Prashant Pillai
Kandeepan Sithamparanathan
Giovanni Giambene
Miguel Ángel Vázquez
Paul Daniel Mitchell (Eds.)

LNICST

231

L N I C S T

Wireless and Satellite Systems

9th International Conference, WiSATS 2017
Oxford, UK, September 14–15, 2017
Proceedings

EAI

Springer

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 231

Prashant Pillai · Kandeepan Sithamparanathan
Giovanni Giambene · Miguel Ángel Vázquez
Paul Daniel Mitchell (Eds.)

Wireless and Satellite Systems

9th International Conference, WiSATS 2017
Oxford, UK, September 14–15, 2017
Proceedings

Ⓩ Springer

Editors
Prashant Pillai
University of Wolverhampton
Wolverhampton
UK

Kandeepan Sithamparanathan
RMIT University
Melbourne, VIC
Australia

Giovanni Giambene
University of Siena
Siena
Italy

Miguel Ángel Vázquez
Centre Tecnològic de Telecomunicacions
  de Catalunya
Barcelona
Spain

Paul Daniel Mitchell
University of York
York
UK

Preface

We are delighted to introduce the proceedings of the ninth edition of the 2017 European Alliance for Innovation (EAI) International Conference on Wireless and Satellite Systems (formerly PSATS), which was held in the historic city of Oxford, UK, during September 14–15, 2017.

The aim of this conference is to bring together researchers, developers, and practitioners from around the world working in the field of wireless and satellite systems. The theme of WiSATS 2017 was on the means of bringing wireless and satellite services directly to the user for personal communications, multimedia, and location identification.

The technical program of WiSATS 2017 comprised 19 full papers, including seven papers in the main conference track. Other papers were presented as part of a series of workshops, including: Next-Generation mmWave and Optical Satellite Systems: From Channel Modelling to System Performance Evaluation (SATPROP: five papers), Unmanned Aerial Systems (IWUAS: five papers), and Communication Applications in Smart Grid (CASG: two papers). In addition to the high-quality technical paper presentations, the technical program also featured two keynote speeches, an invited talk, and a panel session. The two keynote speeches were given by Dr. Hector Fenech (Director of Future Satellite Systems at Eutelsat, France) and Vincenzo Pellegrini (Senior Radiocommunication and SDR Systems Engineer at IDS, Ingegneria dei Sistemi S.p.A, Italy). The invited talk was presented by Prof. Fun Hu from the University of Bradford, UK, and the panel session was on the subject of "SatCom in 5G and Beyond: Opportunities and Challenges."

We would like to thank all members of the Organizing and Steering Committees for their effective cooperation in putting together and delivering the conference. It was a great pleasure to work with such an excellent team. In particular, the Technical Program Committee, led by our TPC co-chairs, Giovanni Giambene and Miguel Ángel Vázquez, who managed the peer-review process. We are also grateful to our conference manager, Lenka Bilska, and her colleagues Erika Pokorna and Dominika Kalafutova for their assistance. We would like to thank all the authors who submitted papers to the WiSATS 2017 conference and contributed to its success. A special thanks also to all the sponsors for their support.

We strongly believe that the WiSATs conference provides a good forum for all researchers, developers, and practitioners to discuss science and technology pertaining to wireless and satellite systems. We expect that the next WiSATS conference will be as successful and stimulating as this edition, as indicated by the contributions presented in this volume.

January 2018

Paul Daniel Mitchell
Prashant Pillai
Kandeepan Sithamparanathan

Organization

Steering Committee

Imrich Chlamtac (Chair)	EAI/Create-Net, Italy
Kandeepan Sithamparanathan	RMIT, Australia
Mario Marchese	University of Genoa, Italy
Agnelli Stefano	ESOA/Eutelsat, France
Prashant Pillai	University of Wolverhampton, UK

Organizing Committee

General Co-chairs

Prashant Pillai	University of Wolverhampton, UK
Kandeepan Sithamparanathan	RMIT University, Australia

Technical Program Committee Chairs

Giovanni Giambene	University of Siena, Italy
Miguel Ángel Vázquez	CTTC, Spain

Local Chair

Shumao Ou	Oxford Brookes University, UK

Workshops Chairs

Daniele Tarchi	University of Bologna, Italy
Sooyoung Kim	Chonbuk National University, South Korea

Publicity and Social Media Chairs

Celimuge Wu	The University of Electro-Communications, Japan
Constantinos T. Angelis	TEIE, Greece

Publications Chair

Paul Daniel Mitchell	University of York, UK

Web Chair

Atm S. Alam	University of Surrey, UK

Panels Chair

Bhavani Shankar   University of Luxembourg, Luxembourg

Industry Chair

Alberto Ginesi   European Space Agency

Tutorials Chair

Aduwati Sali   University of Putra Malaysia, Malaysia

Conference Manager

Lenka Bilska   European Alliance for Innovation (EAI)

Technical Program Committee

Claudio Cicconetti   MBI, Italy
Atm Alam   University of Bradford, UK
Giovanni Giambene   Università degli Studi di Siena, Italy
Raed Abd-Alhameed   University of Bradford, UK
Yongqiang Cheng   University of Hull, UK
Alban Duverdier   CNES, France
Gabriele Oligeri   University of Rome Tre, Italy
Ana Perez   UPC, Spain
Angeles Vazquez-Castro   Universidad Autónoma de Barcelona, Spain
George Eleftherakis   The University of Sheffield International Faculty,
                  CITY College, Greece
Konstantinos Liolis   SES S.A., Luxembourg
Misha Filip   University of Portsmouth, UK
Djuradj Budimir   University of Westminister, UK
Robert Ewatson   University of Bath, UK
Haitham Cruickshank   University of Surrey, UK
Haile-Selassie Rajamani   University of Bradford, UK
Ray Sheriff   University of Bradford, UK
Alister Burr   University of York, UK
Franco Davoli   DITEN, University of Genoa, Italy

Contents

CASG Workshop

SATPROP Workshop Session

Security Issues in Satellite Networking

Security Aware Virtual Base Station Placement in 5G Cloud Radio Access Networks

Tshiamo Sigwele[1(✉)], Prashant Pillai[2], Abimbola Sangodoyin[1], and Yim Fun Hu[1]

[1] Faculty of Engineering and Informatics, University of Bradford, Bradford, West Yorkshire BD7 1DP, UK
{T.Sigwele,a.o.sangodoyin,Y.F.Hu}@bradford.ac.uk
[2] Faculty of Technology, Design and Environment, Oxford Brookes University, Oxford, UK
ppillai@brookes.ac.uk

Abstract. In fifth generation (5G) cloud radio access networks (C-RAN), baseband processing of base stations (BS's) will be processed on virtual machines called virtual BSs (VBS) in the centralized cloud architecture. The existing researches mostly focus on how to maximize resource utilization and reduce energy consumption in 5G C-RAN using VBS placement. However, security issues in the context of VBS placement within 5G C-RAN have been rarely addressed. In this paper, a security aware VBS placement (SAV) scheme within 5G C-RAN is proposed where the placement of VBSs to physical machines (PMs) considers the security levels of both the VBS and the PM. A rigorous simulation study is conducted for validating the proposed scheme, which shows a significant security improvement of 16% compared to the heuristic simulated annealing scheme (HSA).

Keywords: 5G · Cloud computing · C-RAN · Cloud security
Virtual machine placement

1 Introduction

The fifth generation (5G) cellular networks will experience a thousand-fold increase in data traffic with over 100 billion connected devices by 2020 [1]. Such surge in traffic will be from smart-phones, tablets, machine-machine connections and the Internet of Things (IoT). In order to support this sky-rocketing traffic demand, heterogeneous cloud radio access H-CRAN networks has been proposed where macro remote radio heads (RRHs) are overlaid by smaller cell RRHs like femto, pico, micro and relay to increase capacity using spatial frequency reuse. The 5G C-RAN uses cloud computing virtualization techniques to host basestation (BS) functions in virtual machines (VMs) called virtual BSs (VBSs) in the BS cloud [2]. One of the driving forces in the BS cloud is VBS

placement where VBS are migrated among physical machines (PMs) to maximise BS resource utilization and reduce energy consumption. Nevertheless, security issues in the BS cloud in C-RAN have been rarely addressed. The introduction of running BS functions in PMs in the BS cloud brings about security issues. A prevalent VBS image with known vulnerabilities can be instantiated by an attacker in BS cloud, therefore it may generate a large number of security holes for attackers which may include eavesdropping users conversation. The introduction of VBS images with known security vulnerabilities to a PM can lead to security risks to the co-located VBSs in that PM. This is because of the one to many mapping between the PM and the VBSs which makes vulnerabilities propagate rapidly across the entire BS cloud. Some of the attacks that an attacker can introduce include compromising the hypervisor and also side channel attacks to the VBSs co-located within the same PM [3]. A VBS that has been compromised can infect other VBS sharing the same hypervisor, memory or CPU. When a VBS with no security risks is migrated to the PM with a compromised VBS, that VBS will be compromised too. As such there is need for a security aware VBS placement in the BS cloud to avoid the security risks.

In this paper, a VBS and PM security evaluation based on their vulnerabilities is first conducted then, based on these security evaluations, a novel security aware VSB placement is developed which minimises the security risks in the BS cloud. The VBSs and PMs are grouped into vulnerability levels such that a VBS with low vulnerability will be allocated to a PM with low vulnerability and a VBS with high vulnerability will be allocated to a PM with high vulnerability. The rest of the paper is organised as follows. Section 2 describes the related work on VBS placement in C-RAN. The proposed security aware VBS placement framework is described in details in Sect. 3. The results and discussions are presented in Sect. 4. Then Finally the conclusions are presented in Sect. 5.

2   Related Work

Researchers have proposed several VBS placement strategies to improve resouce utilization and improve energy consumption within 5G C-RAN. However, to the best of our knowledge, there are no efforts on VBS placement strategies in 5G C-RAN to minimize the security risks for the BS cloud platform. The author in [4] proposed a VBSs virtualization scheme that minimizes the power consumption with a linear computational complexity order. The scheme is based on a heuristic simulated annealing (HSA) algorithm, which combines a bin packing algorithm with simulated annealing. Simulation results show that the HSA effectively decreases system power consumption when compared to standard approaches. However, the simulated annealing VBS placement scheme does not consider security in migration of VBSs. Authors in [5] proposed a BBU reduction scheme for C-RAN that allocates VBSs to RRHs based on the imbalance of subscribers in office/residential areas. A set of upper limit of VBS utilization is defined to avoid overloading of the VBS. However, the author did not consider security in their VBS consolidation scheme. Namba et al. in [6] proposed a baseband unit (BBU) reduction network architecture called Colony-RAN due to its

ability to flexibly change cell layout by changing the connections of BBUs and RRHs in respect to traffic demand. However, the proposed method has frequent ping pong reselections of RRH to BBU. The author in [7] proposed a model for reducing power consumption in H-CRAN by turning off the BBU in the cloud. However, the author assumes that all the BBUs in the BS cloud operate at full load which is unrealistics. The author in [8] proposed a VBS virtualization scheme that minimizes the power consumption of the BS cloud. The VBS virtualization problem is formulated as a bin packing problem, where each VBS is treated as a bin with finite computing expressed in million Operations per Time Slot (MOPTS). The dynamics of the cell traffic load is treated as an item that needs to be packed into the bins with the size equal to the computing resources in MOPTS, required to support the traffic load. Nevertheless, security has not been considered. Authors in [9] proposed a C-RAN system using virtualization technology on general purpose processors (GPPs) where BBUs are dynamically provisioned according to traffic load. In [10], Cheng et al. developed an energy efficient C-RAN system with a reconfigurable backhaul that allows 4 BBUs to connect flexibly with 4 RRHs using radio-over-fiber technology. The backhaul architecture allows the mapping between BBUs and RRHs to be flexible and changed dynamically to reduce energy consumption in the BBU pool. However, the paper assumes static user traffic whereas in reality BS traffic is dynamic.

3 Proposed Security Aware VBS Placement (SAV)

3.1 System Model

The proposed architecture of SAV is shown in Fig. 1. The system comprise of 5G heterogeneous network which is made up of macro RRH overlaid by femto, pico, micro and relay RRHs. The RRHs from the radio side are connected to the BS cloud via high speed, low latency fiber cables. These connection is called the fronthaul [2]. The fronthaul in the BS cloud is first connected to the dispatcher which route data to their respective VBSs within physical machines (PMs) according to some VBS placement rules from the SAV module in the controller. The SAV module in the controller contains the VBS placement framework which will be explained in details in Sect. 3.3. The SAV takes all VBSs and PMs as inputs and output the VSB-PM map. The BS cloud also comprise of the PMs which host VBSs. The PMs contains hypervisors which runs on top of the PM operating system (OS). The hypervisor is responsible for creating VBS on the PMs. Each RRH has its own VBS.

3.2 VBS and PM Security Evaluation

The security evaluation procedure consists of evaluating the risks of both the VBSs and PMs in the BS cloud.
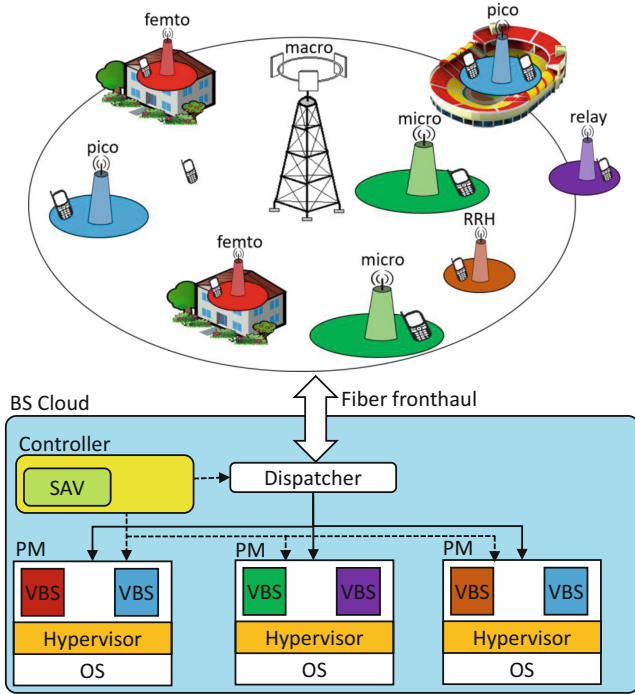
Fig. 1. The proposed SAV architecture.

VBS Security Evaluation: First, we quantify each VBS's vulnerabilities based on the US National Vulnerability Database (NVD) [11], in which all vulnerabilities called common vulnerability and exposures (CVE) are scored according to the Common Vulnerability Scoring System (CVSS) [3]. The NVD is the repository which provides CVSS scores for all CVE vulnerabilities. At present, NVD contains information about over 69,000 CVE vulnerabilities (as of 04/27/2015) [3]. NVD was created by the government of United States to help the Department of Homeland Security to warn public about common computer vulnerabilities. These vulnerabilities now include latest attacks on cloud computing environments. The CVSS base score is the primary metric and describes the severity of the vulnerability. The base score uses an interval scale of (0, 10) to measure the severity of vulnerabilities.

First a check is perfomed with NVD to collect potential vulnerabilities in the VBSs. Vulnerability scanner tools, such as, Nessus and Qualys are available to conduct this job. Since it is possible for a VBS to have more than one vulnerabilities, it is usually desirable to aggregate the scores of individual vulnerabilities for each VBS. The VBS vulnerability can be divided into three descrete levels: low compromise, medium compromise and high compromise. Suppose the collected vulnerabilities of all the VBSs within PM_i are stored in the set $V_{PM_i} = \{V_j | j = 1, 2, ...n\}$. Then the compromise level of a VBS_j denoted C_{VBS_j} can be given as:

$$C_{VBS_j} = \frac{V_j}{\sum_{m=1}^{n} V_m} \qquad (1)$$

Physical Machine Security Evaluation: The VBS compromise level has been computed. Next the probability of survivability for each PM is computed based on the compromise level (C_{VBS_j}) of each hosted VBS. The survivability probability is the possibility that all owned VBSs of a PM can survive during the attacks in one or more of the VBSs. If any of the VBSs in the PM is compromised, then the PM will also be compromised with high probability. Given the compromise level of all VBSs in PM_i as $\{C_{VBS_1}, C_{VBS_2}, ..., C_{VBS_n}\}$, then the survivability score for PM_i denoted as S_{PM_i} is given as:

$$S_{PM_i} = \prod_{m=1}^{n} (1 - C_{VBS_j}) \qquad (2)$$

The survivability quantifies the PM security level which correspond to three discrete states: low survivability, medium survivability and high survivability. Next the VBS placement based on the VBS and PM security levels is described in the next section.

3.3   The SAV Framework

From the previous discussions, we can learn that the success of attacks highly depends on the placement strategy of the cloud. Thus, our approach is to find a systematic solution to place VBSs into PMs which can reduce security risks in the BS cloud. If a VBS has low compromise, it will be infeasible to place it on a PM with low survivability. Also, if a PM is of high survivability, it would be infeasible to place a VBSs with high compromise which may results in a negative survivability impacts on the PM. In the proposed SAV scheme, the compromise level of the VBS and the survivability of the PM are taken into consideration when performing VBS placement. In SAV, it will be reasonable to place a VBS of low compromise level to a PM with high survivability. Also, in SAV, the VBSs are placed to PMs such that the C_{VBS_j} state match the S_{PM_i} state as follows:

 (i) A low compromise VBS is placed on high survivability PM
 (ii) A medium compromise VBS is placed on medium survivability PM
(iii) A high compromise VBS is placed on low survivability PM

Figure 2 shows the flowchart of how the proposed SAV scheme operate. The input to the SAV scheme is all the VBSs and all the PMs in the BS cloud. First for every VBS, the security evaluation is performed by computing the comprise level of each VBS using Eq. (1). Then the compromise level (C_{VBS_j}) is used for calculating survivability of each PM (S_{PM_i}). A test is then performed on each VBS to check whether the compromise state of that VBS match the survivability of the PM. If not, then all the other VBS compromise levels are checked and if they match the PM suravibility state, the VBS is then allocated to the PM. The process will continue untill all VBSs are allocated to PMs securely.
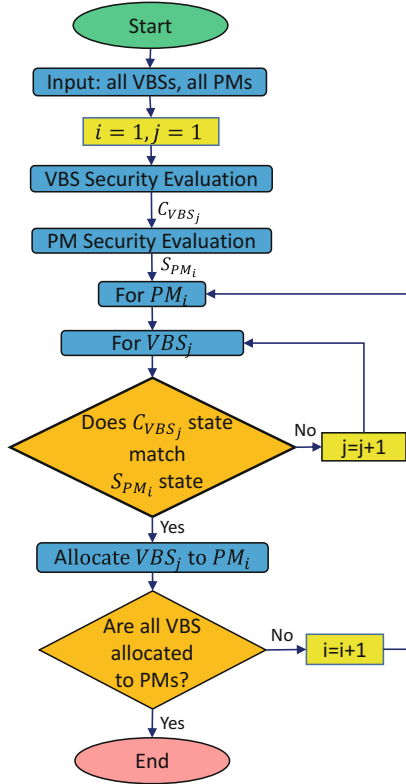
Fig. 2. The flowchart of SAV framework.

4   Results and Discussion

To analyse the proposed SAV framework, a 2 tier H-CRAN long term evolution
(LTE) layout is considered where macro RRHs are overlaid by small cells RRHs.
Up to 200 RRHs are considered and each RRH has its own VBS in the BS cloud.
As explained before, to the best of our knowledge, there are no security aware
VBS placement schemes in 5G C-RAN at the moment, this paper is the first
to proposed such framework. Therefore, the proposed SAV will be compared
with the HSA scheme in [4] which consider VBS placement for saving energy
consumption in the BS cloud without considering security. Figure 3 shows the
effects of increasing the number of VBSs in the network on the PM survivability
for all the schemes. For both the proposed SAV scheme and the HSA schemes,
as the number of VBSs in the BS cloud increases, the PM survivability decreases
because the chances of comprimised VBSs are higher with the increase of VBSs.
Also it can be observed in the diagram than the SAV scheme performs better than
the HSA scheme by 16% with high chances of survivability of an average of 0.65
compared to the HSA scheme with chances of survivability of 0.5. This is because
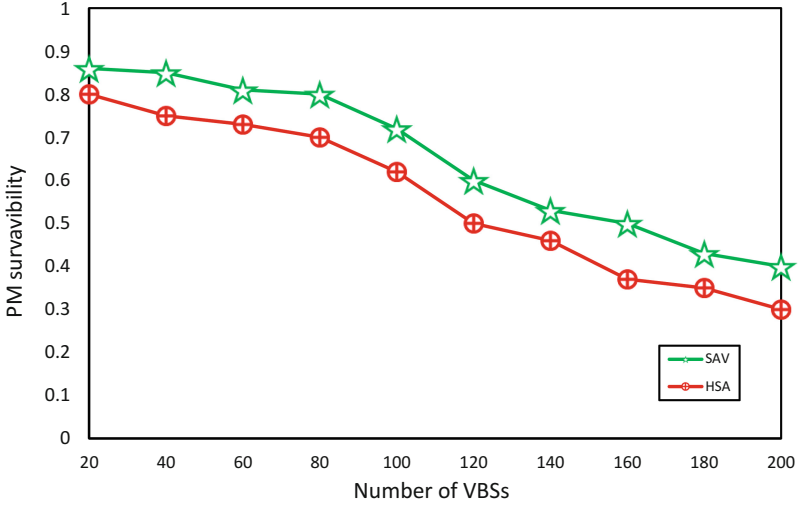
Fig. 3. The effects of VBS variation on PM survivability.
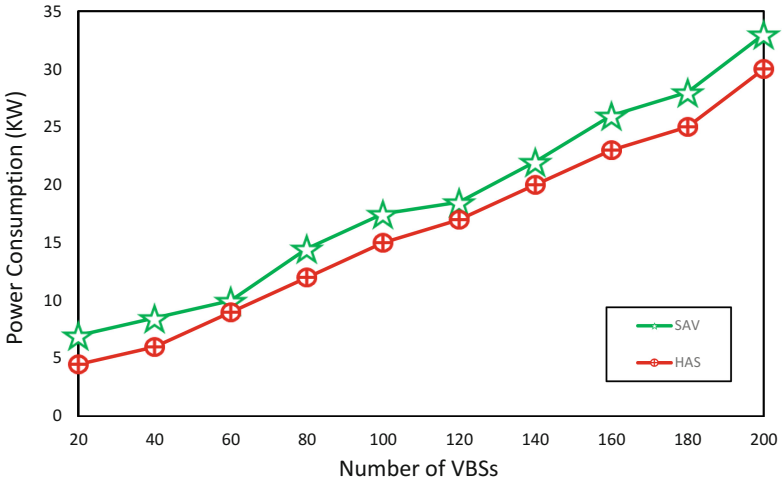
Fig. 4. The effects of VBS variation on power consumption.

the SAV scheme considers security of VBSs and PMs before performing the VBS placement. Figure 4 shows the effects of varying the number of VBSs on the power consumption in the BS cloud. For both the schemes, as the number of VBSs increases, the power consumption also increases as more VBSs requires more PMs which will consume more power. On average, the SAV scheme consumes 15% more energy compared to the HSA scheme. This is because, even though

the SAV scheme provide some security benefits, this results in the penalty of more power of 15% being consumed by the SAV scheme as a results of the VBS placement scheme security overheads which requires more PMs to be deployed.

5   Conclusion

The existing researches mostly focus on how to maximize resource utilization and reduce energy consumption in fifth generation (5G) cloud radio access networks (C-RAN) using virtual base station (VBS) placement. However, security issues in the context of VBS placement within 5G C-RAN have been rarely addressed. In this paper, a security aware VBS placement (SAV) scheme is proposed where the placement of VBSs to physical machines (PMs) considers the security levels of both the VBS and the PM. A rigorous simulation study is conducted for validating the proposed scheme, which shows a significant security improvement since the proposed SAV scheme outperforms the heuristic simulated scheme (HSA) by 16% with power consumption penalty of 15%.

References

1. 5G: A Technology Vision, Huawei Tech. White Paper (2013)
2. Chen, K., Duan, K.: C-RAN: the road towards green RAN. In: China Mobile Research Institute White Paper (2011)
3. Yuchi, X., Sachin, S.: Enabling security-aware virtual machine placement in IaaS clouds. In: Military Communications Conference, MILCOM 2015, pp. 1554–1559. IEEE (2015)
4. Qian, M., Hardjawana, W., Shi, J., Vucetic, B.: Baseband processing units virtualization for cloud radio access networks. IEEE Wirel. Commun. Lett. **4**, 189–192 (2015)
5. Namba, S., Warabino, T., Kaneko, K.: BBU-RRH switching schemes for centralized RAN. In: 2012 7th International ICST Conference on Communications and Networking in China (CHINACOM), pp. 762–766. IEEE (2012)
6. Namba, S., Matsunaka, T., Warabino, T.: Colony-RAN architecture for future cellular network. In: Network and Mobile Summit (FutureNetw), pp. 1–8. IEEE (2012)
7. Domenico, A., Katranaras, E.: Energy efficiency benefits of RAN-as-a-service concept for a cloud-based 5G mobile network infrastructure. IEEE Access **2**, 1586–1597 (2014)
8. Qian, M., Hardjawana, W., Vucetic, B.: Baseband processing units virtualization for cloud radio access networks. IEEE Wirel. Commun. Lett. **4**(2), 189–192 (2015)
9. Kong, Z., Gong, C., Xu, Z.: eBase: a baseband unit cluster testbed to improve energy-efficiency for cloud radio access network. In: IEEE International Conference on Communications (ICC), pp. 4222–4227. IEEE (2013)
10. Liu, C., Sundaresan, K., Jiang, M., Rangarajan, S., Chang, G.K.: The case for re-configurable backhaul in cloud-RAN based small cell networks. In: 2013 Proceedings of IEEE INFOCOM, pp. 1124–1132. IEEE (2013)
11. United State Government: US National Vulnerability Database. https://nvd.nist.gov. Accessed Mar 2017

DoS Attack Impact Assessment
on Software Defined Networks

Abimbola Sangodoyin[1(✉)] , Tshiamo Sigwele[1], Prashant Pillai[2],
Yim Fun Hu[1], Irfan Awan[1], and Jules Disso[3]

[1] Faculty of Engineering and Informatics, University of Bradford,
Bradford, West Yorkshire BD7 1DP, UK
{a.o.sangodoyin,T.Sigwele,Y.F.Hu,i.awan}@bradford.ac.uk
[2] Faculty of Technology, Design and Environment, Oxford Brookes University,
Oxford, UK
ppillai@brookes.ac.uk
[3] Nettitude Limited, Warwickshire CV31 3RZ, UK
jpagnadisso@nettitude.com

Abstract. Software Defined Networking (SDN) is an evolving network
paradigm which promises greater interoperability, more innovation, flex-
ible and effective solutions. Although SDN on the surface provides a
simple framework for network programmability and monitoring, few has
been said about security measures to make it resilient to hitherto secu-
rity flaws in traditional network and the new threats the architecture is
ushering in. One of the security weaknesses the architecture is ushering
in due to separation of control and data plane is Denial of Service (DoS)
attack. The main goal of this attack is to make network resources unavail-
able to legitimate users or introduce large delays. In this paper, the effect
of DoS attack on SDN is presented using Mininet, OpenDaylight (ODL)
controller and network performance testing tools such as iperf and ping.
Internet Control Message Protocol (ICMP) flood attack is performed on
a Transmission Control Protocol (TCP) server and a User Datagram Pro-
tocol (UDP) server which are both connected to OpenFlow switches. The
simulation results reveal a drop in network throughput from 233 Mbps
to 87.4 Mbps and the introduction of large jitter between 0.003 ms and
0.789 ms during DoS attack.

Keywords: Software Defined Networks · DoS · Network security

1 Introduction

Computer networks have become part of our everyday lives from government to
commercial enterprises to individuals [1]. These networks are built from large
number of devices such as routers, switches and middle boxes with complex pro-
tocols running on them. Network administrators are saddled with the responsi-
bility of configuring these vendor-specific devices and configuration policies are

implemented on them. As a result, network management and dynamic response to events and applications are arduous and prone to error.

In addition to the complexity of configuration, operators have little options or mechanisms to respond to difficulties and enforce the required policies in dynamic environments [2]. Similarly, in the face of growing traffic and demand for more data rate from consumers, the service providers need to keep up with the pace through investments in bigger and faster links and edge routers, even though revenues are growing quite slowly [3].

In view of the afore mentioned challenges, the need for a cost effective and programmable network which is robust enough to meet the demand of users is imperative. Thus, the emergence of Software Defined Networks (SDN). SDN has created commendable avenues to overcome age-old problems in networking, while simultaneously enabling the introduction of complex, secure and reliable network policies for next generation networks [4]. As a revolutionary concept, SDN alters existing networks by separating the forwarding functionalities of existing devices, known as data plane from control element, known as control plane [5].

The future of SDN mainly lies in its acceptance and deployment. Technology and its deployment take years before it can be available to end users due to standardisation process and Request for Comments (RFCs). Speculations however remain as to whether same should be expected for SDN or not. According to [1] a proposal for open and programmable network is presented. The need for researchers to run experiment on campus network using an OpenFlow switch is further emphasised in [1]. In line with this, ETHANE, a new network architecture for enterprise was suggested in [6]. For the proposed architecture, in [6], ETHANE switch does not need to learn addresses, support Virtual Local Area Networks (VLANs) or check for source-address spoofing and it has been deployed in a campus environment. The work in [6] was augmented when Google, deployed B4 using OpenFlow switches in their Wide Area Networks (WAN) data centre [7]. Also, with the advent of a Linux foundation collaborative project, OpenDaylight (ODL) [8] platform and VMware NSX virtualisation platform, global acceptance and deployment is envisaged to be no longer far from reach.

In spite of the programmability, flexibility, universal connectivity and decentralised control, which were critical to the success of SDN, these features are at odds with making it more secure. The SDN platform can bring with it several security breaches which include an increased potential for Denial-of-Service (DoS) attacks due to controller centralisation and flow table limitations in network devices [9]. Furthermore, abstraction of flows and underlying hardware resources make it easier for harvesting of intelligence which can be used effortlessly for further exploitation and reprogramming entire network by malicious user [4].

In this paper, the impact of DoS attack on SDN is presented. The simulation has been perfomed using mininet and OpenDaylight controller tools and the simulation result shows that DoS flooding attack on SDN network can degrade network performance by decreasing network throughput and introduce large jitter. This paper is structured as follows: Sect. 2 presents related works on the SDN architecture, vulnerabilities in the SDN architecture and DoS attacks on SDN.
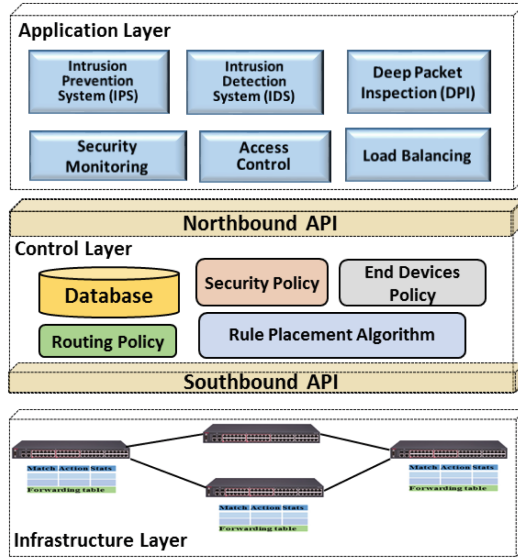
Fig. 1. SDN architecture illustrating the data, control and application layers.

The experimental method and tools are presented in Sect. 3. Then Sect. 4 shows the experimental set-up. The results and analysis are presented in Sect. 5. Finally, the conclusion and future work are presented in Sect. 6.

2 Related Work

2.1 SDN Architecture

SDN architecture encompasses the complete network platform. It is a modular approach that defines chain of command and interoperability within network. Unlike traditional network, the intelligence of data plane devices is removed to a logically centralised control system [10]. Figure 1 presents the SDN architecture showing the data/infrastructure, control and application layer. In an SDN architecture, there are two main elements: the controllers and the forwarding devices. A forwarding device is a hardware or software element specialised in packet forwarding and based on a pipeline of flow tables where each entry of a flow table has: a matching rule, action to be executed on matching packets and counters that keep statistics of matching packets [4]. The controller serves as the brain of the network and it deals with management of network state. Below is a description of various layers:

Infrastructure layer: This layer is also known as data plane. It consists of simple forwarding elements without embedded control or software to take autonomous decisions. It is accessible through the southbound interface and allows packet switching and forwarding.

Control layer: This layer consists of SDN controllers providing a consolidated control functionalities through Application Programming Interfaces (APIs). The crucial value of the controller is to provide abstractions, essential services, and common APIs to developers. Three communication interfaces allows the controller to interact: northbound, southbound and the east/westbound interfaces.

(i) Southbound Interfaces: Southbound interface allows the controller and forwarding elements to interact in the infrastructure layer, thus being the crucial instrument for clearly separating the control and data plane functionality.

(ii) Northbound Interfaces: This interface is the connecting bridge between application layer and control layer. It enables the programmability of the controllers by exposing the data models and other functionalities within the controllers for use by applications at the application layer. The northbound interface is mostly a software ecosystem, hence, a common northbound interface is still an open issue.

(iii) East/Westbound Interfaces: This interface is a special communication interface envisioned for distributed controllers to synchronise state for high availability. Its function include import/export data between controllers and monitoring/notification capabilities to check if a controller is up or notify a takeover on a set of forwarding elements.

Application Layer: The application layer consists of end-user business applications and network services. Example of application that runs here is network virtualisation. Network policy is also defined here.

2.2 Vulnerabilities in SDN Architecture

A number of security analyses has been carried on the vulnerabilities in SDN. Adnan et al. in [5] identified the state of art in SDN security solutions with respect to each layer of SDN architecture. The work focuses on possible security attacks in SDN which could be executed. However, no solution to identified threats is presented. A comprehensive survey of security in SDN is presented in [11,12], the authors identified vulnerabilities introduced by separation of control and data plane. Sandra et al. in [11] presents an overview of SDN security and itemise research work coupled with solution to security issues in SDN. In [12] classification is done using the STRIDE approach and possible SDN security controls is proposed. The concept of offering SDN security as a service is presented in [13].

Kreutz et al. in [2] presents a high level security analysis. Seven main potential threat vectors are presented. Three of the seven identified threat vectors are specific to SDN and relates to the three planes present in SDN architecture. The analysis does not present SDN as a less secure network but triggers the need for innovative ways of responding to the new threats arising from network programmability. The authors state the consequences of these threats in SDN and solutions to the seven threat vectors was proposed.

In [14], a feasibility study on attacking SDN is carried out by fingerprinting to ascertain usage of SDN/OpenFlow switches by the network. The SDN network is then subjected to a specifically crafted flow requests from the data plane to the control plane to exhaust the network resources. Another security vulnerabilities was analysed in ProtoGENI [15]. The authors explored three potential security issues as follows:

(i) **Resource connection:** Once a malicious user obtains access to one experiment node, attacks can easily be launched by utilising the huge ProtoGENI computing resources as a launchpad to harm existing internet users.

(ii) **Wireless Nodes Distribution:** Network sniffing or spoofing can be done here to identify desired node for launching attacks.

(iii) **Virtualization Technology:** In ProtoGENI virtualisation, ProtoGENI resources are shared among as many user as possible. Any bug or compromise from a single device will expose other users in sharing resources to attacks.

The Authors discovered the possibility of using ProtoGENI resources to launch flooding attack to the wider internet. Also, the possibility of compromising confidentiality and availability of other ProtoGENI users is high.

2.3   DoS Attack on SDN

DoS and Distributed DoS (DDoS) attack remains one of the severe network security problems in both traditional network and SDN. Due to separation of control and data plane, an attacker could saturate the controller with malformed packets requiring a flow rule decision. On the other hand, the flow table of the infrastructure device can be overwhelmed with malicious packets. To address bottlenecks of potential saturation attack, AVANT-GUARD [16] introduce connection migration to reduce amount of data-to-control-plane interactions. The method enables the data plane to shield the control plane from saturation attacks. However, the data plane itself is subject to attack. Similarly, a backup strategy which offers resilience against failures in a centralised controlled network is presented in [17]. This approach is an attempt to solve single point of failure bottleneck and it provides seamless transition between primary controller to a back-up controller. However, this solution is limited to centralised implementation and it also raises concern in terms of trust between the east-west interface communications. In addition, Braga et al. [18] proposed lightweight, a new method for detecting DDoS. The proposed method boasts of high rate of true positives and low rate of false alarm using Self Organising Maps (SOM) for flow analysis. The lightweight method consider median values in training the SOM. The drawback of this method is that false negatives are reported when the attack parameter is set to a low value.

The controller has been compared to an operating system capable of managing applications through programmatic interface [19]. Similarly, ETHANE was built to provide network-wide fine-grain policy using a centralised declaration

and enforcing it [6]. While the concept of a centralised controller allow the simplification of policy enforcement and management tasks for network managers, it creates quite a number of bottle necks. In [9], analysis of SDN implementation key challenges has been carried out. The authors opined deployment of SDN technology will contribute to the vision of future communications if outstanding challenges were resolved. In [20] the possibility of DoS attacks and poor rule design that can lead to saturating volumes of controller queries is discussed. Though OpenFlow vulnerabilities in terms of lack of adoption of Transport Layer Security (TLS) for controller-switch communication is highlighted in [20], a number of vulnerabilities proposed was not verified in the work.

3   Experimental Method and Tools

In this experiments, Mininet is used [21]. Mininet is an open source network emulator devoted entirely to OpenFlow architecture and SDN implementation. For the controller, ODL controller is used [22]. ODL integrates open source, open standards and open APIs to deliver SDN platform to make networks more programmable and adaptive. DoS attacks usually engage numerous compromised hosts and a rich topology to launch a successful attack on its victim. While our scenario is much simpler than what is obtainable in real world attacks, we deliberately chose such a low-complexity set-up to expose and analyse the impact of DoS attack on SDN. Common testing tools such as, *ping* and *iperf* are used to generate traffic between host and servers. Figure 2 shows the methodology flowchart with each step explained.
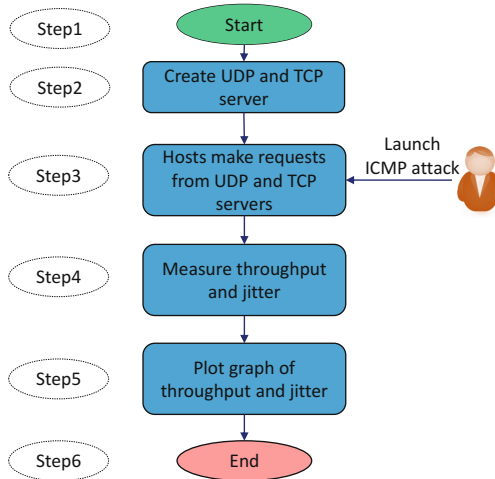
Fig. 2. Methodology flowchart.

Step 1: Start Mininet and ODL controller

```
Sudo mn --custom scenario.py --topo

--controller = remote,ip = x.x.x.x
```

Where x.x.x.x represents the ip address of the remote controller.
Check connectivity using

```
$mininet > net
```

Step 2: Create UDP and TCP server

```
UDP: iperf -s -u -p 5566 -i1
TCP: iperf -s  -p 5566 -i1
```

The TCP server is made to listen on port 5566 with a default window size of 85.3 KB. Similarly, UDP server is made to listen on port 5566 with a default UDP buffer size of 208 KB while receiving 1470 bytes datagrams and the result is monitored every 1 s.

Step 3: Hosts make requests from TCP server and UDP server

```
TCP: -iperf -c x.x.x.x -p5566 -t100
UDP: -iperf -c x.x.x.x -u -t100 -p5566
```

Step 4 and Step 5: Results were extracted using AWK file and results plotted using MATLAB. Then, malicious hosts 5 and 6 launched flooding attack on the servers (similar to step 3). Legitimate traffic is started at the beginning of an experiment, and an attack is launched shortly after for a duration of 100 s.

Step 6: End

```
mininet# ctrl z (end mininet)
Sudo mn -c (clear topology)
```

4   Experimental Set-Up

In this section, a series of experiments are performed to verify the effects of DoS attack in the SDN network. The experimental setup is shown in Fig. 3. To create the scenario in Fig. 3, many software and tools are used as shown in Table 1. There are two servers and four switches in the network. Each switch has a host connected to it. The Transmission Control Protocol (TCP) server is connected to OpenFlow switch1 while User Datagram Protocol (UDP) server is connected to OpenFlow switch3. ICMP flood attack will be launched against both servers by malicious hosts 5 and 6.
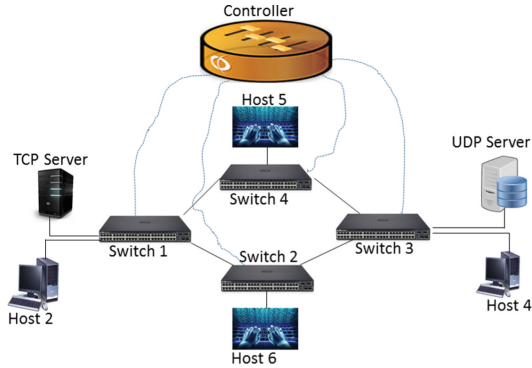
Fig. 3. Experimental setup.

Table 1. Simulation parameters

Simulation	Details
Platform/Enviroment	-Oracle Virtual Box as base environment for simulations
	-Ubuntu OS as base for Mininet v2.2 emulator
	-Ubuntu Server as base for OpenDaylight (Boron) controller
	-Host CPU as intel core i7, 12 G RAM
Attack tool	Hping tool - Hping3 used for flooding attack

5 Results and Analysis

As discussed in the experimental setup, we simulate for two different scenarios; TCP and UDP requests under normal operating condition and under attack. The results for these scenarios are discussed below.

5.1 Effect of DoS Attack on Throughput

Figures 4 and 5 shows a significant drop in throughput due to malicious behaviour (ICMP flood attack) being executed by two attacking nodes. The average throughput for requests made from host 4 to the TCP server is 214 Mbits/s for a total of 2.5 GB of information transferred in 100 s. Similarly, the average throughput of host2 requests from TCP server is 233 Mbits/s for a total of 2.72 GB of information transferred.

Notice that Host 2 shows a better bandwidth utilisation than Host 4 and the reason for this is not far-fetched; they are both connected to OpenFlow switch 1. While the better bandwidth utilisation is seen as an advantage here, it is a major security risk and attractive honeypot to launch attack against the server. The impact of this connection is felt when the server is subjected to ICMP flood attack. During attack, the average throughput dropped to 106 Mbits/s from 214 Mbits/s
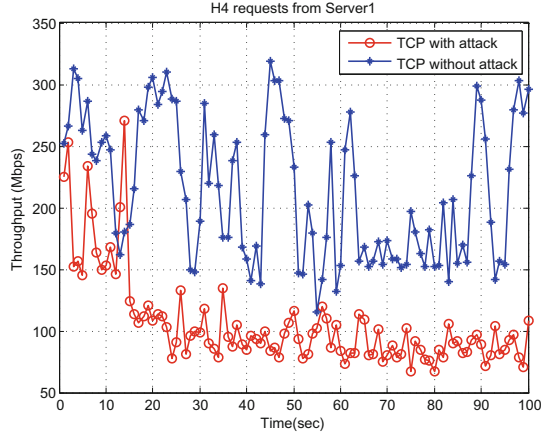
Fig. 4. TCP requests from host 4 to TCP server under ICMP attack.
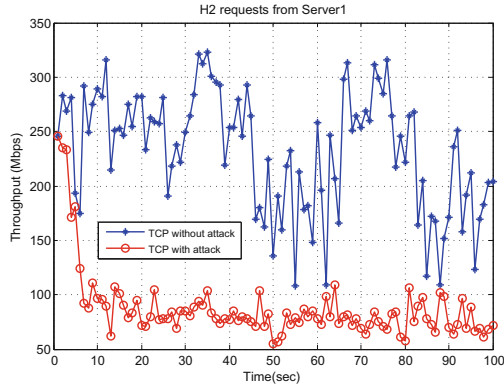
Fig. 5. TCP requests from host 2 to TCP server under ICMP attack.

recorded for H4 requests from server1 without ICMP flood attack. The impact of the attack launched by host 5 and 6 became noticeable after 15 s of transmission and the bandwidth utilisation degraded for the rest of the transmission. The trend is similar for host2 requests from TCP server as degradation started after 8 s of transmission and degraded for the rest of the transmission. The average throughput for h2 requests from TCP server dropped from 233 Mbits/s to 87.4 Mbits/s when the server is under attack. The degradation is more severe for host 2 when under attack even though higher throughput is recorded during normal operation. Hence, the need for better network design, traffic isolation based on priority for mission-critical network and dynamic proactive ways of addressing DoS attacks when the system is under serious attack.

5.2   Effect of DoS Attack on Jitter

Jitter is defined as a variation in the delay of received packets. In Figs. 6 and 7, using UDP buffer size of 208 KB, the jitter varies between 0.003 ms and 0.789 ms. Host 4 Jitter remains within a fair range because it is connected to OpenFlow switch 3 with the UDP server. The spiky delay waveform indicates the presence of congestion in the network. Even though the congestion occurs for a very short period, if the congestion time is more than the scheduled packet transmission time, it will lead to packet drops. Notice that jitter values obtained from host 4 requests to UDP server is better compared to requests from host 2.
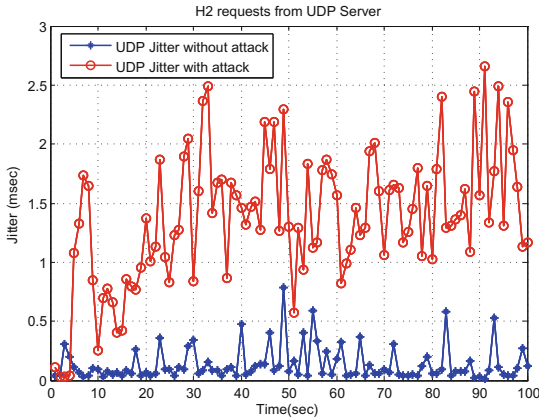
Fig. 6. UDP requests from host 2 to UDP server under ICMP attack.
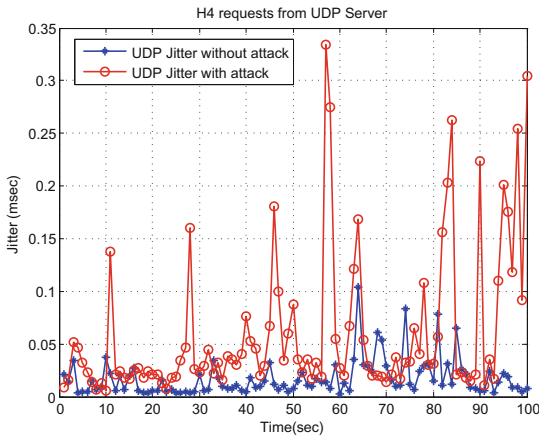
Fig. 7. UDP requests from host 2 to UDP server under ICMP attack.

6   Conclusion and Future Work

In this paper, the impact of DoS attack on SDN has been demonstrated. This study reveals that for a simple network, a DoS attack on the infrastructure plane (UDP and TCP servers) will highly degrade network performance as shown in the performance metrics (throughput and jitter). For a Distributed DoS (DDoS) attack with more active agents, the attack will be more severe. Hence, the need for a robust resilient SDN security architecture. While the evaluation of the impact of DoS attack on SDNs remains a very rigorous endeavour, the work carried out in this paper offers a primer to the objective evaluation of DoS attack on SDNs. The simulation results revealed a drop in network throughput from 233 Mbps to 87.4 Mbps and the introduction of large jitter between 0.003 ms and 0.789 ms during DoS attack. In the future, the mitigation of DoS and DDoS attacks in an exhaustive way at both control and data plane layers will be examined.

References

1. McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L.: Open-Flow: enabling innovation in campus networks. ACM SIGCOMM Comput. Commun. Rev. **38**(2), 69–74 (2008)
2. Kreutz, D., Ramos, F., Verissimo, P.: Towards secure and dependable software-defined networks. In: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in software Defined Networking, pp. 55–60. ACM (2013)
3. Das, S., Parulkar, G., McKeown, N.: Rethinking IP core networks. J. Opt. Commun. Netw. **5**(12), 1431–1442 (2013)
4. Kreutz, D., Ramos, F.M., Verissimo, P.E., Rothenberg, C.E., Azodolmolky, S., Uhlig, S.: Software-defined networking: a comprehensive survey. Proc. IEEE **103**(1), 14–76 (2015)
5. Akhunzada, A., Ahmed, E., Gani, A., Khan, M.K., Imran, M., Guizani, S.: Securing software defined networks: taxonomy, requirements, and open issues. IEEE Commun. Mag. **53**(4), 36–44 (2015)
6. Casado, M., Freedman, M.J., Pettit, J., Luo, J., McKeown, N., Shenker, S.: Ethane: taking control of the enterprise. ACM SIGCOMM Comput. Commun. Rev. **37**(4), 1–12 (2007)
7. Jain, S., Kumar, A., Mandal, S., Ong, J., et al.: B4: experience with a globally-deployed software defined WAN. ACM SIGCOMM Comput. Commun. Rev. **43**(4), 3–14 (2013)
8. VMware: Software-Defined Data Center (SDDC) (2017). http://www.vmware.com/products/nsx/
9. Sezer, S., Scott-Hayward, S., Chouhan, P.K., Fraser, B., Lake, D., Finnegan, J.: Are we ready for SDN? implementation challenges for software-defined networks. IEEE Commun. Mag. **51**(7), 36–43 (2013)
10. Goransson, P., Black, C., Culver, T.: Software Defined Networks: A Comprehensive Approach. Morgan Kaufmann, Burlington (2016)
11. Scott-Hayward, S., Natarajan, S., Sezer, S.: A survey of security in software defined networks. IEEE Commun. Surv. Tutor. **18**(1), 623–654 (2016)

12. Alsmadi, I., Xu, D.: Security of software defined networks: a survey. Comput. Secur. **53**, 79–108 (2015)
13. Ali, S.T., Sivaraman, V., Radford, A., Jha, S.: A survey of securing networks using software defined networking. IEEE Trans. Reliab. **64**(3), 1086–1097 (2015)
14. Shin, S., Gu, G.: Attacking software-defined networks: a first feasibility study. In: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, pp. 165–166. ACM (2013)
15. Li, D., Hong, X., Bowman, J.: Evaluation of security vulnerabilities by using Proto-GENI as a launchpad. In: Global Telecommunications Conference (GLOBECOM 2011), pp. 1–6. IEEE (2011)
16. Shin, S., Yegneswaran, V., Porras, P., Gu, G.: Avant-guard: scalable and vigilant switch flow management in software-defined networks. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, pp. 413–424. ACM (2013)
17. Fonseca, P., Bennesby, R., Mota, E., Passito, A.: A replication component for resilient openflow-based networking. In: Network Operations and Management Symposium (NOMS), pp. 933–939 (2013)
18. Braga, R., Mota, E., Passito, A.: Lightweight DDoS flooding attack detection using NOX/OpenFlow. In: IEEE 35th Conference on Local Computer Networks (LCN), pp. 408–415 (2010)
19. Gude, N., Koponen, T., Pettit, J., Pfaff, B., Casado, M., McKeown, N., Shenker, S.: NOX: towards an operating system for networks. ACM SIGCOMM Comput. Commun. Rev. **38**(3), 105–110 (2008)
20. Benton, K., Camp, L.J., Small, C.: OpenFlow vulnerability assessment. In: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, pp. 151–152. ACM (2012)
21. TeamMininet: Mininet (2017) http://www.mininet.org/download/
22. Linux-Foundation-Collaborative-Projects: ODL (2017) https://www.opendaylight.org

Analysis of the Suitability of Satellite Communication for Time-Critical IoT Applications in Smart Grid and Medical Grade Networks

Frank Ball[1] and Kashinath Basu[2(✉)] [ID]

[1] Frank Ball Consulting, Oxford, UK
frank_ball@ntlworld.com
[2] School of Technology, Oxford Brookes University, Oxford, UK
kbasu@brookes.ac.uk

Abstract. Satellite networks are seen as having the potential to play an important role in machine-to-machine (M2M) communications and the Internet of Things (IoT), which in turn is seen as being important to a number of service sectors. However, certain M2M applications have bounded latency requirements that in some cases may be quite stringent. Satellite networks generally have much higher latency than wired networks and therefore may not be able to meet the requirements of all M2M applications. This paper compares the latency requirements of certain time-critical applications with reported satellite network latency and address the problem of latency evaluation of networks to support these types of applications.

Keywords: Satellite communications · IoT · Time critical application
Monitoring and control · Smart grid · Medical grade network

1 Introduction

The global connective offered by the Internet and the number of heterogeneous, and in many cases smart, devices that are able to connect to it has led to the concept of the Internet of Things (IoT). Some see this concept as being a direction for the evolution of the current Internet, while others view the IoT as separate entity [1, 2]. The important thing from the perspective of this paper, however, is that machine-to-machine (M2M) communication is quite different to that where human interaction is involved. In particular the traffic characteristic and QoS requirements of M2M applications are somewhat different from those of current Internet applications. For certain monitoring and control systems, strict latency requirements will need to be assured [2, 3], and in some cases this assurance may need to be provided over large distances. However, given that propagation delay is a function of distance, then obviously, this sets a limit to what can be achieved, even with minimal switching delays and high bandwidth links.

Global (or very wide area) communication generally involves a combination of wired, wireless, and in many cases satellite networks. Satellite networks are seen as having the potential to play an important role in the IoT and M2M communication

particularly in cases of remotely placed smart objects and group-based communications [1]. The IoT is seen as being important to a number of service sectors, including: Buildings; Energy generation and distribution; Healthcare, e.g. remote patient monitoring and telesurgery; and transport etc.

The control of power grids is one area in which satellite communications already plays a role, and it is expected that they will also make a contribution to the smart grids of the future. However, certain smart grid applications that have been proposed for use in the near future are time-critical and have quite stringent latency requirements [3, 4]. In these cases it seems unlikely that satellite communications could meet their particular requirements. Satellite communications may also be incorporated into Medical Grade networks (MGN) [5]. MGN also need to support time-critical applications, e.g. remote monitoring of a patient's vital signs, although for these applications, a much higher degree of latency is allowable. However, even though the latency requirements are less stringent, they must still be met for the application to function correctly. Also, in addition to bounded delay, both smart grid and MGN control and monitoring applications may also require a high level of reliability. There may also be many other examples of time-critical application for which satellite communication may be considered to be a suitable carrier, however, for economy of space, this paper will only focus on Smart grid control systems and MGN monitoring and diagnostic services.

To provide the necessary guarantees to time-critical applications, performance evaluation must be carried out *a priori* to deployment of the communications system. Therefore, in the case of supporting stringent real-time applications, performance evaluation will need to become an integral part of the system design process. Whilst performance evaluations of existing systems, based on either simulation or empirical testing, can provide useful information, they can only show whether or not a particular system may be able to meet its targets.

The aims of this paper are to address the potential for satellite communication to support time-critical monitoring, and control applications, and to consider how to evaluate the latency of a satellite link for individual cases.

The remainder of the paper is structured as follows: Sect. 2, outlines the characteristics and requirements of time-critical applications; Sect. 3, discuses smart grid applications and their relationship to satellite communications; Sect. 4 presents MGNs and discusses the options for satellite interactions; Sect. 5 presents ongoing work and discusses future directions; and finally Sect. 6 offers some concluding remarks.

2   Time-Critical Applications

Time-critical applications are a class of real-time application in where missed deadlines may result in serious consequences. The degree of seriousness will of course vary between applications, which in extreme cases could include, for example: significant loss of revenue: damage to equipment; and possibly personal injuries [3]. Less serious consequences could include a significant reduction in the efficiency of the control application. The degree of time criticality may also vary between applications depending on the flexibility of the control mechanism. Therefore, the type of real-time guarantee

required will also vary with application. In general, there are three basis classes of real-time guarantee: Hard Real Time, in which information arriving after the deadline T is no longer of any use (i.e. equivalent to packet loss); Firm Real Time, that is similar to Hard Real Time, except that some losses may be tolerable; and Soft Real Time, in which information arriving before a deadline T_1 is fully useful, but will still have a degree of usefulness until a second deadline T_2 has been exceeded.

Proving a Hard-Real-Time service can be quite difficult in practice and the definition of Firm-Real-Time is not particularly explicit. Using a probabilistic interpretation as shown below in Eqs. 1 and 2 offers a more pragmatic definition that can be matched to the tolerances of the control applications.

$$\text{Probabilistic Hard Real Time: } P(t > T) <= 10^{-x} \tag{1}$$

$$\text{Soft Real Time: } P\big(t > T_1\big) <= 10^{-x} \text{ AND } P\big(t > T_2\big) <= 10^{-x}$$
$$\text{with } f_u(t) = \big[1\ t <= T_1; g(t)\ T_1 < t <= T_2; 0 \text{ otherwise}\big] \tag{2}$$

The setting of these parameters would need to be determined by the requirements of the application and the flexibility of the control algorithm. Certain smart grid and MGN control applications are examples of a Networked Control System (NCS) [5], i.e. a closed loop control system that operates over an open network that is shared with other classes of traffic. Designers of NCSs face two challenges: maintaining the appropriate QoS in the networks; and ensuring that the required Quality of Control (QoC) is provided. Therefore research into NCSs focuses on two objectives: Control of the network, to ensure a suitable QoS; and control over the network, that seeks to minimize adverse conditions in the network. Therefore, NCSs methodologies may prove useful for setting the real-time parameter.

The traffic patterns of a monitoring and control application may not only be different from that of other types of applications, but can also vary between different instances of monitoring activity. Communication between a monitoring device and the controller may take the form of: one-to-one, i.e. a single stream; many-to-one, i.e. a number streams fanning in to the controller; and many-to-one with synchronized sources. For the latter, it is possible that the controller may need to wait until a full set of synchronized messages arrive before processing them. In this case the delay bound will apply to the arrival time of the last message from the set. For the other two cases, however, the delay bound will apply to each individual message.

In general, average delays based on long periods of time do not provide a useful QoS metric for time-critical control applications. High percentiles of delay taken over relatively short periods are more useful for identifying adverse transitory conditions.

3   Smart Grid Applications

A communication network is an essential component of a smart grid system. Its role is to support a wide range of applications, many of which have very similar requirements to those of current Internet applications and emerging IoT applications. In particular, they have the general requirements for security, resilience, reliability and wide area

interconnectivity. However, a number of classes of smart grid applications, particularly those intended for controlling smart grids in the near future, have requirements that are significantly different from those of any existing Internet application. Time-critical smart grid applications are responsible for state estimation, control, protection, and ensuring the stability of power generation and distribution. Their domains of operation include both the local area for internal sub-station control, and the wide area for protection, control, and maintaining wide area awareness. Currently, when operating in the wide area, the role of these applications is generally limited to providing visualization for wide area awareness. Applications that provide visualization have near-real-time requirement and can tolerate latencies in the order of 100 ms. Currently, automatic control, for which the more stringent delay requirements apply, is mainly limited to the local area [4].

De Sanctis et al. [1] show that satellite communication based on low earth orbit (LEO) constellations can meet most of current power grid supervisory control and data acquisition (SCADA) requirements. But may not be able to meet the needs of emerging applications. In particular wide area situation awareness (WASA) has latency requirements of <200 ms, which is shown by Yang et al. [2] to be just within the limits of a LEO constellation. It should be noted that the results of both [1, 2] are based on the use of standard IP protocols.

WASA is based on a wide area measurement system (WAMS) that collects synchronised phase measurements of the power cycle, known as Synchrophasors from devices called Phase Measurement Units (PMUs) that are deployed throughout the grid. The result traffic pattern is that of many-to-one with synchronised sources [4]. Currently, in the wide area these measurements are used to provide visualisation for wide area awareness, and control decisions, and activations, are carried out manually. Automatic control is generally limited to internal control within sub-stations. However, in order to meet the smart grid objectives for a greater use of renewable energy and a more efficient of existing non-renewable source the aim is to extend automatic control to the wide area. The ultimate requirement for synchrophasor based wide area control applications is to carry out the measurement-to-decision process within one power cycle [3, 7] which in the case of a 60 Hz power cycle is 16.7 ms. This is an application level target and it has been proposed that the communication latency should be in the order of 2 ms [3].

With such a stringent delay requirement it will not be possible for satellite communications to support this type of application. However, even in the case of a wired network that has high speed links and minimal switching delays, the size of a control domain will still be limited. For example, in the best case a circular domain with a centralized controller will be limited to a radius of less than 400k, based on a propagation delay of 5 us/km. However, given a grid that comprises several control domains, inter-domain control will also be needed, thereby resulting in a hierarchical control system. If this does become the case, and the higher level control systems can operate with less stringent delay bounds, it is possible that satellite communication may be able to support the higher level control communications.

4 MGN Applications

Medical Grade Networks (MGNs) are required to deliver various e-Health services to end users on an anywhere, anytime basis [8, 9]. These services include, for example, health information, electronic health recording, telemedicine and patient monitoring. In common with other communication networks MGNs need to provide reliability, resilience, various levels of security, and to support mobility for a number of applications. In addition, certain classes of these services, e.g. ECG monitoring, IoT based body sensors are delay intolerant and have firm or hard real-time communication requirements. Therefore, a MGNs needs to be an advanced multimedia network that provides the appropriate prioritisation to time critical sources.

Due to the requirements for mobility, local area wireless networks WLANs play an important part for patient monitoring. Generally, they provide a communications path between the monitoring devices and a wider area wired network. Numerous performance studies of IEEE 802.11 WLANs have addressed the problems of prioritization for time-critical applications at MAC and LLC levels, example of which include [9, 10]. These studies emphasise the need for a thorough understanding of the performance issues in both the Link and MAC layers. In particular, for IEEE 802.11 WLANs, both data rates and PER can vary with channel conditions. This relationship offers a more challenging problem to performance evaluation than that experienced in the case of wired networks. Given the influence that channel condition can have on both data rates and packet error rates (PER), performance evaluations based on long term aggregation will most likely produce misleading results. Therefore, short term transitory conditions must also be considered.

Given that satellite communication performance can also be effected by channel conditions the comments made above relating to WLANs will also apply to the satellite link. If the satellite link is used to provide connection to a remote WLAN based monitoring system, then it could be the case that combined WLAN and satellite latency may exceed the delay target. However, with monitoring devices that are able to communicate directly with the satellite system the delays due to the WLAN can be eliminated.

5 Work in Progress and Future Directions

The delay requirements of time critical applications need to be addressed on an end-to-end basis, with the delays resulting from protocol stack operations also being taken into account. However, for practical reasons, some degree of decomposition in the evaluation process is usual needed.

In previous, work we have addressed the problem of post deployment evaluation of a power grid synchrophasor measurement system operating over a wide area wired network, and have carried out the first stage of developing of an evaluation approach based on generic parameterized models [4]. It was shown, that in the case of synchrophasor measurement and control systems, decomposition between the network and the application devices is quite straight forward due to their mode of operation. However,

these initial models were based on high speed links, line rate strict priority switching and relatively deterministic real-time devices.

Currently we are extending these models to accommodate a higher degree of variability in both device output and switching mechanism performance. We have been experimenting with a convolution based approach for evaluating high percentiles of end-to-end delays in cases where decomposition has been applied. Initial results from this work are positive, however, they do show that even a low degree of variability leads to a much higher level of complexity in the final evaluation process.

The focus of our approach has been now widened to include the evaluation process for MGNs and to take into account the effects of wireless LAN communication. This line of the investigation is still in the early stages, however, it has indicated that the Link and MAC layers are appropriate boundaries for decomposition and that there may be a significant advantage in the co-evaluation of the Link and MAC layers due a level of interdependence as identified in [9, 10].

The perceived importance of satellite communications for supporting, the IoT in general, and power grids and MGNs in particular has encouraged us to open our investigation to include satellite communication. Furthermore, as far as the evaluation process is concerned, we can see some commonality between the MGN Link/MAC evaluation and the evaluation of a satellite system. Although we do expect the latter to involve more stages of evaluation.

6   Conclusion

This paper has addressed the potential for satellite communications to support M2M time-critical monitoring and control applications, with particular focus on smart grid and MGN applications including emerging IoT based applications. It has noted that certain control applications proposed for the smart grids of the future have delay requirements that are too stringent to be met by satellite links. However, it has also noted that it may still be able to play a role in the smart grids of the future. It has also outlined the traffic characteristics and requirements of time-critical applications and discussed the need for post deployment evaluation. The paper has also outlined work in progress on the development of a post deployment evaluation technique based on generic and parameterized models. Finally, it has presented an outline plan for the progress of this work.

References

1. De Santus, M., Cianca, E., Araniti, G., Biso, I., Prasad, R.: Satellite Communications Supporting internet of remote things. IEEE Internet of Things J. **3**(1), 113–124 (2016)
2. Yang, Q., Laurenson, D.I., Barria, J.A.: On the use of LEO satellite constellation for active management in power distribution networks. IEEE Trans. Smart Grid **3**, 1371–1381 (2012)
3. Bakken, D.E., Bose, A., Hauser, C.H., Whitehead, D.E., Zweigle, G.C.: Smart generation and transmission with coherent, real-time data. Proc. IEEE **99**(6), 928–951 (2011)
4. Ball, F., Basu, K.: Performance evaluation of time-critical smart grid applications. In: The Proceeding of the Eleventh International Network Conference (INC 2016), Frankfurt, Germany, 13–18 July 2016

5. CISCO Systems White Paper: Cisco Medical-Grade Network: Providing Foundational Architectures for Healthcare. https://www.cisco.com/c/dam/global/en_ca/solutions/strategy/healthcare/assets/docs/09CS2124-MGN.pdf. Accessed 07 April 2017
6. Gupta, R.A., Chow, M.-Y.: Networked control system: overview and research trends. IEEE Trans. Ind. Electron. **57**(7), 2527–2535 (2010)
7. Budka, K.C., Deshpande, J.G., Doumi, T.L., Maddan, M., Mew, T.: Communication network architecture and design principles for smart grids. Bell Labs Tech. J. **15**(2), 205–228 (2010)
8. Skorin-Kapov, L., Matijasevic, M.: Analysis of QoS requirements for e-Health services and mapping to evolved systems QoS classes. Int. J. Telemed. Appl. **2010**, 1–18 (2010)
9. Lin, C.-F.: An advanced wireless multimedia communication application: mobile telemedicine. WESEAS Trans. Commun. **9**(3), 206–215 (2010)
10. Kang, K., Park, K.-J., Song, J.-J., Yoon, C.-H., Sha, L.: A medical-grade wireless architecture for remote electrocardiography. IEEE Trans. Inf. Technol. Biomed. **2**, 260–267 (2011)

Advanced PHY-MAC Schemes
for Future Satellite Systems

Making H-ARQ Suitable for a Mobile TCP Receiver over LEO Satellite Constellations

Bastien Tauran[1]([✉]) [ID], Emmanuel Lochin[1], Jérôme Lacan[1], Fabrice Arnal[2], Mathieu Gineste[2], Laurence Clarac[3], and Nicolas Kuhn[3]

[1] Université de Toulouse, ISAE-SUPAERO TéSA, Toulouse, France
{bastien.tauran,emmanuel.lochin,jerome.lacan}@isae-supaero.fr
[2] Thales Alenia Space, Toulouse, France
{fabrice.arnal,mathieu.gineste}@thalesaleniaspace.com
[3] Centre National d'Etudes Spatiales, Toulouse, France
{laurence.clarac,nicolas.kuhn}@cnes.fr

Abstract. This paper investigates strategies to carry out delay tolerant services over LEO satellite constellations for mobile receiver. In this context, LEO constellations are characterized by important delay variations where propagation impairments are mostly localized on the Land Mobile Satellite (LMS) channel (i.e. on the last hop). To cope with this issue, distinct reliability schemes can be introduced at the physical or link layers. Although their capacity to cope with transmission errors has been demonstrated, these recovery schemes may induce a high jitter that could severely damage TCP's internal timers and reliability schemes. As a matter of fact, transport and link layers' reliability schemes exhibit a clear discrepancy. Following temporal traces representing the delay between a mobile terminal and the last hop satellite from a LEO constellation, we assess how HARQ mechanisms impact on the RTO based retransmission and the duplicate acknowledgments of TCP. Based on ns-2 simulations, we propose a layer-2 buffer that let both link and transport layers to conjointly perform. Our evaluations show an end-to-end data rate increase and more generally illustrate the benefit of re-ordering packets at the link layer when link-layer erasure coding recovery mechanisms are used conjointly with TCP.

Keywords: Constellations · TCP · Latency · Adaptive-HARQ

1 Introduction

Constellations systems such as Iridium connect urban and rural areas to terrestrial Internet broadband. As a result, significant amounts of TCP traffic between end-hosts on the Internet is expected to be carried out by constellations of satellites [1]. However, LEO constellations are characterized by important delay variations [2–4], where propagation impairments are mostly localized on the Land

Mobile Satellite (LMS) channel [5]. In particular, as link errors often occur on the last mile and strongly impact on the end-to-end transmission, link and physical layers error correcting codes are deployed to minimize these errors [6–8]. However, reliable transport protocols remain necessary to ensure end-to-end reliability but the inadequacy of the interactions between both transport and link layers retransmission schemes may result in high end-to-end latency. As opposed to GEO satellite communications, LEO constellations are expected to provide a lower propagation delay. That being said, this may not ensure a lower end-to-end delay and a better quality of experience for latency sensitive applications if, *e.g.*, there is no cross-layer considerations. With LEO constellations, the TCP connection may not require specific tuning and/or to be split using Performance Enhancing Proxies (PEP) as defined in RFC3135.

Considering high delay and its variability in LEO satellite constellations, space and aeronautical communications often lays on delay tolerant services. Such services rely to applications which are bounded by a delivery delay for a certain percentile of messages. For instance, reliable sensor data transfer or aeronautical message services might require that 95% of data messages respect a certain delay threshold [9]. To ensure this kind of service, and to cope with losses that might occur on LMS, several ARQ and Hybrid ARQ schemes have been proposed [8,10]. Basically, these solutions make more robust this last link by adding error correcting, erasure coding, or enhanced Automatic Repeat reQuest (ARQ) schemes. One novel and promising solution is Adaptive-HARQ, which is an evolution of type II HARQ [11]. Once again, although such link layer mechanisms greatly mitigate the percentage of packets loss seen by the transport layer, they do not ensure a fully reliable service.

We acknowledge that multiple studies have analyzed the relation between ARQ and reordering. These investigations have even led to standard documents (RFC3366, 3GPP standard for RLC). These contributions show the interest of introducing a reordering buffer, when layer 2 reliability mechanisms are deployed. However, considering the actual trend in **deploying constellations of LEO satellites, our study fills a gap by assessing the relevance of such solution in this specific context**.

In this paper, we study the performance of bulk data transfers carried out by either TCP NewReno (denoted NR in the results) and CUBIC over a LEO constellation. The objective is to assess the impact of error mitigation mechanisms deployed at the link layer on the transport layer, which may trigger spurious retransmissions when high jitter variation occurs. We first study a set of TCP metrics impacted by this variable delay (number of spurious retransmissions, DUPACK and timeout). Following these preliminary results, we identify the root cause of the problem and propose a re-ordering buffer scheme that greatly improves TCP data delivery ratio. We believe that this solution should be deployed in any case where H-ARQ mechanisms are conjointly working with TCP.

2   Scenario

This section presents how we simulate the satellite environment and the different schemes that are considered through out this paper.

2.1   Satellite Environment

To simulate the satellite environment, we used both Network Simulator 2 (*ns*-2) and `SaVi` [12] to simulate a constellation composed of 66 satellites on Low Earth Orbit, at an altitude of 800 km. This constellation ensures a global coverage of any point of the earth, at any time.

   The transmission delay varies within the satellite constellation because of the satellites' movements and the route changes. Considering the characteristics of our constellation, the time from the gateway to the end user's terminal varies between 70 ms and 90 ms. We consider that apart from the forward LMS links, there are no transmission errors. This assumption eases the impact of reliability schemes analysis on the forward transmission (*i.e.* from the gateway to the terminal) and remains consistent as Inter-Satellite Links (ISL) do not usually exhibit transmission errors. We consider that the routing within the constellation prevents congestion drops inside the constellation. The topology simulated is described in Fig. 1.
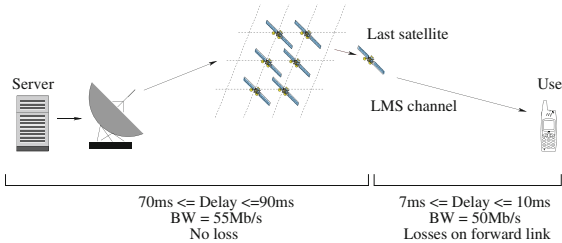
Fig. 1. Model for a satellite constellation   **Fig. 2.** Layering
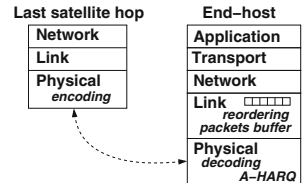
2.2   Hybrid Automatic Repeat ReQuest

The Hybrid Automatic Repeat reQuest (HARQ) mechanism is a physical layer solution which aims at increase both physical and link layers reliability. In particular, this paper considers a novel HARQ scheme, Adaptive-HARQ [11] that has been designed to optimize the bandwidth utilization, as opposed to other HARQ mechanisms. We implemented this scheme as a new module for *ns*-2. This module, working on physical layer, allows the source node to send redundancy bits to the receiver, up to a fixed number of times, until the message is decoded. If the message is not decoded after the maximal number of (re)transmissions, it is then considered as lost and is dropped by the A-HARQ module. Following [11], we authorize up to 3 retransmissions. Figure 3 details how this module is working.
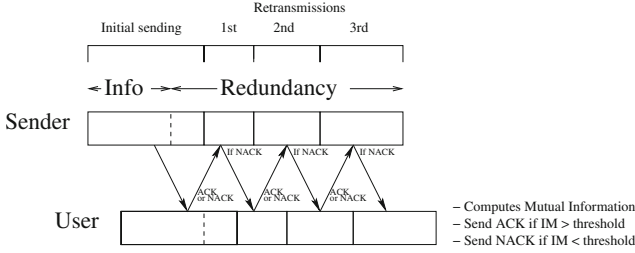
Fig. 3. Description of HARQ

The receiver side of the A-HARQ link uses a buffer to store packets not yet decoded. Each time useful or redundancy bits are received, the module computes the Mutual Information (MI) of the packet, depending on the state of the channel at this moment and the total number of bits already sent. The packet is considered as decoded if the value of the Mutual Information is higher than a given threshold.

As A-HARQ involves both the link and physical layers, we simulate with ns-2 the physical layer using mutual information (MI) and the channel characteristics following data contained in [11]. On the worst case, the packets are recovered at most 70 ms after entering A-HARQ module. This worst case corresponds to the time between a packet is sent from the last hop satellite and the packet is decoded after three A-HARQ retransmissions.

2.3   Version of TCP and Parameters

TCP has been showed to provide reasonable performances over LEO constellation [1,13], without requiring specific PEP optimization mechanisms, we thus consider TCP NewReno (NR) and CUBIC as TCP variants for our evaluations. Even though TCP NewReno is less and less populating the Internet, this protocol is of interest for our problematic since it features basic reliability functionalities of most TCP variants. We also consider CUBIC since (1) its error recovery is more aggressive than TCP NewReno and (2) it is enabled by default in GNU/Linux and OSX systems (since `10.9`).

We test TCP/SACK performance with standard GNU/Linux parameters. The first one is Duplicate SACK (DSACK), which is enabled by default on GNU/Linux systems. DSACK allows TCP sender to differentiate in some cases, if a Fast Retransmit is due to a reordering or to a packet loss. Thus, the TCP sender can adjust the duplicate acknowledgment threshold (defined in RFC2581), which is usually 3 by default in GNU/Linux.

We also evaluate the impact of Delayed Acknowledgment (DelAck) on our proposed solution. Delayed ACK (defined in RFC1123) can combine two in-order packets if they are received within a fixed time window, which is 40 ms in GNU/Linux systems.

2.4   Simulation Scenario

We designed a scenario with three nodes representing the sender, the last satellite on the message route, and the receiver. We simulate the constellation by changing the link delays, using a temporal trace (obtained with `SaVi`) giving at any moment the value of the delay between the sender and the last satellite, and the last satellite and the receiver. These delays have been measured on a initial *ns*-2 simulation using the LEO constellation described in Sect. 2.1. For all simulations, we use these delays traces obtained to mimic the constellation. Each simulation lasts 600 s with one single TCP performing. The simulations are run with a Land Mobile Satellite (LMS) channel [14], [15] between the last satellite and the ground gateway and an ITS (Intermediate Tree Shadowed Environment) environment. The LMS channel enables the Adaptive-HARQ module. We also vary the quality of this channel by setting an average SNR ranging from 7 dB to 13 dB. During the simulations, the link quality changes over time around this SNR average value according to a trace file representing the evolution of the LMS channel. For all other links (return link included), we assume that there are no errors to ease the interpretation of the results.

3   On the Need to Mitigate the Impact of Out-of-Order Packets

In this section and the next ones, the numbers of RTO, DUPACK and spurious retransmission have been normalized by the number of packets sent in order to have comparable results with the different graphs and tables.

Table 1 shows the goodput achieved by TCP without reordering mechanism. The RTO and DUPACK columns represent the proportion of recovery events (due to link errors or congestion events). We divided the number of times TCP received a request for retransmission by the total number of packets sent. For example, when $SNR = 7$ dB, during 600 s, CUBIC transmitted 21406 packets, the RTO timer expired 54 times and 500 retransmissions have been triggered due to DUPACK. We have also measured 480 spurious transmissions. Thus, the proportion for RTO is 0.25%, for duplicate acknowledgments is 2.34% and for spurious is 2.24%.

We observe that we have a low goodput, whatever the value of SNR, whereas we could expect a goodput of 40 Mb/s considering no errors occurred in the network. On the other hand, the number of DUPACK and spurious retransmission remains high, and is not decreasing when the channel quality improves as we could expect. This is mainly due to out-of-order packets interpreted by TCP as congestion losses which trigger spurious retransmissions and halve the congestion window.

To let TCP exploit the available capacity, we need mechanisms that mitigate the effects of out-of-order packets. In the next section, we exploit a possibility that is adding a reordering mechanism after A-HARQ.

Table 1. No reordering mechanism

SNR (dB)	TCP Goodput (kb/s)		A-HARQ Success (%)		RTO (%)		DUPACK (%)		Spurious (%)	
	NR	CUBIC	NR	CUBIC	NR	CUBIC	NR	CUBIC	NR	CUBIC
7	188	265	95.30	95.11	0.62	0.25	2.84	2.34	1.81	2.24
8	242	349	97.22	97.08	0.23	0.05	2.85	2.19	1.86	2.00
9	261	385	97.80	97.72	0.16	0.01	2.77	1.99	1.63	1.82
10	290	431	98.32	98.37	0.08	0	2.72	1.82	1.69	1.75
11	312	469	98.84	98.85	0.04	0	2.59	1.63	1.46	1.71
12	328	494	99.14	99.21	0.02	0	2.56	1.54	1.48	1.73
13	346	511	99.38	99.45	0.01	0	2.44	1.52	1.59	1.75

4   Solution Proposed

The inconvenient of A-HARQ is that packets are delivered out-of-order to the transport layer, due to the changing number of retransmissions by this mechanism, and also by route changes. This strongly impacts on TCP performance and generates DUPACK causing spurious retransmissions. In the next section, we show that adding a reordering mechanism, after A-HARQ and before sending the packets to the transport layer, improves the performance of TCP.

4.1   Adding a Reordering Mechanism

This reordering mechanism is composed of a buffer storing TCP out-of-order packets. A decoded packet from A-HARQ is directly forwarded to the upper layer if in sequence otherwise stored. This mechanism has to deal with the case where the buffer limit is reached and if a packet has not been decoded by A-HARQ. The rationale for the later is to speed up TCP recovery procedure. For both cases:

- we limit the size of the reordering buffer to 125 packets. This value is computed to prevent RTO and results from an heuristic based on throughput and retransmission delay on the last hop. When full, all these out-of-order packets are forwarded to the upper layer. Note this case rarely occurs in our simulation scenario as the traffic load is low (i.e. we only send one long-lived flow) and the route changes in the satellite constellation considered does not lead to several out-of-order packets;
- concerning the case when H-ARQ decoding fails, we propose to flush the buffer even if not full. Thus, all packets are forwarded to the upper layer allowing TCP to quickly recover this missing packet with DUPACK or RTO.

A global view of the last hop is given in Fig. 2, with the A-HARQ module and the reordering buffer.

4.2   Results with Reordering Mechanism

Table 2 shows the results obtained when the reordering mechanism operates conjointly with A-HARQ. We observe an improvement in terms of goodput due to

Table 2. With reordering mechanism

SNR (dB)	TCP Goodput (kb/s)		A-HARQ Success (%)		RTO (%)		DUPACK (%)		Spurious (%)	
	NR	CUBIC	NR	CUBIC	NR	CUBIC	NR	CUBIC	NR	CUBIC
7	409	363	96.06	94.95	0.62	0.59	0.52	0.56	1.45	4.00
8	521	640	97.46	97.18	0.41	0.29	0.40	0.50	1.20	4.65
9	628	753	98.11	97.83	0.30	0.22	0.35	0.45	1.27	4.47
10	749	993	98.72	98.38	0.23	0.16	0.27	0.40	1.21	3.94
11	818	1447	98.97	98.94	0.19	0.09	0.25	0.32	1.13	4.02
12	958	1877	99.27	99.22	0.14	0.06	0.20	0.28	1.20	4.04
13	1134	2367	99.47	99.47	0.10	0.05	0.17	0.25	1.02	3.92
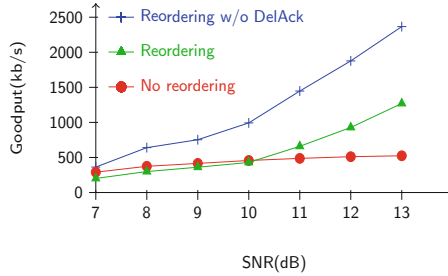
Fig. 4. Impact of reordering mechanism on end to end goodput

(a) Impact on RTO   (b) Impact on DUPACK   (c) Impact on spurious
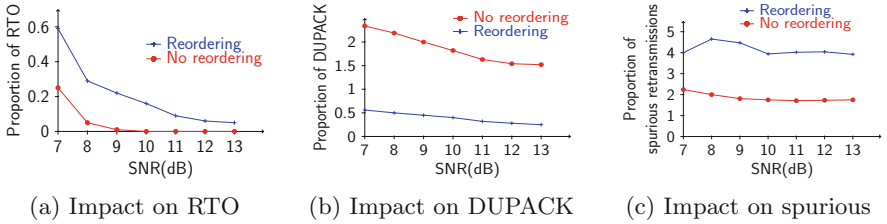
Fig. 5. Impact of reordering mechanism on CUBIC performance

the decreasing number of retransmissions. We also observe a slightly increase of
the number of RTO although the overall performance are much more better than
without reordering as shown in Table 1. Actually, some losses are not recovered
by DUPACK but by RTO because of decoding fails or when the queue is full
as explained in the previous Sect. 4. We also observe an increase of the num-
ber of spurious retransmissions when the TCP goodput increases significantly.
The reason is that TCP becomes more and more opportunistic as the network
conditions are getting better and then congestion appears on other parts of the
networks. As a result, thanks to the conjointly used of the A-HARQ and re-
ordering mechanisms, the bottleneck is not the LMS channel anymore (Fig. 5).

We recall that we only presented results with TCP NewReno and CUBIC. However, the same trend has been observed with TCP Westwood and should be also observed with other TCP variants. CUBIC is well-known to achieve better performance over high-delay bandwidth product networks. This explains the higher performance obtained by this protocol compared to TCP NewReno.

5   Analysis

The results presented in Sect. 3 highlight the interest for mitigating the number of out-of-order packets linked to the use of link layer reliability schemes. As shown in Sect. 4, link layer buffering results in a strong increase of the TCP goodput with a significant decrease of the number of DUPACK and spurious retransmissions for TCP NewReno. This is not only a gain for the use of the expensive satellite capacity, but also for the end-to-end latency. At last but not least, this makes it easier for delay tolerant services to respect their delivery rate constraints.

Concerning standard default TCP parameters tested, using DelAck with the reordering mechanism is counterproductive for TCP performance. As shown in Fig. 4, we observe that the goodput decreases when DelAck is added with the reordering buffer while better without this buffer. The problem is that DelAck might delay the acknowledgment pace while the reordering buffer might also delay the pace of data forwarding to the upper layer. As a matter of fact, both delayed schemes negatively interact between them. Our recommendation is thus to disable DelAck in this context to reach the highest performance if possible. Note that in another context of multipath communications over a satellite constellation, the authors in [1] showed that DelAck has also a negative impact on TCP performance.

With reordering mechanism and when the goodput is high enough, congestion appears on the network, and packets are dropped for reasons that are not due to satellite environment or A-HARQ. This means that A-HARQ is recovering enough packets to allow TCP to work in good conditions. So, TCP metrics are more impacted by congestion in the network than by the few packets dropped

(a) SNR=8   (b) SNR=13
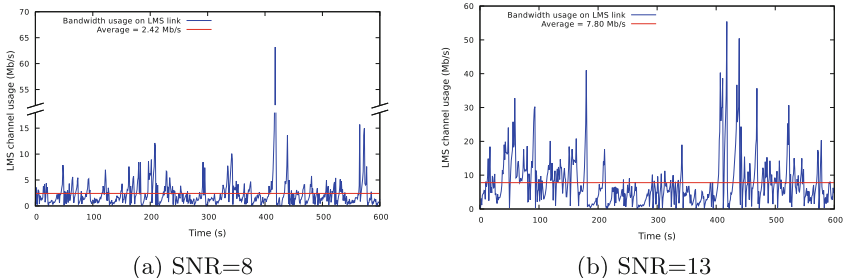
Fig. 6. Usage of the LMS link, using CUBIC
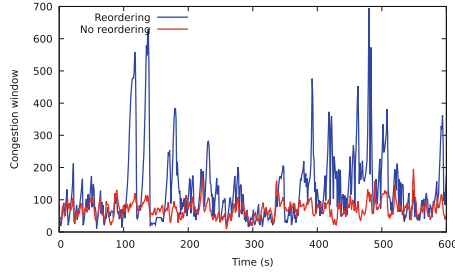
Fig. 7. Evolution of CUBIC congestion window, averaged each second, with SNR = 8

by A-HARQ. In these scenarios, we observe a good usage of the LMS link. We can see in Fig. 6b, corresponding to $SNR = 13$ and with reordering, that the maximum capacity of this link is reached sometimes, with an average usage of 7.8 Mb/s. As a reminder, we set the capacity of this link at 50 Mb/s. This usage corresponds to the real number of bits flowing through this link, including the redundancy bits and the retransmissions by A-HARQ. This usage is even better with higher SNR.

We also observe, as presented in Fig. 7, a real improvement of the congestion window with reordering mechanism. This evolution is correlated with the channel evolution, and confirms the goodput improvement brought out by this mechanism.

To conclude, the results we obtained for A-HARQ can be extended to all HARQ mechanisms, since they all imply varying delays and out-of-order packets in TCP layer. Thus, **adding a reordering mechanism after HARQ should be the standard for all HARQ mechanisms, in order to always guarantee good TCP performance.**

6 Conclusion

We investigated in this paper the performance obtained by TCP within a delay tolerant service over mobile satellite communications. We show that link-layer schemes used in the context of space or aeronautical communications have an impact on the TCP performance. We observe that using TCP over such link-layer schemes leads to weak performance due to high delay jitter. Basically, there is a clear discrepancy between TCP and these link-layer schemes while sharing the same goal. Using temporal traces representing the delay between a mobile terminal and the last hop satellite from a LEO constellation, we identified link-layer parameters that impact on the TCP performance. We proposed a buffer scheme allowing both link-layer and transport layers to conjointly perform. Simulations show that thanks to this simple buffer scheme, we can drastically increase the TCP performance without modifying TCP internal algorithms. In a future work, we seek to extend these preliminary measurements with various kind of traffic and in particular, to assess the impact on short-lived flows over a real satellite link provided by CNES agency.

References

1. Wood, L., Pavlou, G., Evans, B.: Effects on TCP of routing strategies in satellite constellations. IEEE Commun. Mag. **39**(3), 172–181 (2001)
2. Pratt, S.R., et al.: An operational and performance overview of the IRIDIUM low earth orbit satellite system. IEEE Commun. Surv. **2**(2), 2–10 (1999)
3. Taleb, T., Kato, N., Nemoto, Y.: Recent trends in IP/NGEO satellite communication systems: transport, routing, and mobility management concerns. IEEE Wirel. Commun. **12**(5), 63–69 (2005)
4. Diana, R., et al.: DTN routing for quasi-deterministic networks with application to LEO constellations. Int. J. Satell. Commun. Network. **35**, 1–18 (2015)
5. Blaunstein, N., Cohen, Y., Hayakawa, M.: Prediction of fading phenomena in land-satellite communication links. Radio Sci. **45**(6) (2010)
6. Sastry, A.: Performance of hybrid error control schemes of satellite channels. IEEE Trans. Commun. **23**(7), 689–694 (1975)
7. De Cola, T., Ernst, H., Marchese, M.: Application of long erasure codes and ARQ schemes for achieving high data transfer performance over long delay networks. In: Re, E.D., Ruggieri, M. (eds.) Satellite Communications and Navigation Systems. SCT, pp. 643–656. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-47524-0_49
8. Kuhn, N., et al.: On the impact of link layer retransmission schemes on TCP over 4G satellite links. Int. J. Satell. Commun. Network. **33**(1), 19–42 (2015)
9. Gineste, M., Gonzalez, H.: On the usage of a dedicated data link layer and half duplex terminals in satellite systems for future air traffic management. In: 28th AIAA International Communications Satellite Systems Conference (2010)
10. Kuhn, N., et al.: On the impact of link layer retransmissions on TCP for aeronautical communications. In: 5th International Conference on Personal Satellite Services - PSATS (2013)
11. Ahmad, R.A., et al.: Enhanced HARQ for delay tolerant services in mobile satellite communications. In: The Seventh International Conference on Advances in Satellite and Space Communications SPACOMM, Barcelona, ES, pp. 1–6, April 2015
12. Wood, L.: SaVi: satellite constellation visualization. In: First Annual CCSR Research Symposium - CRS 2011 (2012)
13. Chotikapong, Y., Cruickshank, H., Sun, Z.: Evaluation of TCP and Internet traffic via low Earth orbit satellites. IEEE Pers. Commun. **8**(3), 28–34 (2001)
14. Perez-Fontan, F., et al.: S-band LMS propagation channel behaviour for different environments, degrees of shadowing and elevation angles. IEEE Trans. Broadcast. **44**(1), 40–76 (1998)
15. Fontan, F.P., et al.: Statistical modeling of the LMS channel. IEEE Trans. Veh. Technol. **50**(6), 1549–1567 (2001)

Joint Beam Hopping and Precoding in HTS Systems

Alberto Ginesi[✉], Emiliano Re, and Pantelis-Daniel Arapoglou

ESA/ESTEC, Keplerlaan 1, 2200 AG Noordwijk, The Netherlands
{alberto.ginesi,emiliano.re,pantelis-daniel.arapoglou}@esa.int

Abstract. This paper presents a novel concept for offering payload resources flexibility in High Throughput Satellite (HTS) systems. The concepts makes joint use of two advanced techniques, namely beam hopping and precoding. The combination of these two techniques allows the system to really optimize the performance of beam hopping in terms of capability to follow the temporal and spatial variation of user traffic requests within the coverage. The performance of such an approach is demonstrated through computer simulations of an exemplary system. A similar approach can also be used by combing precoding with frequency flexible techniques. Additional combination of on-board power pooling techniques helps to further improve the system performance.

Keywords: HTS · Beam hopping · Precoding · Broadband · Payload flexibility

1 Introduction

The capability to flexibly allocate the satellite payload resources over the service coverage is becoming a must for next generation broadband satellites employing a number of spot beams. Indeed, past and current broadband systems have shown that large multi-beam High Throughput Satellites (HTS) are typically able to fill-up fairly quickly the capacity of some beams, while some others remain (almost) empty over a relatively long part of the satellite life time. The consequence is a loss of satellite operator's revenue due to the number of customers lost within the hot-spots (filled-up beams) and the waste of resources over the empty spots.

The primary goal of flexibility is to minimize the unused and unmet capacity. The introduction of flexibility helps a satellite operator to manage the risks accounted by the unpredicted changes, like regulatory context, competing context, socio-economic context.

Flexibility refers to the ability to change the configuration of the system during the operational life of the satellite.

In the following, we will focus on the flexibility in the forward link of HTS systems. The forward link consists of the uplink between the gateway ground station and the satellite and the downlink between the satellite and the user terminals. The forward (FW) link is anyway the most important in determining the revenues of the operator, as it is the main traffic direction in the network.

1.1   Key Capacity Definitions

In order to correctly characterize the performance of a broadband HTS system the following definitions are in order:

- The capacity/throughput demand is the capacity that is requested by the users which is typically geographically non-uniform and time variant.
- The offered system capacity/throughput represents the maximum capacity of the system, while considering an infinite capacity demand per location.
- The usable system capacity/throughput is the capacity that is really sold taking into account the real capacity demand per location.
- The unused system capacity/throughput is the difference between the offered capacity and the usable system capacity.
- The unmet capacity/throughput demand is the difference between the capacity demand and the offered capacity.

1.2   Flexible Payload Techniques

A number of techniques are available to support flexibility. Ignoring for the moment the case of coverage flexibility, the following is a summary of such techniques:

Flexible power allocation

To better match the capacity demand in each beam, one approach is to distribute the total amount of payload power unevenly across the different beams. Lower power would be assigned to beams with lower capacity demand, while higher power would be given to hot spots. This technique is typically implemented by means of flexible Travelling Wave Tube Amplifier (FlexTWTA) technology [1, 2], where the saturated power of a TWTA is adjusted according to the capacity demand of beams served by the amplified carriers. In case of one High Power Amplifier (HPA) shared between two beams (which is a typical configuration), this technique works if the two beams have similar capacity demand. Alternatively, if the two beams have different capacity demand the power transfer from one beam to the other is done by suppressing part or all the carriers serving the beam of low demand.

An alternative approach for realizing flexible power allocation foresees the exploitation of Multi-Port Amplifiers (MPAs) [3] instead of FlexTWTA.

The drawbacks of flexible power allocation are that any power variation has intrinsically a limited impact to the offered beam capacity due to both the inherent diminishing return behavior of the Shannon function (spectral efficiency versus power), as well as the presence of residual intra-system co-channel interference. Other drawbacks concern the relatively high cost of the FlexTWTA and MPA components.

Flexible Bandwidth allocation

This technique consists in tuning the amount of band that is allocated to a given beam according to the relative capacity demand. Basically, part of the amount of bandwidth that is allocated to low demanding beams gets transferred to high demanding beams. This can be achieved, for example, by splitting unevenly the user bandwidth allocated

to the two beams served by the same on-board HPA (which is a typical configuration for a four color scheme network) and flexibly routing the two portions of the bandwidths to different antenna feeds.

The drawback of such an approach is that in general additional intra-system co-channel interference will be generated due to the possible overlap of the bands assigned to two adjacent co-polar beams. Although some countermeasures can be conceived in order to partially limit the impact of the high co-channel interference in part of the user bandwidth (for example, by a cautious assignment of users to the highly interfered portion of the band), the result of this extra interference limits the efficiency of such technique particularly when considering certain traffic demand distributions.

Flexible time allocation, i.e. Beam Hopping (BH)

This technique [4] is exactly dual w.r.t the flexible bandwidth allocation technique, i.e. it can be explained by replacing time with frequency. Indeed, this solution can be implemented through the so-called BH scheme by which different co-channel beams served by the same HPA get allocated different time slots. By modulating the duration of the time-slots, different offered capacity values can be achieved in different beams. For an uneven capacity demand distribution, adjacent beams might end up being served by different HPAs with overlapping time slots thus generating excessive intra-system interference.

2   The Novel Flexible Payload Technique

Although the frequency flexible techniques can also be addressed, in the rest of the paper we will consider BH as the payload technique offering resource flexibility.

This paper proposes a solution which greatly improves the efficiency of BH by mitigating the co-channel interference that such techniques might end up generating for certain traffic demand distributions. This is graphically illustrated in the next figure where three hot-spot co-polar beams are located geographically close to each other. Due to their high composite user traffic requests, these beams are assigned by the network resource manager long time slots with large overlapping times where high co-channel interference is generated. This situation is typical of a cluster of hot spots within an HTS network.

If the extra generated interference could somehow be cancelled, BH would be able to more efficiently match any capacity demand distribution over the coverage, as the additional time allocation would basically be interference free and thus its benefit to the overall offered capacity would not be only higher but also much more predictable.

Precoding [5, 6] is ground-based technique that is used to pre-cancel co-channel intra-system interference by applying at the Gateway of the broadband network a linear combination of the transmitted signals over the different beams. In practice, the transmitted signals are weighted by complex coefficients from a so-called precoding matrix that performs a sort of inversion of the channel matrix. The coefficients of the linear combination are computed based on feedbacks provided by the user terminals (the channel estimates, including both the amplitude and phase).

According to the novel joint precoding and BH technique [7] proposed in this paper, following a re-configuration of time slots assignment to beams, the user terminals would perform a new channel estimation procedure (training phase). This is necessary as the

precoding matrix is formed for each specific set of served user terminals. After a relatively short amount of time (typically much less than one second), the user terminals would be able to report the new estimates to the GW which in turn would apply precoding thus reducing the interference in situations like the one described in Fig. 1.
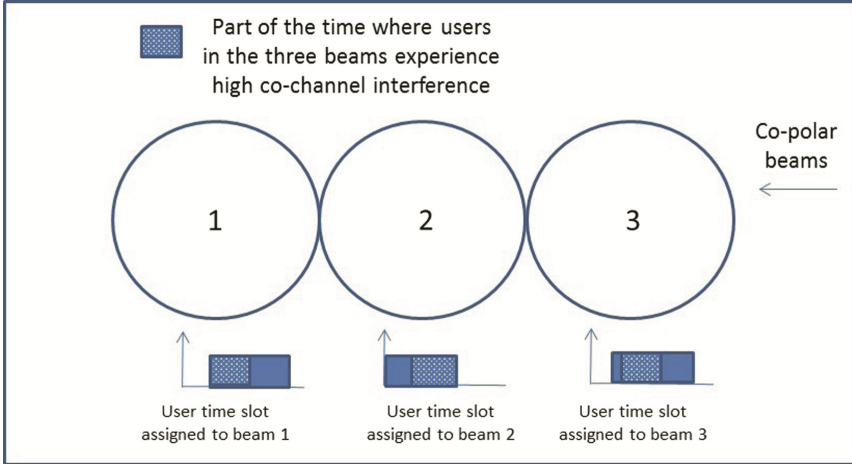
Fig. 1. The excessive co-channel interference issue in flexible payload beam hopping techniques

2.1 Basic Resource Allocation Algorithm

In order to take full advantage of the joint technique, a suitable payload resource allocation algorithm has to be devised.

The problem of resource allocation (RA) belongs to the more general framework of RA for MIMO systems [8]. Here a pragmatic approach for the derivation of the time illumination plan is used. Indeed, the RA algorithm (time illumination plan) is quite simple as the precoded system might be approximated as interference-free. Therefore, the number of illumination time slots to be assigned to each beam can be computed by dividing the traffic request in that beam by the spectral efficiency achievable in absence of interference. Finally, a normalization coefficient is applied to take into account the overall time resources offered by the HPA serving the beam under consideration. The relative equation reads as follows:

$$\tau_j = \frac{R_{req}(j)\big/\eta(j)}{\sum_{k=0}^{K_{HPA(i)}-1}\big(R_{req}(k)\big/\eta(k)\big)}W, j \in HPA_i, \ i = 0, 1, \ldots, K_{HPA} \tag{1}$$

where:

- τ_j represents the number of time slots allocated to beam j
- $R_{req}(j)$ represents the requested traffic throughput for beam j
- $\eta(j)$ represents the average spectral efficiency for beam j

- *W* represents the BH illumination period (total number of time slots)
- HPA_i is the set of indexes identifying the time slot for the i_{th} HPA
- K_{HPA} is the number of HPAs

When considering the scenario where the satellite network is served by a number of gateways (GWs), each GW typically addresses a cluster of 8–16 beams maximum. In this case, precoding can be easily applied as a 16×16 matrix across the full cluster but it cannot mitigate the co-channel interference between two beams belonging to two different clusters. This situation can be addressed and completely solved by using a centralized precoding approach whereby a central processor distributes the precoded signals to all GWs within the network. Alternatively, a mitigation of the issue can be achieved by modifying the resource allocation algorithm in order to make it "GW cluster aware" whereby the allocation of resources (time slots) is done in order to minimize the co-channel interference between adjacent beams belonging to adjacent GW clusters.

2.2 Adding Power Flexibility

An additional well known improvement in flexibility can be provided by flexibly allocating the power to the beams according to their capacity request. Differently to conventional (non-precoded) systems, when using precoding, any beam power unbalance results in direct throughput improvement due to the reduced intra-system interference.

Power flexibility can be achieved by replacing conventional tubes with either FlexTWTAs or MPAs (Multi-Port Amplifiers). An example of a payload architecture using power flexibility by means of MPAs and BH is shown in Fig. 2. MPAs are more suitable as they allow for a larger dynamic range, therefore providing a more flexible power partitioning among the beams. To be noticed that if the scheme with MPAs is used, one has to be careful in assigning beams to MPAs, as a good performance of these devices requires very low correlation among the carriers sharing the MPA. This means that carriers which are precoded together, in principle, should not be using the same
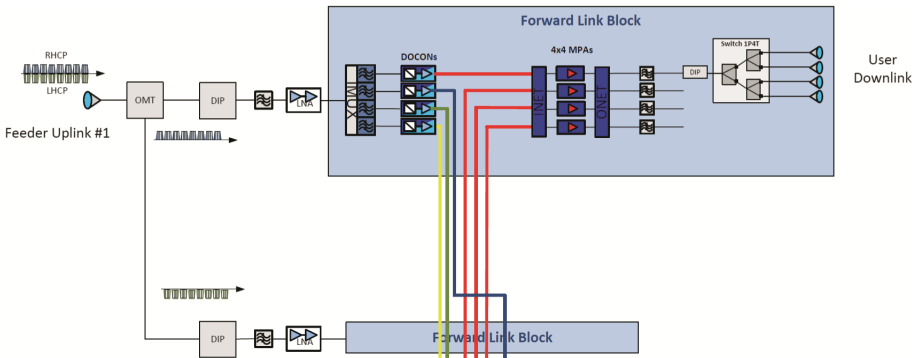

Fig. 2. Block diagram of a BH –based FW link payload with power pooling – only one feeder uplink shown for simplicity

MPA. This might imply that an MPA should be fed only by carriers belonging to distant beams as in this case precoding would only add a low correlation.

The precoding algorithm in the presence of non-uniform power unbalance has to be modified. If a linear precoding algorithm of the MMSE (minimum mean square error) variant is employed to calculate the precoding matrix $\mathbf{W}$ from the channel matrix $\mathbf{H}$ (both with complex elements), the calculation of the precoding matrix in the presence of non-uniform power is as follows:

$$\mathbf{W} = \left[\mathbf{H}^H \mathrm{diag}(\mathbf{P})\mathbf{H} + \mathbf{I} \right]^{-1} \mathbf{H}^H \mathrm{diag}(\mathbf{P}). \tag{2}$$

where $\mathbf{P} = \left[P_i \right]$ and P_i is the power emitted from feed i, $\mathbf{I}$ is the identity matrix and $\mathbf{H}^H$ is the Hermitian of matrix $\mathbf{H}$.

The power unbalance of the system is reflected by the vector $\mathbf{P}$ with one entry per transmit feed. If $\mathbf{W}_i$ is a row of the precoding matrix, to check if any of the rows of $\mathbf{W}$ violates the maximum power available from the totality of on board HPAs, the following norm must be calculated

$$n = \mathrm{norm}(\mathbf{W}_i)^2$$

Then, if $n > 1$, the following normalization is in order to ensure the total power is not exceeding the total available power on board $\mathbf{W}_i = \dfrac{1}{\sqrt{n}} \mathbf{W}_i$.

The optimization of the power allocation algorithm is in general a non-trivial task. Here we have followed a heuristic approach which, after some mathematical manipulations (not shown here for brevity), results in the following equation:

$$\Delta P_i = \frac{\left[2^{\left(\frac{R_{unmet,i} W}{\tau_i B} \right)} (SNR_i + 1) \right] - 1}{SNR_i} \tag{3}$$

Where:

$R_{unmet,i}$ is the unmet capacity of beam i;
B is the band served by each amplifier;
W is the number of time slots in the BH window;
SNR_i is the average SNR of beam i with uniform power.
The power transmitted per beam is then computed as:

$$P_i = P_{un} + \Delta P_i \tag{4}$$

where P_{un} is the power transmitted per beam in the uniform power case.

However, this computation does not yet ensure meeting any constraint concerning the total power. Indeed, we have to make sure that the total power at the level of a single

MPA is constrained. This is achieved by normalizing Eq. 4 w.r.t the total power of the MPA serving the considered beam.

3   System Simulation Results

In order to correctly characterize the benefits of joint BH and precoding for satellite broadband networks, a high number of simulations have been run against different system and traffic distribution assumptions. From an analysis of the results a first conclusion is that these are particularly sensitive to the type of geographical traffic distribution and the relative association of beams to on-board HPAs. In order to understand why, in the following the throughput results are shown for an exemplary broadband network which has been chosen with a relatively limited number of beams (64) in order to allow for an efficient representation of the results. Table 1 below outlines some key system parameters of the case study.

Table 1. Key system parameters of the simulated system

User frequency	Polarization	Number of HPAs	Number of beams	Number of carriers per HPA	On-board downlink EIRP	Air interface	User terminal antenna diameter
Ka-band 500 MHz	Only one used	16 (4 beams per HPA)	64	1	65 dBW	DVB-S2x (roll-off = 20%)	70 cm

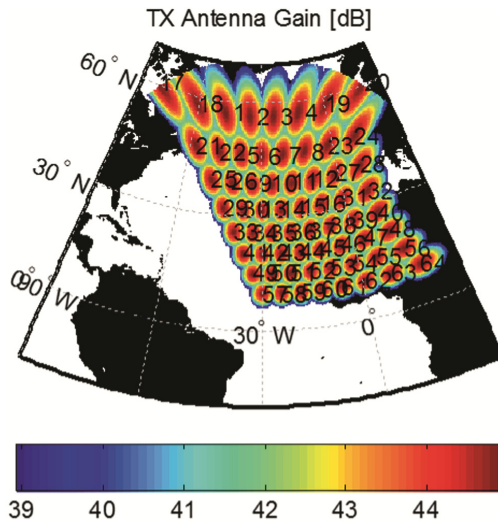The relative user link antenna pattern is shown in Fig. 3.

Fig. 3. Satellite gain antenna pattern in the forward user link of the system case study

The estimated flat FW link throughput of this system is around 15.5 Gbps (i.e. using BH with regular time slot allocation per beam).

We have then assumed to have a traffic distribution with a hot spot configuration whereby out of the 64 beams, 16 central beams are 'hot' beams each requesting 920 Mbps and the remaining 48 beams are 'cold' beams each requesting 10 times less traffic than the hot ones, i.e. 92 Mbps. For this traffic scenario, we assumed two different payload configurations. The first one, which will be dubbed FLEX OPT (as it represents the best configuration from the payload perspective), where each HPA is connected to 3 cold beams and 1 hot beam; the second configuration, which will be dubbed FLEX WRST (as instead represents the worst configuration from the payload perspective) where each HPA is either connected to 4 cold beams of 4 hot beams. It is clear why FLEX OPT is able to deliver the best performance in terms of usable capacity. Indeed, in this payload configuration each HPA resources can be really optimized for the hot spot by allocating the longest time slot to the hot beam and subtracting time resources to the other served low traffic beams. This cannot be done in FLEX WRST as all beams served by each HPA have similar traffic demand (Table 2).

Table 2. Simulation results in terms of offered and usable throughput for the system case study and the two payload configurations

	FLEX OPT offered throughput Gbps	FLEX OPT usable throughput Gbps	FLEX WRST offered throughput Gbps	FLEX WRST usable throughput Gbps
BH	12.5	11.8	9.4	8.8
BH + Precoding	16.9	16.8	9.6	9
BH + Flexible power				9.1
BH + Precoding + Flexible power				9.8

The results shows that joint BH and precoding allows to gain around 40% in usable throughput for FLEX OPT while the gain reduces to around 2% for FLEX WRST. The usage of flexible power has also been tested for FLEX WRST only. In this case, the gain of using precoding in addition to BH and flexible power is about 8%. The higher gain in this case is justified by the consideration that precoding, given the reduction of co-channel interference that it involves, it also allows to better exploit any power flexibility.

Simulation results using other systems showed that joint precoding and BH technique could deliver usable throughputs between 0 to 50% higher than BH alone, depending on the traffic geographical distribution and the relative allocation of HPAs to beams.

4   Conclusions

This paper has described a novel technique for allocating FW link throughput resources. It consists of a combination of well know techniques like Beam Hopping and Precoding. Additional power flexibility can also be added to further enhance the flexibility

performance. The resource allocation algorithms are described in details and the overall system FW link usable throughput has been evaluated for an exemplary satellite broad-band network. In general, the performance gains in terms of usable throughput w.r.t state of the art (i.e. BH only) vary quite significantly with the user traffic distribution within the network. A number of simulation results (not shown in the paper for reason of space limitation) show gains between 0 to 50%. This means that during the satellite life time, depending on how the traffic geographical distribution would change, the usable throughput might be boosted by the joint technique by up to 50%.

References

1. Kaliski, M.: Evaluation of the next steps in satellite high power amplifier technology: flexible TWTAs and GaN SSPAs. In: IEEE International Vacuum Electronics Conference, Rome, Italy (2009)
2. Cuignet, E., Tonello, E., Maynard, J., Boone, P.: Very high efficiency dual flexible TWTA, a flexible concept allowing to deal with performances and schedule constraints of telecommunication payloads. In: 2013 IEEE 14th International Vacuum Electronics Conference (IVEC) (2013)
3. Mallet, A., Anakabe, A., Sombrin, J., Rodriguez, R., Coromina, F.: Multi-port amplifier operation for Ka-band space telecommunication applications. In: 2006 IEEE MTT-S International Microwave Symposium Digest (2006)
4. Angeletti, P., Prim, F.D., Rinaldo, R.: Beam hopping in multi-beam broadband satellite systems: system performance and payload architecture analysis. In: Proceedings of the AIAA Conference, June 2006
5. Angel Vazquez, M., Perez-Neira, A., Christopoulos, D., Chatzinotas, S., Ottersten, B., Arapoglou, P.-D., Ginesi, A., Taricco, G.: Precoding in multibeam satellite communications: IEEE Wirel. Commun. **23**(6) (2016)
6. Arapoglou, P.-D., Ginesi, A., Cioni, S., Erl, S., Clazzer, F., Andrenacci, S., Vanelli-Coralli, A.: DVB-S2X-enabled precoding for high throughput satellite systems. Int. J. Satell. Commun. Network. **34**, 439–455 (2016)
7. Interference-Resilient Flexible Techniques for Payload Resource Allocation in Broadband Satellites, Patent Application number PCT/EP2016/063358
8. Castañeda, E., Silva, A., Gameiro, A., Kountouris, M.: An overview on resource allocation techniques for multi-user MIMO systems. IEEE Commun. Surv. Tutor. **19**(1), 239–284 (2017)

Link Adaptation Algorithms for Dual Polarization Mobile Satellite Systems

Anxo Tato[1(✉)], Pol Henarejos[2], Carlos Mosquera[1], and Ana Pérez-Neira[2,3]

[1] Universidade de Vigo, Vigo, Spain
{anxotato,mosquera}@gts.uvigo.es
[2] Centre Tecnològic de Telecomunicacions de Catalunya, Castelldefels, Barcelona, Spain
pol.henarejos@cttc.es, ana.perez@cttc.cat
[3] Universitat Politècnica de Catalunya, Vigo, Spain
http://www.gpsc.uvigo.es, http://www.cttc.cat

Abstract. The use of dual polarization in mobile satellite systems is very promising as a means for increasing the transmission capacity. In this paper a system which uses simultaneously two orthogonal polarizations in order to communicate with the users is studied. The application of MIMO signal processing techniques along with Adaptive Coding and Modulation in the forward link can provide remarkable throughput gains up to 100% when compared with the single polarization system. The gateway is allowed to vary the MIMO and Modulation and Coding Schemes for each frame. The selection is done by means of a link adaptation algorithm which uses a tunable margin to achieve a predefined target Frame Error Rate.

Keywords: Link adaptation · Adaptive Coding and Modulation
MIMO · Dual-polarization · Satellite communications
Mobile satellite systems

1 Introduction

In recent years, the spectrum saturation and the increasing demand for higher data rates in a ubiquitous way encourages the engineers to design new techniques in order to increase the capacity of communication systems without resorting to expand the occupied bandwidth. Two of these techniques are the leverage of multiple antennas at both transmitter and receiver by means of MIMO (Multiple Input Multiple Output) signal processing techniques and also the so-called Adaptive Coding and Modulation (ACM) or link adaptation. Both are part of many current terrestrial wireless communication standards such as LTE [1] and IEEE 802.16 [2] for cellular technologies and IEEE 802.11 [3] for wireless local area networks.

In this paper the application of MIMO techniques to mobile satellite communication systems by exploiting the polarization domain is proposed. Satellite links operating at low frequency bands such as L- and S-bands, usually rely

on a single circular polarization, either RHCP (Right Hand Circular Polarization) or LHCP (Left Hand Circular Polarization), to avoid the effects of the Faraday rotation [4]. Here, the simultaneous use of both orthogonal polarizations within the same beam to communicate with the users is contemplated following up some preliminary studies presented in [5] within the framework of SatNEx-IV, the European Satellite Network of Experts. An algorithm to switch among different MIMO schemes will be presented. In particular, SISO, Orthogonal Polarization-Time Block Codes (OPTBC), Polarization Modulation (PMod) or V-BLAST (Vertical Bell Laboratories Layer Space-Time) will be chosen as a function of a set of effective Signal to Noise Ratio (SNR) metrics.

On the other hand, the use of ACM techniques allows the system to adapt its instantaneous rate to the current channel capacity without designing the system for the channel worst case. In this paper, following our previous works [6–8], the use of a link adaptation algorithm based on an adaptive margin is proposed in order to select the Modulation and Coding Scheme of each frame (MODCOD or MCS). This algorithm is complemented with the MIMO mode selection scheme to increase the spectral efficiency whilst trying to maintain the Frame Error Rate (FER) at a level specified by the Quality of Service (QoS) parameter of the connection.

The remainder of the article is structured as follows. Section 2 provides a description of the satellite communication system, including the link adaptation algorithms, an overview of the channel model and how the channel series are generated. Section 3 describes the link adaptation algorithms for selecting both MIMO mode and MCS. Section 4 provides the simulation results of the algorithms in a maritime mobile scenario. Lastly, in Sect. 5 the main conclusions are collected and we note the future work on this topic.

2   System Model

A satellite communications system which serves a single mobile user operating in L- or S-band is under consideration here and a link adaptation algorithm is proposed for the forward link, communications from the gateway towards the mobile terminal (MT). Typically, at these bands only one polarization is used to protect from high cross-polarization leakage. For example, the commercial service BGAN (Broadband Global Area Network), standardized by the ETSI as TS 102 744, only employs RHCP [9]. In higher frequency bands, like Ku and Ka, the high XPD (cross polar discrimination) would make it possible to use two independent streams in both polarizations.

The analysed system uses simultaneously both polarizations, yielding a 2×2 MIMO system, is narrow-band and single-carrier and, to some extent, it is inspired by BGAN. The baseline Single Input Single Output (SISO) system is complemented by additional MIMO modes, namely OPTBC, PMod and V-BLAST:

– OPTBC is based on Alamouti space-time coding, used for achieving transmit diversity and introduced in [10]. The spatial components in Alamouti are replaced by the two available polarizations in OPTBC.

- PMod is analogous to Spatial Modulation (SM) techniques [11], with only one polarization per channel use transmitting the symbols from a given constellation. One extra bit of information is conveyed indicating which polarization is chosen. For a QPSK constellation this implies a 50% gain in spectral efficiency.
- In V-BLAST [12] two independent symbols are transmitted per channel use and the receiver is in charge of reducing the cross-stream interference to perform the detection.

All these three MIMO schemes do not require Channel State Information (CSI) at the transmitter and therefore are suitable for the satellite scenario where the CSIT is outdated due to the long round trip time.

In addition to the forward link, where our algorithms operate, there is a feedback channel in the return link used by the MT to inform the gateway about the CSI, the result of the frames decoding (in the form of ACK/NACK) and the optimum transmission mode (which is calculated by the receiver). All these information is used by the gateway, our transmitter, in the selection of the preferred MCS by the link adaptation algorithm.

The signal model for a given time instant n is

$$\boldsymbol{y}_n = \sqrt{P}\boldsymbol{H}_n\boldsymbol{x}_n + \boldsymbol{w}_n \tag{1}$$

where P is the transmitted power, $\boldsymbol{H}_n \in \mathbb{C}^{2\times2}$ is the channel matrix, $\boldsymbol{y}_n \in \mathbb{C}^2$ is the vector of the received signal which has a component per each polarization, $\boldsymbol{x}_n \in \mathbb{C}^2$ is the transmitted signal and $\boldsymbol{w}_n \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_2)$ represents the additive complex white Gaussian noise (AWGN).

The transmitted symbols are grouped into blocks or packets (codewords) which span $N = 2560$ symbols, transmitted with a baud rate of 33600 symbols/s, which gives a frame length of 80 ms. Since the frames of the selected bearer contain just one block, hereafter the term frame will be employed for referring to a codeword. In order to speed up the simulations, the entire signal processing block chain is not implemented and we resort to Physical Layer Abstraction techniques, in particular the Mutual-Information Effective SNR [6]. The comparison of the effective SNR of a received frame with a threshold SNR of the MCS used to transmit that frame allows us to decide if the frame can be decoded (the effective SNR is higher than the corresponding threshold) or not.

The effective SNR of a frame, SNR_{eff}, given the set of the N SNRs of each symbol γ_n can be obtained with

$$\text{SNR}_{eff} = \Phi^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\Phi(\gamma_n)\right), \tag{2}$$

procedure called SNR compression in [13]. The SNR γ_n for each symbol period is calculated for the different MIMO modes as:

- SISO:

$$\gamma_n = \text{SNR}|[\boldsymbol{H}_n]_{11}|^2 \tag{3}$$

– OPTBC:

$$\gamma_n = \frac{\text{SNR}}{2}\|\boldsymbol{H}_n\|_F^2 \tag{4}$$

– PMod:

$$\gamma_n = \text{SNR}\frac{|[\boldsymbol{H}_n]_{11}|^2 + |[\boldsymbol{H}_n]_{22}|^2}{2} \tag{5}$$

– V-BLAST (MMSE (Minimum Mean Square Error) receiver):

$$\gamma_{n,k} = \frac{1}{[(\boldsymbol{I}_2 + \dfrac{\text{SNR}}{2}\boldsymbol{H}_n^H\boldsymbol{H}_n)^{-1}]_{kk}} - 1 \tag{6}$$

In the previous equations $\text{SNR} = P/\sigma^2$, $[\boldsymbol{H}_n]_{ij}$ denotes the coefficient (i, j) of the channel matrix for time instant n, k indicates the number of the stream (1 or 2), $\boldsymbol{I}_2$ is the 2×2 identity matrix and $\|\boldsymbol{H}_n\|_F$ denotes the Frobenius norm of the matrix. Note that in V-BLAST the type of receiver determines the SNR expression, here a MMSE receiver is assumed to be used given its simplicity and robustness against noise.

Table 1 collects all the available transmission modes and MCS, including the required effective SNR for correct decoding (obtained from the curve of SNR vs Mutual Information) and corresponding spectral efficiency in bps/Hz. An important aspect is that in this paper PMod is considered ideal with an efficiency 1.5 times larger than SISO, by assuming that the polarization bit is ideally decoded.

Table 1. Coding rate options for the F80T1Q-1B bearer [9], QPSK constellation.

	L8	L7	L6	L5	L4	L3	L2	L1	R
Coding rate	0.34	0.40	0.48	0.55	0.63	0.70	0.77	0.83	0.87
Threshold SNR (SNR_{th}) (dB)	-2.15	-1.21	-0.09	0.83	1.84	2.74	3.67	4.54	5.19
Spectral efficiency SISO/OPTBC	0.68	0.80	0.96	1.10	1.26	1.40	1.54	1.66	1.74
Spectral efficiency PMod	1.02	1.20	1.44	1.65	1.89	2.10	2.31	2.49	2.61
Spectral efficiency V-BLAST	1.36	1.60	1.92	2.20	2.52	2.80	3.08	3.32	3.48

2.1 Channel Model

The simulation of the mobile satellite dual polarized channel has been done following the work of [14]. The channel follows a Rice distribution and is obtained as the sum of three different components: the Line-of-Sight (LoS), the specular reflected signal and the diffuse components, produced by the scatterers near the MT. The first two components are fixed and the third, the fast fading, is Rayleigh distributed, which altogether gives a Rice distribution. These three components are grouped into the following channel matrix:

$$\boldsymbol{H} = \boldsymbol{\beta}e^{j\phi}\boldsymbol{K}_L + \boldsymbol{\xi}e^{j\phi}\boldsymbol{K}_S + \boldsymbol{D}\boldsymbol{K}_D \tag{7}$$

$\boldsymbol{K}_L$, $\boldsymbol{K}_S$ and $\boldsymbol{K}_D$ are the K-factor matrices which collect the Rice factors of each polarization for the LoS, specular and diffuse components, respectively. The two matrices $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ are related to how the channel mixes the two polarizations and matrix $\boldsymbol{D}$ entries are zero-mean complex Gaussian random variables with a given covariance matrix. The parameters for building all these matrices depend on the considered environment and can be found in [14].

2.2   Channel Generation

Figure 1 shows the block diagram of the channel time series generator. One important aspect of the generated channel series is the time correlation and the Doppler spread. Assuming the Clarke's model, the coherence time can be approximated by

$$\tau_c = \frac{3\lambda}{4v\sqrt{\pi}}, \tag{8}$$

being λ the wavelength and v the MT speed. Firstly, $Q = \lceil N/(\tau_c f_s) \rceil$ independent realizations of the channel matrices are generated from the uniformly distributed random phases and the complex Gaussian random matrix $\boldsymbol{D}$. Then, a linear interpolation is performed to obtain the N channel matrices and a low-pass-filter with cut-off frequency equal to the Doppler spread, $D_s = v/\lambda$, is applied. Lastly, the channel matrices are scaled to yield the given average SNR of the simulation.

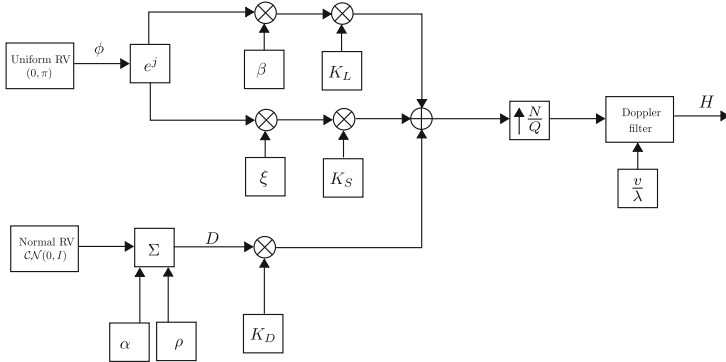**Fig. 1.** Block diagram of the channel time series generator

3   Algorithm for Mode and MCS Selection

In this paper, the link adaptation procedure has to select the MCS and the MIMO mode (SISO, OPTBC, PMod or V-BLAST). These tasks are split between the gateway and the MT. On the one hand, the MT selects the MIMO mode that achieves the highest throughput given the current SNR. The receiver

needs to compute the effective SNR for all modes to choose the most efficient. This task is performed by the receiver to reduce the feedback load. In addition, the transmitter (gateway) is informed by the MT of the ACK/NACK, the preferred MIMO mode and the effective SNR for that mode, and selects consequently the MCS of the following frame.

In order to formalize the procedure let us define the following matrices. S is the matrix with the spectral efficiency of each combination of MIMO mode - MCS. For example, the element S_{ij} is the spectral efficiency of MIMO mode i and MCS number j. The vector $\overline{\text{SNR}}_{eff}$ represents the CSI and its elements are the effective SNR for each one of the transmission modes. Lastly, the vector with the threshold SNR of each MCS is $\overline{\text{SNR}}_{th}$; note that this threshold is the same for all MIMO modes.

The optimization problem to choose the optimum mode $\hat{T}$ can be written as

$$\hat{T} = i \quad \text{s.t.} \quad R_{ij} = \max_{i,j}\{R_{ij}\}, \tag{9}$$

with the corresponding matrix of rates given by $R_{ij} = S_{ij}[\text{SNR}_{eff,i} \geq \text{SNR}_{th,j}]$. If no MCS verifies the required effective SNR, i.e., all R_{ij} are zero, or in case of a tie, the MIMO mode with the highest effective SNR is chosen and reported back to the gateway.

The choice of the MCS at the gateway follows our previous work in [7], which exploits an adaptive margin. Here, for lack of space, the derivation of the algorithm is not replicated; those interested readers are referred to [6,7]. The gateway selects an MCS m_i using a Lookup Table (LUT), represented by means of a function $\Pi(\cdot)$ which maps SNR intervals to MCSs: $m_i = \Pi(\text{SNR}_{i-d} + c_i)$. All this is shown graphically in the diagram of Fig. 2. The value of the SNR reported back by the MT, SNR_{i-d}, plus a margin c_i is introduced in the LUT for selecting the MCS; this margin is updated when new feedback comes in. The round trip time is assumed to be fixed and equal to the duration of d frames. The recursive equation for updating the value of the margin is

$$c_{i+1} = c_i - \frac{\mu}{\theta^2 + \text{SNR}_{i-2d}^2}(\epsilon_{i-d} - p_0)\theta. \tag{10}$$

This recursion is derived in [6] to solve the following optimization problem

$$\min_c J(c) = \min_c |\mathbb{E}[\epsilon] - p_0|^2. \tag{11}$$

The involved variables are SNR_{i-2d}, the effective SNR used by the LUT to decide the MCS of frame $i - d$, and the acknowledgement ϵ_{i-d} (0 for ACK, 1 for NACK). μ and θ are two constants which take the values 1 and 10, respectively. In this work we suppose the objective FER, p_0, has a fixed value, although there are some studies like [15] which show the benefits of having a variable p_0 in order to maximize the throughput.
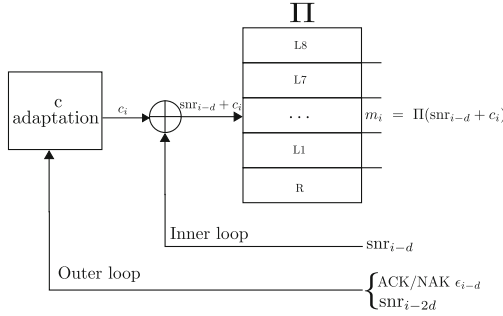
Fig. 2. Block diagram of the MCS selection algorithm with the LUT.

4 Simulation Results

Several simulations are performed to evaluate the spectral efficiency gain from the use of two polarizations simultaneously, and also to understand the potential robustness coming from the use of an adaptive margin in the link adaptation algorithm. All the simulations results presented here are obtained for a maritime scenario with a vessel moving at a constant speed of 50 km/h. The parameters of the channel generator were taken from [14]. The carrier frequency in the simulations is 1.6 GHz, a typical frequency of some Mobile Satellite Systems operating in the L-band. QPSK modulation with frames of 80 ms and 2560 symbols (baud rate of 33600 symbols/s) are used in the physical layer, similarly to the bearer F80T1Q-1B of BGAN [9]. Each simulation comprises the transmission and reception of $M = 60,000$ frames for a specific average SNR, ranging from -5 dB to 25 dB in steps of 2.5 dB. Average spectral efficiency, defined as $1/M \sum_{i=1}^{M}(1-\epsilon_i)r_{m_i}$, and cumulative FER during the whole transmission is also computed for each simulation. In the previous expression r_j is the rate of the jth MCS, m_i is the MCS selected for frame i and ϵ_i is the corresponding ACK. Lastly, the Round-Trip Time (RTT) is set to 7 frames (560 ms) to approximate the feedback delay for GEO satellites. Contrary to other publications which assume instantaneous feedback, here it has a reasonable value and the proposed scheme is robust even with this delay.

The objective of the first set of simulations is to show the benefits of using two orthogonal polarizations simultaneously to serve a mobile user in low bands of the spectrum. Despite the XPD being lower in these bands than in Ku and Ka bands, it provides a significant throughput gain even with the available power split between the two polarizations. In these simulations the SISO case (transmit only in one polarization) is compared with the use of the two polarizations (MIMO) based on the algorithm for choosing dynamically the MIMO mode of Eq. (9). In both cases ACM is employed with a fixed margin $c = -1$ dB in the LUT (see Fig. 2).

The results collected in Table 2 show that use of dual polarization provides a gain higher than 50% for most cases. Interestingly, OPTBC makes it possible

Table 2. Results of the comparison of single polarization (SISO) and adaptive dual polarization (MIMO).

Average SNR	Selected modes	Efficiency (SISO)	Efficiency (MIMO)	Gain
−5 dB	OPTBC	0 bps/Hz	0.68 bps/Hz	N/A
−2.5 to 7.5 dB	PMod	0.68–1.50 bps/Hz	1.02–2.32 bps/Hz	50–55%
10 dB	PMod (30%), BLAST (70%)	1.69 bps/Hz	2.47 bps/Hz	47%
12.5 to 25 dB	BLAST	1.73–1.74 bps/Hz	2.77–3.48 bps/Hz	60–100%

to extend the operation range at low SNR below SISO. As the SNR increases the selected mode turns to be PMod instead of OPTBC with gains of up to 55%. At an average SNR of 10 dB both PMod and V-BLAST are employed and for higher SNRs V-BLAST is mainly used. The V-BLAST MIMO mode, which sends two simultaneous streams of symbols, offers up to 100% gain in terms of throughput, i.e., it can double the capacity.

Next, the margin is allowed to be adaptive in an effort to match a target FER. Figure 3 shows the results obtained in terms of average spectral efficiency and FER. An ACM scheme with a fixed margin of −1 dB is compared with a link adaptation algorithm with adaptive margin for two different values of target FER, 10% and 1%.
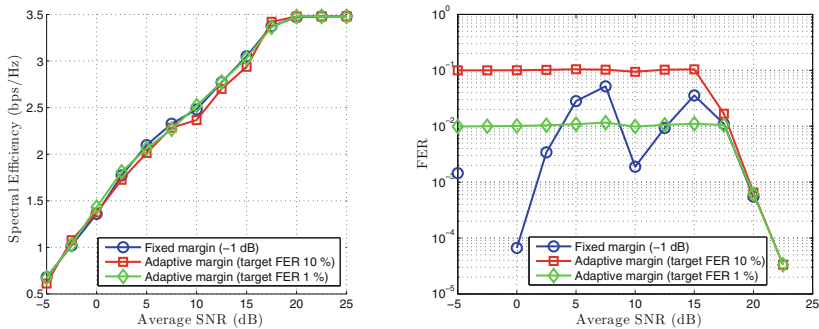
Fig. 3. Average spectral efficiency (left) and FER (right)

In Fig. 3 the behaviour of the algorithm in terms of spectral efficiency and FER for a wide range of average SNRs is illustrated. The spectral efficiency grows linearly with the SNR in dB until it reaches its maximum value of 3.48 bps/Hz. When looking only at the throughput there are not significant differences between using a fixed and an adaptive margin. Furthermore, it is observed how the target FER p_0 influences the final spectral efficiency. For example, the simulations with the target FER 1% have a slightly better efficiency than those with target FER 10%. This dependence of the spectral efficiency with the

target error rate was already studied in [15] where the authors show that there is an optimal FER which maximizes the throughput.

In general, the margin c_i converges to a value and oscillates around it, except for the three higher SNRs where it grows indefinitely. The value of the final margin in each simulation depends on the operation point (LOS SNR), the target FER p_0, the channel conditions and the mobile speed. This last fact is exemplified in [7] where the evolution of the margin in an Intermediate Tree Shadow environment for several MT speeds is shown.

On the other side, the FER behaves very differently when adapting the margin with respect to the fixed case; as shown on the right plot in Fig. 3, a fixed margin cannot guarantee a prescribed FER for different SNR values. As opposed to this, the adaptive algorithm exposed earlier matches the target FER, 0.1 and 0.01 in the example, for a wide range of LOS SNRs. Therefore, the combination of the transmission mode selection and the adaptive margin achieves remarkable throughput gains when compared with the SISO case, whilst guaranteeing a physical layer FER suited to the prescribed QoS.

5   Conclusions and Future Work

The simultaneous use of dual orthogonal circular polarizations in L/S-bands for mobile satellite communication systems is proposed in this paper. The application of MIMO techniques to this satellite scenario, by switching among several MIMO modes, generates significant increments in the spectral efficiency that can get close to 100% for high values of SNR, without using extra bandwidth or power. Moreover, a link adaptation algorithm was proposed to select the MIMO mode and the MCS of each frame based on an adaptive margin. This helps the system to adapt itself to the channel conditions and guarantee a prefixed target Frame Error Rate for a wide range of SNRs. The adaptation scheme is expected to improve through the online computation of the optimal FER which maximizes the spectral efficiency. Additionally, further refinements should be addressed to account for the unavoidable errors in the polarization selection of PMod.

References

1. LTE: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures, ETSI TS 136 213 V8.8.0 (2009–2010)
2. IEEE standard for air interface for broadband wireless access systems. IEEE Std 802.16-2012 (Revision of IEEE Std 802.16-2009), pp. 1–2542 (2012)
3. IEEE standard for information technology-telecommunications and information exchange between systems local and metropolitan area networks-specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007), pp. 1–2793 (2012)
4. Richharia, M.: Mobile Satellite Communications: Principles and Trends, 2nd edn. Wiley, Chichester (2014)
5. Henarejos, P., Perez-Neira, A., Mazzali, N., Mosquera, C.: Advanced signal processing techniques for fixed and mobile satellite communications. In: 8th Advanced Satellite Multimedia Systems Conference and the 14th Signal Processing for Space Communications Workshop (ASMS/SPSC), pp. 1–8, September 2016
6. Rico-Alvariño, A., Arnau, J., Mosquera, C.: Link adaptation in mobile satellite links: schemes for different degrees of CSI knowledge. Int. J. Satell. Commun. Netw. **34**(5), 679–694 (2016). sat.1164. http://dx.doi.org/10.1002/sat.1164
7. Rico-Alvariño, A., Tato, A., Mosquera, C.: Robust adaptive coding and modulation scheme for the mobile satellite forward link. In: IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 530–534, June 2015
8. Tato, A., Mosquera, C., Gomez, I.: Link adaptation in mobile satellite links: field trials results. In: 8th Advanced Satellite Multimedia Systems Conference and the 14th Signal Processing for Space Communications Workshop (ASMS/SPSC), pp. 1–8, September 2016
9. Satellite component of UMTS (S-UMTS): Family SL satellite radio interface, ETSI TS 102 744, October 2015
10. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. IEEE J. Sel. Areas Commun. **16**(8), 1451–1458 (1998)
11. Henarejos, P., Perez-Neira, A.I.: Dual polarized modulation and reception for next generation mobile satellite communications. IEEE Trans. Commun. **63**(10), 3803–3812 (2015)
12. Wolniansky, P.W., Foschini, G.J., Golden, G.D., Valenzuela, R.A.: V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel. In: Proceedings of URSI International Symposium on Signals, Systems, and Electronics (Cat. No.98EX167), pp. 295–300, September 1998
13. Kaltenberger, F., Latif, I., Knopp, R.: On scalability, robustness and accuracy of physical layer abstraction for large-scale system-level evaluations of LTE networks. In: Asilomar Conference on Signals, Systems and Computers, pp. 1644–1648, November 2013
14. Sellathurai, M., Guinand, P., Lodge, J.: Space-time coding in mobile satellite communications using dual-polarized channels. IEEE Trans. Veh. Technol. **55**(1), 188–199 (2006)
15. Park, S., Daniels, R.C., Heath, R.W.: Optimizing the target error rate for link adaptation. In: IEEE Global Communications Conference (GLOBECOM), pp. 1–6, December 2015

Bandwidth Management Using MPLS Model for Future Mobile Wireless Networks

Oba Zubair Mustapha[✉], Ray E. Sheriff, and Felicia L. C. Ong

Faculty of Engineering and Informatics, School of Electrical Engineering and Computer Science,
University of Bradford, Bradford, UK
{O.Z.Mustapha,R.E.Sheriff,F.L.C.Ong}@bradford.ac.uk

Abstract. The recent surge in the development of new technologies, most especially in the field of mobile and wireless communications, requires the adequate maintenance and overall procurement of network infrastructures. This is due to a great deal of accelerating demand from Mobile users having access to real-time information such as data, voice and video services. Therefore, the operators and service providers require seamless integration of network protocols with an improved quality of service (QoS). This paper addresses the performance of multimedia services in Multiprotocol Label Switching (MPLS) nodes and network models design using a simulation approach. MPLS ensures the reliability of the communication minimizing the delays and enhancing the speed of packet transfer. It is valuable in its capability of providing Traffic Engineering (TE) for minimizing the congestion by efficient throughput. The verification of the MPLS model will be the focus of the performance evaluation. An elaborate description of MPLS and its principle of operation will be required. It will eventually address the challenges of packet loss, high latency, high operational cost, more bandwidth utilization, and poor QoS.

Keywords: MPLS network · MPLS-TE · Packet delivery fraction
Multimedia services

1 Introduction

Mobile and Wireless application services form the basis for the success of the technology in the next generation of networks. Therefore, management of the bandwidth has emerged as a powerful platform for controlling the traffic volume of future mobile and wireless networks. A high priority traffic of voice and video has a large proportion of bandwidth consumption in wireless communication systems. Bandwidth is the rate at which data (frame/packet) is transmitted over the network link [1].

Bandwidth management is a generic term that describes the various techniques, technologies, tools and policies employed by an organization to enable the most efficient use of its bandwidth resources [2]. In [3, 4], bandwidth management is defined as a process of allocating bandwidth resources to critical applications on a network.

The key function of bandwidth management is to control the flow of packets on the network links to avoid traffic exceeding the capacity of the network, which would lead

to congestion. This implies that more capacity requires more bandwidth. However, bandwidth is a very limited resource and in most markets is expensive to acquire. Therefore, there is an expectation that demand for data capacity will increase a thousand fold by 2020 [5].

Packets flow would suffer long queuing delays at congested nodes and possibly packet loss if buffers overflow [6]. To solve this problem, managing the available bandwidth would be of benefit to both the users and operators. These, in turn, play a vital role in minimizing the cost of operation rather than demanding for more bandwidth, which will be very expensive. Whilst at the same time used to monitor the effectiveness and efficiency in the performance of the network. The approach stated in [7, 26] is the purpose of supplying bandwidth on a network in order to reserve capacity for users. Nevertheless, the demand is low as compared with the operational capacity of the network.

The aim of this paper is to study the performance of multimedia services in MPLS Network model using a simulation approach. This will eventually proffer a solution to the bandwidth issues of the next generation of Mobile Wireless networks. It will be achievable by proposing design of MPLS node and core networks to manage bandwidth efficiently as possible for the future mobile and wireless networks. Implementation is by performing dynamic and static configuration of the MPLS model network as part of traffic engineering (MPLS-TE). The overall structure of this paper takes form of four sections. Section 1 gives a brief background and related work of the research carried out. Next is the Sect. 2, which entails model design, simulation, and verification of the proposed MPLS technology. Then, Sect. 3 yields the results, and lastly, Sect. 4 concludes the paper.

1.1 Related Work

Distinct bandwidth management techniques have been proposed [6–15]. Research in the area of Multiprotocol Layer Switching (MPLS) technology had been in existence for decades. However, there has been no detailed investigation of using this mechanism for the purpose of bandwidth management to solve the critical problem of delay. In addition, this is a technique that would utilize the available bandwidth to meet the requirement of QoS is required.

Available bandwidth is the maximum throughput of the communication channel without disputing any ongoing flow in the network [16]. This is view according to [17] that many of the bandwidth management techniques proposed have their pros and cons due to its appropriate in a given situation than the other. The bandwidth allocation algorithm proposed in [18] gives the same support. However, the exploration and the proper investigation of other techniques are consider for further works.

George provides a stimulating idea in [2] for the service providers to manage their network efficiently by improving the quality of service to the customer. Further issues is to allocate limited bandwidth with fairness to the users and the application of network management to monitor and control the traffic of multiple applications. Although, there are still many controversial issues yet to be resolved such as increasing network capacity and metered pricing.

Chris and Olov discuss bandwidth management in next generation of packet networks [19]. According to the two authors, there are issues surrounding the bandwidth management for next-generation of voice and multimedia over packet networks. End-to-End QoS requirements for PSTN-grade voice and multimedia service and how it might best support over a packet network infrastructure were investigated. However, the unanswered question of, what amount of bandwidth does each of multimedia services really required?

The authors in [20] proposed Software-based End-to-End Bandwidth Allocation and Reservation for Grid application. The description of the multi-layered architecture that can support the network resource reservation is given. However, there has been little discussion on its interface using networking terms instead of application terms.

In [21], Sanjeev et al. develop ideas of bandwidth management for mobile media delivery. They studied the fundamental problems of mobile broadband networks especially in 3G and 4G technologies such as packet loss and delay. In addition to that, there remains a paucity of performance analysis of the rate control algorithm.

2   MPLS Node and Network Design

Multi-Protocol Label switching (MPLS) is a fast packet forwarding and scalable mechanism that is widely used in the core network [22]. MPLS introduces the Label Switched Path (LSP) tunnel, which provides the mechanism to transport labeled data packets from the source node along the path to the destination node. There are three components in an MPLS network, ingress and egress Label Edge Routers (LER), and Label Switch Routers (LSR). LERs are located at the edges of the MPLS network.

The design of the MPLS node and network models have been implemented as shown in Figs. 2 and 3 respectively. This is in accordance with the configuration of the individual node starting with the application and profile definitions nodes, where the parameters are set out for voice and video. It therefore, seek to obtain statistics collection of the performance behaviour of the model, then running of simulations to view the obtainable results.

The first level of planning design is to create the process model of the peripheral node with define variables, macros, and transitions. A finite state machine (FSM) implements the behaviour of a process model using the states and transitions to determine the actions to be performed [23]. The peripheral nodes proceed to the central hub by point-to-point links, which can be unidirectional, or bidirectional. The main role of the MPLS node model is to simulate packets forwarding from one site node to another site node through packet switching technology. Looking at the nodes shown below in Figs. 1 and 2 indicate the linking of mesh topology using LSRs as the core network. This makes further connection to the sites through LERs both at the ingress and egress points. A source module is another essential aspect of node model, which generate the packets. The processor assign destination addresses to the packets, sends them to a node of the point-to-point transmitter, and retrieve packets arriving from the point-to-point receiver.
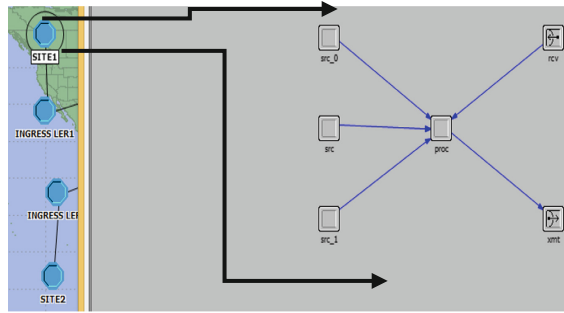
Fig. 1. Source nodes, peripheral nodes, and processor
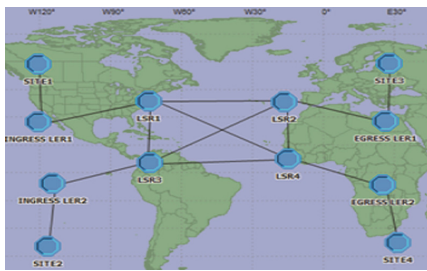
Fig. 2. Nodes model design
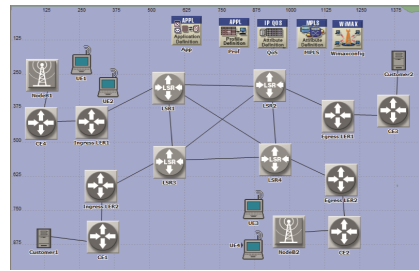
Fig. 3. MPLS network model design

The MPLS network model is shown in Fig. 3, it consists of configuration modules and connectivity of the nodes to generate packet switched data transmission from point-to-point. It is structure to support the availability of resources in order to meet the requirement of the quality of service (QoS). These modules are Application Definition, Profile Definition, IP QoS Attribute Definition, MPLS Attribute Definition and WiMAX. There exist the imposition of the label at the ingress routers LER1 and LER2 called label edge routers, then swapping of the label occurs at the core routers referred to as LSR (label switch routers). Next, is the removal of label and transformation to IP packet format at the egress routers LER1 and LER2 (label edge routers).

Packet Delivery Fraction (PDF). Packet delivery fraction is the ratio of a total number of a data packet received to the total number of data packets sent. This is an estimation of how routing protocol is efficient and effective [24, 25]. The Table 1 shows how the transmission of packets of voice and video conference from the source to the destination. It is justified that a number of packets received are less than the number of packets transmitted. This could be the result of packet drops along the channel of communication. Table 1 illustrates the relationship between packets sent from the source and packets received at the destination using voice and video services.

Table 1. Packet delivery fraction

Multimedia services	Number of packets received (packets/s)	Number of packets sent (packets/s)	Packet delivery fraction
Voice	677	2000	0.3385
Video	14000	19333	0.7241

2.1 MPLS Model Simulation and Verification

OPNET simulator is very useful when working with complex networks with a big number of devices and traffic flows, or in networks where a little change could be critical. Prior to any change in the implementation, it is possible to predict the behaviour and to verify the configurations of the devices [23]. Generally, probability theory and statistics are appropriate for the validation and verification of the network model. As the simulation model gained an improvement, the need to verify and validate the model will be of highly considerable. Verification determines whether the model performs the intended function and meets the required specifications. The fundamental procedure of verification is testing that the OPNET tools and mathematical model are working properly.

Utilization is the ratio of the offered load to the available resource in a given period (instantaneous) [30]. There is provision for Eqs. (1) and (2) as follows:

$$\rho = \frac{\lambda.s}{\varphi} \tag{1}$$

This implies that

$$d = \lambda.s \tag{2}$$

$$\rho = \frac{d}{\varphi}$$

ρ = utilization,
λ = demand for the resource per unit time,
d = demand the resource,
s = waiting or holding time,
φ = supply of the service provided or capacity of the system.

According to [26], utilization is applicable to any resource. If ρ is less than 0.50; this is an under-utilization of the resources. As ρ increases about 0.75, the time spent waiting grows exponentially, asymptotically approaching infinity. If exceeds 1.0, then the number of entities waiting for the resource grows linearly with (d−φ). As indicated below, the mean delay is inversely proportional to the number of packets per second in Eq. (5) while Eq. (6) shows that number of sources is directly proportional to the ratio of number of packets/s and mean service requirement. In Fig. 6, the mean service utilization is inversely proportional to the number of sources. There is variation in the service supplied to the customers at different utilization percentages (25%, 50% and 75%).

Know that 1 byte = 8 bits, let pps = packets/s, bps = bits/s

Therefore, *pps * average_packet_size = bps*

Let the Average rate and utilisation be A_r and ρ respectively

C = buffer service capacity (bits/s)

n = number of sources generating background traffic

$$n = \frac{\rho * C}{A_r} \tag{3}$$

but,

$$C = \frac{\lambda}{\mu}(bits/s)$$

$$\lambda = \frac{no\_of\_packet}{mean\_arrival\_time} = (packets/s) \tag{4}$$

$$Mean\_delay(\bar{w}) = \frac{1}{\mu C - \lambda} = \frac{1}{\dfrac{C}{1/\mu} - \lambda} \tag{5}$$

$$n = \left(\frac{d}{\varphi}\right) * \left(\frac{1}{A_r}\right) * \left(\frac{\lambda}{\mu}\right). \tag{6}$$

3 Results

There is presentation of results of the initial node and process models designed in Fig. 4. There is a constant delay of 0.001 ms and the throughput is increasing rapidly from zero to the level of approximately 16000 packets/s. As for the network designed, the application and profile configurations of voice and video conference are used which yielded results as shown in Figs. 5 and 6 respectively. Subsequently by the discussions of the results obtained from the node and network models. This is likely to be a better communication link between two points or more. As shown in Fig. 4 (left), there is a tremendous increase in the transmission of packets from one end of the site to another end of the site. It is an exponential increment showing a considerable amount of packets transmitted from the source to the destination.
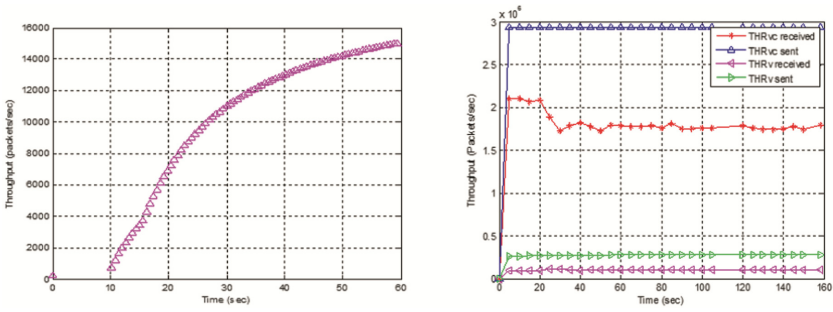
Fig. 4. Throughput of the link for the nodes model *(left)* Throughput sent and received for voice and video *(right)*.
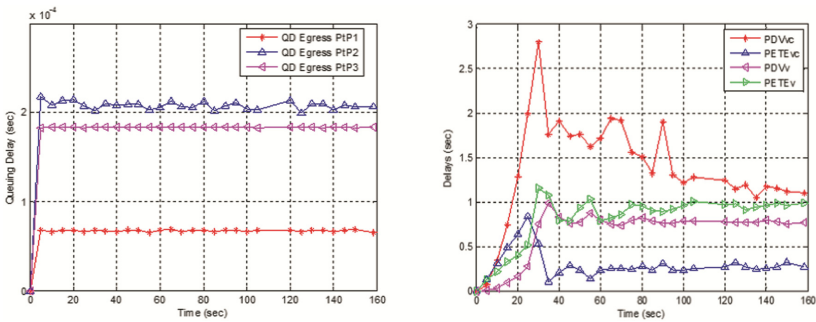
Fig. 5. Queue delay from point-to-point at the egress *(left)*, Packet delay variation and end-to-end delay for voice and video *(right)*.
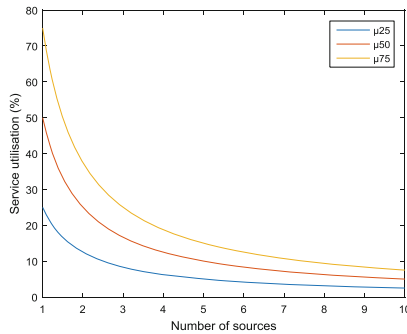
Fig. 6. Service utilization versus number of sources

As can be seen from Fig. 4 (right), the overall throughput is able to stabilize for both voice and video traffics. There appears a wide gap between video traffic (THRvc) sent and video traffic received with the effect of more packets transmitted than packets received. In the case of voice traffic (THRv), a small gap occurs between a number of packets transmitted and packets received which indicates that both have low throughput

within the range of 0 and 500000 packets/s. In general, from Fig. 4 (right), it provides an important opportunity to advance the understanding of the throughput of video conferencing traffic is considerably higher than that of voice traffic due to the level of bandwidth consumption. However, voice transmission is quite easily manageable as compared to video transmission.

Figure 5 (left), the queue delay of point-to-point (PtP) of the different locations of nodes is as shown. The queue delay of PtP1 of 0.08 ms is far lower than that of PtP2 and PtP3 with 2.2 ms and 1.8 ms respectively. In other words, PtP1 has low latency compared to others which imply that the packets will likely move faster in the link with respect to PtP1 whereby enhancing the performance of the network.

The Fig. 5 (right) illustrates the differences between the packet delay variation (jitter) and packet end-to-end delay for both video and voice traffics. As for the packet delay variation, there is an uprising to the peak of about 2.8 s at the initial stage of the video traffic which later drop down by fluctuation to an approximate value of 1.1 s while a remarkable constant delay exists for voice with a minimum of 0 and maximum of about 0.75 s. This is because data rate of the voice is lower than that of the video. In addition, end-to-end delay appears to follow a different pattern in which that of the voice has to reach up to 1.2 s and video has the peak of 0.8 s with little variation.

4   Conclusion and Future Work

In summary, the approach used in this piece of research is similar to that reported in [2], but with different technique and implementation. Therefore, there is need for a thorough study of the performance of the MPLS technology and its implementation that would sustain the future exponential increment in user demand. This observational study of the node model designed are likely to have moderate performance due to low values of end-to-end delay, low queue delay, and high throughput. An implication of the packet delivery fraction is the possibility, which yields a positive performance of the voice packets delivered from the source to the destination due to its low bandwidth consumption as compared with a video packets delivery fraction.

The use of MPLS technology to implement bandwidth management in the future mobile wireless network is reliable and profitable due to its valuable cost to the both operators and service providers. Then it is possible that critical problem of delays such as end-to-end delay, queue delays, and packet delay variation are less likely to occur in the design. However, it will be an additional cost to deploy MPLS technology to the existing network, instead of eliminating existing IP technology. This will enhance the compatibility of the existing facilities.

Performance evaluation of the QoS schemes such FIFO, PQ, and WFQ will be further implementation to assess the services provision to the users. Comparative study of theoretical queue and OPNET queue designs are necessary. More analysis of the MPLS traffic engineering (MPLS-TE) will put into consideration for the next generation of mobile and wireless networks. Further verification, validation, and refinement of the model designed would be required to meet the requirements of the data rates and minimum bandwidth specification for 5G technology.

References

1. Sharma, P., Rathore, V.: Regulating Bandwidth flow estimation and control for Wired/Wireless Networks. Int. J. Soft Comput. Eng. (IJSCE) **1**, 2231–2307 (2012)
2. Ou, G.: Managing broadband networks: a policymaker's guide. ITIF, December 2008
3. Chitanana, L.: Bandwidth management in universities in Zimbabwe: towards a responsible user base through effective policy implementation. Int. J. Educ. Development Using ICT **8**, 62–76 (2012)
4. Kassim, M., Ismail, M., Jumari, K., Yusof, M.I.: A survey: bandwidth management in an IP-based network. World Acad. Sci. Eng. Technol. **62**, 356–363 (2012)
5. Korhonen, J.: Introduction to 4G Mobile Communications. Artech House, Boston (2014)
6. de Veciana, G., Baldick, R.: Resource allocation in multi-service networks via pricing: statistical multiplexing. Comput. Netw. ISDN Syst. **30**, 951–962 (1998)
7. Al-Mosawi, M.A.: Bandwidth estimation and optimisation in rain faded DVB-RCS networks. Ph.D. thesis, University of Portsmouth (2014)
8. Mallapur, J.D., Abidhusain, S., Vastrad, S.S., Katageri, A.C.: Fuzzy based bandwidth management for wireless multimedia networks. In: Das, V.V., et al. (eds.) BAIP 2010. CCIS, vol. 70, pp. 81–90. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12214-9_15
9. Lauwers, J.P.C., Ludwig, L.F.: Network communication bandwidth management. Google Patents (2010)
10. Li, Z.X., Wang, W.-L., Lei, B.-C., Chen, H.-Y.: An approach to bandwidth management based on fuzzy logic. Eng. Sci. **10**, 104–111 (2008)
11. Loh, K.S., LaVigne, B.E., Cavanna, V.V., Thoon, K.Y.: Adaptive bandwidth management systems and methods. Google Patents (2007)
12. Jones Jr., J.K., McLean, S.M., Foley, C.E.: System and method for managing bandwidth utilization. Google Patents (2007)
13. Canova Jr., F.J., Ting, A.H.: Videoconferencing bandwidth management for a handheld computer system and method. Google Patents (2006)
14. Bender, P., Black, P., Grob, M., Padovani, R., Sindhushyana, N., Viterbi, S.: CDMA/HDR: bandwidth efficient high-speed wireless data service for nomadic users. IEEE Commun. Mag. **38**, 70–77 (2000)
15. Al-Majeed, S.S., Hu, C.-L., Nagamalai, D. (eds.): ICCSEA/WiMoA 2011. CCIS, vol. 154. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21153-9
16. Bandung, Y., Langi, A.Z.R., Narendra, A.: Bandwidth Management Technique for Improving Virtual Class in Rural Area Network
17. McGarry, M.P., Maier, M., Reisslein, M.: Ethernet PONs: a survey of dynamic bandwidth allocation (DBA) algorithms. IEEE Commun. Mag. **42**, S8–S15 (2004)
18. Gallon, C., Schelén, O.: Bandwidth management in next generation packet networks. MSF, August 2005
19. Palansuriya, C., Buchli, M., Kavoussanakis, K., Patil, A., Tziouvaras, C., Trew, A., et al.: End-to-end bandwidth allocation and reservation for grid applications, pp. 1–9 (2006)
20. Mehrotra, S., Chen, H., Jain, S., Li, J., Li, B., Chen, M.: Bandwidth management for mobile media delivery, pp. 1901–1907 (2012)
21. Holness, F., Phillips, C.: Congestion control mechanism for traffic engineering within MPLS networks. In: Rao, S., Sletta, K.I. (eds.) INTERWORKING 2000. LNCS, vol. 1938, pp. 254–263. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-40019-2_22
22. Rosen, E., Viswanathan, A., Callon, R.: Multiprotocol label switching architecture. IETF RFC 3031 (Proposed Standard), January 2001
23. Optimisation Network Tools (OPNET). http://www.opnet.com. Accessed 1 Mar 2016

24. Srikanth, T., Narsimha, V.B.: Simulation-based approach to performance study of routing protocols in MANETs and ad-hoc Networks. IJCSNS Int. J. Comput. Sci. Netw. Secur. **11**, 111–115 (2011)
25. Mammeri, Z. (ed.): WMNC 2008. IIFIP, vol. 284. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-84839-6
26. Mitola, J.: Software Radio Architecture: Object-Oriented Approaches to Wireless Systems Engineering, vol. 1. Wiley, New York (2000)

IWUAS Workshop Session

A Survey on Network Architectures and Applications for Nanosat and UAV Swarms

Manlio Bacco[1], Pietro Cassarà[1], Marco Colucci[1], Alberto Gotta[1],
Mario Marchese[2], and Fabio Patrone[2(✉)]

[1] Institute of Information Science and Technologies (ISTI)
of National Research Council (CNR), via Moruzzi, 1, Pisa, Italy
{manlio.bacco,pietro.cassara,marco.colucci,alberto.gotta}@isti.cnr.it
[2] Satellite Communications and Networking Laboratory (SCNL)
of the University of Genoa, via all'Opera Pia, 13, Genoa, Italy
mario.marchese@unige.it, f.patrone@edu.unige.it

Abstract. Nanosatellites and unmanned aerial vehicles are attracting more and more the interest of both industrial and research fields. They are low-cost and easy deployable items, therefore their use is expected to quickly grow in the next few years. This work proposes a survey on the network architectures and the applications for nanosatellite swarms and constellations, as well as for flying ad hoc networks, by characterizing distinctive features and issues yet to be resolved in order to take advantage from both technologies in a joint fashion.

Keywords: UAV · FANET · Nanosatellite
Network architecture and application

1 Introduction

The use of Unmanned Aerial Vehicles (UAVs) and nanosatellites (nSATs) is increasingly common. They provide low-cost support for a large class of applications, making their use appealing for research and market operators. Both of them can be used individually, or in the form of a swarm; when dealing with nSATs, also constellation topologies are possible. While their use as single objects is sufficient in several application fields, it may represent a limitation in others. Using multiple UAVs or nSATs needs coordination and data exchange services among them and with one or more Ground Control Stations (GCSs), thus requiring more complex network architectures. In the last few years, the use of UAVs swarms, referred to as Flying Ad-Hoc Networks (FANETs) in the literature [1], is becoming increasingly of interest. Anyway, their diffusion is limited by the complexity of Command and Control (C2) operations. A FANET is characterized by distinctive features, typically absent in other networks: a high mobility degree, an average and peak movement speed that can challenge the

effectiveness of the communication. Nowadays, those peculiarities still present some challenging research issues to be faced.

Moving to applications, the concept of *servgoods* is described in [2]. Autonomous vehicles, such as UAVs, can be *enveloped* with a service-oriented layer, in order to make the vehicles (more generally, any goods) smarter and more adaptable to particular uses: those entities are defined as servgoods. Technologies like nSATs and UAVs can be considered as the servgoods of the future, being able to sense the environment, to process collected data, to react to events, and to learn from past experiences. Therefore, an overall framework is needed in order to properly deal with such a complex ecosystem, starting from technological and architectural considerations, and then considering privacy, security and liability, as well. In this work, we analyze network architectures and applications in the literature, providing a survey on these topics. Furthermore, we present a preliminary investigation on opportunities and challenges when jointly using nSATs and UAVs.

The rest of this paper is organized as follows: Sect. 2 provides an overview of the state of the art. Section 3 deepens the investigation, by providing technical details on the network architectures and on the applications for FANETs and for nSATs. Particular attention to joint solutions is payed in Sect. 3.3. The conclusions are in Sect. 4, opening to the future work still needed for a joint use of UAVs and nSATs.

2   Related Works

The use of UAVs is becoming very common in several civilian application fields. From time to time, UAVs are used jointly with satellites, in order to exploit the advantages that they can provide. A typical use case is related to disaster scenarios, where satellites can provide services of damage assessment, and UAVs can be used for a closer assessment and for relief actions. For instance, the works in [1,3] show how the use of a FANET or a nSAT constellation can provide communication and remote sensing services, respectively, in a low-cost and fast deploying way, with acceptable accuracy. Apart from disaster scenarios, a wide range of applications can also benefit from using UAVs, and several ones are described in [4–6], such as power lines inspection, monitoring of cultural heritage sites, environmental monitoring, fire and gas detection, as well as precision agriculture. Those scenarios have in common Machine-to-Machine (M2M) traffic profiles in the large majority of the use cases under consideration, and the use of satellites [7] is quite mature when dealing with such a traffic, thus opening to joint uses. In particular, precision agriculture is largely benefiting of the use of UAVs [8], due to low operational costs, high operational flexibility, and high spatial resolution of imagery. The authors show that adopting UAVs is advantageous for small areas, and that a break-even point exists at approximately five hectares; above such a threshold, airborne and then satellite technologies have lower imagery costs. Anyway, the use of nSATs is not investigated in [8], and a system architecture including both UAVs and nSATs, as preliminary discussed

in this work, may represent a real breakthrough for the investments in this fast rising field. If C2 is considered, a GCS can be used to control one or more UAVs via Non Line of Sight (NLoS) and Beyond Line of Sight (BLoS) nSAT links [9]. In fact, if UAVs can be employed to aid communications and extend coverage, by providing relaying, data dissemination and collection services, the covered area can be extended from single-hop scenarios to multi-hop scenarios by using nSATs. In the former case, one or more UAVs are controlled from a GCS, while, in the latter, the use of nSATs can largely extend the coverage by providing intermediate hops.

The use of nSATs can provide several advantages, from low cost and lower delivery delays to fast deploying operations w.r.t. to the use of satellites, thus opening to the possibility of having up-to-date orbiting technology at any time and making this market segment more and more attractive.

3 Network Architectures and Application Fields

Sections 3.1 and 3.2 present networks architectures for nSATs and for FANETs, respectively, highlighting the most relevant features of both. In Sect. 3.3, we identify the most common application scenarios and also discuss how a joint architecture can be profitable in the upcoming future.

3.1 Nanosatellites

About sixty years ago, the first satellite launches took place. Since then, the number of launches has exploded, thanks to the several mission goals that can be accomplished by satellites, such as weather monitoring, disaster prevention, space and Earth observation, and telecommunications [10]. However, the build and launch process of a satellite is extremely expensive (about \$150–\$200 million for a Low Earth Orbit (LEO) satellite and \$300 million for a Geostationary Earth Orbit (GEO) satellite). Such high costs have prevented the access to space to small and medium-sized businesses for a long time. Nowadays, thanks to Micro-Electronics (MEs) and Micro-Systems Technologies (MSTs), the satellite hardware components are decreasing in size, both primary and payload ones [11]. MSTs can provide smaller objects, power savings, and increased robustness. Currently, it is possible to embed all the necessary systems in a single object that weights just a few kilograms (instead of a few tons), which is called *nanosatellite*.

Since 2000, more than 80 universities and several emerging nations have developed programs that provide the realization and launch of nSATs for different purposes [12]. These programs may involve a single or a group of nSATs which can be launched at the same time as secondary payloads of bigger satellite launches. They can constitute a swarm (see Fig. 1a) or a constellation (see Fig. 1a), depending on the deployment strategy. In a *swarm*, all satellites are quite close to each other [13] since they are rapidly deployed one after the other. In a *constellation*, nSATs are equally spaced in the chosen orbital plane (or planes in case of multi-orbit constellations) [14]. Their deployment is sequential

(a) nSAT swarm   (b) nSAT constellation

Fig. 1. Logical representations of common nSAT topologies.

and highly synchronized. In both cases, the use of a set of nSATs leads to some advantages: for instance, in [15] the data gathering, processing and transmission functions towards Ground Stations (GSs) are distributed throughout the whole swarm. The limited resources can be better exploited by sharing the computing power and employing data exchange through Inter-Satellite Links (ISLs). Communication latency decreases thanks to the higher amount of contacts between GSs and nSATs, especially in constellations where these contacts are spread during all the day, which also leads to a considerable improvement in throughput. The employment of more than one nSAT allows achieving a larger *footprint* (area on the Earth's surface covered by nSATs) and providing a higher fault tolerance. Nowadays, there are hundreds of on-going projects which involve nSATs. Thousands of these objects are in orbit and still active, as reported in the online *Nanosatellite Database* at www.nanosats.eu. The most relevant features of possible nSAT network configurations under consideration in this work, i.e., single, swarm, and constellation, are summarized in Table 1.

3.2   Flying Ad-Hoc Networks

In a FANET, the nodes cooperate exchanging data among them, and this can present some challenges: in fact, UAVs can move at high speeds, thus introducing Doppler effects when communicating with GSs. Furthermore, an UAV swarm is generally scattered in space, so that the distance among them can limit the effectiveness of communications. A FANET is controlled from the ground by using a GCS. For the sake of simplicity, we assume that the GCS is also the entity collecting user data[1] The connectivity among UAVs and GCS is of primary importance, especially in the case of C2 links, and should guide in the design of network architectures.

We now describe the most common architectures for FANETs in the literature, which can be seen in Fig. 2. Figure 2a shows one of the simplest architectures for a FANET: each node communicates directly with the GCS. Nodes can move within GCS radio coverage (Line of Sight (LoS) communications). UAV-to-UAV communications suffer a potentially large delay because data need to be routed through GCS. An alternative network architecture relies on the use of fixed terrestrial infrastructure, such as scenarios involving cellular networks [16],

---

[1] C2 and data links should be different physical links for safety reasons.

Table 1. Most relevant features of different nSAT network configurations.

	Single	Swarm	Constellation
Communication latency	**high:** data exchange when nSAT is in the communication range of GSs	**high:** data exchange when nSATs are in the communication range of GSs, and among close nSATs	**low:** data exchange when nSATs are in the communication range of GSs, and among spread nSATs
Fault tolerance	**low:** single nSAT (no backup)	**high:** multiple close nSATs, thus redundant services	**moderate:** widely spread nSATs with redundant services
Throughput	**low:** few contacts between an SAT and each GS per day	**moderate:** few contacts between each nSAT and each GS per day, but high number of overall contacts between nSATs and GSs per day	**high:** high number of overall contacts between nSATs and GSs per day
Available resources	**low:** limitations on on-board HW/SW components: size and weight, computational power, available energy, storage capacity	**high:** each nSAT shares its available resources with close members	**moderate:** each nSAT shares its available resources with other members, but with larger delays than those in swarms
Energy consumption (per nSAT)	**moderate:** a nSAT performs both data collection and data exchange operations with GSs	**low:** several nSATs perform data collection and data exchange within the swarm, while others perform data exchange with GSs	**high:** all nSATs perform data collection and data exchange within the constellation and with GSs
Coverage	**low:** single footprint	**moderate:** several footprints widely overlapping in a small area	**high:** several footprints slightly overlapping in a vast area
Cost	**low:** unitary production and single launch costs	**moderate:** multiple nSATs production and single launch costs	**high:** multiple nSATs production and multiple launch costs

shown in Fig. 2c. Base stations (BSs) can be used to support both UAV-to-UAV and UAV-to-GCS communications. This architecture has some drawbacks: the installation of new BSs for a larger coverage is expensive, and the already existing infrastructure is not designed for air-to-ground communications: thus, high-altitude UAVs may experience a really poor link quality. In addition, each UAV must be within the communication range of at least one BS, which is unlikely to happen in rural areas, thus limiting the use of such an architecture.

In order to overcome the limitations due to LoS communications, NLoS scenarios may be taken into account by relying on satellites [17] or on nSATs. In terms of coverage, both the centralized and the cellular-like architecture may benefit from the use of satellites (see Figs. 2b and d). In the former case, UAV-to-UAV communications are affected by an even larger propagation delay, especially in the presence of GEO satellites. Despite the larger coverage, a satellite-based architecture introduces different design challenges. Propagation delay, fading attenuation and error-prone wireless links must be taken into account, especially in the case of C2. The last solution relies on the definition of a UAV ad-hoc network [18,19]. Each UAV participates in the data forwarding, removing the need for any infrastructure. Within the swarm, one node acts as Cluster Head
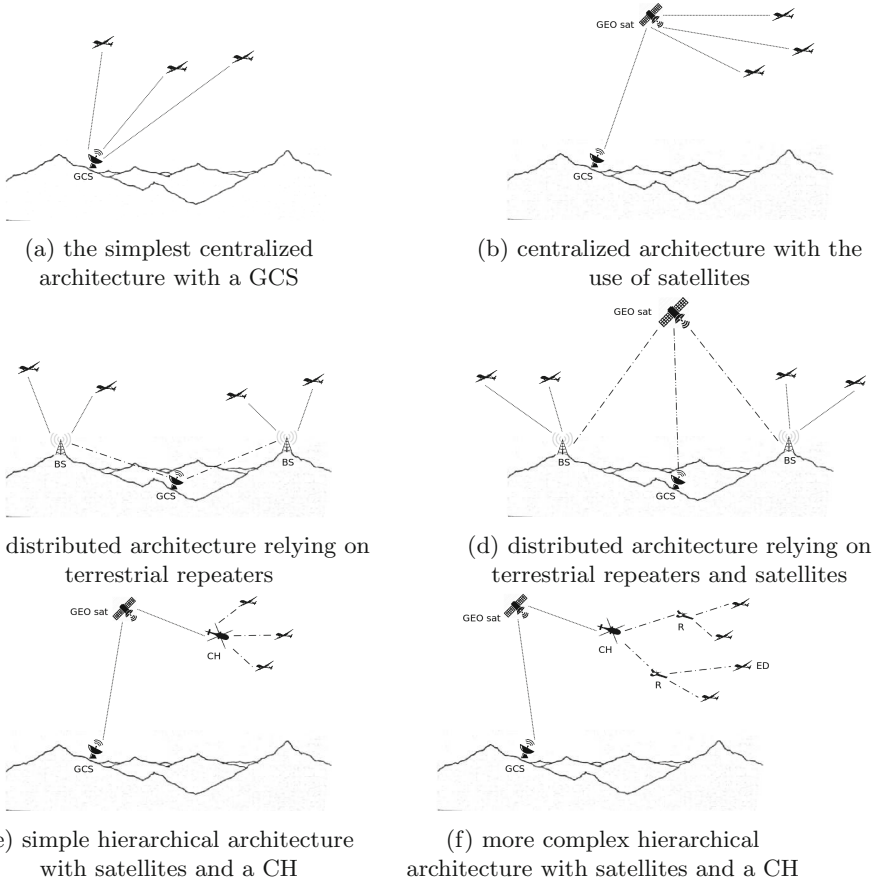
(a) the simplest centralized architecture with a GCS

(b) centralized architecture with the use of satellites

(c) distributed architecture relying on terrestrial repeaters

(d) distributed architecture relying on terrestrial repeaters and satellites

(e) simple hierarchical architecture with satellites and a CH

(f) more complex hierarchical architecture with satellites and a CH

Fig. 2. Network architectures for FANETs.

(CH) and is in charge of forwarding data between nodes and GCS via satellite (see Fig. 2e), while the other members act as slave members. We assume that the CH is able to carry a larger payload and has more available energy than the slave members. The CH needs at least two radio interfaces: one for local transmissions and one for remote transmissions via satellite. Within the swarm, IEEE 802.11-based communications are typically assumed in the literature [20]. In Fig. 2e, a star topology is proposed, namely *simple hierarchical architecture*, thus each UAV is connected only to the CH, which is connected to the GCS via satellite. The main weakness of the last architecture is the lack of robustness: if the CH fails, the entire network is compromised, thus backup CH are required to improve the robustness. In order to overcome such limitations, a possible alternative, namely *complex hierarchical architecture*, consists in relying on a hierarchical network architecture [21], as proposed in Fig. 2f. In the latter, three entities can be recognized: the CH, the routers (Rs) and the end-devices (EDs),

corresponding to three different classes[2] of UAVs. EDs (small class) are connected to the closest router (medium/large class), which in turns is connected to one or more close routers. Each router stores the list of connected EDs and acts as a forwarder for the messages of connected EDs. Each router is equipped with a satellite communication module. The CH acts as *primary router*, with exclusive access to the satellite channel, while the other ones (*secondary routers*) cannot access it: in case of failure of the CH, a secondary router is elected as new CH, thus providing fault-tolerance. The presence of multiple routers also improves the spatial coverage, allowing for multi-hop communications. Several issues must be taken into account with such a hierarchical architecture: UAVs must be able to detect the CH failure, and a CH election algorithm must be designed and implemented, as well as a data synchronization protocol among CH and secondary routers (backup CHs).

Table 2 summarizes some of the most relevant features characterizing the aforementioned network architectures. In particular, we compare the hierarchical architectures with the centralized and cellular-like ones.

Table 2. Comparison among network architectures for FANETs.

	Centralized	Cellular-like	Hierarchical (simple)	Hierarchical (complex)
Communication latency	**Low/medium** propagation delay (typically LoS)		**High** propagation delay (BLoS, NLoS)	
Fault tolerance	**Very limited:** centralized solution	Roughly **proportional** to the number of BSs	**Very limited:** centralized solution	**Proportional** to the number of routers
Scalability and performance issues	**Limited** by the number of UAVs contemporarily controllable by GCS via a single LoS link	**Moderate** scalability due to the infrastructure; larger delay w.r.t. the centralized solution	**High** scalability due to the hierarchical architecture, but a single CH may represent a bottleneck in case of high traffic rates	
Coverage	**Very limited:** UAVs must move within radio coverage of GCS	Roughly **proportional** to the number of BSs	**Limited:** each UAV must fly within radio coverage of the CH	**Scalable:** the larger the number of Rs, the wider the covered area
Cost	**Roughly proportional** to the number and class of UAVs to be deployed: small and low-cost UAVs carry more limited payloads than larger UAVs			
	Low: single LoS link	**Low:** use of existing BSs	**High** cost due to the additional HW/SW modules to be installed on each router	
Energy consumption	**Limited/moderate** power consumption: UAVs directly communicate with a GCS and operate independently, thus energy-saving mechanisms can be adopted		**High** power consumption: routers always active for traffic forwarding	

---

[2] In this work, the class of an UAV describes the amount of available resources on it, such as energy or computational power: larger classes have more available resources than smaller classes.

3.3   Joint Architectures and Application Fields

As we summarize in Table 3, the combined use of a FANET and of nSATs can enrich the available services in different application scenarios. We consider three reference scenarios: search and rescue, surveillance and monitoring, and goods delivery. For instance, in a Search And Rescue (SAR) scenario in case of large wildfires, the use of both solutions can provide, at the same time, an overview of the whole situation from nSATs and a closer look from UAVs. The latter can be also used to follow operators or civilians in danger, and to timely deliver medical supplies, for instance, while the former help in assessing the whole situation in order to support real-time rescue operations and the decision process. UAVs can be seen as a *mobile extension* of the footprint of a nSAT, a sort of *additional logical beam*. One of the most relevant advantages of multi-UAV systems is the coverage: the larger is the number of UAVs, the wider is the covered area, especially in case of a hierarchical network architecture. According to the application domain, additional components can be installed on-board of UAVs: satellite/radio communication modules, high-resolution cameras, and chemical detectors are just some examples.

Mission requirements and typical performance indicators [22] for the three reference scenarios under consideration are reported in Table 4.

A key issue for both nSATs and UAVs is the energy consumption: while the former ones are equipped with solar panels for battery recharging in order to ensure proper functioning of on-board systems at all time, the latter ones land

Table 3. Assessment of the benefits of an integrated platform composed of nSATs and UAVs in three reference scenarios.

	Safety	Target identification	UAV preservation	Integration between UAVs and nSATs
Search and rescue	Strong safety requirements, for instance in urban scenarios where buildings, trees and other obstacles can block the operations or the swarm can itself be a danger	nSATs can provide information on the target and UAVs can confirm the identification through on-board equipment	nSATs can provide information on buildings, trees, and obstacles in order to avoid collisions with UAVs	nSATs can extend the capabilities of UAV swarms by helping to identify targets in vast areas and by providing map information
Surveillance and monitoring	nSATs can quickly notify any events of interest, while UAVs can provide close details when deployed	nSATs can search a vast area for target(s). UAVs can provide actual identification and *follow me* services	Need of a continuous estimation of the residual energy of UAVs	nSATs can strongly extend the capabilities of UAV swarms in vast areas, while UAVs can provide on-demand services closer to ground
Goods delivery	nSATs can provide (quasi) real-time information on the delivery area/target to assist an UAV during a safe items delivery	nSATs continuously track the position of a (mobile) target avoiding failed UAV-assisted deliveries	nSATs can confirm the operativeness of deployed UAV and the position of goods	nSATs and UAV swarms can cooperate in perform challenging deliveries, for instance to mobile/maritime destinations

Table 4. Mission requirements and performance indicators in three reference scenarios.

	Mission requirements	Performance indicators
Search and rescue	*Time-critical:* hazards and/or victims must be timely identified	*Response time* (time between target identification and rescue operations)
Surveillance and monitoring	*Target identification and tracking*: the target must be correctly identified and tracked	*Identification delay and reaction time*: rapid target identification and prompt reaction
Goods delivery	*Goods tracking and safety*: goods position and integrity must be known	*Delivery time and reliability*: goods should arrive as soon as possible in a consistent state

when the mission is completed or the available energy is almost depleted. Joint scenarios require a policy for the overall energy management, in order to ensure a working system at each time. Strategies and policies to deal with the latter are left to future studies.

4   Conclusions

In this work, we describe feasible network architectures for nanosatellites and FANETs, pointing out the potential strengths and weaknesses of each considered solution. We identified some plausible application scenarios involving the combined use of UAVs and nSATs, thus preliminarily discussing the advantages of a hybrid FANET-nSAT architecture. Low cost and reduced propagation delay are some of the advantages that make a nSAT-based solution appealing w.r.t to the use of LEO/GEO satellites. For instance, C2 links require reliability, and low propagation delays: the latter can be fulfilled by using nSATs in place of LEO/GEO satellites when BLoS/NLoS scenarios are considered. Several limitations, such as limited bandwidth, absence of fault-tolerance, unreliability, lack of coverage, and energy issues must be taken into account, thus requiring further investigations on both architectural and performance aspects of a joint architecture.

References

1. Bekmezci, I., Sahingoz, O.K., Temel, Ş.: Flying ad-hoc networks (FANETs): a survey. Ad Hoc Netw. **11**(3), 1254–1270 (2013)
2. Tien, J.M.: The sputnik of servgoods: autonomous vehicles. J. Syst. Sci. Syst. Eng. **26**, 1–30 (2017)
3. Nasseri, S.A., Asrar, F.M.: An integrated space-based solution in support of public health during disaster relief. In: 2015 7th International Conference on Recent Advances in Space Technologies (RAST), pp. 531–537. IEEE (2015)

4. Bacco, M., Ferro, E., Gotta, A.: UAVs in WSNs for agricultural applications: an analysis of the two-ray radio propagation model. In: IEEE SENSORS 2014 Proceedings, pp. 130–133. IEEE (2014)

5. Daponte, P., De Vito, L., Mazzilli, G., Picariello, F., Rapuano, S., Riccio, M.: Metrology for drone and drone for metrology: measurement systems on small civilian drones. In: Metrology for Aerospace (MetroAeroSpace), pp. 306–311. IEEE (2015)

6. Bacco, M., Caviglione, L., Gotta, A.: Satellites, UAVs, vehicles and sensors for an integrated delay tolerant ad hoc network. In: Bisio, I. (ed.) PSATS 2016. LNICST, vol. 148, pp. 114–122. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47081-8_11

7. Bacco, M., De Cola, T., Giambene, G., Gotta, A.: Advances on elastic traffic via M2M satellite user terminals. In: 2015 International Symposium on Wireless Communication Systems (ISWCS), pp. 226–230. IEEE (2015)

8. Matese, A., Toscano, P., Di Gennaro, S.F., Genesio, L., Vaccari, F.P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., Gioli, B.: Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. Remote Sens. **7**(3), 2971–2990 (2015)

9. Zeng, Y., Zhang, R., Lim, T.J.: Wireless communications with unmanned aerial vehicles: opportunities and challenges. IEEE Commun. Mag. **54**(5), 36–42 (2016)

10. Woellert, K., Ehrenfreund, P., Ricco, A.J., Hertzfeld, H.: Cubesats: cost-effective science and technology platforms for emerging and developing nations. Adv. Space Res. **47**(4), 663–684 (2011)

11. Gill, E., Monna, G., Scherpen, J., Verhoeven, C.: MISAT: designing a series of powerful small satellites based upon micro systems technology. In: 58th International Astronautical Congress, vol. 6, pp. 1–6 (2007)

12. Bouwmeester, J., Guo, J.: Survey of worldwide pico-and nanosatellite missions, distributions and subsystem technology. Acta Astronaut. **67**(7), 854–862 (2010)

13. Felicetti, L., Santoni, F.: Nanosatellite swarm missions in low earth orbit using laser propulsion. Aerosp. Sci. Technol. **27**(1), 179–187 (2013)

14. Gill, E., Sundaramoorthy, P., Bouwmeester, J., Zandbergen, B., Reinhard, R.: Formation flying within a constellation of nano-satellites: the QB50 mission. Acta Astronaut. **82**(1), 110–117 (2013)

15. Budianu, A., Rajan, R., Engelen, S., Meijerink, A., Verhoeven, C., Bentum, M.: OLFAR: adaptive topology for satellite swarms. In: International Astronautical Congress, pp. 3–7 (2011)

16. Frew, E.W., Brown, T.X.: Networking issues for small unmanned aircraft systems. J. Intell. Robot. Syst. **54**(1–3), 21–37 (2009)

17. Frew, E.W., Brown, T.: Airborne communication networks for small unmanned aircraft systems. In: Proceedings of the IEEE, vol. 96 (2008)

18. Bekmezci, I., Sahingoz, O.K., Temel, C.: Flying ad-hoc networks (FANETs). Ad Hoc Netw. **11**(3), 1254–1270 (2013)

19. Lav Gupta, R.J., Vaszkun, G.: Survey of important issues in UAV communication networks. IEEE Commun. Surv. Tutorials **18**, 1123–1152 (2016)

20. Ilker Bekmezci, I.S., Erkalkan, E.: Flying ad hoc networks (FANET) test bed implementation. In: 7th International Conference on Recent Advances in Space Technologies (RAST). IEEE (2015)

21. Sahingoz, O.K.: Mobile networking with UAVs: opportunities and challenges. In: International Conference on Unmanned Aircraft Systems (ICUAS) (2013)
22. Hayat, S., Yanmaz, E., Muzaffar, R.: Survey on unmanned aerial vehicle networks for civil applications: a communications viewpoint. IEEE Commun. Surv. Tutorials **18**(4), 2624–2661 (2016)

Command and Control of UAV Swarms via Satellite

Pietro Cassarà[✉], Marco Colucci, and Alberto Gotta

Institute of Information Science and Technologies (ISTI) of National Research
Council (CNR), via Moruzzi, 1, Pisa, Italy
{pietro.cassara,marco.colucci,alberto.gotta}@isti.cnr.it

Abstract. Unmanned Aerial Vehicles (UAVs) are attracting an increasing interest from both the industrial and the research fields, because of the large number of scenarios and applications that they can support. One of the big challenges of the next future is the use of UAV swarms, in order to exploit the advantages that coordinated actions of multiple drones can provide. In this work, we propose an analytical framework to evaluate the probability of a reliable command and control message delivery from a Ground Control Station to a UAV swarm via satellite, also exploiting intra-swarm gossiping.

1 Introduction

Nowadays, UAVs, or drones, are attracting a lot of attention from industrial and research fields. They are suited to a large number of applications, from military to civilian ones, thus the advantages that they can provide are eagerly exploited by several actors. UAVs have been largely used in the military field in the past, but, nowadays, the cost reduction makes them of interest also in several civilian fields: for instance, in the precision agriculture [1,2], in the surveillance field or environmental monitoring [3], and in search and rescue applications [4]. The use of UAVs has proved to be particularly effective in otherwise impervious areas, or each time their use can remove the need for expensive temporary scaffolding, such as in the case of the inspection of historical or cultural areas and buildings. In the latter scenarios, UAVs are typically equipped with the needed sensors in order to facilitate an inspection: for instance, cameras, but also short-range communication radios, in order to collect data from previously installed sensors or to deliver commands, in the case of actuators. Up to now, the largest fraction of civilian applications is based on the use of Line of Sight (LoS) communications, so that the operator remotely controlling the drone can avoid any close obstacles. In fact, a strict regulation is quickly spreading in several European countries, in order to control the use of these devices, mainly in areas where poor experienced personnel can improperly use UAVs, such as close to airports or in the presence of crowd, with possibly disastrous consequences. When considering Non Line of Sight (NLoS) or Beyond Line of Sight (BLoS) communications, the use of the satellites is a possibility, but the following limiting factors should be taken into

account: a larger delay than in LoS communications; the absence of a direct visual feedback; the need of always available bandwidth, in order to control the drone and to collect data; furthermore, the availability of automatic collision avoidance systems to compensate the operators' maneuvering delay. The aforementioned requirements make more expensive the design, the manufacturing, and the use of these devices. In several contexts, the use of a single UAV can be a limiting factor; for instance, in search and rescue applications, if several drones can be rapidly deployed, the probability of a successfully rescue mission may increase. The use of UAV swarms is of interest in several fields, if the task previously assigned to a single drone can be parceled and parallelized. A number of advantages are provided by the use of UAV swarms, as pointed out in [5]: (*i*) likely, the overall cost of acquisition and maintenance of several small Commercial Off-the-Shelf (COTS) UAVs is lower than the overall cost of a single large UAV; (*ii*) scalability, which is a key feature of UAV swarms, instead absent in single UAV missions; (*iii*) fault-tolerance, because a single malfunction has a limited impact on the swarm; (*iv*) faster operations, thanks to the parallelization of the work.

The scenario under consideration in this work is built upon UAV swarms remotely controlled via satellite. Although some works in the literature deal with the use of the drone swarms [6–8], the issues posed by Command and Control (C2) via satellite require further investigations. The main contribution of this work is providing an analytical framework to estimate both coverage probability and delivery delay, when an UAV swarm receives C2 data via satellite, which can be further forwarded (*gossiped*) inside the UAV swarm, in order to increase the probability of a reliable data delivery. The rest of the paper is structured as follows: Sect. 2 provides some background and discusses the related works. Section 3 deals with the description of the problem and of the analytical model needed to address it. Section 4 provides some preliminary numerical results; the conclusions are in Sect. 5.

2   Related Works

Several works can be identified in the literature on Flying Ad-Hoc Networks (FANETs), mainly focusing on the communications within the swarm and on the issues posed by the communications with a terrestrial station. The survey in [5] provides a very valuable overview of both the issues and the advantages provided by FANETs. The communication link quality within a swarm exhibits a complex behavior that depends on several factors: the distance between any couple of nodes, the shadowing due to the UAV itself, the drone attitude, and the environmental conditions. The last ones play a major role in the link quality, if UAVs fly above or below the clouds. Typically, small UAVs fly below the clouds, therefore the rain fading can impair the communications. Furthermore, because of the wind, small UAVs can frequently change their attitude, thus modifying the relative orientation on the pitch, roll and yaw angles; this impacts on the link quality, as well, because the power loss due to the antenna mismatch can be sometimes severe [5].

In [6], four basic communication architectures for Unmanned Aerial Systems (UASs) are discussed: direct link, satellite, cellular, and mesh networking. Satellite-based and mesh networks may be the most promising solutions. In a mesh network, each node acts as a relay to forward data, thus a control station can be reached via several intermediate hops. Anyway, it requires a path, i.e., nodes in place, in order to work, thus it can be feasible only in areas with a large nodes' density. The use of satellites can provide a better coverage than the use of the direct links, so that the UAV swarm remains well connected. The typical limited bandwidth in satellite links does not really pose here an issue, because C2 protocols should not require large amount of available bandwidth. On the other hand, if user data were to be delivered, larger bandwidth may be required to meet the requirements of high data rate applications. Geostationary Orbit (GEO) and Low Earth Orbit (LEO) satellites can be employed; if considered, a large delay should be taken into account in the former case, while temporary disconnections are expected in the latter case. Despite the challenges characterizing satellite-controlled UAV systems (especially for civilian purposes), research and industrial communities are still investigating the feasibility of the introduction of UAVs in non-segregated airspace. The DeSIRE project[1] is aimed at demonstrating maritime surveillance services using Remotely Piloted Aircraft System (RPAS), by exploiting BLoS communications.

3   System Model

In the scenario under consideration, visible in Fig. 1a, a Ground Control Station (GCS) transmits C2 messages, each composed of k control blocks, towards a swarm composed of n drones. Therefore, we assume that each C2 message (for instance, new navigation data) is split in k fragments that must be successfully received. An average loss probability PL_{SAT} is considered, in order to take into account possible impairments on the satellite channel. Therefore, a successful
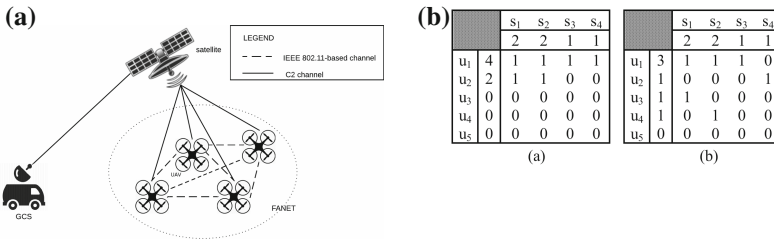

Fig. 1. (a) The scenario under consideration in this work. (b) Examples of received blocks per UAV in the swarm, when $h = 6$ (sum of the margins), $k = 4$ columns (number of blocks), and $n = 5$ rows (number of UAVs).

---

[1] DeSIRE stands for *Demonstration of Satellites enabling the Insertion of RPAS in Europe*, a joint ESA-EDA initiative.

transmission probability P_{SAT} of k blocks (i.e., a single C2 message) towards each node of the swarm can be written as $P_{SAT} = (1-PL_{SAT})^k$. In the scenario under consideration (see Fig. 1a), the UAVs cooperate in propagating C2 data, in order to increase the probability that each UAV in the swarm can correctly decodes C2 messages. If an UAV has correctly received the C2 message (i.e., all k blocks), then it can *gossip* them to the neighbors. We call *gossiping user* an UAV that forwards data within the swarm. For the sake of clarity, all drones (neighbors) are potentially gossiping users. The gossiped blocks can be lost with an average loss probability $PL_{FAN} = PL_{fs} + PL_{coll} - PL_{fs} PL_{coll}$, where PL_{fs} is the average free-space loss probability at a given distance d between UAVs, and PL_{coll} is the average collision probability due to the medium contention. Therefore, PL_{FAN} is the average loss probability when UAVs cooperate in gossiping C2 messages received via satellite, in order to compensate for any losses.

In the following, we derive the analytical formulation of the throughput within the FANET. Then, we estimate the coverage probability and the average delivery delay of a C2 message. The medium access mechanism among UAVs is here based on the use of the 802.11 standard: 802.11 frames are transmitted on the channel, and each frame is composed of several time-slots. Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) mediates the access to the shared medium. CSMA/CA uses a backoff time randomly chosen from a contention window of length W [slots], whose range is $[W_{min}, W_{max}]$. W is doubled after each unsuccessful transmission, up to the maximum value equal to $(W_{max} + 1)$. A two-dimensional Markov chain of $(b+1)$ stages is used in [9] to model the backoff time of a node. That model assumes that each block collides in a time-slot with constant and independent probability PL_{coll}, whereas τ, which is the stationary probability that the node transmits a packet in a generic randomly chosen time-slot, is derived as a function of the number of backoff stages b, of the minimum contention window value W_{min}, and of the collision probability. In [9], a system of non linear equations allows a unique solution (PL_{coll}, τ), that in turns is used to compute the normalized throughput:

$$Thr = \frac{P_s P_{tr} L}{P_s P_{tr} T_s + P_{tr}(1 - P_s)T_c + (1 - P_{tr})T_{id}}, \qquad (1)$$

where $P_{tr} = 1-(1-\tau)^n$ is the probability that there is at least one transmission in a time-slot; $P_s = (n\,\tau(1-\tau)^{n-1})/(1-(1-\tau)^n)$ is the probability of a successfully transmission on the channel, given that at least one node has transmitted; T_{id} is the duration of a time-slot; L is the payload size; $T_s = MAC_{header} + L + SIFS + 2T_{id} + ACK + DIFS$ is the average duration of the busy period of the channel because of a successful transmission; $T_c = MAC_{header} + L + DIFS + T_{id}$ is the average duration of the busy period of the channel because of a collision. $SIFS$, $DIFS$ and ACK parameters in use in this work are provided in Table 1.

By using (1), we can compute the delivery delay of the gossiped blocks on the 802.11 channel. We are interested in evaluating the coverage probability P^{cov}, which is the probability that each node receives all the k blocks composing a C2

message. It is defined by the following equation:

$$P^{cov} = P_{SAT} + (1 - P_{SAT}) P_{FAN}^{cov}. \tag{2}$$

Equation (2) is composed of two terms: the probability P_{SAT} that a node successfully receives the blocks via satellite, and the probability P_{FAN}^{cov} that the blocks are successfully received from the gossiping neighbors. In order to estimate P_{FAN}^{cov}, we need to enumerate the coverage events. *A coverage event* occurs if all k blocks are received by a single UAV thanks to gossiping. In order for a coverage event to occur, the UAV must receive at least once each of the k control blocks[2]. Assuming that a coverage event has occurred, the total number of control blocks received by the UAV, namely h, is bounded as follows: $k \leq h \leq k(n-1)$.

In order to enumerate the coverage events, we need the following definitions: a *row* (column) *margin* is defined as the sum of the entries, rows by rows (columns by columns). The coverage events can be enumerated by counting the number of receiving matrices $\mathcal{M}(R(h), C(h))$. A receiving matrix is composed of vectors $R(h) = \{r_1(h), , ..., r_n(h)\}$ and $C(h) = \{c_1(h), , ..., c_k(h)\}$, which are the row and column matrix margins, respectively, for a given h. Matrices $\mathcal{M}(R(h), C(h))$ are $(n-1) \times k$ binary matrices, where the entry (i, j) is 1 or 0, if the i-th neighbor successfully transmits (or not) the j-th block. An example of a receiving matrix $\mathcal{M}$ is shown in Fig. 1b(a), for $k = 4$ and $n = 5$, with margins $R(h) = \{4, 2, 0, 0, 0\}$, $C(h) = \{2, 2, 1, 1\}$. The sum of the elements of $R(h)$ and $C(h)$ in Fig. 1b(a) is $h = 6$. The set of margins $\mathcal{C}(h)$, corresponding to the coverage events, can be enumerated by evaluating all the k integer partitions of h with $h = k, \cdots, k(n-1)$. We recall that the integer partitions of h are the ways of writing h as a sum of k positive integers $C(h) = \{c_1(h), \cdots, c_k(h)\}$. Some of the partitions may not be feasible; a partition is said to be *feasible* if and only if $\sum_i c_i(h) \geq k$ and $c_i(h) \geq 1$, for $i \in [1, k]$. In order to enumerate only the feasible partitions, some constraints are needed. Those constraints can be written as follows:

$$h = \sum_{i=1}^{k} c_i(h), \forall C(h) \in \mathcal{C}(h), \ h = k, \cdots, k(n-1);$$
$$1 \leq c_i(h) \leq n - 1, \ \forall C(h) \in \mathcal{C}(h). \tag{3}$$

The constraints in (3) guarantee that the maximum number of transmissions per block is equal to the number of the gossiping nodes, and that the minimum number of successfully transmitted blocks must be k for a coverage event to occur. For instance, the matrix in Fig. 1b(a) shows the symbols s_k received (entries equal to 1) or not received (entries equal to 0) from each user u_n for one of the two feasible column margins $C_1(6) = \{2, 2, 1, 1\}$ and $C_2(6) = \{3, 1, 1, 1\}$ for $h = 6$. Furthermore, given that any column margin $C_l(h)$ has a finite number of entries, namely $c_{li}(h)$, $c_{li}(h)$ can appear with multiplicity m_i, thus all the permutations with repetition ϵ_r of $C_l(h)$ must be evaluated, because these margins

---

[2] Because of gossiping, each block can be received more than once.

still represent feasible solutions. These permutations can be calculated as:

$$\epsilon_r := \begin{pmatrix} k \\ m_1, m_2, \cdots, m_j \end{pmatrix} = \frac{k!}{m_1! m_2! \cdots m_l!} \tag{4}$$

Note that P_{FAN}^{cov} depends on the number of nodes that transmit with success. Therefore, for any $C_l(h) \in \mathcal{C}(h)$, several configurations of successfully transmitting neighbors $\mathcal{R}_l(h) = \{R_{l,j}(h), \ j = 1, \cdots, J\}$ are possible, where $R_{l,j}(h) = \{r_{l,j,1}(h), \cdots, r_{l,j,n-1}(h)\}$.

Figure 1b shows two of the possible configurations of transmitting users for the column margin $C_1(6) = \{2,2,1,1\}$: $R_{1,1}(6) = \{4,2,0,0,0\}$ and $R_{1,2}(6) = \{3,1,1,1,0\}$. Again, we can notice that $R_{1,2}(6)$ can be achieved from $R_{1,1}(6)$ as follows: $\{a_{1,4} \to a_{2,4}, a_{2,1} \to a_{3,1}, a_{2,2} \to a_{4,2}\}$. Hence, the set of feasible row margins in $\mathcal{R}_l(h)$ can be enumerated by evaluating the permutations with repetition ϵ_r of the binary column vectors of the receiving matrix. Then, arranging those permutations in groups large k, the disposition groups are obtained from the obtained $R_{l,j}(h)$ row margin. A *disposition group* is a permutation group where each permutation is taken only once. For the sake of clarity, in Fig. 1b(a), we have $s_1 = \{1,1,0,0,0\}$, $s_2 = \{1,1,0,0,0\}$, $s_3 = \{1,0,0,0,0\}$, $s_4 = \{1,0,0,0,0\}$, with sets of permutations $\epsilon_r(s_1)$, $\epsilon_r(s_2)$, $\epsilon_r(s_3)$, $\epsilon_r(s_4)$. The number of permuted elements is 6 for $\epsilon_r(s_1)$ and $\epsilon_r(s_2)$, 4 for $\epsilon_r(s_3)$ and $\epsilon_r(s_4)$. The total number of arrangements in the groups of 4 permuted vectors is $6 \cdot 6 \cdot 4 \cdot 4 = 576$, but several of these arrangements provide the same row margin $R_{l,j}(h)$, which must be counted only once in order to obtain a disposition group. For each column margin $C_l(h)$, the set of row margins $\mathcal{R}_l(h)$ can be now evaluated. However, we still need to enumerate all the possible matrices $M_i(C_l(h), R_{l,j}(h))$ for a given pair of margins $C_l(h)$, $R_{l,j}(h)$. The matrices $M_i(C_l(h), R_{l,j}(h))$ correspond to the feasible coverage events. In fact, there exists a set of matrices $\mathcal{M}(C_l(h), R_{l,j}(h))$, where $M_i(C_l(h), R_{l,j}(h))$ differs from $M_j(C_l(h), R_{l,j}(h))$ by a sequence of *elementary circuit sub-matrices*, as proved in [10]. In Fig. 1b(b), the sub-matrix $\{\{u_1, u_2\}, \{s_3, s_4\}\}$ can be changed in another one through an elementary circuit matrix, as follows:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the relative margins $C_l(h)$, $R_{l,j}(h)$ do not change. An algorithm to efficiently count the matrices of the set $\mathcal{M}(C_l(h), R_{l,j}(h))$, hereafter referred to as $|\mathcal{M}(C_l(h), R_{l,j}(h))|$, is provided in [11]. To summarize, the coverage events can be evaluated according to the following steps:

1. enumerate all feasible column margins in $\mathcal{C}(h)$ and its permutations $\epsilon_r(\mathcal{C}(h))$ for a fixed h value;
2. enumerate all feasible row margins $R_{i,j}(h)$, $\forall C_i(h) \in \mathcal{C}(h)$;
3. count the number of matrices $\mathcal{M}(C_i(h), \mathcal{R}_i(h))$ that correspond to the coverage events in the sets $(C_i(h), \mathcal{R}_i(h))$;

4. count all the matrices corresponding to the coverage events related to all the feasible partitions. The number of those matrices is Ω, and can be estimated as in the following formula:

$$\Omega(h, \mathcal{C}(h), \mathcal{R}(h)) = \epsilon_r(\mathcal{C}(h)) \sum_{C_l(h) \in \mathcal{C}(h)} \sum_{R_{l,j} \in \mathcal{R}_l(h)} |\mathcal{M}(C_i(h), R_{i,j}(h))|.$$

So far, we are able to enumerate all the possible coverage events for the parameters h and the number of gossiping users U_{tx}. Finally, the occurrence probability of a coverage events is:

$$Pr(h|U_{tx}) = \Omega(h, \mathcal{C}(h), \mathcal{R}(h))(1 - PL_{FAN})^h (PL_{FAN})^{U_{tx}k-h}. \tag{5}$$

In other words, (5) allows evaluating the probability that an UAV receives h blocks given U_{tx} transmitting neighbors. The remaining $(n - 1 - U_{tx})$ UAVs do not contribute to h, because the blocks transmitted via satellite and via gossiping have been lost. Let $Pr(U_{tx})$ be the probability of having U_{tx} gossiping neighbors. The coverage probability can be written as:

$$P_{FAN}^{cov} = \sum_h Pr(h|U_{tx}) \, Pr(U_{tx}),$$

$$Pr(U_{tx}) = (P_{SAT})^{U_{tx}} \left[(1 - P_{SAT}) + P_{SAT} PL_{FAN}^k \right]^{n-U_{tx}-1}.$$

4   Numerical Results

In this section, we provide some numerical results for the scenario under consideration. We consider an UAV swarm composed of n nodes, $PL_{SAT} \in [0.01, 0.1]$, and an average free-space loss probability in the FANET $PL_{fs} = 0.07$. It is worth noting that the largest contribution to the average loss rate PL_{FAN} within the swarm is due to the collision probability PL_{coll}. Table 1 provides the settings used in the performance evaluation, which provide the following results. Two performance metrics are here considered: the coverage probability and the delivery delay. In fact, when dealing with C2 data, a timely and correct reception of the commands is of primary importance.

Table 1. Settings of the 802.11-based intra-UAVs channel.

Carrier frequency	Channel bit rate	Propagation delay	Slot time	SIFS/DIFS	Payload	MAC/PHY header	ACK	$W_{min}/$ W_{max}	b
2.4 [GHz]	1 [Mbps]	1 [μs]	50 [μs]	28/128 [μs]	1024 [B]	272/128 [b]	112 [b] + PHY header	15/1023 [time-slots]	$\log_2 \dfrac{W_{max}+1}{W_{min}+1}$

Figure 3 shows that the coverage probability vs. the number of gossiping neighbors, when a C2 packet composed of $k = 4$ blocks is sent. The use of a

gossiping algorithm can significantly increase the coverage probability. However, when considering severe impairments on the satellite channel ($PL_{SAT} = 0.1$), the coverage cannot be guaranteed, as shown in Fig. 3 with 10 UAVs. Different approaches, such as either forward error correction techniques for real-time traffic as in [12,13], or the use of Network Coding (NC) as in [14], can be employed to increase the coverage probability; such investigation is left out for future works.
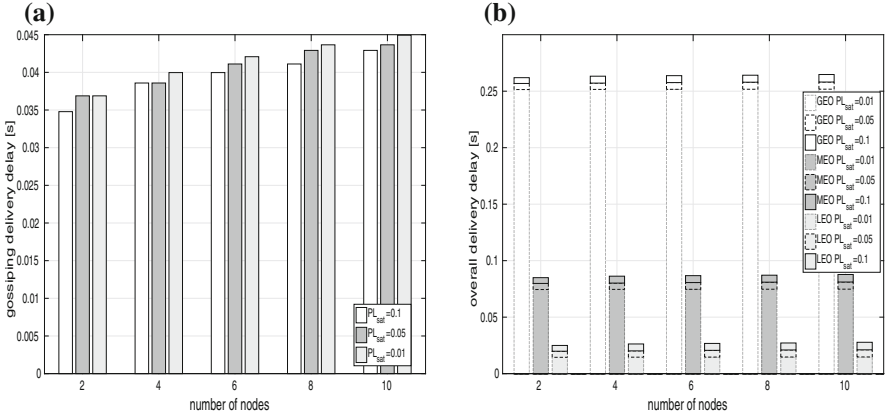

Fig. 2. (a) 802.11 delivery message delay vs. number of gossiping nodes (b) Overall message delivery delay vs. number of gossiping nodes
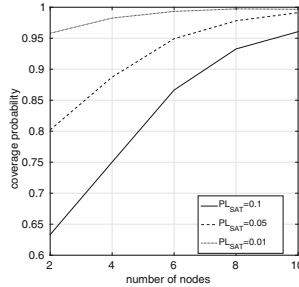

Fig. 3. Coverage probability vs. number of gossiping nodes

Figure 2a shows the impact (defined as *gossiping delay*) of increasing the number of gossiping nodes on the average delivery delay, due to 802.11 backoff mechanism, which reduces the time a node contends for the medium, to reduce the collision probability. Figure 2b shows the one-way-delay for delivering a C2 message from the GCS to the UAV swarm, in case of GEO, MEO and LEO satellite. This delay is the sum of two components: the satellite latency, weighted by P_{SAT}, and the gossiping delay (in Fig. 2a), when the message is received via

gossiping. The use of a gossiping algorithm does not significantly impact on the delivery delay; in fact, the largest component is due to the latency of the satellite. However, the channel latency is still higher than the gossiping delay in case of LEOs, but the latter is no more so negligible w.r.t. the former.

5   Conclusions

This study provides an analytical framework for evaluating the probability of a reliable C2 message delivery via satellite, as well as the relative delivery delay, when a gossiping algorithm is used within the swarm to increase the probability of a successfully transmission. Such a framework provides a simple but effective tool for future studies, both theoretical and empirical ones. Future works will be devoted to the development of an actual test-bed, in order to validate the results provided in this work and, furthermore, to the extension of the proposed analytical framework to the case of an UAV swarm controlled by a single GCS, when LoS communications are considered.

References

1. Bacco, M., Ferro, E., Gotta, A.: Radio propagation models for UAVs: what is missing? In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 391–392. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2014)
2. Bacco, F.M., Ferro, E., Gotta, A.: UAVs in WSNs for agricultural applications: an analysis of the two-ray radio propagation model. In: Proceedings of IEEE SENSORS, pp. 130–133. IEEE (2014)
3. Chmaj, G., Selvaraj, H.: Distributed processing applications for UAV/drones: a survey. In: Selvaraj, H., Zydek, D., Chmaj, G. (eds.) Progress in Systems Engineering. AISC, vol. 366, pp. 449–454. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-08422-0_66
4. D'Auria, S., Luglio, M., Roseti, C., Strollo, R., Zampognaro, F.: Real time transmission of cultural heritage 3D survey in case of emergency. In: 3rd International Conference on Information and Communication Technologies for Disaster Management, Vienna, Austria, December 2016
5. Bekmezci, I., Sahingoz, O.K., Temel, Ş.: Flying ad-hoc networks (FANETs): a survey. Ad Hoc Netw. **11**(3), 1254–1270 (2013)
6. Frew, E.W., Brown, T.X.: Networking issues for small unmanned aircraft systems. J. Intell. Robot. Syst. **54**(1–3), 21–37 (2009)
7. Corner, J.J., Lamont, G.B.: Parallel simulation of UAV swarm scenarios. In: Proceedings of the 36th Conference on Winter Simulation. Winter Simulation Conference, pp. 355–363 (2004)
8. Wei, Y., Blake, M.B., Madey, G.R.: An operation-time simulation framework for UAV swarm configuration and mission planning. Procedia Comput. Sci. **18**, 1949–1958 (2013)

9. Bianchi, G.: Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J. Sel. Areas Commun. **18**, 535–547 (2000)
10. Brualdi, R.A.: Matrices of zeros and ones with fixed row and column sum vectors. Linear Algebra Its Appl. **33**, 159–231 (1980)
11. Miller, J., Harrison, M.: Exact sampling and counting for fixed-margin matrices. Ann. Stat. **41**(3), 1569–1592 (2013)
12. Barsocchi, P., Gotta, A., Potortì, F., González-Castaño, F., Gil-Castiñeira, F., Moreno, J., Cuevas, A.: Experimental results with forward erasure correction and real video streaming in hybrid wireless networks. In: 2nd International Symposium on Wireless Communications Systems - Conference Proceedings, ISWCS 2005, vol. 2005 (2005)
13. Gotta, A., Barsocchi, P.: Experimental video broadcasting in DVB-RCS/S2 with land mobile satellite channel: a reliability issue. In: 2008 International Workshop on Satellite and Space Communications - Conference Proceedings, IWSSC 2008 (2008)
14. Giambene, G., et al.: Network coding applications to high bit-rate satellite networks. In: Pillai, P., Hu, Y.F., Otung, I., Giambene, G. (eds.) WiSATS 2015. LNICST, vol. 154, pp. 286–300. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25479-1_22

Toward Decentralised Consensus and Offloading for Area Coverage in a Fleet of Drones

Hanna Kavalionak, Emanuele Carlini[(✉)] [iD], Pietro Cassarà, and Carlo Meghini

Institute of Information Science and Technologies (ISTI) of National Research Council (CNR), via Moruzzi, 1, Pisa, Italy
{hanna.kavalionak,emanuele.carlini,pietro.cassara,
carlo.meghini}@isti.cnr.it

Abstract. A precise and dynamic visual coverage of a given area is an essential task in many smart contexts, ranging from civil communities to military applications. Due to the last years advancement in hardware miniaturization and efficiency, area coverage is often performed with a combination of static and moving devices, such as unmanned aerial vehicles (drones). Drones are useful to cope with the highly unpredictability and dynamicity of environments, but require specific and efficient solutions toward and efficient area coverage. In this paper we proposes an initial work toward a drone-based approach for the task of area coverage. In particular, we focus our analysis on the following points: (i) decentralized consensus for movement planning, and (ii) the integration of cloud computing infrastructures and technologies for computation offloading, both for image analysis and movement planning.

Keywords: Decentralized consensus · Drones · Distributed tracking Dynamic environments

1 Introduction

The active monitoring of a geographical area through sensors is a fundamental and widespread aspect for a wide range of applications including private surveillance, crowd tracking, and public security. Active monitoring can be exploited in a number of both military and civil applications, such as surveillance of national borders to control immigration/emigration or controlling the flow of tourists in large cities. A core aspect of active monitoring is represented by the task of *area coverage*, i.e. the ability to place sensors in order cover the considered area in an optimal way. Since in many applications the condition of the monitored area can change abruptly, the best way to place sensors also change over time. An infrastructure made of static sensors can be not enough to cope with unexpected events that can result from the inherent unpredictability of crowd behaviour and the environment, such as for example a broken camera or unexpected visual obstacles. In addition, ground sensors takes time to be installed,

and therefore cannot be deployed in an unexpected situation if not foreseen in advance. Also, the monetary investment for the monitoring of a single event can not be justified in certain scenarios (e.g. research activities).

Therefore, a crucial aspect is the degree of adaptability that sensors (e.g. cameras, temperature, sound, etc.) are able to exploit. In order to cope with highly unpredictability and dynamicity of coverage activities, recently several approaches exploits Unmanned Aerial Vehicles (UAVs, informally known as *drones*) as a valid option to carry out many different kind of monitoring [15]. The technological advancements of UAVs rapidly increased in the recent years mostly due to military reasons, nevertheless most of the technology is also available to civil and research purposes. Drones can be deployed to different locations on demand, with a very short notice and without requiring a dedicated static infrastructure placed beforehand. Their behaviour can be reprogrammed while in mission, making them suitable to adapt to fast and unpredictable events within the same mission.

In this paper we consider the challenge of an *active area coverage* by means of a fleet of UAVs. For the purpose of this paper, we assumes UAVs are equipped with means to communicate to each other and with the sensors necessary for their mission. The usage of a fleet of drones presents multiple benefit when compared to a single UAV in terms of: (i) size of the coverable area, (ii) duration of coverage, (iii) coverage redundancy. However, a careful orchestration is required in order to accurately plan the movements of UAVs in order to obtain the best coverage possible. A straightforward way to organize the movement of UAVs is to employ a centralized entity (e.g. a server) that continuously collects their position and generate the new movement plan. However, the effectiveness of this solution is limited, as it suffers in terms of robustness (what if the server crashes?) and scalability (the frequency and the number of communication can saturate the server, which is not able to produce the new movement plans in time). Therefore, in we advocate a decentralized and distributed approach, in which the drones self-organize their movement toward an effective active area coverage.

Besides the self organization of the fleet, we also draw several considerations in relation to the utilization of cloud computing technologies and mechanisms for the image analysis activity of the drones. In particular, we consider the case in which specific image analysis techniques are too computationally heavy to be executed by drones (e.g. due to battery constraint) and it would be more effective to communicate the images to be analyzed to a cloud server which can return back the results of the computation. We analyze this approach according to the work already done in the field of mobile cloud, in which mobile devices (typically smartphones) offload their computation to nearby cloud computation units (cloudlet).

The paper is organized as the following. Section 3 presents the reference architecture and the envisioned scenario considered in the paper. Section 4 discusses a preliminary model for the area coverage and its exploitation in a best-of-n problem formulation. Section 5 elaborates the possible integration of cloud computing technologies for the offloading of computation in the considered scenario. Finally, Sect. 6 concludes the paper.

2   Related Work

The task of self-organize the movements of a number of entities in a decentralise fashion is not new, and it might be referred to as *flocking* [11]. Generally, such task can be abstract as a specialized version of distributed consensus, and it has been tackled in many research fields, although often with different nomenclature and purpose. For example, in the field of multi-agent systems, holonic systems define an organizational model of agents based on self-similarity into "super-agents"? that are seen as single agents from the outside [7]. In peer-to-peer many approaches rely on self-organization techniques to organize the peers of the network in overlay for multiple purposes, such as area coverage [3] or in order to estimate a distribution of network parameters [12].

Recently, decentralized and peer-to-peer flocking algorithms have been applied to drone networks with the aim of self-organizing a fleet of drone toward the completion of complex task. The literature about this topic is very vast; hereby we provide several pointers to recent results. For example, Vásárhelyi et al. [17] provides an algorithm solution based on short-term repulsion and long-range attraction of drones, and it is validated via a numerical simulation. Another recent approach, Yuan et al. [18] proposed a decentralized model predictive control (DMPC) flocking algorithm to self-organize the movement planning of drones, by using the XBEE communication technology.

Several recent works specifically dealt with some version of the problem of area coverage and in specific application domains. For example, [2] propose the usage of drones for area coverage in agricultural applications. Rosalie et al. [13] proposed an ant-colony algorithm paired with a way-points based mobility model to improve the area coverage of drones. The approach of Schleich et al. [15] is to maintain a connected network among the drone by exploiting a tree-based overlay network, as well a mechanism that allows drone to predict positions of one-hop neighbours in the tree.

Most of the above works do not consider the remaining level of the battery for the drones in the system when planning for movement or actions. However, there are few approaches that take batter in consideration for different purposes. For example, Messous et al. [10] propose an approach that tries to keep network connectivity taking into account the battery level of the drones in a fleet. Although in a preliminary shape, our approach differentiates from the above ones as it embeds the following features: (i) consideration of battery consumption in the decentralized modeling for area coverage, and (ii) integration of cloud computing technologies for the offloading of computation devoted to area coverage activity.

3   Reference Scenario and Architecture

The envisioned scenario is depicted in Fig. 1. In such scenario, groups of persons move across the considered area. The area is potentially large and can contains obstacles such as trees or buildings. The area can already be equipped with static ground sensors devoted to area coverage that are interconnected via a wireless
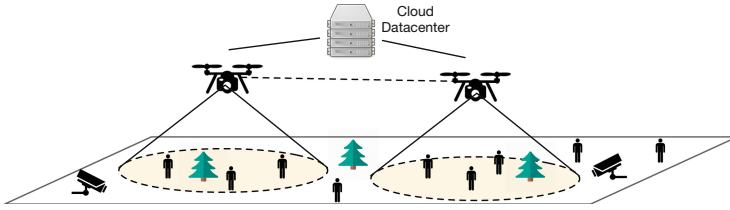
Fig. 1. The envisioned drone-assisted cloud-based crowd control scenario

or wired network. The typical operations conducted by these sensors include the estimation of the crowd density, motion and behaviour. However, an infrastructure made solely of static sensors can be not enough to cope with unexpected events that can result from the inherent unpredictability of crowd behaviour and the environment, such as for example a broken camera or unexpected visual obstacles. In addition, ground sensors takes time to be installed, and therefore cannot be deployed in an unexpected situation if not foreseen in advance. Also, the monetary investment for the monitoring of a single event can not be justified in certain scenarios (e.g. research activities). Therefore we advocate a scenario in which sensor-equipped drones complement with the ground sensor network in order to resolve many of the aforementioned issues. In such scenario a fleet of drones flies above the area, each drone connected to each others and with the network of ground sensors. Drones can be used as highly-moving computational and storage units, allowing for a dynamic access point toward remote cloud datacenters. They can be deployed to different locations on demand, with a very short notice and without requiring a dedicated static infrastructure placed beforehand. Their behaviour can be reprogrammed while in mission, making them suitable to adapt to fast and unpredictable events within the same mission.

In the light of aforementioned vision, the project focuses on two tightly connected aspects:

– a scalable and decentralized support for drones-to-drones and drone-to-ground communication, with the aim of disseminate information about both the state of the sensor and behaviour of the crowd in the drone-assisted area coverage network.
– an effective and QoS-aware orchestration of the computation related to area coverage in terms of computational resource selection, task management, and offloading to remote computational resources, organized by means of the Cloud Computing paradigm.

An high level overview of a reference architecture for the internal software of the drone is depicted in Fig. 2. On the bottom level of the architecture lies the drone hardware. We assume drones to be equipped with sensor for manoeuvrability (e.g. GPS, rotors controller, etc.) and image acquisition (e.g. cameras). We also assume they are equipped with relatively high battery capacity and computational power.

The *communication manager* module will take into account the management of the drone-to-drone and drone-to-ground communications. Since communication is a costly operation, a particular care will be taken such that information dissemination will be done in an effective way, maximising the usefulness of information sent. The component will also take into account the unreliability aspects of the communication channels. The information obtained by means of the communication module will feed the *local context manager*. The context models the view of a drone about its surrounding, and contains information of other drones, ground sensor and about the crowd. The information of drones range from their positions, direction and speed, to battery level and computational capacity.

An important features of the context is the exploitation of prediction algorithms to predict ahead the context, which will allow the drones to plan in advance their behaviour so to possibly anticipate or avoid critical situations. On the top of the stack, the *application manager* orchestrates the computational aspect of the drones. The computational tasks can be related to the decentralized organization of the fleet and area coverage activities. The *movement planner* decides the trajectory of the drone considering the local context, and in such a way to globally optimize the area covered by the fleet. The *crowd tracking* module will employ image recognition algorithms already

Fig. 2. High-level architectural view

existing in literature in order to acquire information and build models of crowd behaviour. The application manager will coordinate the computation underlying these modules by deciding whether to execute the related tasks locally or remotely according to the local context.
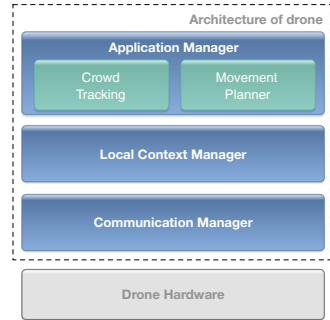
4   A Model for Area Coverage

In general terms, the targeted objective is to *cover* as much area as possible by exploiting a fleet of drones, while minimizing battery consumption. The exact meaning of cover depends on the specific application scenario; In this context we consider an area coverage devoted to crowd control, in which the main task carried out by drones drones acquire images of a certain area and perform some analysis with the aim of recognize certain items (i.e. crowds, people, objects). Therefore, our problem can be formulated as: for a given accuracy associated with the recognition/classification of the items, we have to find/compute a combination of states for the drones that maximizes the surveillance area coverage while minimizing their battery energy consumption.

We consider each drone having a state composed by three parameters: (i) its geographical (GPS) position (x, y), where x and y are linear coordinates; (ii) the altitude h (distance from the ground), which also specify the area coverage S; and (iii) energy level of the battery E. The task is then to find the proper

combination of the states that would maximize the area coverage for the given accuracy while minimizing the battery level consumption.

The proposed approach considers to divide the behaviour of the drones in two phases:

1. A local adjustment phase in which drones applies the necessary changes in order to respect the given accuracy;
2. A phase in which (i) drones exchange each other possible option plans for their positioning, and (2) reach a consensus on which plan to apply. This phase exploit a best-of-n formulation to reach the consensus.

In the following we provide a preliminary analysis and modelling of the two phases.

4.1   Local Adjustments for Accuracy

The given recognition accuracy is reached by the drones by operating locally on two parameters:

– The altitude of the drone h (Fig. 3). According to its altitude, a drone can cover a different portion of an area with a different level of details. A lower altitude allows to receive more details of the items of interest and, presumably, to perform a more detailed and precise recognition/analysis of such items. Nevertheless, in this case the coverage area of the drone decreases together with the attitude $S = \pi r^2$, where r is the radius of the coverage circle. Considering the ϕ as angle of horizontal field of view of the drone to be fixed, while h changes in time we can derive the area coverage radius as:

$$r(t) = h(t) * \tan(\frac{\phi}{2}) \tag{1}$$

– The specific algorithm used for image recognition. This choice has an impact on battery consumption due to different complexity of the computation. Normally, the more accurate algorithm is for the recognition, the more energy consumption computation has to be executed. Hence decreasing the accuracy of the recognition can decrease the required energy consumption.

Therefore, we can derive the accuracy of the recognition by the drone as a function of its attitude and energy consumption. In other words given the required accuracy and the local state of the battery the drone can compute in what way to satisfy the accuracy, by: (i) decreasing the covered area; (ii) increasing the computation complexity of the algorithm increasing battery consumption; or (iii) applying a combination of both these approaches.

4.2 Best-of-n Formulation

Self-organization is a popular research topic in robot
swarm, especially in its Best-of-n Problem formulation
[16]. In particular, the best-of-n problem refers to prob-
lem of collective decision making done by a set of agents.
According to Valentini et al. [16] *"The best-of-n problem
requires a swarm of robots to make a collective decision
over which option, out of n available options, offers the
best alternative to satisfy the current needs of the swarm"*.
As a consequence, a decision among the options is taken
according to the concept of majority (i.e. when a suffi-
cient number of agents favour a specific option), which
generally depends on the specific application.

Fig. 3. Area coverage
for a drone according
to its height

Valentini et al. also categorize the best-of-n prob-
lems according to two specific characteristics of an
options, namely *quality* and *cost*. Both these characteris-
tics depend on the application scenario considered. A best-of-n problem is then
categorized according to the symmetry or asymmetry of both quality and cost. If
in a given problem all the options have the same quality, the n-problem is sym-
metric with respect to quality. Otherwise if at least two options have different
quality, then the problem is asymmetric. The same reasoning goes for the cost.

Few assumptions are necessary to frame the active area coverage defined in
this paper into a discrete best-of-n problem. The first assumption is done by
making discrete the problem of coordinate a set of drones (flocking). This can
be done by considering the following two factors: (i) limit the area of actions
of drones and (ii) divide this area into a grid of tiles, and the movement of
the drones are defined as movement from one tile to another. With these two
assumptions, the flocking problem goes to continuous to discrete, as the possible
actions (movements) of the drones are finite. The second assumption is that the
drones have in place the proper protocol an technology (such as XBEE [18])
to communicate to each other in order to exchange the computed plans and to
select the one that best satisfy the area coverage problem.

By applying the same criteria used by Valentini et al. we characterize the area
coverage as a best-of-n problem, by associating the quality of a solution with the
amount of area covered (at the given quality) by the fleet, and the cost with the
amount of energy spent by drones in order to provide such coverage. According to
this formulation, both quality and cost are asymmetric. The interaction between
cost and quality can be defined as synergic or agnostic: in the former case the
best option has the best quality with the minimum cost, in the latter the best
option results in a tradeoff between cost and quality. In our case the interaction
between area coverage and energy consumption is synergic, as the best solution
would be the one that maximizes the area coverage while minimize the energy
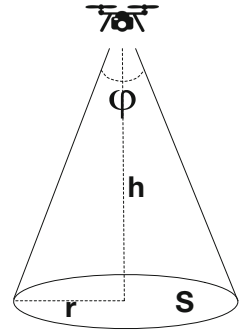consumption.

5   Offloading to Cloud Computing

In the last years, many approaches have dealt with scenario in which computation is offloaded from mobile devices to cloud datacenters [6]. The benefit of such offloading is to improve the capacity of mobile and thin devices, usually limited in terms of CPU, memory and battery life, so that even simple devices can run complex and demanding applications. Among the many proposals, MAUI [5] and CloneCloud [4] are based on virtual machine migration and focuses on offloading of computation from mobile devices to remote servers at execution time, allowing the developers of applications to decide which computation can be offloaded.

In terms of computational resources, Cloud computing could represent an ideal back-end solution to manage the computation related to crowd tracking and image processing [8,9]. However, due to the large amount of data collected, which needs to be transferred to the cloud, and the inherent dispersion of entities that performs data collection, it can be infeasible or inconvenient to transfer the computation toward a large remote datacenter. This is specially true in out envisioned scenario, as the behaviour of the crowd for the purpose of area coverage shall be identified fast such to allow the drone fleet to adjust their position.

This scenario points toward the case in which several ground sensors, or some powerful drone, assume the role of *cloudlet* [14], while normal drones the role of mobile devices. In the cloudlet model, drones would offload their computation to cloudlets, which are relatively small computational units connected with the full blown remote cloud server. Cloudlets are deployed locally to the area of interest and often placed in common and crowded areas to achieve physical proximity with mobile devices. This aspect provides devices with low latency and high bandwidth connections, thereby allowing an interactive response for demanding applications.

The approaches defined for cloudlet currently developed target mobile devices like smart-phones or laptops. The difference with respect to our scenario is the fact that offloading from smart-phones does not affect the context of the cloudlets or the devices. Instead, in our scenario the offloading also affects the behavior of a drone, which in turns can affect the whole fleet. In other terms, the decision whether to offload is not only affecting the quality of the application but potentially affects the area coverage scenario as a whole, for example by modifying the behavior of the other drones in the fleet. Therefore, we plan to adapt existing or design new distributed algorithms that: (1) orchestrate the computation also considering the effect that offloading can have in all the entities related to the crowd tracking, and (2) perform fast and effective brokering of cloudlet resource [1], in order to guarantee the quality of service demands from the crowd tracking tasks.

6   Conclusion

The organization of the activities of a fleet of drones is a relevant task in many of today's smart environment. In this paper, we present several initial considerations

about the area coverage, i.e. the activity devoted to the analysis of an area through image analysis. Specifically, we analysed the problem of decentralized consensus for movement plan in a best-of-n problem formulation, and we reviewed the current trends and approaches for computation offloading, specifically for image analysis, in the frame of cloud computing technologies. As future work, we plan to integrate the analysis provided in the paper into a concrete proposal, both from a technological and algorithmic viewpoints.

References

1. Anastasi, G.F., Carlini, E., Coppola, M., Dazzi, P.: QBROKAGE: A genetic approach for QoS cloud brokering. In: 2014 IEEE 7th International Conference on Cloud Computing, pp. 304–311. IEEE (2014)
2. Barrientos, A., Colorado, J., del Cerro, J., Martinez, A., Rossi, C., Sanz, D., Valente, J.: Aerial remote sensing in agriculture: a practical approach to area coverage and path planning for fleets of mini aerial robots. J. Field Robot. **28**(5), 667–689 (2011)
3. Carlini, E., Ricci, L., Coppola, M.: Integrating centralized and peer-to-peer architectures to support interest management in massively multiplayer on-line games. Concurr. Comput. Pract. Exp. **27**(13), 3362–3382 (2015)
4. Chun, B.-G., Ihm, S., Maniatis, P., Naik, M., Patti, A.: CloneCloud: elastic execution between mobile device and cloud. In: Proceedings of the sixth conference on Computer systems, pp. 301–314. ACM (2011)
5. Cuervo, E., Balasubramanian, A., Cho, D.-K., Wolman, A., Saroiu, S., Chandra, R., Bahl, P.: MAUI: making smartphones last longer with code offload. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 49–62. ACM (2010)
6. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. Wirel. Commun. Mob. Comput. **13**(18), 1587–1611 (2013)
7. Esmaeili, A., Mozayani, N., Motlagh, M.R.J., Matson, E.T.: A socially-based distributed self-organizing algorithm for holonic multi-agent systems: Case study in a task environment. Cogn. Syst. Res. **43**, 21–44 (2017)
8. Kavalionak, H., Carlini, E., Lulli, A., Gennaro, C., Amato, G., Meghini, C., Ricci, L.: A prediction-based distributed tracking protocol for video surveillance. In: 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), pp. 140–145, May 2017
9. Kavalionak, H., Gennaro, C., Amato, G., Meghini, C.: Dice: A distributed protocol for camera-aided video surveillance. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), pp. 477–484. IEEE (2015)
10. Messous, M.-A., Senouci, S.-M., Sedjelmaci, H.: Network connectivity and area coverage for UAV fleet mobility model with energy constraint. In: 2016 IEEE Conference on Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE (2016)
11. Olfati-Saber, R.: Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Trans. Autom. Control **51**(3), 401–420 (2006)

12. Payberah, A.H., Kavalionak, H., Montresor, A., Dowling, J., Haridi, S.: Lightweight gossip-based distribution estimation. In: 2013 IEEE International Conference on Communications (ICC), pp. 3439–3443. IEEE (2013)
13. Rosalie, M., Dentler, J.E., Danoy, G., Bouvry, P., Kannan, S., Olivares-Mendez, M.A., Voos, H.: Area exploration with a swarm of UAVs combining deterministic chaotic ant colony mobility with position MPC. In: 2017 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 1392–1397, June 2017
14. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The case for VM-based cloudlets in mobile computing. IEEE Pervasive Comput. **8**(4), 14–23 (2009)
15. Schleich, J., Panchapakesan, A., Danoy, G., Bouvry, P.: UAV fleet area coverage with network connectivity constraint. In: Proceedings of the 11th ACM International Symposium on Mobility Management and Wireless Access, pp. 131–138. ACM (2013)
16. Valentini, G., Ferrante, E., Dorigo, M.: The best-of-n problem in robot swarms: formalization, state of the art, and novel perspectives. Front. Robot. AI **4**, 9 (2017)
17. Vásárhelyi, G., Virágh, C., Somorjai, G., Tarcai, N., Szörényi, T., Nepusz, T., Vicsek, T.: Outdoor flocking and formation flight with autonomous aerial robots. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3866–3873, September 2014
18. Yuan, Q., Zhan, J., Li, X.: Outdoor flocking of quadcopter drones with decentralized model predictive control. ISA Trans. **71**(1), 84–92 (2017)

How to Support the Machine Learning Take-Off: Challenges and Hints for Achieving Intelligent UAVs

Patrizio Dazzi[(✉)] and Pietro Cassarà

Information Science and Technologies Institute,
National Research Council of Italy, Pisa, Italy
{patrizzio.dazzi,pietro.cassara}@isti.cnr.it

Abstract. Unmanned Aerial Vehicles (UAVs) are getting momentum. A growing number of industries and scientific institutions are focusing on these devices. UAVs can be used for a really wide spectrum of civilian and military applications. Usually these devices run on batteries, thus it is fundamental to efficiently exploit their hardware to reduce their energy footprint. A key aspect in facing the "energy issue" is the exploitation of properly designed solutions in order to target the energy- and hardware-constraints characterising these devices. However, there are not universal approaches easing the implementation of ad-hoc solutions for UAVs; it just depends on the class of problems to be faced. As matter of fact, targeting machine-learning solutions to UAVs could foster the development of a wide range of interesting application. This contribution is aimed at sketching the challenges deriving from the porting of machine-learning solutions, and the associated requirements, to highly distributed, constrained, inter-connected devices, highlighting the issues that could hinder their exploitation for UAVs.

Keywords: Machine learning · UAV · Decentralized intelligence
Machine-to-machine · IoT

1 Introduction

Unmanned Aerial Vehicles (UAVs) [12] are arousing a growing interest from both industrial and scientific communities. This is mainly due to their flexibility: UAVs can be used for a really wide spectrum of civilian and military applications, either for supporting or replacing humans in dangerous and insalubrious environments.

Flexibility that is not only rooted in the large set of potential ways of using such vehicles, but also deriving from the almost endless possibilities of customization, personalization and configuration (i.e. UAVs range from very cheap drones to military planes, from devices having a reduced set of sensors to complete video or meteo flying stations).

Even more, UAVs promise to have a tremendous impact on many areas, also from an economic perspective, many UAVs (even some powerful ones) are built by exploiting commodity hardware. It is quite common to see UAV which installed hardware depends on the specific goals to be pursued by that device. Typically, the preferred key for the hardware selection process is performance/energy trade-off [17].

In fact, usually this devices run on batteries, thus it is fundamental to efficiently exploit the hardware to reduce the energy consumption, e.g. using GPU, APU or accelerators that demonstrated to be very effective solutions in pursuing such goal.

Beyond a careful hardware selection, a key aspect in facing the "energy issue" is the exploitation of proper algorithmic solutions. Properly conceived, designed and implemented to target energy- and hardware-constrained devices and, more in details UAVs. To achieve this goal, many different approaches have been proposed so far [29, 36], however two of the most promising strategies rely on approximation and collaboration.

The former as a way to reduce the energy footprint by accepting results that are not exact but still having enough significance to be useful and/or valuable for the purpose of the UAV. The latter consisting in an active collaboration occurring between UAVs to pursue altogether a common goal.

As matter of fact, there is no standard or universal approach and/or guideline to adopt for implementing the aforementioned strategies; it just depends on the class of problems to be faced. Much research effort need to be spent to innovate solutions in order to target UAV, in fact, as we mentioned before, due to their flexibility, there are many different application domains in which it is worth to exploit UAVs. In spite of this, there are a few classes of solutions that, if properly targeted to UAVs, would be beneficial for many different kinds of applications. Actually, such applications could fruitfully exploit these solutions to achieve an efficient, smart behavior of UAVs. Among them, there are a few machine-learning solutions that could be good candidates to that aim.

The aim of this paper is to outline the challenges that need to be faced to achieving machine learning solutions on a decentralized context and more in particular UAVs. Along the challenges, the paper provides some hints and suggestions as derived from the existing literature, borrowed from other scientific fields.

2 Challenges

Standard machine learning approaches require centralizing the training data on one machine or in a datacenter. However, recently, many approaches [22, 23, 30] have been proposed aiming to provide machine-learning capabilities on a decentralized scenario. In spite of this plethora of proposal, achieving an highly decentralized machine-learning requires to overcome many algorithmic and technical challenges.

In fact, with a typical machine learning system, an optimization algorithm, such as the Stochastic Gradient Descent (SGD) [8], runs on a large dataset appropriately partitioned across servers in a large datacenter or in a cloud. These algorithms require low-latency, high-throughput connections to the training data.

However, as can be easily noticed in a distributed scenario, data is distributed across a large set of devices in a highly unpredictable way. In addition, these devices have higher-latency, lower network bandwidth and are not always online, making them not always available to be involved in the training process. These bandwidth and latency limitations clearly motivate the need for properly defined approaches, tools, methodologies and protocols.

Achieving the computation of these algorithms by means of a large set of distributed, heterogenous and dynamic devices requires an innovative and sophisticated technology stack. Classical solutions need to be re-thought from this actual perspective. Design of algorithms, programming models, runtime supports, network stack, will require a paradigm shift to embrace the peculiarities characterising UAVs.

On device training need to exploit ad-hoc programming tools, properly tuned to the features and capabilities of the devices. Specifically conceived scheduling and runtime resource management ensures that training will happen only when the device is idle, fully charged, or on a high-capacity wireless connection, to limit the impact on the device performance. It is also fundamental that the system communicates and aggregates updates to the originally computed model in an efficient and fault-tolerant way. Communication shall be compressed and provided at irregular time intervals, the whole infrastructure supporting the interactions among UAVs and the computational back-end need to be properly tuned and designed.

Inter-UAVs connections and interaction are also important to achieve a more effective exploitation of the decentralized resources, on the one hand resulting into a reduced amount of data to be sent via the uplink but, on the other hand requiring a careful design of the interaction protocol.

2.1   Communication and Collaboration

Following from their distributed deployment, it is of paramount importance to properly design the communication processes occurring among "intelligent UAVs" to drive the information process exchange in a way that will allow to build a distributed knowledge to exploit for the jointly distributed decision making process. To this end many different approaches could be considered, ranging from technologies borrowed from Peer-to-peer computing, to approaches derived from Agent-based computing.

Smart UAV-2-UAV Computing. As aforementioned, one of the key requirements for enabling the development of smart/intelligent solutions, based on ML technologies, targeting fleets of UAVs, stands in the ability of leveraging the distributed knowledge owned by a fleet of UAVs as it was a unique knowledge base.

To this end, good candidates are the technologies originally developed in the context of P2P computing [2,9,21,24,34]. In particular, can be considered those approaches aimed at enabling a totally decentralized support for indexing and accessing data [16,28,33].

These approaches work by defining logic overlay networks, either structured or unstructured, depending by the degree of dynamicity characterising data, each aimed at supporting the distributed and decentralized execution of a specific set of operations, of various kinds, performed by relying on a subset of the peers composing the network. Regardless the limited involvement of the peers composing the network, the system is able to produce results that consider the whole set of data belonging to the entire system.

As two examples consider:

– a DHT indexes the data belonging to the entire system, then it is able resolve queries without involving the complete set of peers, but just a subset of them, without hindering the quality of the final result.
– a protocol like GROUP [3,4,10] is able to create homogeneous cluster of data, without any need of collecting all the pieces of information in a unique place.

These solutions demonstrated to be very effecting in allowing efficient and effective access to distributed data, often located in remote locations. Even more, these solutions are usually conceived, developed and optimized assuming unstable communications, churn (namely, nodes that disconnect from the network without providing any kind of notification).

A further interesting aspect of these technologies are their ability of managing very different kind of dynamic data, produced at a very different paces, depending on the application and context in which such information is generated.

As matter of fact **P2P computing-based approaches** could be good candidates for providing technological solutions supporting the achievement of Smart UAV-2-UAV computing, by enabling an efficient and effective communication between UAVs. In spite of the significant amount of research to be conducted to achieve a proper adaptation of these solutions to the UAVs, they can be a valid starting point.

2.2 Modelling and Orchestrating

Another important item in the achievement of a distributed, decentralized support to realize the smart UAVs, is the proper way to adopt for the overall modelling of the orchestration affecting the achievement of ML-based solutions. In particular, the way in which is such orchestration/interaction model the overall management of communication, collaboration and data exchange taking place among the UAVs. This is the keystone on which the decentralization is build and achieved. There are not much approaches in literature that specifically deal with such issues.

Federated learning [18,32] is probably the most effective approaches, existing so far, that is conceived and build on the idea of achieving a decentralized ML solution. It consists of a machine learning approach where the goal is to train models that are conceptually centralized but computed in a decentralized way, involving data distributed over a huge amount of computing nodes each with unreliable and relatively slow network connections. Federated learning considers learning algorithms in which, at each round, each node independently computes an update to the current model based on its local data, and communicates this update to a central server, where client-side updates are aggregated to compute a new global model.

In its current form and adoption, the typical clients are mobile phones, so communication efficiency is of utmost importance. In fact, in the paper in which such solution is presented, the authors discuss two ways to reduce the uplink communication costs able to reduce the upload communication required to train a reasonable model by two orders of magnitude w.r.t. other existing approaches.

Federated learning is nowadays the best candidate to be the starting point on which to build the orchestration and modelling of decentralized ML targeting UAVs. By adopting this approach, the ML models driving UAVs could benefit from a large knowledge base at the cost of a limited amount of communication.

2.3   Programmability

After defining the way UAVs follow to communicate one each others, and the information that are expected to exchange in this process, a further complex issue arise: How to program smart UAVs? Namely, how could be achieved their programmability without charging the programmers of the complex, error-prone, development of the entire process discussed so far, aimed at the management of the actual interaction occurring among UAVs? How this could be obtained in a high-level way?

In the scientific literature there is not much work specifically focused on UAVs, however, there are many interesting approaches dealing with pretty similar problems organized in a pretty similar way. It is the case of the solutions aimed at describing, programming and evaluating systems composed by large sets of interacting entities. There are essentially two main sectors in which these issues have been faced so far:

- Agent based models [6,14,19,26]. An agent-based model is a computational model aimed at simulating the actions and interactions resulting by the interplay of autonomous agents, focusing on the assessment of their effects on the system as a whole. It combines elements of game theory, complex systems and evolutionary programming.
- Multi-agent systems [11,25,31,35]. A multi-agent system is a computerized system composed of multiple interacting agents embedded with some form of intelligence, operating within a given target environment. Multi-agent systems are used to solve problems that are difficult for an individual agent or a single system to solve.

Both these approaches, can offer useful hints for the definition of programming models targeting fleets of smart UAVs. By following these agent-oriented views to organise the computation it could be possible to have a well-organized way to structure and organize the development process following a well-studied, high-level and highly-tested approach, that match the underlying deployment architecture of UAVs.

From a more technological viewpoint, one tool that could be considered for the programming of UAVs is Akka [7,13,27]. Akka is a free and open-source toolkit aimed at easing the construction of concurrent and distributed applications on the Java Virtual Machine (that is more and more supported also by the single-board devices installed on UAVs).

Akka supports multiple programming models for concurrency, but it emphasizes actor-based concurrency, essentially porting to an imperative-based languages level the features that were used to characterise Erlang [1].

2.4   Even More Efficient: Let's Go to the Edge

As aforementioned, in several ways and forms, the achievement of ultimate smart UAVs requires effective and efficient collaboration among UAVs, proper modelling of the algorithms and effective way to program them.

The overall, implicit, assumption has been that in this way UAVs can interact in a fully decentralized way to achieve their goal. As an alternative, we pointed out how federated learning could ensure, at cost of a very reduced communication footprint the generation (possibly storing it on a cloud) and assessment of a global ML model, resulting by the single actions performed by each single UAV. This essentially bi-partite the potential system organizations between: (i) totally decentralized models and (ii) cloud-based models.

However, recently is getting momentum the idea of blurring this distinction by introducing a hierarchy of computing devices standing in the middle between clients and the cloud. Depending on the pervasiveness of these additional devices, their aim and focus, we can distinguish between Edge [5,15] (additional devices only at the edge of the network) and Fog [20] computing (additional devices placed along all the network path from the cloud to the clients).

More specifically, it can be envisioned for fleets of UAVs the possibility of relying on properly defined and placed "base stations" aimed at caching data, performing complex computation, etc. The challenge will be to make these stations "invisible", meaning that UAVs (and their "application" programmers) do not need to be aware of the existence of such station; it would be the system itself able to manage the offloading of the computation, the management of the caching and all the activity requested to optimize the work of UAVs by means of the adoption of Edge computing.

The adoption of such deployment strategies can provide useful benefits in terms of energy saving and reduced network latencies.

3   Further, Non-functional, Challenges

Located at downstream of these challenges, there is also yet another key element to take into account: the actual testing of these solutions. Regardless the usual solutions, the behavior of these systems needs to be evaluated and assessed in a realistic, physical, environment before allowing the flight to the developed approaches.

One example of the current existing solution on this extent is the Flying Machine Arena (FMA) developed by the ETH of Zurich. FMA is a portable space devoted to autonomous flight. Measuring up to $10 \times 10 \times 10$ m, it consists of a high-precision motion capture system, a wireless communication network, and custom software executing complex algorithms for estimation and control.

The motion capture system can locate multiple objects in the space at rates exceeding 200 fps. While this may seem extremely fast, the objects in the space can move at speeds in excess of 10 m/s, resulting in displacements of over 5 cm between different snapshots. This information is then joined with other data and models related to the system dynamics in order to predict the state of the objects into the next future.

The system uses this knowledge to determine what should be the commands for the vehicles, that should be executed to achieve their desired behavior. Then, via wireless links, the system sends the commands to the vehicles, which execute them with the aid of on-board computers and sensors such as rate gyros and accelerometers.

Somehow this recalls the idea of federated learning, but in this context is not necessarily related to the achievement of the orchestration itself but mainly to its following evaluation and estimation.

4   Conclusion

In this paper we present a set of relevant challenges to make possible the decentralized execution of ML-based solutions on fleets of UAVs. Along of such challenges we also provide some suggestions about the potential solution to adopt for addressing such issues. As matter of fact a direct adoption may not be possible in all the cases, but represent a useful starting point for investigating and developing solutions addressing the aforementioned challenges.

The presentation mainly focused on the collaboration among UAVs, the definition of a decentralized ML model, the problem related to their programmability and the possibilities deriving from exploiting solutions based on Edge and Fog computing.

Finally the paper briefly observed that another key aspect in the achievement of ML-based solutions for UAVs is related to their actual evaluation, on an actual setup in a real environment. Even in this case the paper reports an existing solution that is worth to consider for the validation of next generation smart UAVs.

References

1. Armstrong, J., Virding, R., Wikström, C., Williams, M.: Concurrent Programming in Erlang. Prentice Hall, Hertfordshire (1993)
2. Baraglia, R., Dazzi, P., Guidi, B., Ricci, L.: GoDel: Delaunay overlays in P2P networks via gossip. In: IEEE 12th International Conference on Peer-to-Peer Computing (P2P), pp. 1–12. IEEE (2012)
3. Baraglia, R., Dazzi, P., Mordacchini, M., Ricci, L.: A peer-to-peer recommender system for self-emerging user communities based on gossip overlays. J. Comput. Syst. Sci. **79**(2), 291–308 (2013)
4. Baraglia, R., Dazzi, P., Mordacchini, M., Ricci, L., Alessi, L.: GROUP: a gossip based building community protocol. In: Balandin, S., Koucheryavy, Y., Hu, H. (eds.) NEW2AN/ruSMART 2011. LNCS, vol. 6869, pp. 496–507. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22875-9_45
5. Beck, M.T., Werner, M., Feld, S., Schimper, T.: Mobile edge computing: A taxonomy (2014)
6. Benenson, I., Torrens, P.M.: Geosimulation: Automata-based Modeling of Urban Phenomena. Wiley, Chichester (2004)
7. Bernhardt, M.: Reactive Web Applications: Covers Play, Akka, and Reactive Streams. Manning Publications Co., Greenwich, CT (2016)
8. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Lechevallier Y., Saporta G. (eds.) Proceedings of COMPSTAT 2010, pp. 177–186. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-7908-2604-3_16
9. Carlini, E., Coppola, M., Dazzi, P., Laforenza, D., Martinelli, S., Ricci, L.: Service and resource discovery supports over P2P overlays. In: International Conference on Ultra Modern Telecommunications & Workshops, ICUMT 2009, pp. 1–8. IEEE (2009)
10. Dazzi, P., Felber, P., Leonini, L., Mordacchini, M., Perego, R., Rajman, M., Rivière, É.: Peer-to-peer clustering of web-browsing users. In: Proceedings of LSDS-IR, pp. 71–78 (2009)
11. Ferber, J.: Multi-agent Systems: An Introduction to Distributed Artificial Intelligence. vol. 1. Addison-Wesley, Reading, MA (1999)
12. Gallington, R.W., Berman, H., Entzminger, J., Francis, M.S., Palmore, P., Stratakes, J.: Unmanned aerial vehicles. Future aeronautical and space systems (A 97–26201 06–31), Reston, VA, American Institute of Aeronautics and Astronautics, Inc., Progress. Astronautics and Aeronautics. **172**, 251–295 (1997)
13. Gupta, M.: Akka Essentials. Packt Publishing Ltd, Birmingham (2012)
14. Helbing, D.: Agent-based modeling. In: Social self-organization, pp. 25–70. Springer (2012)
15. Hu, Y.C., Patel, M., Sabella, D., Sprecher, N., Young, V.: Mobile edge computing-a key technology towards 5G. ETSI White Paper **11**(11), 1–16 (2015)
16. Kaashoek, M.F., Karger, D.R.: Koorde: a simple degree-optimal distributed hash table. In: Kaashoek, M.F., Stoica, I. (eds.) IPTPS 2003. LNCS, vol. 2735, pp. 98–107. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45172-3_9
17. Kestur, S., Davis, J.D., Williams, O.: BLAS comparison on FPGA, CPU and GPU. In: 2010 IEEE Computer Society Annual Symposium on VLSI, pp. 288–293, July 2010
18. Konečnỳ, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: strategies for improving communication efficiency (2016). arXiv preprint arXiv:1610.05492

19. Ligtenberg, A., Bregt, A.K., Van Lammeren, R.: Multi-actor-based land use modelling: spatial planning using agents. Landscape Urban plann. **56**(1), 21–33 (2001)
20. Luan, T.H., Gao, L., Li, Z., Xiang, Y., Wei, G., Sun, L.: Fog computing: Focusing on Mobile Users at the Edge (2015). arXiv preprint arXiv:1502.01815
21. Marzolla, M., Mordacchini, M., Orlando, S.: Resource discovery in a dynamic grid environment. In: Sixteenth International Workshop on Database and Expert Systems Applications, pp. 356–360. IEEE (2005)
22. McMahan, B., Moore, E., Ramage, D., Hampson, S., Aguera y Arcas, B.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282 (2017)
23. McMahan, B., Ramage, D.: Federated Learning: Collaborative Machine Learning without Centralized Training Data (2017)
24. Mordacchini, M., Dazzi, P., Tolomei, G., Baraglia, R., Silvestri, F., Orlando, S.: Challenges in designing an interest-based distributed aggregation of users in P2P systems. In: International Conference on Ultra Modern Telecommunications & Workshops, ICUMT 09, pp. 1–8. IEEE (2009)
25. Olfati-Saber, R., Fax, J.A., Murray, R.M.: Consensus and cooperation in networked multi-agent systems. Proc. IEEE **95**(1), 215–233 (2007)
26. Pahl-Wostl, C.: Actor based analysis and modeling approaches. Integrated Assessment, **5**(1) (2005)
27. Roestenburg, R., Bakker, R., Williams, R.: Akka in Action. Manning Publications Co., Greenwich, CT (2015)
28. Rowstron, A., Druschel, P.: Pastry: scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, pp. 329–350. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45518-3_18
29. Sarma, S., Muck, T., Bathen, L.A.D., Dutt, N., Nicolau, A.: SmartBalance: a sensing-driven linux load balancer for energy efficiency of heterogeneous MPSoCs. In: 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC), pp. 1–6. IEEE (2015)
30. Scardapane, S., Wang, D., Panella, M.: A decentralized training algorithm for echo state networks in distributed big data applications. Neural Netw. **78**, 65–74 (2016)
31. Schumacher, M.: Multi-agent Systems. Objective Coordination in Multi-agent System Engineering: Design And Implementation, pp. 9–32 (2001)
32. Smith, V., Chiang, C.-K., Sanjabi, M., Talwalkar, A.: Federated Multi-task Learning (2017). arXiv preprint arXiv:1705.10467
33. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., Balakrishnan, H.: Chord: a scalable peer-to-peer lookup service for internet applications. SIGCOMM Comput. Commun. Rev. **31**(4), 149–160 (2001)
34. Trunfio, P., Talia, D., Papadakis, H., Fragopoulou, P., Mordacchini, M., Pennanen, M., Popov, K., Vlassov, V., Haridi, S.: Peer-to-peer resource discovery in grids: models and systems. Future Gener. Comput. Syst. **23**(7), 864–878 (2007)
35. Van der Hoek, W., Wooldridge, M.: Multi-agent systems. Found. Artif. Intell. **3**, 887–928 (2008)
36. Woithe, H.C., Kremer, U.: TrilobiteG: a programming architecture for autonomous underwater vehicles. In: Proceedings of the 16th ACM SIGPLAN/SIGBED Conference on Languages, Compilers and Tools for Embedded Systems 2015 CD-ROM, LCTES 2015, pp. 14:1–14:10, New York. ACM (2015)

UAVs and UAV Swarms for Civilian Applications: Communications and Image Processing in the SCIADRO Project

Manlio Bacco[1], Stefano Chessa[2], Marco Di Benedetto[1], Davide Fabbri[3],
Michele Girolami[1], Alberto Gotta[1(✉)], Davide Moroni[1],
Maria Antonietta Pascali[1], and Vincenzo Pellegrini[3]

[1] Institute of Information Science and Technologies (ISTI) of National Research
Council (CNR), via G. Moruzzi, 1, Pisa, Italy
{manlio.bacco,marco.dibenedetto,michele.girolami,alberto.gotta,
davide.moroni,maria.antonietta.pascali}@isti.cnr.it
[2] Department of Computer Science, University of Pisa, Pisa, Italy
stefano.chessa@unipi.it
[3] I.D.S. S.p.A., via E. Calabresi, 24, Pisa, Italy
{v.pellegrini,d.fabbri}@idscorporation.com

Abstract. The use of unmanned aerial systems is increasingly common in both research and industrial fields. Nowadays, the use of single unmanned aerial vehicles is quite established and several products are already available to consumers, while swarms are still subject of research and development. This work describes the objectives of a research project, namely *SCIADRO*, which deals with innovative applications and network architectures based on the use of single unmanned aerial vehicles and of swarms in several civilian fields.

Keywords: UAV · Swarm · Multipath · Crowd-sensing · Imaging

1 Introduction

Nowadays, Unmanned Aerial Vehicles (UAVs) are attracting a lot of attention from industrial and research fields. They are suited to a large number of applications, thus making them of interest for commercial and research purposes. The use of a single UAV provides several services to consumers and to industry because of the low-cost and the reliability that UAVs can provide. In those applications, UAVs are typically used in a Line of Sight (LoS) fashion, i.e., the pilot can Command and Control (C2) the UAV from the ground without losing sight of it. In fact, national regulamentation are quite severe because of safety reasons, and the use of UAVs for civilian applications must adhere to them, thus making their use more common outside of urban areas. An upcoming scenario for UAVs, quite challenging, is the use of *swarms*, or Flying Ad-Hoc Networks (FANETs): several drones, from tens to hundreds, jointly used in order to execute a given task. The joint use of multiple UAVs poses several challenges that

must be met, and the SCIADRO[1] research project deals with some specific application scenarios involving the use of swarms. As pointed out in [1–3], the number of application scenarios for UAVs is rapidly increasing, involving power line inspection, monitoring of cultural heritage sites, environmental monitoring, fire and gas detection, as well as precision agriculture. Several advantages can be brought by the use of multiple UAVs in those scenarios: for instance, it is likely that the overall cost of acquisition and maintenance of several small UAVs can be lower than the overall cost of a single large UAV needed for the same task [4]. Furthermore, fault-tolerance is inherently provided by the use of swarms, because a single drone can be removed with a limited impact on the overall formation. Swarms can also provide scalability, i.e., adding or removing drones from a swarm, in order to better adapt to changing conditions or to simply replace one or more UAVs experiencing issues or just battery depletion.

The rest of this paper is organized as follows: Sect. 2 provides a brief overview of the state of the art. Then, Sect. 3 describes the main application scenarios considered in the SCIADRO project, also proposing several approaches. Eventually, the Conclusions are in Sect. 4.

2   Related Works

A wide range of applications can benefit from the use of UAVs, as in [1–3], such as power lines inspection, monitoring of cultural heritage sites, environmental monitoring. In particular, precision agriculture is largely benefiting of the use of UAVs [5], due to low operational costs, high operational flexibility and high spatial resolution of imagery. The use of UAVs in this field is expected to grow faster in the next years and it has proved to be particularly effective in otherwise impervious areas, or each time their use can remove the need for expensive temporary scaffolding, such as in the case of the inspection of historical or cultural areas and buildings. In the latter scenarios, UAVs are typically equipped with the needed sensors in order to facilitate any inspections: for instance, cameras, but also short-range communication radios, in order to collect data from previously installed sensors or to deliver commands, in the case of actuators. A survey of UAV usage for imagery acquisition in the field of disaster research and management in provided in [6]. The authors focus on the inspection of pipelines, in order to quickly provide damage survey thanks to UAVs. In fact, they provide a low-cost solution for imagery collection, and the small size and maneuverability makes UAVs a viable and low-cost option.

The use of UAV swarms, on the other side, is less established than the use of single UAVs. This depends on the issues that the contemporary use of multiple drones poses to C2, among other issues, in order to avoid any collisions among the members of the swarm, or the need of intelligent algorithms to exchange data, coordinate the swarm, and process collected/generated data. From the network perspective, an UAV swarm is typically referred to as FANET. A FANET has

---

[1] SCIADRO is the acronym of *SCIAme di DROni*, which translates into *UAV swarm*.

distinctive features w.r.t. other networks: a high mobility degree, a flight forma-tion, and average and peak movement speeds of that must be carefully taken into account. For instance, aerial mobility experiences less constraints than the terres-trial one, but higher speeds are expected. Nonetheless, their use is attracting the attention of both the industrial and the scientific world in several fields [1, 6, 7], and this work proposes several civilian applications that can largely benefit of the use of UAV swarms.

3   The SCIADRO Project

Co-funded by the Tuscany region, Italy, the SCIADRO project aims at devel-oping the enabling technologies that are key to accomplishing a rather rich and diverse span of missions through the use of a coordinated drone swarm for civil-ian purposes. Goals of the foreseen missions include environmental monitoring, first response to natural disasters, monitoring of social events, safety inspection of public utility grids or other critical infrastructures. The project team includes seven partners, being either research institutions or Small and Medium-sized Enterprises (SME) with permanent operations in Tuscany. More in detail, the project aims at: (i) developing sensors to monitor the presence of potential pol-lutants within surveyed environments; (ii) achieving computer vision techniques and algorithms which can detect complex objects and extract information on local anomalies which might affect them; (iii) developing suitable logics and algorithms which can effectively organize and guide the overall swarm motion and actions during a mission; (iv) studying, developing and demonstrating net-work architectures and protocols which can allow communication among multiple drones within a swarm while also increasing communication reliability towards the ground segment and reducing the *Size, Weight and Power* (SWaP) require-ments of in-flight radio-communication equipment.

In the following, we present an overview of the activities on communication and image processing currently carried out within the SCIADRO framework.

3.1   Inspection of Aerial Power Lines

The inspection of aerial power lines using UAVs is a service offered by several companies, in order to acquire high quality videos both in the visible light spec-trum and in the infrared spectrum. The SCIADRO project aims at automatizing the monitoring process by using a single UAV or a swarm, in a cost-effective man-ner. The complete automation of such a task requires to empower scene under-standing capabilities on board of the UAV, sharing processing results with peers. The specific use case of the inspection of aerial power lines is taken into account. The UAV swarm is made of drones equipped with different payloads: infrared and RGB cameras, with acquisition parameters set differently, depending on which task should be carried out. The acquired sequences of images undergo 2D and 3D processing, summarized in Fig. 1. The 3D scene reconstruction uses
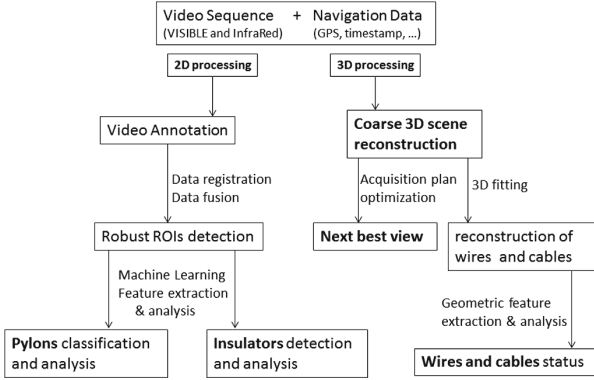
Fig. 1. The processing pipeline for 2D and 3D video processing.

large views of the scene, with single frame preferably including at least two subsequent pylons; the coarse 3D scene may be used for a real-time re-planning of the path of the swarm, e.g. in order to reduce data redundancy or to trigger a focused inspection where possible anomalies are detected, e.g. excess vegetation, or missing/loose/broken components. On the other hand, 2D image processing applied to both RGB and infrared video sequences would benefit a lot from the consistent video annotation of the two sequences of data. The features extracted from each sequence are used to define a fusion map of the inspected power line, possibly refining the video annotation, and supporting classification of relevant components (tower structure, hanging point, insulator, cables) and evaluation of their possible defects. In this case, the swarm advantage is to have heterogeneous data of the same target (RGB and IR data), able to make detection and classification tasks more efficient and robust.

Image Understanding and Recognition. SCIADRO aims at enabling an UAV to perform very complex tasks such as object detection, recognition, and analysis in an unknown environment; it requires fast and robust algorithms, and there are no standard approaches in the literature. Also, image processing should be specifically designed for the collaborative setting of an UAV swarm. Even if the collaborative setting poses a number of issues (e.g. regarding the information sharing and processing), it could be seen as a strength, if a data fusion step is properly implemented on the different data flows: this kind of processing may efficiently increase in quantity and quality the information extracted by the diverse sensors hosted by the UAVs. In the scenario of inspection of aerial power lines, the following set of actions is necessary: (i) detection of wires and cables; (ii) analysis of wires and cables; (iii) detection and classification of electric towers; (iv) analysis of tower components (insulators, hanging points). In more details, the detection of wires and cables is inspired by Candamo et al. [8], whose method builds a feature map on the basis of a pixel motion estimation, the morphological properties and the linear patterns, computed via a multi-window

Hough space. In our setting, the detection will also benefit of the thermal imaging and 3D fitting, in order to increase the robustness against scene complexity, occlusion, noise, and environmental clutter. Once wires and cables are correctly detected, the thermal data and the high resolution images are used to assess their health status. Few related works can be found about detection and classification of electric towers, apart from some preliminary results in [9]. The most recent approaches, such as [10], consider this task as a supervised learning problem. In particular, we will use a multi-layer perceptron neural network both to predict whether the region inside an image is a tower (or not), and to distinguish the tower type on the basis of a dataset of training. A rich dataset of images is needed to train the neural networks. The correct classification of the tower would improve the performance of the automatic detection and analysis of the tower sub-components (i.e., insulators and hanging points), which will use a region-based segmentation and template matching.

Scene Acquisition and 3D Reconstruction. As anticipated in Sect. 3.1, the physical status of the surface of wires and insulators can be evaluated by relying on Computer Vision techniques. However, there are parameters that require accurate geometric measurements, such as checking for proper hanging of the wire by modeling it as a mathematical function (i.e., the *catenary curve*), and ensuring that the space surrounding the wire is clear of obstructions. SCIADRO will use Computer Graphics techniques, to accurately acquire a representation of the 3D scene, which can be analyzed to detect anomalies. In fact, a 3D representation of the aerial power lines and of the surrounding environment are generated, and analyzed by *photogrammetry* algorithms, or through active acquisition devices (e.g., UAV-mountable 3D scanners, namely LiDAR devices). The target area is *sampled* by several UAVs, and each sample produces a 3D georeferenced point, by collectively creating a so called *point cloud*. By analyzing the point density in the cloud and combining it with the path of UAVs, the next navpoints can be estimated, allowing for the reconstruction of so-far unknown areas of the scene. The integration of such an algorithm, known as *next best view* selection, will produce, as time passes, a better and better reconstruction of the scene. Once enough data has been collected, a combination of Computer Vision and Computer Graphics techniques will be used to detect the aerial power lines. At first, points are classified based on their confidence in belonging to an elongate structure (i.e., a wire) by analyzing the anisotropy of the spatial distribution of the neighboring points with the eigenvalues of the covariance matrix of the 3D coordinates. An experimental plug-in has been developed in MeshLab [11] to harness the algorithmic correctness (see Fig. 2). By observing that a catenary seen from above is a straight line, a Hough Transform [12] is applied to the projection of candidate points onto the ground plane; each detected segment is then projected in its containing vertical plane and a RANSAC fitting procedure is applied to compute the parameters of the catenary equation that will be used to test whether the cable is hanging properly. But, even if the wire catenary itself complies to construction constraints, obstructions such as tree branches
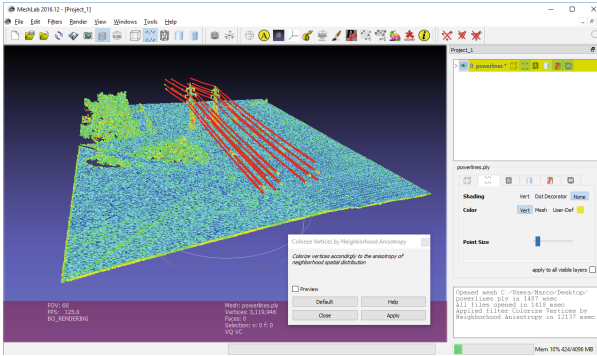
Fig. 2. Powerline 3D points classification in the MeshLab plug-in.

should trigger a maintenance action. To cope with these scenarios, 3D analysis will exploit technologies such as voxel coloring [13] to check whether power lines have a surrounding adequately clear of obstructions. Even if an automatic analysis is possible, it is often very useful to make the surveyor or the operator aware of the swarm configuration and the actual condition of the scene of interest. Hence, the need for a visualization tool that is able to cope with multiple video stream and present them to the user in an effective way. The problem of assembling and presenting multiple views of the same scene has been tackled for still images in [14] and for video streams in [15,16], with the goal of generate smooth transitions from one point of view to the other. Building on these ideas, in the context of SCIADRO, we are currently developing a visualization tool that helps the user understand the current spatial configuration of the swarm and assist her in inspecting the scene with a constrained navigation metaphor.

3.2   Sensing the Crowd with UAV Swarms

Even if the current regulation limits using UAVs in urban and populated areas, such limitations may be mitigated in the near future, thus opening to some possible application scenarios where UAV swarms could be employed. According to the latter, in SCIADRO it is investigated how swarms can be employed to achieve crowdsensing tasks. The term *crowdsensing* is referred to the possibility of collecting data produced by people devices in urban (or rural) areas. In fact, the great majority of today's smart devices provide advanced sensing and computational capabilities that can be exploited to collect data generated directly from the crowd. The ParticipAct living lab [17] must be cited among the successful crowdsensing campaigns, which was designed to gather environmental and multimedia data generated by smartphones or smartwatches (i.e., temperature, noise intensity, user surveys, pictures or movie clips). We plan to employ an UAV swarm as mobile agents that collect specific kinds of data from people's devices, by flying over the region of interest. Such devices already have the possibility of a direct interaction, giving rise to the so-called Mobile Social Networks (MSN) [18].
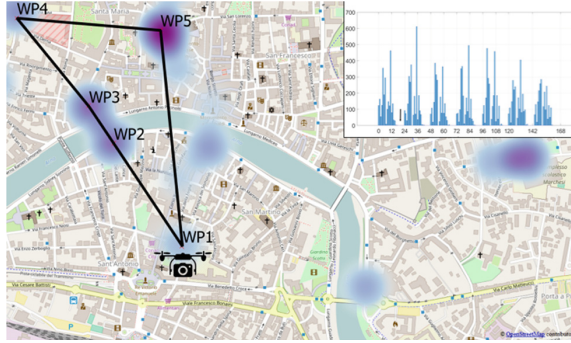
Fig. 3. The crowdsensing scenario under consideration.

UAVs could be analogously seen as peers of a MSN. In fact, they can be equipped with short range communication interfaces in order to join the MSN formed on the ground, similarly to other devices. In this context, UAVs act as mobile collectors of data. The design of optimized trajectories is a key element that allows the swarm to efficiently overfly the most crowded locations. Therefore, understanding the way the people move and the way the people interact is crucial for an efficient use of UAVs in crowdsensing scenarios.

Figure 3 shows the map of Pisa with an overlay heatmap of the most crowded areas. The UAV overflies a number of way-points (WP1 to WP5), since we learned from history of users' mobility that those areas become crowded at specific times. In order to measure the effectiveness of the proposed trajectory, a methodology is necessary to measure the amount of data that an UAV can collect. To this purpose, we propose some metrics useful to determine: (*i*) the quality of the interactions among the users; (*ii*) the fraction of the users *sensed* by the UAV. In particular, we are interested in measuring the number of interactions of the UAV swarm with other devices (also referred to as contact number), and the duration of such interactions, also referred to as *contact duration*. The inset in the right side of Fig. 3 shows the average number of hourly contacts of the UAV with other ground devices on a hourly basis. We can compare such values to those obtained with UAV trajectories built without any notions of the human sociality. We refer to the last ones as *social-oblivious* paths. Preliminary results show a remarkable increase on the number of UAV swarm contacts through social-aware trajectories, by reducing the flight time.

3.3 MP-RTP-Based Multimedia Data Transmission from UAVs to Ground

In this project, also the use of multipath techniques is of interest, in order to deliver quasi real-time multimedia data from one or more UAVs toward a terrestrial Ground Control Station (GCS). The data link is assumed separated from the C2 link for safety reasons.

We focus on the more challenging case of a live video rather than still frames, which require less bandwidth. An UAV can experience bad channel conditions towards a GCS due to obstacles or channel fading; furthermore some Doppler effect must be taken into account, in case of high speeds. Moreover, bandwidth requirements must be tackled since multimedia flows are bandwidth-eager, but broadband links can be very expensive. We propose the use of MPRTP (Multi-Path Real-time Transport Protocol) to support High Throughput (HT) multimedia applications.

Multiple physical links are aggregated by MPRTP, in order to provide an efficient, reliable and cost-effective HT logical channel. Having more than one channel available at any time, the better performing channel (or channels subset) can be used, excluding (or limiting the use of) bad performing ones. The use of MPRTP provides a way to *schedule* the transmission of a multimedia flow over multiple links, according to the implemented scheduling policy. Figure 4 sketches the use of MPRTP between an UAV and its GCS in the proposed scenario. A preliminary implementation of MPRTP protocol is already available, built upon the open source multimedia framework GStreamer. The scheduler is the core module, highlighted in Fig. 4 as part of the MPRTP implementation. The scheduler core functionalities can be summarized as: (*sender side*) splitting of the multimedia flow in n sub-flows, and transmission of the i-th subflow on the i-th physical link; (*receiver side*) aggregation of the n subflows to reconstruct the original multimedia flow, and transmission of feedback data (RTCP).
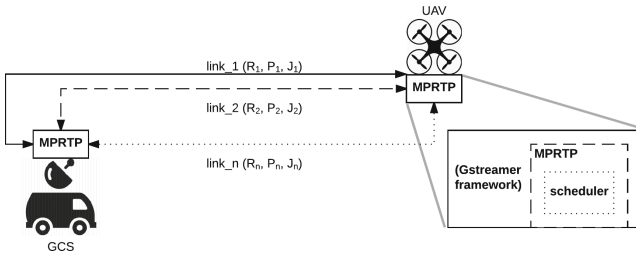
Fig. 4. An exemplary scenario involving the use of MPRTP for multimedia transmissions from an UAV to a GCS.

The current scheduler implementation estimates the overall available bandwidth, and splits the multimedia flow, accordingly. The source video bitrate is then set, by providing the maximum achievable video quality, according to the available bandwidth. However, the current implementation does not provide a way to reduce the load on lossy and/or high-delay sub-links, thus causing corrupted video at the receiver.

We are currently implementing a modified version of the scheduler, which reacts in real-time to time-varying channel statistics, and that is able to properly choose subset of the physical links. Figure 4 shows how our scheduler reacts to Round-Trip Time (RTT) R_i, packet loss rate P_i, and jitter J_i variations on the

i-th link. A preliminary demonstration of what our implementation provides is available at wnlab.isti.cnr.it/ncmprtp.

4   Conclusions

In this work, we presented an overview of the activities currently ongoing in the SCIADRO project. The use of UAVs and UAV swarms is under investigation, and several applications have being taken into account, as exemplary use cases of interest. While the use of single drones is quite established, the use of swarms still require some investigations, and the SCIADRO project will shed some lights on still open issues and possible solutions.

References

1. Daponte, P., De Vito, L., Mazzilli, G., Picariello, F., Rapuano, S., Riccio, M.: Metrology for drone and drone for metrology: measurement systems on small civilian drones. In: IEEE Metrology for Aerospace, pp. 306–311. IEEE (2015)
2. Bacco, M., Caviglione, L., Gotta, A.: Satellites, UAVs, vehicles and sensors for an integrated delay tolerant ad hoc network. In: Bisio, I. (ed.) PSATS 2016. LNICST, vol. 148, pp. 114–122. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47081-8_11
3. Bacco, M., Ferro, E., Gotta, A.: UAVs in WSNs for agricultural applications: an analysis of the two-ray radio propagation model. In: IEEE SENSORS, pp. 130–133. IEEE (2014)
4. Bekmezci, I., Sahingoz, O.K., Temel, Ş.: Flying ad-hoc networks (FANETs): a survey. Ad Hoc Netw. **11**(3), 1254–1270 (2013)
5. Matese, A., Toscano, P., Di Gennaro, S.F., Genesio, L., Vaccari, F.P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., Gioli, B.: Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. Remote Sens. **7**(3), 2971–2990 (2015)
6. Adams, S.M., Friedland, C.J.: A survey of unmanned aerial vehicle (UAV) usage for imagery collection in disaster research and management. In: 9th International Workshop on Remote Sensing for Disaster Response, p. 8 (2011)
7. Menouar, H., Guvenc, I., Akkaya, K., Uluagac, A.S., Kadri, A., Tuncer, A.: UAV-enabled intelligent transportation systems for the smart city: applications and challenges. IEEE Commun. Mag. **55**(3), 22–28 (2017)
8. Candamo, J., Kasturi, R., Goldof, D., Sarkar, S.: Detection of thin lines using low-quality video from low-altitude aircraft in urban settings. IEEE Trans. Aerosp. Electron. Syst. **45**(3), 937–949 (2009)
9. Tilawat, J., Theera-Umpon, N., Auephanwiriyakul, S.: Automatic detection of electricity pylons in aerial video sequences. In: International Conference on Electronics and Information Engineering, vol. 1, pp. 342–346, August 2010
10. Sampedro, C., Martinez, C., Chauhan, A., Campoy, P.: A supervised approach to electric tower detection and classification for power line inspection. In: International Joint Conference on Neural Networks, pp. 1970–1977, July 2014

11. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G.: MeshLab: an open-source mesh processing tool. In: Eurographics Italian Chapter Conference. The Eurographics Association (2008)
12. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**(1), 11–15 (1972)
13. Seitz, S.M., Dyer, C.R.: Photorealistic scene reconstruction by voxel coloring. Int. J. Comput. Vis. **35**(2), 151–173 (1999)
14. Brivio, P., Benedetti, L., Tarini, M., Ponchio, F., Cignoni, P., Scopigno, R.: Photocloud: interactive remote exploration of large 2D–3D datasets. IEEE Comput. Graph. Appl. **33**(2), 86–96 (2013)
15. Waschbüsch, M., Würmlin, S., Gross, M.: 3D video billboard clouds. Comput. Graph. Forum **26**(3), 561–569 (2007)
16. Lipski, C., Klose, F., Magnor, M.: Correspondence and depth-image based rendering: a hybrid approach for free-viewpoint video. IEEE Trans. Circuits Syst. Video Technol. **24**(6), 942–951 (2014)
17. Cardone, G., Cirri, A., Corradi, A., Foschini, L.: The participact mobile crowd sensing living lab: the testbed for smart cities. IEEE Commun. Mag. **52**(10), 78–85 (2014)
18. Girolami, M., Chessa, S., Caruso, A.: On service discovery in mobile social networks: survey and perspectives. Comput. Netw. **88**, 51–71 (2015)

CASG Workshop

A Comparative Assessment of Embedded Energy Storage and Electric Vehicle Integration in a Community Virtual Power Plant

Oghenovo Okpako[1(✉)], Haile-Selassie Rajamani[2], Prashant Pillai[3], Ugonna Anuebunwa[1], and K. Shanti Swarup[4]

[1] Faculty of Engineering and Informatics, University of Bradford, Bradford, UK
{ookpako,U.R.Anuebunwa}@bradford.ac.uk
[2] Faculty of Engineering and Information Sciences, University of Wollongong, Dubai, UAE
HaileRajamani@uowdubai.ac.ae
[3] Faculty of Technology, Design and Environment, Oxford Brookes University, Oxford, UK
ppillai@brookes.ac.uk
[4] Electrical Engineering Department, Indian Institute of Technology, Madras, India
swarup@ee.iitm.ac.in

Abstract. Among the key objectives of the smart grid technology are to foster the grid integration of renewable energy as well as market participation of domestic energy consumers through demand response program. Energy storage remains a key component of the smart grid. Past works on integration of energy storage at the domestic side of the electricity grid has identified the electric vehicle technology (EV) and the embedded energy storage (EES) technology, etc. However, it was difficult to compare between these technologies in terms of business incentives and technical performance. This was investigated in this work, and the results are presented. It was propose to use percentage difference to compare between VPP with EES and VPP with EV. The results shows that the difference in prosumers incentives between VPP with EES and VPP with EV is very low. It is approximately 0.89%. However, the percentage difference in VPP operator profit between VPP with EES and VPP with EV is very high. It is approximately 85.3%. The VPP makes very high profit in the VPP EES case compared to VPP EV case. The same also applies to the VPP cumulative performance where the percentage difference in the VPP cumulative performance between VPP with EES and VPP with EV is approximately 10.9%. This has implication on the storage mechanism to be integrated in to a VPP at the domestic level as well the business model to be adopted.

Keywords: Prosumer · Battery · Virtual Power Plant (VPP)
Genetic Algorithm (GA) · Smart grid · State of charge

1 Introduction

Sustainable energy production and the efficient utilization of available energy resources remains one of our greatest challenge towards minimizing carbon emission from fossil

fuel. In the UK, the government is keen to reduce its carbon footprint by 80% by the year 2050.

Energy storage is a key option in tackling the global challenge of climate change, as it can help in increasing the use of clean energy from intermittent renewable energy sources. Grid integration of energy storage would help in improvement of power quality; provision of peak shaving and valley filling services; deferral of investment on grid's enhancement; demand side management; spinning reserve; black start services; and provision of load following services, etc. [1–7].

The ongoing global restructuring of electric power utilities is expected to create a competitive market. The consumer role is envisage to change to that of a prosumer who both consume and produce energy. Energy storage becomes an asset to the prosumer, which can be use for selling demand response in the power market. The use of energy storage at the domestic (prosumer) side of the electricity grid can be in form of embedded energy storage (EES), and electric vehicle (EV). In EES, the storage is place in the house of the prosumer and is stationary. In EV, the storage is inside the vehicle and can only be connected to the grid when the vehicle is idle. EV can be classify as a mobile storage. The use of EV for participation in the power market and for providing grid support is known as vehicle to grid (V2G). Kempton proposed this concept [8–10].

Due to the small power capacity of EES and EV, prosumers cannot participate directly in the power market, as this is done at the wholesale level. Prosumers would require the service of a third party called the Virtual Power plant (VPP). The VPP is an aggregator and a third party agent that combine large number of units of prosumers EVs, EESs and other energy resource device for participation in the wholesale power market. Participation of the VPP in the power markets, involves the VPP establishing contract with EV's and EES's owners to allow them use their storage device for energy/market transaction while providing support to the grid support. A prosumer who allows the VPP to use their storage for market participation gets a financial compensation.

The business potentials of EES as part of a VPP have been researched by these authors: [11–13]. In [11], EES can maximize prosumers feed-in tariff reward. Also, multiple power service provision by EES can increase the financial value of EES under a time of use tariff as well as a dynamic use of system pricing [11]. According to [12, 13] on EES, it is possible to have a pricing regime that simultaneously incentivize both the VPP operator and the prosumer while meeting the VPP objectives. In addition, the business potentials of EVs in the power market have been investigated by these authors [10, 14, 15]. According to [10], EV owners would get more financial incentives when there is capacity payment for provision of ancillary service.

Intelligent control strategies has been proposed as way to deal with large scale integration of energy storage in to the grid [16, 17]. This is to prevent increase in load at transmission, sub transmission and distribution level [18], need for an additional investment in grid infrastructure as a result of prosumers discretion for fast charging and charging time, as well as the coincidence between peak load of EV and non-EV on grid's reliability [19].

While existing literatures on grid integration of EES and EV has explored its business potential and its energy management strategy. There is no work known by this author, which has compared between the EES and EV in terms of technical performance for

provision of dynamic load levelling and the business case. There is no clarity on whether EES offers a better business incentives and technical performance compared to EV vis-à-vis. This was investigated in this work.

This paper is organized as follows; Sect. 2 is a description of the model, Sect. 3 is the mathematical modelling, Sect. 4 is the genetic algorithm optimization implementation, Sect. 5 is the developed cumulative performance index for measuring VPP technical performance, Sect. 5 is the simulation parameters, Sect. 6 is the results and discussion, and Sect. 7 is the conclusion.

2   Framework of Virtual Power Plant Model

In this work, the VPP model was considered under two scenarios of storage integration. These includes: EES and EV.

2.1   Embedded Energy Storage Scenario

In EES, each of the prosumers where assumed to have battery storage embedded inside their house. The batteries are aggregated in to the VPP, and are use for participation in the power market at the wholesale level. Figure 1, is the diagram showing the VPP model under EES scenario.
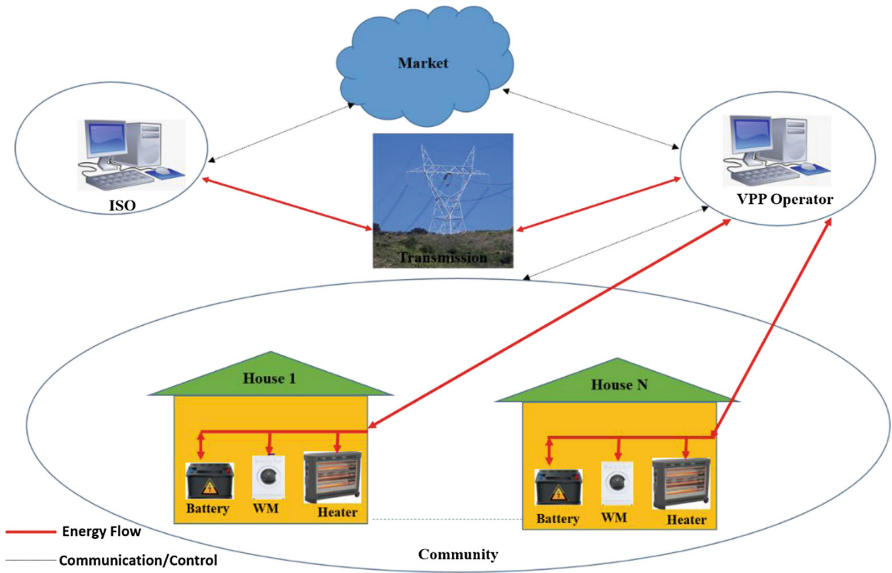
Fig. 1.   Architecture of the Virtual Power Plant model (using EES).

In Fig. 1, N is the total number of houses within the community aggregated in to the VPP. Each house is own by a prosumer. Each prosumer sells energy to the VPP operator through discharging of their battery. In this work a community consisting of three

prosumers is investigated (N = 3). Each prosumer buy energy from the VPP operator to charge its battery as well as meet its load demand. The prosumers load can consist of a washing machine (WM), electric heater, etc. as shown in Fig. 1. The day ahead hourly load profile of each prosumer is shown in Fig. 2 as follows.
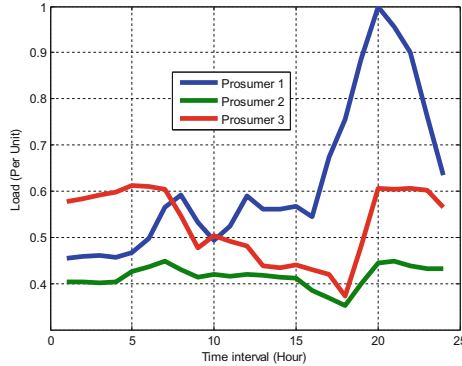
Fig. 2. Forecasted hourly load profile of each prosumer.

The hourly load data was obtained from Xcel Energy [20]. In Fig. 3, each of the prosumer has a different hourly load profile. The motivation of the prosumer as a participant in its local community VPP is to get incentive.
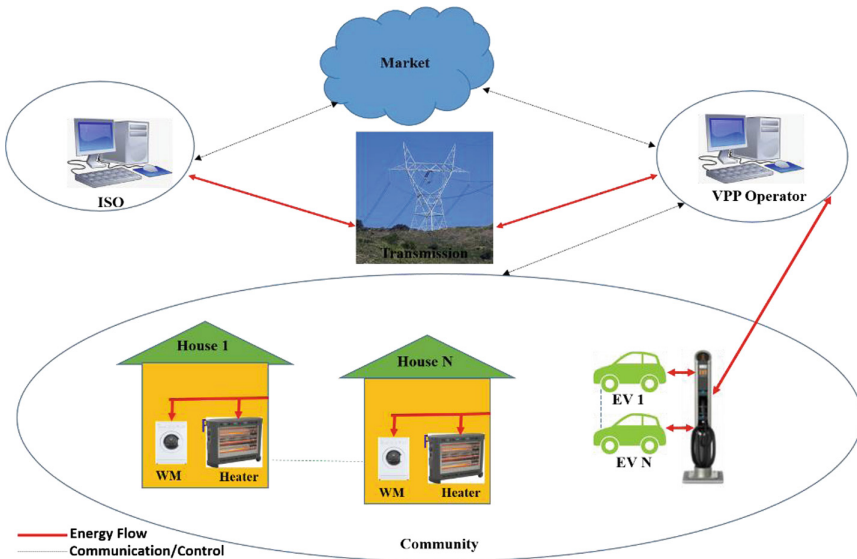
Fig. 3. Architecture of the Virtual Power Plant model (using EV).

The VPP operator can buy energy from the grid and from the prosumers. The energy bought from the grid is use to meet the prosumer's energy demand respectively. The VPP buys power in bulk from the power market to meet its prosumer's load demand, as well as for charging the prosumers. In this model, the VPP can combine both energies from the grid and the prosumer's battery to meet the load demand of each prosumer respectively. The energy bought from each prosumer's battery are aggregated by the VPP. The aggregated energy is first use within the community to meet each prosumer's load demand respectively before it can be traded in the power market (exported to the external grid) by the VPP on behalf of the prosumers.

In this work, the VPP operator was considered as having a day ahead forecast of each prosumer hourly load profile respectively. In addition, the VPP operator has a day ahead forecast of the price at which the external grid would buy its energy (i.e. the day ahead forecast price paid by the grid to the VPP for exporting energy), as well as the day ahead forecast of the price at which the grid would sell energy (i.e. the day ahead forecast price paid by VPP to the external grid for importing energy). Both import and export prices for energy are agreed between the VPP operator and the grid in the wholesale power market. Based on the day ahead import and export price, the VPP operator agrees a day ahead prosumer buy and sell price of energy. Thereafter, the VPP operator has to optimally allocate energy resource by determining the day ahead schedule assuming no error band during forecasting. The day ahead energy resource allocation is done by determining the day ahead charge/discharge energy from each prosumer battery. The charge/discharge energy is use to control the amount of energy to be imported from grid and exported to the grid, and also to provide the balancing services required by the ISO (independent system operator). The financial reward for encouraging the VPP operator is profit.

2.2   Electric Vehicle Scenario

In this model, the prosumer uses its EV for participation in the power market. The difference between this model and that of the EES is that the battery is sometimes mobile. EV can only be use for market participation when it is idle. Figure 3, is the diagram showing the VPP model under EV.

In Fig. 3, each EV is labelled in accordance to the house number. Each house is own by a prosumer. The VPP combines in cloud each prosumer EV. The EV battery storage is regarded as virtual storage. This is because even though the vehicle is elsewhere within the community, it is still virtually connected to the VPP and the prosumer's house through smart plug and communication network. This allows the VPP to have information regarding the energy use by the EV for driving, as well as the energy traded and exchange between the VPP operator and prosumer's EV. The VPP uses the EV to participate in the wholesale power market only at time period when the prosumers plugged in their EV via the EVSE (Electric Vehicle Supply Equipment). The plugging operation can be done at the home and work location of the prosumer. The prosumer's home and work location are within the same community controlled by the VPP. During time period when the prosumers are driving their cars within their community (i.e. from home to work location, and from work location back to home) they are still virtually

connected to the VPP, however during that time period the VPP cannot make use of the prosumers battery for energy transaction. At that time period the VPP can only transact energy based on the house load (domestic load) of the prosumer. Each prosumer house load pattern are the same as that of the EES scenario in Fig. 2. The driving load profile of the prosumer is shown in Fig. 4, as follows.
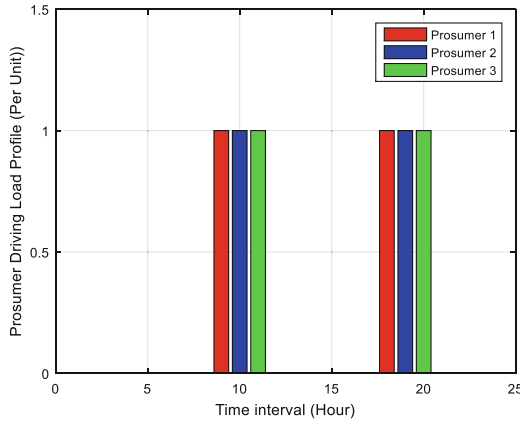

Fig. 4. Forecasted hourly driving load profile of each prosumer.

Figure 4, is the driving load profile of each prosumer. It shows the energy consumed (discharged) from each prosumers EV during time period when the EV is been driven to work and back home. Each prosumer where assumed to have different driving pattern. In this work, it was considered that each prosumer EV is not available for VPP operation in terms of energy and business transaction for approximately 8% of the day total number of time interval (24 time interval).

3   Mathematical Tool Used in Modelling

3.1   The VPP Net Dynamic Load

The VPP net dynamic load at time interval t on the grid is calculated in (1) as follows.

$$E_{g_t} = \sum_{i=1}^{N} \left(L_{i,t} + \alpha_{i,t} \cdot Ecd_{i,t} \right) \tag{1}$$

$$Ecd_{i,t} = \begin{cases} E_{c_{i,t}}, & if\ charging\ occur \\ -E_{d_{i,t}}, & if\ discharging\ occur \\ 0, & if\ battery\ is\ idle \end{cases}$$

$$\alpha_{i,t} = \begin{cases} 1, & for\ VPP\ with\ EES \\ [0,1], & for\ VPP\ with\ EV \end{cases}$$

Where N is the total number of prosumers. $\alpha_{i,t}$ is the connection state of the prosumer's storage, which is either 0 or 1. For VPP with EES, $\alpha_{i,t}$ is always 1, this is because EES is stationary and always connected to the grid. For VPP with EV, $\alpha_{i,t}$ is 1 when EV is plugged in (idle) at t, and 0 when EV is mobile at t. $E_{imp_{i,t}}$ and $E_{exp_{i,t}}$ are the amount of import and export energy in per unit respectively. $Ecd_{i,t}$ is the amount of energy exchange in per unit between prosumer i battery and the VPP at time interval t. $E_{c_{i,t}}$ and $E_{d_{i,t}}$ are the amount of charge and discharge energy in per unit respectively allocated to prosumer i battery. $L_{i,t}$ is the load of prosumer i in per unit at t. If E_{g_t} is greater than zero, the VPP is importing energy, and if less it is exporting energy.

3.2 VPP Operator Profit

The VPP operator profit Vpp_{profit}, at each time interval t over the day's total number of time interval (T) is calculated as follows.

$$\sum_{t=1}^{T} Vpp_{profit_t} = \sum_{t=1}^{T} \left(Vpp_{rev_t} - Vpp_{\cos t_t} \right) \tag{2}$$

Where Vpp_{rev_t} and $Vpp_{\cos t_t}$ are the VPP revenue and cost respectively during the time interval t. T is the day's total number of time interval. Both VPP revenue and cost are calculated in (3) and (4) as follows.

$$\sum_{t=1}^{T} Vpp_{rev_t} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\begin{array}{c} Pp_{buyt} \cdot (L_{i,t} + \alpha_{i,t} \cdot E_{c_{i,t}}) \\ + Pv_{exp_t} \cdot E_{exp_{i,t}} \end{array} \right) \tag{3}$$

$$\sum_{t=1}^{T} Vpp_{\cos t_t} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Pp_{sell_t} \cdot \alpha_{i,t} E_{d_{i,t}} + Pv_{imp_t} \cdot E_{imp_t} \right) \tag{4}$$

Where Pp_{sell_t}, Pp_{buy_t}, Pp_{imp_t}, and Pp_{exp_t} are the prosumer selling price of energy to the VPP, prosumer buy price of energy from the VPP, the VPP import price of energy, and the VPP export price of energy respectively at t. These prices are measured in pence/per unit. $\alpha_{i,t}$ is the connection state of the storage, and is already explained in detail in Sect. 3.1.

3.3 Prosumer Net Cost

The prosumer's net cost Pp_{cost}, at each time interval t over the day's total number of time interval T is calculated as follows:

$$\sum_{t=1}^{T} Pp_{\cos t_t} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\begin{array}{c} Pp_{buy_t} \cdot \left(L_{i,t} + \alpha_{i,t} \cdot E_{c_{i,t}} \right) \\ - Pp_{sell_t} \cdot \alpha_{i,t} \cdot E_{d_{i,t}} \end{array} \right) \tag{5}$$

3.4 Battery State of Charge

The battery state of charge (SOC) gives an information on the battery energy level. In this work, the battery energy level is measured in per unit. Usually the battery SOC cannot be measured directly, but can be inferred from the battery energy level. Therefore, the battery state of charge of charge is a measure of the battery energy level in comparison to the battery actual capacity, assuming an ideal battery with no peukert effect, no losses (self-discharge) and whose actual capacity is the same as its nominal capacity. The SOC is measured in percentage. It gives an information on the battery depth of discharge. The battery energy level measured during t is calculated as follows.

$$E_{stored_{i,t}} = E_{o_i} + \sum_{t=1}^{T} Ecd_{i,t} \tag{6}$$

$E_{stored_{i,t}}$ is prosumer i battery energy level in per unit measured at t. E_{o_i} is prosumer i initial battery energy level in per unit before participation in the day ahead power market. Each prosumer battery SOC at t is calculated as follows.

$$SOC_{i,t} = 100 \frac{E_{stored_{i,t}}}{E_{batt_i}} \tag{7}$$

$SOC_{i,t}$ is the state of charge of prosumer i battery measured in percentage at t. E_{batt_i}, is the actual battery capacity in per unit of prosumer i.

3.5 Battery Constraints

Each prosumer battery discharge constraint is represented as follows.

$$E_{d,\min_i} \leq E_{d_{i,t}} \leq E_{d,\max_i} \tag{8}$$

Where $E_{d,\min_i}$ and $E_{d,\max_i}$ are the minimum and maximum discharge energy that can be allocated to prosumer i battery. Each prosumer battery charge constraint is represented as follows.

$$E_{c,\min_i} \leq E_{c_{i,t}} \leq E_{c,\max_i} \tag{9}$$

Where $E_{c,\min_i}$ and $E_{c,\max_i}$ are the minimum and maximum charge energy that can be allocated to prosumer i battery. Each prosumer battery state of charge constraint is represented as follows.

$$SOC_{\min_i} \leq SOC_{i,t} \leq SOC_{\max_i} \tag{10}$$

Where $SOC_{\min_i}$ and $SOC_{\max_i}$ are the minimum and maximum SOC limit of prosumer i battery.

4   Optimizations of the Community Virtual Power Plant

To understand the optimization problem, the number of prosumer chosen to participate in the community VPP was kept at three. The optimization function is the prosumer net cost. This is gotten from (5) and is represented as follows.

$$[Min]F = \sum_{t=1}^{T} Pp_{\cos t_t} \tag{11}$$

F is the objective function to be minimize. It represents both the prosumer net cost. Genetic algorithm was used to solve the optimization problem [12, 13].

5   Cumulative Performance Index for Measuring VPP Technical Performance

The grid requires energy balancing service from the VPP. Energy balancing service is provided for by the VPP through peak and off-peak service. During off-peak period, the grid requires the VPP to increase its load, by importing energy from the grid to meet its prosumers load as well charging of battery. During peak period, the grid requires support from the VPP. The VPP can support the grid by discharging of its prosumer battery to meet the prosumers load demand, and also by exporting energy to the grid. The grid energy balancing need are reflected by the prices. More clarity on this is presented in [21].

The VPP performance is determined by comparing both the VPP dynamic load and the energy balancing need of the grid.

The VPP dynamic load at t, E_{g_t} is the energy imported from the grid by the VPP or the energy exported to the grid by the VPP at t. The energy balancing need of the grid at t is determined by the exchange price (δ_t) at time t. This is calculated as the difference between the import and export price of electricity at t. This is mathematically represented as follows:

$$\delta_t = P_{imp_t} - P_{\exp_t} \tag{12}$$

$$\begin{cases} if \ \delta_t > 0, \ grid \ requires \ off-peak \ service \\ if \ \delta_t < 0, \ grid \ requires \ peak \ service \end{cases}$$

When δ_t is positive, the grid requires the VPP to provide off-peak service by increasing its load. The VPP can increase its load by importing energy from the grid. When δ_t is negative the grid requires the VPP to reduce its load. The VPP can reduce its load by discharging the prosumer's battery to support the load, as well as to export energy to the grid. In this work, when energy is imported at t, E_{g_t} (from (1)) is greater than zero. When energy is exported at t, E_{g_t} is less than zero. Both δ_t and E_{g_t} are represented with the logic input A_t and B_t respectively, such that:

$$\begin{cases} \text{if } \delta_t > 0, & A_t = 1 \\ \text{if } \delta_t < 0, & A_t = 0 \\ \text{if } E_{g_t} > 0, & B_t = 1 \\ \text{if } E_{g_t} < 0, & B_t = 0 \end{cases}$$

If grid requires off-peak service, it means the dynamic load must be increased. On the other hand, if the grid requires peak service, the dynamic load must be decreased. Based on this criteria, a performance state C_t at time t is derived as follows:

$$C_t = \overline{(A_t \oplus B_t)} \tag{13}$$

C_t is the output of an EX-NOR logic combination of inputs A_t and B_t. Based on C_t, a new performance index called the "Cumulative Performance Index" of the VPP (CPI) over the day is computed. *CPI* over *T* is formulated from (13) as follows (Table 1):

$$CPI = \frac{100 \sum\limits_{t=1}^{T} C_t}{T} \tag{14}$$

Table 1. Simulation Parameters

Parameters	Specification
Ebatt	24 per unit
E_o	12 per unit
Initial *SOC*	50%
$E_{d,min}$	0 per unit
$E_{d,max}$	1 per unit
$E_{c,min}$	0 per unit
$E_{c,max}$	1 per unit
T	24
N	3

6   Results and Discussion

Figure 5, is the modified pricing used in this model. This pricing scheme allows both the VPP and the prosumer to get business incentives.
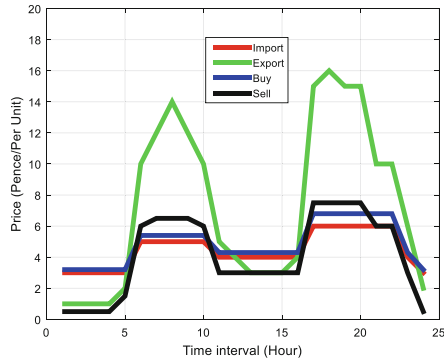
Fig. 5. Modified pricing scheme.

In Fig. 5, it is seen that the prices are set based on the grid requirement for dynamic load levelling. This pricing is detailed in [12, 13]. GA was used to optimized the energy transaction of the community. Figure 6, is the cumulative prosumers incentive in both VPP with EES and VPP with EV scenario.
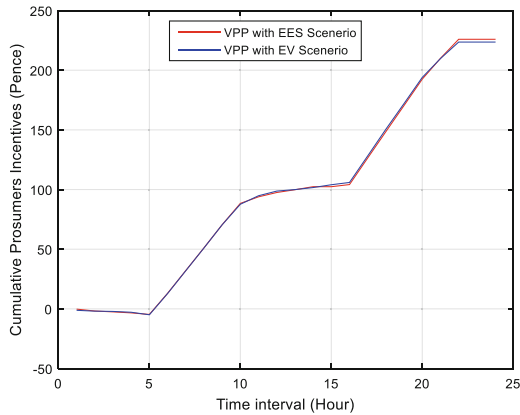
Fig. 6. Cumulative prosumers business incentives.

In Fig. 6, it is seen that the prosumers get incentive under both scenario. In the VPP with EES scenario, the cumulative prosumers incentive was approximately 226 pence. In the VPP with EV scenario, the cumulative prosumers incentive was approximately 224 pence. The prosumers incentive is slightly higher for the VPP with EES scenario scenario as compared to the VPP with EV scenario. The higher incentive for VPP with EES is because the battery storage is 100% available (connected to the grid) for the prosumer to use for energy and business transaction with VPP. Using the percentage difference, the difference in prosumers incentives between VPP with EES and VPP with

EV is very low. It is approximately 0.89%. This is attributed to both the objective function and the pricing regime which favors the prosumers in both scenario. Figure 7, is the cumulative VPP profit for both VPP with EES and VPP with EV scenario.
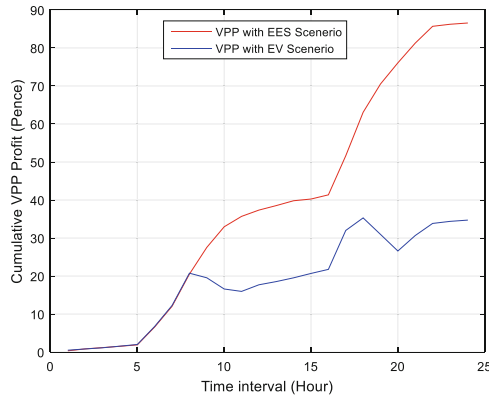
Fig. 7. Cumulative VPP business incentives

In Fig. 7, it is seen that the VPP makes profit on both scenario. In the VPP with EES scenario the VPP makes a profit of approximately 87 pence. In the VPP with EV scenario the VPP makes a profit of approximately 35 pence, which is about 40% of what it makes in the EES scenario. The VPP operator profit under the EES scenario is much higher compared to that of the EV scenario. The percentage difference in VPP operator profit between VPP with EES and VPP with EV is approximately 85.3%. This is unlike the prosumers incentives where the difference is low (approximately 0.89%). This high difference in VPP profit is due to VPP inability to fully utilize the prosumers battery for market participation during time periods when the EV owners is mobile; the battery discharge energy use for mobile activities; and the optimization function which favor's the prosumers incentive.

Figure 8, is the cumulative performance of the under both VPP with EES and VPP with EV scenario.

In Fig. 8, it was observed that the VPP performance under the VPP with EES scenario is approximately 88%. The VPP performance under the VPP with EV scenario is approximately 79%. The percentage difference in the VPP cumulative performance between VPP with EES and VPP with EV is approximately 10.9%. The VPP performs better with EES compared to EV because in the EES case scenario, the battery is always available to the VPP, and can be use by it for dynamic load levelling. Unlike the EV scenario, each of the prosumer's EV is only connected to the grid via the VPP for only 92% of the day total number of time interval (22 h out of 24 h). The remaining time period when the EV is not connected to the grid actually prevents the VPP from being able to provide dynamic load levelling. Thus resulting to a lower performance for the VPP.

In this work, the use of energy storage at the domestic level as part of a community VPP has been compared under two scenarios, which are the EES scenarios and the EV
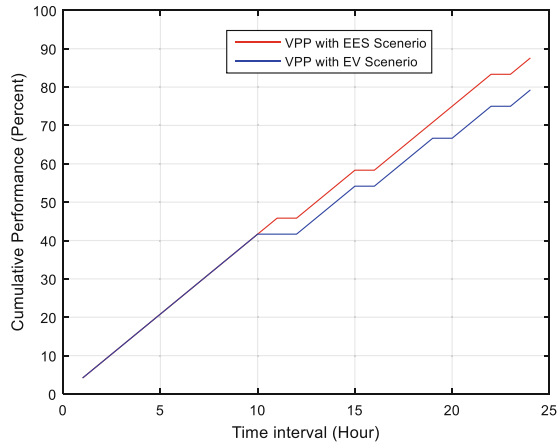
Fig. 8. Cumulative performance of the VPP.

scenarios. Using percentage difference to compare both scenarios, the results clearly demonstrates that the best business case for prosumers may not necessarily mean the best business case for VPP operator. This is because even though the prosumers where allowed to use their battery energy to meet their driving need in the EV scenario, the percentage difference in their incentive as compared to the EES scenario is very low. This is unlike the VPP operator where the percentage difference is much higher between VPP with EES and VPP with EV (VPP operator profit too low for VPP with EV in comparison to VPP with EES). The same also applies to the best technical case for grid operators in terms of their requirement for dynamic load levelling. These factors should be considered by electric utilities and grid operators when selecting suitable business model and method of storage integration for VPP at the domestic level. For clarity purpose, Table 2, is presented as follows.

Table 2. Comparison between VPP with EES and VPP with EV.

	VPP operator profit	Prosumers incentive	Cumulative performance
EES	87 pence	226 pence	88%
EV	35 pence	224 pence	79%
Percentage difference	85.3%	0.89%	10.9%

7   Conclusion

In this study, we have compared a community VPP with EES and a community VPP with EV in terms of business incentive and technical performance. Using the percentage difference, the results showed that the difference in prosumers incentives between VPP with EES and VPP with EV is very low. It is approximately 0.89%. This is attributed to both the objective function and the pricing regime which favors the prosumers. While,

past research works has demonstrated that most private cars are mobile for approximately 4% of the day total number of time interval and are idle for 94%, we have considered EVs which are mobile for 8% of the day's total number of time interval in this work. Using the percentage difference, it is easy to understand that at 4% total number of time interval when EVs are mobile and disconnected from the grid, it is possible to have a minimal negative impact on the prosumers incentive if the prices are set correctly and the objective function is made to favor the prosumer. However, the percentage difference in VPP operator profit between VPP with EES and VPP with EV is very high. It is approximately 85.3%. The same also applies to the VPP cumulative performance where the percentage difference in the VPP cumulative performance between VPP with EES and VPP with EV is approximately 10.9%. Furthermore, there is a current demand towards a paradigm shift from fossil fuel vehicles towards EVs, therefore, it becomes critical to understand that having the best business case for prosumers may not necessarily mean the best business case for VPP operator. The same also applies to the best technical case for the grid operator in terms of its requirements for dynamic load levelling. Using percentage difference in Comparing VPP with EES and VPP with EV helps understand this. Future work would investigate different initial battery conditions and pricing models.

References

1. Chen, H., Cong, T.N., Yang, W., Tan, C., Li, Y., Ding, Y.: Progress in electrical energy storage system: a critical review. Prog. Nat. Sci. **19**, 291–312 (2009)
2. Ferreira, H.L., Garde, R., Fulli, G., Kling, W., Lopes, J.P.: Characterisation of electrical energy storage technologies. Energy **53**, 288–298 (2013)
3. Akhil, A.A., Georgianne, H., Currier, A.B., Kaun, B.C., Rastler, D.M., Chen, S.B.: DOE/EPRI Electricity storage handbook. Sandia national laboratories (2013)
4. Masaud, T.M., Lee, K., Sen, P.: An overview of energy storage technologies in electric power systems: What is the future?. In: North American Power Symposium (NAPS), pp. 1–6 (2010)
5. Boicea, V.A.: Energy Storage Technologies: The Past and the Present. Proc. IEEE **102**(11), 1777–1794 (2014)
6. Medina, P., Bizuayehu, A., Catalão, J.P., Rodrigues, E., Contreras, J.: Electrical energy storage systems: technologies' state-of-the-art, techno-economic benefits and applications analysis. In: 2014 47th Hawaii International Conference System Sciences (HICSS), pp. 2295–2304 (2014)
7. IEA.: Technology roadmap: energy storage. Paris (2014)
8. Kempton, W., Tomic, J., Letendre, S., Brooks, A., Lipman, T.: Vehicle-to-grid power: battery, hybrid, and fuel cell vehicles as resources for distributed electric power in California. Institute of Transportation Studies (2001)
9. Kempton, W., Letendre, S.E.: Electric vehicles as a new power source for electric utilities. Transp. Res. Part D Transp. Environ. **2**(3), 157–175 (1997)
10. Kempton, W., Tomic, J.: Vehicle-to-grid power fundamentals: calculating capacity and net revenue. J. Power Sources **144**, 268–279 (2005)
11. Fei, T., Strbac, G.: Business cases for energy storage with multiple service provision. J. Mod Power Syst. Clean Energy **4**, 615–625 (2016)

12. Okpako, O., Rajamani, H., Pillai, P., Anuebunwa, U., Swarup, K.: Investigation of an Optimized energy resource allocation algorithm for a community based virtual power plant. In: Proceedings of 2016 IEEE PES Power Africa Conference, Livingstone, Zambia, pp. 153–157 (2016)
13. Okpako, O., Rajamani, H., Pillai, P., Anuebunwa, U., Swarup, K.: Evaluation of community virtual power plant under various pricing schemes. In: Proceedings of 2016 IEEE Smart Energy Grid Engineering Conference, Oshawa, Canada, pp. 72–78 (2016)
14. Guille, C., Gross, G.: A conceptual framework for the vehicle-to-grid (V2G) implementation. Energy Policy **37**, 4379–4390 (2009)
15. Jingshan, L., Shiyu, Z., Yehui, H.: Federal and state incentives heighten consumer interest in electric vehicles. In: Advances in Battery Manufacturing, Service, and Management Systems, p. 416. Wiley-IEEE Press (2017)
16. Lopes, J.A.P., Soares, F.J., Almeida, P.M.R.: Integration of electric vehicles in the electric power system. Proc. IEEE **99**, 168–183 (2011)
17. Iria, J., Soares, F., Franchin, I., Silva, N.: Development of a novel management system for electric vehicle charging. In: 2014 IEEE International Electric Vehicle Conference (IEVC), pp. 1–7 (2014)
18. Heydt, G.T.: The impact of electric vehicle deployment on load management straregies. IEEE Trans. Power Appar. Syst. 1253–1259 (1983)
19. Wu, D., Aliprantis, D.C., Ying, L.: Load scheduling and dispatch for aggregators of plug-in electric vehicles. IEEE Trans. Smart Grid **3**(1), 368–376 (2012)
20. Xcel Energy Hourly Load Profile. https://www.xcelenergy.com/staticfiles/xe/Corporate/Corporate%20PDFs/AppendixD-Hourly_Load_Profiles.pdf
21. Okpako, O., Rajamani, H., Pillai, P., Anuebunwa, U., Swarup, K.: A new performance index for evaluating community virtual power plant with domestic storage. In: Proceedings of 2017 IEEE PES General Meeting, Chicago, USA (2017)

Investigating the Impact of Cyber-Attack on Load Profile of Home Energy Management System

Ugonna Anuebunwa[1]([✉]) [ID], Haile-Selassie Rajamani[2], Prashant Pillai[3], and Oghenovo Okpako[1]

[1] Electrical Engineering and Computer Science, University of Bradford, Bradford, UK
{U.R.Anuebunwa,O.Okpako}@bradford.ac.uk
[2] Faculty of Engineering and Information Science, University of Wollongong, Dubai, UAE
HaileRajamani@uowdubai.ac.ae
[3] Faculty of Tecchnology, Design and Engineering, Oxford Brookes University, Oxford, UK
ppillai@brookes.ac.uk

Abstract. Load profile for a household is key to understanding and applying automated load scheduling executed by the Home Energy Management Systems (HEMS). The provision of securing this basic domestic information as well as preventing intruders from being able to accessing and modifying them should be a matter of high priority. Any malicious attack on this data will have serious impact on the performance of the load scheduling algorithms. This paper is an investigation of how the scheduled load profile of a household can be deformed due to false data injection on the original load profile as a result of cyber-attack on the HEMS. Various incremental false data levels are introduced during an optimization process and the corresponding effect on the overall scheduled load profile is evaluated to understand the actual impact of the cyber-attack. Results show that as noise attack level increases, the optimized load profile shrinks and approaches a straight line which is equivalent to the average value of the original load profile. The implication of having such a load profile as a schedule is the obvious excessive disruption of a household's energy use which results to having appliances switched ON or OFF at highly undesired times of the day thereby exacerbating user inconvenience.

Keywords: Cybersecurity · Demand Response
Dynamic pricing *Home* automation · Internet of Things

1 Introduction

On 23rd December 2015, the Ukrainian reginal electricity distribution company was subject to a critical cyber-attack. On this occasion, seven 110 kV and twenty three 35 kV substations were disconnected for three hours from the Ukrainian grid network resulting in a huge economic and societal impact. This attack was attributed to foreign government-sponsored cyber-criminals, who remotely controlled the SCADA distribution management system and caused a blackout for approximately 225,000 customers [1]. A similar incident is the case of cyber-attack in October 2016 when a Distributed Denial of Service (DDoS) attack knocks heating system offline in at least two housing bocks

in the city of Lappeenranta, leaving their residents in subzero weather conditions. In an attempt to fight back the attack which was only short-lived, the automated systems rebooted which unfortunately got stuck in an endless loop that kept restarting and shutting down. This scenario lasted for over a week but returned to normal service by 3rd November afternoon [2]. As much as researchers and scientists are working hard every day to improve living standards as well as efficiency of system designs, criminals are also attempting and finding ways to interfering with these systems thereby sabotaging, thwarting and frustrating their operations. While cyber-attacks are a persistent threat to internet users, they are also becoming a cause for concern in other areas like power grids, healthcare networks, transportation, that are more and more using ICT technologies. The aim of this paper is to investigate the impact of a cyber-attack on Home Energy Management Systems (HEMS) and then propose means of mitigating the disruptions occasioned by these attacks, which is capable of distorting and impacting on the efficiency of the normal operations of a domestic load scheduler. This investigation is carried out by introducing noise signals on the load profile data and then observing the effect on the optimized load profile. A comparison of the outcome with the expected load profile when there is no attack is also made and appropriate deductions carried out.

The energy load profile represents the electrical load consumption of residential, commercial or industrial consumers over a period of time. The energy usage will vary over different days (weekend and weekdays) and also over different seasons (winter and summer). Such load profiles can be used to determine energy allocation and planning benefits depending on how much power is available for distribution as well as where the priority lies during peak and off-peak periods. A forecasted load profile is usually obtained from historical load profile data and it is very useful for planning the schedule for the next day. Using these load profiles, HEMS run load scheduling algorithms to propose a scheduled load usage for consumers for improved grid balancing and reduction of consumer's energy costs. This means that load profiles are invaluable and any successful attack on it can disrupt power supply and network by producing inaccurate forecasted load profiles as well as invalidating the outcome of the proposed load allocation and efficient energy planning schedule. In this paper the forecasted load profile used is hourly-based over a 24-h period and it is computed using moving average forecasting methods. The optimized load profile is obtained using Genetic Algorithm (GA) optimization with input variables such as pricing and occupancy data in order to determine the optimized output whose data is influenced by the amount of noise present at the forecasted load profile.

In any smart home, the attack vector to a load profile data stored in the HEMS is via the communication network that connects the HEMS to the internet as well as through smart appliances and any wireless device which connects the load to the HEMS. This connection is to enable the HEMS obtain pricing data online as they are published by the energy providers, or to communicate with the load. Unfortunately, this exposes the communication link to become a target for an attack of which cyber-criminals could break into. This is where appropriate security design should be enforced because as much as the load profile data is vulnerable; virtually any data on the HEMS can be attacked. However, this paper will limit the investigation on attack on load profile, while attack on other aspects is considered out of scope. Furthermore, it is assumed that the

prices are known a day before but they could be more real time with forecasting being carried out to predict the price for the next 24 h.

2   Related Work

Protection of vulnerable load and other related components of the smart grid from cyber-attack keeps attracting interests from researchers around the globe due to the numerous challenges facing the internet world. The authors in [3, 4] discussed the importance of detecting cyber-attacks in energy consumption data of power systems as provided by smart meters, and suggested schemes for adequate protection. Such attacks on dynamic loads known as: dynamic load altering attacks (D-LAA), was considered because, the possibility to control loads dynamically implies also the possibility to attack loads dynamically [4]. This is in contrast to Static load altering attacks (S-LAA) which is more common and is based on changing the volume of certain vulnerable loads, usually in an abrupt fashion. The paper suggested the detection D-LAAs is possible by applying frequency domain analysis of the load profile using Fast Fourier Transform (FFT) of the original load profiles via spectral analysis [5]. Another detection technique includes Real-time detection in frequency domain using Windowed-FFT (W-FFT), and detection based on both load and frequency signals [6]. The paper suggested optimization problem formulation, solution method and protection system design under uncertainty as approaches towards applying adequate protection schemes to hinder successful attacks on the load data.

As much as these authors have identified means to detecting these attacks on various types of loads which includes FFT analysis, they did not investigate or analyze the impact of such attacks on the load profile. It is acknowledged that it is important to identify that there has been an attack, but the other important issue is to understand what impact such an attack is capable of causing on the infrastructure under protection, before applying any sort of solution. This scenario can be referred to the attack in Finland as mentioned in [1] because a DDoS attack which was only short-loved was identified. But this may not be as bad as requiring a system restart because, restarting the system was what actually kept it in a perpetual reboot mode thereby making it impossible for the heating supply to continue. Therefore this paper is centered in understanding the impact of such attacks first, so that appropriate steps can be taken in attempting to solve the problem which may offer other ways to solving the problem depending on the identified impact on the system.

3   Proposed Approach

For effective demand response (DR) system, the HEMS is designed to retain a record of historical load profiles and generates up-to-date forecasted load profile for daily optimization process using any modern forecasting techniques available. In this work, moving average forecasting technique is applied because it is easier to implement and most especially because this paper is not about implementing the bests of load forecasting techniques. The attacker is therefore assumed to have access to this data by

injecting false data onto the forecasted load profile. This data is thereafter simulated and the impacts analyzed. Using the load scheduling technique as proposed in [7, 8] to define the objective function, the following equation is obtained:

$$F_{j,i} = w_a * \sum A_{j,i} - w_b * \sum B_{j,i} + w_c * \sum C_{j,i} - w_d * \sum D_{j,i}.$$ (1)

Where:

A (Change on occupants) = Change in Energy ($\Delta\mathcal{E}$) * Occupancy
B (Cost) = Optimized Load (x) * Dynamic Pricing (α)
C (Discomfort) = $\Delta\mathcal{E}$/Standard deviation of Load Profiles (σ)
D (Optimization Factor) = Optimized Load (x)/Forecast Load (e)
e = Forecasted load profile.
i = Iteration number
j = hourly time interval in a day.
w = Weighting factor
x = randomly generated load profile for optimization.
$\Delta\varepsilon = e - x$

The objective function as given by Eq. 1 is optimized using GA to find the best solution (load Profile) that satisfies the conditions as defined by the input variables. By attacking the load it is expected that a direct impact will be is primarily felt on input variables A, C and D because they are the variables that have load data as a component. By modifying the affected input variable as they relate to the attack on the load, we can therefore be able to observe the impact on the forecasted scheduled load profile as the optimization process is run in order to meet the already defined objective function.

3.1 Defining Optimization Constraints

The proposed constraints for the objective function as given in Eq. 1 are defined as: energy limitation constraint and energy conservation constraint. Equation 2 is an energy limitation equation whereby both the maximum and minimum energy level of every randomly-generated load profile sample remains within the limit of the forecasted load profile.

$$e_{min} \leq x \leq e_{max} .$$ (2)

On the other hand, Eq. 3 is an energy conservation equation whereby the total energy of each randomly-generated load profiles samples is equal to the total energy consumed in any given day. This samples are taken hourly over a 24-h period in a day.

$$\sum_{j=1}^{24} x_j = \sum_{j=1}^{24} e_j$$ (3)

The results from this analysis will be compared with the normal scheduling operations which is used as the control model to show an attack-free optimization with secured data, in order to ascertain the impact of the on the household as well as the grid.

3.2   False Data Injection Attack on Load Profile

The model of cyber-attack on Load Profile as presented on this paper, is defined as an injection of false data on the original load profile data with the aim to cause the generation of random and unpredictable results thereby presenting a scheduled load which is not a true reflection of the consumer's choice as well as the market events. Let us consider an attack scenario whereby the forecasted load profile e_j, is injected with some discrete randomly generated noise η_j to create some form of distortion thereby creating a new forecast load profile as shown in Fig. 1. The new load profile q_j over a 24 h interval j, is given as:

$$q_j = e_j + \eta_j. \tag{6}$$
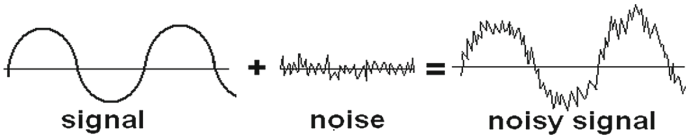
Fig. 1.   Noise injection attack on forecast load profile

The false data signal could also be a sinusoidal wave form, which may be out of phase with the original signal. The impact on load profile with a variation of false data levels is evaluated in this paper. Any amount of false data injection is possible although in this paper, it is assumed that the maximum noise that can be introduced is up to 100% of the mean load profile value. Therefore by definition, q_j is bound by a maximum allowable proportion of the forecast load profile and for only positive load profile values. This is given as:

$$e_{j_{\min}} \le q_j \le 2e_{j_{\max}}. \tag{7}$$

The load profile data was obtained from [9] and 10 iterations of increasing noise level manner from zero up till 100% of the mean load value was introduced. In order to derive the corresponding objective function, the new load profile q_j as it affects A, C and D is substituted in Eq. 1.

Therefore;

$$F_{t,i} = w_a * \sum A_{newj,i} - w_b * \sum B_{j,i} + w_c * \sum C_{newj,i} - w_d * \sum D_{newj,i}. \tag{8}$$

Where:

$$A_{newj,i} = (q_{j,i} - x) * \text{Occupancy}. \tag{9}$$

$$C_{new\,j,i} = (q_{j,i}\, x)/\sigma. \qquad (10)$$

$$D_{new\,j,i} = x\big/ q_{j,i}. \qquad (11)$$

4   Simulation and Results

Three input data which are as controlling variables are used in the optimization process and they include: the Price Profile, Occupancy and Standard Deviation of Load Profile as shown in Fig. 2. The controlled variable is given as the Original Load Profile which is attacked with false data injection, while the output is the Optimized Load Profile. The Load Profile and Price data were obtained from [9, 10] respectively.
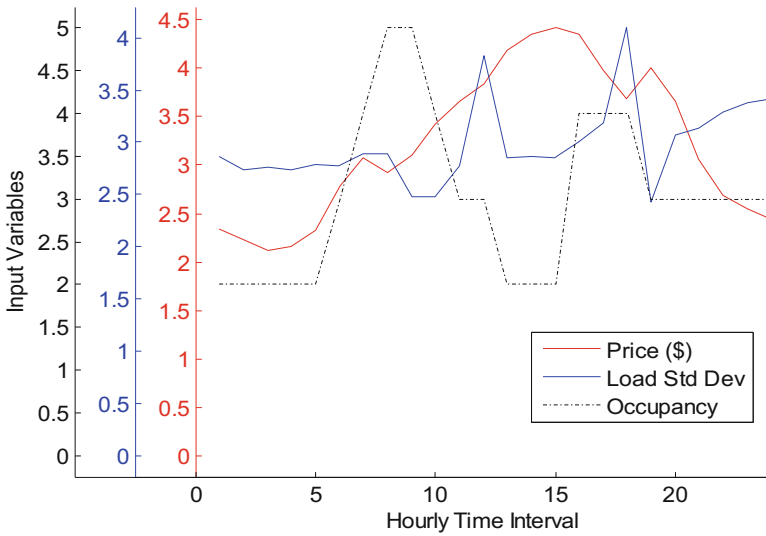
Fig. 2.   Principal input variables for load scheduling

Figures 3, 4, 5, 6 and 7 shows the Original Load Profile data and various optimized Load Profile generated using Eq. 8 as the objective function. The Original Load Profile is used as the main Load data which is subject to attack and the optimization result produced the Optimized Load Profile (With Attack) as shown in Fig. 3. The Optimized Load profile (W/out Attack) was generated using the same Original Load profile but with no false data injection. It therefore indicates the response assuming there was no attack and hence, acts as the baseline result. The Forecast Optimized Load profile is generated from Forecast Load Profile (which acts as the back-up) and is obtained from historical data of the previous 4 days of the same day of the week, as recorded within the HEMS. In this illustration, the retailer supplies the required original Load profile data, while the back-up Load Profile data was generated locally, although these positions can be interchanged as desired.
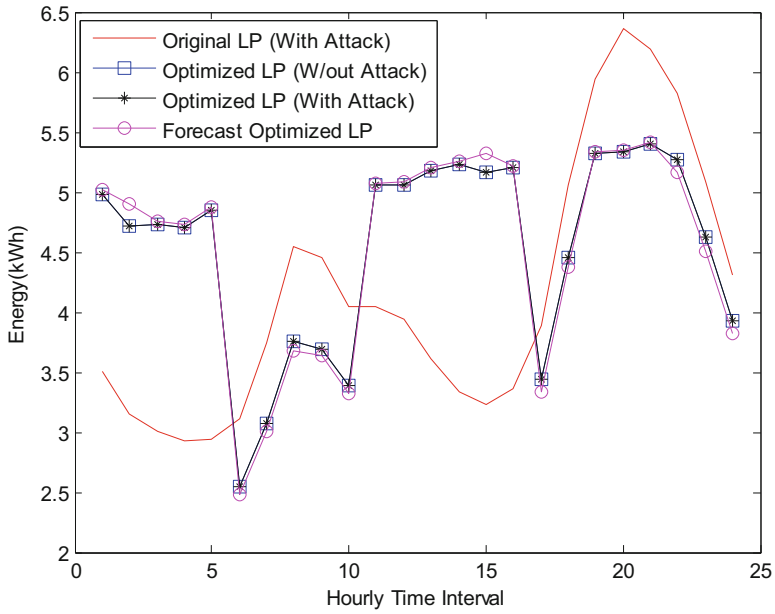
Fig. 3. Load schedule with 0% noise content
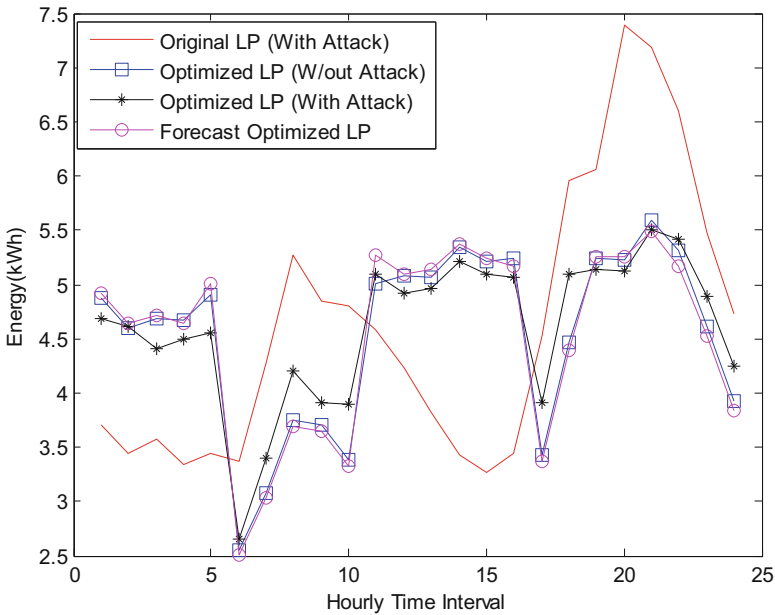
Fig. 4. Load schedule with 20% noise content
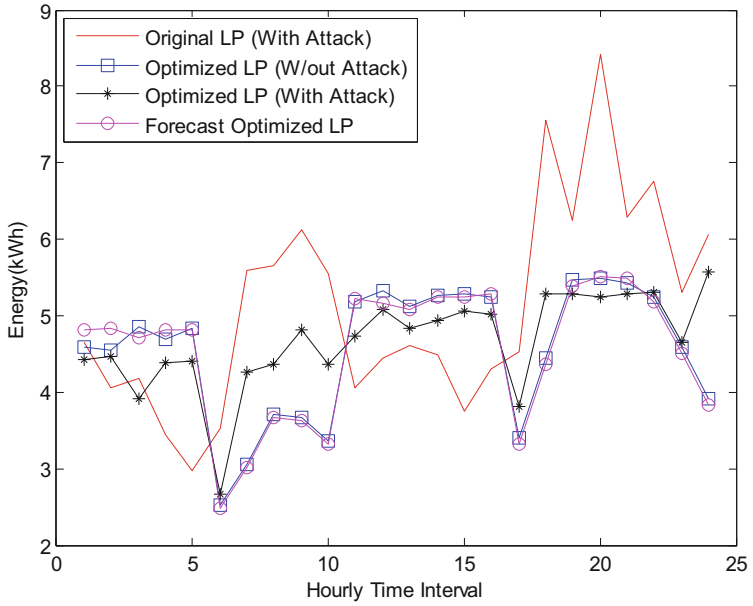
Fig. 5. Load schedule with 50% noise content
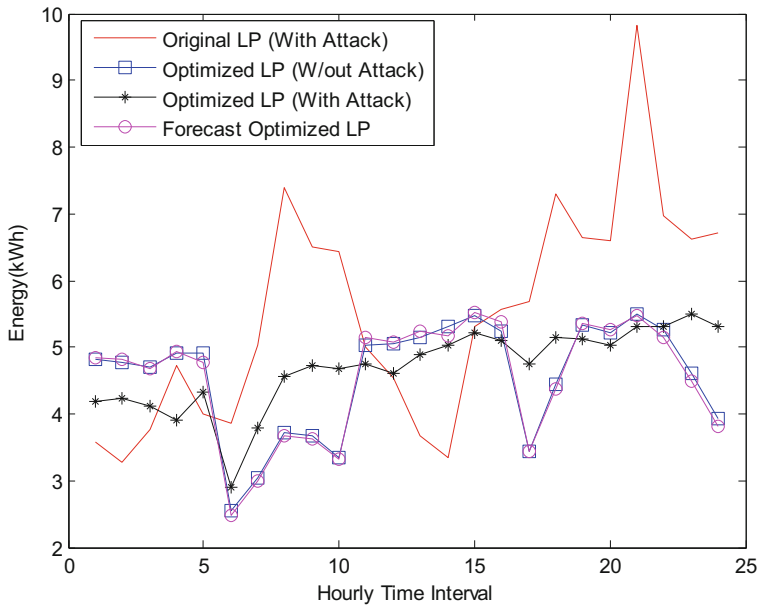
Fig. 6. Load schedule with 80% noise content
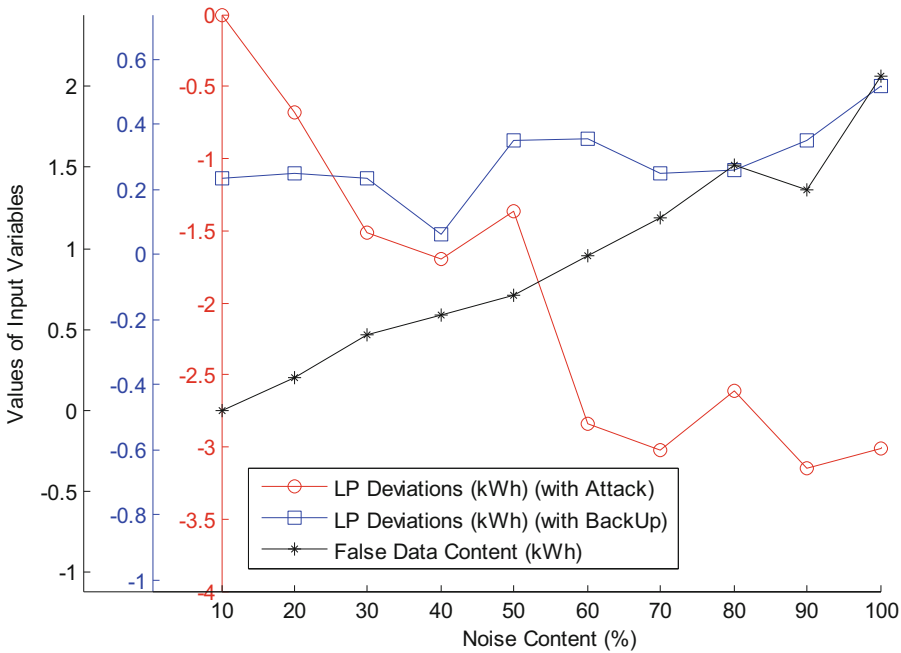
Fig. 7. Convergence of price and load profile deviations as false data injection increases

Four scenarios are shown whereby in the first one, there is no attack involved. Here, the Optimized Load Profile is exactly same for both with attack and without attack. The forecast Optimized Load Profile deviates only a bit from the two already mentioned. The second scenario as shown in Fig. 4 shows a false data injection of 20%. Here, the Optimized Load Profile (with Attack) begins to pull away from the Optimized Load Profile (without Attack) and this continues consistently at 50% as shown in Fig. 5. At 80% of false data injection as shown in Fig. 5, it can be observed that the Optimized Load Profile (with Attack), is nearly flattened out except for the spike at 6:00 h. This is a very key result as it clearly shows the effect of introducing false random data to the Original Load Profile, while the Forecast Optimized Load Profile shows that it is a reasonably good back-up to rely upon incase the HEMS detects irregular random or unexpected data within Load profile data. It is also appropriate to mention that the Optimized Load profile (without Attack) which is the baseline result remained the same with minimum and maximum values of about 2.5 kWh and 5.5 kWh respectively.

Two key observations are derivable from these results and they include:

- It is observable that as noise levels increases, the optimized load profile shrinks towards a straight line which represents the average value of the original load profile. This is an interesting result because we could see the effect of noise in diminishing the efficiency of the load scheduling process.

- The implications as deducible from the graphs is that such shrinking is capable of effecting significant load shifting whereby several loads could be turned ON when they are expected to be OFF and vice versa.

Therefore, this can be a worrisome scenario for consumers who participate in demand response programs and may not realize that their load profile has come under attack. They may conclude that participating in DR programs is highly discomforting or that perhaps, their HEMS system is dysfunctional.

Figure 7 shows the sum of deviations of Optimized Load Profile (with Attack) from the base Load Profile, which is the Optimized Load Profile (without Attack). It also shows the sum of deviations of the Forecast (Back-up) Load Profile from base Load Profile for a day, as false data content increases. It can be seen that the sum of deviations of the Back-up Load profile is fairly stable but, the Optimized Load Profile (with Attack) continues to increase significantly. This is shows the impact of such an attack on Load scheduling mechanism as well as a means of mitigating such an attack, by the provision of an effective back-up system.

5   Discussion

The key to an efficient and active DR participation is on provision of a secured network with the correct and up-to-date levels of authentication and malware security applied, in order to prevent intrusion. In a case where an attack on the load profile is successfully achieved, the response by any installed security mechanisms becomes critical. It can be observed that there is no disruption of the optimization process due to the attack which means that neither the HEMS nor the users will be able to detect any anomalies by themselves since there will always be optimized load as results. It is therefore obligatory for the designers of load optimization algorithms to include means of flagging any unexpected results and as well, include means of deriving instant solutions.

The metering system could be a reliable source to detecting anomalies within the HEMS which is in view of the availability of the historical load consumption stored in the HEMS. So if an unexpected scheduling pattern which has no resemblance and differs remarkably with the historical load profile is generated, the system could call for a reassessment and vetting of all the input data. For instance in this case as presented in this paper, having an untrusted result could require the HEMS to use the last accurate load data and apply it with the current price and occupancy data, assuming they too are not affected by the attack.

Furthermore, since the attack produced results which lead to reduction in customer satisfaction, any affected consumers will most likely be discouraged to continue with their engagements and may withdraw from active participation in demand response programs. This will therefore defeat the aim for its design but with improved security, a long term benefit and advancement of the grid can be assured.

6   Conclusion

Cyber-attacks on HEMs are a real possibility and care should be taken towards ensuring the protection of the infrastructure that constitutes the network. In this paper, there has

been a presentation and simulation of a cyber-attack on the HEMS which was modelled as false data injection onto the load profile data. Results obtained showed that such an attack diminishes the optimization mechanism as well as the system performance by forcing the load profile to flatten out. Having such a flattened load demand throughout the day may seem to be the most optimal energy supply for a community from the grid perspective in terms of ensuring a supply of constant energy capabilities thereby eliminating peak load demand. But in practice, a flattened load profile is neither realistic nor comfortable for the users.

Finally, using previously known accurate data can help minimize the impact of such attacks. This means that the HEMS should keep a record of recent data and as well, perform some forecasting mechanisms on all data available. Nevertheless, preventing unauthorized access remains the best possible solution. Access authentication is naturally a key part for any proposed solution as it is important to ensure consumer confidence otherwise the dream of having a robust smart home which consumers will appreciate, will become unachievable.

References

1. Lee, R.M., Assante, M.J., Conway, T.: Analysis of the cyber attack on the Ukranian power grid. E-ISAC (2016). https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf. Assessed 31 Dec 2016
2. Kumar, M.: DDoS Attack takes down central heating system amidst winter in Finland. http://thehackernews.com/2016/11/heating-system-hacked.html. Accessed 21 June 2017
3. Amini, S., Pasqualetti, F., Mohsenian-Rad, H.: Detecting Dynamic Load Altering Attacks: A data-driven time-frequency analysis (2016)
4. Amini, S., Pasqualetti, F., Mohsenian-Rad, H.: Dynamic load altering attacks against power system stability: attack models and protection designs. IEEE Trans. Smart Grid (Submitted) (2016)
5. Duhamel, P., Vetterli, M.: Fast fourier transforms: a tutorial review and a state of the art. Signal Process. **19**, 259–299 (1990)
6. Zhang, F., Geng, Z., Yuan, W.: The algorithm of interpolating windowed FFT for harmonic analysis of electric power system. IEEE Trans. Power Deliv. **16**(2), 160–164 (2001)
7. Anuebunwa, U., Rajamani, H.S., Pillai, P., Okpako, O.: Novel genetic algorithm for scheduling of appliances. In: Proceedings of the 2016 IEEE PES Power Africa Conference, Livingstone, Zambia, pp. 57–61 (2016)
8. Anuebunwa, U., Rajamani, H.S., Pillai, P., Okpako, O.: Investigating the impact of discomfort in load scheduling using genetic algorithm. In: IEEE International Conference on Power Systems Technology (IEEE POWERCON2016), Australia (2016)
9. U.S Department of Energy.: Commercial and residential hourly load profiles for all TMY3 Locations in the United States, OpenEI, 2nd July 2013
10. Illinois, A.: Day Ahead Pricing used for billing RTP and HSS service, Ameren, 1st May 2015

SATPROP Workshop Session

Propagation Elements for the Link Budget
of Broadband Satellite Systems in Ka and Q/V Band

Spiros Ventouras[1(✉)], Paul S. Crawford[2], and Charilaos I. Kourogiorgas[1]

[1] RAL Space, Harwell, Oxfordshire, UK
{spiros.ventouras,charilaos.kourogiorgas}@stfc.ac.uk
[2] University of Dundee, Dundee, UK
psc@sat.dundee.ac.uk

Abstract. In this paper we discuss two propagation elements which are crucial in the design - from both economical and technological perspective - of HTS systems utilizing Ka-, Q/V- or even W- frequency bands. These are the statistical combination of rain and cloud attenuation statistics and the inter-annual variability of the attenuation statistics. Also the recently proposed model for the prediction of annual rain statistics (ITU-R Rec. 837-7) is compared with rain data of 13 consecutive years at Chilbolton, UK.

Keywords: HTS systems · Total attenuation · Ka band · Q band
Inter-annual variability · System risk assessment · Rec 837

1 Introduction

The choice of operating frequency for a satellite radio link is constrained by a number of factors. Some are fundamental physical properties (e.g. bandwidth being limited to a moderate fraction of the chosen centre frequency, tropospheric impairments), while others are constrained by current technology (e.g. cost of ground based hardware, available space-based transmitter power, etc.), and of course the regulatory frameworks in place to manage the use of the radio spectrum. As a general rule, moving to higher frequencies provides greater bandwidth and thus potential data rates but is more demanding in terms of technology (and thus hardware cost), and ultimately more prone to the effects of atmospheric gases, clouds, rain and tropospheric turbulence.

The majority of the atmospheric phenomena exhibits a stochastic behaviour both in time and space and therefore differ from all other deterministic factors (e.g., free space loss) that affect the satellite link under clear sky link conditions. Furthermore, since the power resources are limited and bearing in mind (a) the levels of atmospheric attenuation which have to be coped with and (b) the higher performance requirements of these higher frequency systems, more accurate link budget calculations are required.

This means that the possible fluctuations of the various contributions to the link budget are to be considered to comply with the requirements in all conditions. One cause of these fluctuations is the prediction error of the propagation elements of the link budget (e.g. fade margin at a specific probability level). Therefore this propagation prediction

error has to be modelled in order to provide a system risk assessment (i.e. risk of not reaching the availability target) and an indication of the system reliability.

Section 2 of this paper provides the background regarding the concept of the Link Margin, whereas in Sect. 3 the prediction of total attenuation statistics that required for the Link Budget calculations is discussed; a new prediction method is explained. The uncertainty in the system design because of the variability of the annual attenuation statistics is discussed in Sect. 3. Also, in this section, the recently proposed model for the prediction of annual rain statistics (ITU-R Rec. 837-7), required for the prediction of attenuation statistics, is compared with rain data of 13 consecutive years from Chilbolton, UK.

2   Link Budget

The link performance is evaluated by the parameter Carrier-to-Noise Density ratio $\left(\dfrac{C}{N_O}\right)$ given by the Eq. (1):

$$\left(\frac{C}{N_O}\right) = EIRP + \left(\frac{G}{T}\right) - \left(L_{FS} + \Sigma OtherLoss\right) + 228.6 \,(\text{dBHz}) \tag{1}$$

Where C is the average RF carrier power, N_O the noise power density, EIRP the effective isotropic radiated power (dBW), $\left(\dfrac{G}{T}\right)$ the figure of merit of the receiving terminal (dBK^{-1}), L_{FS} the free space loss (dB), $\Sigma OtherLoss$ are the additional losses e.g. due to troposphere and 228.6 is the dB value of the inverse of Boltzmann's constant.

2.1   Link Margins

For a link, with $\left(\dfrac{C}{N_O}\right)_{Req}$ as the minimum value of the carrier-to-noise density ratio the available margin M is given by the formula:

$$M = \left(\frac{C}{N_O}\right) - \left(\frac{C}{N_O}\right)_{Req} \tag{2}$$

where $M \geq 0 \, dB$ for the link to be capable of operation.

Depending on what losses are included in the link budget calculation we have different margins. The simplest case is where we only have the free space loss, leading to the in-vacuum margin:

$$M_V = \left(\frac{C}{N_O}\right) - \left(\frac{C}{N_O}\right)_{Req} = EIPR + \left(\frac{G}{T}\right) - L_{FS} + 228.6 - \left(\frac{C}{N_O}\right)_{Req} \tag{3}$$

However, the in-vacuum margin is of little use for a ground station as there is the always present attenuation due to the atmospheric gases, A_G. Measurements have shown [2] than A_G experiences a small variability around its mean value and can be consider fixed either equal to its mean value or for a more conservative approach equal to the exceeded level for the annual percentage 99%. Both these values can be predicted by the ITU-R Rec. 676. So, the simplest useful margin is the "clear sky" margin:

$$M_{CS} = M_V - A_G = EIRP + \left(\frac{G}{T}\right) - \left(L_{FS} + A_G\right) + 228.6 - \left(\frac{C}{N_O}\right)_{Req} \tag{4}$$

Again we require a positive margin for the link to be capable of any operation in the real world. However, we have other sources of attenuation that are variable and models (e.g. ITU-R) the loss for a given link availability p (in %) based on:

$$M_T = M_{CS} - A_{VT}(100 - p) \tag{5}$$

Where M_T is the total link margin (taking in to account all losses) and $A_{VT}(100 - p)$ is the excess tropospheric attenuation threshold exceeded for the $(100 - p)\%$ of the average year (i.e., the variable part: rain, cloud, scintillation).

The traditional approach for computing a link budget then is to evaluate the engineering and site-related parameters (EIRP, G/T, L_{FS} and required C/N_0), and then to compute the gaseous attenuation A_G to yield the clear-sky margin M_{CS} from Eq. (3). Then for a given availability specification p_0, we can compute the corresponding excess attenuation A_{VT} and from that we have the total margin M_T for the link using Eq. (4). If we find that $M_T \geq 0\ dB$ then we are satisfied that the link will meet or exceed the requirements.

If we find that $M_T < 0$ dB, or M_T is significantly above 0 dB, then normally we will change one or more of the engineering parameters to either enable reliable operation (e.g. increasing EIRP if the margin is below 0 dB) or to reduce the overall system cost (e.g. reducing EIRP if a substantial positive margin is found).

Therefore, it is clear that the accurate prediction of Total attenuation statistics (i.e., Gaseous and Excess attenuation) is crucial for the optimum design of the satellite system.

3   Total Attenuation

The current ITU-R method for the prediction of total attenuation statistics - Rec. P. 618-12, Annex 1 Sect. 2.5 - combines the statistics of the individual tropospheric effects on an equiprobable basis. This combination is very convenient from an engineering point of view, however, it is not correct and can lead to the system overdesign [2, 3]. In particular, it has been proved to be inappropriate for higher frequency or low elevation angle systems where the effects from gases, clouds, light rain and scintillation can be significant.

Total attenuation can be considered as the sum of two components: the gaseous attenuation due to oxygen and water vapour, which is always present, and the excess attenuation due to clouds, rain and scintillation. In beacon attenuation measurements it

is difficult to distinguish rain attenuation from its associated raining-cloud attenuation. As a consequence rain attenuation prediction models or experimental rain attenuation data include also raining clouds attenuation. Therefore, the combined attenuation due to rain and clouds can be written as the sum of two components: the rain and raining clouds attenuation component, and the non-raining clouds attenuation component.

In contrast to cloud and rain attenuation, gaseous attenuation is always present. Measurements and studies have shown that the spread of gaseous attenuation around its mean value is less than ~1 dB across Europe for the frequency range of interest [2]. Therefore, this method estimates the total attenuation level exceeded for a given percentage from: the gaseous attenuation $m_{gaseous}$ (mean or the exceeded level for the annual percentage 99%), the rain attenuation statistics (which includes the precipitating cloud attenuation), the cloud attenuation statistics (which refers to non-precipitating clouds) and the scintillation statistics.

Step 1: Rain and cloud attenuation statistics: Obtain the rain and cloud attenuation statistics either from experimental data or prediction procedures and construct the pairs:

$\{P_i(A_R \geq a_i), a_i\}$ where $P_i(A_R \geq a_i)$ is the probability the rain attenuation (including the precipitating cloud attenuation) a_i is exceeded.

$\{P_i(A_C \geq a_i), a_i\}$ where $P_i(A_C \geq a_i)$ is the probability the cloud attenuation (only of non-precipitating clouds) a_i is exceeded.

The values of the attenuation thresholds a_i can be determined in steps of 1 dB (or 0.5 dB for a better accuracy), {0,1, 2, 3,…, a_max}. The value a_max is determined by the maximum measurable attenuation in case of experimental data or the attenuation area of interest in case of predictions.

NOTE 1. If the rain or cloud attenuation statistics are given in terms of percentages (i.e., exceedance probability) they can be derived in terms of attenuation thresholds by interpolation (ITU-R Rec. P. 1144).

NOTE 2. The rain statistics can be obtained from the ITU-R Rec. 618.

NOTE 3. Due to the lack of a model for the prediction of non-precipitating cloud attenuation statistics (or more accurately due to the lack of predictions of total columnar content of reduced cloud liquid water only for non-precipitating clouds) the required non-precipitating cloud statistics can be approximated by the ITU-R Rec. 840). This might be established after the validation of the method with experimental data.

Step 2: Gaseous attenuation: Obtain the mean gaseous attenuation $m_{gaseous}$ from experimental data or prediction procedures (e.g., ITU-R P. Rec. 676).

Step 3: Contribution from rain and cloud attenuation: The probability, $P_i(A_{R+C} \geq a_i)$, that the attenuation due to rain and clouds, A_{R+C}, exceeds the attenuation threshold of a_i dB, is given by the sum of the exceedance probabilities of rain and cloud attenuation for the attenuation threshold of a_i dB.

$$P_i(A_{R+C} \geq a_i) = P_i(A_R \geq a_i) + P_i(A_C \geq a_i) \tag{6}$$

Step 3 provides the pairs $\{P_i(A_{R+C} \geq a_i), a_i\}$ where a_i as defined in step 1 (e.g., {0, 1, 2, 3,…, a_max}).

Step 4: The required attenuation exceeded level a_{R+C} due to cloud and rain for a specific percentage of interest p of an average year can be obtained as in step 1 by

performing interpolation on the estimated values of probability $P_i\left(A_{R+C} \geq a_i\right)$ versus a_i (p the abscissa value for a_i).

Step 5: Total attenuation ($A_T(p)$) exceeded for the percentage of interest p:

$$A_T(p) = m_{gaseous+} \sqrt{a_{R+C}(p)^2 + A_S(p)^2} \tag{7}$$

where $m_{gaseous}$ is the gaseous attenuation, mean value or the exceeded level for the annual percentage 99% as estimated in step 2, $a_{R+C}(p)$ the attenuation level exceeded for the percentage p due to rain and cloud as estimated in step 4 and $A_S(p)$ the attenuation level exceeded for the percentage p due to scintillation either from experimental data or prediction procedures (e.g. ITU-R Rec.618).

The method has been validated against data collected in the Southern of England during the ITALSAT propagation campaign [2] (see Fig. 1). In addition the method presented in the document CP(15)16 3J/130, 3M/227: Proposed Modification to ITU-R Rec. p. 618, Propagation data and prediction methods for the design of Earth-Space telecommunication systems, allocated to SWG 3M-2 in the ITU-R meeting in April 2015. The outcome was: 3 M/TEMP/85 Working document towards revision of Rec. ITU-R P.618-11 to be attached to the WP 3 M CR.
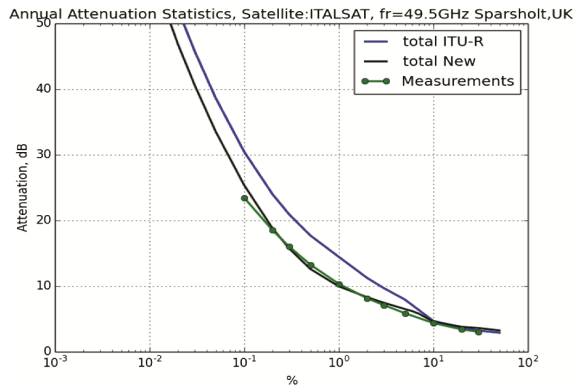
Fig. 1. Annual measured and predicted total attenuation statistics at 49.5 GHz.

4   Risk Associated with Propagation Margin

Measurements have shown that a propagation parameter, for example atmospheric attenuation at a given probability level, shows a variability which increases with the decreasing probability. Similar behaviour has been also observed for the rain statistics. As stated in [4] this variability is not to related to climatic variability [5, 6] or the up-time of an experiment [7]. In fact it can be demonstrated that the statistical estimator used to compute the empirical complementary cumulative distribution function has a significant variance due to the finite duration of the measurements [4].

Figure 2 shows the annual rain rate variability at Chilbolton for consecutive 13 years of measurements in comparison with the ITU-R predictions (Rec. 837), the current (836-6) and the recently proposed modification (837-7).
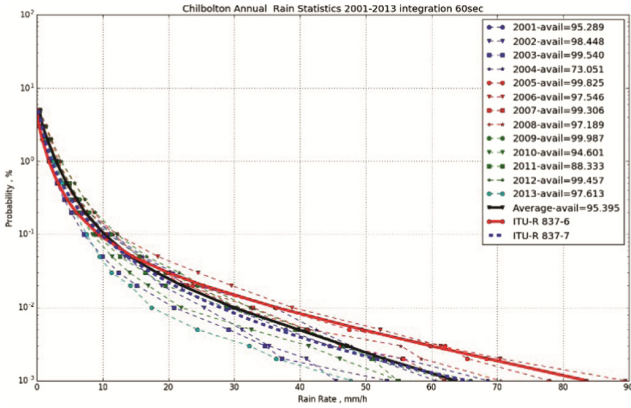
Fig. 2. Annual measured and predicted Rain rate statistics at Chilbolton, UK.

The recent ITU-R Rec. P.678-3 (Annex 2) provides a methodology based on study [4] for estimating the inter-annual variability of rainfall rate and rain attenuation statistics around the long-term Complementary Cumulative Distribution Function (CCDF). In particular, for a given location, the inter-annual fluctuations of rainfall rate and rain attenuation statistics around the long-term Complementary Cumulative Distribution Function (CCDF) p are normally distributed with mean p and yearly variance σ^2 so that:

$$\sigma^2(p) = \sigma_C^2(p) + \sigma_E^2(p) \tag{8}$$

where σ_E^2 is the variance of estimation and $\sigma_C^2(p)$ is the inter-annual climatic variance.

There is an uncertainty in the system design because of the variability of the annual attenuation statistics. In particular, the total attenuation margin M_T and consequently the achieved availability Pa depend on the Variable Loss, A_{VT} (see Eq. 4), which undergoes an annual variability for a given target availability P_T. The system does not meet its specifications when the Variable Loss for a specific year exceeds the estimated value.

From an engineering point of view the following approach is proposed to consider the variability of the annual statistics.

- Estimation of the clear sky Margin M_{CS} (Eq. 4).
- According to ITU-R Rec. 678 the exceedance probability for the attenuation threshold of M_{CS} experiences an inter-annual variability with mean value $p_{cs,m}$ and standard deviation σ_{CS} (see Eq. 8). Estimation of the exceedance probability, $p_{cs,m}$ from the long-term total attenuation statistics (minus the fixed gaseous attenuation) and the standard deviation σ_{CS} (ITU-R Rec. 678).

- For a given target availability P_t (i.e. exceedance probability $1 - P_t$) the system with margin M_{CS} meets its specifications if the relevant exceedance probability p_{cs} is less than the target exceedance probability $1 - P_t$. Therefore the probability $P[p_{cs} \leq 1 - P_t]$ gives a measure that the system meets its specification whereas the probability $P[p_{cs} \geq 1 - P\_t]$ gives a measure of the risk, R, that the system will not meet its specification. Following that the variability of annual attenuation statistics follows the normal distribution (ITU-R Rec. P.678), the risk R is estimated from:

$$R = P(p_{CS} \geq 1 - P_t) = \frac{1}{2} erfc(\frac{1 - P_t - p_{CS.m}}{\sqrt{2}\sigma_{CS}}) \tag{9}$$

Figure 3 illustrates the above procedure.
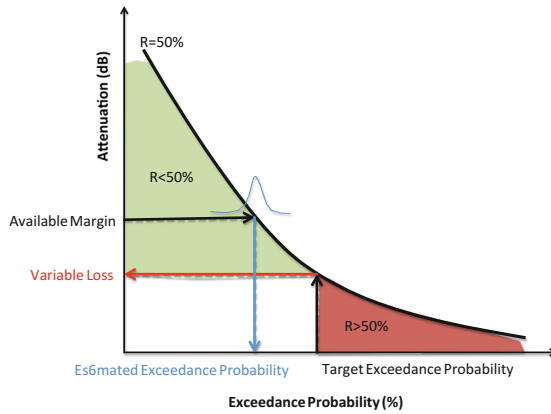
Fig. 3. Estimation of the risk associated with propagation margin

5 Conclusion

In this paper the prediction of total attenuation statistics that required for the Link Budget calculations was discussed. The current ITU-R prediction it seems that overestimates the total attenuation statistics and can lead to the system overdesign, in particular for Ka and Q/V band systems. A new prediction method based on rigorous mathematical foundations and experimental observations was explained and was in good agreement with the limited, however, experimental data at higher frequencies.

There is an uncertainty in the system design because of the variability of the annual attenuation statistics. A procedure, based on the ITU-R Rec. 678, was presented in order the variability of the annual statistics to be considered and the risk associated with the system performance to be evaluated.

Also the recently proposed model for the prediction of annual rain statistics (ITU-R Rec. 837-7) was compared with rain data of 13 consecutive years at Chilbolton, UK.

The comparison showed a great agreement with the data and a remarkable improvement compared with the Rec. 837-6.

References

1. Ippolito, L.J.: Satellite Communications Systems Engineering: Atmospheric Effects, Satellite Link Design and System Performance. Wiley, New York (2008)
2. Ventouras, S., Callaghan, S.A., Wrench, C.L.: Long-term statistics of tropospheric attenuation from the Ka/U band ITALSAT satellite experiment in the United Kingdom. Radio Sci. **41**, RS2007 (2006)
3. Ventouras, S., Wrench, C.L.: ITU-R total attenuation predictions in comparison with slant path measurements in Southern England. Electron. Lett. **38**(18), 1058 (2002). https://doi.org/10.1049/el:20020696
4. Jeannin, N., Boulanger, X., Féral, L., Castanet, L., Lacoste, F.: Inter-annual variability, risk and confidence intervals associated with propagation statistics. Part I: theory of estimation. Int. J. Satell. Commun. Netw. **32**(5), 407–421 (2013). https://doi.org/10.1002/sat.1060
5. Poiares Baptista, J.P.V., Zhang, Z.W., McEwan, N.J.: Stability of rain-rate cumulative distributions. Electron. Lett. **22**(7), 350–352 (1986)
6. Paulson, K.S.: Trends in the incidence of rain rates associated with outages on fixed links operating above 10 GHz in the Southern United Kingdom. Radio Sci. **45**, RS1011 (2010)
7. Matricciani, E., Riva, C.: The search for the most reliable long-term rain attenuation CDF of a slant path and the impact on prediction models. IEEE Trans. Antennas Propag. **53**, 3075–3079 (2005)

Monthly and Seasonal CFLOS Statistics for Optical GEO Feeder Links Design

Nikolaos K. Lyras[1], Charilaos I. Kourogiorgas[1], Athanasios D. Panagopoulos[1(✉)] [iD], and Konstantinos P. Liolis[2]

[1] School of Electrical and Computer Engineering,
National Technical University of Athens, Athens, Greece
lyrasnikos@central.ntua.gr, harkour@mail.ntua.gr,
thpanag@ece.ntua.gr
[2] SES S.A., Betzdorf, Luxembourg
Konstantinos.Liolis@ses.com

Abstract. A methodology for the generation of seasonal and monthly cloud-free-line of sight (CFLOS) statistics for single optical satellite slant paths and joint CFLOS statistics for multiple spatially separated optical satellite links is proposed and evaluated with numerical results. It is assumed that the blockage of the satellite link is considered with the occurrence of clouds. The reported methodology is based on the stochastic dynamic modeling of Integrated Liquid Water Content (ILWC) and uses as inputs monthly statistical parameters of the logarithm of (ILWC) to produce CFLOS statistics. This methodology is an extension of an existing model that captures very well both the seasonal and monthly variability of clouds. The proposed model takes into account the temporal and spatial correlation of clouds, the elevation angle and the altitude of the ground station among others. The numerical results are concentrated on for hypothetical satellite links with optical gateways located in North and South hemisphere respectively are presented. The numerical results are compared with the corresponding annual results.

Keywords: Optical satellite communications
Cloud-Free-Line of Sight probability · CFLOS · Cloud coverage
Seasonal and monthly variability

1 Introduction

Nowadays, Earth-space optical communication links are gathering more and more of the attention of the communication society. The use of optical communications has additional advantages with respect to the radio frequencies (RF) communications such as: (a) absence of frequency regulation constraints due to the highly directive antennas, (b) smaller systems with potentially lower power consumption, (c) enhanced security due to directivity of the beams and (d) ability to completely reallocate RF frequency bands to the user link [1]. These advantages have motivated the investigations of optical satellite communication systems. Already, a bunch of studies have been conducted for

the migration of feeder links from RF [2] to the optical band in the context of satellite communication systems [3–5].

Although optical links have a great variety of advantages, they are severely affected by several atmospheric phenomena like precipitation, atmospheric gases, liquid water particles and atmospheric turbulence [1]. Among these impairments clouds contain the dominant fading mechanism. Clouds induce tens of dBs attenuation in optical signal. Therefore, the blockage of the link can be considered with the occurrence of clouds [6]. For the mitigation of clouds site diversity technique is employed. Multiple ground stations shaping a ground station network are placed in large distances to cope with cloud coverage and to guarantee the desired availability of the system. Methodologies for the estimation of Cloud Free Line of Sight (CFLOS) probability for both single links and for a ground station network have been developed [6]. These methodologies have to take into account the temporal and spatial correlation of clouds, the elevation angle and the altitude of the stations among others. In addition, for the dimensioning of an optical ground station network, it is important to examine the monthly and the seasonal variability of the clouds occurrence. As it is expected the clouds are highly seasonal and monthly dependent. Even for regions which are favorable for optical communications (low cloud coverage among others) the percentage of cloud coverage greatly depending on the month and season. Moreover, significant differences can be identified between stations that are located in different hemispheres. For example, in Fig. 1 the cloud coverage per month (monthly variability) for year 2009 for one station in north hemisphere, Madrid Spain, and another in south hemisphere Santiago, Chile is presented.
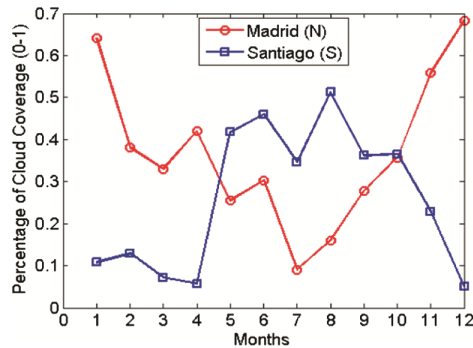
Fig. 1. Percentage of cloud coverage (0–1) for year 2009 for Madrid (North Hemisphere) and Santiago (South Hemisphere)

For the reliable and the accurate design of Cloud-Free Line-of-Site probability for the optical communication network, monthly and seasonal variability of clouds must be taken into account. The main objective of this contribution is to exhibit monthly and seasonal CFLOS statistics for single optical links and for optical satellite links separated in spatial domain. Moreover, the necessity of using such statistics for the dimensioning of an optical satellite communication system is revealed while, the seasonality differences and the benefits of using stations in different hemisphere are investigated.

The remainder of the paper is structured as follows: In Sect. 2 the methodology for the estimation of single and joint CFLOS statistics is presented. In Sect. 3 the exhibited methodology is evaluated for various single and spatial diversity scenarios. In addition, single and joint CFLOS statistics using both monthly and annual statistical parameters are synthesized and compared. Finally, in Sect. 4 the conclusions of this contribution are presented.

2 Methodology

In this Section, the methodology for the generation of seasonal and monthly CFLOS statistics will be exhibited. This methodology is based on the stochastic dynamic modeling of ILWC fields which is presented in [7]. The proposed methodology except from the spatial and temporal variability which is also included in [7] could be able to capture the seasonal and monthly variability of both single and joint CFLOS statistics. Moreover, this methodology takes into account the elevation angle and the altitude of the ground station for the evaluation of CFLOS statistics. To begin with, the main assumptions of the proposed methodology will be reported. Firstly ILWC can be sufficiently described by lognormal distribution [8] while Liquid Water Content (LWC) can be modeled taking into account the height and ILWC using the expression proposed in [9]. Moreover the statistical parameters of the logarithm of ILWC (mean value, standard deviation and probability of cloud coverage) can be derived from Numerical Weather Prediction databases on yearly, monthly, daily or hourly base and will be considered constant for the whole slant path.

For the generation of ILWC fields correlated both on temporal and spatial domain for either single or spatial diverse links, the methodology exhibited in [7] using stochastic differential equations is applied. However, for this contribution, as statistical parameters of the logarithm of ILWC, monthly parameters are used. For the calculation of these parameters the methodology presented in SMOC [9] will be followed. Firstly, from Numerical Weather Prediction databases, in our case ERA Interim database the monthly values of total cloud cover (probability of cloud coverage) and total column cloud liquid water for the places of interest are extracted. Then, according to SMOC the mean value and standard deviation of the log normal distribution of ILWC are derived. Thus, the monthly statistical parameters of logarithm of ILWC have been computed. It is assumed that these parameters are constant for the whole slant path. Now the main steps for the generation of CFLOS time series will be exhibited. The following steps are followed for each month:

1. ILWC time series are computed [7] using the monthly statistical parameters. More particularly, SDEs are employed and ILWC grids 1 km × 1 km correlated on temporal and on spatial domain as it is demonstrated in Fig. 1 of [7] are synthesized. It is assumed that ILWC is constant for each grid [6, 7, 9].
2. ILWC time series are converting into cloud vertical extent time series through the expression presented in [9] and also used in [7].
3. Time series of cloud coverage are estimated taking into account the elevation angle and the altitude of the ground station. Clouds block the link if there is a cloud

formation with such a vertical extent so as the link is impaired. The link is not impaired either when the top level of the cloud is under the slant path or the bottom of the cloud is above the slant path. The configuration of slant path taking into account the vertical extent of clouds is demonstrated in [6].

4. Time series of cloud coverage are converting into time series of CFLOS *CFLOS = 1− Cloud_Coverage.*

In Fig. 2 a 2-dimensional field of ILWC for a specific snapshot (step 1) is exhibited.
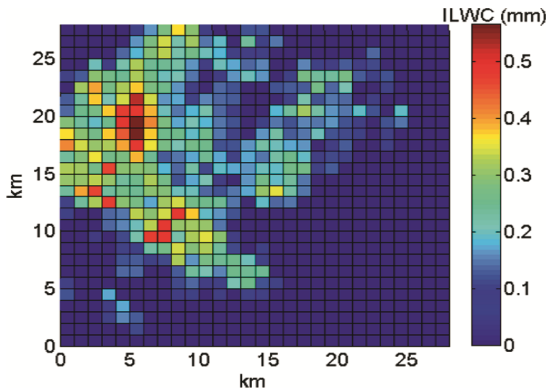
Fig. 2. ILWC 2-dimensional field for a specific time instance

3   Numerical Results and Discussion

In this Section, the methodology reported before will be evaluated for the generation of single and joint CFLOS statistics, in North and South America. The statistical parameters have been derived from the monthly database of Era-Interim for the period from 1/1/2009 to 31/12/2015 while the numerical calculation software has been developed in MATLAB. The hypothetical optical links which that are used for the numerical results are shown in Table 1. As space segment, the QuetzSat-1 GEO satellite at 77degW owned and operated by SES is considered along with a hypothetical optical payload on-board the satellite.

To begin with, following the reported methodology, the mean monthly CFLOS probability from 2009 to 2015 for the hypothetical stations at Santiago, Malargüe Steele Valley and Vernon are presented in Fig. 3. For the x label of Fig. 3, 1 means January, 12 December etc. From this Figure it can be easily observed both the monthly variability of each station and the variability between the two different hemispheres. For example, for the stations located in North the best months are from May to October while the opposite occurs for the ones located in South.

Table 1. Hypothetical optical satellite links

Station	Latitude (deg)	Longitude (deg)	Elevation Angle (deg)	Altitude (m)	Hemisphere
Santiago, Chile	−33.44	−76.68	50.51	600	South
Malargüe, Argentina	−35.483	−69.58	48.04	1400	South
Calama, Chile	−22.508	−68.95	62.1	2200	South
Steele Valley, California, USA	33.76	−117.32	32.0	612	North
Vernon, Texas, USA	34.218	−99.40	43.57	400	North
Kitt Peak, USA	31.96	−11.60	37.39	2000	North
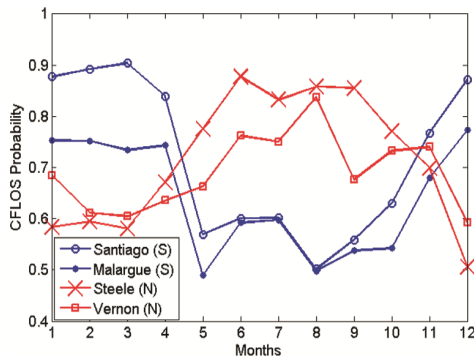
Fig. 3. Monthly CFLOS Probability

Now the same results are presented for the high altitude stations of Kit Peak and Calama. The reported methodology can capture the effect of the Altitude of the ground stations in the CFLOS statistics, among others (Fig. 4).

To continue, in Fig. 5 the mean Seasonal CFLOS statistics are reported. Again the seasonal variability in each station and between the stations of different hemisphere is made explicit.
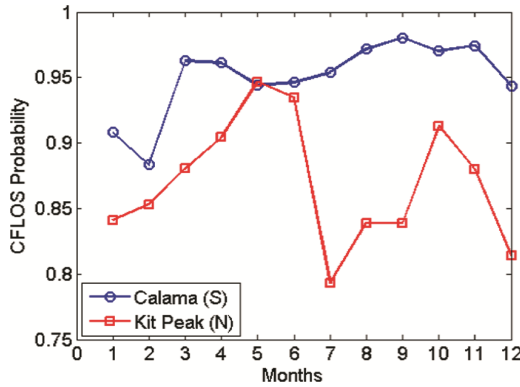
Fig. 4. Monthly CFLOS Probability, high altitude stations
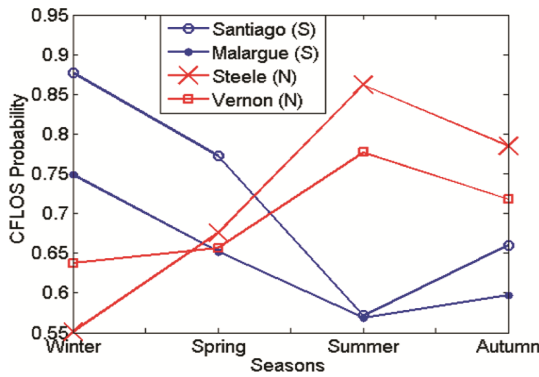
Fig. 5. Seasonal CFLOS Probability

At this point, with the following Figures the differences between using monthly and annual statistical parameters are demonstrated. Finally the following results are computed using the monthly statistics, and the annual statistics. The stations are the same as before and the period is 6 years (from 2009 to 2015).

In Figs. 6 and 7 the monthly and seasonal CFLOS statistics are depicted. For both figures the statistics are computed for two stations one in the north and another in the south hemisphere using monthly and annual statistical parameters. The results reveal that using annual statistical parameters, the monthly and seasonal variability of CFLOS probability and the variability between the two hemispheres cannot be captured. Thus, the use of annual statistical parameters cannot guarantee a robust optical ground station dimensioning.
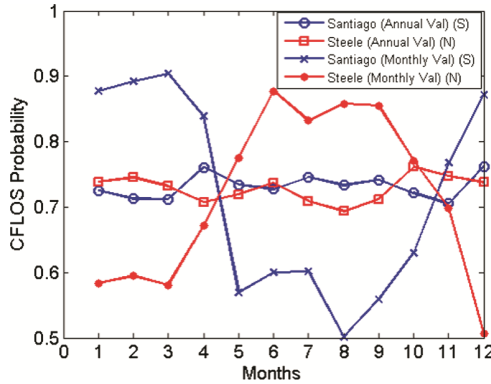
Fig. 6. Monthly CFLOS Probability using annual and monthly statistical parameters
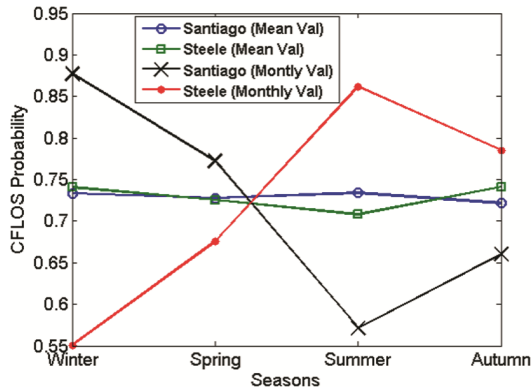
Fig. 7. Seasonal CFLOS Probability using annual and monthly statistical parameters

Except from the statistics for the single links, joint CFLOS statistics of a spatial diversity scenario are also presented. A double triple etc. diversity scenario means that at least one of the two three etc. stations respectively is not blocked. For the double site diversity scenario the stations of Santiago and Steele Valley are used, for the triple the station of Malargüe is added to the previous ones and finally for the quadruple scenario the station at Vernon is added. You can observe that for double scenario there is one station from the south hemisphere and another from the north etc.

In Figs. 8 and 9 the monthly joint CFLOS probability for the double, triple and quadruple spatial diversity scheme, is computed using annual and monthly statistical parameter.
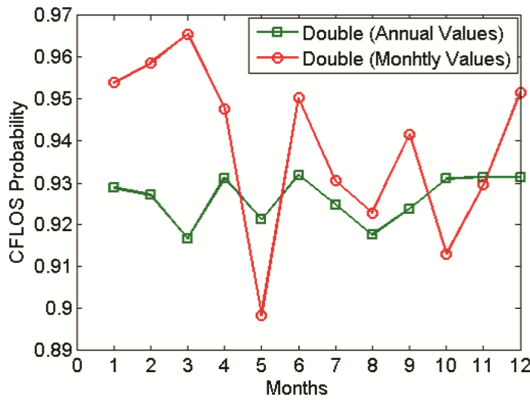
Fig. 8. Joint CFLOS Probability for double diversity scenario using annual and monthly statistics
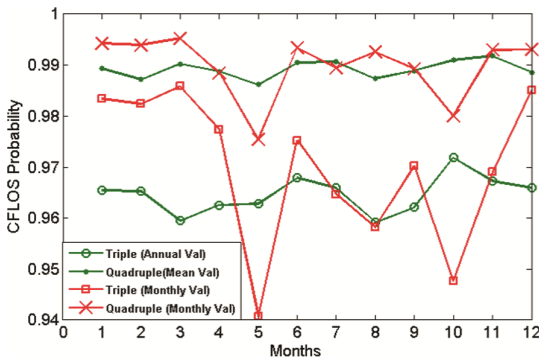
Fig. 9. Monthly joint CFLOS Probability for triple and quadruple spatial diversity scenario using annual and monthly statistics

In Fig. 10 the seasonal joint CFLOS probability for the double, triple and quadruple spatial diversity scheme, is computed using annual and monthly statistical parameter.

Finally in Figs. 11 and 12 the monthly statistics for the quintuple and sextuple spatial diversity scheme are exhibited. In Fig. 11 we have pinpointed the 99.9% CFLOS probability as reference. In addition, it can be easily observed that in case of using annual statistical parameters, the monthly CFLOS is always higher than the 99.9% reference level but in case of using monthly statistical parameters it is not true. Thus, depending on the system requirements, monthly CFLOS statistical parameters result on more reliable CFLOS results. For the quintuple scheme the station of Calama has been added to the before mentioned station while for the sextuple except from Calama the station in Kit Peak is also added.
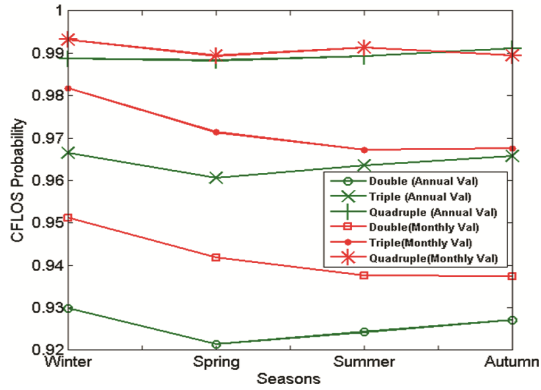
Fig. 10. Seasonal joint CFLOS Probability for triple and quadruple spatial diversity scenario using annual and monthly statistics
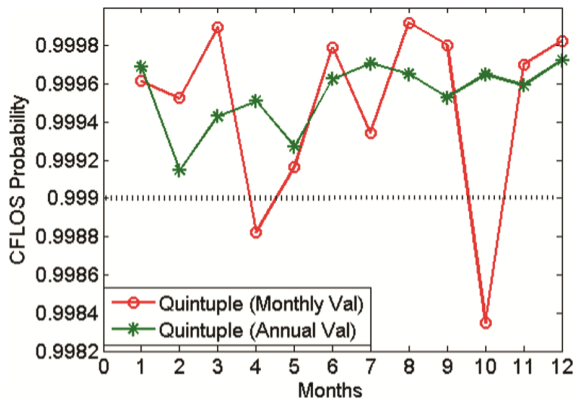
Fig. 11. Monthly joint CFLOS Probability for quintuple spatial diversity scenario using annual and monthly statistics
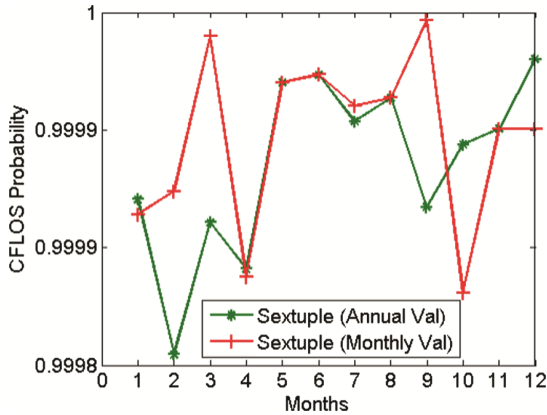
Fig. 12. Monthly joint CFLOS Probability for sextuple spatial diversity scenario using annual and monthly statistics

4   Conclusions

In this paper, a stochastic dynamic methodology for the generation of single and joint CFLOS statistics is presented taking into account the temporal, spatial, seasonal and monthly variability of clouds, the elevation angle and the altitude of the ground stations among others. This methodology is evaluated for hypothetical single optical GEO feeder links and for multiple spatially separated GEO feeder links in North and South America. In this contribution, the necessity taking into account the seasonal and monthly variability of clouds for the dimensioning of an optical ground station network is highlighted while the employment of optical ground stations located in different hemispheres is investigated.

References

1. Kaushal, H., Kaddoum, G.: Optical communication in space: challenges and mitigation techniques. IEEE Comm. Surv. Tutorials **99**, 1 (2016)
2. Panagopoulos, A.D., Arapoglou, P.-D.M., Cottis, P.G.: Satellite communications at Ku, Ka and V bands, propagation impairments and mitigation techniques. In: IEEE Communication Surveys and Tutorials, 3rd Quarter, pp. 1–13 (2004)

3. Giggenbach, D., et al.: A high-throughput satellite system for serving whole Europe with fast internet service, employing optical feeder links. In: Broadband Coverage in Germany, 9th ITG Symposium, Berlin, pp. 1–7 (2015)
4. Perlot, N., et al.: Optical GEO feeder link design. In: Future Network & Mobile Summit (FutureNetw), pp. 1–8 (2012)
5. Mengali, A., Kourogiorgas, C.I., Lyras, N.K., Shankar, B., Kayhan, M.R F., Panagopoulos, A.D., Baeumer, T., Liolis, K.P.: Ground-to-GEO optical feeder links for very high throughput satellite networks: accent on diversity techniques. Int. J. Satell. Commun. Network. Wiley (2017, Special Issue)
6. Lyras, N., et al.: Cloud free line of sight prediction modeling for optical satellite communication networks. IEEE Commun. Lett. **99**, 1 (2017)
7. Lyras, N., et al.: Cloud attenuation statistics prediction from Ka band to optical frequencies: integrated liquid water content field synthesizer. IEEE Trans. Antennas Propag. **65**(1), 319–328 (2017)
8. ITU-R P.1853-1 Tropospheric attenuation time series synthesis, Geneva (2012)
9. Luini, L., Capsoni, C.: Modeling high-resolution 3-D cloud fields for Earth-Space communication systems. IEEE Trans. Antennas Propag. **62**(10), 5190–5199 (2014)

Large Scale Site Diversity Experimental Campaign Between Greece and UK Using ALPHASAT: First Results

Apostolos Z. Papafragkakis[1], Charilaos I. Kourogiorgas[1],
Athanasios D. Panagopoulos[1(✉)] ⓘ, Spiros Ventouras[2], and Pantelis-Daniel Arapoglou[3]

[1] School of ECE, National Technical University of Athens, Athens, Greece
{apapafrag,harkour}@mail.ntua.gr, thpanag@ece.ntua.gr
[2] STFC Rutherford Appleton Laboratory, RAL Space, Oxford, UK
spiros.ventouras@stfc.ac.uk
[3] ESTEC, European Space Agency, Noordwijk, Netherlands
pantelis-daniel.arapoglou@esa.int

Abstract. There is an imminent migration of satellite communications to higher frequency bands in order to support the next generation services; this however, poses new challenges in terms of system design. With increasing frequency signal propagation becomes more prone to atmospheric phenomena and therefore accurate channel modeling is required. In this paper an ongoing propagation campaign in Ka and Q band taking place in Greece and UK using ALPHASAT's beacons is outlined. Some first results from the acquired measurements are presented, indicating the advantages of using site diversity techniques in feeder links; finally, a preliminary evaluation of large scale site diversity gains in terms of outage capacity is accomplished by means of simulated data.

Keywords: Ka-band · Q-band · Propagation · Campaign · Measurements
ALPHASAT · Beacon · Site diversity · Outage capacity

1 Introduction

The increasing demand for very high data rate services has imposed new requirements in satellite communication systems; although current satellite networks can offer capacities in the order of hundreds of Gbps, it is estimated that next generation satellite systems shall be able to support Tbps of throughput to accommodate the upcoming multimedia services [1]. The imminent migration of services to the Ka and Q/V bands is therefore well justified, particularly when taking into consideration the already congested frequency spectrum. These bands offer massive bandwidth (allegedly up to 5 GHz for the V-band feeder-link case [1]), nevertheless up the ante on system design as signal propagation at these bands can be severely impaired by the various atmospheric phenomena (gases, clouds, rain and tropospheric turbulence) [2].

To facilitate the design of high availability systems, accurate propagation modeling is of utmost importance; such models provide the means to quantify and potentially mitigate the impact of atmospheric effects on signal propagation, ensuring that the

required Quality of Service (QoS) requirements are met in an efficient manner. Producing propagation models is nevertheless a non-trivial task; it is key that they are validated against real long-term data from diverse climatic regions obtained from experimental campaigns. In the past two major experimental campaigns were conducted in Europe, one using the ESA OLYMPUS satellite (targeting Ku and Ka bands) and another one using the Italian ITALSAT F1 satellite (targeting the Q/V bands) [3]; although high-quality measurement data were acquired during these campaigns, the spatio-temporal correlation characteristics of signal propagation were not thoroughly investigated. Since 2013 the European Space Agency (ESA) coordinates the transmission of a payload dedicated to propagation measurements in Ka and Q bands using the ALPHASAT satellite. Many measurement campaigns making use of this payload are ongoing across Europe, aiming to populate the scientific databases with new, valuable data. The National Technical University of Athens (NTUA) in Greece and the STFC Rutherford Appleton Laboratory – RAL Space in the UK are currently conducting measurements in both Ka and Q bands using ALPHASAT's payload, with the objective -among others- to study frequency and site diversity schemes as will be explained in the following. Both NTUA and RAL Space are also members of the relevant ESA contract "ASALASCA" [4].

2   Gateway Site Diversity Concepts

It is well known that with increasing frequency signal propagation becomes more prone to atmospheric phenomena; even under non-rainy conditions a high fade margin is oftentimes required to account for gases, cloud attenuation and scintillation particularly for the Q/V bands. Adding rain attenuation on top of these effects lowers the system availability even further, rendering it troublesome at the very least, unless a Fading Mitigation Technique (FMT) [5] or some form of redundancy is employed.

Considering feeder links (gateway – satellite links) system availability is of over-riding significance as an outage can result into service disruption for a vast number of users; the typical availability requirement is in excess of 99.9% and the typical FMT used is uplink power control. The feeder link is designed with a high link margin (in the order of a few dB) so as to be able to compensate for possible fade attenuation effects; however, depending on the climatic region and the elevation angle, fade attenuation events can be in the order of 15–20 dB [2] for a non-negligible percentage of time. It is therefore evident that setting up a link margin alone is insufficient to maintain the high availability requirement.

A technique that can be used to mitigate the impact of severe propagation effects on feeder links is site diversity [6]. This involves either the installation of complementary, spatially separated back-up (redundant) gateways to support each already existing gateway, or in its "smart" version [7], the exploitation of already existing gateways to reroute the traffic dynamically in order to achieve the required fading link availability (Smart Gateway Diversity). In this paper, the conventional site diversity will be examined for the case of Ka- and Q-band.

In a conventional site diversity scheme, a primary (master) gateway is interconnected using a high-capacity, high-availability network (e.g. optical fiber) to a pool of auxiliary, back-up gateways each of which is spatially separated from one another. The separation distance between each pair of gateways should be greater than the maximum expected rain cell radius so as to ensure the atmospheric phenomena are uncorrelated at each site. When the master gateway experiences a deep fade e.g. as a result of intense rainfall, a controller shall register the event, enable a back-up gateway (the one experiencing the lowest atmospheric attenuation) and reroute all traffic through this gateway until the master station's signal level recovers (e.g. attenuation falls below a predefined threshold). This technique is transparent to the satellite and can dramatically increase the availability of the system (the outage probability theoretically equals the product of the single probabilities for each station, if statistical independence occurs). The overall system availability shall then be bottlenecked by the user links that traditionally merely use Adaptive Coding and Modulation (ACM) techniques.

3   Campaign Details

The novel feature of this campaign [8] is that concurrent measurements are acquired across two locations with vastly different topography and climatic characteristics, Greece and the UK. Greece is characterized by its Mediterranean climate with long-lasting, hot summers and relatively mild winters; during the winter season, very intense rainfalls (convective precipitation) usually take place at irregular intervals, however, they last no more than a few consecutive days. Also, diurnal temperature variations are very common during the whole year but particularly pronounced during the spring and summer seasons (i.e. very high temperatures during the day that drop drastically at night). On the other hand, the temperate oceanic climate experienced in the UK is characterized by cool and cloudy weather and frequent showers during almost the entire year (stratiform precipitation); temperature fluctuations across different seasons are considered small and extreme weather phenomena are infrequent.

The currently ongoing propagation measurements take place at four sites, two in Greece (Athens and Lavrion, about 36.5 km apart) and two in the UK (Chilbolton, Chilton, 47.8 km apart); at each site both Ka and Q band measurements are conducted. This configuration allows for the study of frequency diversity as well as site diversity schemes in small and large scale distances. As per common practice, the measurements are carried out using beacon signals, i.e. Continuous Wave (CW) signals of constant power and frequency; the beacon signals received at 19.701 and 39.402 GHz for the Ka and Q bands respectively are transmitted by the ALPHASAT satellite at 25.0º E. ALPHASAT -although a commercial satellite (commercial name Inmarsat-4A F4) carries payloads for experimental purposes under the coordination of the European Space Agency. Among them is the so-called Aldo Paraboni Technology Demonstration Payload 5 (TDP #5), carrying two fully redundant, coherent beacons at the aforementioned frequencies.

Despite being a geosynchronous satellite, ALPHASAT is placed on a slightly inclined orbital plane to prolong its life (maximum expected inclination less than 3.0º).

As a result, its apparent position as observed from the ground varies over time necessitating the use of antenna tracking systems. The tracking systems can take advantage of the precalculated positions of the satellite provided by means of Orbit Ephemeris Messages (OEM) files, to account for the azimuth and elevation angle changes over time at the various locations (Fig. 1).

Fig. 1. Locations of the beacon receivers used in the campaign

To allow for truly concurrent measurements, all receivers are synchronized using GPS time as reference and the received signal samples are timestamped, converted to a unified form and stored in a common database for further consolidation and processing; this allows for quick data retrieval and direct comparison between data across the different sites.

3.1 NTUA Receivers

NTUA has designed, tested and deployed two identical Ka-band beacon receivers at the NTUA Campus and at the Lavrion Technological and Cultural Park; they are currently fully operational. Another two receivers targeting Q-band have also been designed and tested and are in the process of deployment at the time of writing this. All receivers are based on the Software Defined Radio (SDR) principle, making use of high-grade off-the-shelf parts whenever possible to lower the procurement and service costs and to allow for quick provisioning.

The Ka-band receivers consist of 1.2 m offset dish antennas while the Q-band ones consist of 60 cm shrouded parabolic antennas. After undergoing filtering, amplification and down-conversion at the Low Noise Block (LNB) units, the signals are fed to a Universal Software Radio Peripheral (USRP) which further samples them, digitizes them and passes them to a computer for further processing; the computer runs a custom software based on the popular GNU Radio framework. The signal power estimation is done using a real-time FFT algorithm, providing 10 Hz of sampling rate. Apart from the beacon signal measurements, the noise power in the same bandwidth is measured and recorded [9].

All oscillators used in the receiver chains are locked to an external GPS Disciplined Oscillator (GPSDO) to ensure that extremely high frequency stability is achieved. Also, all antennas are equipped with an accurate tracking system designed in-house. The measured dynamic range for the Ka-band receivers is in excess of 39 dB; for the Q-band case it is estimated at 35 dB.

3.2   RAL Space Receivers

RAL Space has already deployed four receivers targeting both the Ka- and Q-band ALPHASAT's beacons. They are located at Chilbolton and Chilton and comprise of equipment already used successfully in past propagation measurement campaigns; all equipment is located indoors, in a specially adapted Portakabin with woven PTFE windows.

The antennas used are 50 cm in diameter, lens horn type and cassegrain type for the Ka- and Q-band respectively. Each antenna has its own (commercially procured) tracking system, ensuring that the movement of the satellite does not influence the measurements. The beacon reception is based on conventional techniques, i.e. Phase Locked Loop (PLL) envelope detection using a PLL tracking receiver configured with a noise bandwidth of 300 Hz and a tracking range of 100 kHz. The output of the PLL receiver is fed to a computer-attached Analog to Digital Converter (ADC) for further processing (sampling, digitization etc.).

The available dynamic range for the Ka-band receivers is 19 dB while for the Q-band ones is 22 dB; all receivers operate at a 10 Hz sampling rate and measure only the co-polar component, the latter being the case for NTUA receivers as well.

3.3   Ancillary Instrumentation

At all receiver locations ancillary instrumentation has been in operation since the beginning of the campaign; this allows for in-situ meteorological measurements, allowing for further study of the correlation between the fading statistics and the observed meteorological events. All ancillary measurements are synchronized to the receivers and saved with the appropriate meta-tags for later processing. Besides the recorded measurements, data from other local stations or the national weather service agencies can be accessed should they be required.

4   First Site Diversity Results Using Measurements

The following constitutes a first attempt to evaluate a small/large scale diversity scheme using concurrent measurements from the four receiver sites. The configuration is as follows: in Greece the NTUA Lavrion (LTCP) is considered the primary gateway whereas the NTUA Athens (Campus) the secondary one. Until now only Ka-band measurements are available in Greece and therefore the diversity scenario involves only Ka stations (at 19.701 GHz); in the UK the primary gateway is considered to be in Chilton while the secondary one in Chilbolton. In the UK Q-band measurements (at 39.402 GHz) are also available and therefore are the ones presented. Two days have been selected for these scenarios, namely 12/09/2016 and 28/11/2016.

4.1   Performance for September 12, 2016 (No Fades)

Regarding this particular day no fades are experienced by any station. Both primary and secondary gateways can support other gateways that are located elsewhere in the world and experience fading (Fig. 2).
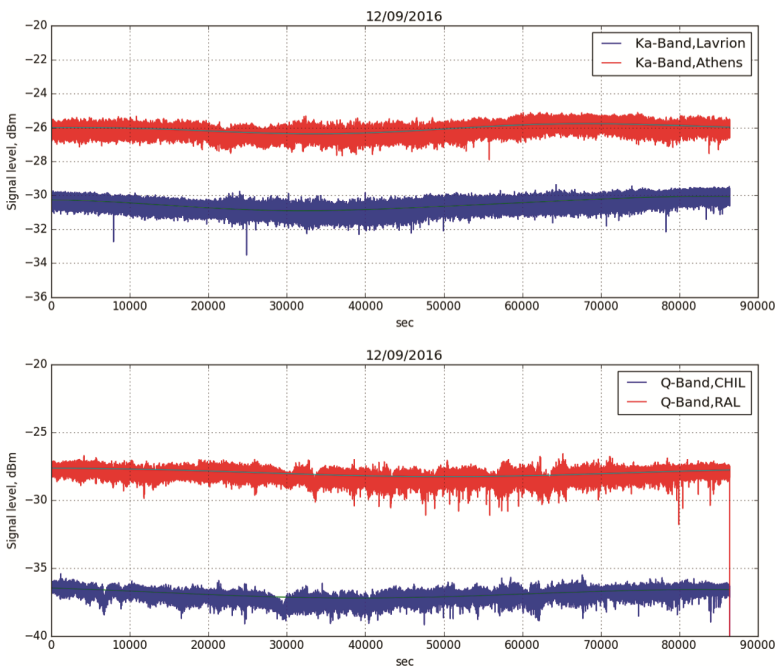
Fig. 2.   Beacon received power time series for the diversity scenario on 12/09/2016.

4.2   Performance for November 28, 2016 (Rain in Greece)

On this day the stations located in Greece experienced high attenuation as a result of intense rainfall; on the other hand, the weather in the UK remained dry and therefore the stations did not exhibit any significant fading (Fig. 3).
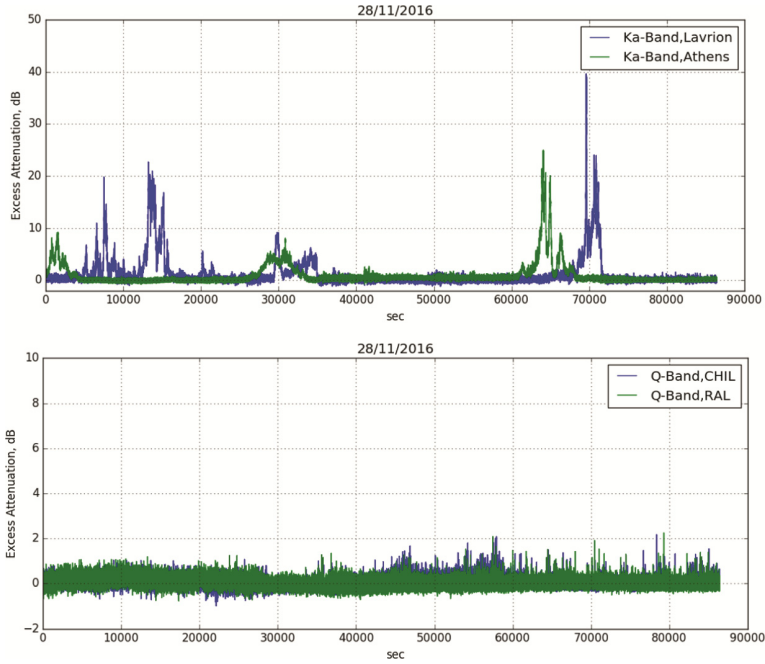
Fig. 3.   Excess attenuation time series for the diversity scenario on 28/11/2016.

With a fade margin of 10 dB for the primary station in Greece (NTUA Lavrion) and 5 dB for the secondary station (NTUA Athens) the Greek gateway achieves 100% availability on 28/11/2016 (perfect switching is assumed). During the same day the English station could support other gateways experiencing fading.

5   Capacity Evaluation Using Simulation

As the campaign is still ongoing and measurements are acquired, it is useful to have a reference baseline against which results can be compared. Using the well-established multidimensional stochastic dynamic modeling in [10], 1 year of rain attenuation data have been generated for two of the four locations, namely Athens, GR (primary gateway) and Chilton, UK (secondary gateway) at both Ka- and Q-band. This provides a clear picture of what one should expect in terms of large scale diversity gains; finally, using the methodology in [11], the expected improvement in outage capacity is presented (Figs. 4, 5, 6 and 7).
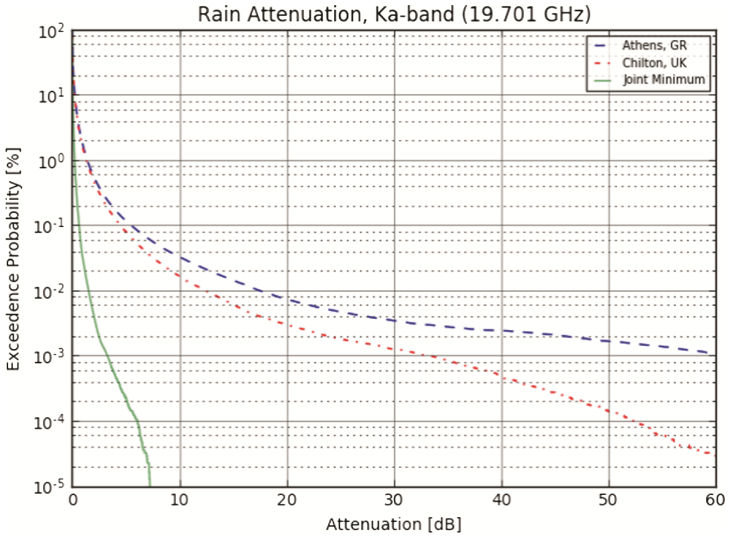
Fig. 4. Rain attenuation exceedance probabilities for Ka-band (19.701 GHz), single and joint minimum scenarios.
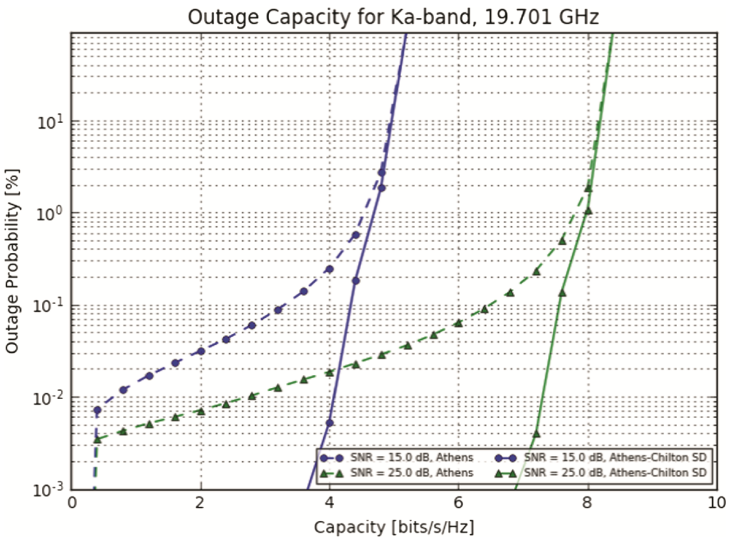
Fig. 5. Comparison between the resulting Ka-band outage capacities: a. using only the primary gateway in Athens, GR and b. using Site Diversity with the secondary station in Chilton, UK.
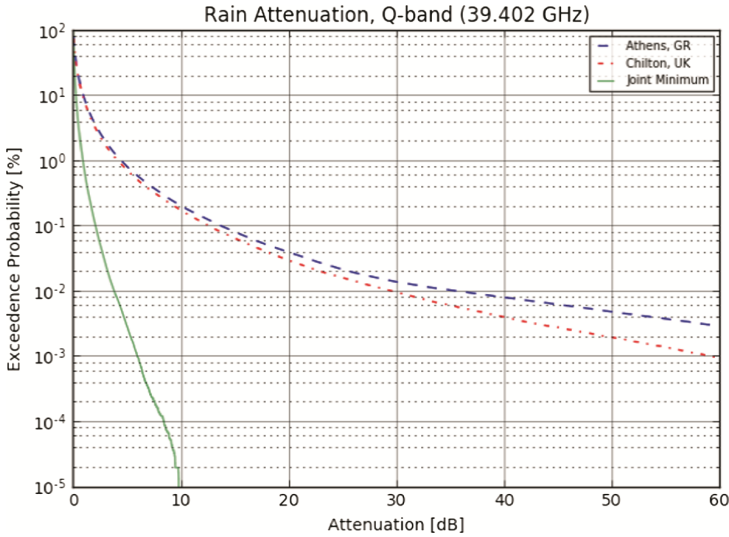
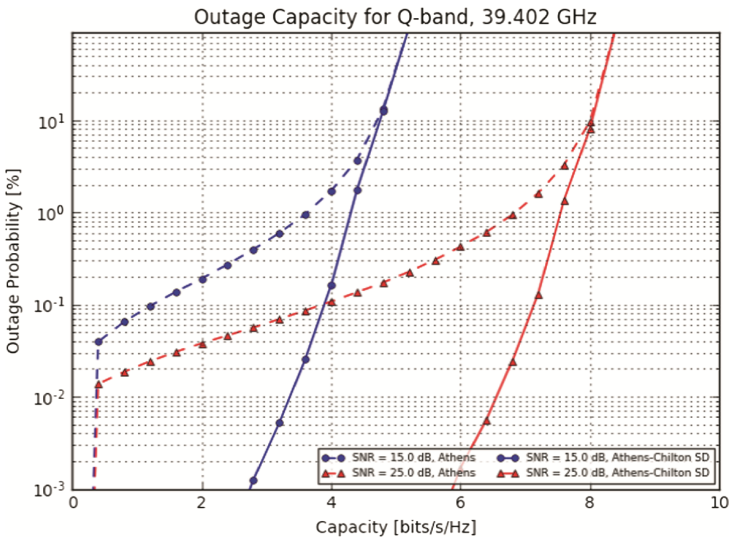**Fig. 6.** Rain attenuation exceedance probabilities for Q-band (39.402 GHz), single and joint minimum scenarios.

Fig. 7. Comparison between the resulting Q-band outage capacities: a. using only the primary gateway in Athens, GR and b. using Site Diversity with the secondary station in Chilton, UK

6   Conclusions and Future Work

The ongoing measurement campaigns in Greece and UK have been presented; as more data are collected, joint statistics on the frequency, temporal and spatial domains will

be derived; the data shall be ultimately used to populate scientific databases (e.g. ITU-R) and to develop and validate new propagation models, FMTs and small/large site diversity techniques.

References

1. Kyrgiazos, A., Evans, B., Thompson, P., Jeannin, N.: Gateway diversity scheme for a future broadband satellite system. In: 2012 6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), Baiona, pp. 363–370 (2012)
2. Arapoglou, P.-D., Shankar, M.R.B., Panagopoulos, A.D., Ottersten, B.: Gateway diversity strategies in Q/V band feeder links. In: 17th Ka Broadband Communications Conference, Palermo, Italy (2011)
3. Ventouras, S., Wrench, C.L.: Long-term statistics of tropospheric attenuation from the Ka/U band ITALSAT satellite experiment in the United Kingdom. Radio Sci., 41, RS2007, https://doi.org/10.1029/2005RS003252 (2006)
4. Ventouras, S., et al.: Large scale assessment of Ka/Q band atmospheric channel across Europe with ALPHASAT TDP5: the augmented network. In: 2017 11th European Conference on Antennas and Propagation (EuCAP), Paris, March 2017
5. Panagopoulos, A.D., Arapoglou, P.D.M., Cottis, P.G.: Satellite communications at Ku, Ka and V bands: propagation impairments and mitigation techniques. IEEE Commun. Surv. Tutor. **6**(3), 2–14 (2004)
6. Gharanjik, A., Shankar, M.R.B., Arapoglou, P.D., Ottersten, B.: Multiple gateway transmit diversity in Q/V band feeder links. IEEE Trans. Commun. **63**(3), 916–926 (2015)
7. De Gaudenzi, R., Re, E., Angeletti, P.: Smart gateways concepts for high-capacity multi-beam networks. In: 30th AIAA International Communications Satellite System Conference (ICSSC), International Communications Satellite Systems Conferences (ICSSC) (2012)
8. Papafragkakis, A.Z., Kourogiorgas, C.I., Panagopoulos, A.D., Ventouras, S.: Site diversity experimental campaigns in Greece and UK using ALPHASAT at Ka and Q band. In: Loughborough Antennas and Propagation Conference LAPC 2016, Loughborough (2016)
9. Papafragkakis, A.Z., Panagopoulos, A.D., Ventouras, S.: Combined beacon and noise satellite propagation measurements using software defined radio. In: 11th European Conference on Antennas and Propagation (EUCAP), Paris, March 2017
10. Karagiannis, G., Panagopoulos, A.D., Kanellopoulos, J.D.: Multi-dimensional rain attenuation stochastic modelling: application to earth-space diversity systems. IEEE Trans. Antennas Propag. **60**(1), 5400–5411 (2012)
11. Liolis, K.P., Panagopoulos, A.D., Arapoglou, P.D.M.: An analytical unifying approach for outage capacity achieved in SIMO and MISO broadband satellite channel configurations. In: 3rd EUCAP, pp. 2911–2915, Berlin, March 2009

Architectural Design of the Q/V Band Site Diversity Experiment Between Austria and Hungary

Michael Schmidt[1(✉)], Laszlo Csurgai-Horvath[2], Peter Horvath[2], Balint Horvath[2], Antonio Martellucci[3], and Juan Rivera Castro[3]

[1] Joanneum Research, Graz, Austria
Michael.schmidt@joanneum.at
[2] Budapest University of Technology and Economics, Budapest, Hungary
{laszlo.csurgai,peter.horvath,
balint.horvath}@hvt.bme.hu
[3] ESTEC, Noordwijk, The Netherlands
{Antonio.Martellucci,Juan.Rivera.Castro}@esa.int

Abstract. The architectural design of the Q/V band site diversity experiment between the transmit/receive ground station in Graz/Austria and the receive only ground station in Budapest is described in this paper. The ground station Budapest is connected via Internet to transfer information about the reception quality of the received signals and to request to adjust the modulation and coding. The uplink power is controlled by the station in Graz, based on the local received beacon signal from the satellite. The planed experiments and the expected results are described in the final paragraph of this paper.

Keywords: Q/V band · Site diversity · Availability
Satellite ground-station design

1 Introduction

Future High Throughput Satellites (HTS) can provide capacity in the hundreds of Gigabit/s to the user and need a broadband bidirectional access to the terrestrial network. Therefore feeder-links to and from the satellite are necessary. The Q/V band can fulfil this role perfectly because the bandwidth in this band is sufficient and no Ka band capacity between the satellite and the user terminal is needed for the feeder link. The challenge in the operation of the Q/V band is the higher fade dynamics which require fast fade mitigation techniques like Adaptive Coding and Modulation (ACM), uplink power control (UPC) and site diversity. All three will be investigated in the described system. The results shall help in the dimensioning of the future Q/V band ground stations and defining the fade mitigation parameters. This paper describes the architecture of the experiment performed via the Q/V-Band Aldo Paraboni Payload on Alphasat and the expected results.

2   Architecture

The two ground stations are about 275 km apart and have therefore no correlated weather events (non correlated weather events require at least 40 km distance [8]), which make site diversity feasible (Fig. 1). The station in Graz described in Sect. 2.1. was built in the framework of an ESA, ARTES8 projects and is operational since 2013. The station has gathered already 3 years of operational data for fade mitigation techniques and channel modeling statistics.
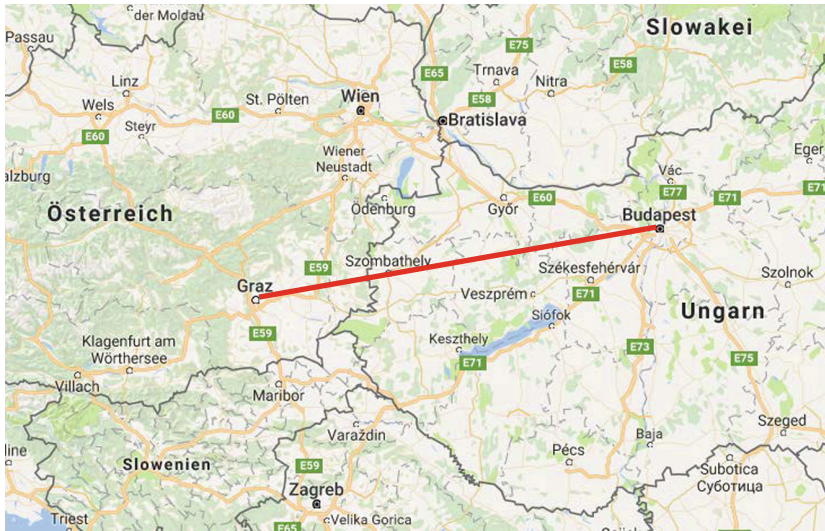

Fig. 1.   Position of the two ground-stations with a distance of 275 km (Google maps)

In Fig. 2 is the overall architecture shown. The GS in Graz transmits a DVB-S2 signal [9] with a ModCod based on the reception quality, the SNR, of the signal at the GS in Budapest. The ModCod can be changed by every DVB-S2 frame if necessary, to adapt to the changing channel conditions. It is important to mention that the dead time between the measurement and the final reception of a new ModCod has to be taken into account of the control algorithm of the ACM system. This is described in detail in [5]. The feedback from the GS Budapest to the GS Graz is via terrestrial Internet and covers the following parameters (Table 1).

The timestamp is the local time of the GS Budapest synchronized to the station's GPS-based NTP server and helps to take the dead-time correctly into account. Based on the loopback reception of the secondary beacon signal (see Sect. 2.1) at the GS in Graz, it is possible to distinguish between the uplink and downlink fading, already described in [6]. With the information about the uplink fading be can compensate fading of about 13 dB with the Up-link power control.

Based on the feedback information from the GS Budapest we can set the required ModCod to compensate the downlink fading. All transmit and reception parameters are
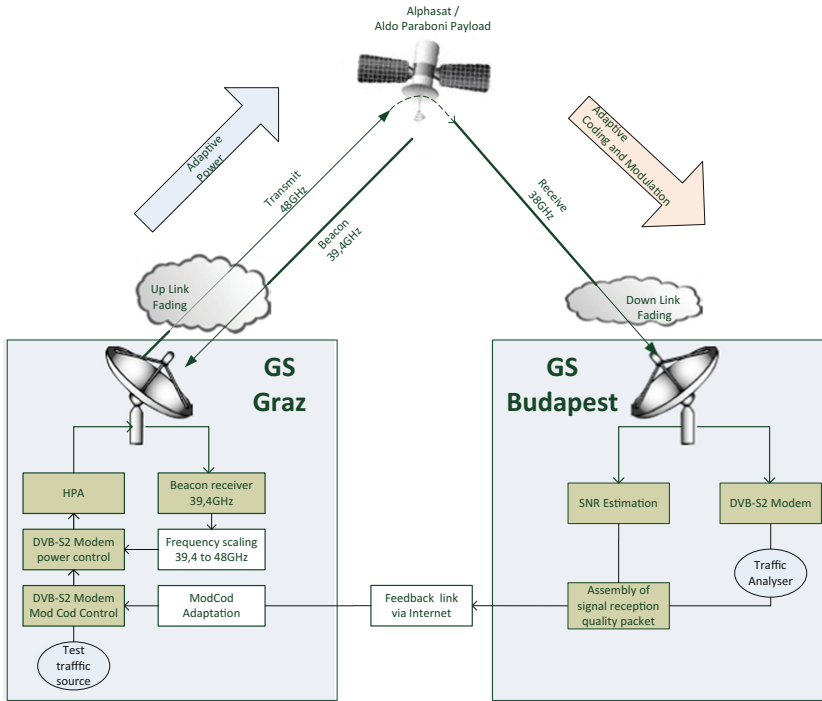
Fig. 2. Architecture of site diversity experiment between Graz and Budapest

Table 1. Exchange parameters between GS Budapest and Graz

Name	Type received	Bytes
Timestamp [s]	Unsigned int	4
Timestamp [µs]	Unsigned int	4
Received Beacon Power [dBm]	Float	4
Desired Modcod []	Unsigned int	4
Signal-to-Noise-Ratio [dB]	Float	4

recorded 24/7 and can be used for further post processing, like simulation of new improved fade mitigation techniques.

2.1 Ground Station in Graz

The communication ground-station in Graz is placed on top of a 35 m high historical tower, called Hilmwarte. The antenna is a Cassegrain system with shaped main and sub-reflector contours which meet the specified performance of a G/T of 33,8 dB/K and a gain of 61,2 dB @ 48 GHz.

In Fig. 3 you can see the block diagram of the GS in Graz [7]. Starting with the transmit path, the test data is generated in the self-designed packet generator and sent to

the transmit port of the Newtec DVB-S2 modem EL470. The HPA supplied from CPI provides a nominal power of 50 W (400 W peak) with an EIK (Extended Interaction Klystron). The HPA is connected to the V-Band up-converter via a WR22 waveguide flange. An additional monitor output is provided by the HPA. This output is connected then to a test translator (48/38 GHz), which is required for operating the ground station for test purpose in loop-back mode. The amplified signal from the HPA can either be switched to a dummy load for measurement purpose or sent to the diplexer. This module provides an isolation of 100 dB between the TX and RX port. The final module before the feed is a motor controlled polarization adjustment.
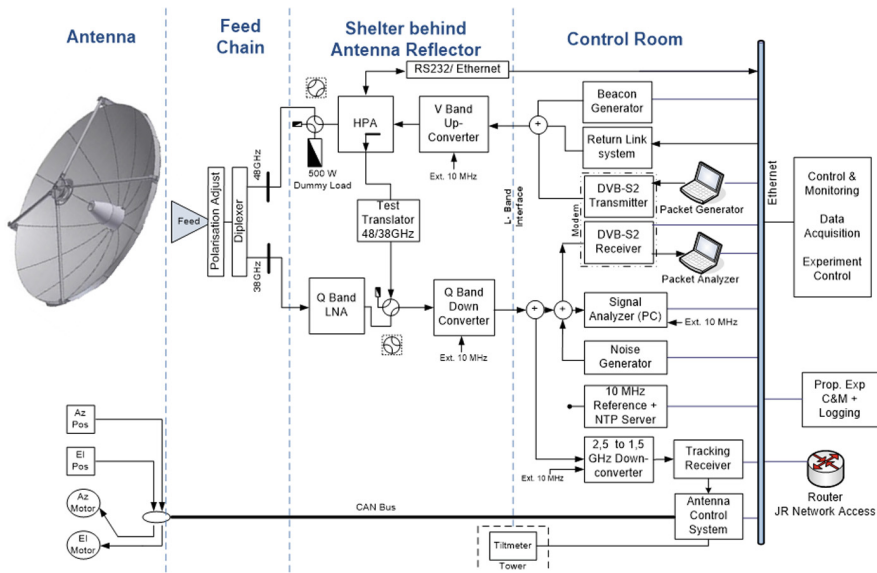
Fig. 3. Architecture of the ground station Graz

The receive part operates on the air-interface from 37,9 to 39,4 GHz. The signal passes the diplexer and arrives at the LNA (50 dB, 270 K). The down-converter can now select between the satellite signal from the LNA, and the signal from the HPA via the test translator. All waveguide switches are controlled by the control and monitoring program via a web-controlled I/O switch unit. The down-converter from Miteq allows adjustment of the drive output level in 0.1 dB steps.

Currently there are 3 experiments from the GS Graz operational on the 10 MHz transponder (Fig. 4): (1) The already mentioned DVB-S2 fade Mitigation experiment, which is also used for the site diversity experiment with Budapest. (2) We transmit a beacon from our GS and measure the reception signal power. This allows us then together with the reception of the satellite beacon to distinguish between uplink fading and downlink fading of our station in Graz. And finally (3) a low operational-point return link system which will be described in a separate paper in the future.
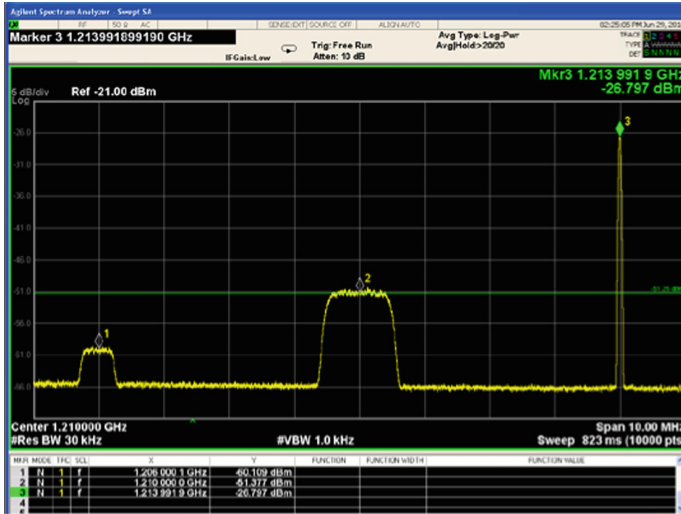
Fig. 4. Screen shot of the L-band IF spectrum transmitted from the GS Graz. The transmit frequency is 46,590 GHz higher from the values seen on the spectrum analyser markers. The 10 MHz transponder is used for 3 experiments. Left carrier (1): Low operational point, return-link system. Middle (2): DVB-S2 spectrum used for the site diversity experiment and other fade mitigation experiments. Right (3): Clean carrier transmitted from the GS which helps to distinguish between uplink and downlink fading

The 10 MHz Reference is a GPS-disciplined rubidium frequency standard source with an integrated NTP-Server. All devices will be synchronized to UTC with an average accuracy of 1 ms.

A self-designed packet generator and analyser software is executed on a PC. The packet generator loads the carrier with defined traffic pattern. The packet analyser finally measures the packet error rate (PER). The same packet analyser is used in the GS in Budapest.

The function of the signal analyser is to measure the SNR and to track the MODCOD of the Physical Layer Frames (PL-Frame) of the DVB-S2 generic stream. An accurate SNR measurement is essential for the analysis and development of Adaptive Coding and Modulation (ACM) algorithms, which can be selected as data-aided and non-data-aided approaches. To this end, we have developed a solution, which offers the possibility to test and optimize various SNR algorithms by means of software. The signal analyser is based on the GNU radio platform based on the Ettus Research HW. The HW provides an L-band interface with a down-converter and an AD converter with 100 MHz. All processing-critical operations which are based on sample information are processed in the FPGA and the symbol based signals on the attached PC platform.

2.2   Ground Station in Budapest

Figure 5 depicts the block diagram of the receive-only terminal in Budapest for the Alphasat Q/V Band Communications Experiment. The system was built to receive the satellite's EU1 beam and demodulate the DVB-S2 stream transmitted from Graz. The receiver is also capable of measuring the Q-band beacon signal transmitted at 39.402 GHz by the Alphasat Aldo Paraboni Payload and the unmodulated pilot carrier transmitted from the GS Graz along with the DVB-S2 signal.
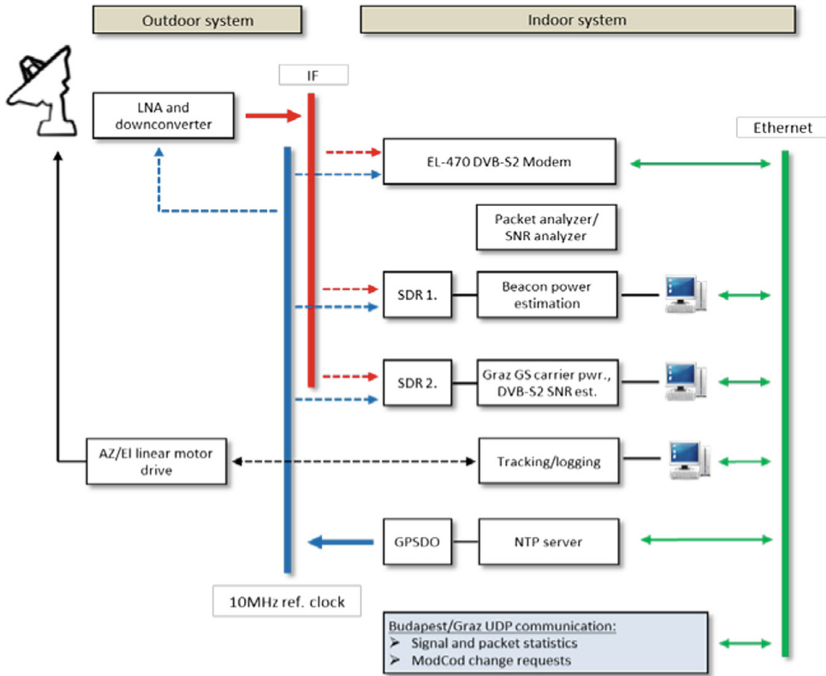
Fig. 5. System structure of the receiver terminal in Budapest

The preliminary link budget calculations that will be detailed in Sect. 2.3 have demonstrated that the system will be capable of receiving the DVB-S2 stream even under moderate/light rain conditions. This has already been proven experimentally during the initial system tests.

The main components of the receiver system are the following:

The antenna is an 1.8 m diameter, 52 dBi gain, Cassegrain type dish covered with radome [1]. The tracking system employs Az/El linear motors with program tracking, yielding $\pm 5°$ range with 0.005° resolution.

A Low Noise Block with an estimated noise figure of 2.6–3.2 dB converts the 37.85–38.15 GHz transponder band to the IF band between 1420–1620 MHz, and the 39.4 GHz beacon to 2.9 GHz [2]. The GPSDO provides 10 MHz reference clock and also serves as a Stratum-1 NTP server for data timestamping. The DVB-S2 signal is

demodulated and decoded using a Newtec EL470, DVB-S/S2 satellite modem [3]. Beacon signal power estimation, Graz GS carrier power estimation and the DVB-S2 SNR estimation are all performed using nuand bladeRF software defined radios [4]. The PC-based signal processing relies on the GNU Radio framework. One SDR receives the 10 MHz transponder passband, while a second receiver-is tuned to the down-converted beacon frequency. The packet analyzer software conveys information about the current reception conditions at the Budapest station to the experiment control PC in Graz, based on the number of correctly received DVB-S2 packets and the instantaneous SNR. The beacon power levels are also shared with the experiment control.

Figure 6 shows the spectrum of the received DVB signal in Budapest. At 2 MHz below center, the DVB-S2 spectrum can be seen, whereas the unmodulated pilot carrier is visible 2 MHz above center. The Alphasat Q-band beacon signal of the Aldo Paraboni experiment cannot be seen in this plot as it is down-converted to a frequency 1 GHz above the DVB spectrum of Fig. 6.
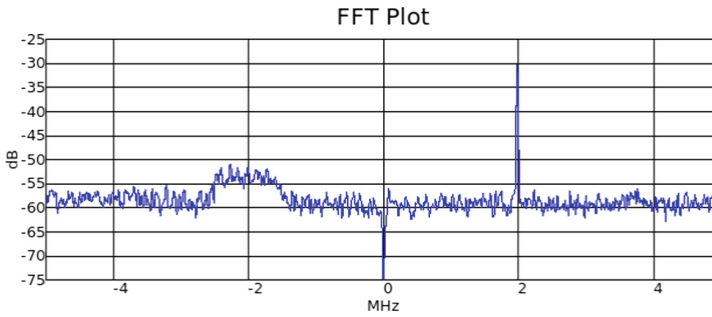

Fig. 6. The L-band IF spectrum of the received signal transmitted by the GS Graz in Budapest

2.3 Link Budget

In the receiver station in Budapest a Newtec EL470 DVB-S2 modem is applied to demodulate the signal transmitted by Alphasat. Table 2 is a link budget calculation for the Budapest GS, considering the satellite transponder parameters and the receiver factors. The contour loss is an estimation of the EU1 beam signal degradation at Budapest. Our measurements proved that the applied 2 dB is a realistic value.

The station is fully compatible with the DVB-S2 standard and because the Graz uplink station is using the same type of modem to modulate the signal, the stations are hardware compatible.

Nevertheless, the applicable ModCod combinations and baud rate depend on the station parameters and are also influenced by the propagation conditions, especially by the rain intensity at the uplink and downlink sites. In order to estimate the limitations of our system, we performed a link budget calculation, which is summarized in Table 2, assuming a symbol rate of 1 MSym/s. Clearly, in order to offset the smaller antenna gain in Budapest, one has to resort to the more robust ModCod combinations and use a smaller channel bandwidth.

Table 2. Link budget for the Budapest ground station

Parameter	Value	Unit
Frequency	38.1	GHz
Guaranteed EIRP	38	dBW
Earth–satellite distance	35756.0	km
Free-space attenuation	215.79	dB
Atmospheric loss	0.5	dB
Ionospheric loss	0.8	dB
Contour loss	2.0	dB
Receive antenna gain	52	dBi
Receiver bandwidth	1.25	MHz
Receiver noise figure	3.0	dB
Received signal power	−98.39	dBm
Received noise power	−109.03	dBm
Es/N0	10.64	dB
C/N0	71.61	dB

In the modem datasheet the manufacturer provides the required minimum E_S/N_0 values for the different ModCod combinations at 10^{-5} packet error rate. From Table 3 one can see that the station with the estimated parameters above can receive (at most) the lower 10 ModCod combinations, if the weather conditions are not too bad, at 1 Msym/s.

Table 3. Link margins for different ModCod

ModCod	Required Es/N0 [dB]	Receiver fade margin[dB]
QPSK 1/2	1.46	9,18
QPSK 3/5	2.86	7,78
QPSK 2/3	3.66	6,98
QPSK 3/4	4.36	6,28
QPSK 4/5	5.16	5,48
QPSK 5/6	5.56	5,08
QPSK 8/9	6.66	3,98
8PSK 2/3	7.16	3,48
8PSK 3/4	8.46	2,18
16APSK 2/3	9.66	0,98

3 Experiments

This project has various goals. First of all, we proved with calculations and later with test measurements that the EU1 beam can be received in Budapest with a relatively small, 1.8 m antenna. Reception test were successful, and we are currently evaluating the actual capabilities of the Budapest station compared to the predictions in Table 3.

The project is entering its operational phase in April 2017 and different experiments will be conducted to test the diversity operation and the Adaptive Coding and Modulation (ACM) mode.

One of the two SDRs in the receiver system is detecting the signal power of the TDP#5 Q-band beacon and the Graz GS unmodulated carrier in order to estimate the fading on the channel and, in addition, to distinguish the Graz uplink and Budapest downlink conditions.

The role of the packet analyzer in the Budapest receiver station is to support the ACM mode. The packet analyzer receives the IP packets from the DVB-S2 modem after the modem performed the DVB-S2 demodulation. The Graz GS transmits a well-defined packet structure, relayed by the satellite to Budapest. Analyzing the received packets permits to qualify the receiving conditions from a different aspect.

A third tool in the receiver station that supports the ACM operation is the SNR estimation, performed by the second SDR in the receiver system.

The beacon signal levels, the packet error statistics and the actual SNR value permit to implement an ACM algorithm and drive the Graz GS to switch between the available ModCod from the lowest QPSK1/2 up to the highest 16APSK2/3 allowed by the receiver system in Budapest. The ACM operation can take into account the diversity weather conditions and always ensures the optimal data transmission speed and quality between the uplink and downlink stations. The following figure is an example on the diversity conditions where the packet error rate in Budapest was influenced by the local and the remote weather conditions, respectively. The measurement was performed during the test phase of the station on 17/09/2016 (Fig. 7).

In order to ensure further data analysis the Budapest GS is logging in portable HDF5 files the following parameters (3 separated daily files are generated):

- Received packet statistics:
  - timestamp
  - number of corrected packets in the last time slot (~ 0.5 s)
  - last packet number
  - missing packet number in the last time slot
  - received packet number in the last time slot
- Signal:
  - timestamp
  - Q-band beacon power
  - Graz GS beacon power
  - Budapest GS DVB-S2 SNR
- ModCod:
  - timestamp
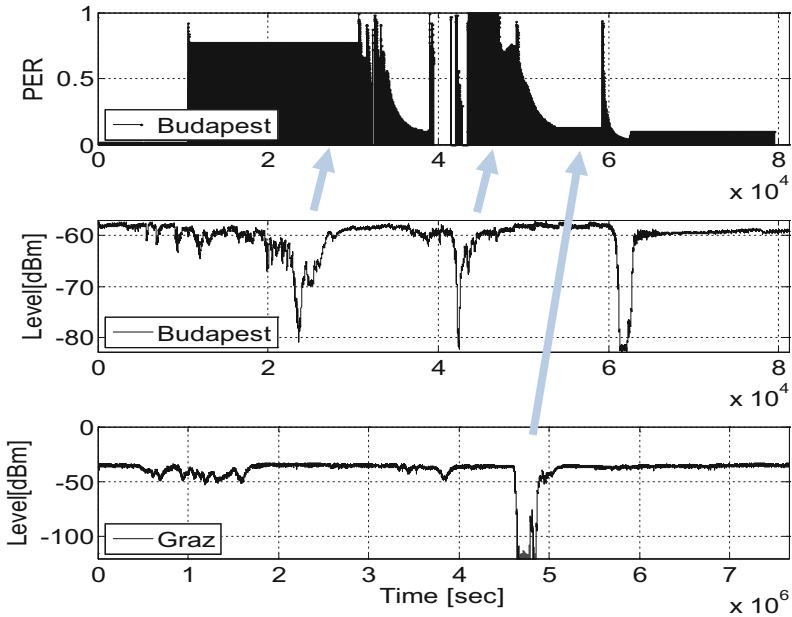 - actual ModCod at Budapest GS

Fig. 7. PER and received power in Budapest and Graz

4 Conclusion

We described in this paper the architecture of the overall site diversity experiment between the Q/V band GS in Graz and Budapest and the in detail the architectures of the two GS. The results of the measurement campaign will help in the dimensioning of future Q/V band ground stations and will contribute in a better understanding of the channel. We will further publish our results after we have gathered statistically enough data for analysis.

References

1. GRANTE Antenna Development and Production Corporation, HPA…380 Series Antenna Specifications. http://www.grante.hu
2. Totaltel Telecom Techniques Ltd. http://www.totaltel.hu
3. Newtec EL470 Satellite Modem manual ver. 3.0, Newtec Cy N.V. (2010)
4. Nuand bladeRF USB 3.0 Software Defined Radio manual, Nuand (2016)
5. Ebert, J., Schmidt, M., Kastner, S., Rivera-Castro, J.: ACM strategies for the high fade dynamics in Q/V-BAND. In: Ka Band Conference, Bologna, October 2015

6. Schmidt, M., Schlemmer, H., Ebert, J., Kastner, S., Rivera Castro, J.: Up-link power control strategies for a Q/V band ground station. In: Ka Band Conference (2015)
7. Schmidt, M., Schlemmer, H., Ebert, J., Kückelheim, M., Rivera Castro, J.: Q/V band ground station for Alphasat TDP5 telecommunication experiment-design and verification. In: 20th Ka and Broadband Communications, Navigation and Earth Observation Conference, Salerno, Italy, October 2014
8. ITU-R P. 618-5, Propagation data and prediction methods required for the design of earth-space telecommunication systems
9. ETSI EN 302 307-1 V1.4.1, Digital Video Broadcasting (DVB-S2)

Flexible Capacity Allocation in Smart Gateway Diversity Satellite Systems Using Matching Theory

Anargyros J. Roumeliotis[1], Charilaos I. Kourogiorgas[1],
Argyrios Kyrgiazos[2], and Athanasios D. Panagopoulos[1(✉)] [ID]

[1] National Technical University of Athens, Athens, Greece
{aroumeliot,harkour}@mail.ntua.gr,
thpanag@ece.ntua.gr
[2] University of Surrey, Surrey, UK
a.kyrgiazos@surrey.ac.uk

Abstract. In this paper, we investigate the performance of multi-beam high throughput satellite systems that employ the smart gateway diversity concept. The multi-beam satellite system is examined under the satellite system's capacity losses, the impact of the channels' atmospheric effects, the offered capacities of gateways and finally the requested capacities of users' beams. To ameliorate the system's performance, a novel flexible resource allocation algorithm is proposed which is based on the matching theory concept. Gateways and users' beams are matched according to the aforementioned factors applying the deferred acceptance algorithm. Simulation results illustrate that the proposed matching scheme outperforms a less sophisticated fixed, i.e. without resource management, allocation scheme in terms of system's capacity losses, rendering the former a very promising allocation mechanism for future satellite systems.

Keywords: Smart gateway diversity · Resource allocation
Multi-beam satellite · Matching theory · Deferred acceptance algorithm

1 Introduction

The wireless users' demands for broadband internet services are growing extremely due to the enormous technological progress in devices such as smartphones, tablets and laptops. Cisco predicts that 78% of the world's mobile data traffic will be video by 2021 [1]. The terrestrial LTE-Advanced (4G) cellular systems are not able to follow the tremendous demand in video services. Afterwards, to fulfill the prospects of the aforementioned synchronous challenges in wireless communication networks a turn into the exploitation of satellite networks has been made. Especially, satellite systems are ideal for rural and suburban areas where the terrestrial networks have limited connectivity due to the less number of subscribers.

The satisfaction of users' high bit rate requirements gives rise to multi-beam satellite systems in which each gateway (GW) serves a number of user beams and makes feasible the exploitation of gateway diversity techniques. The application of

satellite communication networks for broadband internet satellite services has been included in the second generation of Digital Video Broadcasting for Satellite Transmission DVB-S2 standard and in its extension DVB-S2X [2]. The exploitation of Ka Band (20–30 GHz) for satellite transmission and the multi-beam earth coverage are efficient ways to support the users' high throughput demands that need extra bandwidth resources at a significant lower cost for satellite operators.

The huge traffic demands lead to very high data rates for future satellite communication systems which target to values of hundreds Gigabits/s up to Terabit/s [3, 4]. Hence, higher frequency bands, such as Q/V Band (40–50 GHz), are also investigated. Especially, in these systems the use of Q/V Band at the feeder link (the link between the gateway and the satellite) and the Ka Band at the user link (the link between the satellite and the user (UE) beam) are examined. However, in case that the operating frequency is above 10 GHz, the performance of the system is deteriorated by the atmospheric effects [5]. Rain attenuation is the dominant fading mechanism and can exhibit several dBs for a small but crucial for the availability of the systems time percentage.

To deal with the severe signal degradations in Ka and Q/V Bands and for a more efficient use of ground and space resources, smart gateway diversity paradigms are proposed [4, 6, 7] and developed for the mitigation of attenuation taking into account the spatial and temporal behavior of the atmospheric effects and radio channel conditions. In the smart gateway techniques the service of a user beam by a specific gateway can change and be routed to another gateway, if the former's link is unable to serve the user beam's requirements. More specifically, there are two types of smart gateway diversity schemes. The first scheme is called N + P scheme where there are N active GWs and P redundant. A GW of the P redundant is set on operation in case that one of the first N gateways in on outage due to the atmospheric attenuation.

The second scheme is called N-active scheme, all gateways are active, but in case that one of the gateways goes in outage, the traffic is served forwards to the other active GWs while the bit rate decreases since smaller bandwidth is used. The N-active scheme is separated into frequency multiplexing in which each UE beam is served simultaneously by different GWs in different frequency carriers and time multiplexing in which each UE beam is served by different GWs in different time slots. The latter architecture is similar to the implemented Satellite Switched TDMA (SS-TDMA).

In [6] the N-active time multiplexing scheme is investigated under a flexible resource allocation management in a multi-beam SS-TDMA system with four GWs and four UE beams. Particularly, the matching among the UE beams and GWs is examined based on the UEs' capacity requirements and the GWs' offered capacities. Three different allocation algorithms are proposed targeting to the maximization of throughput and to balance the traffic between the UE beams considering the feeder links' propagation conditions. The simulation results prove the outperformance of proposed algorithms in terms of system's losses compared with a fixed allocation scheme. In [7] besides the performance investigation of the aforementioned resource allocation algorithms, the availability performance in N + P and N-active schemes is also examined.

Recent literature, such as [6, 7], shows the importance of flexible resource allocation algorithms for having better performance towards to the direction of High Throughput Satellite (HTS) communication networks. Considering the above aspect, this paper investigates the use of matching theory in satellite communication systems

which is applied in such systems for first time according to authors' best knowledge. Matching theory [8, 9] is a mathematical tool which describes a distributed and self-optimizing approach to take resource management decisions efficiently. This theory, based on the two sided approach, contains two different sets of agents and creates mutually beneficial relations among them through the application of an efficient stable matching algorithm, known as deferred acceptance algorithm (DAA) [9]. In our case the satellite networks can be modeled as matching markets, where each agent of one set, i.e. GW/UE, ranks the agents of the other set, i.e. UEs/GWs (correspondingly), using a preference relation which is based on their utility functions. Moreover, it is noticeable that matching theory has a wide practical impact through its implementation in many real world systems, such as the U.S. National Resident Matching Program.

In this paper the flexible resource allocation, based on GWs' offered capacities and UE beams' requirements, in a multi-beam HTS SS-TDMA network is investigated using the matching theory. In Sect. 2, a brief description of time series synthesizer proposed in [10] is given and in Sect. 3, the basic concepts of matching theory are described to have a better understanding. In Sect. 4, the system model is described in detail while in Sect. 5, numerical results of the performance of our proposed matching algorithm are given. Finally in Sect. 6, conclusions derived from the present paper are described.

2 Multi-dimensional Rain Attenuation Synthesizer

In this Section we describe the generation of time series of rain attenuation which has been considered for modeling the links' propagation conditions and extracting the corresponding Carrier-to-Noise-and-Interference Ratio (CNIR).

In [11] a first-order stochastic differential equation (SDE) is proposed for the generation of time series of rain attenuation, called as the Maseng-Bakken model (M-B). The main assumptions of the model are that rain attenuation follows the lognormal distribution and that the rate of change of rain attenuation is proportional to the instantaneous value of rain attenuation. In [10], based on the SDE of [11], a multidimensional SDE is presented for simulating correlated in spatial domain rain attenuation time series at multiple sites which has the following form:

$$d\mathbf{a}_t = \mathbf{F}(\mathbf{a}_t)dt + \mathbf{Z}(\mathbf{a}_t)d\mathbf{W}_t \tag{1}$$

where $\mathbf{a}_t = \left[a_t^1, \ldots, a_t^n\right]^T$ is the vector of rain attenuation at the different sites. Rain attenuation on each link $a_t^i, i = 1, \ldots, n$ follows the lognormal distribution with standard deviation of its natural logarithm σ_{a_i} and median value a_{m_i}. The joint distribution is joint lognormal with covariance matrix $\mathbf{C}$. The other coefficients of (1) are:

$$\begin{aligned}
\mathbf{F}(\mathbf{a}_t) &= \left[F_1\left(a_t^1, \ldots, a_t^n\right), \ldots, F_n\left(a_t^1, \ldots, a_t^n\right)\right]^T \\
\mathbf{Z}(\mathbf{a}_t) &= \left[z_{ij}\left(a_t^1, \ldots, a_t^n\right)\right]_{1 \le i,j \le n} \\
\mathbf{W}_t &= \left[W_t^1, \ldots, W_t^n\right]^T
\end{aligned} \tag{2}$$

with $\mathbf{W_t}$ the n-dimensional Brownian Motion, which means that $W_t^1, \ldots, W_t^n$ are independent Brownian Motions [12]. According to the model presented in [10], the coefficients $\mathbf{F}$ and $\mathbf{Z}$ must be chosen in such way that the diffusion coefficient of each stochastic process is proportional to the instantaneous value of rain attenuation on the specific link and second the transformation of (3) will yield to the form of (4).

$$X_t^i = \ln\left(\frac{a_t^i}{a_{m_i}}\right), i = 1, \ldots, n \tag{3}$$

$$d\mathbf{X}_t = \mathbf{B}\mathbf{X}_t dt + \mathbf{S}d\mathbf{W}_t \tag{4}$$

where $\mathbf{X}_t = \left[X_t^1, \ldots, X_t^n\right]^T$ is the vector of the random process X_t^i, with initial values $\mathbf{X}_0 = \mathbf{x}_0$ and $\mathbf{B}$ is a n $\times$ n matrix with elements $b_{ij} = -\beta_i \delta_{ij}$ where δ_{ij} is the Kronecker delta function and β_i is the dynamic parameter as this is defined on the M-B model. The analytical solution of (4) is [10, 12]:

$$\mathbf{X}_t = e^{t\mathbf{B}}\mathbf{X}_0 + e^{t\mathbf{B}} \int_0^t e^{-s\mathbf{B}}\mathbf{S}d\mathbf{W}_s \tag{5}$$

Assuming a correlation coefficient as the one presented in [13] the covariance matrix of the random process $\mathbf{X}_t$ is:

$$\mathbf{C_X} = \begin{bmatrix} \sigma_{a_1}^2 & \cdots & \rho_{n,1n}\sigma_{a_1}\sigma_{a_n} \\ \vdots & \ddots & \vdots \\ \rho_{n,n1}\sigma_{a_n}\sigma_{a_1} & \cdots & \sigma_{a_n}^2 \end{bmatrix} \tag{6}$$

Therefore, for (4) denoting:

$$\mathbf{G} = \mathbf{S}\mathbf{S}^T \tag{7}$$

it holds from [10], (5) and the theory of multidimensional SDEs [10] that:

$$[G]_{ij} = \left(\beta_i + \beta_j\right)[C_X]_{ij} \tag{8}$$

So, in order to fulfill the requirements denoted at the beginning of this Section the coefficients $\mathbf{F}$ and $\mathbf{Z}$ must be:

$$F_i(\mathbf{x}) = x_i\left[-\beta_i \ln\frac{x_t^i}{a_{m_i}} + \frac{1}{2}\sum_{j=1}^n s_{ij}^2\right] \tag{9}$$

and

$$z_{ij}(\mathbf{x}) = x_i s_{ij} \tag{10}$$

The methodology for the calculations of the rain attenuation time series is fully described in [11].

3   Basic Concepts of Matching Theory

In this Section some basic concepts of matching theory are provided in order to be able to understand thoroughly the proposed matching scheme which leads to the mutually acceptable and efficient GW-UE beam pairing. In following notions we use the terms UE and UE beam interchangeably. The matching theory is applied to two different sets of agents, i.e. GWs and UE beams and is based on specific utility functions, defined below as U, according to their preferences. Due to the fact that in our scenario each GW can serve at most one UE beam and each UE beam can be served by at most one GW, the investigated matching is called as one-to-one matching approach.

The fundamental element of matching theory is the matching function μ, which is formally defined as follows for the investigated system scenario:

Definition 1   [8]: A matching μ is a function from the set **GWs**$\cup$**UEs** into the set of unordered families of elements of **GWs**$\cup$**UEs** such that:

1. $|\mu(UE)| = 1$ for every UE $\in$ **UEs** and $\mu(UE) = $ UE if $\mu(UE) \notin$ **GWs**.
2. $\mu(UE) = $ GW if and only if UE is in $\mu(GW)$.

Definition 2: Let M be the number of GWs and N the number of UEs. For any UE_n, a preference relation $\succ_n$, which consists a complete, reflexive and transitive binary relation between GWs and UEs, is defined over the set of M GWs such that, for any two GWs $i, j \in M$, $i \neq j$ and two matchings $\mu, \mu' \in M \, x \, N$, $i = \mu(n)$, $j = \mu'(n)$:

$$(i, \mu) \succ_n (j, \mu') \Leftrightarrow U_{UE_n,i}(\mu) > U_{UE_n,j}(\mu') \tag{11}$$

The preference relation for GWs $\succ_m$ is defined similarly.

Each GW (UE) constructs a preference list of the UEs (GWs) based on its utility function according to (12), (13). Especially, the GWs' and UEs' preference vectors include the corresponding ordered lists $prefGW_k = \{UE_{k(i)}\}_{i=1}^{N}$ and $prefUE_j = \{GW_{j(i)}\}_{i=1}^{M}$, that are constructed according to the following conditions respectively:

$$U_{GW,k}(UE_{k(1)}) > \ldots > U_{GW,k}(UE_{k(N)}) \tag{12}$$

$$U_{UE,j}(GW_{j(1)}) > \ldots > U_{UE,j}(GW_{j(M)}) \tag{13}$$

The solution of the matching game is based on the DAA algorithm, described briefly in Sect. 4, according to which in a two-sided, one-to-one matching game a stable matching always exists [9]. It should be noted that a matching can be regarded as stable only if it leaves no pair of players on opposite sides who were not matched to each other but would both prefer to be. Finally, the runtime complexity of this algorithm is $O(MN)$ [9].

4   System Model

In this paper, we consider a multi-beam HTS SS-TDMA system with M GWs and N UEs' beams as depicted in Fig. 1. In a specific time instance each UE beam can be served by one GW and each GW can serve only one UE beam. The system scenario is investigated per a SS-TDMA frame period of 1 s similar to [6, 7], considering that the links' propagation conditions change slowly during the duration of 1 s [14]. Furthermore, the links between the satellite and UEs' beams are considered ideal and identical to investigate exclusively the performance of feeder links [6, 7].
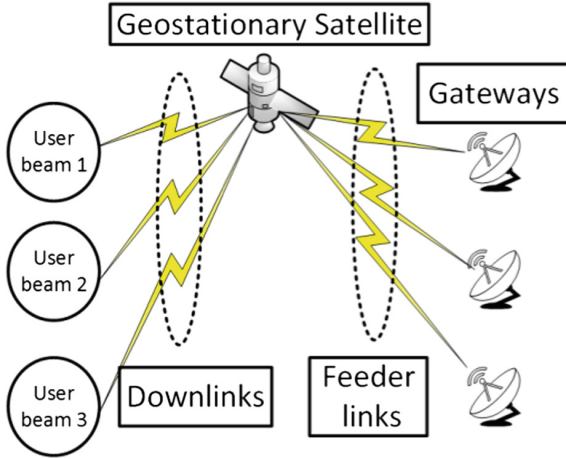
Fig. 1. System model

In each frame period each GW_i, $i = 1, \ldots, M$, offers capacity, defined as OC_i in bps, to UEs' beams. The offered capacity OC_i is expressed by the Shannon formula as follows:

$$OC_i = B_C \log_2(1 + \gamma_i) \tag{14}$$

where B_C is the carrier bandwidth and $\gamma_i^{-1} = CNIR_{up,i}^{-1} + CNIR_{dn}^{-1}$ is the total or end-to-end $CNIR_i$ including both the CNIR of feeder link i and the CNIR of downlink which is invariant in our scenario. Especially $CNIR_{up,i} = CNIR_{CS,i} 10^{-A_i/10}$, the $CNIR_{CS,i}$ is the CNIR in clear sky conditions for the feeder link i and A_i is the rain attenuation of the same link. Finally, we must consider that in above notions $CNIR = C/(N + I)$. In terms of UEs' beams, in each frame period the requested capacity of UE_j $j = 1, \ldots, N$ is defined as RC_j. Afterwards, the capacity losses are expressed as [6, 7]:

$$L = \sum_{j=1}^{N} \max\{RC_j - OC_j, 0\} \tag{15}$$

According to our proposed matching algorithm and based on (12), (13) and the expressions of offered and requested capacities, each GW sorts the UEs in a descend order, from larger to smaller values, based on latter's requested capacities. Hence, GWs are more willing to match with the UEs which request higher capacities. On the other hand, each UE sorts the GWs in a descend order based on latter's offered capacities. Consequently, UEs are more willing to match with GWs that offer higher capacities. This process results in the construction of preference vectors for each GW and UE beam.

We apply the DAA algorithm [9] which involves a number of "iterations". In first iteration each GW proposes to its most favorite UE, the first UE in the GW's preference list. Each UE matches with its most preferable GW, the first GW in the UE's preference list and rejects the proposals of the rest GWs. In each subsequent iteration, every rejected GW from the previous iteration proposes to the most preferred UE to whom it has not yet proposed. Each UE rejects proposals from unacceptable GWs and rejects also its current paired GW, if there is a proposal from a more preferable GW. The algorithm finishes when either all the GWs have been matched with UEs, or there are GWs that were rejected by all UEs acceptable to them.

Afterwards, we result in the GWs-UEs pairs and compute the system losses considering the requested capacity of each UE and the offered capacity to this UE, according to (15), after the matching process. Finally, our proposed pairing scheme is compared with a less sophisticated fixed allocation scenario where each feeder link is connected for equal number of frames to each downlink, without requiring any resource management.

5 Simulation Results and Discussion

For the numerical results we assume that there are four gateways in locations that can be found in [6, 7] and four UE beams. Every link between the gateway stations and the satellite operate at 50 GHz, the bandwidth $B_C = 1$ GHz, the uplink CNIR in clear sky conditions is 25 dB, assuming that it is the same for all the four gateways, while the corresponding CNIR for the downlink is 13 dB. Furthermore, the UEs' requested capacities are drawn from a uniform distribution in the range $(0, B_C \log_2(1 + CNIR_{\max}))$ bps where $CNIR_{\max}^{-1} = CNIR_{CS,up}^{-1} + CNIR_{CS,dn}^{-1}$. For the simulation of the rain attenuation time series, the statistical parameters of the lognormal distribution for each site are derived from the methodology described in [15] and the dynamic parameter of rain attenuation from [16, 17]. Finally, the system's performance in terms of capacity losses is investigated over a year.

Time series of rain attenuation for four GWs' locations similarly with the papers [6, 7] are evaluated and the outage probability of each GW may be computed by $P_{out} = P[CNIR \leq CNIR_{th}]$ indicating the time percentage that the CNIR of the GW is below a given threshold. Moreover, from the time series of GWs' CNIR, originated by the rain attenuation time series through the Shannon formula, the capacity time series for every link can be derived.

Figure 2 presents the gateways' transmitted capacities and the system's capacity losses based on (15). In the first subfigure the total offered capacity by gateways is depicted over time. In the second subfigure the capacity losses of proposed matching algorithm are illustrated, while the third subfigure shows the losses of the fixed allocation scheme. We remark that, generally, in deep fades the losses increase for both schemes, but the proposed matching mechanism results in fewer losses' events with lower values of capacity losses compared with the fixed mechanism.
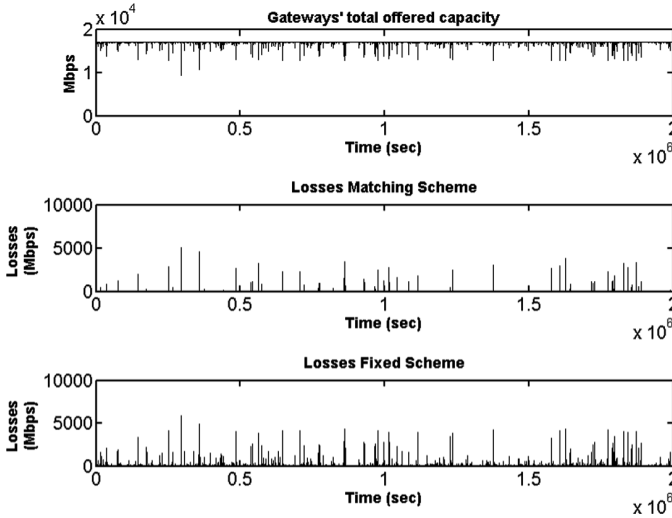
Fig. 2. A time series snapshot of gateways' offered capacities and system's capacity losses under matching and fixed allocation schemes.

Figure 3 presents the complementary cumulative distributions of the losses for both mechanisms as a percentage of instantaneous traffic demands. In Fig. 3(a) we examine the system without considering $CNIR_{th}$, assuming all the values of capacity losses over a year. The superiority of the proposed matching mechanism, resulting in much smaller probability losses than the fixed scheme, is obvious. In Fig. 3(b) we investigate the performance of both schemes for different probabilities $P = \{10\%, 1\%, 0.1\%\}$ resulting in $CNIR_{th} = \{12.63, 11.86, 3.72\}$ dB respectively evaluated from the corresponding capacity time series. The distributions contain the annual time instances for which $min\{CNIR_{GW_i}\} < CNIR_{th}$, $i = 1, \ldots, 4$. The probability that the minimum of CNIR is less than a threshold gives the probability that at least one of the link has CNIR less than a defined threshold. It is remarkable that the matching scheme outperforms the fixed allocation for all different values of $CNIR_{th}$. Finally we obtain that as $CNIR_{th}$ decreases, resulting in more severe fading conditions, the probability of system's capacity losses increases for both mechanisms.
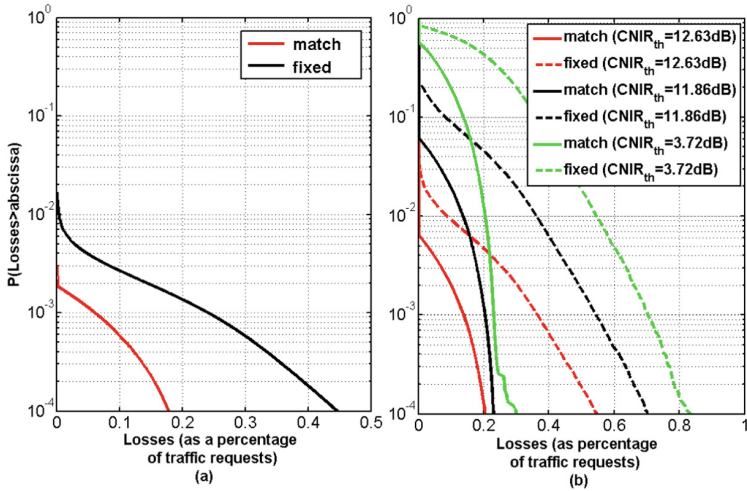
Fig. 3. Complementary cumulative distributions of the losses for both mechanisms, (a) without considering $CNIR_{th}$ and (b) with three different $CNIR_{th}$ values.

6 Conclusion

In this paper a novel flexible capacity allocation mechanism, based on the matching theory concept, in a multi-beam HTS SS-TDMA system has been proposed. The proposed mechanism relies on a smart gateway N-active time multiplexing scheme. For making appropriate and efficient matching among gateways and user beams the channels' propagation conditions, originated by a multi-dimensional rain attenuation synthesizer, the offered gateways' capacities and requested users' capacities are considered. The matching between the gateways and users' beams applies the deferred acceptance algorithm. Finally, the simulation results prove the superiority of the proposed scheme compared with a fixed allocation scheme in terms of system's capacity losses. The matching scheme results to fewer losses events with lower values of capacity losses compared with the fixed scheme.

References

1. Cisco visual networking index: Global Mobile Data Traffic Forecast Update 2016–2021. Cisco White Paper (2017)
2. ETSI EN 302307-2: Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications; Part II: S2-Extensions (S2-X) (2014)
3. Kyrgiazos, A., Evans, B., Thompson, P., Mathiopoulos, P.T., Papaharalabos, S.: A terabit/second satellite system for European broadband access: a feasibility study. Int. J. Satell. Commun. Netw. **32**(2), 63–92 (2014)

4. Jeannin, N., Castanet, L., Radzik, J., Bousquet, M., Evans, B., Thompson, P.: Smart gateways for terabit/s satellite. Int. J. Satell. Commun. Netw. **32**(2), 93–106 (2014)
5. Panagopoulos, A.D., Arapoglou, P.D.M., Cottis, P.G.: Satellite communications at Ku, Ka and V Bands: propagation impairments and mitigation techniques. IEEE Commun. Surv. Tutor. **6**(3), 2–14 (2004)
6. Kyrgiazos, A., Thompson, P., Evans, B.G.: Gateway diversity via flexible resource allocation in a multibeam SS-TDMA system. IEEE Commun. Lett. **17**(9), 1762–1765 (2013)
7. Kyrgiazos, A., Evans, B.G., Thompson, P.: On the gateway diversity for high throughput broadband satellite systems. IEEE Trans. Wirel. Commun. **13**(10), 5411–5426 (2014)
8. Roth, A.E., Sotomayor, M.A.O.: Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Cambridge University Press, Cambridge (1992)
9. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. Am. Math. Mon. **69**(1), 9–14 (1962)
10. Karagiannis, G.A., Panagopoulos, A.D., Kanellopoulos, J.D.: Multidimensional rain attenuation stochastic dynamic modeling: application to Earth–space diversity systems. IEEE Trans. Antennas Propag. **60**(11), 5400–5411 (2012)
11. Maseng, T., Bakken, P.M.: A stochastic dynamic model of rain attenuation. IEEE Trans. Commun. **29**, 660–669 (1981)
12. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Springer, New York (2005). https://doi.org/10.1007/978-1-4612-0949-2
13. Paraboni, A., Barbaliscia, F.: Multiple site attenuation prediction models based on the rainfall structures (mes- or synoptic scales) for advanced TLC or broadcasting systems. In: XXVII URSI General Assembly (2002)
14. Van de Kamp, M.M.: Statistical analysis of rain fade slope. IEEE Trans. Antennas Propag. **51**(8), 1750–1759 (2003)
15. Kanellopoulos, J.D., Koukoulas, S.G.: Analysis of the rain outage performance on route diversity systems. Radio Sci. **22**(4), 549–565 (1987)
16. Kourogiorgas, C.I., Panagopoulos, A.D., Kanellopoulos, J.D., Arapoglou, P.D.M.: Rain attenuation time series synthesizer for leo satellite systems operating at Ka Band. In: Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), pp. 119–123. IEEE (2012)
17. Panagopoulos, A.D., Kanellopoulos, J.D.: On the rain attenuation dynamics: spatial-temporal analysis of rainfall-rate and fade duration statistics. Int. J. Satell. Commun. Netw. **21**(6), 595–611 (2003)

Author Index