

How Far Are Wireless Networks from Being Truly Deterministic?

Yan Li, Yunquan Dong, *Member, IEEE*, Pingyi Fan *Senior Member, IEEE* and Khaled Ben Letaief, *Fellow, IEEE*

Abstract—With the rapid development of Internet-of-Things (IoT) technology and machine-type communications, various emerging applications appear in industrial productions and our daily lives. Among these, applications like industrial sensing and controlling, remote surgery, and automatic driving require an extremely low latency and a very small jitter. Delivering information deterministically has become one of the the biggest challenges for modern wire-line and wireless communications. In this paper, we present a review of currently available wire-line deterministic networks and discuss the main challenges to build wireless deterministic networks. We also discuss and propose several potential techniques enabling wireless networks to provide deterministic communications. By elaborating the coding/modulation schemes of the physical layer and managing the channel-access/packet-scheduling at the media access control (MAC) layer, it is believed that wireless deterministic communications can be realized in the near future.

Index Terms—Internet of Things, deterministic communications, on-time, wireless networks.

I. INTRODUCTION

WITH the rapid development of the Internet of Things (IoT) and Industry 4.0, an increasingly large number of smart devices are connected through the internet. This has spawned a lot of time-critical applications, such as the smart grid, the remote control and the factory automation [1]. Most of these time-critical applications require the smart devices to complete information deliveries with deterministic delays, which is critical to the system performance but difficult to achieve in practice. Note that traditional wire-line and wireless networks (e.g., Ethernet) only provide best-effort transmissions, in which the end-to-end latency may be as high as 50 ~ 500 ms and the packet loss rate may be up to 5% ~ 50% [2]. In order to support information transmissions with low latency, low jitter, and high reliability, therefore, many deterministic networking technologies have been developed [2].

The deterministic property of networks can be interpreted in the following two ways. First, it is widely recognized that in a deterministic network, the packets should be delivered with *small delays* (e.g., end-to-end delay being smaller than 10 ms) and *near-zero jitters* (approximately several milliseconds) [2].

Y. Li and Y. Dong are with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: {yanli, yunquandong}@nuist.edu.cn).

P. Fan is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: fpy@tsinghua.edu.cn).

Khaled B. Letaief is with the Department of Electrical and Computer Engineering, HKUST, Clear Water Bay, Kowloon, Hong Kong (e-mail: eekhaled@ust.hk).

In a traditional soft real-time system, the packets are required to be received before a deadline with few violations. In a hard real-time system, the packets must be received before the deadline without any violations. In contrast, a deterministic network tries to complete each packet delivery at its scheduled deadline with near-zero divergence. Second, there are also deterministic networks where the packets are expected to be received *exactly at some planned epochs*, i.e., be received on-time [3]. In a nutshell, these deterministic networks concentrate on the *predictability* more than the immediacy of transmissions. Thanks to the ability to provide bounded end-to-end latency and jitter, the deterministic networks have a wide range of applications in various areas of industrial automation. A few examples include:

- *real-time monitoring systems* like programmable logic controller (PLC) controlled systems and unmanned logistics systems have a strong demand on the real-time status of the operational processes and equipment status within the monitoring field. Specifically, thousands of field sensors transmit their sensing information on the equipment status and task execution progress to the control center, so that it can control actuators, initiate new control tasks, arrange maintenance or trigger alarms in an automatic or human intervention manner. To guarantee the timeliness and reliability of the received data, a 1~10 ms deterministic delay is required for the data exchanging among the sensors, actuators, and the control center [4].
- *in-vehicle networks* should provide real-time and reliability assurance for the transmission of braking/steering commands and advanced driver assistance systems (ADAS) type data for vehicles. Thus, a series of deterministic networking mechanisms (e.g., the controller area network-bus (CANbus), automotive Ethernet and audio video bridging (AVB) system) were developed to meet the strict requirement on transmission delay and delay jitter [5].
- *precise space-time positioning system*, in which the satellites provide precise location and timing for ground users with errors less than 1 cm and 1 μ s, is a key enabler of real-time digital twins and meta-verse [6]. Note that as a bridge between the physical world and a digital world, the virtual representations of the objects, events, and actions are expected to follow the same spatial and temporal statistics as their physical counterpart. Thus, the positioning for the physical world is meaningful only if the transmission of corresponding timing information is precise and deterministic.

Although deterministic networks have been widely used, most practical applications are based on wire-line networking technologies (e.g., the Ethernet, Profinet, EtherCAT, Powerlink) other than wireless ones. This is because the randomness in wireless channels has prevented the wireless technologies such as Bluetooth (IEEE 802.15.1), ZigBee (IEEE 802.15.4), Wi-Fi (IEEE 802.11) and cellular networks (3GPP) from providing the required low latency and high reliability. For example, super-critical industrial control applications require communications networks to provide 100 μ s cycle times for dozens of nodes with extremely limited jitter [7]. While real-time Ethernet can meet this requirement, the most advanced wireless solutions are hardly able to provide a cycle time on the order of 1 ms. However, wireless networks are much cheaper to build and more flexible to changes. In this paper, therefore, we shall survey the existing wire-line solutions of deterministic networks, discuss the main challenges of replacing wire-lines with wireless channels, and propose several possible techniques enhancing the determinacy of wireless communications.

This rest of the paper is organized as follows. In Section II, we discuss some existing solutions to deterministic communications. In Section III, we analyze the difficulties and challenges in providing deterministic communications through wireless based techniques. In Section IV, we propose several physical layer and medium access control (MAC) layer solutions to ensure deterministic transmissions over wireless networks. Finally, we conclude the paper in Section V.

II. EXISTING DETERMINISTIC NETWORKING TECHNOLOGIES

The objective of deterministic networks is to provide deterministic latency, low jitter, low packet loss, and high reliability for information delivery. Although the wire-line networks have relatively stable transmission capabilities, the corresponding end-to-end delays also depend on the burstiness of the source traffic. For example, many bursts of data may arrive at a transmitter within a short period even if the overall traffic load is low. In this case, the output link would be congested, resulting in large latencies and a high packet loss rate. That is, traditional Ethernet networks cannot provide quality of service (QoS) guarantees for real-time and deterministic applications without using some advanced high-layer scheduling. To this end, several deterministic networking technologies have been developed, including the time-sensitive networking (TSN) [8], the deterministic networking (DetNet) [9], and the deterministic Wi-Fi (Det-WiFi) [10].

Specifically, TSN enhances the deterministic assurance of networks mainly through MAC layer techniques, such as the precise network synchronization (IEEE 802.1 AS), shaping mixed traffic of different priorities, the flow reservation mechanism (IEEE 802.1 Qcc), and time-sensitive traffic controlling mechanisms (IEEE 802.1 Qch, cyclic queuing) [8]. By eliminating micro-bursts and providing deterministic QoS guarantees, TSN has become the basis of time sensitive applications like industrial automation and automotive driving.

The Internet Engineering Task Force (IETF) proposed a DetNet system to provide ultra-low end-to-end latency and high

reliability services through an optimized network layer routing scheme. Particularly, DetNet enables the co-networking of those real-time traffic with some less time-sensitive traffic flows. Key features of DetNet include explicit routing, multi-path routing, resource reservation, jitter reduction, flow replication/merging, packet sequencing, and so on [9]. However, most of the proposed DetNet standards draw their basis on TSN and suffer from difficulties in strict time synchronization, long transmission delays, and high complexity of traffic scheduling. Thus, practical implementations of DetNets need further investigations and improvements.

The IEEE 802.11be task group (TGbe) has developed a series of protocols to provide deterministic services over wireless access networks, focusing on PHY and MAC layer techniques and amendments [11]. In particular, 802.11be has integrated multiple solutions, such as the access to the medium, scheduling/preemption of packets, and the coordination among multiple access points. In this way, 802.11be is expected to be effective in reducing the worst-case end-to-end delay of wireless local area networks (WLANs) at a peak throughput of 30 Gbps. Thus, 802.11be is considered as the successor of the IEEE 802.11ax and a key element of next generation Wi-Fi, i.e., Wi-Fi 7 [12]. There were also works running software TDMA MAC over the high-speed 801.11 physical layer, which is named as Det-WiFi [10]. As shown in the paper, Det-WiFi support high-speed applications and provide better deterministic services in practical multi-hop industrial environments.

III. MAIN CHALLENGES IN MOVING FROM WIRE-LINE TO WIRELESS DETERMINISTIC NETWORKING

To meet the required ultra-low latency and ultra-high reliability in time-critical applications, the vast majority of communications technologies in industrial IoT are wire-line based. However, the wire-line networks are less flexible and more expensive to build, as well as to maintain for long-term reliability. Therefore, wireless networking is expected to be used to improve the flexibility of industrial productions. Nevertheless, wireless communications are inherently less reliable and less efficient than wire-line communications for the following reasons.

- Susceptibility to external interference and multi-path fading. Due to the shared nature of wireless channels, each transmitted packet suffers from the interference from other users and adjacent frequency bands. Thus, wireless channels are prone to errors, which further leads to packet loss. In addition, the frequency selective property and the multi-path propagation property reduce the reliability of wireless transmissions significantly. This further increases the corresponding end-to-end delays and jitters.
- The half-duplex constraint limits efficiency. For wireless transmissions, most of the nodes are not allowed to transmit and receive simultaneously in the same frequency band, since the transmitted signals will cause strong self-interference to the received signals at the same node. Under the half-duplex constraint, the efficiency of the wireless channels are quite limited, and thus increasing

the latency of transmissions. Although some preliminary results on full-duplexing have been available in wireless systems, the cost of full-duplex devices is relatively high and the size of full-duplex devices is large. Thus, the full-duplex technology is not mature for industrial environments at present [13], [14].

- **Limited packet length.** In industrial Internet of Things (IIoT) networks, the packets exchanged among smart devices are usually short, with a payload size of a few bytes. This brings great difficulty for the channel coding and frame design for wireless transmissions. Specifically, the control overhead may have a greater impact on the delay than the payload itself. The transmission rate of short packets is also very limited due to the short coding block length.

Moreover, the transmitted packets are easily eavesdropped during transmissions due to the broadcasting nature of wireless channels. The scarcity in transmit power and frequency bandwidth also bring great challenges to the protocol design for wireless networks.

IV. ENABLING DETERMINACY BY WIRELESS NETWORKS

Due to the fading property and the attenuation property, the information transmissions over wireless channels are unreliable and rate-varying, i.e., indeterministic. For wireless channels, the key solution to provide deterministic transmissions lies in compensating the randomness of channels and shaping the source traffic flows according to the time-varying transmission capacity of the channels. For example, we can increase the reliability of each link, shape the arriving traffic, and delay (or even eliminate) some flows. Fig. 1 presents some MAC/PHY layer techniques that have been or can be used in networks/wireless networking standards. These techniques can provide deterministic transmission for data by reducing latency, improving reliability, and exploiting diversity gain and bandwidth utilization. In this section, we focus on these techniques and review some potential techniques for wireless deterministic networking.

A. MAC Layer Techniques

A major task of MAC layer protocols are to coordinate the access of nodes to the common communications channel and prevent conflicts among them. Due to the broadcasting nature of wireless channels, if two nodes are located in the same area and transmit at the same time in the same frequency band, their signals will be added up and conflict each other at the receiver, which prevents the correct decoding of packets and increases transmission latency. Thus, it is crucial for the MAC layer to adequately manage the conflicts among nodes and transmit real-time traffic with higher priority.

1) **TDMA:** In time division multiple access (TDMA), each node access the channel in a dedicated time slot of each frame while all other nodes remain silent, and thus providing the nodes a conflict-free MAC. As a well-established method for providing deterministic communications, TDMA plays a key role in reducing channel access latency and has been applied in many wireless systems [15]. The drawback of TDMA is

	Technology	Networking Standards
MAC layer	TDMA	5G URLLC, Det-WiFi, TSN
	OFDMA	802.11ax, 802.11be
	NOMA	5G URLLC
	ARQ	802.11ax
	HARQ	802.11be
	TSCH	802.15.4
	Bandwidth allocation	TSN, DetNet
	Protection mode (“1+1” and “1:1”)	Flexible Ethernet (FlexE)
	Dynamic priority assignment	TSN, DetNet
	IEEE 802.1 functionality	TSN, DetNet, Det-WiFi, WTSN
PHY layer	mmWave communications	5G URLLC
	(massive) MIMO	5G URLLC, 802.11ax, 802.11be
	LDPC, polar codes	802.11ax/802.11be, 5G URLLC
	Analog fountain codes	TBD

Fig. 1. Technologies that have been or can be used in time sensitive wire-line/wireless networking standards.

that its implementation requires the precise synchronizations among all the nodes, which imposes additional overhead to the network. Besides, the frame length increases linearly with the number of nodes in the network, which introduces large delays in large-scale networks. Also, the TDMA based networks are not very flexible to the changes in nodes, i.e., joining the network or leaving the network. Thus, it is difficult for TDMA to be used directly in scalable networks.

2) **OFDMA:** Before the advent of Wi-Fi 6 (802.11ax), WLAN has been using the orthogonal frequency division multiplexing (OFDM) based modulations. In OFDM, the frequency bandwidth is divided into several orthogonal sub-channels, so that a high-speed data traffic flow is separated into a set of parallel low-speed sub-data traffic flows (i.e., sub-flows). By mapping the sub-flows to sub-channels, the frequency selective fading suffered by single-carrier transmissions could be avoided. However, only one flow can be transmitted over each sub-channel, which reduces the spectrum efficiency. To address this issue, 802.11ax proposes to use orthogonal frequency division multiple access (OFDMA), as shown in Fig. 2. In OFDMA, the frequency and time resources are divided into time-frequency resource units (RU) of fixed-size, so that the data of different users are carried on RU resource blocks other than sub-channels. Therefore, OFDMA enables higher flexibility in resource allocation and improves reliability while reducing queuing delay. It is also noted that OFDMA is more suitable for downlink communications other than uplink transmissions.

3) **NOMA:** Non-orthogonal multiple access (NOMA) use the power multiplexing technique which allow multiple users to transmit in the same frequency band simultaneously. By using the successive interference cancellation (SIC) method at the receiver side, the receiver decodes the data of different users successively. Once the data of a user is decoded, its interference to remaining users would be subtracted from the received signal, so that the signal-to-noise ratio of the

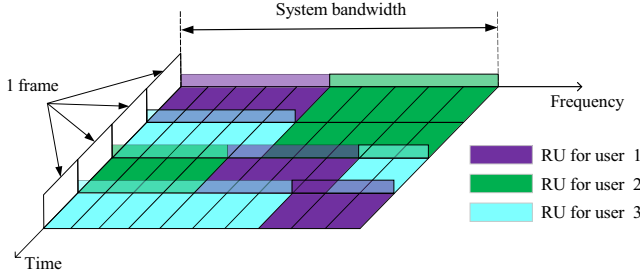


Fig. 2. Resource structure in OFDMA.

remaining signals can be increased. It has been proved that the orthogonal multiple access (OMA) techniques (e.g. TDMA, OFDMA) performs slightly better than NOMA in terms of latency when the number of users is small, while NOMA performs better than OMA when the number of users is large [20]. This is because when the number of users is small, each user can be allocated with a separated and orthogonal piece of resource so that spectral efficiency is higher when transmitting through OMA; when the number of devices is large, there would be not enough resources to maintain the orthogonality among users. This leads to many unavoidable collisions and large latencies. On the contrary, NOMA allow a large number of devices to share the same radio resources and provide them with the desired latency. Another advantage of NOMA is that it does not require a separate authorization and random access phase, as devices can send data at any time. This is the scenario of interest for most IoT use cases, and for URLLC scenarios where a large number of users joins and leaves the network frequently (e.g., intelligent traffic systems and logistics monitoring systems). In fact, NOMA has been selected as a key technique for 5G URLLC use cases. However, the implementation of NOMA technology still faces some difficulties. For example, the power multiplexing technology is not mature; the non-orthogonal receiver is quite complex; interferences from other nodes; the channel states are needed to restore and remove; designing the SIC NOMA-receivers also depends on further advancement in signal processing chips.

4) *CSMA/TDMA based Hybrid MAC*: The carrier sense multiple access with collision avoidance (CSMA/CA) protocol is a contest-based MAC protocol with high flexibility and scalability. However, the CSMA protocol is inefficient in terms of throughput due to the time wastage caused by the back-off process. On the contrary, the CSMA/TDMA hybrid MAC protocol improves determinacy by combining the advantages of CSMA and TDMA. Specifically, CSMA/TDMA does not require the frame synchronization among nodes. In case the traffic load is relatively light, the devices access the channel using the CSMA/CA scheme. In case the traffic load is relatively heavy, the devices would send a slot request (indicating how many slots are needed) to a central coordinator, which allocates some reserved slots for the node. In doing so, higher channel utilization and lower latency can be guaranteed. In CSMA/TDMA, however, all nodes must perform low-power

listening in all the time slots to monitor the possibly incoming data, which is challenging for most battery-powered sensor nodes [16], [17].

5) *ARQ and HARQ*: In wireless networks, the reliability of data transmissions is guaranteed by retransmission mechanisms. In the conventional automatic retransmission request (ARQ) mechanism (e.g., stop-and-wait ARQ), the receiver returns an ACK to the transmitter if the decoding of the received signal is successful and returns a NACK to start a retransmission otherwise. The continuous ARQ protocol allows the transmitter to send packets continuously without waiting for a confirming feedback, including the go-back-N ARQ protocol and the selective repeat ARQ protocol. In the go-back-N ARQ protocol, the receiver sends a feedback (NACK) only for unsuccessful decodings. Upon receiving a NACK, the transmitter would retransmit the corresponding packet and all the following packets. In the selective repeat ARQ protocol, each packet is labeled with a timer, which indicates the number of allowed slots to wait for the feedback of the packet. If an ACK is not received within the period indicated by the timer, the transmitter will retransmit the packet. On the contrary, in a super ARQ protocol, the transmitter keeps retransmitting a packet until its ACK packet is received. In each slot, we denote the probability of successfully transmitting a packet over the fading channel as p and denote the probability of successfully transmitting an ACK/NACK over the feedback channel as p_f . We set the timer to C slots. It assumes that the latter packet is transmitted immediately after completing the transmission of the previous packet. Denote the number of slots required for a packet to complete its transmission as T_p , which includes the time to transmit and retransmit the packets, as well as the time to transmit the ACK and NACK. Specifically, T_p can be expressed as

$$T_p = T_t + \sum_{i=1}^M T_i, \quad (1)$$

in which T_t is the number of slots to transmit and retransmit the packet, T_i is the time to feedback an ACK or a NACK, and M is the number of used ACKs and NACKs. It is clear that T_t follows the geometric distribution with parameter p and T_i follows the geometric distribution with parameter p_f . In the stop-and-wait ARQ, the next packet cannot be started until the ACK of current packet. Note that the transmission and retransmission consume T_t slots means that there are one slot to transmit the packet (which fails) and $T_t - 1$ slots to retransmit the packet (only the last retransmission is successful). Thus, the receiver will feedback $T_t - 1$ NACKs and one ACKs, i.e., $M = T_t$. In the go-back-N ARQ, we have $M = T_t - 1$, since the receiver only feeds back the NACKs. In

the super-ARQ, the transmitter keeps retransmitting a packet until it receives the feedback ACK. Thus, we only need to consider the delay generated by feedbacking ACKs, i.e., $M = 1$. For the selective repeat ARQ, if the transmitter does not receive an ACK within the period defined by the timer of C , the packet would be retransmitted, and thus T_p can be expressed as

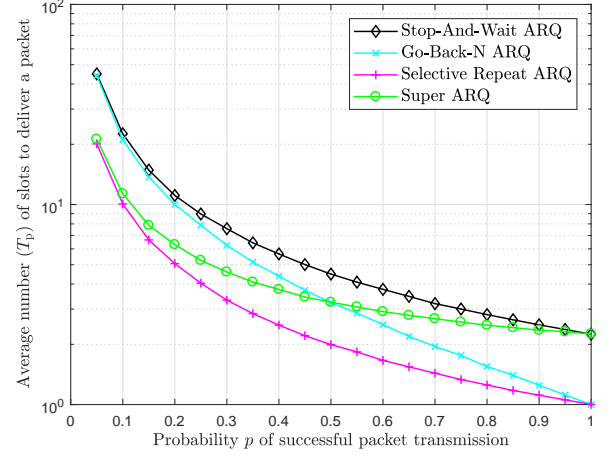
$$T_p^i = \begin{cases} T_t^i, & 1 + T_f^i \leq C \\ T_t^i + T_p^{i+1}, & 1 + T_f^i > C, \end{cases} \quad (2)$$

$$T_p = T_p^1, \quad (3)$$

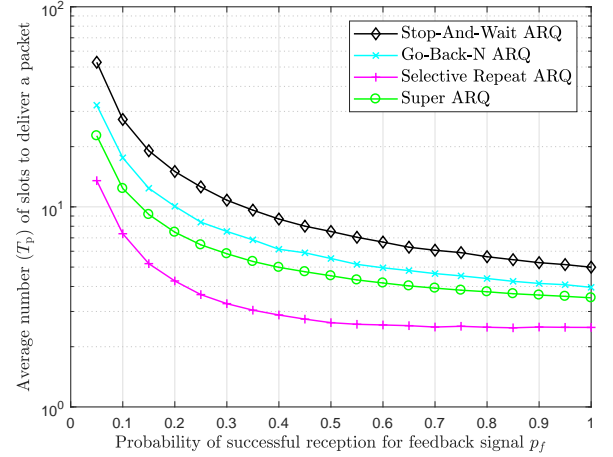
in which T_p^i is the number of time slots for the packet to complete its transmission by the i -th retransmission, T_t^i is the number of time slots to deliver the packet during the i -th retransmission, T_f^i is the number of time slots for feedbacking the ACK in the i -th retransmission. Note that T_t^i and T_f^i obey geometric distributions with parameters p and p_f , respectively. In Fig.3(a) and Fig.3(b), we compare the number of transmission slots to deliver a packet over a fading channel for the above mentioned protocols by simulation. From the Fig.3(a) and Fig.3(b), we observe that the transmission delay of the selective repetitive ARQ is the smallest. However, the packets may not be received with correct order and the packet repetition probability is relatively large. In addition, the go-back-N ARQ protocol is a satisfying choice in case the probability of successful decoding is large, i.e., the channel condition is good.

Another widely used retransmission protocol is the hybrid ARQ (HARQ), which combines the forward error correction coding (FEC) and ARQ. Specifically, the transmitter encodes each packet with a certain FEC coding scheme. When the receiver fails to decode a packet, it saves the received signal and feedbacks the transmitter a NACK to start a retransmission. Upon receiving the retransmitted signal, the receiver decodes the packet based on an optimized combination of the retransmission and previous unsuccessful attempts other than the retransmission itself. Compared with other ARQ protocols, HARQ can efficiently reduce the probability of unsuccessful decoding, and thus increases the transmission rates while reduces the transmission delay. However, the use of FEC coding requires additional overhead bits in each packet. Thus, a reasonable balance between transmission reliability and the transmission efficiency should be considered.

6) *TSCH*: Time synchronous channel hopping (TSCH) is a MAC layer scheduling scheme specified by IEEE 802.15.4 [18], [19]. It is designed to provide high reliability channel access for low power IoT networks. Specifically, TSCH combines the TDMA, frequency hopping, and the HARQ retransmission mechanism. When a packet transmission fails, it would be retransmitted in the next time slot on a different frequency with the HARQ protocol, which provides a higher probability of success than using the same frequency again. Thus, TSCH exploits the frequency diversity and time diversity effectively. More, TSCH can also be combined with path di-



(a) Average number (T_p) of slots to deliver a packet versus the probability p of successful packet transmission, in which $p_f = 0.8$ and $C = 5$ slots.



(b) Average number (T_p) of slots to deliver a packet versus the probability p_f of successful feedback transmission, in which $p = 0.4$ and $C = 5$ slots.

Fig. 3. Change in the average number (T_p) of slots to deliver a packet.

versity techniques (e.g., packet replication) that allows packets to be transmitted over multiple independent paths. In doing so, TSCH reduces the packet loss and the end-to-end delay of the network effectively. Different from CSMA/CA, the low-power feature of TSCH allows routing nodes to reduce power consumption and ensures better latency control. In addition, TSCH provides a resource reservation mechanism that allows slots to be reserved in advance for time-critical data traffic, and thus ensures better determinacy.

7) *Black Bursts*: In case there are both high-priority real-time traffic and low-priority non-real-time traffic in a network, Sobrinho and Krishnakumar proposed an Ethernet quality of service using black bursts (EQuB) mechanism to provide bounded delay for high-priority traffic [21]. Specifically, the low-priority nodes tries to access the channel by using the binary exponential back-off algorithm. In case a high-priority node participates the channel contention, it transmits a jamming sequence (i.e., a black burst) with a pre-defined length to stop other node from accessing the channel. The use of black bursting requires mandatory modifications to the MAC

and physical layers, which inevitably changes the use of components off-the-shelf (COTS). The black bursting mechanism has been proved to be effective in providing high QoS and has been applied in various real-time wireless applications.

8) *Transmission Protections*: Most existing real-time transmission schemes focus on high-priority real-time traffic without considering the influence on those low-priority traffic. For example, the preemption mechanism cannot guarantee the worst-case end-to-end delay (WCD) for the low-priority traffic. Therefore, it is necessary to guarantee the WCD of individual data traffic through adjustable protected transmission mechanisms, as detailed in the following.

- **Power adjustment**: Note that by adjusting the instant transmit power, the probability of successful packet transmission can be changed on demand. When the (re)transmission of some packets is close to its deadline, we can increase its transmit power so that the transmission packets could be completed before the deadline.
- **Bandwidth allocation**: To meet the requirement of real-time applications in wireless networks, the limited bandwidth needs to be allocated for high-priority traffic flows first and allocated for low-priority traffic flows when there is spare bandwidth. Among the TSN-related protocols, IEEE 802.1 Qci defines a flow filter that can block all future frames of a particular traffic if unexpected/excessive bandwidth usage is detected; IEEE 802.1Qcc improves the stream registration protocol (SRP) by automatically reducing the bandwidth and frame size of flows if there is no sufficient bandwidth resources for its reservation. Bandwidth resources can also be allocated and optimized according to the probability distribution of traffic flows with different priority levels, so that when bandwidth for high-priority traffic is insufficient, the bandwidth allocation between the low-priority flows and the high-priority traffic flows can be adjusted. However, the throughput achieved through bandwidth allocation is still limited, as the available sub-6-GHz spectrum is insufficient to meet the extreme bandwidth demands of data-intensive IIoT applications. Therefore, a beyond sub-6-GHz band is required for IIoT.
- **The 1:1 and 1+1 protection**: The “1:1” protection mode uses two transmission channels, i.e., a primary channel and a backup channel. During normal operations, data traffic is transmitted through the primary channel while the standby backup channel can be shared with other low-priority traffic [12]. In case a transmission failure occurs on the primary channel, the transmitter and receiver would negotiate and make a decision to switch the transmission from the primary channel to the backup channel. The “1+1” protection mode allows the traffic flow to be transmitted on the both channels at the same time. The receiver monitors the quality of the channels and decodes packets from the one with a higher reliability.
- **Dynamic priority assignment**: Variable priority transmission allows the transmitter to increase the priority of a low-priority traffic flow when its packets are approaching their deadlines. This dynamic priority assignment can reduce the end-to-end delay of those initially assigned

low-priority traffic effectively.

9) *WTSN*: Recent advances in 5G and IEEE 802.11 based wireless accessing technologies have boosted significant interest to wireless time-sensitive networking (WTSN). Note that some MAC layer protocols associated with TSN can effectively reduce end-to-end latency and improve transmission reliability. For example, IEEE 802.1Qbv meets the strict delay bounds of high-priority traffic through the gate control list (GCL) mechanism and reduces the interference between traffic flows with different real-time priorities by separating them with protection bands; IEEE 802.1CB increases transmission reliability by transmitting redundant packets over independent paths and eliminating some duplicated packets at or near the receiver; With some proper modifications, some TSN functionalities can also work well over wireless media and wireless networks. In fact, a portion of the 802.1 TSN standards (e.g., 802.1AS (time synchronization), 802.1Qav (credit-based traffic shaping), 802.1Qbv (time-aware scheduling), and 802.1Qca (path control and reservation)) have been shown to be applicable to wireless standards (e.g., 802.11 and 5G) [22]. In addition, 802.11be (defined by TGB) have provided several new features on better integration with 802.1 TSN standards. For example, the newly defined multi-link/channel operation reduces congestion by separating traffic of different priority levels; the multi-AP functionality enables multi-AP to improve link reliability by exploiting spatial diversity gain.

B. Physical Layer Techniques

In order to connect smart devices (both fixed and mobile) to distributed industrial data centers, the 5G-based IIoT networks need to provide reliable wireless accessing services. Specifically, a large amount of data needs to be transmitted at a very high rate (up to 40 Gb/s for stationary devices and 5 Gb/s for mobile devices), while centimeter-level positioning accuracy needs to be guaranteed under multi-path conditions [23]. This has driven extensive research in advanced signal processing methods at the physical layer, aiming at transmitting information over the channels with high reliability and high data rates.

1) *mmWave Communications*: With advantages in larger bandwidth, lower interference, and more abundant spectrum, the mmWave communication is widely considered as a prospective solution to future wireless networks with demand on increasingly higher data rates and lower latency [24]. For example, the capacity of mmWave-based 5G system is ten times larger than traditional systems, and thus is applicable for future IoT-based factory deployments [25]. Note that the mmWave antennas are much smaller than those traditional antennas using lower frequency bands, since the minimum length of an antenna increases linearly with the wave-length. Thus, many mmWave antennas can be placed in small size devices (e.g., smart phones) as mmWave-mMIMO antenna arrays to combat path loss [26]. Modern on-chip mmWave-mMIMO arrays have also proved to be cost effective and power efficient for industrial IoT devices [27]. Moreover, by using the unlicensed millimeter wave spectrum and the TDMA scheduling in ultra-dense radio access networks (RAN), smaller delays and

jitters can be realized at a near-minimum transmission power in Wi-Fi systems [28], which provides a feasible solution for time-critical industrial wireless communications. Nevertheless, there are still some issues that need to be considered in applying mmWave communications in time-critical industrial IoT networks.

- In case the used frequency bandwidth is very large (e.g., over 1 GHz), the corresponding A/D converter and signal processing unit would be much more expensive and less energy-efficient [29].
- The path-loss attenuation is large for mmWave communications. In order to reduce interference and attenuation, it is critical to perform directional beamforming and transmit on LoS path in mm-Wave communications [30].
- While beamforming provides higher antenna gain by using narrower beams, the corresponding multiplexing gain is smaller due to less multi-path components. In fact, parallel multi-path transmissions between multiple transmitter-receiver pairs are crucial to increase the reliability and data rate, in terms of diversity and multiplexing gain, respectively. Therefore, a balance between array gain (beamforming) and spatial multiplexing gain should be considered [31].
- The penetration capability of mm-Wave signals is poor due to its high frequency. Furthermore, since the scattering level is high and the objects are mobile in industrial environments, the probability of LoS blocking is high. To this end, we need to provide effective beam management (e.g. tracking, serving) in mmWave communications [32].

2) *Modulation and Coding Schemes:* In industrial IoT networks, the packets are usually short, with a payload size of a few bytes. Thus, the controlling overhead (e.g., synchronization information, channel status) is comparable to the payload or even higher [33]. On the one hand, the code block length is very limited in industrial applications, so that the coding gain is also limited, which reduces the reliability of transmissions. On the other hand, by using high complexity decoding algorithms and the HARQ scheme, although higher reliability of transmissions can be guaranteed, and the corresponding latency will inevitably be increased. Therefore, efficient modulation and coding schemes are essential to achieve deterministic communications.

- Fixed-rate channel codes: The BCH codes, tail-biting convolutional codes, low density parity check (LDPC) codes, and polar codes are considered as strong candidates for URLLC. Specifically, although the minimum distance of BCH codes is relatively large and the block error rate is smaller, the code block length cannot be chosen arbitrarily and is not so flexible; while the tail-biting convolutional codes solve the rate loss problem due to its zero-tailed termination in short block lengths, the encoding/decoding complexity is high; the coding rate of LDPC codes can closely approach the Shannon Limit, but the decoding complexity increases with flexibility; while the polar codes perform better than LDPC codes in the short block length regime and are not limited by any error lower bounds, the decoding complexity is higher as the

block length becomes larger.

- Rate adaptive schemes: Modern channel estimation techniques are providing transmitters with more and more accurate CSI, with which the transmitter can determine its coding rate adaptively so that the desired block error rate can be guaranteed. The main challenge for using rate adaptive schemes in IoT networks is that the overhead is relatively large due to the limited block length.
- Rateless code: A rateless code encodes the information to be transmitted to a large number of short packets without requiring the CSI of the channel. These packets are then transmitted to the receiver without requiring feedback for each packet. Once the receiver has successfully received a certain number of these packets, the original information can be decoded successfully. Thus, the coding rate of rateless codes is adaptive to the channel automatically. This feedback-free property makes rateless codes a prospective scheme to meet the low latency and high reliability requirements of URLLC applications. Conventional rateless code, such as LT codes and Raptor codes [34], have also been shown to approach infinite block length channel capacity over binary erasure channels. In the low SNR regime, however, the achievable rate of rateless codes degrades significantly, so that they have to be used jointly with some channel codes.
- Analog fountain code: The analog fountain code (AFC) extends traditional rateless codes by coding over the Euclidean space other than the Hamming space. Thus, AFC can be readily used for adaptive modulations. It was also shown that the capacity of AFC is close to that of rateless codes for a wide range of SNR and the coding complexity is linear with arbitrary code rate. These features enables AFC a prospective solution for the physical layer design of URLLC [35].

Coding schemes that are adaptive for a wide range of channel conditions could reduce end-to-end delay and improve transmission reliability effectively. For rate adaptive schemes (even if rateless codes or AFC is used), the receiver still requires the CSI to decode the received packets, so that the transmitter has to insert pilot symbols in each block, which may lead to serious performance loss. The above mentioned self-adaptive channel codes also have some challenges and problems. For example, the decoding delay of AFC and that of other URLLC candidates (such as LDPC codes and polar codes) are comparable.

C. An End-to-End Perspective

1) *Time-triggered Mechanism:* In time-triggered mechanisms, the activities are initiated periodically, which reduces the randomness in the system *in the first place* [5]. On the contrary, traditional Ethernet uses an event-triggered mechanism, in which end systems/devices can access the network at any time. Since the events arrive at these end systems randomly and their communications are carried out by the same medium, the accumulation of delays and jitter are inevitable and relatively large. Therefore, it is easier to provide deterministic communications by replacing event-triggered networks with time-triggered networks (e.g., TTP, TTCAN, TTE, FlexRay).

2) *Time Synchronization*: In industrial communications networks, multiple traffic flows of different priorities share the communications channels to improve efficiency and reduce costs. It is necessary to ensure the time synchronization among these traffic flows. In wire-line networks, IEEE 802.1AS provides bounded delay and extremely low delay variation for TSN applications through precise time synchronization for all the nodes, which is a fundamental requirement for implementing other standard functions. In wireless networks, it was shown that precise time synchronization and precise time protocol (PTP) applications can be supported by using hardware time-stamping in an open source testbed called *openWIFI* [36].

3) *Traffic Shaping*: In case that some devices generate burst of the traffic over the network, traffic shaping *at the sources* should be used to smooth the bursts according the available bandwidth. For example, IEEE 802.1Qcr reshapes the flows based on their priorities at each hop of transmission using an asynchronous traffic shaper (ATS); IEEE 802.1Qbv provides a time-aware shaper (TAS) through the GCL and a credit-based shaper (CBS) to prevent low-priority traffic shortages. For wireless networks, traffic shaping and scheduling is even more important. By further considering the randomness of fading channels, extending ATS and TAS to wireless networks is necessary and feasible.

4) *Scheduling Transmissions and Decision-Makings*: In addition to providing reliable transmission links, optimizing the desired decision epochs *at the destinations*, i.e., the time to use the received information, is also crucial to deterministic networking. As is shown in [37], making decisions periodically leads to the smallest average age upon decisions (AuD), i.e., the average age of received update when they are used for decisions. Moreover, for a sequence of periodic decision epochs, we can maximize the probability for the packets to be received at their respective decision epochs (i.e., on-time) by scheduling the *transmission of packets* such as repeating, delaying and dropping the packets [3]. Suppose a sequence packets are transmitted to the receiver through a Rayleigh fading channel, the pre-defined reception slot interval of adjacent packets is T_{tgt} , and the m -th packet is expected to be received at the mT_{tgt} -th slot with a deviation no larger than δ slots. We present how the on-time reception rate changes when the reception probability p increases in Fig. 4, in which the on-time reception rate is obtained by the Algorithm 2 in [3]. From Fig. 4, we observe that by optimally scheduling the packets, the on-time reception rate approaches 100% as the probability of successful decoding from each transmission approaches $p = 0.5$, which indicates a very unreliable channel. On the contrary, if no packet scheduling is used (i.e., using the random transmission scheme), the on-time reception rate is close to zero, regardless of p .

V. CONCLUSION

This article analyzes the possibility to provide deterministic communications with wireless networks. On the one hand, traditional wireless channels using omni-directional antennae are highly unreliable while industrial IoT networks raised

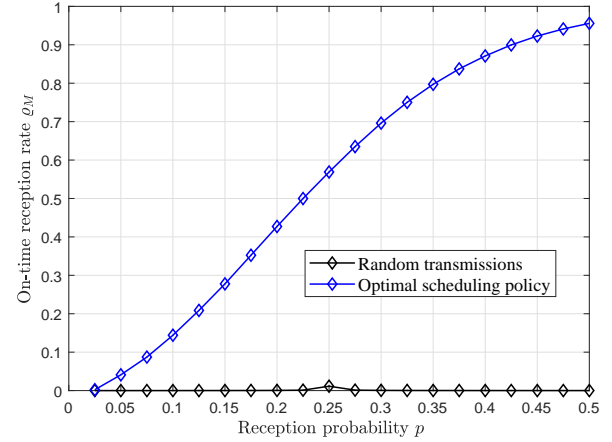


Fig. 4. On-time reception rate versus the probability of success decoding of the fading channel, in which $T_{\text{tgt}} = 4$ and $\delta = 2$.

strict real-time requirements for information transmissions. On the other hand, recent progress in wire-line deterministic networking technologies provided valuable benchmarks or references for wireless communications. By optimally choosing the MAC layer techniques (e.g., TDMA, packet scheduling, and transmission protection) and physical layer techniques (e.g., OFDMA and AFC), it is believed that wireless deterministic networks will come to us in the near future. It is also expected that when some machine learning method such as the Q-learning, multi-agent reinforce learning, or the federated learning is used, the system design would be more flexible for deterministic communications scenarios.

REFERENCES

- [1] A. Nasrallah et al., "Ultra-Low Latency (ULL) Networks: The IEEE TSN and IETF DetNet Standards and Related 5G ULL Research," in *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 88–145, Firstquarter 2019.
- [2] T. H. Szymanski, "Supporting Consumer Services in a Deterministic Industrial Internet Core Network," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 110–117, 2016.
- [3] Y. Li, Y. Dong, B. Shim (2022), "On-time communications over fading channels [Online]," Available: <https://arxiv.org/abs/2202.08646>.
- [4] Weilan Huang, Xiaofei Zhao, et al. *The convergence of 5G deterministic networking and industrial Internet white paper*, [Online]. Available: <https://www-file.huawei.com/-/media/corporate/pdf/news/5gdn-based-industrial-internet-white-paper.pdf>
- [5] H. Gharavi and B. Hu, "Deterministic Intra-Vehicle Communications: Timing and Synchronization," *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, 2020.
- [6] J.Y. Lee and R. A. Scholtz, "Ranging in a dense multi-path environment using an UWB radio link," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1677–1683, 2002.
- [7] M. Luvisotto, Z. Pang and D. Dzung, "High-Performance Wireless Networks for Industrial Control Applications: New Targets and Feasibility", *Proc. IEEE*, 2019.
- [8] N. Finn, "Introduction to Time-Sensitive Networking," *IEEE Commun. Standards Mag.*, vol. 2, no. 2, pp. 22–28, Jun. 2018.
- [9] X. Yang, D. Scholz, and M. Helm, "Deterministic networking (DetNet) vs time sensitive networking (TSN)," in *Seminar IITM SS'19: Netw. Architect. Services*, pp. 79–84, Oct. 2019.
- [10] Y. Cheng, D. Yang, and H. Zhou, "Det-WiFi: A multihop TDMA MAC implementation for industrial deterministic applications based on commodity 802.11 hardware," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–10, Apl. 2017.

- [11] Purple Mountain Laboratory, et al. "Deterministic arithmetic network white paper," [Online]. Available: <https://www.ambchina.com/zfc/uploads/1660612978488.pdf>
- [12] Shuo Wang, Jiasen Wang, et al. "Future Network White Paper on Deterministic Networking Technologies," [Online]. Available: <https://www-file.huawei.com/-/media/corporate/pdf/news/future-network-whitepaper.pdf>
- [13] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Conf. Rec. Forty 4th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1558–1562.
- [14] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
- [15] B. Anjum, "Optimizing channel access for bandwidth versus latency in 802.11ac WLANs," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Garden Grove, CA, USA, Feb. 2015, pp. 113–128.
- [16] Naguib, Ahmed, Waleed Saad, and Mona Shokair. "Remaining Energy Aware ML-CSMA/TDMA Hybrid MAC Protocol for LTE-M2M Wireless Network," 2019.
- [17] W. Wang, H. Wang, D. Peng, and H. Sharif, "An energy efficient preschedule scheme for hybrid CSMA/TDMA MAC in wireless sensor networks," in *Proc. 10th IEEE Singapore Int. Conf. Commun. Syst.*, Oct. 2006, pp. 1–5.
- [18] D. Dujovne, T. Watteyne, X. Vilajosana, and P. Thubert, "6TiSCH: deterministic ip-enabled industrial internet (of things)," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 36–41, December 2014.
- [19] C. Orfanidis, A. Elsts, P. Popy, and X. Fafoutis, "TSCH evaluation under heterogeneous mobile Scenarios," *IoT*, vol. 2, no. 4, pp. 656–668, Feb. 2021.
- [20] 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.
- [21] J.L. Sobrinho, A. Krishnakumar, "EQuB-Ethernet quality of service using black bursts," *Conference on Local Computer Networks (1998)* 286–296.
- [22] D. Cavalcanti, "Wireless TSN-Definitions, Use Cases and Standards Roadmap [Online]," Available: <https://avnu.org/wireless-tsn-paper/>.
- [23] A. V. Lopez, A. Chervyakov, G. Chance, S. Verma, and Y. Tang, "Opportunities and challenges of mmWave NR," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 4–6, Apr. 2019.
- [24] Sur, S., Venkateswaran, V., Zhang, X., Ramanathan, P., "60 GHz Indoor Networking through Flexible Beams: A Link-Level Profiling," in *Proc. of the ACM SIGMETRICS* (2015).
- [25] B. M. Lee and H. Yang, "Massive MIMO for industrial Internet of Things in cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2641–2652, Jun. 2018.
- [26] J. Zhang, E. Bjornson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.
- [27] W. Hong, "Solving the 5G mobile antenna puzzle: Assessing future directions for the 5G mobile antenna paradigm shift," *IEEE Microw. Mag.*, vol. 18, no. 7, pp. 86–102, Nov./Dec. 2017.
- [28] T. H. Szymanski, "Strengthening security and privacy in an ultra-dense green 5G radio access network for the industrial and tactile Internet of Things," in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 415–422.
- [29] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.
- [30] C. Cano, G. H. Sim, A. Asadi, and X. Vilajosana, "A channel measurement campaign for mmWave communication in industrial settings," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 299–315, Jan. 2021.
- [31] A. Mahmood et al., "Industrial IoT in 5G-andbeyond networks: Vision, architecture, and design trends," *IEEE Trans. Ind. Informat.*, early access, 2021.
- [32] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "Standalone and non-standalone beam management for 3GPP NR at mmWaves," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 123–129, Apr. 2019.
- [33] M. Luvisotto, Z. Pang, D. Dzung, M. Zhan, and X. Jiang, "Physical layer design of high-performance wireless transmission for critical control applications," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2844–2854, Dec. 2017.
- [34] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.
- [35] M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Near-capacity adaptive analog fountain codes for wireless channels," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2241–2244, Dec. 2013.
- [36] M. Aslam, W. Liu, X. Jiao, and J. Haxhibeqiri, "High precision time synchronization on Wi-Fi based multi-hop network," in *Proc. IEEE INFOCOM WKSHPs*, Vancouver, BC, Canada, May 2021.
- [37] Y. Dong, Z. Chen, S. Liu, P. Fan, and K. B. Letaief, "Age-upon-decisions minimizing scheduling in Internet of Things: To be random or to be deterministic?" *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1081–1097, Feb. 2020.