **b** Time delay rate

41.5 Conclusions

It focuses on building orbit determination filter with dynamic model compensation to process the consecutive mild and large maneuvers during the brake of CE-1. The attitude control force and the orbit control force are modeled with the uncertain dynamic parameters into extended states along with position velocity vectors to be estimated. EKF has been developed to estimate forwardly uncertain dynamic parameters in real time, also the variation equation about state vector with dynamic parameters has been put forward in realizing the linearization of the non-linear dynamic system. The third lunar orbital brake process of CE-1 has been processed to evaluate that the algorithm developed here can estimate the acceleration precisely during the continuous thrust maneuver process.

References

1. Guo L (2007) Reduction of the instantaneous state vectors of probe based on VLBI tracking data. Shanghai astronomical observatory, Chinese academy of sciences
2. Li JL, Guo L, Qian ZY et al (2009) The application of the instantaneous state vectors reduction on the pivotal section. Sci China 39(10):1393–1399
3. Li PJ, Hu XG, Huang Y et al (2010) Kinematic statistical method using in the orbit determination of CE-1. Prog Astron 28(3):290–300
4. Huang Y, Hu XG, Huang C et al (2008) The orbit determination of maneuvering GEO based on distance data of CAP. Sci China 38(12):1750–1758
5. Chen SJ, Du L, Danzeng Q et al (2011) Monitoring lunar orbit insertion period for time of CE-1 probe based on regional VLBI network. Deep Space Explor 9(3):21–26
6. Shi X, Wang MY, Jian NC et al (2004) Monitoring technology for lunar orbit insertion of CE-1. Deep Space Explor 24(5):354–359

7. Tang GS, Chen LD, Liu Y (2009) Analysis and implement of determining the orbit maneuver effect of CE-1 satellite. Chin Space Sci Technol 12(6):1–6
8. Li HN, Li JS, Huang YX (2010) A dynamic model compensation orbit determination method for maneuvering satellite. J Astronaut 31(10):2269–2275
9. Ouyang ZY (2007) Fly in the sky: from earth satellite to lunar probe. Shanghai science education press, Shanghai, pp 100–104
10. Yang XH, Li ZG, Feng CG (2008) Fastly resuming GEO ephemeris algorithm after manuvering course G. Sci China 38(12):1765–1795
11. Li JS (1995) Satellite precision orbit determination. PLA publishing house, Beijing, pp 79–187
12. Chen M, Tang GS, Cao JF et al (2011) Precision orbit determination of CE-1 lunar satellite. Geomat Inf Sci Wuhan Univ 36(2):212–217
13. He YF, Wang JS, Wang XH (2010) Precise orbit determination fit for maneuvering GEO satellite. CNSS: 191
14. Li HN, Li JS, Huang YX (2002) Minimum-variance estimation of trust acceleration during orbit transfer process. Chin J Space Sci 22(4):583–588
15. Li HN, Li JS, Huang YX (2003) Tracking and navigating filter for maneuvering target. Chin Space Sci Technol (3):13–18
16. Yang YX (2006) Adaptive navigation and kinematic positioning. Geomatic Science Press, Beijing, pp 78–90
17. Montenbruck O, Gill E (2000) Satellite orbits printing. Mercedes-Druck, Berlin, pp 240–244
18. Huang Y (2006) Orbit determination of the first Chinese exploration probe CE-1. Shanghai astronomical observatory, Chinese academy of sciences, pp 45–67

Chapter 42
A Modified IAE Algorithm for GNSS and IMU Integration

Peng Li, Chan Li, Xiangjun Wu and Zhonggui Chen

Abstract Innovation-based adaptive estimation (IAE), which is one of the proved Adaptive Kalman Filter (AKF) algorithms, can improve the accuracy of GNSS/IMU combined navigation system based on the condition that the received GNSS measurements are independently accurate enough. However, IAE is more likely to be subjected to bias and non-convergence with degraded measurements accuracy due to GNSS signal outage or low-cost receiver. In order to maintain the performance of integrated GNSS/IMU system with the coexistence of less accurate measurements, a modified IAE algorithm is proposed in this paper. The algorithm, named "IAE with measurements discarding strategy" (IAE-D), monitors the quality of the estimations and measurements in real-time, and discards the measurements when the estimations are accurate enough or the measurements are less qualified. Field test was carried out. Noise was imported to simulate different levels of measurements deterioration. Performance comparison between Extended Kalman Filter (EKF), IAE and IAE-D has been executed with real data. The results demonstrate that IAE-D has magnificent advantage over EKF and IAE in regard to stability and accuracy when the power level of measurements interference is relatively high.

Keywords GNSS/IMU IAE · AKF · Data fusion

P. Li (✉) · X. Wu · Z. Chen
Institute of Spacecraft System Engineering, China Academy of Space Technology, Beijing 100094, China
e-mail: richardlee0914@gmail.com

C. Li
National University of Defense Technology, Changsha 410073, China

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_42,

42.1 Introduction

Global Navigation Satellite System (GNSS) and Inertial Navigation System (INS) can compensate for each other in an integrated navigation system, to provide synergistic performance improvements. When GNSS receiver is integrated with INS sensor, which is also called Inertial Measurement Unit (IMU), the constructed navigation system is possible to provide uninterrupted, high sampling rated and fully functional navigation solutions with both short term precision and long term stability [1, 2]. The architectures of GNSS/INS integration have been explored for over 30 years. The associated literature is available in [2–5]. It shows that data fusion is always the main challenge in different integration architecture.

The Extended Kalman filter (EKF) is an ideal tool to carry out the task of data fusion for nonlinear dynamic systems by linearization [6], and has been used in the GNSS/IMU integration for a long time. Conventional EKF uses constant parameters for system modeling and constant updating frequency. System noise matrix $\boldsymbol{Q}$ and measurement noise matrix $\boldsymbol{R}$ are two major modeling parameters in EKF. They are generally set according to the priori information of the process noise and measurement errors. However, system dynamic model and measurement error stochastic model are changing in most kinematic applications. As a result, in the kinematic scenarios, the preset $\boldsymbol{R}$ and $\boldsymbol{Q}$ have bias to the ideal values, which consequently depredates the performance in EKF application [7], especially with the existence of the GNSS signal interference.

In order to address the issue of insufficient priori statistics, Adaptive Kalman Filter (AKF) is proposed to allow the filter stochastic information to be modified on-line. The accuracy of the integrated navigation system is consequently improved. The existing AKF algorithms are roughly classified into three categories. They are Multiple Model Adaptive Estimation (MMAE), Covariance Scaling (CS) and Innovation-based adaptive estimation (IAE). The MMAE method processes a bank of Kalman filters with different parameters in parallel, computes the posteriori probabilities for each of the filter modes, determines and updates the weight for each of them. The adaptive optimal estimation can be obtained as the weighted sum of the estimation from each filter [8]. The CS method introduces a multiplication factor S ($S > 1$) to the state covariance matrix $\boldsymbol{P}$ in order to increase the weight factor of the new measurements [9, 10]. IAE is a maximum likelihood estimator. It updates the Q, R according to the whiteness of the filter innovation sequence. Test has proved that IAE can improve the performance of GNSS/IMU navigation system by 50 % compared with the conventional EKF [11]. However, IAE is subjected to the quality of IMU and provides less accurate estimates of the attitude parameters than EKF when low cost IMUs are used [12].

Although IAE algorithms have been proven to be applicable in high accurate kinematic applications, the choice of the estimation window size is a dilemma in practice. Short window size may give rise to the problem of divergence; while long window size will lower the sensitivity of the system to the dynamic of R depending on the GNSS signal interference. As a result, the accuracy of the

integrated system would be vulnerable to the accuracy degradation of GNSS measurements [11]. Furthermore, the impact of non-convergence on IAE with the coexistence of less accurate measurements has been proven to be unacceptable for GNSS/IMU integration [13].

In order to improve the performance of IAE, this paper introduces a modified IAE algorithm called IAE with measurements discarding strategy (IAE-D). The measurements discarding method was firstly provided to improve the EKF performance [14], which monitors the quality of the estimations and the measurements in real-time, and discards the measurements when the estimations are accurate enough or the measurements are less qualified.

To evaluate the proposed algorithm, a test method is used by combining data from field test and off-line simulation. Firstly, field test was carried out to collect real data including GNSS and IMU data. Simulated noise was then imported to the GNSS data off-line to simulate the measurement accuracy degradation affected by interfered GNSS signal. And then the integration is executed with different algorithm including EKF, IAE and IAE-D.

42.2 Innovation-Based Adaptive Estimation

Innovation-based Adaptive Estimation is one of the adaptive Kalman filters and frequently performs data fusion function in GNSS/IMU integrated navigation. It has been proven to be able to improve the accuracy of Kalman filter in dynamic applications effectively by adapting $\boldsymbol{R}$ and $\boldsymbol{Q}$ [11]. IAE can adapt $\boldsymbol{R}$ or $\boldsymbol{Q}$ or both $\boldsymbol{R}$ and $\boldsymbol{Q}$ in practice. This paper concerns about adapting $\boldsymbol{R}$ because the measurement degradation is mainly reflected in the change of $\boldsymbol{R}$.

Substituting (42.1) to z_k^{INS} and z_k^{GNSS} in Eq. (42.6) gives,

$$\delta z_k = z_k^{INS} - z_k^{GNSS} = H_k \delta x_k + v_k - H_k \delta \hat{x}_k^- = H_k (x_k - \hat{x}_k^-) + v_k \tag{42.1}$$

Where δz_k is the difference between the estimated measurements and the actual measurement. It can be defined as innovation sequence. Combining Eqs. (42.10) and (42.8), the covariance of δz_k can be expressed as,

$$E(\delta z_k \delta z_k^T) = H_k \hat{P}_k^- H_k^T + E(v_k v_k^T) = H_k \hat{P}_k^- H_k^T + R_k \tag{42.2}$$

The innovation autocorrelation matrix $C_{\delta z_k}$ can be defined as the average value of m innovation sequence autocorrelation matrices,

$$C_{\delta z_k} = \frac{1}{m} \sum_{i=0}^{m-1} \delta z_{k-i} \delta z_{k-i}^T \tag{42.3}$$

where m is the window size for innovation sequences. Estimating the innovation sequences with yields,

$$\hat{E}(\delta z_k \delta z_k^T) = H_k P_k^- H_k^T + E(v_k v_k^T) = H_k P_k^- H_k^T + \hat{R}_k = C_{\delta z_k} \tag{42.4}$$

Rearranging Eq. (42.4) yields the following equation,

$$\hat{R}_k = C_{\delta z_k} - H_k \hat{P}_k^- H_k^T = \frac{1}{m} \sum_{i=0}^{m-1} \delta z_{k-i} \delta z_{k-i}^T - H_k \hat{P}_k^- H_k^T \tag{42.5}$$

Equation (42.5) is the adaptive equation for measurement noise matrix $\boldsymbol{R}$. Similarly, the adaptive equation for system noise matrix $\boldsymbol{Q}$ can be obtained [11],

$$\hat{Q}_k = \frac{1}{m} \sum_{i=0}^{m-1} \delta x_{k-i} \delta x_{k-i}^T + P_k^+ - \varphi_k P_{k-1}^+ H_k^T \tag{42.6}$$

42.3 Measurements Discarding Method

IAE has been proven to have better performance than traditional EKF, if high accurate measurements are available [11]. However, if the accuracy of measurements degrades, $\boldsymbol{R}$ will change rapidly, and the stability of IAE integrated system will be subjected to instability. Especially when the GNSS signal outage happens, the measurements will be deteriorated temporarily and greatly. It will be happen that unnecessary update with less qualified measurements is imposed on accurate estimation. As a result, in the worse case, the non-convergence happens, and the filter performance degrades rapidly. In order to mitigate the impact of less qualified measurements, the measurements should be monitored before using and the discarding method is therefore introduced [14].

Measurements discarding method, in which both the state estimation and measurements are monitored, discards the measurements when the estimations are accurate enough or the measurements are less qualified and leave the integrated system work solely with INS. In other word, the whole integrated system is working independent from the GNSS measurements, when the state estimation is accurate enough or the measurements are unqualified. The proposed discarding method algorithm contain two thresholds based on the development of tolerant estimation error level and is provided here.

Defining error cumulating rate V_{P_k} covariance as,

$$V_{P_k} = \frac{1}{n} \sum_{j=0}^{n-1} P_{k-i} \tag{42.7}$$

V_{P_k} indicates the accumulated estimation error level in a period of time, and can be calculated by averaging inside a moving estimation window of size n. Defining error cumulating rate factor as $\hat{V}_k$ the trace of V_{P_k} as,

$$\hat{V}_k = tr(V_{P_k}) = tr(\frac{1}{n}\sum_{j=0}^{n-1} P_{k-i}) \tag{42.8}$$

where $\hat{V}_k$ indicates the estimations error cumulating rate. If there is presenting a threshold V_{thr} properly, one can obtain a criteria for estimation correction,

$$\hat{V}_k > V_{thr} \tag{42.9}$$

Equation (42.9) indicates that if the accumulated error of the estimations is relatively small, the GNSS/IMU system can work with INS without correction. In other words, the measurements can be discarded.

Similarly, one can define measurements deterioration factor $\hat{O}_k$ as,

$$\hat{O}_k = tr(\delta z_k \delta z_k^T - C_{\delta z_{k-1}}) \tag{42.10}$$

It indicates the average error covariance between estimation and measurements in a period of time. The represents the $\hat{O}_k$ difference between the measurements error covariance obtained in a single time instant, and its estimation. If the measurements are less qualified due to additive interference, $\hat{O}_k$ is expected to be abnormally increased. If presetting the threshold properly, one can therefore obtain the second requirement for measurements correction.

$$\hat{O}_k < O_{thr} \tag{42.11}$$

Equation (42.11) indicates that although the error of estimations is large enough for correction, the measurements should be discarded temporarily because the measurements are unqualified. "The double thresholds measurements discarding method" expressed by Eqs. (42.9) and (42.11) mitigate the impacts of less qualified measurements on the performance of the integrated system through monitoring quality of the estimations and measurements, and discarding the less qualified measurements, which is proven by following test.

42.4 Test Setup and Data Processing Steps

Dynamic experiment has been carried out and simulated noise has been imported into the real data to investigate the performance of different Kalman filters EKF, IAE and proposed IAE-D. The experimental equipments include the Leika GPS receiver 'GPS 1200', the C-MIGITS II IMU and data collection laptop computers. The antenna and IMU were mounted on the roof of a car. The lever arm between the IMU and GPS was approximately 46 cm. Figure 42.1 illustrates the experimental setup.

C-MIGITS II is an integrated GPS/INS system, and contains a Digital Quartz IMU (DQI) and a GPS receiver. However, C-MIGITS II was used as an IMU

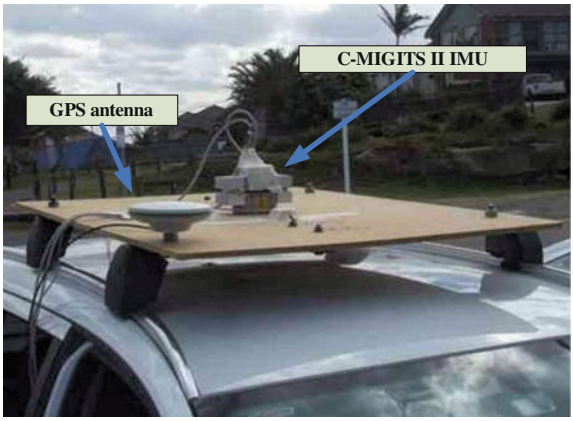

Fig. 42.1 GNSS/IMU test setup

Table 42.1 C-MIGITS II DQI error sources

Error sources	Labeled value
Gyro bias	5 deg/h (1σ)
Gyro scale factor	500 ppm (1σ)
Gyro random walk	0.09 deg/root-h (max)
	0.035 deg/root-h (normal)
Accelerometer bias	500 μg (1σ)
Accelerometer scale factor	800 ppm (1σ)
Accelerometer random walk	180 μg/root-Hz (max)
	60 μg/root-Hz (normal)

sensor and only the raw IMU outputs from the DQI were collected in the test. C-MIGITS II DQI is a tactical grade IMU, the major coefficients are shown in Table 42.1 [15].

The dynamic raw data were collected by field test carried out at a sea beach road. Nearly two circles around the test field were made and more than four satellites in good view of line-of-sight during the test. The real data collected in the test included GPS data, IMU data collected by Leika receiver and C-MIGITS II IMU respectively and the differential data from the reference station.

The post-processing steps are listed as follows:

1. Process the GPS data and differential data by using LEICA Geo OfficeTM software to get the centimeter level GNSS measurements.
2. Integrate GNSS measurements and the IMU data by the traditional EKF to get the 100 Hz trajectory, in which the estimating rate and updating rate are 100 Hz and 1 Hz respectively. The results could be regarded as the centimeter level reference trajectory.
3. Import the white noise into the centimeter level GNSS measurements to simulate the additive white noise interference to the measurements so that the accuracy degrades. The degraded measurements were then integrated with the

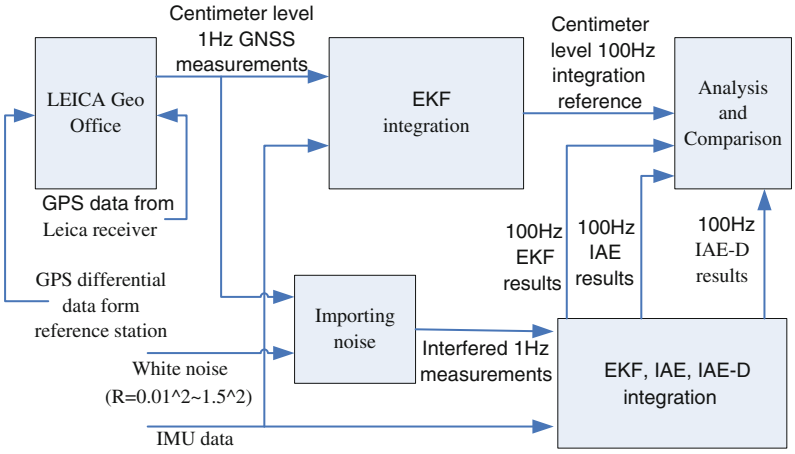

Fig. 42.2 Data processing flow chart

IMU data by EKF, IAE and IAE-D. Differentiate the results with the reference in step 2 to obtain the error values.

4. Repeat step 3 using the noise with the autocorrelation R ranging from 0.012 to 1.52 and stepping 0.012. Analyze the results and compare the performance among EKF, IAE and IAE-D in different measurements accuracy. Figure 42.2 shows the data processing flow chart.

42.5 Results and Analysis

42.5.1 Reference Trajectory

The positioning result integrating centimeter level GPS measurements with IMU by EKF is shown in Fig. 42.5. It is regarded as the reference and all the following integrating results will be compared with it. The experimental vehicle had collected the static data at point A for 3 min, and then moved clockwise around the field. The white noise was imported from point B, after the initial alignment completed. There were approximately 3-min dynamic data collected for integrating calculation from point B to point C. (Fig. 42.3)

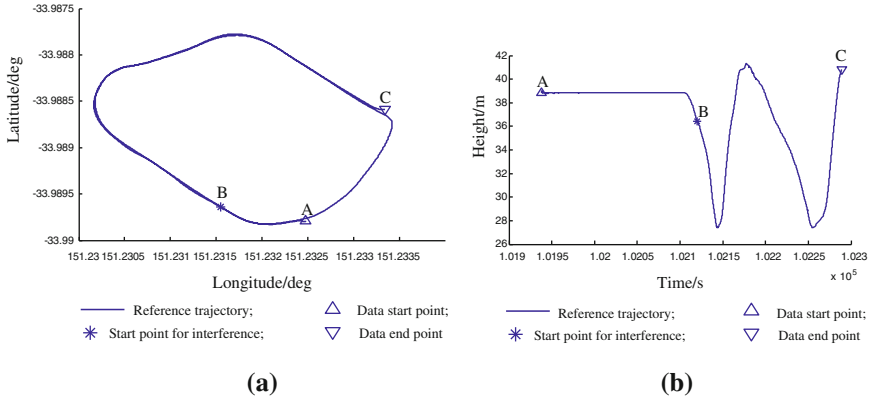

Fig. 42.3 Experimental reference trajectory, **a** Reference ground trajectory **b** Reference height

42.5.2 Overall Performance Comparison

In order to analyze the algorithms performance in different interference levels, this integration process is then repeated using different additive noise level which leads to different autocorrelation $\boldsymbol{R}$ ranging from $0.1^2/m^2$ to $1.5^2/m^2$ and stepping $0.01^2/m^2$. The results show the performance comparison among these three algorithms in different interference power levels. The results are plotted in Fig. 42.4. In addition, the discarding rate for IAE-D is plotted in Fig. 42.4.

Figure 42.4a shows the positioning performance comparison. When $\boldsymbol{R}$ is less than $0.9^2/m^2$, IAE has less positioning error than EKF, and the error of IAE is only 50 % of that of EKF when $\boldsymbol{R}$ is about $0.6^2/m^2$. These results illustrates that IAE can mitigate the interference of the measurements deterioration to the positioning performance of the integrated system when the noise level is relatively low. However, the IAE positioning error would exceed EKF and grow rapidly when interference $\boldsymbol{R}$ is larger than $0.9^2/m^2$. The IAE positioning error would be increased to about four times as that of EKF when $\boldsymbol{R}$ equals to $1.5^2/m^2$.

When $\boldsymbol{R}$ is in the range of 0 to $0.6^2/m^2$, IAE-D positioning error is between that of EKF and IAE, and the discarding rate is kept less than 0.05. Nevertheless, as shown in Fig. 42.5, the discarding rate of IAE-D will be increased to about 0.32 rapidly while $\boldsymbol{R}$ grows from $0.6^2/m^2$ to $1.2^2/m^2$. Moreover, the positioning error of IAE-D is the smallest and maintains the estimation error at about one third of that of EKF on condition that the noise level is above $0.6^2/m^2$. The Fig. 8a illustrates that IAE-D resistant to the impact on the positioning performance degradation due to additive interference.

As the velocity error is concerned, Fig. 42.4b illustrates that IAE and IAE-D have advantages over EKF in the range of simulated noise level. IAE velocity error maintains at about one fourth or one third of that of EKF. IAE-D velocity error is slightly larger than that of IAE when $\boldsymbol{R}$ is in the range of $0–0.8^2/m^2$ and smaller than that of IAE when $\boldsymbol{R}$ grows beyond $0.8^2/m^2$.

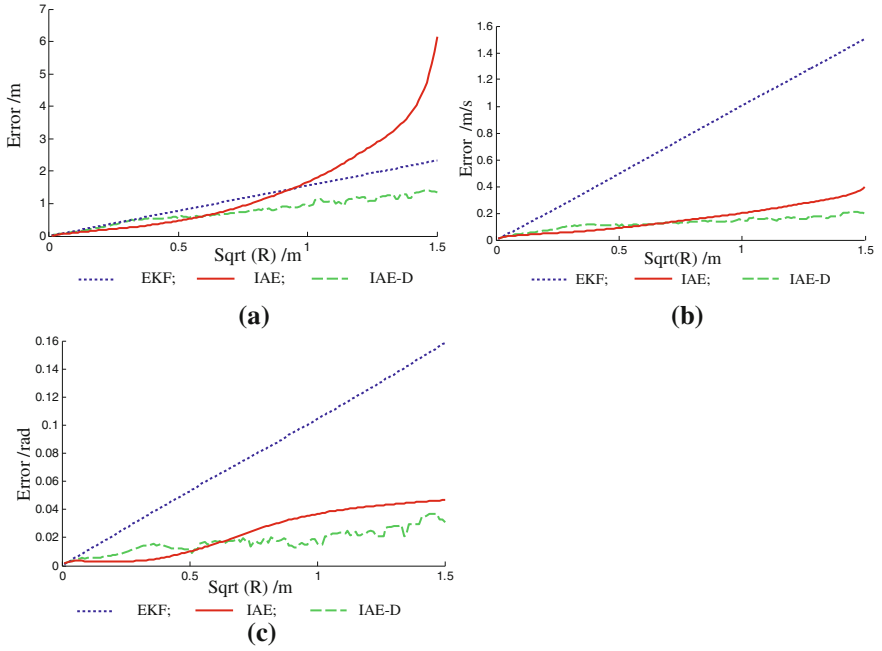

Fig. 42.4 The integrated results comparison ($R = 0.01^2$–$1.5^2/m^2$), **a** Position error **b** Velocity error, **c** Altitude error

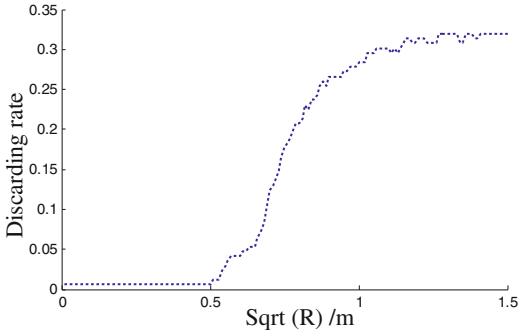

Fig. 42.5 Discarding rate for IAE-D ($R = 0.01^2$–$1.5^2/m^2$)

Figure 42.4c shows the attitude comparison and it indicates that the attitude error of IAE and IAE-D are always smaller than that of EKF in the range of simulated noise level. IAE attitude error maintains at about one fourth to half of that of EKF. The attitude error of IAE-D is slightly larger than IAE when $\boldsymbol{R}$ is in the range of 0–$0.7^2/m^2$ and smaller than IAE when $\boldsymbol{R}$ grows beyond $0.7^2/m^2$.

42.6 Concluding Remarks

IAE improves the accuracy of GNSS/IMU integrated navigation system compared with the traditional EKF if the high accurate GNSS measurements are available. When measurements accuracy degrades due to GNSS signal interference, IAE could suffer from bias and non-convergence. Furthermore, when the additive interference increases to some level ($\boldsymbol{R} = 0.9^2/\mathrm{m}^2$ in the test of this paper), the error of GNSS measurements can be larger than estimations error, which on the other hand leads to the dramatic deterioration of navigation.

In order to improve the performance of IAE on condition that less accurate measurements are used, a modified IAE algorithm (IAE-D) is proposed. The algorithm monitors the quality of the estimations and the measurements in the real-time, and discards the measurements when the estimations are accurate enough or the measurements are less qualified. The results of field test and simulation comparisons have demonstrated that IAE-D has magnificent advantage over EKF and IAE in regard to the performance of position, velocity and attitude when the measurements deterioration level is relatively high. The encouraging results show the provided IAE-D has potential application in GNSS/IMU navigation using less accurate GNSS receiver or in the environments with high additive interference.

References

1. Grewal MS, Weill LR, Andrews AP (2007) Global positioning systems inertial navigation and integration. Wiley, New Jersey
2. Groves PD (2008) Principles of GNSS, inertial, and multisensor integrated navigation systems. Artech House Press, Boston
3. Cox DB (1980) Integration of GPS with inertial navigation systems. The ION "Redbooks" 1:144–153
4. Greenspan RL (1996) GPS and inertial integration. Global Position Syst Theory Appl 1:87–220
5. Li P, Wang J, Feng Z (2008) GNSS/pseudolite signal re-acquisition with the aid of INS in short signal blockage scenarios. In: Proceedings of the international symposium on GPS/GNSS 2008
6. Grewal MS, Andrews AP (2008) Kalman filtering: theory and practice using MATLAB, 2nd edn. Wiley, New York
7. Toda NF, Schlee FH, Obsharsky P (1969) Region of Kalman filter convergence for several autonomous navigation modes. AIAA J 7:622–627
8. White NA, Maybeck PS, Devilbiss SL (1996) MMAE detection of interference/jamming and spoofing in a DGPS-aided inertial system. Master thesis of Ohio University, Ohio
9. Hu C, Chen Y, Chen W (2001) Adaptive Kalman filtering for DGPS positioning. In: Proceedings of the international symposium on Kinematic systems in geodesy, geomatics and navigation (KIS)
10. Ding WD, Wang J, Chris R et al (2007) Improving adaptive Kalman estimation in GPS/INS integration. J Navig 60:517–529
11. Mohamed AH, Schwarz KP (2003) Adaptive Kalman filtering for INS/GPS. J Geodesy 73:193–203

12. Hide C, Moore T, Smith M (2003) Adaptive Kalman filtering for low cost GPS/INS. J Navig 56:143–152
13. Li P, Lu M, Feng Z (2010) Positioning accuracy analysis for adaptive Kalman filtering in measurements degradation. Syst Eng Electron 32(7):1489–1492
14. Lu M, Li P, Feng Z (2009) Adaptive updating strategy for GNSS/IMU integration. In: Proceedings of the 22nd international technical meeting of the satellite division of the institute of navigation, Savannah, Georgia, pp 750–758
15. Boeing North American Inc (1997) C-MIGITS II user's guide. Boeing defense & space group

Errata to—Chapter 1: Space-Based MA TT&C System and Technologies; Chapter 15: Flexible Hemispherical Simultaneous Multi-Beam TT&C Technology

Jianping Hu, Hongjun Yang, Maoge Xu, Pengyi Wang, Yongfei Kong and Haizhou Wu

Errata to:

Chapter 1: Space-Based MA TT&C System and Technologies, doi:10.1007/978-3-642-33663-8_1
Chapter 15: Flexible Hemispherical Simultaneous Multi-Beam TT&C Technology, doi:10.1007/978-3-642-33663-8_15

Pages	Item or line	Corrections
9	Chapter 1, Section 1.3.4, first paragraph	Replace "There are many methods for multi-beamforming on ground, such as main beam control(open loop), scan (open loop), and adaptive beamforming (close loop) [9]". by "There are many methods for multi-beamforming on ground, such as main beam control(open loop), scan (open loop), and adaptive beamforming (close loop) [6]"
11	Chapter 1, Reference Zhang JS (1979)	The reference "Zhang JS (1979) Tracking and data relay satellite overview. J Spacecraft TT&C Technol" should be removed from the reference list
11	Chapter 1, Reference Hu JP (2004)	The reference "Hu JP (2004) Research of TDRSS user transponder technology. Telecommun Eng 44(3) (Supp.)" is being renumbered as [8] reference
147	Chapter 15, Figure. 15.2	Replace old Fig. 15.2 by new Fig. 15.2 (shown below)

The online version of the original chapters can be found at DOI: 10.1007/978-3-642-33663-8_1 and 10.1007/978-3-642-33663-8_15.

J. Hu (✉) · H. Yang · M. Xu
Southwest China Institute of Electronic Technology, Chengdu 610036, China
e-mail: jphu63@sina.com

P. Wang (✉) · Y. Kong · H. Wu
The 54th Research Institute, CETC, Shijiazhuang 050081, China
e-mail: wpybox1000@tom.com

R. Shen and W. Qian (eds.), *Proceedings of the 26th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 187,
DOI: 10.1007/978-3-642-33663-8_43,

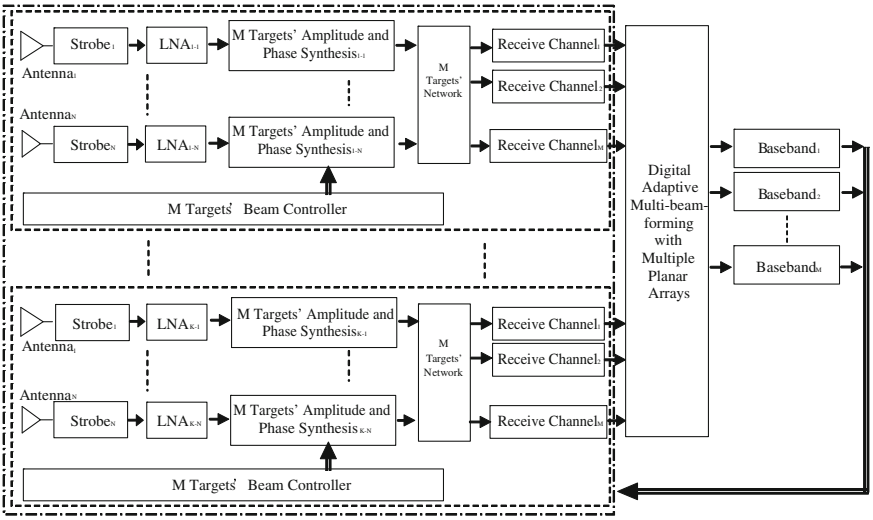

Fig. 15.2 Receiving framework of active phased array with N channels and M targets