2023 58th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST) | 979-8-3503-1073-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEST58410.2023.10187341

IoT Networks QoS Guarantee

Georgi Balabanov[1] and Stanyo Kolev[2]

Abstract –**Buffer space management is critical to ensuring full utilization of network resources in present IoT networks. It allows the multiplexing of services with different quality of service requirements. Current IoT technologies supports a wide range of services including medical treatment, preventive equipment maintenance, remote operation of machinery, environmental monitoring, video surveillance and real-time alerts, connected transport as well as multiplexed services consisting of combinations of these. Traffic prioritization plays an important role in quality of service management in all mobile networks. From a network perspective, the aim is to ensure a minimum waiting time for packets of higher priority classes in the buffers of these networks. In order to use the resources correctly, it is necessary for each type of traffic or service to share the traffic capacity depending on the quality of service requirements they have.**

Keywords – **IoT, QoS, 5G, Buffer management, Traffic priority.**

I. Introduction

The fundamental concept behind the Internet of Things (IoT) revolves around the widespread existence of diverse objects that possess the ability to interact and collaborate with one another in order to achieve a shared goal [1]. The IoT poses numerous obstacles that need to be addressed, including security and privacy concerns, participatory sensing, management of large amounts of data, and architectural considerations, in addition to the well-known challenges associated with Wireless Sensor Networks (WSNs), such as energy efficiency, protocols, and ensuring Quality of Service (QoS) [2].

The utilization of IoT technology presents us with possibilities to enhance operational productivity and elevate speed/bandwidth. This progress is anticipated to result in advancements in the quality of services offered, in addition to the creation of novel services. The emergence of 5G technology is causing a shift in the IoT landscape, leading to groundbreaking IoT applications for numerous industries and diverse use cases. The dynamic nature of 5G technology offers us the chance to incorporate a broad and varied array of end devices, enabling the introduction of innovative applications and breakthroughs in various sectors [3].

Remote monitoring: IoT technology enables the use of drones or remote cameras to monitor various civil infrastructure, such as power grids, bridges, tunnels, highways, and railways. These devices can transmit images or video, providing valuable information for automated problem detection. By utilizing near-real-time IoT data from stationary and mobile sensors, this solution can help identify issues promptly.

Process inspection: This application is suitable for industries that rely on IoT sensors to inspect their operational assets and production lines. The crucial factor to consider is the network's low latency. As products move along the production line, cameras can capture images, and the resulting data can aid in process and product quality control. Artificial Intelligence powered software can analyze these images to detect any quality issues or defects.

One more application of IoT based on 5G networks is a saturating factory with a huge number of sensors without requiring the construction and maintenance of expensive IT infrastructure.

Building and Facilities Management: The advent of 5G technology promises to revolutionize building and facility management by facilitating the easy connectivity of sensors. This advancement opens up new possibilities for creating innovative solutions that promote energy efficiency, efficient occupancy management, and enhanced visitor experience. Additionally, this technology will be especially beneficial for larger facilities with numerous assets that require monitoring and maintenance, including stadiums, airports, malls, and schools. By leveraging the power of 5G, these facilities can track and analyze data in real-time to streamline operations, improve security, and enhance the overall experience for their visitors and occupants. The efficient connectivity of sensors can help to optimize building systems and save energy, while enhancing comfort and convenience. Overall, the integration of 5G technology in building and facility management promises to bring a multitude of benefits that will help to create smarter and more sustainable environments.

Smart Vehicles: Contemporary automobiles gather significant amounts of data to ensure safety. The faster response time of 5G, together with increased bandwidth and dependable connectivity, will facilitate various smart vehicle applications.

Smart Agriculture: A new technology that allows producers to adequately manage cultivated areas depending on spatially differentiated information. It is an innovative, technologically and information-based, intelligent approach to identify, analyze and manage variables to obtain cost-effective production with optimal output and resource conservation. The point of the approach is to make the right management decisions in agriculture based on the variable characteristics of the field and obtain maximum yields. Drones, employing visual analytics, will independently examine crops in the field, operating in close proximity to the area of interest. Their primary purpose will be to evaluate various factors such as growth rates and losses caused by pests.

[1]Georgi Balabanov is with Technical University of Sofia, Faculty of Telecommunication, E-mail: grb@tu-sofia.bg.

[2]Stanyo Kolev is with Technical University of Sofia, Faculty of Telecommunication, E-mail: skolev@tu-sofia.bg.

In the realm of shipping and logistics, it becomes possible to maintain constant surveillance and supervision over numerous shipments and shipping containers, ensuring enhanced precision and predictability in the delivery process.

Based on the brief analysis made above, it can be seen that IoT offers a large range of services that have different requirements in terms of maintaining quality of service. In this regard, IoT networks must provide mechanisms to guarantee quality of service, meeting the requirements of individual service types.

Ensuring the optimal performance of IoT services is a major challenge as it relies on the performance of multiple components within IoT systems. To meet the demands of users, these systems must continuously evolve and enhance their services. This requires constant monitoring and improvement of each component to ensure that the system as a whole operates efficiently and effectively. In short, achieving and maintaining high-performance standards in IoT services is a complex and ongoing process that requires careful attention and constant effort.

II. QoS Guarantee in IoT

Many models and methods have been developed to QoS guarantee in IoT networks. We can define several aspects related to guaranteeing QoS, such as availability, reliability, mobility, performance, scalability and interoperability, most of these issues have been identified by [3]-[5].

TABLE I
QoS parameters for Smart transport and mobility

IoT Service	Traffic speed	QoS	Power source
Vehicle automation	regular, 1 message every day per vehicle, irregular and frequent when communicating between vehicles	low, allowable delay close to real time	powered by the vehicle battery
Vehicle localization and monitoring	1 message every 30 sec per vehicle	regular medium, allowable delay 10 sec	powered by the vehicle battery
Monitoring the quality of sending conditions	regular, 1 message every 15 min per vehicle	medium, allowable delay 15s	powered by battery
Dynamic control of traffic lights	regular, 1 message every 1 min per vehicle	medium, permissible delay 5 sec	powered by the network
Road condition control	irregular	average, allowable delay 30s	powered by battery

With the advent of countless machine-type devices interconnected by the Internet of Things, communication networks are facing unprecedented challenges to efficiently support the dynamic growth of traffic. To optimize current and future IoT communication systems, it is essential to first understand and model the specific traffic patterns generated by IoT data.

In congestion mode, where the incoming flow of services is greater than the bandwidth of the radio interface, the buffers in the mobile device and in the serving nodes play an essential role in guaranteeing the quality of service. However, if the overload also covers the buffer spaces, the only solution is to "throw away" some of the packets stored in the buffers. Many methods have been developed, such as Drop-tail, Random Early Detection (RED), Packet Discard Prevention Counter (PDPC), which determine which packets from the buffer should be discarded in order to prevent the buffer from overflowing and to maintain the quality of service parameters within the set limits [6].

The domain of smart transport and mobility presents specific challenges related to traffic characteristics (as shown in Table 1). The increasing number of sensors in vehicles enables a network capable of handling a large volume of connected mobile devices with extensive coverage. Depending on the service, traffic speed varies, with some requiring frequent message transmission, such as predicting public transport arrivals, while others may only need data collection for analytical purposes, allowing for less frequent messaging, even up to once every 24 hours. Quality of Service (QoS) requirements are generally high, with the need for near-real-time communication in vehicle-to-vehicle interactions, although most services do not demand instantaneous delivery and accept delays of a few seconds. Smart transport and mobility devices primarily rely on vehicle batteries or grid power for roadside devices/sensors, prioritizing precision and accuracy over energy savings. In smart agriculture, traffic characteristics align with the requirements of sending small amounts of data about environmental conditions over extended periods. The network size in this context can range from small to medium, depending on the area to be covered by the application. Traffic speed is typically regular, although changes in environmental conditions tend to occur slowly. However, in animal monitoring scenarios, the traffic rate can vary irregularly based on the specific implementation. QoS requirements are low, with delays of up to 1 minute considered acceptable. Real-time transmission is crucial in certain cases, such as when an animal crosses the boundaries of a pasture. Power source demands are significant in smart agriculture as devices heavily rely on battery power. The traffic characteristics in wearable devices and healthcare applications emphasize the efficient implementation of tracking using a small number of devices, resulting in a small network size. However, patient tracking in a hospital setting may require a medium-sized network with hundreds of nodes to cover an entire building. Traffic rates for healthcare applications are typically regular but with a high frequency, as health statuses can change rapidly. QoS requirements are undoubtedly high, as message loss or significant delays are unacceptable, particularly during emergencies. Wearable devices are powered by batteries, posing a challenge for their longevity due to high traffic. Nevertheless, they can be easily recharged. In addition to dynamic nodes, static nodes with a mains power source can also be employed, particularly in services like patient monitoring [7].

III. IoT Dynamic Queue Management

The 3GPP specifications for LTE and 5G do not include algorithms for the management of buffer spaces in mobile terminals or IoT devices, thus leaving mobile operators free to implement solutions leading to guaranteeing quality of service and optimizing network usability resources.

The dwell time of packets in the IoT devices during upstream transmission is a component of end-to-end QoS. This problem is more pronounced when competing real-time and non-real-time services. To guarantee this QoS, it is necessary to introduce buffer space management mechanisms and methods in the mobile terminal. In addition, we have to consider the fact that the packet delays in the buffer spaces are much larger compared to the transmission times in the radio access network especially in heavily loaded cells. Then the dwell time in the buffer spaces is critical to guarantee the desired quality of service. The introduction of adequate methods and mechanisms for managing the buffer spaces in the mobile terminal during uplink transmission will lead to an improvement in the utilization of radio resources, which in turn will lead to an increase in cell capacity [8].

Real-time services have certain limits on delay sensitivity and loss tolerance, while non-real-time services have limits on loss sensitivity and delay tolerance. By analyzing these requirements alongside buffer space management methods and the characteristics of uplink transmission channels in mobile networks, we can propose a new teletraffic system that can handle both real-time and non-real-time teletraffic flows equally. This system is called IoT Dynamic queue management (IoTDQM), which is a hybrid system designed to prioritize traffic flows and provide joint control of quality of service for competing real-time and non-real-time services transmitted on a shared channel on Uplink. The IoTDQM system uses a common shared buffer, and Fig. 1 displays a block diagram of the system.

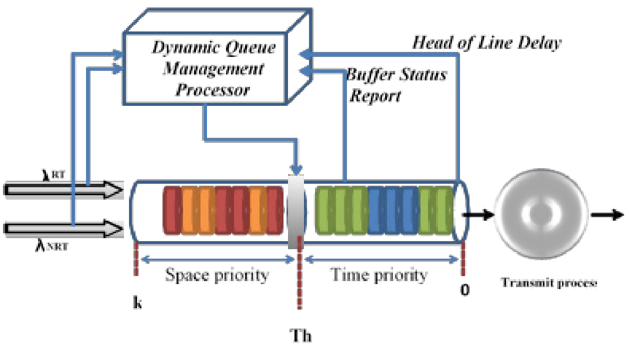

Fig. 1. IoT Dynamic Queue Management.

IoT Dynamic Queue Management (IoTDQM) is a queuing system with a capacity of k packets that allows joint quality of service (QoS) control for real-time and non-real-time flows in user terminal uplinks over a shared wireless channel. The IoTDQM system uses a single queue with a threshold Th to control the admission of real-time packets. This flow is prioritized over non real-time flow and follows a FIFO method, as it is sensitive to delay. However, because real-time flow is somewhat tolerant of losses, the threshold is used to limit the total number of ela-time packets in the queue and prioritize non-rela-time packets for buffer space. This is referred to as class 2. As a result, real-time delay is minimized with time priority, while non-real0time loss is minimized with the space priority mechanism. Real-time packets are placed at the front of the buffer queue ahead of non-real-time packets until the threshold limit Th is reached. Fig. 1 shows the IoTDQM queue system.

The IoTDQM queuing system has an essential characteristic where real-time packets are given a higher priority up to a certain threshold limit (Th) when the buffer is full. This is accomplished through a Last in First Drop policy, which drops the non-real-time packet at the tail of the queue to allow the arriving real-time packet into the full buffer, as long as the total number of real-time packets is less than the threshold Th. It should be noted that this displacement policy is expected to cause a minor increase in non-rela-time packet loss if the threshold Th is small compared to the total IoTDQM size k. The IoTDQM queuing system shares the entire allocated buffer space between real and non-real time packets, irrespective of the threshold Th, ensuring high buffer utilization and minimizing overall packet loss probability. Additionally, reducing non-real-time packet loss in the user buffer enhances network resource utilization and improves the performance of higher layer protocols, as lost non-real-time packets usually require retransmission via error control mechanisms, leading to a decline in end-to-end throughput and a waste of transmission resources. The Dynamic Queue Management Processor is responsible for maintaining a dynamic threshold value, which ensures continuous QoS provision for both real-time and non-real-time traffic classes. To achieve this, it's necessary to jointly optimize the QoS parameters for a given set of system and traffic parameters. This is done by calculating the optimum threshold value using a cost function called the weighted grade of service cost function G. The cost function determines the best operating threshold position based on the given traffic and system parameters. In order to optimize IoT services in semi real-time, the Dynamic Queue Management processor needs to be invoked to update the threshold value. There are several possible methods for updating the threshold value, including:

- The Time Period Update method is a technique for updating the threshold value of the Dynamic Queue Management processor at regular intervals. This method involves updating the input traffic and service process parameters periodically to allow the analytic model to find the value of Th that minimizes the cost function G. Once the optimal value of Th is calculated, it will be used as the threshold value for the time space priority buffer management algorithm until the next update. By implementing Time Period Update, the IoTDQM can adapt to changing network conditions and traffic patterns, ensuring continuous QoS provisioning for both real-time and non-real-time traffic. In addition, TPU can be an effective method for optimizing the performance of IoT services.

• The QoS requirements update method involves initially setting a value for the threshold Th based on the user's traffic profile. If the system detects that a QoS performance measure such as blocking or delay exceeds a certain allowable value, it will then update the threshold Th accordingly. This ensures that the system can continuously adjust to meet the user's QoS requirements.

• Arrival pattern update: The system can be prompted to calculate a new optimum value for the buffer optimization when there are changes in the arrival pattern of traffic. For instance, if there are variations in the arrival rates of real-time and non-real-time traffic, the system can recalculate the optimum value of Th.

A simulation model of IoTDQM has been created to ensure quality of service in IoT networks. This model includes a functional representation of the time-space priority buffer and dynamic queue management processor. The model was designed using both Python and C programming languages.

The fundamental assumptions for the system model are that it consists of two services which maintain a session that comprises two classes of flows, namely a real-time flow and a non-real-time (NRT) flow. Independent packet arrivals to the buffer are allowed. To simulate RT flows, three different RT services models are used: Smart Health (Quality of Service Class Identifiers, QCI=1), Smart Vehicle (QCI=2), and Remote Monitoring (QCI=3). The packets are queued according to the IoTDQM mechanism described earlier, and the queue has a capacity of 1200 packets (k=1200) for the connection.

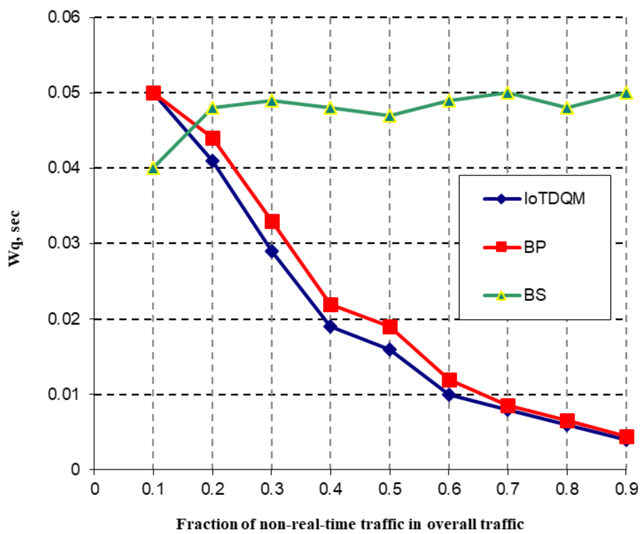

Fig. 2. Mean waiting time (Wq) as a function of arrivals component.

Fig. 2 shows that both the IoTDQM and the Buffer Partitioning (BP) schemes achieve low mean real-time flow delay by prioritizing real-time services. The IoTDQM scheme, which also uses a threshold limit Th, outperforms the Buffer Sharing (BS) scheme because it employs precedence queuing and time priority for real-time flows. These features minimize real-time queuing delay and facilitate the transfer of rela-time packets from the user terminal to eNode-B. It is important to note that the BS scheme does not have these features.

IV. Conclusion

The purpose of the research presented in this paper is to enhance the quality of service assurance techniques in IoT networks by implementing the IoT Dynamic Queue Management queuing system. This system prioritizes delay-sensitive and loss-tolerant mixed traffic with time transmission priority, while allocating space priority to loss-sensitive and delay-tolerant mixed traffic, and actively managing the time queue length. Selecting the optimal threshold value, Th, is a challenging problem in dynamic environments where traffic and channel conditions are constantly changing. The outcomes presented in this paper provide some general conclusions regarding quality of service parameters in IoT networks.

References

[1] D. Giusto, *The Internet of Things: 20th Tyrrhenian Workshop on Digital Communications*, Springer, New York, 2010.

[2] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, „Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions“, Future Gener. Comput. Syst., vol. 29, no. 7, pp. 1645-1660, 2013.

[3] M. Attaran, „The impact of 5G on the Evolution of Intelligent Automation and Industry Digitization“, J Ambient Intell Human Comput., 2021, doi: 10.1007/s12652-020-02521-x.

[4] R. Khan, S. U. Khan, R. Zaheer, and S. Khan, „Future Internet: the Internet of Things architecture, Possible Applications and Key Challenges“, pp. 257-260, 2012.

[5] L. Atzori, A. Iera, and G. Morabito, „The Internet of Things: A Survey“, Comput. Netw., vol. 54, no. 15, pp. 2787-2805, 2010.

[6] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, „Internet of Things: Vision, Applications and Research Challenges“, Ad Hoc Netw., vol. 10, no. 7, pp. 1497-1516, 2012.

[7] J. Mocnej, A. Pekar, Winston K.G. Seah, I. Zolotova, „Network Traffic Characteristics of the IoT Application Use Cases“, School of Engineering and Computer Science, Victoria University of Wellington, 2018.

[8] S. Yerima and K. Al-Begain, „Dynamic Buffer Management for Multimedia QoS in Beyond 3G Wireless Networks“, IAENG International Journal of Computer Science, 2009.