Chapter 3
QoS Aware Service Provisioning and Resource Distribution in 4G/5G Heterogeneous Networks

Rintu Nath

Introduction

The 4G technology gives enhanced features to mobile communication and foster creation of associated applications and service ecosystem to revolutionize ICT. Growth of Internet of Things (IoT) is generating high volume data and causing congestion of network. Unpredictable data flow from billions of connected devices will affect throughput and may cause deterioration of the Quality of Service (QoS) [1]. Due to large scale deployment of IoT technologies and ever-increasing number of mobile users on 4G and 5G networks, increasing channel efficiency while ensuring QoS constraints is challenging. The new age mobile communication system is driven by heterogeneous applications, variable user requirement and unpredictable data rate. Deployment of services with guaranteed QoS and increased spectrum efficiency are often conflicting and, hence challenging.

Advancement in 5G network is expected to increase spectrum efficiency with higher throughput. However, design constraints for QoS provisioning need to be addressed for a reliable 5G network. Architecture, framework, and scheduling algorithm imposing new design challenges for 5G. Delay bounded QoS requirements for high volume data is difficult to fulfill in a 4G network. To characterize QoS for delay bound high-volume data, homogeneous statistical provisioning is done that guarantees QoS for each link. However, for a 5G network, different delay bound heterogeneous data traffic having different QoS constraints can be guaranteed.

R. Nath (✉)
Vigyan Prasar, Department of Science and Technology, AI Building, Technology Bhawan, New Mehrauli Road, New Delhi 110016, India
e-mail: Rnath@vigyanprasar.gov.in

V. Kadyan et al. (eds.), *Deep Learning Technologies for the Sustainable Development Goals*, Advanced Technologies and Societal Change,
https://doi.org/10.1007/978-981-19-5723-9_3

When a network is configured for a large buffer, it becomes unstable and subsequent delay in packet-switched network cause bufferbloat. Variation in packet delay causes jitter and overall throughput reduces. Applications like online gaming, Voice over IP (VoIP), online transactions become unreliable due to bufferbloat and jitter. To overcome these problems and to ensure QoS, several solutions exist. Solutions are primarily categorized in two groups, end-to-end and in-Network solutions. Congestion Control Algorithms (CCA) in 4G and 5G networks try to ensure high throughput and low latency. However, CCAs need to have fairness while interacting with different networks and should only be deployed after QoS parameters are ensured.

The aim of this chapter is to discuss various service provisioning techniques and throughput constraint resource allocation that guarantees QoS for a diverse set of services, applications, and user requirement in 4G and 5G networks. Simulation of heterogeneous statistical delay bounded QoS provisioning with differential baud rate and fading parameters are done and results discussed.

Related Work

Challenges of 4G and 5G networking for QoS provisioning, resource allocation, information flow, coding and modulation schemes, and resource management are listed in [2, 3].

There is a fundamental tradeoff between QoS provisioning and effective throughput. Tang et al. [4] reported power and rate adaptation scheme to maximize system throughput for a given delay QoS constraint. The scheme is applied on a block fading channel model in which higher channel correlation gives faster convergence of power-control policy and stringent adaptation of QoS. Call Admission Control (CAC)-based QoS provisioning in a resource management framework is reported by Inaba et al. [5]. Beshley et al. [6] proposed another resource management framework with end-to-end QoS management algorithm in 4G/5G networks. The authors proposed a modified architecture of Long-Term Evolution (LTE) for IoT services. Haile et al. [7] discussed an end-to-end congestion control approach for 4G and 5G networks that guarantees high throughput. Various dynamic network slicing and resource allocation techniques are presented. The authors, however, did not present any data on overall spectrum efficiency with QoS provisioning.

Beshay et al. [8] presented a framework for link-coupled TCP for 5G networks. It is a transport layer solution that can take advantage of 5G architecture trends. Without modifying TCP clients, LCTCP can be deployed. However, network latency and associated bottleneck is not addressed by LCTCP. A machine learning-based 5G architecture is proposed by Zhu et al. [9]. The authors proposed supervised learning-based QoS assurance protocols to mitigate network congestions and to improve channel throughput. Park et al. [10] proposed an improved version of congestion control algorithm ExLL The primary objective of the algorithm is to ensure congestion

reduction and considers latency as one of the QoS constraints. For downlink packet reception, ExLL utilizes cellular bandwidth inference, and need minimum round-trip time (RTT). One of the drawbacks of ExLL is that it is not able to improve channel throughput effectively. Improvement in channel throughput by statistical-QoS driven resource allocation is reported by Zhang et al. [11]. The proposed allocation policies are applicable in both asymptotic and non-asymptotic regimes of mmWave-based 5G network.

Xie et al. [12] proposed PBE-CC algorithm for congestion control that improves channel throughput significantly by precise measurement of rise and fall of wireless capacity demand. Another end-to-end congestion control approach is reported by Haile et al. [13]. The authors presented a congestion control algorithm that ensures low latency and high throughput in a highly variable network links. In addition, the authors discussed deployability of the algorithm. A service oriented transmission protocol in 5G network is discussed by Chen et al. [14]. Sharma et al. [15] presented a review of wireless backhaul networks and emerging trends of 5G networks. Ahmad et al. [16] discussed QoS constrained dynamic spectrum sharing using cognitive radio networks.

QoS Provisioning for M2M

The 4th Generation Long-Term Evolution (LTE) wireless network has dominant share of uplink data for Machine to Machine (M2M) communication. LTE uses shared radio channel based Single-Carrier Frequency Division Multiple Access (SC-FDMA) technique for uplink and Orthogonal FDMA (OFDMA) for downlink for M2M terminals. The focus of this section is on time domain scheduling using QoS class Identifier (QCI).

Enhanced LTE Network Architecture for 5G Networks

The existing LTE architecture of 4G network is not able to handle growing data rate of M2M communication. This chapter gives an overview of the improved architecture for 5G network that enable mass deployment of M2M and H2M data.

5G network is expected to cater three types of user classes, namely, massive Machine Type Communications (mMTC), enhanced Mobile Broad Band (eMBB), and Ultra-Reliable and Low Latency Communications (URLLC). Applications of mMTC include smart cities, smart homes, and office automation. Online gaming, ultra-high definition videos need eMBB support of 5G. Applications like autonomous vehicles, robots need reliable communication network and 5G URLLC is expected to deliver that.

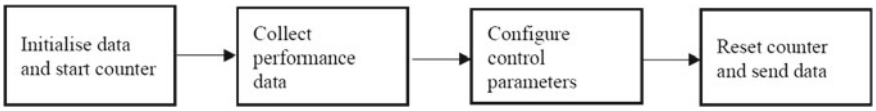

Fig. 3.1 Generalized congestion control algorithm flowchart

End-to-end cellular congestion control algorithms may be classified in three categories (i) predictive algorithm, (ii) reactive algorithm, and (iii) network assisted algorithm. Predictive algorithms may be fixed or dynamic interval. Reactive algorithms may be based on loss, delay, or rate triggered. Network assisted algorithms are based on in-band or out-band signaling. A generalized congestion control algorithm flowchart is illustrated in Fig. 3.1. For CCA, data is initialized, and counter starts. In the next stage, performance data is collected, and control parameters are configured. Finally, counter is reset, and data sent. QoS constrained congestion control algorithms are mentioned in Table 3.1.

Two primary constraints of CCAs are cross traffic and cellular bottleneck. Cross traffic is reduced by implementing per-user-queue. Cellular bottleneck is reduced by moving content nearer to the end users. Moving content closer to cellular access link reduces cost of bandwidth and reduces delay.

Resource Allocation to Ensure QoS

Link adaptation, admission control, removal and handover management, packet switching are some of the key features of resource management. For LTE and LTE-A, packet switching is done at MAC layer with the objective of increasing spectral efficiency with larger throughput. Scheduling decision is based on QoS requirement and Channel State Information (CSI).

Mobile devices should be able to connect to multiple access points, e.g., 5G, WLAN, LTE. Dual connectivity can exploit diversities of access points and deliver high data rate. Another important concept of resource allocation is network slicing by dividing physical network in multiple logical entities. Network slicing enables dynamic allocation of resources with different use cases.

Adaptive Channel Bandwidth Selection in LTE 4G/5G Networks

For fading channels 3G communication, generalized finite-state Markov channel (FSMC) is useful. However, for 5G network, Filter Bank Multi Carrier (FBMC) is utilized. Orthogonal Frequency Division Multiplexing (OFDM) is primarily based on FBMC.

Table 3.1 QoS constrained congestion control algorithms (CCAs)

Algorithm	Type	Features	Design criteria	Applications
C2TCP [17]	End-to-end solution	Delay sensitive, network state profiling not required	Sits on top of loss-based TCP	High throughput and low delay applications
X-TCP [18]	5G networks and 3GPP new radio	Provides large bandwidth, increased cell throughput	High variability	Congestion control with TCP CUBIC
PBE-CC [12]	Physical-layer bandwidth measurements	5G new radio innovations, increased wireless capacity	Cellular aware, congestion control protocols	Software-defined radios, PBE sender and receiver
DL-TCP [19]	Deep-learning-based TCP	TCP congestion window adjustment	Mobility information, signal strength	5G mmWave network
LCTCP [8]	Link-coupled TCP	Transport layer solution	Isolation of 5G access link, lightweight signaling	Link buffer, application server
DRWA [20]	Dynamic receiver window adjustment	Solves bufferbloat problem, reduces latency	TCP modification	Over the air (OTA) updates
CQIC [21]	Cross-layer congestion control	Physical layer information exchange	Adjustment in packet sending behavior	Cross-layer optimization
QTCP [22]	Reinforcement Q-learning framework	High throughput, low transmission latency	Hard-coded rules not required	Optimal congestion control
CDBE [23]	Client driven bandwidth estimation	Down-stream performance enhancement	Varying cellular BW	Variation of down-stream delay
ExLL [10]	Extremely low latency congestion control	Controls congestion window	Continuous bandwidth probing	Dynamic cellular channels

System Architecture

To overcome bufferbloat and subsequent jitter, a point-to-point Congestion Control Algorithms (CCA)-based solution is preferred. The system architecture is illustrated in Fig. 3.2. Distributed data and per-user-data are the two main reasons of bottleneck. CCA feedback ensures lesser congestion in network traffic. Data source consists of datalink layer. Theses datalink layers are stored in input buffer. In the next stage, physical layer splits datalink layers into bit streams. Based on QoS constraints and

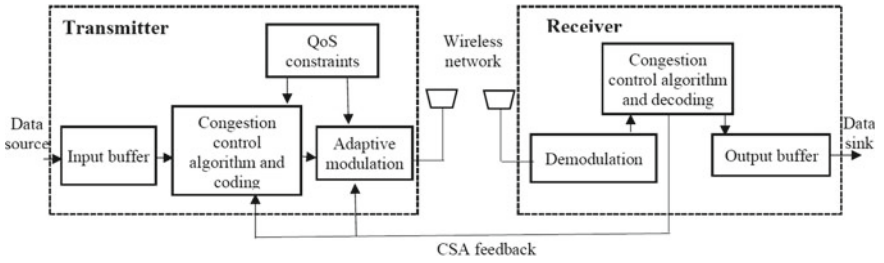

Fig. 3.2 System architecture

CCA feedback, adaptive modulation is done. In the receiving end reverse process is followed to reconstruct original message signal. Demodulation consists of an intermediate frequency stage that reduces modulated signal frequency.

TCP/IP transport layer supports CCAs. This signify that transport layer information and QoS constraints are the deciding factors for bottleneck link access. This in turn has preferential selection for top layer applications. Channel State Information (CSI) is updated based on the feedback received from receiver. Deployment of this model is possible even for applications where link layer information is not accessible. CSI of some applications may be dependent of APIs to interact with lower levels. API interaction allows continuous information flow for CCAs. However, ties of CSI and API may be loosely bound or tightly bound, depending on specific technologies for which applications are designed. Coordination among different functional blocks of transmitter and receiver is important for effective communication channel that is devoid of any bottleneck. CSA feedback introduce modifications in the input stack and subsequently all the down-stream layers of transmitter. Similarly, any modification at the receiving end would update network stack of users.

Delay Bounded QoS Provisioning

For delay bound Content Delivery Networks (CDN), performance tradeoff is done between available capacity and bit rate. Larger queue is required when capacity is increased. Underutilization of capacity reduces throughput. In some cases, some minimum delay may have to be guaranteed, while CSI may not be able to provide that delay. Hence CCAs prioritizes QoS delay constraint through underutilization of queue stack. CSI mostly receives information like delay, loss and packet count from transport layer. Based on CSI feedback, congestion control algorithms changes control decisions. However, delay due to adaptive provisioning of lower layers may cause bottleneck. In such cases, transport layer will not be able to differentiate

between delay caused by link rescheduling and queuing delay. In such cases cross-layer feedback, i.e., feedback received from network stack becomes dependable and gaining popularity in 5G networks. This approach also gives a reliable estimate about congestion.

Hybrid Scheduler with QoS Class Identifier

M2M communication model using LTE wireless network communicate with several servers. Hence data flow is more in the uplink. In general, M2M traffic is delay tolerant and designed for end-to-end QoS. Till 3G network, there was no specific QoS class for M2M communication. However, 4G and 5G network introduced QoS Class Identifier (QCI), which is used for admission control, queuing decision and congestion control algorithm. Shared radio channels between LTE access network and user equipment communicate by OFDMA and SC-FDMA. Admission control, mobility monitoring, and resource allocation are done via enhanced Node B (eNodeB). Both time domain and frequency domain subcarrier allocation are done for LTE subscribers. Optimal data packet set for transmission is time domain scheduling, whereas mapping data packets to resource blocks is frequency domain scheduling.

Conclusion

A delay bound QoS provisioning technique and performance tradeoff for content delivery networks is discussed in the chapter. Large input buffer causes bufferbloat and subsequent jittering the network. Congestion control algorithm along with Channel State Information (CSI) feedback plays important role to mitigate bottleneck and guarantee QoS. Further research on efficient congestion control algorithm, delay bounded QoS provisioning, effective protocols for steering antenna beamforming, and improved error correction coding will make 5G network more useful and user friendly. Advancement in technologies like, Cloud Radio Access Networks (C-RAN), Heterogeneous C-RAN, Software-Defined Networking (SDN), Massive MIMO, Self-Organizing Network (SON), and opportunistic network will lead to a new generation of communication network. Energy efficiency should ideally be one of the important constraints toward a sustainable 5G network. It is almost impossible for a single technology to meet all criteria and converge—hence, a sequential roll-out of 5G will be able to meet-up with the user expectations and will be sustainable. Cooperation and coordination among industries, academia and regulatory bodies would play an important role in balancing QoS, energy efficiency and effective spectrum utilization.

References

1. Sharma, T., Chehri, A., Fortier, P.: Review of optical and wireless backhaul networks and emerging trends of next generation 5G and 6G technologies. Trans. Emerg. Telecommun. Technol. **32**(3), 1–16 (2021). https://doi.org/10.1002/ett.4155
2. Abdalla, I., Venkatesan, S.: A QoE preserving M2M-aware hybrid scheduler for LTE uplink. In: International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT), vol. 7, pp. 127–132 (2013). https://doi.org/10.1109/MoWNet.2013.6613808
3. Akhtar, T., Tselios, C., Politis, I.: Radio resource management: approaches and implementations from 4G to 5G and beyond. **27**(1) (2021)
4. Tang, J., Zhang, X.: Quality-of-service driven power and rate adaptation for multichannel communications over wireless links. IEEE Trans. Wirel. Commun. **6**(12), 4349–4360 (2007). https://doi.org/10.1109/TWC.2007.06031
5. Inaba, T., Sakamoto, S., Oda, T., Barolli, L., Takizawa, M.: A new FACS for cellular wireless networks considering QoS: a comparison study of FuzzyC with MATLAB. In: Proceedings of the 18th International Conference on Network-Based Information Systems. NBiS 2015, pp. 338–344 (2015). https://doi.org/10.1109/NBiS.2015.52
6. Beshley, M., Kryvinska, N., Seliuchenko, M., Beshley, H., Shakshuki, E.M., Yasar, A.U.H.: End-to-end QoS 'Smart Queue' management algorithms and traffic prioritization mechanisms for narrow-band internet of things services in 4G/5G networks. Sensors (Switzerland) **20**(8) (2020). https://doi.org/10.3390/s20082324
7. Haile, H., Grinnemo, K.J., Ferlin, S., Hurtig, P., Brunstrom, A.: End-to-end congestion control approaches for high throughput and low delay in 4G/5G cellular networks. Comput. Networks **186**, 107692 (2021). https://doi.org/10.1016/j.comnet.2020.107692
8. Beshay, J.D., Nasrabadi, A.T., Prakash, R., Francini, A.: Link-coupled TCP for 5G networks. In: IEEE/ACM 25th International Symposium on Quality of Service (IWQoS) (2017). https://doi.org/10.1109/IWQoS.2017.7969170
9. Zhu, G., Zan, J., Yang, Y., Qi, X.: A supervised learning based QoS assurance architecture for 5G networks. IEEE Access **7**, 43598–43606 (2019). https://doi.org/10.1109/ACCESS.2019.2907142
10. Park, S., Lee, J., Kim, J., Lee, J., Lee, K.: ExLL: an extremely low-latency congestion control for mobile cellular networks. In: CoNEXT 2020: Proceedings of the 16th International Conference on emerging Networking EXperiments and Technologies, pp. 307–319 (2020)
11. Zhang, X., Wang, J., Poor, H.V.: Heterogeneous statistical-QoS driven resource allocation over mmWave massive-MIMO based 5G mobile wireless networks in the non-asymptotic regime. IEEE J. Sel. Areas Commun. **37**(12), 2727–2743 (2019). https://doi.org/10.1109/JSAC.2019.2947941
12. Xie, Y., Yi, F., Jamieson, K.: PBE-CC: congestion control via endpoint-centric, physical-layer bandwidth measurements. In: SIGCOMM 2020: Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 451–464 (2020). https://doi.org/10.1145/3387514.3405880
13. Haile, H., Grinnemo, K.J., Ferlin, S., Hurtig, P., Brunstrom, A.: End-to-end congestion control approaches for high throughput and low delay in 4G/5G cellular networks. Comput. Networks **186**, 107692 (2021). https://doi.org/10.1016/j.comnet.2020.107692
14. Chen, J., et al.: SDATP: an SDN-based traffic-adaptive and service-oriented transmission protocol. IEEE Trans. Cogn. Commun. Netw. **6**(2), 756–770 (2020). https://doi.org/10.1109/TCCN.2019.2963149
15. Sharma, T., Chehri, A., Fortier, P.: Review of optical and wireless backhaul networks and emerging trends of next generation 5G and 6G technologies. Trans. Emerg. Telecommun. Technol. **32**(3), 1–16 (2021). https://doi.org/10.1002/ett.4155
16. Ahmad, W.S.H.M.W., et al.: 5G technology: towards dynamic spectrum sharing using cognitive radio networks. IEEE Access **8**, 14460–14488 (2020). https://doi.org/10.1109/ACCESS.2020.2966271

17. Abbasloo, S., Li, T., Xu, Y., Chao, H.J.: Cellular controlled delay TCP (C2TCP). In: IFIP Networking Conference (IFIP Networking) and Workshops, pp 118–126 (2018). https://doi.org/10.23919/IFIPNetworking.2018.8696844
18. Azzino, T., Drago, M., Polese, M., Zanella, A., Zorzi, M.: X-TCP: a cross layer approach for TCP uplink flows in mmwave networks. In: 16th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), pp. 1–6 (2017). https://doi.org/10.1109/MedHocNet.2017.8001650
19. Na, W., Bae, B., Cho, S., Kim, N.: DL-TCP: deep learning-based transmission control protocol for disaster 5G mmWave networks. IEEE Access **7**, 145134–145144 (2019). https://doi.org/10.1109/ACCESS.2019.2945582
20. Jiang, H., Wang, Y., Lee, K., Rhee, I.: DRWA: a receiver-centric solution to bufferbloat in cellular networks. IEEE Trans. Mob. Comput. **15**(11), 2719–2734 (2016). https://doi.org/10.1109/TMC.2015.2510641
21. Lu, F., Du, H., Jain, A., Voelker, G.M., Snoeren, A.C., Terzis, A.: CQIC: revisiting cross-layer congestion control for cellular networks. In: HotMobile 2015: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, pp. 45–50 (2015). https://doi.org/10.1145/2699343.2699345
22. Li, W., Zhou, F., Chowdhury, K.R., Meleis, W.M.: QTCP: adaptive congestion control with reinforcement learning. IEEE Trans. Netw. Sci. Eng. **4697**, 1–13 (2018). https://doi.org/10.1109/TNSE.2018.2835758
23. Zhong, Z., Hamchaoui, I., Ferrieux, A., Khatoun, R., Serhrouchni, A.: CDBE: a cooperative way to improve end-to-end congestion control in mobile network. In: International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), vol. 2018, pp. 216–223 (2018). https://doi.org/10.1109/WiMOB.2018.8589175