IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 6, 2020, accepted March 9, 2020, date of publication March 20, 2020, date of current version March 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982357

Media Access Control in Large-Scale Internet of Things: A Review

PROSPER Z. SOTENGA[1], (Member, IEEE), KARIM DJOUANI[2], (Member, IEEE), AND ANISH M. KURIEN[1], (Member, IEEE)
[1]Department of Electrical Engineering, Tshwane University of Technology, Pretoria 0183, South Africa
[2]LISSI Lab, Université Paris-Est Créteil, 94000 Créteil, France

Corresponding author: Prosper Z. Sotenga (zanuzy05@hotmail.com)

This work was supported in part by the National Research Foundation (NRF) of South Africa under Grant 90604. The opinions, findings and conclusions or recommendations expressed in this publication are those of the authors alone, and the NRF accepts no liability whatsoever in this regard.

ABSTRACT The Media Access Control (MAC) serves as an imperative part of wireless communication networks which enables efficient provisioning of communication resources for device interconnectivity and ensures Quality of Service (QoS). The emergence of Large-Scale Internet of Things (LS-IoT) networks is characterised by a multi-domain distributed wireless communication network that provides end-to-end connectivity for a multitude of active heterogeneous Machine-to-Machine (M2M) devices. The nature of LS-IoT networks requires robust and scalable MAC protocols to manage concurrent, dynamic and massive media access and resource allocation. Several reviews have been conducted on MAC protocols with a handful of them focused on LS-IoT networks or massive M2M networks. In this paper, the characterisation of LS-IoT networks and the MAC component are extensively discussed delineating the impact of LS-IoT on the MAC. Emerging research issues on the MAC for LS-IoT networks including high collision probability, high control overheads, spectrum constraints, timing constraints, hardware constraints, energy consumption and hidden terminals are discussed. Some recently proposed solutions in literature for enhancing the MAC in LS-IoT networks are discussed under LS-IoT specific MAC protocol enhancement classifications. The contributions and drawbacks of the recent solutions in literature are discussed and summarised. The Future direction that is oriented towards the use of virtualisation within the framework of Network Function Virtualisation (NFV) and Software-Defined Networking (SDN) approaches is proposed. This is aimed at providing a dynamic distributed multidimensional MAC resource scaling approach for LS-IoT access network devices to ensure a robust and effective MAC for massive M2M devices in LS-IoT networks.

INDEX TERMS Large-scale Internet of Things (LS-IoT), machine-to-machine (M2M) communication, media access control (MAC), wireless networks.

I. INTRODUCTION

The classical Internet infrastructure was initially designed as a global system for interconnecting computers using the standard Internet protocol suite [1]. On the other hand, the Internet of Things (IoT) is characterised by the interconnection of ubiquitous and heterogeneous devices for machine-type data exchange over the classical Internet infrastructure [2]. Conceptualised by Kevin Ashton in 1999, the IoT has received significant interest by stakeholders. This has resulted in the implementation of the IoT in a commercial and industrial scale consequently leading to the explosion of devices connected to the Internet. The stimulation of such explosive growth can be attributed to the tremendous advancement in sensing capabilities, control systems, cloud computing, processing and storage capabilities for autonomous, seamless and intelligent applications. Additionally, the growth in the number of devices connected to the Internet has been triggered by the continuous reduction in the cost of sensors, microcontrollers, microprocessors, actuators and radio chips used for the fabrication of Machine to Machine (M2M) devices [3]. Large-Scale Internet of Things (LS-IoT) networks are typically associated with the dense deployment of heterogeneous M2M devices interconnected using the Internet infrastructure over multiple network domains. Concerns about the enormous number of devices that will be connected

The associate editor coordinating the review of this manuscript and approving it for publication was Noor Zaman.

to the Internet in the near future have been raised with a forecast of annual M2M device growth reaching over 20% globally [4].

The allocation of shared resources facilitated by the Media Access Control (MAC) protocol is a critical element of the LS-IoT network that ensures effective connectivity. With the increasing growth of M2M devices, LS-IoT networks could face serious inevitable congestion which could compromise the Quality of Service (QoS) requirements in the LS-IoT network. The primary constraint for accommodating large-scale M2M devices in LS-IoT networks is the lack of effective scalability of the MAC protocol. The scheduling and allocation of resources for devices in the network hinders end-to-end connectivity. Industrial stakeholders and standardisation bodies indicate that media access for scenarios with a very large-scale of devices leads to a bottleneck [3]. It is therefore critical that the MAC protocol designs address the effective control and management of massive M2M devices that share communication resources. Therefore, it is imperative to study the characteristics of LS-IoT, the MAC related issues, and some existing solutions.

Comprehensive reviews on the MAC focused on LS-IoT networks or massive M2M deployment are not common. Studying the characteristics of LS-IoT is imperative for understanding the MAC protocol design challenges involved in achieving massive M2M communication in LS-IoT networks given the dynamic, heterogeneous and random transmissions of data frames in LS-IoT networks. It is also important to study some proposed strategies that seek to provide effective scalable MAC protocol design solutions for LS-IoT networks with stable or improved QoS parameters. This helps in understanding the status-quo of the MAC in LS-IoT networks and provide future directions.

The key objectives of this paper are as follows: The first objective is to present a comprehensive description of the characteristics of a LS-IoT network based on literature, and to discuss the impact of some of the LS-IoT characteristics on the network and the MAC; the second objective is to highlight and discuss the key technical issues faced with the design of an effective scalable MAC protocol for LS-IoT networks while contributing to the understanding of a MAC protocol and the various classifications. The aim is to establish the relationship between the MAC issues, the LS-IoT network characteristics and the impact on QoS; the final objective is to provide a comprehensive review of the most current approaches aimed at mitigating some of the MAC issues related to providing effective scalable MAC protocols for LS-IoT networks and to delineate their strengths and weaknesses presenting some possible research gaps through comparative analysis. Based on the identified research gaps, the aim is to finally recommend new research areas that may mitigate some of the research gaps of the current strategies presented.

The unique contributions provided in this paper are: (1) To the best of the authors' knowledge, this paper is the first to provide a synthesized description and characterisation of LS-IoT networks based on the functional and physical elements of the network by looking at the different domains, identifying the core bottleneck domain and using literature to associate the bottleneck phenomena with the performance of the MAC protocol; (2) this paper comprehensively discusses the issues related to the design and implementation of MAC protocols that are specifically intended at improving network scalability in LS-IoT networks or massive M2M communication. To the best of the authors' knowledge, the discussions on the MAC issues presented in this paper provide the best real-life and detailed challenges that will easily enable the reader to formulate, identify or investigate some research patterns or new design approaches to address some of the MAC design issues; (3) this study discusses, compares and contrasts the most current approaches proposed in literature specifically oriented at improving massive M2M communication in LS-IoT networks taking into account the dynamics of LS-IoT. To the best of the authors' knowledge, this is the best broadly examined comparative analysis of the most current approaches aimed specifically at improving massive M2M communications and more broadly LS-IoT under LS-IoT specific MAC classifications.

The remainder of the paper is structured as follows: Section II presents an overview of the IoT network architecture and domains and the characteristics of the LS-IoT; Section III presents an overview of the general MAC layer and the related classifications of the MAC protocols; Section IV highlights the impact of the LS-IoT characteristics on the network and the MAC protocol and presents the MAC design issues; Section V presents currently proposed solutions in literature to mitigate the issues faced with a scalable MAC protocol design for LS-IoT; Section VI presents proposed future directions for the improvement of the MAC protocol for LS-IoT networks. Section VII concludes the paper by consolidating all the key aspects studied in this paper.

II. THE LARGE-SCALE IoT NETWORK

A. OVERVIEW OF THE IoT NETWORK ARCHITECTURE AND DOMAINS

Various architectures and descriptions of the IoT network have been presented in literature. A typical high-level architecture and functional description of the IoT network is specified by the European Telecommunications Standards Institute (ETSI) [5], [6]. A classical IoT architecture is illustrated in Fig. 1 which highlights the following domains: the M2M local network or capillary network [7], the access network or network edge, the core network or backbone, the cloud data centre, and the application domain.

The M2M local network domain consists of the M2M devices which provide autonomous sensing or actuation mechanisms and autonomously generate machine-type data for transmission to other M2M devices [8] or to end-user applications. The concept of machine-type data exchange in the IoT network is commonly articulated as M2M communication. The M2M local network is a sub-network within

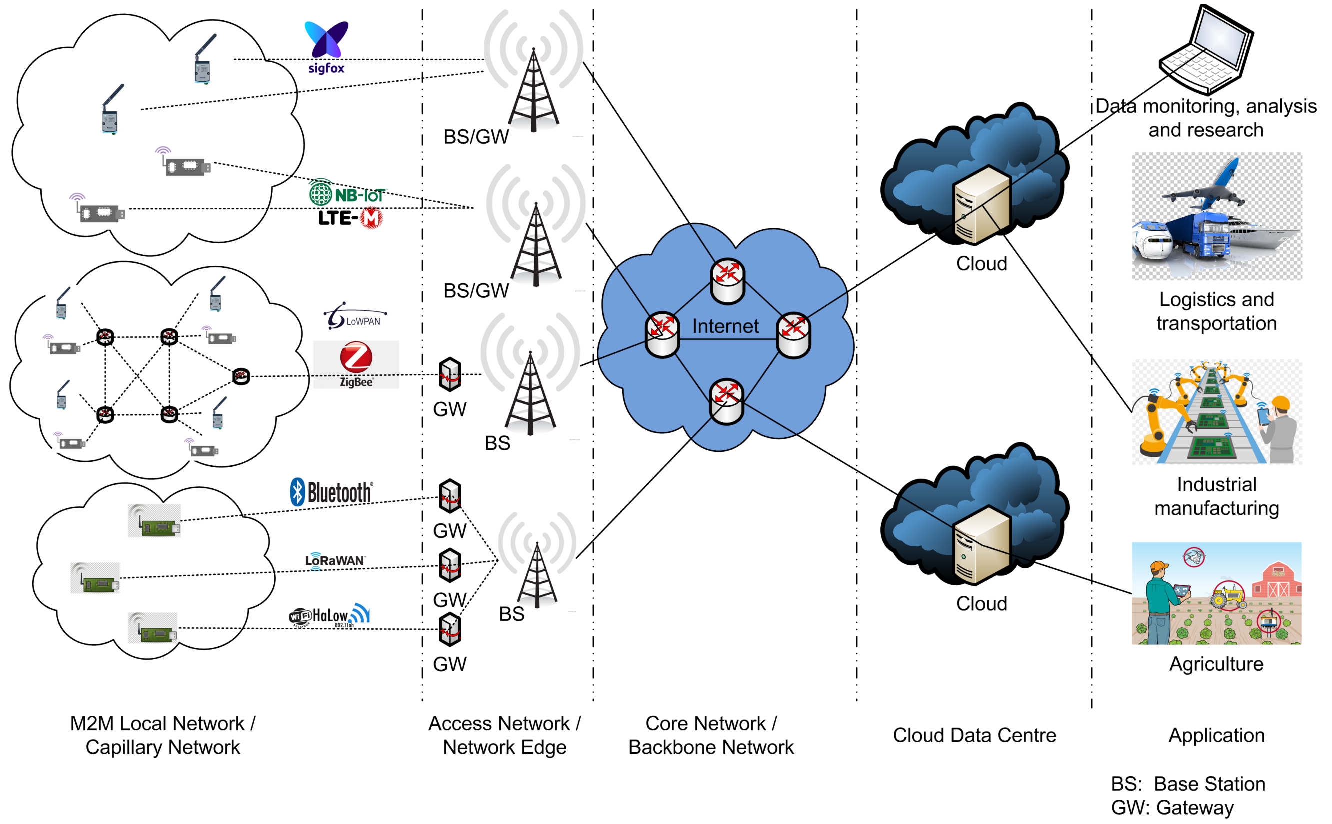


FIGURE 1. An illustration of the IoT network with multiple domains, heterogeneous devices, technologies and applications.

the IoT network which can be configured based on two approaches. The M2M devices can be connected directly to the Gateway (GW) or Base Station (BS) without an auxiliary GW. One example of such a configuration is the Cellular M2M network where the devices are directly connected to a cellular network BS. The M2M devices can also be interconnected in a star or mesh topology to form a capillary network which is coordinated by a GW device acting as a proxy for a remote BS. This configuration is commonly used for integrating Wireless Sensor Networks (WSN) in IoT networks.

The access network is made up of GW and BS devices collectively called the access network devices. They provide access to a wider third-party network which is the Internet in the context of IoT. In this paper, the GW and BS devices are considered as one device except when explicitly regarded as separate devices in coexistence. When the GW and BS coexist, the GW device acts merely as a proxy of the BS to manage access for the capillary network and to translate one communication protocol to a compatible protocol understood by the BS.

The core network domain on the other hand represents the classical Internet infrastructure which routes packets between different networks based on the Internet Protocol (IP). The core network enables an M2M local network to be transparently connected to other M2M local networks. It also enables connection from an M2M local network to networks other than an M2M local network including the cloud data centre. The core network is also known as the backbone.

The cloud data centre is an infrastructure consisting of a network of hardware equipped with computing, storage and networking capabilities. The hardware that constitute the cloud data centre are distributed and provisioned remotely. The cloud data centre and the application domain operate closely together. The application domain enables logical human interaction with the M2M local network through specialised services. An Application Programming Interface (API) is typically used to provide the interaction between the cloud data centre and the application services.

The domains discussed are important components that can be used to adequately qualify an IoT network as large-scale. The different domains help identify some fundamental problems and challenges in LS-IoT as well as identify and examine the bottleneck in the LS-IoT network.

B. LS-IoT NETWORK CHARACTERISATION

Contrary to application-specific networks or networks with minimal diverse applications such as cellular networks and Wi-Fi (IEEE 802.11) networks, the IoT network builds on many diverse applications, functions and network components. To provide adequate context to the meaning of LS-IoT networks in this paper, the description of the LS-IoT network characteristics is presented in two dimensions based on the functional and physical architecture of the LS-IoT network.

The structure of the elements related to the processes such as computing, sensing, protocols, transmissions, etc., in the LS-IoT network constitutes the functional architecture. The structure of the actual physical devices on which all the functional elements are executed such as the M2M devices, GW, BS, Routers, Data Centres, etc., constitutes the physical architecture of the network. Both aspects are significantly influenced by the magnitude of applications and services.

Moreover, the characterisation of LS-IoT networks is underpinned by M2M communications which form the core principle behind the entire IoT network concept. M2M communication is described by Alam *et al.* [9] as the utilisation of wireless data connection to establish a link between systems, remote devices, locations and individuals to enable the collection of information, setting of parameters and sending or receiving indications of some phenomena on an enormous number of machines. The concept of M2M is driven by the idea that multiple interconnected devices are highly beneficial and valuable because they easily enable the support for ubiquitousness, pervasiveness and autonomousness of systems and processes using machines [10], [11]. The pervasive, ubiquitous, smart, intelligent and autonomous characteristics of M2M communication manifests into LS-IoT.

Therefore, all the above culminates into providing context to the large-scale nature of IoT networks. The key network domains and the two-dimensional characteristics of the large-scale nature of the IoT network are discussed below with reference to Fig. 1. Table 1 also provides a summary of the LS-IoT characteristics and is presented for each LS-IoT network domain based on their functional or physical characteristics.

1) FUNCTIONAL CHARACTERISATION OF LS-IoT

LS-IoT networks involve massive distributed sensing and actuation which render large amounts of data, requiring enormous on-board signal and data processing and short-term storage. Depending on the use case, the M2M local network involves frequent node-level media access scheduling, radio transmissions and reception for data exchange in a large-scale context. The M2M local network may also require dense and complex wireless network routing of frames given the geographic large-scale nature of the M2M local network.

The access network domain of a LS-IoT network involves the processing of massive heterogeneous traffic arrival patterns such as short traffic, burst traffic, aperiodic traffic and periodic traffic. The scale of processing relates to the complex multiplexing of traffic arrivals onto the much wider IP based network. The access network's large-scale attributes also include the extremely high communication resource management, utilisation and scheduling over constrained physical resources for a very large number of M2M devices. Thus, to manage many devices in the access network, complex algorithms and processes are required, and hence, the reason for the high utilisation of computing, memory and communication resources.

TABLE 1. A summary of the characteristics of LS-IoT networks based on the respective domains, their functional and physical characteristics.

Type of Characteristics	**M2M Local Network**	**Access Network**	**Core Network, Cloud and Application**
Functional	Massive distributed sensing and actuation. Large amounts of data acquisition. Enormous on-board data processing and storage. Frequent node-level media access scheduling, radio transmission and reception. Dense and complex routing of frames in the capillary network.	Processing of massive dynamic and heterogeneous M2M traffic pattern arrivals. Complex multiplexing of traffic onto the core network. High communication resource management, utilisation and scheduling.	Allocation of many IP addresses. Enormous data storage and computational capabilities in the cloud. Complex and large data analysis for application logic. Dynamic and frequent API interaction for service applications.
Physical	Enormous magnitude of wirelessly interconnected M2M devices. Existence of many heterogeneous M2M devices. Existence of multiple local M2M networks (capillary networks).	Moderate number of access network devices. Coexistence of heterogeneous access network devices.	High capacity connectivity infrastructure with worldwide access to M2M packets. Distributed data centres consolidated into one cloud platform infrastructure. Large number of devices supporting IoT logical IoT applications and services such as smart-phones and computers connected to the cloud.

In the Core Network, the LS-IoT involves the allocation of many IP addresses and packet routing over the Internet with the reliance on a highly constrained IP address availability. The scale of IP addresses is directly associated with the number of actively connected devices (GWs, BSs or M2M devices) communicating through the core network. The cloud data centre and application domains involve the influx of large and complex data that need complex systematic approaches to establish substantial logic for the various services and application through complex and dynamic API deployment and execution in the context of LS-IoT.

2) PHYSICAL CHARACTERISATION OF LS-IoT

In terms of the physical characteristics of the LS-IoT network, the M2M local network is attributed to the existence of an enormous magnitude of wirelessly interconnected heterogeneous M2M devices. Multiple local M2M networks of devices may coexist. Such characterisation is widely sensitised by many stakeholders such as the 3rd Generation Partnership Project (3GPP) which predicted that the number of devices in a cell of a typical Long-Term Evolution (LTE) network will exceed 30000 compared to the current average of 50 devices per cell [12]. This is attributed to the proliferation of M2M devices in the M2M local network.

The access network domain comprises one or more GW or BS devices for adequate long-range coverage in LS-IoT. However, one main characteristic is the coexistence of heterogeneous access network devices for LS-IoT networks. Due to the diverse applications and use cases of LS-IoT networks such as agricultural and industrial production monitoring, heterogeneous wireless communication technologies are prominent in different M2M devices, and hence, require association with compatible network technologies in the access network.

The core network, cloud data centre and application domain are by nature already large-scale infrastructure for traditional networks, and therefore, their characterisation in terms of LS-IoT networks are more adequately defined in terms their functional architectural characteristics in LS-IoT networks. However, the core network, cloud data centre and the application domain can still be characterised as a high capacity network infrastructure, a consolidation of data centres to form a cloud platform and a large number devices connected to the cloud for application services respectively.

III. THE MEDIA ACCESS CONTROL

A. OVERVIEW OF THE MAC PROTOCOL

There are two core functions required between the access network devices and the M2M devices. The first core function is the management of the transmission medium for multiple devices by employing multiple access techniques to ensure that the constrained transmission resources are effectively utilised for a successful transmission [13], [14]. The multiple access techniques are based on rules that indicate how the transmission resources are allocated and controlled. These rules are stipulated by the MAC protocol also called the MAC scheduling protocol [15]. This ensures that multiple frames from or to the different devices do not encounter interference over the media when communicating with the access network devices. All these functions are provided by the dedicated MAC sub-layer of the Data Link layer. The second core function is to ensure that all MAC frames are logically multiplexed to the network layer which may consist of multiple network protocols. This is achieved by including network protocol-specific headers to the MAC frames to enable multiple network layer protocol transmissions to the core network. This function is provided by the Logical Link Control (LLC) sub-layer of the Data Link layer within the GW and BS devices. There are other functions provided by the GW and BS but are out of the scope of this paper.

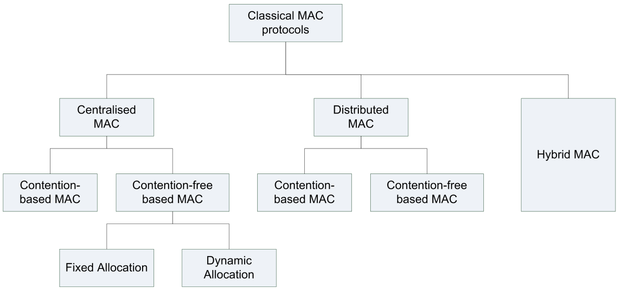


FIGURE 2. **A classic classification of MAC protocols in a typical wireless network.**

The MAC protocol relies on the Physical (PHY) layer's multiplexing mechanism where multiple signals from different devices are multiplexed onto a constrained spectrum at the same time or different periods. While the PHY layer multiplexing deals with the combining of signals onto the transmission resources, the MAC logically deals with the devices itself using the defined media access rules. Therefore, the MAC protocol abstracts the access network based on all the connected devices and all the available transmission resources to provide the media access rules. From a low-level perspective, this is seen as the scheduling of frames from multiple M2M devices implemented by a scheduler. The scheduler performs the assignment of transmission time and channels for a frame transmission from all devices distributively or centrally [16].

B. CLASSIFICATION OF MAC PROTOCOLS IN WIRELESS NETWORKS

The scheduling or allocation of transmission resources for frame transmission may be controlled by devices themselves without any co-ordination by the GW or BS. Such a strategy is known as a distributed MAC scheme. On the other hand, the allocation and scheduling of transmission resources may be controlled centrally by the GW or BS which is known as the centralised MAC scheme. Both distributed and centralised schemes can be applied together to provide a hybrid scheme.

Simultaneous media access by multiple devices in any wireless network can lead to collisions which affects throughput, network efficiency, and other network quality indicators. The management of communication resources using the MAC protocol is aimed at either avoiding collisions or resolving them. Based on this concept, MAC protocols can be further sub-divided into contention-based protocols, contention-free based protocols, and hybrid-based protocols [17], [18] as shown in Fig. 2. The hybrid-based protocol in this case may be based on the combination of contention-based and contention-free based protocols.

In contention-based protocols, devices randomly or sporadically attempt media access which exposes the devices to possible simultaneous media access. However, contention-based protocols employ a Collision Resolution

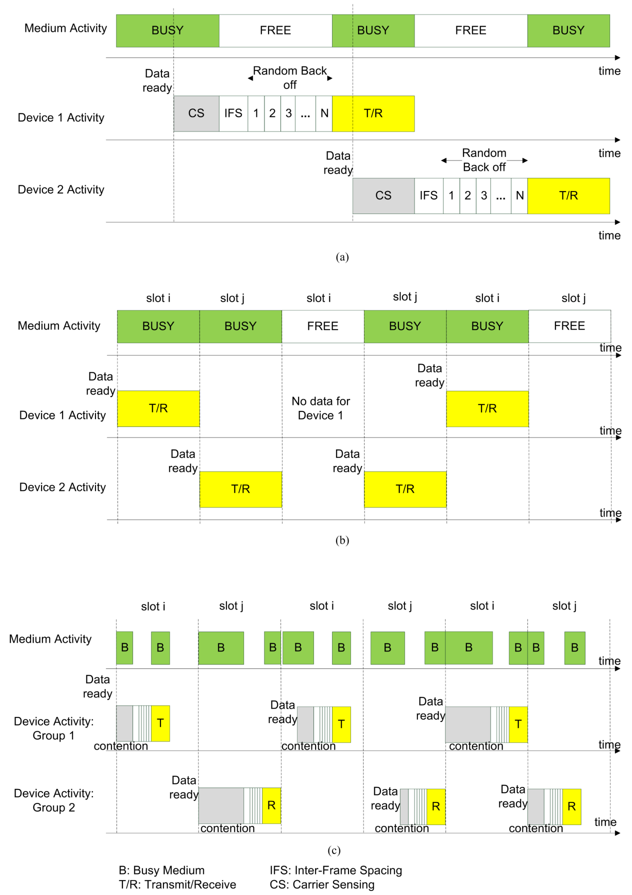


FIGURE 3. **(a) Contention-based MAC Protocol. (b) Contention-free based MAC protocol.(c) Hybrid MAC protocol.**

Algorithm (CRA) to ensure that the detected simultaneous media access is resolved. CRAs are classically employed distributively. A distributed CRA requires devices themselves to execute randomised re-transmission media access schedules or reservations. This concept is illustrated in Fig. 3 (a) whereby device 1 has some data ready to be transmitted. However, the device senses a busy medium until a time when the medium is free. When the medium is free, the device then waits for a fixed duration called an Inter-Frame Space (IFS) plus a random back-off delay time. This process randomly schedules device 1 for transmission. However, if the transmission is unsuccessful, the random back-off delay is computed again until a successful transmission occurs. The same process is executed by device 2. The contention-based scheduling is widely known for its ease and flexibility of implementation as well as reduced overhead [19]. However, it cannot scale due to the high collision probability with many devices.

In contention-free based protocols, a scheduler is distributively or centrally employed to allocate the communication resource. In contention free-based MAC scheme, a device may be permanently or dynamically allocated some communication resource. Devices only use their allocated communication resource to transmit, which may be coordinated by a coordinating device such as a GW or BS. The basic operation of the contention-free based approach is depicted in Fig. 3 (b). Device 1 only transmits within the slot labelled i whereas device 2 only transmits during the slot labelled j. During the assigned slot duration, the medium is busy for the entire transmission period. It should be noted that, the way time slots are allocated is generic and may differ depending on the network performance metrics. Numerous algorithms exist for the pre-assignment or scheduling of timeslots for devices. Some of these algorithms include BodyMAC, S-MAC and UWAN, and can be found in [20]–[22] respectively. Contention-free based protocols avoid collisions in time and spectrum. However, these schemes cannot scale since more resources are needed in proportion to the number of devices. Thus, more time slots and more channels are required.

In a typical hybrid MAC protocol, a coordinator provides synchronised time slots over a time frame for groups of devices. Each group of devices then employs contention-based approach to access the media within a given time slot. This process is illustrated in Fig. 3 (c). Device 1 which belongs to a group of devices that can contend for media access in slot i contends for media access using the contention-free based approach. The busy medium depicts a transmission or reception activity either by device 1 or other devices within that slot. Likewise, device 2 also competes for channel access with a group of stations in slot j and the slots are repeated periodically. Such hybrid approaches are sometimes called synchronous contention-based protocols [23] and are the basis for many MAC scheduling protocols in existing wireless communication protocols and standards for IoT and M2M communication.

C. CLASSIFICATION OF IMPROVED MAC PROTOCOLS FOR LS-IoT NETWORKS

The characterisation of LS-IoT networks requires specialised protocols to deal with the scalability issues that may arise as a result of the massive number of M2M devices and the heterogeneous traffic patterns. Some approaches are presented in literature to deal with large-scale media access of M2M devices. These approaches use concepts that fall under one or more of the general classification in Fig. 2. In this work, another level of classification specific to concepts used to improve MAC in LS-IoT and massive M2M is presented. By studying the technical objectives and the core approaches of some current MAC strategies of LS-IoT networks, the MAC protocol can be categorised as indicated in Fig. 4.

1) HYBRID FRAME-BASED APPROACH

Hybrid frame-based approaches are based on dividing the available transmission time into independent service periods whereby each service period is allocated for a MAC specific function and is orchestrated either by the device or by the GW. Typically, the individual frames are made up of contention-free periods and contention periods [24]. The contention for media access is executed during the Contention Only Period (COP) and transmission is executed during the

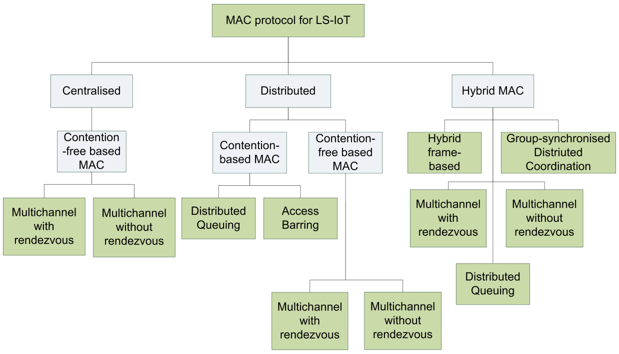


FIGURE 4. **The classification of MAC protocols based on the reviewed improved MAC strategies for LS-IoT and massive M2M networks.**

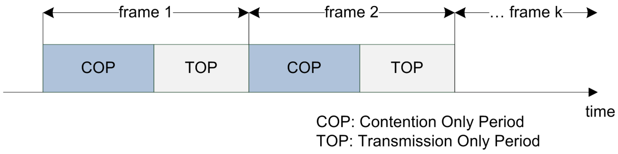


FIGURE 5. **Basic hybrid frame-based MAC for LS-IoT.**

Transmission Only Period (TOP). This approach is a variant of hybrid MAC protocols which provides the combined benefits of contention-based and contention-free based protocols to enable scalability in LS-IoT networks. However, the time constraints imposed by the framing of MAC service periods is a major concern. A typical illustration of this class of MAC protocols is presented in Fig. 5.

2) MULTICHANNEL HYBRID APPROACH WITH OR WITHOUT RENDEZVOUS

In multichannel hybrid approaches, multiple access is granted through multiple frequency channels. Devices may choose frequency channels to transmit on but may still employ contention-based or contention-free based protocols to reduce or avoid possible frame collisions as a result of simultaneous channel access. This type of scenario is illustrated in Fig. 6. This scheme may also be based on using a rendezvous whereby devices may negotiate for media access on a dedicated common channel at any time (illustrated in Fig. 7 (a)) or alternating times (illustrated in Fig. 7 (c)) before switching to different channels for transmission using a hybrid approach. The process of negotiation on a given channel is referred to as a rendezvous. Another scenario based on the rendezvous is based on two devices switching together from one channel to another (illustrated in Fig. 7 (b)) or separately (illustrated in Fig. 7 (d)) until the transmission is agreed upon. A rendezvous algorithm is required to provide coordination of the search for a common channel to exchange frames. A comparison of these scenarios is presented in [25]. The multichannel approach is known to provide improved capacity in wireless networks [26], [27]. Therefore, the use of multichannel with or without rendezvous for MAC scheduling has been widely used to address MAC related challenges in LS-IoT networks.

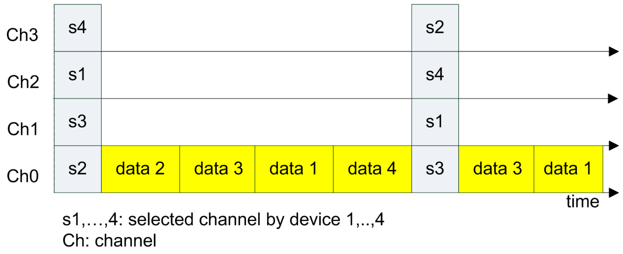


FIGURE 6. **Multichannel without Rendezvous.**

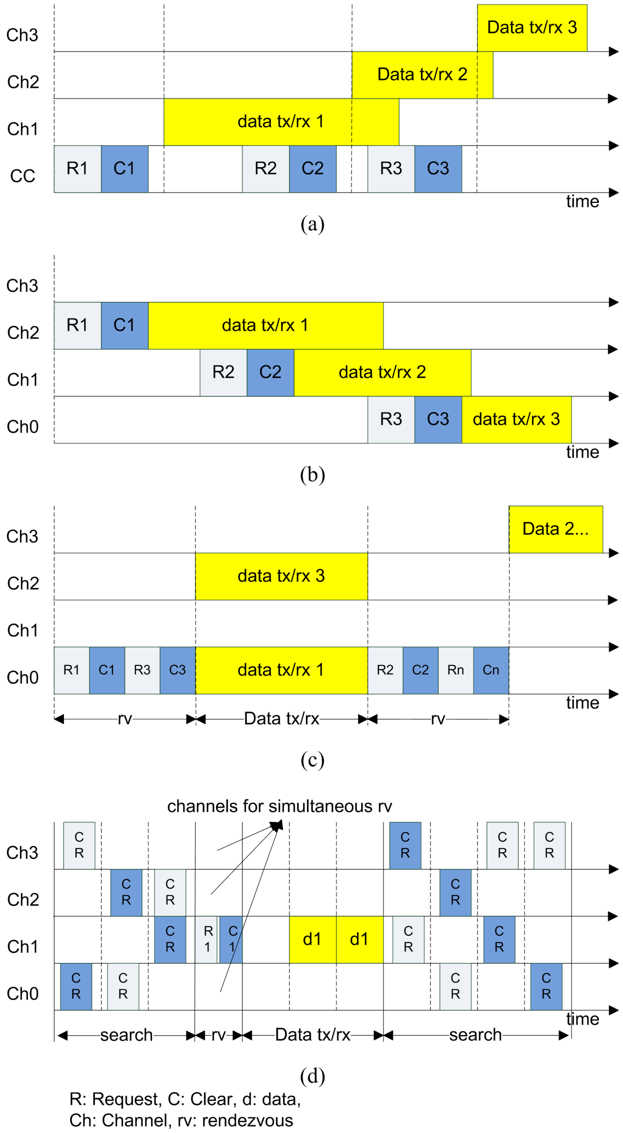


FIGURE 7. **Multichannel with rendezvous: (a) Using a dedicated control channel. (b) Using multiple channels switching together. (c) Split into phases. (d) Using multiple channels switching separately.**

3) DISTRIBUTED QUEUING RANDOM ACCESS APPROACH

The Distributed Queuing (DQ) is based on multiple queues (typically two queues) whereby devices either enter a queue to resolve a collision or enter a queue to perform transmission. This type of MAC scheme is employed to eliminate the need for back-off periods and to enable scheduling based

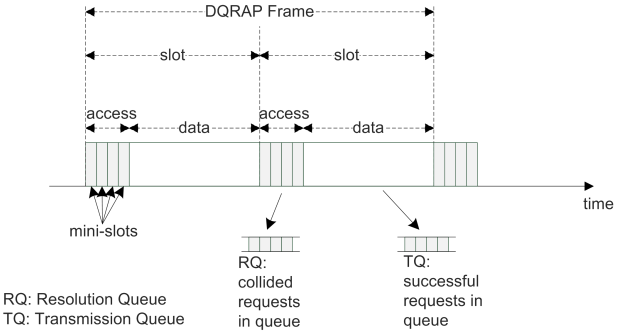


FIGURE 8. **The frame structure of a classical distributed queueing random access protocol (DQRAP).**

on QoS requirements [28]. The typical Distributed Queuing Random Access Protocol (DQRAP) is a classical technique used for providing a superior performance of MAC scheduling by minimising access delays and increasing throughput for large networks such as LS-IoT networks. As depicted in Fig. 8, the DQRAP divides the time frame into slots. In each slot, a part of the slot is reserved for media access while another is reserved for data transmission. The access part is further divided into a number of mini-slots whereby each mini-slot represents an access opportunity. The DQRAP uses a collision Resolution Queue (RQ) and a Transmission Queue (TQ) in its operation. A device randomly selects a mini-slot and tries to request for access. If a collision occurs in the mini-slot, the collided device is queued in the RQ in the order of the selected mini-slot. A collision resolution algorithm is then employed to process the RQ queue. If the access request was successful, the device is queued in the TQ and transmitted during the next scheduled transmission. This process improves media access through the queuing process. As a result, this algorithm has been extended for various applications such as WLAN [29] and cellular networks [30]. A number of authors have employed this approach and proposed unique modifications to address MAC scheduling in M2M and LS-IoT networks.

4) ACCESS BARRING APPROACH

The concept of access barring is based on grouping devices into certain access classes and blocking some class of devices from performing Random Access (RA) procedures. The BS broadcasts a bitmap frame to all associated devices indicating the devices that are barred from performing a RA procedure. The BS updates the bitmap after a barring check (usually based on traffic conditions) and pages each device in a predefined order to receive an updated bitmap. The barring process may be evoked depending on the number of devices and the number of access requests. The default access mechanism is applied if the barring procedure is not evoked. Fig. 9 illustrates the access barring process based on the Extended Access Barring (EAB) which is a variant of the Access Class Barring (ACB) scheme used to deal with increased contention and congestion.

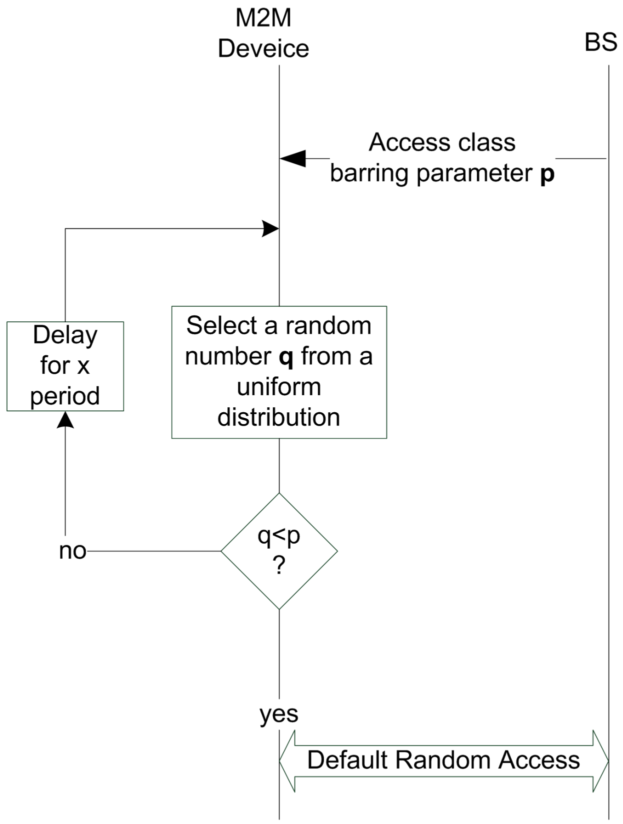


FIGURE 9. **Basic access barring procedure.**

5) GROUP-SYNCHRONISED DISTRIBUTED COORDINATED APPROACH

Group-Synchronised Distributed Coordinated scheduling approach is based on restricting media access to groups of devices within a period. The restricted period is known as the Restricted Access Window (RAW). Within the restricted period, devices contend for channel access in a distributed manner. Restricting media access to a group of devices reduces the average collision probability which enables the network to achieve stable throughput with an increasing number of devices. Group synchronisation is studied in literature and is considered as being very useful for improving the MAC scheduling performance in terms of scalability of the LS-IoT where simultaneous media access contention is highly probable [31], [32]. This is a key motivation for the adoption of group-synchronised distributed media access in the 802.11ah standard to support many M2M devices [33]. Fig. 10 shows the time frame structure between two beacons with multiple RAWs. Each RAW is made up of RAW slots which a group of devices use to contend for channel access. The beacon is broadcasted by the GW.

IV. IMPACT OF LS-IoT CHARACTERISTICS ON THE NETWORK AND MAC

A. LS-IoT IMPACT ON THE NETWORK

The large-scale characteristics discussed in the previous sections have some impact on network performance. It can be

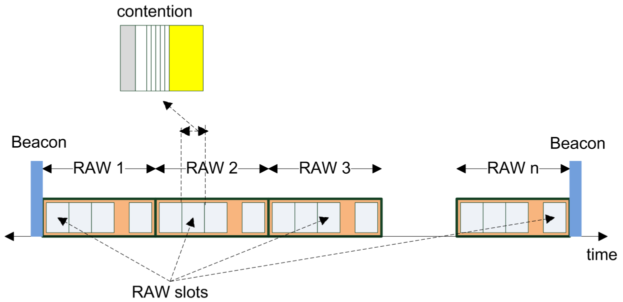


FIGURE 10. **Beacon interval representation of a group-synchronised distributed coordinated MAC approach.**

argued that the impact is largely determined by the M2M local network domain. Moreover, to facilitate the end-to-end exchange of data through the core network, the access network devices play a central role in ensuring that all the M2M devices within the local network are effectively managed and coherently interconnected with the other domains.

In LS-IoT, the sporadic and enormous deployment of M2M devices is beyond the control of the network. However, imposing regulatory limitations on the size of the M2M local network could hinder the ability to harness the advantages of IoT such as ubiquitousness, pervasiveness and autonomousness. Therefore, the growing trend of M2M devices inevitably makes the M2M local network uncontrollably large and results in enormous data generation. The impact of such synopsis is reflected in the core network, cloud data centre, and application domains which are typically characterised as very large infrastructure and are managed in large-scale. The emergence of the IoT paradigm has resulted in the dynamic optimisation of such infrastructure and in some cases, have been regarded as being underutilised. As a result, the paradigm of cloud computing has surfaced in order to harness such underutilised resources in the core network and in the cloud. This phenomenon has been highlighted in several studies including [34]–[38].

This underutilised core network and cloud physical resources is testimony to the constriction effect which hinders effective data flow from the large number of M2M devices in the M2M local network to the core network and further to the cloud. If the enormous number of M2M devices in the M2M local network are uncontrollably increasing, the access network devices need to deal with the complexities of managing the exchanging of data to and from the core network. The access network devices create a barrier for end-to-end traffic delivery. The exchange of data between the M2M local network and the core network onwards is constrained by the GW and BS creating a bottleneck effect in the LS-IoT network. Such phenomena is depicted in Fig. 11. The authors in [39]–[41] analysed and articulated a similar phenomenon in different network types including wireless sensor networks, whereby amongst all the identified bottleneck points in the network, the GW or BS devices are seen to have the greatest bottleneck effect.

B. LS-IoT IMPACT ON THE MAC PROTOCOL

Studies such as [42] show that the very large number of M2M devices in a capillary M2M network or a cell network contributes to the most demanding challenges in IoT networks in general. The MAC protocol is typically designed by considering the number of possible interconnected devices in the network. This means that the size of the M2M local network plays a critical role in developing an effective MAC solution. The LS-IoT's peculiar characteristics associated with the pervasive and ubiquitous nature of M2M device deployment leads to serious challenges in M2M networks [43]. The impact of LS-IoT is evident on the shared media resource which is essential for guaranteed connectivity. In LS-IoT networks, the shared media resource is regarded as an extremely scarce resource due to the massive number of M2M devices that require media access. Therefore, access to this resource always needs to be managed effectively to ensure adequate connectivity.

It could be argued that increasing the number of access network devices in proportion to the growing size of the M2M local network coupled with an interference mitigation strategy may address the bottleneck effect. While this may be true, the GW and BS are either fixed infrastructure or based on dedicated hardware with a lack of dynamism. Increasing the number of GWs and BSs require mores greater deployment of dedicated hardware resources which contributes to high Capital Expenditure (CapEx). A cost-effective option is usually to improve the functional elements associated with the MAC protocol design to achieve the required elasticity to avoid a possible bottleneck effect. Thus, the MAC protocol can be optimised to provide scalability in a network. The goal in designing a suitable MAC protocol for LS-IoT networks is primarily to achieve scalability for a dynamically increasing number of M2M devices. Some of these issues that affect the scalability of the MAC scheduling protocol are discussed further.

C. MAC DESIGN ISSUES FOR LS-IoT

MAC scalability is the ability of the MAC protocol to efficiently accommodate an increasing number of frame transmissions, reception, and resource allocation. A scalable MAC ensures seamless multiple media access scheduling for an extremely large and dynamically growing number of M2M devices while effectively maintaining QoS requirements of the LS-IoT network. The core challenges faced with achieving such MAC scalability for LS-IoT networks are discussed comprehensively below.

1) COLLISIONS

Scalability of the MAC protocol in LS-IoT networks is significantly affected by the high collision probability in contention-based MAC schemes. The collision probability is established based on the likelihood that there will be at least one or more devices that may randomly select or compute the same back-off schedule following the detection of a free

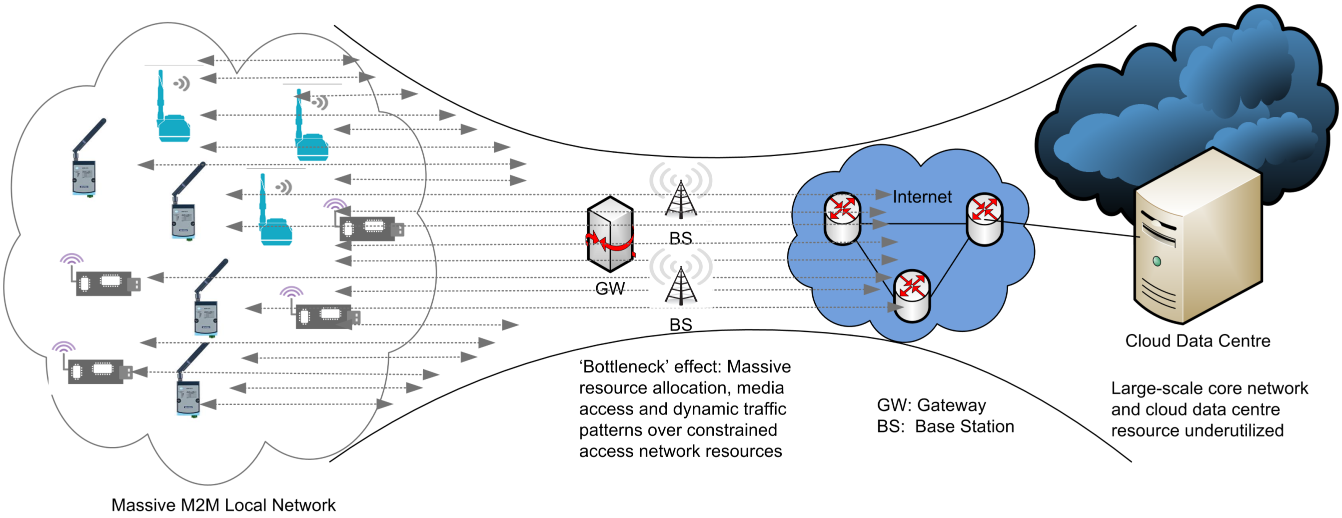


FIGURE 11. **A depiction of the bottleneck effect in a LS-IoT network caused by the GW and BS.**

medium for a transmission attempt [44]. This likelihood certainly increases for LS-IoT networks because many devices lead to a high likelihood of transmission or reception of frames. To support this assertion, a number of models have been formulated in literature based on different approaches such as in [45]–[49], to establish the relationship between the conditional collision probability of a frame transmission and the number of contending devices. In [50], the number of contending devices n is modelled as a function of the probability of a collision p as given in (1) and based on a classical contention-based scheduling method, Distributed Coordination Function (DCF). CW_{min} and m represent the back-off parameters related to the minimum contention window and an exponential doubling factor for determining maximum contention window $CW_{max} = 2^m CW_{min}$ respectively.

$$n = 1 + \frac{\log(1-p)}{\log\left(1 - \frac{2(1-2p)}{(1-2p)(CW_{min}+1)+pCW_{min}(1-(2p)^m)}\right)} \tag{1}$$

Using (1), the relationship between the number of contending devices and the conditional collision probability is illustrated in Fig. 12. Fig. 12 shows the rapid increase in the conditional collision probability as the number of contending devices increases for the different contention window values. When collisions become highly probable, delays emanate, and to some extent, the transmission fails inevitably after many retransmission attempts. This affects effective throughput significantly while trying to accommodate many devices. Thus, the scalability of the MAC protocol is greatly affected by the high rate of collision probability in LS-IoT.

2) CONTROL OVERHEADS

Generally, wireless MAC protocols necessitate the exchange of control frames for signalling to make the protocol practically possible [51]. These control overheard frames have an impact on the MAC protocol. In the widely adopted

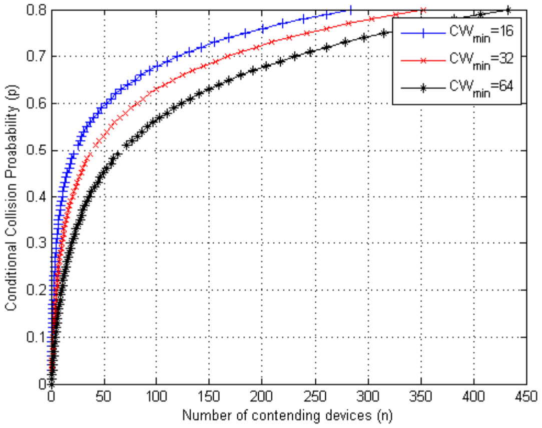


FIGURE 12. **The effect of rising number of contending devices on the probability of a collision due to multiple transmission attempts at a given time based on contention-based MAC scheduling approach using DCF.**

hybrid-based MAC protocols and the contention-free based protocols, synchronisation between devices is necessary. The synchronisation of devices implies that the duration of the signalling information needs to be compensated for within the time frame duration that is to be allocated for transmission. As a result, there is a consequence in terms of reduced available time for contention and frame transmission.

In terms of LS-IoT networks, the need for the coexistence of dynamic traffic patterns and heterogeneous device types such as Human-Type Communication (HTC) and M2M communication makes large control overheads inevitable [52]. The MAC protocol needs to facilitate media access to ensure reliable transmission for a multitude of unrelated devices. Achieving such cohesion to minimise complexity and ensure seamless coexistence in LS-IoT networks require reliable control information. However, this comes at a huge cost leading to increased control overheads for the MAC protocol. This significantly affects the scalability of the

MAC protocol. Some LS-IoT networks that aim to achieve coexistence at the expense of increased control overheads can be found in [53]. A classic case of the impact of LS-IoT on control overhead issues is the Physical Random Access Channel (PRACH) approach which uses 64 preambles for contention-free and contention-based media scheduling [54]. The Third Generation Partnership Project (3GPP) group has identified the RACH procedure as a major issue for the existence of M2M communication and HTC devices due to the surge in signalling overhead caused by the increasing number of M2M devices attempting media access at the same time [55].

3) TIMING CONSTRAINTS

The MAC service time is a critical component of the resource scheduling process. The service time indicates how fast, how effective, and how scalable transmission and reception may be achieved when a device is saturated with data. For a typical contention-based MAC protocol, the MAC service time is the interval between the start of contention and the end of a successful transmission (thus, including acknowledgement) or until the frames are discarded due to a failed transmission [56]. The MAC service time T_{mac_srv} can be represented using (2), where T_{suc} is the duration of a random successful transmission, T_{col} is the duration of a random collision and T_{bof} is the duration for the back-off process.

$$(T_{suc} + T_{col} + T_{bof}) \leq T_{mac\_srv} \tag{2}$$

For contention-free based protocols, T_{col} and T_{bof} could be negligible. From a general perspective, T_{mac_srv} may also accommodate for T_{grd}, T_{prc} and T_{phy} such that (2) becomes (3).

$$(T_{suc} + T_{col} + T_{bof} + T_{grd} + T_{prc} + T_{phy}) \leq T_{mac\_srv} \tag{3}$$

In (3), T_{grd} is a guard time to accommodate any synchronisation offset. The correction of synchronisation offsets can be achieved using synchronisation protocols such as those mentioned in [57]. T_{prc} is the duration for any computational processing of frames that are necessary to establish critical parameters for resources provisioning. This includes time slot computation, contention windows, synchronisation times and other complex mandatory or optional computational functions. The computational processing delays during MAC scheduling is also directly affected by the PHY and lower MAC computational processes such as Clear Channel Assessment (CCA) for checking a busy medium and decoding for detecting errors to be able to generate an ACK or NACK. Such delays must be accounted for. Some analysis on the delays due to the processing of frames in some wireless technologies can be found in [58]–[60].

The inter-frame arrival rate is also important for discussing the timing constraint of the MAC protocol. In [61] some simulation and analysis are conducted to establish the relationship between frame inter-arrival time and the collision probability for contention-based MAC scheduling. It shows that when the average inter-arrival time increases, the collision probability drops exponentially and so does the average delay. This simply means that for low inter-arrival times, the T_{mac_srv} is usually able to accommodate all the timing delays before the next schedule begins which reduces the collision probability.

With LS-IoT, low inter-arrival times may not always be the case. In LS-IoT, industrial-based M2M connectivity contributes significantly to the large-scale nature of IoT networks and sometimes referred to as Industrial IoT (IIoT) [62]. Automation and control application in mega-factories, aviation, road transportation and shipping, collectively require millions of sensors. Such applications contribute to the greatest number of sensors in a confined place at a given time with a typical example being a commercial aircraft embedded with about 6000 sensors. Such LS-IoT networks rely on real-time data and strict latency requirements of the wireless connection due to strict regulations on safety and other factors. For such LS-IoT network applications to be possible, high inter-arrival times of frames are inevitable and the T_{mac_srv} must be strictly bounded. To allow the multitude of massive sensors to communicate in scale with such high inter-arrival time, the T_{mac_srv} must be reduced significantly. However, this relies on several delay factors defined earlier which in some cases are not easily reduced or are randomly defined. Also, the underlying hardware and firmware factors such as interrupts, multi-thread processing, and switching affect the ability to reduce T_{grd} and T_{prc} delays [63]. This hinders the possibility of achieving a scalable MAC protocol.

4) SPECTRUM CONSTRAINTS

The design of a scalable MAC may require the need for adequate bandwidth. Several traditional wireless communication technologies still rely on the licensed spectrum due to the signal fidelity and minimal spectrum interferences that come with using the licensed band. However, MAC protocols that rely on the licensed band are highly constrained due to the limited availability of spectrum. On the other hand, many IoT standards are moving to the unlicensed spectrum band to provide connectivity for M2M devices since it is free of monetary cost. The fact that the unlicensed band may be used without formal allocation means that there could be high interference.

In the case of LS-IoT, the existence of a multi-spectral access network is a reality. LS-IoT access networks are meant to provide a converged multi-standard connectivity of devices. Given that most MAC protocols that rely on the licensed band are optimised for HTC communication, LS-IoT networks increase the resource utilisation of the MAC protocols optimised for the licensed spectra. As such, the scalability of the MAC protocol is greatly affected especially with multichannel based MAC protocols for LS-IoT.

Having a massive number of heterogeneous devices converged within the access network and operating in the unlicensed band causes huge congestion. Therefore, the MAC protocol could be constrained as a result of channel bandwidth and channel interference. The effects of spectrum constraints and interference on the scalability of a link

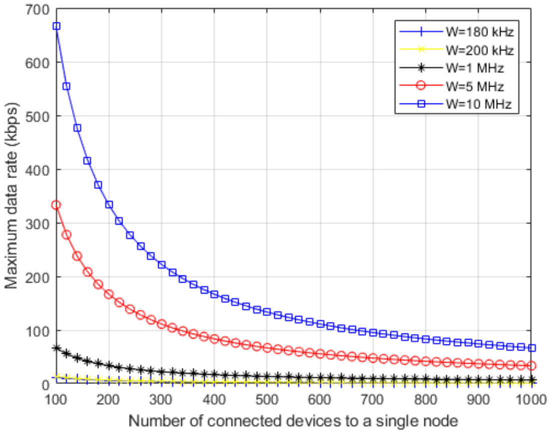


FIGURE 13. **The effect of a large number of devices and spectrum bandwidth constraints on the maximum data rate. SINR = 20 dB.**

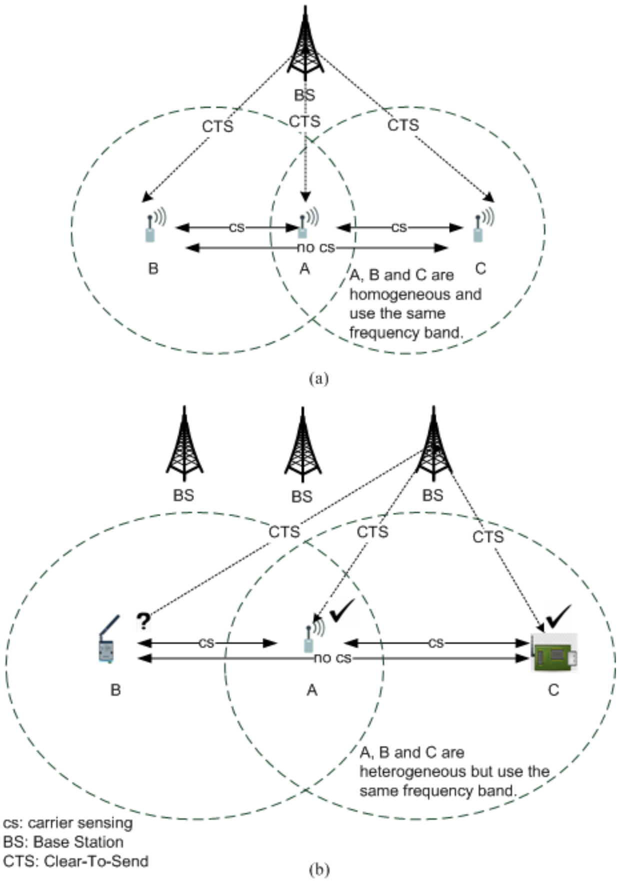
between device *i* and device *j* in terms of the maximum data rate R_{ij} as provided in [64] is expressed in (4).

$$R_{ij} = T_{ij} \times Wlog_2(1 + SINR_{ij}) \qquad (4)$$

The transmission opportunities for devices in the link is denoted as T_{ij} which depends on the total number of device connections to device i or vice versa. W represents the bandwidth of the channel and $SINR_{ij}$ represents the Signal to Interference and Noise Ratio of the link (i, j). From (4), LS-IoT networks with a constrained spectrum bandwidth (thus, a limited W), a large number of device connections (thus, a lower T_{ij}) and a high level of interference (thus, a lower $SINR$) will result in a reduction in the maximum data rate. This phenomenon is also depicted in Fig. 13 based on a SINR of 20 decibels (dB).

5) HIDDEN TERMINAL

The hidden terminal problem generally exists in wireless networks. MAC protocols that rely on the ability of a device to sense the spectrum to detect any possible collision before transmitting (such as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)) are exposed to the hidden terminal problem. The hidden terminal problem is illustrated in Fig. 14 (a) whereby device B is unable to sense a carrier from device C and vice versa. This results in the false sensing of a free medium by device B when device C is busy on the medium and vice versa. The hidden terminal problem in wireless networks is generally mitigated using a synchronisation signal based on the Request to Send (RTS) and Clear to Send (CTS) handshaking process [65], also called virtual carrier sensing. Upon receiving an RTS frame from a device that has ceased the media, the BS broadcasts a CTS frame which indicates the duration of the frame transmission of the device. This enables the devices that are hidden to enter a deferral period before media access [66].



(a)

(b)

FIGURE 14. **(a) Classical Hidden node problem with homogeneous devices mitigated using virtual carrier sensing. (b) LS-IoT Hidden node problem still exists with heterogeneous devices.**

In LS-IoT, the use of virtual carrier sensing may not mitigate the hidden terminal problem. The unlicensed band attracts a lot of IoT applications and services due to the low cost and flexible utilisation of the spectrum. This leads to the LS-IoT network being characterised by heterogeneous devices that operate within the same frequency band and may exist in proximity together. With this characterisation, Fig. 14 (b) illustrates the hidden terminal problem that may exist even with the use of synchronisation signals for virtual carrier sensing. In Fig. 14 (b), device A, B and C operate in the same band but employ heterogeneous PHY and MAC technologies. In this case, sensing a carrier in the same band is possible even though they employ different technologies. However, with device B hidden from device C and using a different technology, a CTS frame broadcast by the BS may not be decoded by device B, and therefore, device B cannot defer its transmission in that channel, leading to a false assumption of a free medium. In this case, the hidden terminal problem may still exist even with virtual carrier sensing. This indicates how interference is eminent in the coexistence of heterogeneous access networks in LS-IoT which results in significant collisions which affect the scalability of the MAC protocol.

6) ENERGY CONSTRAINTS

M2M devices often depend on batteries for energy. However, batteries do not provide guaranteed lifetime energy for devices, and therefore, there is a need to conserve energy. Such constraints on energy consumption may influence designing a scalable MAC protocol. In dealing with collision, some MAC protocols need to logically control the active status of the physical radio module. However, one of the contributing factors to high energy consumption in IoT is the prolonged active status of the radio module for transmission and reception of data. For example, after a collision, a device attempts to retransmit its data which means that the device must contend for media access to be scheduled again. As a result, the radio may stay active for the duration of the retransmission attempt. Depending on the protocol used, the retransmission attempts can be as many as 5 times or more. Every attempt to retransmit data after a transmission failure results in the accumulation of control frames. This essentially contributes to very high energy consumption. Typically, in the IEEE 802.11 MAC protocol, the control frame overheads contribute to almost 50% of the total energy consumption [67].

In LS-IoT networks, the scalability of the MAC protocol is influenced by the energy constraints of the device. This is because LS-IoT networks deal with ubiquitous and pervasive devices with huge constraints on battery energy. After all, some of these devices are autonomous and may not be able to be recharged or maintained frequently. Due to such constraints, the scalability of the MAC protocol must be sensitive to the energy consumption profile of the LS-IoT. This affects the ability to implement some collision resolution algorithms and other algorithms. The constraints on energy consumption affect how the MAC protocol logically controls the PHY layer to achieve the required scalability of the MAC protocol. For example, in [68] and [69], the authors proposed an approach based on an energy-efficient clustering for massive access in M2M networks using an incentivised cooperative approach. Such an approach will require the use of many control overheads to perform clustering and initiate incentives which will directly affect the MAC protocol. The trade-off between energy efficiency and MAC scalability needs to be addressed to achieve optimal performance of the MAC scheduling protocol for massive devices in LS-IoT.

7) HARDWARE CONSTRAINTS

The advent of Software-Defined Radio (SDR) brings together parts of the PHY layer and the MAC layer onto one System on Chip (SoC) platform. In this type of modern design approach, the PHY, MAC and Network layer functions are easily implemented on a modular platform. The SoC architecture typically involves a distributed processing of PHY layer and MAC functions on Field Programmable Gate Arrays (FPGAs) and microprocessors respectively [70], [71]. The SoC also includes a Radio front-end (for radio frequency conversion), an Analog-to-Digital Conversion (ADC), a Digital-to-Analog Conversion (DAC), a flash memory (for storing the firmware that runs the MAC protocol) and a Random Access Memory (RAM) (for temporary storage during firmware execution and computations). The constraints imposed by the hardware components that directly affect the scalability of the MAC scheduling protocol are based on timing. A scalable MAC protocol depends on the interfacing between all the components of the SoC platform. The interconnection between the different PHY processing units in the FPGA and the microprocessor affect the timing of the MAC protocol. The different PHY processing units cause significant delays in the path between the different PHY processes and the MAC scheduling functions [72].

The MAC protocol may require the radio to switch between sleep, transmit and receive modes frequently. This process introduces some amount of delay which affects the ability to provide timely MAC scheduling functions to support scalability [73]. The synchronisation of devices during MAC scheduling requires the acquisition of timing information through timestamping. The accuracy of the timestamp is critical in ensuring an effective scalable MAC protocol.

The RAM-Processor combination enables the temporary storage and accessibility and of data during the MAC protocol execution. The RAM is usually a resource shared amongst different layers of the communication protocol stack. For example, the network layer of the GW may use the same RAM resource for processing the network layer packets such as the assignment of IP and other associated tasks.

For LS-IoT networks, the massive frame exchanges result in the culmination of all the processing path delays and queuing which causes a slowdown of the MAC scheduling of frames and affects the time constraint requirements. On the other hand, if the processing of frames during the execution of the MAC protocol is not fast enough due to the microprocessor incapability, new frames arriving from the PHY may result in memory clogging leading to a bottleneck situation which in turn affects scalability. Such hardware constraints are very critical in LS-IoT networks since the arrival rate of data at the radio front-end may rise with an increasing number of connected devices which leads to a bottlenecking.

Moreover, LS-IoT networks have the potential to generate a large number of frames arriving at the MAC layer for scheduling which may trigger interrupts and several other instructions [63]. This may cause delays in capturing timestamps leading to synchronisation problems and affecting the scalability of the MAC protocol of the LS-IoT network.

D. SUMMARY OF THE IMPACT OF LS-IoT ON THE MAC

The MAC design issues are all attributed to some of the LS-IoT characteristics. Therefore, the impact of such characteristics on the MAC protocol in terms of the related MAC design issues is presented in Table 2 showing how the design issues can be related to the LS-IoT characterisation and MAC performance.

V. CURRENT PROPOSED IMPROVEMENT STRATEGIES ON MAC PROTOCOLS FOR LS-IoT

In this section, several proposed strategies in literature for addressing the scalability of the MAC protocol in LS-IoT and massive M2M networks are reviewed. The proposed

TABLE 2. Summary of the relationship between the LS-IoT network characteristics, the MAC design issues and the impact on the MAC protocol's performance.

LS-IoT characteristics	MAC design issues	Impact on MAC performance
Frequent node-level media access scheduling, radio transmission and reception. Enormous magnitude of wirelessly interconnected M2M devices.	Collisions.	High collisions, increased delays and reduced throughput.
Processing of massive dynamic and heterogeneous M2M traffic pattern arrivals. High communication resource management, utilisation and scheduling. Coexistence of heterogeneous access network devices.	Control overheads.	High control overheads for signalling different devices, increased delays and reduced throughput.
Processing of massive dynamic and heterogeneous M2M traffic pattern arrivals. High communication resource management, utilisation and scheduling.	Timing constraints.	High frame inter-arrival times, increased delays and reduced throughput.
Frequent node-level media access scheduling, radio transmission and reception. High communication resource management, utilisation and scheduling.	Spectrum constraints.	Increased spectral resource utilisation, high congestion and reduced throughput.
Existence of many heterogeneous M2M devices. Existence of multiple local M2M networks (capillary networks). Coexistence of heterogeneous access network devices.	Hidden terminal.	Incompatible CTS frames, increased hidden terminals and reduced device access.
Massive distributed sensing and actuation. Large amounts of data acquisition. Frequent node-level media access scheduling, radio transmission and reception.	Energy constraints.	Affects collisions resolution functionality, high overheads and reduced throughput.
Enormous on-board data processing and storage. Processing of massive dynamic and heterogeneous M2M traffic pattern arrivals. High communication resource management, utilisation and scheduling. A moderate number of access network devices.	Hardware constraints.	Increased delays and reduced throughput

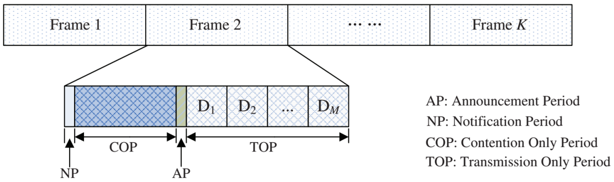


FIGURE 15. Frame structure of the scalable hybrid MAC protocol for massive M2M networks as proposed in [74].

strategies in literature differ based on the strategy employed, the technical objectives, and QoS requirements. The proposed strategies presented are reviewed under the various LS-IoT-specific classification discussed earlier. In some of the approaches presented, some common technologies are adopted and improved or applied by considering improvements. As such, some reviewed works may employ multiple strategies based on common technologies. However, the focus area upon which the need for improvement towards a scalable MAC is derived is used to classify the reviewed work. Table 3 presents the summary of the reviewed strategies under the various classification. The Improvement strategy results that were obtained and comments on the drawback based on LS-IoT networks are provided for each reviewed proposed solution.

A. HYBRID FRAME-BASED APPROACH

1) SCALABLE HYBRID MAC USING OPTIMISATION

In [74], the authors propose a scalable frame-based hybrid MAC scheduling protocol for massive M2M networks. The time frame is divided into four distinct sub-frame periods as depicted in Fig. 15. A Notification Period (NP) and an Announcement Period (AP) is proposed to notify devices of the start of a COP and to announce the transmission schedules in the TOP respectively. Contention in the COP is based on the classical p-persistent CSMA/CA media access method. Successful devices after contention are scheduled for frame transmissions in the TOP based on Time Division Multiple Access (TDMA) approach.

The duration of COP and TOP is a critical factor in ensuring scalability in this scheme because increasing the COP duration reduces the collision probability but reduces the frame transmission time. Likewise, reducing the COP increases the collision probability and allows for more data transmission. The authors address this challenge by developing a hybrid scheduling scheme based on formulating an optimisation problem which sought to maximise the throughput and finding the optimal value of the contention probability and the optimal number of transmissions during COP and TOP respectively.

The results presented in their work shows better throughput, delay, and frame utilisation when compared with s-ALOHA and TDMA based approaches. However, the proposed approach may result in high overhead cost given the fact that the BS needs to transmit synchronisation frames,

TABLE 3. A summary of current proposed solutions for Improving MAC scalability performance in LS-IoT networks.

Classification	Related work	Improvement strategy	Results obtained	Comments and possible drawbacks
Hybrid Frame-Based Approach.	[74]	Scalability using optimisation.	Improved throughput, delay and frame utilisation over s-ALOHA and TDMA.	High control and computation overhead.
	[75]	Continuous Transmission in Multiple Frames.	Best throughput over to NP-CSMA, S-ALOHA and p-ALOHA.	Possible frame congestion.
	[76]	Clock synchronisation and device control.	Reduced delay over s-ALOHA.	High control overheads and device limitation.
	[77]	Scalability using Machine Learning.	Near stable throughput.	High communication and computational overheads. Off-line training required.
	[78]	QoS priority grouping.	Best throughput over CSMA and [19].	No relationship between the number of possible successful devices and the number of available time slots.
Multichannel Hybrid Approach with and without Rendezvous.	[79]	Distributed negotiation using a common control channel.	Near optimal channel utilisation and better channel utilisation than [99].	Possible channel wastage. Estimation phase duration is assumed to be 0. Possible increased in overheads.
	[80]	Orthogonal Codes with Prioritisation.	Better throughput and packet delivery ratio over classical DQCA.	Excessive delays for devices waiting in the queue. Possible false prioritisation of devices.
	[81]	RTS/CTS Overhead Reduction.	Improved saturation throughput and delay over single band approach.	Rapid channel scanning required. The entire frame may be utilised by one device.
	[82]	Channel allocation by device pairs	Improved throughput and delay performance.	Continuous transmit-receive switching. Increased search duration for a large number of channels.
	[83]	Optimisation of logical channel allocation using Network Topology.	Improved throughput over classical hybrid MAC.	Complex computation, channel assignment delays and increased cross-layer overhead
Distributed Queuing Approach.	[84]	Collision resolution using virtual LTE frame and DQ.	Access delay is the same as EAB	Multiple layers of increased delay and complexity.
	[85]	Collision resolution using grouped virtual DQ.	Higher device access rate over ACB method.	Possibility of very long queues.
	[86]	Contention Resolution using DQ.	Better access delay over the classical RA process.	Increased overheads due to counter update frames.
	[87]	DQRAP using downlink feedback.	Improved cell capacity.	Increased length of feedback messages.
Access Barring Approach.	[88]	Two phase ACB with joint optimisation.	Improved transmission success and reduced grant time.	Computationally intensive. Many iterations required for accuracy.
	[89]	Dynamic ACB assisted by MME.	Improved throughput over ACB.	More effective for static devices. More time required for RA procedure.
	[90]	Prioritized random access with dynamic Access Barring	Improved access probability and average delay over EAB scheme.	Possible high utilisation of RBs by HTC devices.
	[91]	Cooperative ACB with adaptive resource management.	Improved access delay.	Possible high utilisation of RBs by HTC devices.
	[92]	Barring-Enhanced DCF.	Improved scalability over DCF.	Lack of granularity of the Back-off factor. Complex computation required.
Group-Synchronised Distributed Coordinated Approach.	[33]	TIM and RAW.	-	No device grouping strategy.
	[94]	RAW Optimisation for 802.11ah.	Better success probability over fixed RAW.	Computational time not accounted for within the beacon period.
	[95]	Uniform Grouping in RAW for 802.11ah.	Better throughput over equal RAW slot allocation.	Full DTIM beacon interval not considered.
	[96]	Spatial Grouping in RAW for 802.11ah.	Better throughput over DCF.	High polling overhead. Requires frequent computation for mobile M2M devices.
	[97]	Group Renewal Access.	Better success probability over DC.F	No constraints on slot duration. Increased waiting time for frozen devices.
	[98]	Real-Time Dynamic Grouping in RAW for 802.11ah.	Better average throughput over ED-CA/DCF.	High memory overhead for historic data. Real-time computational delays.

IDs of the successful devices and the transmission schedules during the AP.

In LS-IoT or massive M2M networks, many devices need to be scheduled for transmission after a contention. Hence, the number of devices' ID and scheduling times that need to be announced may result in large overheads resulting in increased delays and affecting the scalability of the protocol. This issue could be considered in their optimisation and fairness problem as an improvement. Moreover, the computation needed for computing the optimal parameters is complex and requires considerable processing capabilities as well as memory overheads.

2) CONTINUOUS TRANSMISSION IN MULTIPLE FRAMES

Hegazy *et al.* proposed a simple strategy as a scalable solution to accommodate massive M2M devices in [75]. Their approach uses the same frame structure as in Fig. 15. However, the devices use a non-persistent CSMA procedure to contend for the reservation of transmission, and the BS allocates time slots to all the successful devices and shares this information with devices. The difference in this approach is that devices transmit their data during their allocated time slot in each TOP of all the subsequent frames until there is no data left to be transmitted.

The results presented showed improved throughput and delay performance when compared to classical np-CSMA, s-ALOHA and p-ALOHA methods. However, in their proposed strategy, if all the time slots are fully occupied by devices that have large amounts of data to be transmitted, new devices arriving during the next COP may have to wait for a number of frame periods for the previously allocated time slots to be free. This may affect the protocol's ability to scale well in a very dense and saturated LS-IoT network. Also, if the number of occupied slots is high, devices may delay in transmitting subsequent data in the next frame which could affect time-critical IoT applications.

3) ROBUST HYBRID MAC WITH CLOCK SYNCHRONISATION AND DEVICE CONTROL

Another frame-based approach is proposed in [76] with the aim of combating clock synchronisation failures between a large number of M2M devices which affects scalability and robustness. The technique is based on the same principle presented in [74]. However, the authors optimised the TOP by using the IEEE 802.11 DCF mechanism (including handshaking) in each TOP time slot to alleviate communication failures due to lack of clock synchronisation. Their proposed approach also controls the number transmitting devices during a TOP period as the number of nodes increases. The control mechanism used in their proposed approach is not explicitly presented.

The results presented however show that the proposed approach performs better than s-ALOHA and TDMA based approaches in terms of throughput. The authors acknowledged that by introducing the DCF scheme with handshaking during the TOP slot the number control overheads increases which further imposes delays during the TOP slot. The reduction of the number of communication failures at the expense of reduced transmission period may not guarantee scalability in massive M2M or LS-IoT networks. It is also worth mentioning that controlling the number of devices during a TOP based on the increase in the number of devices may improve throughput and delays. However, it may limit the transmission to a few devices which may not be a true reflection of a scalable MAC protocol for massive M2M device access.

4) SCALABLE MAC ASSISTED BY MACHINE LEARNING

Yang *et al.* proposed the use of machine learning to optimise the frame length in a frame-based hybrid approach to achieve a scalable MAC scheduling for IoT networks in [77]. Their proposed strategy is based on predicting the number of IoT devices actively connected to the network and using that information to optimally and dynamically adjust the length of the frame to achieve stable throughput with an increasing number of devices.

Though the authors presented a near stable throughput performance with increasing devices, this approach incurs very large overheads due to the exchange of control frames for performing the prediction. Such overheads may cause an increased delay which may contribute to the delay already imposed by the online estimation process. In addition, the proposed approach requires some off-line training which makes the management of such a network difficult.

5) QoS-AWARE GROUPING

The authors in [78] use a frame-based approach with added QoS awareness to support LS-IoT networks. In their approach, the time frame is divided into a Beacon Period (BP), TDMA reservation Period (TDMA-P), CSMA Contention Period (CSMA-P) and a Burst Period (Relay-P). To provide a scalable solution, devices are grouped according to priority. In the CSMA-P period, devices contend for channel access based on their priority using different back-off time windows. Devices that are unable to secure channel access during the CSMA-P period increase their priority level and wait for the next beacon frame to begin a new contention process. The BP is used for notifying the devices that succeeded in the previous CSMA-P. A device computes the beginning of the time slot to transmit their data during the TDMA-P period based on (5), where $WakeTime_i$ is the wake-up time of the i^{th} device, T_{slot} is the duration of each slot in the TDMA-P period and $N_{slot}[i]$ is the number of slots per device.

$$WakeTime_i = T_{slot} \times \sum_{i=1}^{Prev-1} N_{slot}[i] \tag{5}$$

The results presented in their work shows a relative improvement in throughput and delay performance for an increasing number of devices. However, there is no mechanism to control the number of possible successful devices in the CSMA-P in proportion with the number of available TDMA-P timeslots. Therefore, a device may reserve a channel in the CSMA-P period. However, it may fail to transmit its data during the TDMA-P if all the time slots are occupied by other devices. This may lead to reduced performance in LS-IoT networks where more devices joining the network with high priority may not be able to transmit their frames during the TDMA-P period.

B. MULTICHANNEL HYBRID APPROACH WITH OR WITHOUT RENDEZVOUS

1) DISTRIBUTED NEGOTIATION USING A COMMON CONTROL CHANNEL

An adaptive multichannel approach to MAC scheduling in large-scale M2M networks is proposed by authors in [79]. The proposed method is based on a distributed negotiation of media access using a Common Control Channel (CCC). The authors attempted to address the channel utilisation problem in order to provide a scalable MAC scheduling protocol for massive M2M networks. In their approach, the time frame is divided into an estimation phase, negotiation phase and

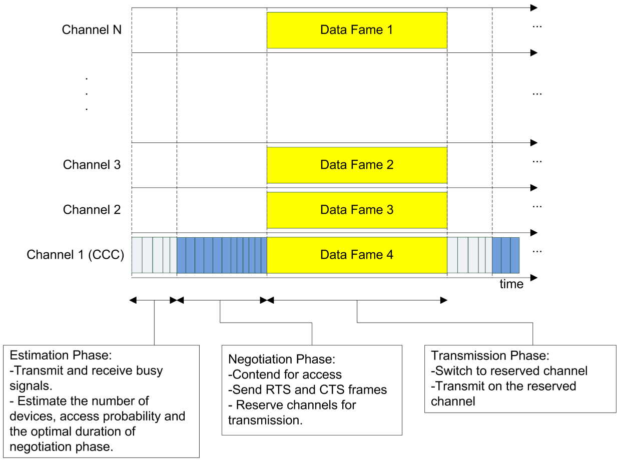


FIGURE 16. **Timing diagram illustration of the proposed MAC scheduling protocol approach in [79].**

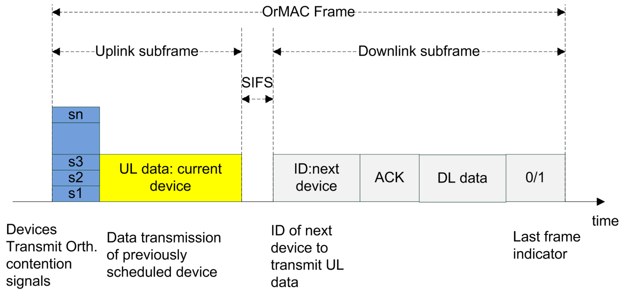


FIGURE 17. **The frame structure of the Orthogonal code MAC scheduling approach for M2M as proposed in [80].**

transmission phase as illustrated in Fig. 16. Devices estimate the number of devices, access probability and optimal duration of the negotiation phase using busy signals transmitted in each time slot of the CCC. In the negotiation phase of the CCC, transmitting devices contend and reserve channels with their intended receiving devices using RTS/CTS frames in each time slot. Devices that have experienced collision try again to negotiate with an access probability which is recomputed using the fact that a device that successfully negotiates need not negotiate again. Successful devices transmit simultaneously during the transmission phase using the reserved channels established during the negotiation phase.

The presented protocol performs well in terms of near-optimal channel utilisation compared to other references from literature. However, if different transmitting devices reserve different channels with a single receiving device, only one channel may be utilised during the transmission phase. This may cause wastage of channels and may result in data frame drops which consequently affects the scalability of the protocol.

The authors also assumed the estimation phase duration to be 0 while analysing the impact of the negotiation phase's duration. The estimation phase is very critical as it serves as an entry point for devices in the proposed approach. Therefore, in LS-IoT networks, the number of busy signals transmitted may be large and the delay in computing the estimated number of devices may increase. Therefore, if the duration for the estimation phase is very short, only a few devices out of the very large number of devices may successfully transmit a busy frame given the sporadic nature of M2M devices. This may deprive some devices access to the transmission resource and may also affect the optimal duration of the negotiation phase as well as the number of devices that can negotiate and transmit data.

The other drawback is the use of extensive signalling information for the estimation phases and negotiation phase. In this approach, signalling overheads may increase for many devices and this may affect the estimation and negotiation phases.

2) ORTHOGONAL CODES WITH PRIORITISATION

A unique approach to MAC scheduling based on orthogonal codes called OrMAC for M2M networks is proposed in [80]. The concept is based on the pre-assignment of orthogonal codes to devices which are used as a media access key, representing orthogonal contention signals. The orthogonality of the codes ensures that two or more codes are not the same. This eliminates collisions during contention. The concept is illustrated in Fig. 17. The time frame is partitioned into an uplink sub-frame and a downlink sub-frame. The devices also indicate the urgency of the data transmission by varying their transmission power. Devices transmit their contention signals and are then put in a queue until the currently scheduled transmission is completed. In the downlink sub-frame, the BS transmits the ID of the next scheduled device in the queue followed by an ACK of the immediate completed transmission and then a bit field to indicate if the completed transmission was the last frame for that device or not.

While the proposed approach performs relatively well in comparison with the classical DQCA, in a large-scale or dense environment, if a scheduled device has a very large amount of data frames to be transmitted, it will cause excessive delays for devices waiting in the queue. Therefore, though the collision probability problem is resolved, the scalability issue may still exist in terms of delays and low throughput for a large number of devices. Also, the use of orthogonal signals for each device limits the ability to accommodate many devices when using a spectrum that is highly constrained.

Moreover, while a device may transmit a high-power contention signal, the separation between the device and the BS, however, may result in the power received being lower than the transmitted power. Therefore, this may create false prioritisation of devices in the MAC protocol scheme proposed.

3) RTS/CTS OVERHEAD REDUCTION

In [81], the authors used a distributed multichannel approach to reduce the overhead caused due to the transmission of RTS and CTS frames in CSMA/CA MAC protocol. In their approach, each device chooses a channel and transmits an

RTS frame on that channel after the back-off timer has expired. This is contrary to the classical approach whereby only one device may transmit an RTS frame in a single channel. In this case, the RTS frame will not collide. The BS receives and decodes all the RTS frames on all the channels. The BS broadcasts a CTS over all the channels which provides information about the devices with a successfully decoded RTS. The devices that had a successfully decoded RTS send their data frames followed by an ACK frame over a single channel based on priority levels indicated also in the CTS frame by the BS. This multichannel approach increases the data transmission period to allow for the scheduling of multiple devices before the back-off period begins.

Through simulation, the proposed strategy shows improved saturation throughput and delay compared to approaches that use a single band approach for transmitting an RTS frame. However, the ability for the BS to rapidly scan through all the channels and decode all the RTS frames is not considered in this approach. This is important because it may affect the response time of the BS which could decrease the data transmission time period and counter the gains of this strategy. Moreover, in a massive M2M network with dynamic traffic patterns, long frame transmission from devices may result in inevitably reserving the transmission period for only one device before the next back-off begins unless a frame fragmentation strategy is employed. This may affect massive device access in LS-IoT networks.

4) CHANNEL ALLOCATION BY DEVICE PAIRS

The authors in [82] deal with the distributed channel allocation strategy between a pair of transmit-receive devices in a multichannel approach to attempt to improve M2M communication in high-density IoT networks. They deal with addressing distributed channel allocation with a rapidly growing number of M2M devices. A device hops from one channel to another based on a randomly defined sequence. In each channel, a device listens for a 'HELLO' signal frame from other devices over a period. When a 'HELLO' frame is not received on that channel, the devices secure that channel and begin transmitting 'HELLO' frames in the secured channel. However, if the device receives a 'HELLO' frame in the secured channel, the device switches to the next channel in a sequence with a probability. A device which needs to transmit to the receiver which has secured a channel performs a search based on its own sequence until it finds the channel of the receiver using the 'HELLO' signal frames. RTS and CTS frames are then used to agree on that channel.

Their proposed strategy shows relatively better throughput and reduced delays. The strategy presented requires continuous transmit-receive switching which means that the algorithm may not provide similar results in practice due to potential switching delays. In addition, the transmitter's search for a receiver may be quite long if there are many channels to scan through. This may increase the delay of a frame that needs to be transmitted. This approach may not be suitable for real-time LS-IoT applications.

5) OPTIMISATION OF LOGICAL CHANNEL ALLOCATION USING NETWORK TOPOLOGY

In [83] the authors use a multichannel approach to achieve improved throughput and accommodate devices in an interference range for M2M networks. In their approach, logical channels are created by having sub-frames and time slots within the sub-frames for each channel. All the logical channels are combined to create multiple super-frames. The logical channels are assigned to devices by the BS. They the information from the network layer to optimise the allocation of the logical channels to devices. The BS abstracts the network's routing topology from the routing table information and optimally allocates all devices with a logical channel.

The performance of the proposed strategy is evaluated in simulation and shows a relatively improved throughput for an increasing number of devices when compared to a hybrid MAC in literature. However, given the fact that the protocol relies on the upper network layer to perform optimal channel assignment, significant time may be wasted during configuration since a lot of cross-layer interactions have to be performed to achieve topology optimisation before channel assignment. Such an approach may be feasible, provided the network's topology does not change frequently so that the previously obtained topology information may still be used without the need for reconfiguration.

C. DISTRIBUTED QUEUING APPROACH

1) COLLISION RESOLUTION USING VIRTUAL LTE FRAME AND DQ

In [84], authors extended the DQRAP technique to improve the RA process in LTE to support many M2M devices for enabling LS-IoT. The aim of the proposed method is to reduce collision and access delays. In their approach, a virtual LTE frame structure is created based on the modification of the actual LTE frame such that an access part is created to implement the DQRAP. The preambles used in LTE for media access are then divided into virtual groups with each group containing a certain number of preambles. The preambles are virtually mapped to the mini-slots of the DQRAP. A device that has data to be transmitted first selects one of the preambles of a selected virtual group. Devices keep track of the number of collisions in the RQ and the position of each device in the RQ based on the status of the detected preamble, whether it is successful, empty or collided. The devices in the RQ are resolved and the position of the device in the RQ is updated. A new device may select a preamble in a different virtual group based on a proposed request transmission rule to reduce the collision probability in the RQ.

The simulated results presented show improvement in the access delay for many devices when compared to the classical EAB procedure in LTE advanced (LTE-A). However, the approach results in multiple layers of increased delay and complexity added to the existing LTE RA process. The devices need to perform the RA process in the first layer before switching to the second layer to perform the DQRAP process within a predefined time frame. The timing constraint imposed by the LTE RA frame structure may not allow this method to scale very well in a very dense M2M scenario giving the amount of signalling and handshaking that needs to be performed. Another drawback is that the collision may still be imminent for LS-IoT networks in the sense that the selected virtual preambles still need to be based on Resource Blocks (RBs) before entering the DQRAP layer. This still limits the resource allocation for a large number of devices.

2) COLLISION RESOLUTION USING GROUPED VIRTUAL DQ

The authors in [85] considered the use of DQ approach with a proposed modification to eliminate the need for employing class barring in 5G and LTE-A networks for M2M devices. Their objective was to deal with massive and bursty M2M device access that may be delayed during the barring process. They proposed multiple virtual queues based on groupings to resolve a collision. The proposed concept estimates the number of collided devices during a random access process. Based on the estimated collided devices, the devices are divided into multiple groups. The groups are then added to a logical access queue and the groups that arrive first at the head of the queue then exits the queue by performing a random access procedure. If the random access contention is unsuccessful the device randomly selects a group to join in the queue. The length of the queue is updated by a counter in the BS and the position of the device in the queue is updated in the device itself. A device that arrives during a random procedure is either suspended until the current contention period has finished or is simply added to the end of the queue.

The results presented for their approach show improved performance when compared with the classical access baring scheme. However, a new device joining the end of the queue coupled with a collided device randomly joining any group in the queue may lead to continuous increased queue length based on individual devices and groups. This may cause delays for other devices in the group. For a massive number of devices that access the network, the likelihood of having a device arrive during the contention period is high, and therefore, applying the rule for newly arriving devices and currently collided devices may further increase the number of devices per group in the queue which leads to a recurring high probability of collisions. This may affect the protocols ability to scale with an increasing number of devices.

3) CONTENTION RESOLUTION USING DQ

In [86], a strategy is proposed based on DQ to provide a scalable MAC strategy for massive M2M devices in LTE. When a collision is detected during an RA process, the collided devices are split into virtual distributed queues using a proposed algorithm before being scheduled for the transmission opportunity. The updates are performed upon receiving information from the BS. To provide massive device access, when a certain number of devices choose the same preamble, an RB is allocated to the group of devices that collided and are immediately put into a CRQ. Internal counters within each device are updated to keep track of the virtual queue length and the device's position in the queue using the RB status information broadcast by the BS.

The devices resolve a collision in the CRQ by selecting a different preamble within the RB that was reserved after the collision. If the new preamble is not in collision, the device transmits its data in the next occurrence of the RB. If a collision is detected, the devices repeat the preamble selection until there is no collision and the device then transmit during the next round of the RB.

The proposed strategy shows improvement in the access delay for a relatively large number of simultaneous message arrivals over the standard RA process in LTE. However, the number of virtual queues is based on the number of preamble collisions per given time. In an extremely large-scale scenario, the queues may rise rapidly. This may lead to more overheads due to the inclusion of sub-headers in response frames transmitted by the BS for updating counters.

4) DQRAP USING DOWNLINK FEEDBACK

In [87], a downlink feedback process is proposed to eliminate the need for two queues in the DQRAP for large-scale M2M communication in 5G technology. Downlink feedback is issued to devices which contains two variables indicating the length of the queue (the number of devices in the RQ) and the status of an access opportunity. The status of an access opportunity is indicated in the feedback frame using a bitmap of length equivalent to the total number of available access opportunities, with a '1' indicating an access collision and a '0' indicating empty access. The devices then use the feedback information to determine their transmission time interval and small packet block. The scheduling of devices in subsequent transmission time intervals may be granted in this approach.

The results obtained for this approach shows better access delay over the classical RA process in LTE. In LS-IoT networks with a massive number of M2M devices and sporadic request for access opportunity, the frequent transmission of feedback and an increased length of the feedback messages may be required due to the length of the bitmap feedback message being proportional to the number of access opportunities. This may critically affect the scalability of the MAC approach proposed.

D. ACCESS BARRING APPROACH

1) TWO PHASE ACB WITH JOINT OPTIMISATION

In [88], the improvement of access success with efficient radio resource utilisation in the classical ACB approach is proposed. The proposed strategy considers executing the standard ACB algorithm in two phases. In the first phase, the ACB process is executed by all devices with pending data to be transmitted to the BS before entering a second phase of the ACB. In the second phase, a set of two different types of preambles are created and devices randomly choose a preamble which automatically groups the devices into two groups. One set of preambles represent successful preambles and the other set is regarded as unsuccessful preambles.

The ACB in the second phase varies the access barring factor based on the number of successful and unsuccessful preambles and optimally allocates RBs. A joint optimisation is used to find the optimal access baring factor and the number of RBs to be allocated for both RACH and Physical Uplink Shared Channel (PUSCH). The PUSCH is used to effectively accommodate other devices that selected a successful and unsuccessful preamble.

The simulation and analytical results of the proposed strategy in [88] shows a high number of possible successful transmissions and reduced grant time when compared with two other references. Though very good results were achieved, the proposed algorithm is computationally intensive and requires several cycles to achieve accurate estimations. Therefore, obtaining similar results in a typical large-scale IoT environment with thousands of devices may require a greater number of cycles which will affect the overall throughput of the network.

2) DYNAMIC ACB ASSISTED BY MME

In [89], an approach for enhancing the barring mechanism for massive M2M access is proposed. The Mobility Management Entity (MME) which has information about the congestion status and access barring factor of all BSs connected to it manages the traffic congestion for each BS. Using the access barring factor, α, which may represent the success probability of a particular BS, the MME balances the M2M traffic load amongst the connected BSs by averaging the i^{th} BS's access barring factor over the sum of barring factors as in (6). In (6), α_i is the newly calculated balanced access barring factor for the i^{th} BS and N is the number of BSs connected to the MME.

$$\alpha_i^* = \frac{\alpha_i}{\sum_{i=1}^{N} \alpha_i} \tag{6}$$

The α_i parameter is periodically updated based on any changes to the traffic load. When the BSs gets the α_i value, the classical access barring procedure depicted in Fig. 9 is executed.

The results obtained in their work shows relative improvement in throughput when compared to the classical ACB which does not change its barring factor based on load balancing. However, the scheme depends on the ability of an M2M device to receive signals from multiple BSs. Therefore, for a massive number of M2M devices that are not mobile and that can only receive signals from one BS, the strategy may not be beneficial. Also, this approach increases the total time required to finish an RA procedure due the data exchange, computation and random selection of an access value that must be completed. Therefore, in a very dense scenario, the average access delay may rise significantly which may affect scalability.

3) PRIORITIZED RANDOM ACCESS WITH DYNAMIC ACCESS BARRING

In [90], an improved EAB scheme is proposed which uses Prioritised Random Access (PRA) and Dynamic Access Barring (DAB) to deal with scalability in terms of massive M2M access in the LTE-A networks. Virtual resources are created and are mapped to a pool of RA opportunities varying across different sub-frames of the LTE-A frame structure as shown in Fig. 18. HTC devices may use all the allocated RA opportunities. Emergency and timed M2M devices may use one virtual slot where the devices may employ a back-off procedure to resolve collisions in the virtual slot. Priority-based (High/Low) M2M devices use the remaining virtual slots and employ a back-off procedure for collision resolution. The virtual resource allocation is based on statistical information about the M2M devices stored in the BS and is occasionally broadcast to the M2M devices.

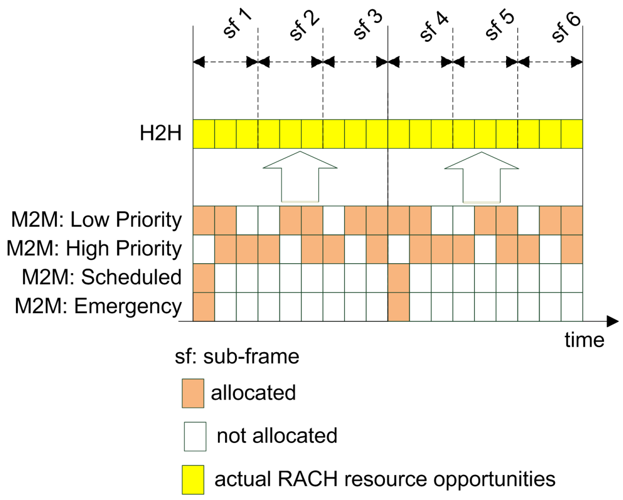


FIGURE 18. **Resource allocation strategy as proposed in [90].**

To further reduce collision for a massive number of devices in the proposed approach, the BS calculates the number of the successfully decoded first message of the device and determines three possible congestion states which are low, medium and high. When a high state is detected, the M2M devices are then notified to defer from transmitting the first message.

The results presented for this approach shows improvement in the access probability and average delay for all classes of devices when compared to the EAB scheme. However, the simulation results do not consider a situation whereby HTC devices use a large portion (90% to 100%) of the RA opportunities. This is important because if all RA opportunities are used by the HTC devices, the M2M devices cannot map their virtual resources to a RACH opportunity, which may result in very low access success probability and large delays leading to packet drops.

4) COOPERATIVE ACB WITH ADAPTIVE RESOURCE MANAGEMENT

The authors in [91] tried to improve the throughput of the ACB/EAB scheme for massive M2M device access over LTE-A network with the key objective of reducing the random access delay of the M2M device by means of an enhanced cooperative approach. In their work, the BS uses the total number of devices attached to it to establish the probability of an M2M device accessing the BS. It uses this information to provide a set of access class barring parameters. The uniqueness of this approach is that BSs are assisted by neighbouring BSs to achieve accurate estimation of the number of devices and the access probability of an M2M device. The M2M devices get to choose the optimal BS based on a unique rule before employing the ACB procedure as in Fig. 9.

To reduce collisions, the authors proposed that M2M devices and HTC devices may use different RBs and preambles. However, the RB allocation is dynamic such that the unused RBs of the HTC devices may be allocated to the M2M devices to enhance the throughput of the massive number of M2M devices.

The results presented in their work shows relative improvement in the access delays for many M2M devices. However, this approach may affect M2M device scalability when the HTC devices occupy all the RBs.

5) BARRING-ENHANCED DCF

The access barring approach is applied in the classical DCF approach of an 802.11 network to support massive access by M2M devices in [92]. In a similar manner proposed in [91], the Barring Enhanced DCF (BE-DCF) proposed in [92] estimates the contention level which is used to establish a barring factor. Given the fact that the DCF operation is distributed, estimating the contention level based on the number of active devices is very complex. However, the authors used the event of the first slot after a Distributed Inter-frame Frame Spacing (DIFS) period to trigger the estimation of the level of contention for adjusting the access barring factor. This strategy is used because the first slot after the DIFS period is accessed by all contending devices. The devices can determine the access barring factor that will be used to perform the access barring procedure. The access barring factor is decreased by half for every collision detected in the first slot. However, if the frame is transmitted successfully, the access barring factor is set to the highest value. A random number is uniformly selected and if the random number is greater than the access barring factor, the device defers its back-off process to the next DIFS. Otherwise, the back-off procedure is executed immediately.

The results of the proposed BE-DCF depict improved performance in terms of scalability when compared to the classical DCF process. However, due to the two levels of granularity of the back-off factor, the improved performance may not extend to an extremely large-scale network scenario. Achieving finer granularity of the barring factor may improve scalability and fairness, however, it may be complex to achieve that given that devices themselves cannot estimate the number of devices in the network.

E. GROUP-SYNCHRONISED DISTRIBUTED COORDINATED APPROACH

1) TIM AND RAW IN 802.11ah STANDARD

One of the standardised applications of the Group-synchronised distributed coordinated approach is in the proposed 802.11ah standard for LS-IoT [33]. In the 802.11ah standard, the RAW is divided into slots within which the classical CSMA/CA is used to schedule transmissions distributively. A Traffic Indication Map (TIM) is used in conjunction with the RAW to provide synchronisation of groups for both Up Link (UL) and Down Link (DL) transmissions. The devices are grouped based on a hierarchical structure made up of Pages, TIM groups, sub-blocks and devices. The hierarchical grouping of the devices is aimed at UL and DL scheduling based on groups of devices within a RAW period to effectively support many devices.

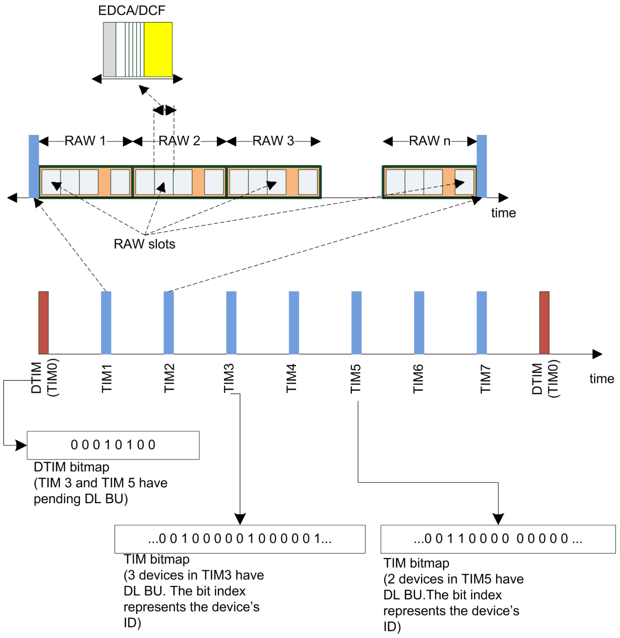


FIGURE 19. The RAW operation as proposed in the 802.11ah standard [33].

As illustrated in Fig. 19, the BS first broadcasts a Delivery TIM (DTIM) beacon to devices at the Target Beacon Transmission Time (TBTT). All devices need to be awake at the start of the TBTT to receive the DTIM beacon. The DTIM beacon announces to devices, which TIM groups have pending DL Buffer Units (BU) in the BS. The devices that belong to that group then go to sleep and only wake up at a Target Short Beacon Transmission Time (TSBTT) for that group where a TIM beacon is broadcasted to specify which stations in that TIM group have pending DL frames. Only the devices that are notified then contend for media access in a RAW slot. The devices use (7) to map their Association Identity (AID) or TIM bitmap, x, to a specific RAW slot, (i_{slot}). N_{RAW_SLOT} is the number of RAW slots and N_{offset} is an offset parameter.

$$i_{slot} = (x + N_{offset}) mod N_{RAW\_SLOT} \tag{7}$$

This approach reduces overhead in the beacon frames and reduces collision probability. However, the strategy for grouping devices is not addressed. Also, in a very dense network, the RAW slot duration may shrink due to many RAWs and RAW slots which may make it difficult to accommodate RTS/CTS signalling overhead. Also, crossing a RAW slot may be allowed in this protocol. However, this causes high collision probability in the slot that has been crossed as emphasised in [93].

2) RAW OPTIMISATION FOR 802.11ah

In [94], Park *et al.* proposed a mechanism for estimating the size of the RAW of the proposed 802.11ah standard in terms of the number of RAW slots with the aim of improving the probability of a successful transmission for massive M2M networks. The BS of the network uses a statistical approximation of the success probability to estimate the number of devices that have buffered data for uplink transmission. The BS finds the optimal size of the RAW in terms of the number of RAW slots by maximising the success probability objective function. This process is computed in the next beacon interval.

The results presented in their work show improved success probability for a large number of devices compared to a pre-assigned fixed RAW value. Though relatively superior results were achieved, there may be an issue regarding the time available to perform the estimation. The computational time is not accounted for within the beacon interval which could affect the duration of the beacon interval.

3) UNIFORM GROUPING IN RAW FOR 802.11ah

In [95], the authors propose a strategy for the allocation of a RAW slot duration based on the size of the group. Larger groups of devices are allocated a relatively short RAW slot duration per RAW slot group and vice versa. The entire RAW frame is divided into sub-frames. Two sets of equally sized groups are created based on a scheme that uniformly distributes the devices amongst the number of available groups. The proposed grouping scheme of a device, n, follows the following order: $n_k, n_{k+K}, n_{k+2K}, \ldots$, where $k = 1, 2, \ldots, K$ is the k^{th} group and K is the total number of groups. Thus, a set of $NmodK$ and $K - (NmodK)$ sized groups are created at any point in time with N being the total number of devices. Each sub-frame is mapped to one of the two sets of groups. Each sub-frame also contains several RAW slots within which devices belonging to the RAW slot performs a DCF procedure. The duration of each RAW slot in both sub-frames are determined by the BS based on the size of the group such that the longer sub-frames have RAW slot duration greater than that of the RAW slots in the shorter sub-frame. Effectively the BS increases the duration of the RAW slot if the collision probability is low for the lesser devices allocated to a RAW slot.

A throughput simulation and analysis on the approach in [95] shows that the proposed strategy improves the throughput when compared to the conventional allocation strategy of allocating equal RAW slot duration for all group sizes. However, this approach only focuses on the throughput of a RAW frame. Therefore, the aggregated throughput across the entire beacon interval with multiple RAW may influence the performance of the strategy given the fact that the RAW slots with larger devices have a shorter duration to perform DCF.

4) SPATIAL GROUPING IN RAW FOR 802.11ah

In [96], the allocation of RAW slots is based on a spatial grouping algorithm. The authors avoid the use of (7) for the RAW slot allocation and propose the provisioning of the slot index based on the geographic location of the device. The coverage area is divided into the same number of RAW slots present in the RAW. The BS demarcates its coverage area based on the furthest detection range of a device which represents the radius of the coverage area. A large-scale fading model is used to estimate this range based on the farthest distance between two stations.

The performance of the proposed strategy is simulated and compared with the conventional DCF process. The simulation results show throughput improvement. However, this approach may incur high delays for LS-IoT networks in the sense that a lot of polling needs to be done for a very large number of devices to estimate the geographic location of devices. Also, for each polled device, the distance must be computed. If a device is mobile, the location must be updated frequently. This consequently affects the throughput performance in LS-IoT.

5) GROUP RENEWAL ACCESS

A grouping strategy is proposed in [97] called the Grouped Renewal Access Process. The grouping of devices based on transmission attempts is proposed in this approach in order to deal with RAW slot crossing challenges in super dense IoT networks. A slot is adopted such that each slot is defined by the number of transmissions within a group of which the end of a slot is the end of the last transmission. In the last transmission, a new back-off value is chosen by the device involved in the last transmission. At the same time, all other devices pause their back-off timer values until their next assigned slot period. During the next slot period, devices with frozen back-off timer values resume their back-off process from the paused value.

The success probability results presented show better success probability when compared to the conventional DCF. However, this strategy does not impose a constraint on the slot duration which means that for devices with long frames in a slot, the slot duration may expand quite significantly which will mean that the frozen devices may have to wait longer before resuming a back-off. This may have an impact on the overall throughput of the network.

6) REAL-TIME DYNAMIC GROUPING IN RAW FOR 802.11ah

The authors in [98] propose a strategy to optimise the RAW parameters dynamically based on traffic conditions in real-time. The BS records the number of successful transmissions and the status of a transmission result for a given device. The BS continuously computes the estimated transmission interval for a station and determines the optimal beacon interval, the number of RAW groups or RAW slots and the duration of the RAW slot. Based on the transmission frequency of the device, which is assumed to be predictable, the BS assigns a device to a RAW slot using the optimised values.

The results obtained shows relative improvement. However, the BS is burdened by the need to store a lot of historic data about each device as well as compute in real-time the optimisation problem. This may affect strict timing requirements needed by the BS to ensure synchronisation of the groups in all the RAWs.

VI. FUTURE DIRECTION FOR MAC SCALABILITY IN LS-IoT NETWORKS

Virtualisation could be employed as a novel strategy to create multiple virtual instances of the MAC protocol and its related functions that are within the LS-IoT GW device and to improve the scalability of the MAC. The authors in [100] indicate how virtualisation of the radio or media access network resources in future mobile networks could deal with the capacity demand. Virtualisation optimises resources utilisation and increases the whole system's performance [101]. However, the use of virtualisation to provide a scalable MAC has not been exploited in wireless network technologies for LS-IoT networks. The benefits of virtualisation which can be extended to LS-IoT networks depends on how the virtualisation framework and algorithms are established.

The virtualisation of the MAC protocol could be employed within the framework of a Network Function Virtualisation (NFV) and Software-Defined Networking (SDN) approach to create, orchestrate and distribute virtual MAC instances over a set of edge data centres with greater processing, memory, and bandwidth resource capabilities. However, the virtualisation of the MAC protocol in the GW of the LS-IoT network requires the consideration of several factors to obtain the best possible solution. These factors include heterogeneous M2M device traffic patterns, resource requirements of frames from M2M devices (e.g., processor, memory, and bandwidth), front-haul network latency, and residual resource capacity in the edge data centre.

The M2M devices and their traffic patterns can be modelled using queuing approaches in order to establish the different MAC resource requirements for each MAC instance. A heuristic algorithm coupled with an edge data centre placement cost model with constraints could be considered in order to find the optimal placement location to host the virtual MAC instances given a number edge data centres with varying processor, memory and network bandwidth capabilities. A residual resource de-fragmentation algorithm could also be considered to improve the acceptance rate of virtual MAC instances in the edge data centre. A front-hauling technology such as the Common Public Radio Interface (CPRI) which guarantees extremely low latency could be considered in the NFV-SDN framework to provide real-time interaction between the virtual MAC instances in the edge data centre and the PHY layer in the GW device.

Considering all the above, MAC protocols, including the legacy protocols, could be implemented to address the scalability issue of the MAC layer in LS-IoT networks. Thus, a stable MAC layer throughput could be achieved with a growing number of connected devices to the GW device.

VII. CONCLUSION

In this paper, the characteristics of LS-IoT networks, the impact of LS-IoT on the network and the MAC, the related issues, current solutions and future direction for MAC improvement in LS-IoT networks were presented and discussed. The fundamental characteristics of LS-IoT networks were divided into the various network domains and further grouped based on the functional and physical architecture of the LS-IoT network. The characterisation of the LS-IoT network highlights the impact of the large-scale nature of the network on the access network domain which is depicted as a bottleneck effect. This is characterised by the coexistence of a relatively few heterogeneous access network devices and resources but with a large number of dynamic media access, resource provisioning and utilisation required by a massive number of devices.

The MAC protocol which is the fundamental function of the access network and M2M local networks component was discussed under a general wireless network classification and LS-IoT specific classification. This paper was able to establish and discuss the relationship between the LS-IoT characteristics, the MAC protocol design issues and the impact on the performance of the network. The design issues related to collisions, control overheads, timing constraints, spectrum constraints, hidden terminal, energy constraints and the hardware constraints are influenced by some of the LS-IoT characteristics, that typically increases delays, increases resource utilisation and reduces the throughput scalability performance of the LS-IoT network.

Some proposed solutions in literature for enhancing MAC protocols for LS-IoT networks were discussed under the different classifications based on the strategy employed or the target approach to be enhanced. Most of the solutions reviewed under the hybrid-frame based approaches presented throughput improvements. However, they require intensive computation and large control signal overheads. The reviewed multichannel approach solutions also presented improvement in throughput and channel utilisation. However, they mostly required rapid channel scanning and transmit-receive switching. The distributed queuing approaches reviewed mostly presented improved performance in device access delays. However, most of the solutions suffer from high control overheads for queue updates and feedbacks. For the reviewed solutions under the access barring approach, improvements on access delays were mostly presented. However, most of the approaches do not provide guaranteed access for M2M devices and they also require M2M devices to be attached to multiple BSs. The group synchronised coordinated solutions generally demonstrated improved throughput, but they are difficult to implement due to time constraints, computational complexity and the high control overheads. The use of a virtualisation approach for enhancing the MAC performance in LS-IoT was suggested as a future direction.

In summary, though there are more than one factors that may affect the scalability of the LS-IoT network, the throughput performance of the MAC protocol contributes significantly to the overall LS-IoT network throughput performance. Therefore, there is a need to address the complexities of enhancing the MAC protocol in LS-IoT networks by using a multi-faceted approach, and state-of-the-art technological advances such as NFV and SDN to expand the scheduling of devices onto a virtual platform as a future direction.

REFERENCES

[1] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A literature review," *J. Comput. Commun.*, vol. 3, no. 5, p. 164, 2015.

[2] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "A comparative study of LPWAN technologies for large-scale IoT deployment," *ICT Express*, vol. 5, no. 1, pp. 1–7, Mar. 2019.

[3] A. Rajandekar and B. Sikdar, "A survey of MAC layer issues and protocols for machine-to-machine communications," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 175–186, Apr. 2015.

[4] T. Adame, A. Bel, B. Bellalta, J. Barcelo, and M. Oliver, "IEEE 802.11AH: The WiFi approach for M2M communications," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 144–152, Dec. 2014.

[5] M. B. Alaya, Y. Banouar, T. Monteil, C. Chassot, and K. Drira, "OM2M: Extensible ETSI-compliant M2M service platform with self-configuration capability," *Procedia Comput. Sci.*, vol. 32, pp. 1079–1086, Jan. 2014.

[6] *Machine-to-Machine Communications (M2M); Functional Architecture*, Eur. Telecommun. Standards Inst., Sophia Antipolis, France, 2013.

[7] T. Macaulay, *RIoT Control: Understanding and Managing Risks and the Internet of Things*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[8] M. Chen, J. Wan, S. Gonzalez, X. Liao, and V. C. M. Leung, "A survey of recent developments in home M2M networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 98–114, 1st Quart., 2014.

[9] M. Alam, R. H. Nielsen, and N. R. Prasad, "The evolution of M2M into IoT," in *Proc. 1st Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Jul. 2013, pp. 112–115.

[10] M. Chen, V. C. M. Leung, R. Hjelsvold, and X. Huang, "Smart and interactive ubiquitous multimedia services," *Comput. Commun.*, vol. 35, no. 15, pp. 1769–1771, Sep. 2012.

[11] J. Wan, M. Chen, F. Xia, L. Di, and K. Zhou, "From machine-to-machine communications towards cyber-physical systems," *Comput. Sci. Inf. Syst.*, vol. 10, no. 3, pp. 1105–1128, 2013.

[12] *Study on Ran Improvements for Machine-Type Communications*, document 3GPP TR 37.868, 2011.

[13] H. K. Patil and T. M. Chen, "Wireless sensor network security: The Internet of Things," in *Computer and Information Security Handbook*. Amsterdam, The Netherlands: Elsevier, 2017, pp. 317–337.

[14] D. Rountree, *Security for Microsoft Windows System Administrators: Introduction to Key Information Security Concepts*. Amsterdam, The Netherlands: Elsevier, 2011.

[15] D. L. Shinder, T. W. Shinder, C. Todd, and L. Hunter, "MCSA/MCSE 70-291: Reviewing TCP/IP basics," in *MCSA/MCSE (Exam 70-291) Study Guide*. Rockland, MA, USA: Syngress, 2003, ch. 1, pp. 1–96. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978193183692050007X

[16] M. Thoppian, S. Venkatesan, R. Prakash, and R. Chandrasekaran, "MAC-layer scheduling in cognitive radio based multi-hop wireless networks," in *Proc. Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, 2006, p. 10.

[17] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[18] P. Suriyachai, U. Roedig, and A. Scott, "A survey of MAC protocols for mission-critical applications in wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 2, pp. 240–264, 2nd Quart., 2012.

[19] Y. Liu, C. Yuen, X. Cao, N. U. Hassan, and J. Chen, "Design of a scalable hybrid MAC protocol for heterogeneous M2M networks," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 99–111, Feb. 2014.

[20] G. Fang and E. Dutkiewicz, "BodyMAC: Energy efficient TDMA-based MAC protocol for wireless body area networks," in *Proc. 9th Int. Symp. Commun. Inf. Technol.*, Sep. 2009, pp. 1455–1459.

[21] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proc. 21st Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 3, Jun. 2002, pp. 1567–1576.

[22] Y. Guan, C.-C. Shen, and J. Yackoski, "MAC scheduling for high throughput underwater acoustic networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2011, pp. 197–202.

[23] M. Doudou, D. Djenouri, N. Badache, and A. Bouabdallah, "Synchronous contention-based MAC protocols for delay-sensitive wireless sensor networks: A review and taxonomy," *J. Netw. Comput. Appl.*, vol. 38, pp. 172–184, Feb. 2014.

[24] X. Yu, P. Navaratnam, and K. Moessner, "Delay model for super-frame based resource reservation in distributed wireless networks," in *Proc. Int. Conf. Mobile Multimedia Commun.* Springer, 2011, pp. 89–104.

[25] J. Mo, H.-S.-W. So, and J. Walrand, "Comparison of multichannel MAC protocols," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 50–65, Jan. 2008.

[26] P. Kyasanur, J. So, C. Chereddi, and N. H. Vaidya, "Multichannel mesh networks: Challenges and protocols," *IEEE Wireless Commun.*, vol. 13, no. 2, pp. 30–36, Apr. 2006.

[27] H.-J. Lei, G. A. O. Chao, Y.-C. Guo, and Z.-Z. Zhang, "Survey of multi-channel MAC protocols for IEEE 802.11-based wireless Mesh networks," *J. China Univ. Posts Telecommun.*, vol. 18, no. 2, pp. 33–44, 2011.

[28] L. Alonso, R. Ferrus, and R. Agusti, "WLAN throughput improvement via distributed queuing MAC," *IEEE Commun. Lett.*, vol. 9, no. 4, pp. 310–312, Apr. 2005.

[29] J. Alonso-Zarate, C. Verikoukis, E. Kartsakli, A. Cateura, and L. Alonso, "A near-optimum cross-layered distributed queuing protocol for wireless LAN," *IEEE Wireless Commun.*, vol. 15, no. 1, pp. 48–55, Feb. 2008.

[30] L. Alonso, R. Agusti, and O. Sallent, "A near-optimum MAC protocol based on the distributed queueing random access protocol (DQRAP) for a CDMA mobile communication system," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, pp. 1701–1718, Sep. 2000.

[31] L. Tian, J. Famaey, and S. Latré, "Evaluation of the IEEE 802.11 ah restricted access window mechanism for dense IoT networks," in *Proc. IEEE 17th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2016, pp. 1–9.

[32] A. Šljivo, D. Kerkhove, L. Tian, J. Famaey, A. Munteanu, I. Moerman, J. Hoebeke, and E. De Poorter, "Performance evaluation of IEEE 802.11 ah networks with high-throughput bidirectional traffic," *Sensors*, vol. 18, no. 2, p. 325, 2018.

[33] *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 Amendment 2: Sub 1 GHz License Exempt Operation*, IEEE Standard 802.11ah-2016 (Amendment to IEEE Std 802.11-2016, as amended by IEEE Std 802.11ai-2016), 2017, pp. 1–594.

[34] A. Marinos and G. Briscoe, "Community cloud computing," in *Proc. IEEE Int. Conf. Cloud Comput.* Berlin, Germany: Springer, 2009, pp. 472–484.

[35] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "Above the clouds: A berkeley view of cloud computing," Dept. Electr. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. UCB/EECS, 2009, p. 2009, vol. 28, no. 13.

[36] A. Schuldt, K. Hribernik, J. D. Gehrke, K.-D. Thoben, and O. Herzog, "Cloud computing for autonomous control in logistics," in *Proc. INFORMATIK Service Science—Neue Perspektiven für die Informatik. Band 1*, 2010, pp. 305–310.

[37] C. Teixeira, R. Azevedo, J. S. Pinto, and T. Batista, "User provided cloud computing," in *Proc. 10th IEEE/ACM Int. Conf. Cluster, Cloud Grid Comput.*, May 2010, pp. 727–732.

[38] G. Kirby, A. Dearle, A. Macdonald, and A. Fernandes, "An approach to ad hoc cloud computing," 2010, *arXiv:1002.4738*. [Online]. Available: http://arxiv.org/abs/1002.4738

[39] X. Wu, J. Liu, and G. Chen, "Analysis of bottleneck delay and throughput in wireless mesh networks," in *Proc. IEEE Int. Conf. Mobile Ad Hoc Sensor Sysetems*, Oct. 2006, pp. 765–770.

[40] E. S. Hashem, "Analysis of random drop for gateway congestion control," Massachusetts Inst. Technol. Cambridge Lab. Comput. Sci., Cambridge, MA, USA, Tech. Rep. MIT/LCS/TR-465, 1989.

[41] S. M. Das, H. Pucha, and Y. C. Hu, "Mitigating the gateway bottleneck via transparent cooperative caching in wireless mesh networks," *Ad Hoc Netw.*, vol. 5, no. 6, pp. 680–703, Aug. 2007.

[42] A. Alexiou and A. Gotsis, "Packet scheduling strategies for machine-to-machine (M2M) communications over long-term evolution (LTE) cellular networks," in *Machine-to-Machine (M2M) Communications*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 173–186.

[43] J. Kim, J. Lee, J. Kim, and J. Yun, "M2M service platforms: Survey, issues, and enabling technologies," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 61–76, 1st Quart., 2014.

[44] Y. Kwon, Y. Fang, and H. Latchman, "A novel MAC protocol with fast collision resolution for wireless LANs," in *Proc. 22nd Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2, Mar. 2003, pp. 853–862.

[45] H. Ma, X. Li, H. Li, P. Zhang, S. Luo, and C. Yuan, "Dynamic optimization of IEEE 802.11 CSMA/CA based on the number of competing stations," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Jun. 2004, pp. 191–195.

[46] H. L. Vu and T. Sakurai, "Collision probability in saturated IEEE 802.11 networks," in *Proc. Austral. Telecommun. Netw. Appl. Conf.*, 2006.

[47] Y. C. Tay, K. Jamieson, and H. Balakrishnan, "Collision-minimizing CSMA and its applications to wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1048–1057, Aug. 2004.

[48] P. K. Sahoo and J.-P. Sheu, "Modeling IEEE 802.15. 4 based wireless sensor network with packet retry limits," in *Proc. 5th ACM Symp. Perform. Eval. Wireless Ad Hoc, Sensor, Ubiquitous Netw.*, 2008, pp. 63–70.

[49] P. Chatzimisios, V. Vitsas, and A. C. Boucouvalas, "Throughput and delay analysis of IEEE 802.11 protocol," in *Proc. 3rd IEEE Int. Workshop Syst.-Chip Real-Time Appl.*, 2002, pp. 168–174.

[50] G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in *Proc. 22nd Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2, Mar. 2003, pp. 844–852.

[51] T. Van Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proc. 1st Int. Conf. Embedded networked sensor Syst.*, 2003, pp. 171–180.

[52] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[53] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 184–192, Jul. 2012.

[54] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digit. Commun. Netw.*, vol. 1, no. 1, pp. 1–19, Feb. 2015.

[55] A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, "Enhanced LTE-advanced random-access mechanism for massive machine-to-machine (M2M) communications," in *Proc. 27th World Wireless Res. Forum (WWRF) Meeting*, 2011, pp. 1–5.

[56] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of IEEE 802.11 MAC protocols in wireless LANs," *Wireless Commun. Mobile Comput.*, vol. 4, no. 8, pp. 917–931, Dec. 2004.

[57] *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 Standard 802.11, 2016.

[58] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.

[59] *Study on Latency Reduction Techniques for LTE (Release 14)*, document TR 36.881, 3GPP, 2016, pp. 1–300.

[60] N. A. Vasanthi and S. Annadurai, "AWS: Asynchronous wakeup schedule to minimize latency in wireless sensor networks," in *Proc. IEEE Int. Conf. Sensor Netw., Ubiquitous, Trustworthy Comput. (SUTC)*, vol. 1, Jun. 2006, pp. 1–7.

[61] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE INFOCOM*, vol. 2, Mar. 2004, pp. 1404–1413.

[62] S. Mumtaz, A. Alsohaily, Z. Pang, A. Rayes, K. F. Tsang, and J. Rodriguez, "Massive Internet of Things for industrial applications: Addressing wireless IoT connectivity challenges and ecosystem fragmentation," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 28–33, Mar. 2017.

[63] X. Jiang, Z. Pang, R. N. Jansson, F. Pan, and C. Fischione, "Fundamental constraints for time-slotted MAC design in wireless high performance: The realistic perspective of timing," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2018, pp. 4135–4140.

[64] R. Han, Y. Gao, C. Wu, and D. Lu, "An effective multi-objective optimization algorithm for spectrum allocations in the cognitive-radio-based Internet of Things," *IEEE Access*, vol. 6, pp. 12858–12867, 2018.

[65] K. Bian, J.-M. Park, L. Chen, and X. Li, "Addressing the hidden terminal problem for heterogeneous coexistence between TDM and CSMA networks in white space," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4450–4463, Nov. 2014.

[66] L. Du and L. Chen, "Receiver initiated network allocation vector clearing method in WLANs," in *Proc. Asia–Pacific Conf. Commun.*, 2005, pp. 615–619.

[67] F. De Rango, A. Perrotta, and S. Ombres, "A energy evaluation of E-TDMA vs IEEE 802.11 in wireless ad hoc networks," in *Proc. Int. Symp. Perform. Eval. Comput. Telecommun. Syst. (SPECTS)*, Jul. 2010, pp. 273–279.

[68] J. W. Raymond, T. O. Olwal, and A. M. Kurien, "Cooperative communications in machine to machine (M2M): Solutions, challenges and future work," *IEEE Access*, vol. 6, pp. 9750–9766, 2018.

[69] R. W. Juma, A. M. Kurien, and T. O. Olwal, "Energy-efficient coalition games with incentives in machine-to-machine communications," *J. Comput. Netw. Commun.*, vol. 2019, pp. 1–16, Jun. 2019.

[70] G. Nychis, T. Hottelier, Z. Yang, S. Seshan, and P. Steenkiste, "Enabling MAC protocol implementations on software-defined radios," in *Proc. NSDI*, vol. 9, 2009, pp. 91–105.

[71] G. Panic, D. Dietterle, Z. Stamenkovic, and K. Tittelbach-Helmrich, "A system-on-chip implementation of the IEEE 802.11 a MAC layer," in *Proc. Eur. Symp. Digit. Syst. Design*, 2003, pp. 319–324.

[72] T. Schmid, O. Sekkat, and M. B. Srivastava, "An experimental study of network performance impact of increased latency in software defined radios," in *Proc. 2nd ACM Int. Workshop Wireless Netw. Testbeds, Exp. Eval. Characterization (WinTECH)*, 2007, pp. 59–66.

[73] K. Langendoen, "Medium access control in wireless sensor networks," in *Medium Access Control in Wireless Networks*, vol. 2. Hauppauge, New York, NY, USA: Nova Science Publishers, 2008, pp. 535–560.

[74] Y. Liu, C. Yuen, J. Chen, and X. Cao, "A scalable hybrid MAC protocol for massive M2M networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 250–255.

[75] E. Hegazy, W. Saad, M. Shokair, S. El halafawy "Proposed MAC protocol for M2M networks," *Int. J. Comput. Digit. Syst.*, vol. 5, no. 4, pp. 357–363, Jul. 2016.

[76] P. K. Verma, R. Tripathi, and K. Naik, "A robust hybrid-MAC protocol for M2M communications," in *Proc. Int. Conf. Comput. Commun. Technol. (ICCCT)*, Sep. 2014, pp. 267–271.

[77] B. Yang, X. Cao, and L. Qian, "A scalable MAC framework for Internet of Things assisted by machine learning," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5.

[78] A. Kalita, N. Ahmed, H. Rahman, and M. I. Hussain, "A QoS-aware MAC protocol for large-scale networks in Internet of Things," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2017, pp. 1–6.

[79] C.-Y. Hsu, C.-H. Yen, and C.-T. Chou, "An adaptive multichannel protocol for large-scale machine-to-machine (M2M) networks," in *Proc. 9th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jul. 2013, pp. 1223–1228.

[80] E. Shitiri, I.-S. Park, and H.-S. Cho, "OrMAC: A hybrid MAC protocol using orthogonal codes for channel access in M2M networks," *Sensors*, vol. 17, no. 9, p. 2138, 2017.

[81] B. Mawlawi, J.-B. Doré, N. Lebedev, and J.-M. Gorce, "CSMA/CA with RTS-CTS overhead reduction for M2M communication," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Mar. 2015, pp. 119–124.

[82] H. Zhao, Q. Dong, L. Chen, Y. Liu, and H. Hua, "Offloading channel allocation to terminals in high-density IoTs for next generation networks," in *Proc. Int. Conf. Commun. (COMM)*, 2018, pp. 265–270.

[83] E.-J. Kim, S.-P. Heo, H.-J. Chong, and H.-W. Jung, "Integrated hybrid MAC and topology control scheme for M2M area networks," in *Proc. 14th Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, 2012, pp. 1–5.
[84] A. Samir, M. M. Elmesalawy, A. S. Ali, and I. Ali, "An improved LTE RACH protocol for M2M applications," *Mobile Inf. Syst.*, vol. 2016, Jul. 2016, Art. no. 3758507.
[85] A.-T. H. Bui, C. T. Nguyen, T. C. Thang, and A. T. Pham, "Free access distributed queue protocol for massive cellular-based M2M communications with bursty traffic," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5.
[86] A. Laya, L. Alonso, and J. Alonso-Zarate, "Contention resolution queues for massive machine type communications in LTE," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug. 2015, pp. 2314–2318.
[87] S. Saur, A. Weber, and G. Schreiber, "Radio access protocols and preamble design for machine type communications in 5G," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, 2015, pp. 8–12.
[88] Y. Wu, N. Zhang, and G. Kang, "A new hybrid protocol for random access and data transmission based on two-phase ACB mechanisms for M2M communications," *Mobile Inf. Syst.*, vol. 2017, Apr. 2017, Art. no. 1567494.
[89] L. Ferdouse and A. Anpalagan, "A dynamic access class barring scheme to balance massive access requests among base stations over the cellular M2M networks," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug./Sep. 2015, pp. 1283–1288.
[90] J.-P. Cheng, C.-H. Lee, and T.-M. Lin, "Prioritized random access with dynamic access barring for RAN overload in 3GPP LTE-A networks," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Dec. 2011, pp. 368–372.
[91] Y.-H. Hsu, K. Wang, and Y.-C. Tseng, "Enhanced cooperative access class barring and traffic adaptive radio resource management for M2M communications over LTE-A," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Oct./Nov. 2013, pp. 1–6.
[92] L. Zhong, Y. Shoji, K. Nakauchi, and S. Eum, "BE-DCF: Barring-enhanced distributed coordination function for machine type communications in IEEE 802.11 networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Jun. 2014, pp. 467–471.
[93] L. Zheng, M. Ni, L. Cai, J. Pan, C. Ghosh, and K. Doppler, "Performance analysis of group-synchronized DCF for dense IEEE 802.11 networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6180–6192, Jul. 2014.
[94] C. W. Park, D. Hwang, and T.-J. Lee, "Enhancement of IEEE 802.11 ah MAC for M2M communications," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1151–1154, Jul. 2014.
[95] N. Nawaz, M. Hafeez, S. A. R. Zaidi, D. C. McLernon, and M. Ghogho, "Throughput enhancement of restricted access window for uniform grouping scheme in IEEE 802.11 ah," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[96] M. Dong, Z. Wu, X. Gao, and H. Zhao, "An efficient spatial group restricted access window scheme for IEEE 802.11 ah networks," in *Proc. 6th Int. Conf. Inf. Sci. Technol. (ICIST)*, May 2016, pp. 168–173.
[97] Y. Kim, G. Hwang, J. Um, S. Yoo, H. Jung, and S. Park, "Throughput performance optimization of super dense wireless networks with the renewal access protocol," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3440–3452, May 2016.
[98] L. Tian, E. Khorov, S. Latré, and J. Famaey, "Real-time station grouping under dynamic traffic for IEEE 802.11 ah," *Sensors*, vol. 17, no. 7, p. 1559, 2017.
[99] J. So and N. H. Vaidya, "Multi-channel MAC for ad hoc networks: handling multi-channel hidden terminals using a single transceiver," in *Proc. 5th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2004, pp. 222–233.
[100] X. Costa-Pérez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 27–35, Jul. 2013.
[101] S. S. Zaghloul, "The mutual effect of virtualization and parallelism in a cloud environment," in *Proc. Africon*, Sep. 2013, pp. 1–5.

PROSPER Z. SOTENGA (Member, IEEE) received the M.Tech. degree in electrical engineering from the Tshwane University of Technology (TUT), Pretoria, South Africa, in 2017, and the M.Sc. degree in electrical and electronic systems engineering from ESIEE, Paris, France, in 2019, in collaboration with the French South African Institute of Technology (F'SATI). He is currently pursuing the D.Eng. degree in electrical engineering with TUT and the Ph.D. degree in computer science (on the co-tutelle program) with Université Paris-Est Créteil (UPEC), Créteil, France.

He is a Lecturer with TUT. His research areas include next-generation wireless protocols for the IoT and M2M communication, wireless sensor network design and algorithms for context-aware based applications, ubiquitous and pervasive computing, and embedded systems.

KARIM DJOUANI (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Université Paris-Est Créteil, Créteil, France, in 1994.

He was a Manager of the National and European Projects at the UPEC LISSI Laboratory until 2008. He is currently a Full Professor with the French South African Institute of Technology (F'SATI), Tshwane University of Technology, Pretoria, South Africa, and with UPEC, Creteil, France. He also acts as an Expert for different public and private institutions (ANR- France, NRF- South Africa, UNU, EU, Algerian Ministry of Research and Education, and major companies in telecommunication). He has authored/coauthored about 200 articles in archival journals and conference proceedings, 18 chapters in edited books, and two books. His research interests include the development of novel and highly efficient algorithms for reasoning systems with uncertainty as well as optimization for distributed systems, networked control systems, wireless ad-hoc networks, wireless and mobile communication, wireless sensors networks, and robotics.

Dr. Djouani is a member of the IEEE Communication, Computer, Robotics and Automation, and Artificial Intelligence Societies, the European Centre of Excellence in Complexity, and several National Research Task Groups (GDR-MACS and GDR-ISIS).

ANISH M. KURIEN (Member, IEEE) received the D.Tech. degree in electrical engineering from the Tshwane University of Technology (TUT), Pretoria, South Africa, in 2012, and the Ph.D. degree in computer science (through co-tutelle) from the University of Paris-Est, Champs-sur-Marne, France, in 2012.

He is currently the Node Director of the French South African Institute of Technology (F'SATI), TUT, and is responsible for postgraduate programs in the Department of Electrical Engineering. He is also involved in research projects in wireless communications, radio resource management, and mobile network optimization, and in several industrial projects related to wireless networks and technology development. His research interests include feature extraction and pattern recognition algorithms applied to mobile network subscriber classification.

• • •