Statistical Learning-based Adaptive Network Access for the Industrial Internet-of-Things

Muhammad Ahmad Raza, *Graduate Student Member, IEEE,* Mehran Abolhasan, *Senior Member, IEEE,* Justin Lipman, *Senior Member, IEEE,* Negin Shariati, *Member, IEEE,* Wei Ni, *Senior Member, IEEE,* and Abbas Jamalipour, *Fellow, IEEE*

***Abstract*—Industrial Internet-of-Things (IIoT) applications generate data in varying amounts with diverse quality of service requirements. The adaptive network access approach and distributed resource management in IIoT networks can reduce the communication overheads caused by centralized resource management approaches. In this regard, statistical learning is a promising tool for addressing decision-making problems in a dynamic environment. This paper considers uplink dominant IIoT networks in which massive devices generate delay-sensitive and delay-tolerant data and communicate over shared radio resources. We propose a novel grant-free access scheme using a statistical learning approach that enables IIoT entities to perform delay-sensitive and delay-tolerant transmissions over dynamically partitioned resources in a prioritized manner. In order to improve utilization of available radio resources, we design an adaptive network access mechanism operating in a semi-distributed manner. This mechanism enables end devices to use their transmission history to choose between static and dynamic resource allocation-based grant-free schemes in a dynamic environment. Simulation results show that average latency and resource utilization vary in grant-free access schemes employing static and dynamic resource allocations. Thus, compared to a single transmission scheme, the proposed adaptive network access offers better channel utilization while meeting the application-specific latency bound in IIoT networks.**

***Index Terms*—Industrial Internet of Things, Adaptive networks, Statistical learning, Grant-free access.**

I. Introduction

FUTURE wireless networks are envisioned to support Industrial Internet of Things (IIoT) applications that generate network traffic with diverse quality of service (QoS) requirements [1-3]. Massive machine type communication (MTC) devices deployed in the industrial environment generate delay-sensitive, and delay-tolerant data in cyclic and acyclic manners [2]. Latency requirements for different industrial process automation applications, including safety, control, and monitoring, are described in [4]. Delay-sensitive data needs to be delivered under strict latency and reliability constraints compared to the transmission of delay-tolerant data. Moreover, the amount of data being generated can vary from application to application. The use of automated guided vehicles and mobile robots has gained considerable attention in industrial operations, including warehouse operations [5]. IIoT networks are used to operate the unmanned vehicles performing time-sensitive (TS) and non-TS tasks in smart warehouses [6]. Furthermore, vehicular IoT entities generate maintenance-critical (delay-sensitive) and maintenance non-critical (delay-tolerant) data of significantly different amounts [7]. The use of multimedia applications in IIoT networks requires the availability of low latency communication links to transmit large amounts of delay-sensitive data. Along with the latency and reliability, the number of successful delay-sensitive transmissions per second and the portions of the bandwidth and computing resources available for delay-tolerant transmission are among the key performance indicators of IIoT networks [3], [6]. Fulfillment of application-specific diverse QoS requirements is challenging when communication is performed over limited radio resources and different network parameters change dynamically [8].

Long Term Evolution (LTE) and LTE-Advanced (LTE-A) technologies provide a four-stage grant-based network access mechanism for communication over limited radio resources [9]. In this approach, the devices first undergo a random access channel (RACH) phase where each device transmits a preamble selected randomly from a pool of available preambles. Collisions happen when multiple devices select the same preamble resulting in RACH failures. The successful devices are granted dedicated resources to transmit their data. Implementing appropriate transmit power control and back-off strategies in successive RACH attempts can improve the probability of success in the RACH phase [10]. However, due to the inherent control signaling overheads, the grant-based network access approach is more suitable for delay-tolerant transmission. Moreover, the reservation of available resources in grant-based access methods makes it less efficient for networks with massive devices. On the other hand, grant-free access avoids long scheduling delays by allowing transmissions over shared resources without going through a request-grant phase. Grant-free access has gained considerable attention to support IIoT applications in 5G and future wireless networks [11]. However, simultaneous transmissions from two or more devices over the same channel can impact the system efficiency and the reliability of delay-sensitive data transmission. Therefore, to evaluate the performance of a particular grant-free access mechanism, both the latency and channel

Muhammad Ahmad Raza, Mehran Abolhasan, Justin Lipman, and Negin Shariati are with the RF and Communication Technologies (RFCT) research laboratory, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia. (email: muhammadahmad.raza@student.uts.edu.au; mehran.abolhasan@uts.edu.au; justin.lipman@uts.edu.au; negin.shariati@uts.edu.au).

Wei Ni is with the Commonwealth Scientific and Industrial Research Organization, Sydney, NSW 2122, Australia (email: wei.ni@data61.csiro.au).

Abbas Jamalipour is with the School of Electrical & Information Engineering, the University of Sydney, NSW 2006, Australia (email: a.jamalipour@ieee.org).

utilization need to be considered.

While communicating over shared radio resources, the number of active devices is one of the critical factors that govern the behavior of the random component of latency experienced by a data packet. Since the number of active devices in IIoT networks can change over time, we can exploit this change to accommodate the transmission of delay-tolerant data resulting in enhanced utilization of the available shared radio resources. Thus when the network load is at a level where transmission of delay-tolerant data does not impact the QoS of delay-sensitive data, the devices can utilize the available resources for the delay-tolerant transmissions. Such an effort would require an adaptive network access mechanism in which end devices can utilize their transmission history to predict the corresponding latency and resource utilization of the network access mechanism under variable network load. Moreover, to avoid the control signaling overheads, end devices need to be able to partition the available resources to accommodate two different types of network traffic without requiring additional feedback information from the BS. The design of such semi-distributed adaptive network access mechanisms involving dynamic resource allocation is challenging under time-varying network conditions. Therefore, distributed computing needs to be explored to address IIoT applications where a large number of devices generate delay-sensitive and delay-tolerant data in a dynamic environment.

The above discussion indicates that the adaptive network access mechanism and device-level learning approach are crucial in designing IIoT networks under limited radio resources and variable network load. The need for adaptive network access approaches is further strengthened when the probability distribution of the time-varying network load is unknown. Moreover, the proportions of the available time, frequency, and energy resources consumed to provide control information to massive end devices become more significant in the BS-centered resource management approaches. On the one hand, existing physical layer enhancements combat the channel's time-varying nature to provide the desired application-specific QoS requirements. On the other hand, statistical learning can be used at the MAC layer to potentially address the problem of adaptive network access to support IoT applications with diverse latency-reliability constraints [12], [13].

In this paper, inspired by the need for adaptive network access mechanisms for future IIoT applications, we consider the uplink dominant IIoT networks employing grant-free access. The large number of devices in these networks generate delay-sensitive and delay-tolerant data in random and deterministic manners with varying amounts and diverse QoS requirements. We aim to design a statistical learning-based semi-distributed adaptive network access mechanism that can enable devices to adapt to the network dynamics with limited assistance from the BS. For that purpose, we address the challenge of device-level resource partitioning to accommodate delay-sensitive and delay-tolerant transmissions in IIoT networks. The required parameters related to the device-level resource partitioning are obtained through a statistical learning based-*Network Exploration Phase* designed in [13]. The key contributions of this paper are summarized as follows:

- We propose a statistical learning-based novel grant-free access scheme in which end devices are enabled to dynamically partition the available channels for two different types of transmissions. The proposed scheme prioritizes delay-sensitive transmissions over delay-tolerant transmissions and provides an almost constant collision probability for delay-sensitive transmissions in each slot.
- We design an adaptive network access mechanism that enables end devices to choose between the grant-free access schemes with fixed and dynamic resource allocations under dynamically varying network load. The devices are also enabled to predict outages where current network load is too high to meet desired latency bound for delay-sensitive data, and devices perform a random back-off.
- The proposed adaptive network access operates in a semi-distributed manner and relies on the transmission history of end devices. It avoids additional feedback information from the BS, reducing control signaling overheads while efficiently managing transmission of different types of network traffic in a dynamic environment.

Through simulations, we show that average latency and channel utilization vary in different access mechanisms under the given network load. Therefore, the device-level decision-making capability enables end devices to adapt to the network dynamics and efficiently utilize available radio resources.

The rest of the paper is organized as follows: Section II provides a brief review of the related works. The system model is described in Section III, while Section IV outlines the steps of the proposed adaptive network access mechanism. Section V presents the design and analysis of the proposed grant-free access with dynamic resource allocation. Device-level prediction of different parameters related to the proposed scheme is discussed in Section VI. Simulation results are presented in Section VII, and the paper is concluded in Section VIII while highlighting future research directions. Table I contains the notations used in this paper.

II. Related Works

In this section, we briefly review some of the recent works which propose to use the grant-free access mechanism for a wide range of IoT applications, including IIoT applications. Moreover, we highlight the gaps in literature motivating us to perform the work presented in this paper.

Liu et al. [14] presented a comparative analysis for different types of grant-free access schemes supporting ultra-reliable and low latency communications (URLLC) services. This work used latent access failure probability to measure the performance of each grant-free scheme. Gao et al. [15] proposed a distributed coordination-based grant-free access for heterogeneous IIoT networks in which massive devices are deployed in a relatively small geographical area. This work used time-based access of a single channel for uplink data transmission. Gao et al. [16] developed a centralized scheduling control for the grant-free access proposed in [15]. Gharbieh et al. [17] proposed a grant-free access mechanism that enables end devices to perform uplink data transmission when their channel gains are above a prescribed threshold. The

TABLE I
LIST OF NOTATIONS

Notation	Definition
$\mathcal{C}$	Set of available channels
$\mathcal{C}_{m,n}^{(ds)}$	Subset of $\mathcal{C}$ for delay-sensitive transmissions
$\mathcal{C}_{m,n}^{(dt)}$	Subset of $\mathcal{C}$ for delay-tolerant transmissions
K	No. of orthogonal channels in $\mathcal{C}$
$K_{m,n}$	No. of channels available for delay-sensitive transmissions
$\mathcal{L}_{\text{max}}$	Application specific latency bound
M	No. of active devices
N	No. of slots per round
R	No. of rounds
$\mathcal{S}_{\text{ds}}^{(\text{DRA})}$	Average successful delay-sensitive transmissions per round
$\mathcal{S}_{\text{dt}}^{(\text{DRA})}$	Average successful delay-tolerant transmissions per round
$V_{m,n}$	No. of devices transmitting delay-sensitive data
$W_{m,n}$	No. of devices transmitting delay-tolerant data
$\alpha_{m,n}$	Probability of collision for delay-sensitive transmission in the SRA scheme
$\beta_{m,n}$	Probability of collision for delay-sensitive transmission in the DRA scheme
$\gamma_{m,n}$	Probability of collision for delay-tolerant transmission in the DRA scheme
$\eta^{(\text{SRA})}$	Channel utilization in the SRA scheme
$\eta^{(\text{DRA})}$	Channel utilization in the DRA scheme
$\mu_{\text{ds}}^{(\text{SRA})}$	Average latency in the SRA scheme
$\mu_{\text{ds}}^{(\text{DRA})}$	Average latency in the DRA scheme

proposed scheme aims to enhance the utilization of device-level harvested energy. Choi et al. [18] proposed two grant-free access schemes named power domain multiple access (PDMA) and a rate domain multiple access (RDMA). Multiple users share a single channel in these schemes employing a hybrid automatic repeat request with incremental redundancy-based re-transmit diversity. At the same time, a successive interference cancellation (SIC) based receiver is used. The grant-free access approach, when combined with the non-orthogonal multiple access (NOMA), can potentially address the problem of massive connectivity in delay-sensitive IoT applications [19]. Therefore, Tegos et al. [20] presented two power domain NOMA schemes for slotted ALOHA systems. In this work, one scheme uses an SIC-based receiver while the other scheme performs joint decoding at the receiver. Since the network traffic in IIoT networks has both delay-sensitive and delay-tolerant components with varying amounts of data, the coexistence of grant-free and grant-based approaches can potentially address these diverse QoS requirements [21].

The optimal allocation of available time and frequency resources in a selected network access mechanism is essential to meet the application-specific QoS requirements. The PHY-layer considerations play a significant role in adapting to the network dynamics [22], [23]. Along with the physical channel characteristics, the number of active devices at a given time

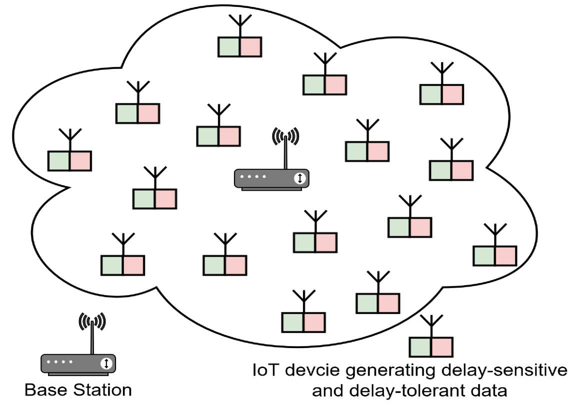

Fig. 1. IoT devices in an industrial environment generating delay-sensitive and delay-tolerant data and communicating with a single BS over shared radio resources in a grant-free manner.

also impacts the efficiency of resource utilization. Since the network load can change over time, the optimal resource allocation strategy being used should be updated dynamically. Significant research has been performed to address the problem of monitoring the network traffic and optimizing the resource allocation strategies accordingly [24], [25]. In these works, the probability of access class barring (ACB) plays a major role in controlling network congestion. However, the network traffic-aware radio resource management in existing grant-free and grant-based access mechanisms is primarily BS-centered and involves excessive computation and control signaling overheads. Therefore, new intelligent resource management methods are required for the uplink dominant IIoT networks, which can avoid the control signalling overheads caused by centralized resource management approaches.

Distributed computing enables on-device network exploration and thus can assist the BS in allocating radio resources for delay-sensitive and delay-tolerant transmissions much more efficiently [26], [27]. Statistical learning theory can be used for designing data-driven strategies to predict desired parameters in wireless networks when complete information on the probability distributions of these parameters is not available [12]. However, device-level acquisition of the knowledge regarding the statistical parameters is challenging when network load is variable. Therefore, Raza et al. [13] proposed a statistical learning-based grant-free access mechanism for delay-sensitive IoT applications. The proposed network access mechanism enables end devices to use their transmission history to learn different dynamic network parameters without relying much on the BS. The BS can utilize the knowledge shared by end devices regarding the current network to optimize the resource allocation strategy. Moreover, the proposed network exploration strategy identifies different IoT groups based on their latency and reliability requirements. However, the work in [13] considers only the transmission of delay-sensitive data, and the proposed grant-free access mechanism uses static resource allocation.

The existing centralized grant-free access mechanisms primarily employ static resource allocation to accommodate data transmissions having different latency requirements. However,

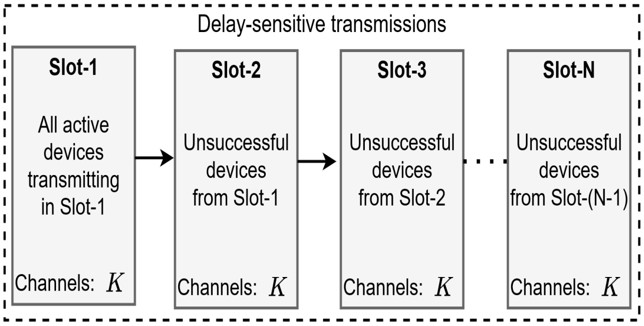

Fig. 2. The status of different slots in the m^{th} round of *Network Exploration Phase* with $m = 1, 2, ..., R$. In each round devices transmit delay-sensitive by employing the SRA scheme.

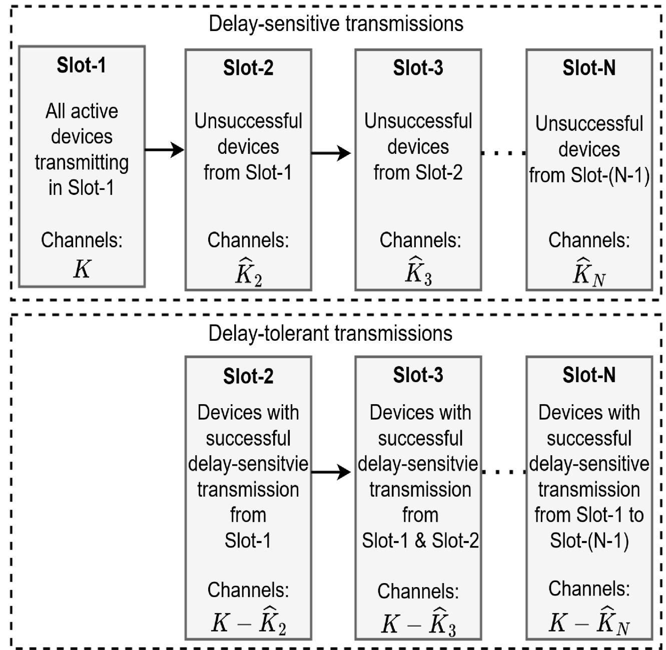

Fig. 3. Under the condition: $\widehat{\mu}_{\text{ds}}^{(\text{DRA})} \leq \mathcal{L}_{\max}$, status of different slots in each round for $m = R+1, R+2, ..., 2R$ where devices transmit delay-sensitive and delay-tolerant data by employing the DRA scheme.

to provide the desired QoS requirements for different data transmissions, the number of resources can be determined dynamically, thus improving the utilization of available shared radio resources. Moreover, to reduce the control signaling overheads, end devices need to be equipped with the capability of adapting to the network dynamics with limited feedback from the BS. Motivated by these required enhancements in the existing literature, we aim to design an adaptive network access mechanism enabling end devices to dynamically determine the resources required for delay-sensitive and delay-tolerant transmissions in the IIoT networks.

III. System Model

As shown in Fig. 1, we consider uplink communication scenario in IIoT networks composed of massive devices and a single base station (BS). The devices, shown by using two colors in Fig .1, generate delay-sensitive and delay-tolerant data. Each device has a scheduler to prioritize the transmission of delay-sensitive data, and all devices perform uplink data transmissions to the BS. Time is divided into slots. A sequence of N consecutive slots makes one round, and a window comprises $2R$ rounds.

At the start of each window, M devices become active and communicate with the BS over shared radio resources. The number of active devices remains fixed in a given window; however, the value of M is unknown to the BS. Moreover, M can randomly change from one window to the other while the probability distribution of M is unknown. Active devices use the first half of each window, for rounds $m = 1, 2, ..., R$, to explore the network, including the network load prediction. While the second half of each window, for rounds $m = R+1, ..., 2R$, is used to adapt to the network dynamics. It is assumed that each device can completely transmit its data packet in one slot.

Each slot is equipped with K orthogonal channels $\mathcal{C} = \{C_1, C_2, ..., C_K\}$, where each channel in $\mathcal{C}$ can be a frequency or code-based resource. The uplink data transmission in each slot follows the grant-free approach where each active device selects a channel randomly and independently from other devices. Moreover, the channel selection is uniform across the available channels. We consider a physical layer abstraction to the MAC layer in which a transmission is unsuccessful if two or more devices select the same channel in a given slot. On the other hand, if a device does not collide with any other devices, the transmission is successful, i.e., the BS can decode the message, and an acknowledgment is sent by the BS. In this paper, we consider the reactive grant-free approach in which a device retransmits its data packet only if it does not get an acknowledgment from the BS. Throughout this paper, we use the terms resource and channel interchangeably.

We consider the random component of latency $(\mathcal{L})$ introduced by the number of (re)transmissions for a successful transmission in the contention-based grant-free access mechanisms. It is assumed that each device can store a newly generated data packet until it is transmitted successfully. Thus, the average latency (μ_L) experienced by a data packet is defined as the average number of (re)transmissions required for successful transmission. In this paper, we use μ_L to evaluate the possibility of executing a particular IIoT application. Let $\mathcal{L}_{max}$ be the application-specific maximum affordable average latency. The current network conditions are considered feasible for the desired IIoT application as long as $\mu_L \leq \mathcal{L}_{max}$. On the contrary, when $\mu_L > \mathcal{L}_{max}$, we call this event an outage. The channel utilization (η_r) is defined as the average number of successful transmissions per channel per slot, where $0 \leq \eta_r \leq 1$.

Our objective in this paper is to design a grant-free access mechanism in which end devices can dynamically partition the resource set $\mathcal{C}$ to accommodate the delay-sensitive and delay-tolerant transmissions under the variable network load. For that purpose, end devices undergo a statistical learning-based *Network Exploration Phase* to predict the required parameters of the proposed scheme by employing a static resource allocation-based grant-free access scheme of [13] called the SRA scheme. This leads to an adaptive network access mechanism in which end devices can choose an appropriate grant-free access mechanism under a dynamic environment while avoiding additional feedback information from the BS. In the following section, we provide an overview of the proposed adaptive network access mechanism, followed by the design of grant-free access with dynamic resource allocation called the DRA scheme.

IV. Adaptive Network Access

Algorithm-1 explains different steps of the proposed adaptive network access to be executed by each device. As shown in Fig. 2, using the entire set $\mathcal{C}$, active devices perform delay-sensitive transmissions during the first R rounds of each window, called *Network Exploration Phase*. For this phase, active devices apply the SRA scheme in which every active device can have only one successful delay-sensitive transmission per round. As explained in Section-VI, the *Network Exploration Phase* enables end devices to predict the current network load ($\widehat{M}$), average latency in the delay-sensitive transmission $\left(\widehat{\mu}_{\text{ds}}^{(\text{SRA})}\right)$, and channel utilization ($\widehat{\eta}^{(\text{SRA})}$) offered by the SRA scheme.

On the other hand, as shown in Fig. 3, under the DRA scheme, channels available for delay-sensitive transmissions in the n^{th} slot of a round are given as: $\widehat{\mathcal{C}}_n^{(\text{ds})} = \left\{C_1, C_2, ..., C_{\widehat{K}_n}\right\}$, where $\left|\widehat{\mathcal{C}}_n^{(\text{ds})}\right| = \widehat{K}_n$ and $\widehat{\mathcal{C}}_1^{(\text{ds})} = \mathcal{C}$. While the channels available for delay-tolerant transmissions in the n^{th} slot of a round are given as: $\widehat{\mathcal{C}}_n^{(\text{dt})} = \left\{C_{\widehat{K}_n+1}, C_{\widehat{K}_n+2}, ..., C_K\right\}$, where $\left|\widehat{\mathcal{C}}_n^{(\text{dt})}\right| = K - \widehat{K}_n$ and $\widehat{\mathcal{C}}_1^{(\text{dt})} = \{\emptyset\}$. Thus, in each slot, the available channels are partitioned into two disjoint subsets, i.e., $\mathcal{C} = \widehat{\mathcal{C}}_n^{(\text{ds})} \cup \widehat{\mathcal{C}}_n^{(\text{dt})}$. The resource set $\mathcal{C}$ is partitioned in each slot of a round such that the probability of collision for delay-sensitive transmission remains almost the same as in the first slot. The *Network Exploration Phase* also enables the active devices to predict the vector parameter $\widehat{\Gamma} = \left[\widehat{K}_1, \widehat{K}_2, ..., \widehat{K}_N\right]$, average latency in the delay-sensitive transmission ($\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$), and the channel utilization ($\widehat{\eta}^{(\text{DRA})}$) offered by the DRA scheme.

After executing the *Network Exploration Phase* if the active devices observe that the current network load is at level where $\widehat{\mu}_{\text{ds}}^{(\text{SRA})} \leq \mathcal{L}_{\max}$, the devices can evaluate the possibility of transmitting delay-tolerant data in the remaining R rounds of the current window by employing the DRA scheme. If the DRA scheme does not cause an outage, active devices transmit delay-sensitive and delay-tolerant data in the next R rounds by employing the DRA scheme. Otherwise, the devices continue to transmit delay-sensitive data in the remaining R rounds of the current window following the SRA scheme.

On the other hand, when current network load is too high to meet the prescribed latency bound for the delay-sensitive data, the *Network Exploration Phase* yields $\widehat{\mu}_{\text{ds}}^{(\text{SRA})} > \mathcal{L}_{\max}$. Therefore, in case of an outage, the devices transmit only delay-tolerant data in each slot of the next R rounds using the multichannel slotted ALOHA (MSA) scheme. Under the MSA approach, in each slot, every active device selects a channel from $\mathcal{C}$ randomly and independently from other devices and

Algorithm 1 Adaptive Network Access

Require: $\mathcal{C}$, R, N, $\mathcal{L}_{\max}$, $W_{\max}$
Run Algorithm-2 to obtain $\widehat{\mu}_{\text{ds}}^{(\text{SRA})}$, $\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$, $\widehat{\Gamma}$
if $\left(\widehat{\mu}_{\text{ds}}^{(\text{SRA})} \leq \mathcal{L}_{\max}\right)$ **then**
  if $\left(\widehat{\mu}_{\text{ds}}^{(\text{DRA})} \leq \mathcal{L}_{\max}\right)$ **then**
    for $m = R+1$ to $2R$ **do**
      Delay-Tolerant-Flag $:= 0$
      for $n = 1$ to N **do**
        Determine $\widehat{\mathcal{C}}_n^{(\text{ds})}$ and $\widehat{\mathcal{C}}_n^{(\text{dt})}$
        if Delay-Tolerant-Flag $== 0$ **then**
          Select a channel randomly from $\widehat{\mathcal{C}}_n^{(\text{ds})}$
          Transmit/retransmit delay-sensitive data
          if success **then**
            Delay-Tolerant-Flag $:= 1$
          end if
        else
          Select a channel randomly from $\widehat{\mathcal{C}}_n^{(\text{dt})}$
          Transmit/retransmit delay-tolerant data
        end if
      end for
    end for
  else
    for $m = R+1$ to $2R$ **do**
      for $n = 1$ to N **do**
        Select a channel randomly from $\mathcal{C}$
        Transmit/retransmit delay-sensitive data
        if success **then**
          Stop transmitting in current round: $n := N+1$
        end if
      end for
    end for
  end if
else
  Transmit delay-tolerant data in each slot of the next R rounds using the MSA scheme
  Select W_b from $\{1, ..., W_{\max}\}$ randomly
  Skip next W_b windows
end if
return

(re)transmits the delay-tolerant data. Moreover, the devices perform a random-back-off where each device decides to skip next W_b windows independently from other active devices, where W_b is a random number selected from $\{1, ..., W_{\max}\}$ following uniform distribution. The parameter $W_{\max}$ shows the maximum number of windows a device can skip, and BS periodically broadcasts the parameter $W_{\max}$. The random-back-off strategy provides a fair opportunity for the newly active devices to perform the uplink data transmission.

In summary, the proposed adaptive network access mechanism enables end devices to choose an appropriate grant-free access scheme under a dynamic environment. The only feedback devices need from the BS is the outcome of their transmissions. The end devices use their transmission history to predict different parameters employing the statistical learning approach. Therefore, the proposed adaptive network access mechanism operates in a semi-distributed manner, and avoids excessive control signaling overheads. We first discuss the design of the proposed DRA scheme. Later, we explain the *Network Exploration Phase* to predict the required parameters for the SRA and DRA schemes.

V. Grant-Free Access with Dynamic Resource Allocation

The proposed grant-free access scheme with dynamic resource allocation aims to enable end devices to transmit the delay-sensitive and delay-tolerant data over non-overlapping groups of channels. As shown in Fig. 3, the delay-sensitive transmissions are prioritized over the delay-tolerant transmissions. Under this scheme, the devices after having a successful delay-sensitive transmission in the n^{th} slot perform delay-tolerant transmission in the remaining $N-n$ slots of a given round. Thus, each slot of a round, except Slot-1, can carry the both types of data. Active devices keep the history of their transmission outcomes. Let the random variable $B_{m,n}$ show the outcome of an indented device's transmission in the n^{th} slot of the m^{th} round for $m=R+1,...,2R$ defined as:

$$B_{m,n} := \begin{cases} 1, & \text{Successful transmission;} \\ 0, & \text{Collision with other device/s.} \end{cases} \tag{1}$$

The channels available for delay-sensitive transmissions in the n^{th} slot of the m^{th} round are given as: $\mathcal{C}^{(\text{ds})}_{m,n} = \{C_1, C_2, ..., C_{K_{m,n}}\}$, where $\left|\mathcal{C}^{(\text{ds})}_{m,n}\right| = K_{m,n}$ and $\mathcal{C}^{(\text{ds})}_{m,1} = \mathcal{C}$. While the channels reserved for delay-tolerant transmissions comes out to be: $\mathcal{C}^{(\text{dt})}_{m,n} = \{C_{K_{m,n}+1}, C_{K_{m,n}+2}, ..., C_K\}$, where $\left|\mathcal{C}^{(\text{dt})}_{m,n}\right| = K - K_{m,n}$ and $\mathcal{C}^{(\text{dt})}_{m,1} = \{\emptyset\}$. We define a vector parameter $\Gamma_m := [K_{m,1}, K_{m,2}, ..., K_{m,N}]$. Thus, in Slot-2 to Slot-N, the available channels are partitioned into two disjoint subsets, i.e., $\mathcal{C} = \mathcal{C}^{(ds)}_{m,n} \cup \mathcal{C}^{(\text{dt})}_{m,n}$. As explained below, different elements of Γ_m are determined such that the delay-sensitive transmissions can get the desired QoS in each slot of a round.

A. Device-Level Resource Partitioning:

In this subsection, we discuss the computation of Γ_m. Under the DRA scheme, the number of devices transmitting delay-sensitive data in the n^{th} slot of the m^{th} round is represented by $V_{m,n}$ and defined as:

$$V_{m,n} := \begin{cases} M & n=1, \\ M - \sum_{j=1}^{n-1} V'_{m,j} & n=2,3,...,N. \end{cases} \tag{2}$$

where $V'_{m,j}$ is the number of devices having successful delay-sensitive transmissions in the j^{th} slot. The probability that delay-sensitive transmission of an intended device faces a collision is given as:

$$\begin{aligned}\beta_{m,n} &:= \Pr\left(B_{m,n}=0 \mid \text{delay-sensitive transmission}\right) \\ &= 1 - \left(1 - \frac{1}{K_{m,n}}\right)^{V_{m,n}-1}, \quad n=1,2,...,N. \end{aligned} \tag{3}$$

Active devices begin to transmit delay-tolerant data after successful delay-sensitive transmission. The number of devices transmitting delay-tolerant data in each slot is given as:

$$W_{m,n} = \begin{cases} 0 & n=1, \\ \sum_{j=1}^{n-1} V'_{m,j} & n=2,3,...,N. \end{cases} \tag{4}$$

Given that an intended device has a successful delay-sensitive transmission, the probability that its delay-tolerant transmission remains unsuccessful in a given slot is computed as:

$$\begin{aligned}\gamma_{m,n} &:= \Pr\left(B_{m,n}=0 \mid \text{delay-tolerant transmission}\right) \\ &= 1 - \left(1 - \frac{1}{K-K_{m,n}}\right)^{W_{m,n}-1}, \quad n=2,3,...,N. \end{aligned} \tag{5}$$

While considering PHY-layer abstraction, different dynamic resource allocation schemes can provide different latency for delay-sensitive and delay-tolerant transmissions under the given set of channels, round size, and the number of active devices. In the proposed DRA scheme, the resource set $\mathcal{C}$ is partitioned in each slot such that all active devices can experience almost the same probability of collision for delay-sensitive transmissions throughout R rounds. We define α_1 as:

$$\alpha_1 := 1 - \left(1 - \frac{1}{K}\right)^{M-1}. \tag{6}$$

In this paper, we compute Γ_m by keeping the probability of collision in each slot equal to the probability of collision in the first slot, i.e., $\beta_{m,n} = \alpha_1,\ \forall n = 1,2,...,N$. Thus, the number of channels available for the delay-sensitive-data transmission in each slot comes out to be:

$$K_{m,n} = \begin{cases} K, & n=1; \\ \left\lceil \left\{ 1 - \left(1 - \frac{1}{K}\right)^{\frac{M-1}{V_{m,n}-1}} \right\}^{-1} \right\rceil, & n=2,3,...,N. \end{cases} \tag{7}$$

where $V_{m,n} > 1\ \forall n = 1,2,...,N$, and the ceiling function $\lceil x \rceil$ represents the smallest integer value larger than or equal to x. Due to the discretization in (7), we have $\beta_{m,n} \simeq \alpha_1,\ \forall n = 1,2,...,N$.

B. Channel Utilization:

The computation of channel utilization in the DRA scheme requires availability of average number of successful delay-sensitive and delay-tolerant transmissions per round. The average successful delay-sensitive transmissions per round is computed as:

$$\mathcal{S}^{(\text{DRA})}_{\text{ds}} := \frac{1}{R} \sum_{m=1}^{R} \sum_{n=1}^{N} V'_{m,n}. \tag{8}$$

While, the average successful delay-tolerant transmissions per round is computed as:

$$\mathcal{S}^{(\text{DRA})}_{\text{dt}} := \frac{1}{R} \sum_{m=1}^{R} \sum_{n=2}^{N} V''_{m,n}. \tag{9}$$

where $V''_{m,n}$ is the number of successful delay-tolerant transmissions in the n^{th} slot of the m^{th} round. Thus, the average number of successful transmissions per round comes out be:

$$\begin{aligned}\mathcal{S}^{(\mathrm{DRA})} &:= \mathcal{S}_{\mathrm{ds}}^{(\mathrm{DRA})} + \mathcal{S}_{\mathrm{dt}}^{(\mathrm{DRA})}, \\ &= \frac{1}{R}\sum_{m=1}^{R}\left(\sum_{n=1}^{N} V'_{m,n} + \sum_{n=2}^{N} V''_{m,n}\right). \end{aligned} \tag{10}$$

Finally, the channel utilization for the DRA scheme is computed by:

$$\eta^{(\mathrm{DRA})} := \frac{\mathcal{S}^{(\mathrm{DRA})}}{KN}. \tag{11}$$

By applying (10) in (11), $\eta^{(\mathrm{DRA})}$ comes out to be:

$$\eta^{(\mathrm{DRA})} = \frac{1}{KNR}\sum_{m=1}^{R}\left(\sum_{n=1}^{N} V'_{m,n} + \sum_{n=2}^{N} V''_{m,n}\right). \tag{12}$$

In Section VI-A, we explain the prediction of channel utilization and average latency offered by the DRA scheme which involves the prediction of Γ_m, $\beta_{m,n}$, and $\gamma_{m,n}$.

C. Probabilistic Analysis:

In this subsection, we perform the probabilistic analysis for the different events related to the DRA scheme. Under this scheme, in each round, a device can undergo one of the three mutually exclusive events: E_1, E_2, and E_3. The event E_1 is defined as the case where a device of interest, after successful delay-sensitive transmission, gets at least one successful delay-tolerant transmission. The probability for the occurrence of E_1 in the m^{th} round is computed as:

$$\mathcal{P}_m(E_1) := \sum_{n=1}^{N-1} \underbrace{(1-\beta_{m,n}) \prod_{\substack{j=1\\ n>1}}^{n-1} \beta_{m,j}}_{\mathrm{I}} \underbrace{\left(1 - \prod_{\substack{j=n\\ n>1}}^{N} \gamma_{m,j}\right)}_{\mathrm{II}}. \tag{13}$$

The expression I in (13) is the probability that the intended device gets successful delay-sensitive transmission in the n^{th} slot, while II in (13) shows the probability of having at least one successful delay-tolerant transmission in the remaining $N-n$ slots of the m^{th} round.

On the other hand, event E_2 arises when the intended device has one successful delay-sensitive transmission but does not get any successful delay-tolerant transmissions. The probability for the occurrence of E_2 in the m^{th} round is computed as:

$$\begin{aligned}\mathcal{P}_m(E_2) := &\sum_{n=1}^{N-1} \underbrace{(1-\beta_{m,n}) \prod_{\substack{j=1\\ n>1}}^{n-1} \beta_{m,j}}_{\mathrm{I}} \underbrace{\prod_{\substack{j=n\\ n>1}}^{N} \gamma_{m,j}}_{\mathrm{II}} \\ &+ \underbrace{(1-\beta_{m,N}) \prod_{n=1}^{N-1} \beta_{m,n}}_{\mathrm{III}}. \end{aligned} \tag{14}$$

Algorithm 2 Network Exploration Phase

Require: $\mathcal{C}$, R, N

Ensure: $\widehat{\mu}_{\mathrm{ds}}^{(\mathrm{SRA})}$, $\widehat{\mu}_{\mathrm{ds}}^{(\mathrm{DRA})}$, and $\widehat{\Gamma}$.

for $m = 1$ to R **do**
    for $n = 1$ to N **do**
        Select a channel randomly from $\mathcal{C}$
        Transmit/retransmit delay-sensitive data
        if success **then**
            if $n == 1$ **then**
                $A_{m,1} := 0$
            end if
            Stop transmitting in current round: $n := N + 1$
        else
            if $n == 1$ **then**
                $A_{m,1} := 1$
            end if
        end if
    end for
end for
Predict $\widehat{M}$ using (24)
Predict $\widehat{\mu}_{\mathrm{ds}}^{(\mathrm{SRA})}$ from Eq. (26)
Predict $\widehat{\Gamma}$ using (32)
Predict $\widehat{\mu}_{\mathrm{ds}}^{(\mathrm{DRA})}$ from Eq. (34)
return $\widehat{\mu}_{\mathrm{ds}}^{(\mathrm{SRA})}$, $\widehat{\mu}_{\mathrm{ds}}^{(\mathrm{DRA})}$, and $\widehat{\Gamma}$.

The expression I in (14) is the same as in (13), while II in (14) is the probability that after having a successful delay-sensitive transmission, all delay-tolerant transmissions from the intended device are unsuccessful in the remaining slots of that round. The expression III in (14) shows the probability of having successful delay-sensitive transmission in the last slot of the m^{th} round.

The event E_3 arises when the intended device does not get any successful delay-sensitive transmission and ultimately no successful delay-tolerant transmissions. The probability for the occurrence of E_3 in the m^{th} round is computed as:

$$\mathcal{P}_m(E_3) := \prod_{n=1}^{N} \beta_{m,n}. \tag{15}$$

The probabilities $\mathcal{P}_m(E_1)$, $\mathcal{P}_m(E_2)$, and $\mathcal{P}_m(E_3)$ satisfy $\mathcal{P}_m(E_1) + \mathcal{P}_m(E_2) + \mathcal{P}_m(E_3) = 1$.

By using $\{\beta_{m,n}\}_{n=1}^{N} \simeq \alpha_1$, $\forall m = 1, 2, ..., R$, the probabilities $\mathcal{P}_m(E_1)$, $\mathcal{P}_m(E_2)$, and $\mathcal{P}_m(E_3)$ are approximated as follows:

$$\mathcal{P}_m(E_1) \simeq \sum_{n=1}^{N-1} (1-\alpha_1)\, \alpha_1^{n-1} \left(1 - \prod_{\substack{j=n\\ n>1}}^{N} \gamma_{m,j}\right). \tag{16}$$

$$\mathcal{P}_m(E_2) \simeq \sum_{n=1}^{N-1} (1-\alpha_1)\, \alpha_1^{n-1} \prod_{\substack{j=n\\ n>1}}^{N} \gamma_{m,j} + (1-\alpha_1)\, \alpha_{m,1}^{N-1}. \tag{17}$$

$$\mathcal{P}_m(E_3) \simeq \alpha_1^{N}. \tag{18}$$

The average probability that a device of interest has a successful delay-sensitive transmission, whether it gets any delay-tolerant transmissions or not, is given as:

$$\mathcal{P}_{\text{ds}}^{(\text{DRA})} := \frac{1}{R}\sum_{m=1}^{R}\left(\mathcal{P}_m(E_1)+\mathcal{P}_m(E_2)\right) = 1-\frac{1}{R}\sum_{m=1}^{R}\mathcal{P}_m(E_3). \tag{19}$$

By using (18) in (19), $\mathcal{P}_{\text{ds}}^{(\text{DRA})}$ is computed as:

$$\mathcal{P}_{\text{ds}}^{(\text{DRA})} \simeq 1-\alpha_1^N. \tag{20}$$

While the average probability that a device of interest will have at least one delay-tolerant transmission, denoted by $\mathcal{P}_{\text{dt}}^{(\text{DRA})}$, is given by:

$$\mathcal{P}_{\text{dt}}^{(\text{DRA})} := \frac{1}{R}\sum_{m=1}^{R}\mathcal{P}_m(E_1). \tag{21}$$

By applying (16) in (21), $\mathcal{P}_{\text{dt}}^{(\text{DRA})}$ is computed as:

$$\mathcal{P}_{\text{dt}}^{(\text{DRA})} \simeq \frac{1}{R}\sum_{m=1}^{R}\sum_{n=1}^{N-1}\left(1-\alpha_1\right)\alpha_1^{n-1}\left(1-\prod_{\substack{j=n\\n>1}}^{N}\gamma_{m,j}\right). \tag{22}$$

In the following section, we discuss the prediction of different parameters related to the SRA and DRA schemes.

VI. Network Exploration Phase

As shown in Fig. 2, during *Network Exploration Phase*, all the active devices perform delay-sensitive transmissions following the SRA scheme of [13]. Under this scheme, each active device can have only one successful transmission per round. The probability that an intended device's transmission faces a collision with one or more other transmitting devices in the n^{th} slot of the m^{th} round is denoted by $\alpha_{m,n}$ and computed by:

$$\alpha_{m,n} = 1-\left(1-\frac{1}{K}\right)^{M_{m,n}-1} \tag{23}$$

where $M_{m,n}$ is the number of transmitting devices in the n^{th} slot of the m^{th} round, and we have $M = M_{m,1}\ \forall m = 1,2,...,2R$. Thus for $n=1$, (6) and (23) generate the same expressions.

During the *Network Exploration Phase*, all active devices maintain the history of their transmissions outcomes in a vector: $\mathbf{h} = [A_{1,1}, A_{2,1}, ..., A_{R,1}]$, where each element of $\mathbf{h}$ is a Bernoulli random variable i.e, $A_{m,1} \in \{0,1\}$, $\forall m = 1,2,...,R$. The random variable $A_{m,1} = 1$ if the intended device's transmission faces a collision with one or more other active devices in Slot-1 of the m^{th} round.

Finally using the statistical learning approach, end devices are enabled to predict the number of transmitting devices ($\widehat{M}_n$) and the corresponding collision probability ($\widehat{\alpha}_n$) in different slots of each round of the *Network Exploration Phase* as follows [13]:

$$\widehat{M}_n = \begin{cases} 1+\dfrac{\ln\left(1-\frac{1}{R}\sum_{m=1}^{R}A_{m,1}\right)}{\ln\left(\frac{K-1}{K}\right)}, & n=1; \\ \left[1+\dfrac{\ln\left(1-\frac{1}{R}\sum_{m=1}^{R}A_{m,1}\right)}{\ln\left(\frac{K-1}{K}\right)}\right]\prod_{j=1}^{n-1}\widehat{\alpha}_j, & n=2,3,...,N. \end{cases} \tag{24}$$

$$\widehat{\alpha}_n = \begin{cases} \frac{1}{R}\sum_{m=1}^{R}A_{m,1}, & n=1; \\ 1-\left(1-\frac{1}{K}\right)^{\widehat{M}_n-1}, & n=2,3,...,N. \end{cases} \tag{25}$$

where $\widehat{M}_n$ and $\widehat{\alpha}_n$ are the predictions of $M_{m,n}$ and $\alpha_{m,n}\ \forall m=1,2,...,R$, respectively, and $\widehat{M}=\widehat{M}_1$.

Consequently, active devices predict the average latency of delay-sensitive transmissions in the SRA scheme using [13, Eq. (30)] as follows:

$$\widehat{\mu}_{\text{ds}}^{(\text{SRA})} := \frac{1}{1-\prod_{n=1}^{N}\widehat{\alpha}_n}\left(1+\sum_{\substack{z=1\\N>1}}^{N-1}\prod_{j=1}^{z}\widehat{\alpha}_j\right). \tag{26}$$

Along with $\widehat{\mu}_{\text{ds}}^{(\text{SRA})}$, end devices predict the channel utilization ($\widehat{\eta}^{(\text{SRA})}$) of the SRA scheme as follows:

$$\widehat{\eta}^{(\text{SRA})} := \frac{\widehat{M}\left(1-\prod_{n=1}^{N}\widehat{\alpha}_n\right)}{KN}. \tag{27}$$

where $\widehat{M}\left(1-\prod_{n=1}^{N}\widehat{\alpha}_n\right)$ is the prediction of average successful devices per round in the SRA scheme.

A. Prediction of Average Latency in the DRA Scheme

If the current network conditions are such that $\widehat{\mu}_{\text{ds}}^{(\text{SRA})} \leq \mathcal{L}_{max}$, end devices can further explore the possibility of transmitting delay-tolerant data followed by the delay-sensitive transmissions using the DRA scheme. For this purpose, end devices need to predict the vector parameter $\widehat{\Gamma}$ and the average latency ($\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$) offered by the proposed DRA scheme. The end devices apply their knowledge of the current network load prediction, obtained from (24), in (7) and predict the number of channels reserved for delay-sensitive transmissions in the n^{th} slot of a round as follows:

$$\widehat{K}_n = \begin{cases} K, & n=1; \\ \left\lceil\left\{1-\left(1-\dfrac{1}{K}\right)^{\frac{\widehat{M}-1}{\widehat{V}_n-1}}\right\}^{-1}\right\rceil, & n=2,3,...,N. \end{cases} \tag{28}$$

where $\widehat{V}_n$ is the prediction of $V_{m,n}\ \forall m=1,2,...,R$ and it is computed by:

$$\widehat{V}_n := \widehat{V}_1\prod_{\substack{j=1\\n>1}}^{n-1}\widehat{\beta}_j, \quad n=1,2,...,N. \tag{29}$$

where $\widehat{V}_1 = \widehat{M}$ and $\widehat{\beta}_j$ is the prediction of $\beta_{m,j}$, $\forall m = 1, 2, ..., R$. Since the proposed DRA scheme aims to keep the probability of collision for delay-sensitive transmissions almost the same in all slots of a round while accommodating the transmission of delay-tolerant data; by using $\widehat{\beta}_n \simeq \widehat{\alpha}_1$, $\forall n = 1, 2, ..., N$; $\widehat{V}_n$ is approximated as follows:

$$\widehat{V}_n \simeq \widehat{M}\widehat{\alpha}_1^{n-1} \tag{30}$$

By substituting (24) and (25) in (30), $\widehat{V}_n$ is computed as:

$$\widehat{V}_n \simeq \left\{1 + \frac{\ln\left(1 - \frac{1}{R}\sum_{m=1}^{R} A_{m,1}\right)}{\ln\left(\frac{K-1}{K}\right)}\right\}\left(\frac{1}{R}\sum_{m=1}^{R} A_{m,1}\right)^{n-1}. \tag{31}$$

Finally, by substituting (30) in (28), different elements in $\widehat{\Gamma}$ are computed by:

$$\widehat{K}_n := \begin{cases} K, & n = 1; \\ \left\lceil \left\{1 - \left(1 - \frac{1}{K}\right)^{\frac{\widehat{M}-1}{\widehat{M}\widehat{\alpha}_1^{n-1}-1}}\right\}^{-1} \right\rceil, & n = 2, 3, ..., N. \end{cases} \tag{32}$$

Thus by using (31) and (32) in (3) end devices can compute $\widehat{\beta}_n$ as follows:

$$\widehat{\beta}_n = 1 - \left(1 - \frac{1}{\widehat{K}_n}\right)^{\widehat{V}_n - 1} \qquad n = 1, 2, ..., N. \tag{33}$$

Once end devices have computed $\widehat{\Gamma}$ and $\widehat{\beta}_n$, the average latency in delay-sensitive transmission for the DRA scheme is computed as:

$$\widehat{\mu}_{\text{ds}}^{(\text{DRA})} \simeq \frac{R}{R - \sum_{m=1}^{R} A_{m,1}}. \tag{34}$$

Please see Appendix-A for the derivation of (34).

After completing the exploration phase, end devices compute, $\widehat{\mu}_{\text{ds}}^{(\text{SRA})}$ and $\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$ through (26) and (34), respectively. If the average latency in the delay-sensitive transmission offered by the SRA scheme does not exceed the latency bound, i.e., $\widehat{\mu}_{\text{ds}}^{(\text{SRA})} < \mathcal{L}_{max}$, the devices can evaluate the possibility of accommodating the delay-tolerant transmissions in the next R rounds. Under the given current network load, if the DRA scheme follows the latency bound i.e., $\widehat{\mu}_{\text{ds}}^{(\text{DRA})} < \mathcal{L}_{max}$, the devices select the DRA scheme for the next R rounds. Otherwise, the devices continue using the SRA scheme during the second half of the current window. Thus, the possibility of executing the DRA is determined before executing it with the help of an exploration phase. Moreover, as demonstrated in the following section, allowing the transmission of delay-tolerant data increases channel utilization. The relationship between $\widehat{\mu}_{\text{ds}}^{(\text{SRA})}$ and $\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$ is described in Proposition 1.

Proposition 1. *The average latency of a successful delay-sensitive transmission with static resource allocation under the SRA and with dynamic resource allocation under the DRA are related as:*

$$\widehat{\mu}_{ds}^{(SRA)} \leq \widehat{\mu}_{ds}^{(DRA)}. \tag{35}$$

Proof of Proposition-1 is given in Appendix-B.

1) Latency for delay-tolerant transmissions: Although the proposed adaptive network access mechanism requires end devices to predict the average latency for delay-sensitive transmissions to adapt to the network dynamics, we also analyze the latency experienced by delay-tolerant data. For that purpose, we recall that each active device begins to transmit delay-tolerant data packets after having a successful delay-sensitive transmission. Therefore, the latency experienced by an intended device's first delay-tolerant data packet is composed of the delay introduced by the successful delay-sensitive transmission and the number of (re)transmissions performed to transmit the delay-tolerant packet successfully.

Active devices can use their transmission history from the second half of the current window to predict the average latency for delay-tolerant transmissions. Let t_m denote the slot number in the m^{th} round in which the delay-sensitive packet from the intended is transmitted successfully, where $1 \leq t_m \leq N$. The intended device then performs $N - t_m$ delay-tolerant-transmissions in the m^{th} round. Thus, the average latency for a delay-tolerant transmission can be predicted as follows:

$$\widehat{\mu}_{\text{dt}}^{(\text{DRA})} := \widehat{\mu}_{\text{ds}}^{(\text{DRA})} + \frac{\sum_{m=R+1}^{2R}(N - t_m)}{\sum_{m=R+1}^{2R}\sum_{x=t_m+1}^{N} B_{m,x}}. \tag{36}$$

In (36), the term $\sum_{m=R+1}^{2R}(N - t_m)$ provides the total number of delay tolerant transmissions performed by the intended device, while the term $\sum_{m=R+1}^{2R}\sum_{x=t_m+1}^{N} B_{m,x}$ provides the number of successful delay-tolerant transmissions obtained by the intended device in R rounds. Thus the term $\frac{\sum_{m=R+1}^{2R}(N-t_m)}{\sum_{m=R+1}^{2R}\sum_{x=t_m+1}^{N} B_{m,x}}$ shows the average number of (re)transmissions per successful delay-tolerant transmission. By substituting (34) in (36), $\widehat{\mu}_{\text{dt}}^{(\text{DRA})}$ is computed by:

$$\widehat{\mu}_{\text{dt}}^{(\text{DRA})} = \frac{R}{R - \sum_{m=1}^{R} A_{m,1}} + \frac{\sum_{m=R+1}^{2R}(N - t_m)}{\sum_{m=R+1}^{2R}\sum_{x=t_m+1}^{N} B_{m,x}} \tag{37}$$

Thus we can see from (37) that end devices can use the transmission history of each window to predict the average latency of successful delay-tolerant transmission.

B. Prediction of Channel Utilization in the DRA Scheme:

In order to predict the channel utilization, end devices need to predict the average successful delay-sensitive and delay-tolerant transmissions per round defined in (8) and (9), respectively. Once end devices have computed $\widehat{\Gamma}$, they can predict the average number of successful delay-sensitive transmissions per round as follows:

$$\widehat{\mathcal{S}}_{\text{ds}}^{(\text{DRA})} := \sum_{n=1}^{N} \widehat{V}_n\left(1 - \widehat{\beta}_n\right). \tag{38}$$

On the other hand, the devices can predict the number of delay-tolerant transmissions in each slot as follows:

$$\widehat{W}_n := \sum_{i=1}^{n-1} \widehat{V}_i\left(1 - \widehat{\beta}_i\right), \quad n = 2, 3, ..., N. \tag{39}$$

By applying (29) in (39), $\widehat{W}_n$ is computed as:

$$\widehat{W}_n = \widehat{M}\left(1 - \prod_{i=1}^{n-1} \widehat{\beta}_i\right). \tag{40}$$

The number of channels reserved for delay-tolerant transmissions in each slot comes out be: $K - \widehat{K}_n,\ \forall n = 2, 3, ..., N$. The probability that the transmission of an intended device collides with at least one of the other devices transmitting their delay-tolerant data in the n^{th} slot, is predicted as:

$$\widehat{\gamma}_n = 1 - \left(1 - \frac{1}{K - \widehat{K}_n}\right)^{\widehat{W}_n - 1}, \quad n = 2, 3, ..., N. \tag{41}$$

Thus the average number of successful delay-tolerant transmissions per round is predicted as:

$$\widehat{\mathcal{S}}_{\text{dt}}^{(\text{DRA})} := \sum_{n=2}^{N} \widehat{W}_n \left(1 - \widehat{\gamma}_n\right). \tag{42}$$

The channel utilization in the DRA is predicted as follows:

$$\begin{aligned}
\widehat{\eta}^{(\text{DRA})} &:= \frac{\widehat{\mathcal{S}}_{\text{ds}}^{(\text{DRA})} + \widehat{\mathcal{S}}_{\text{dt}}^{(\text{DRA})}}{KN}, \\
&= \frac{1}{KN}\left\{\sum_{n=1}^{N}\left(\widehat{V}_1 \prod_{\substack{j=1\\ n>1}}^{n-1} \widehat{\beta}_j\right)\left(1 - \widehat{\beta}_n\right)\right. \\
&\qquad \left. + \sum_{n=2}^{N} \widehat{M}\left(1 - \prod_{i=1}^{n-1} \widehat{\beta}_i\right)\left(1 - \widehat{\gamma}_n\right)\right\}.
\end{aligned} \tag{43}$$

Finally, end devices can approximate $\widehat{\eta}^{(\text{DRA})}$ using $\widehat{\beta}_n \simeq \widehat{\alpha}_1,\ \forall n = 1, 2, ..., N$ in (43) as follows:

$$\begin{aligned}
\widehat{\eta}^{(\text{DRA})} \simeq \frac{1}{KN}&\left\{\widehat{M}\sum_{n=1}^{N} \widehat{\alpha}_1^{n-1}\left(1 - \widehat{\alpha}_1\right) + \right. \\
&\left. \widehat{M}\sum_{n=2}^{N}\left(1 - \widehat{\alpha}_1^{n-1}\right)\left(1 - \widehat{\gamma}_n\right)\right\}
\end{aligned} \tag{44}$$

(44) is simplified to:

$$\begin{aligned}
\widehat{\eta}^{(\text{DRA})} \simeq \frac{\widehat{M}}{KN}&\left[N - \sum_{n=1}^{N} \widehat{\alpha}_1^{n} - \sum_{n=2}^{N}\left(1 - \widehat{\alpha}_1^{n-1}\right)\right. \\
&\left. \times\left\{1 - \left(1 - \frac{1}{K - \widehat{K}_n}\right)^{\widehat{M}\left(1 - \widehat{\alpha}_1^{n-1}\right) - 1}\right\}\right].
\end{aligned} \tag{45}$$

Since computation of $\widehat{K}_n$ also depends on $\widehat{M}$ and $\widehat{\alpha}_1$, we can readily show that for the given values of N, K, and R, the channel utilization of the DRA in (45) becomes a function of $\widehat{M}$ and $\widehat{\alpha}_1$. While the closed-form expressions of both $\widehat{M}$ and $\widehat{\alpha}_1$ use the on-device transmission history. Therefore, the statistical learning-based network exploration enables end devices to predict each scheme's average latency and channel utilization in closed form.

Accuracy in predicting average latency and channel utilization for the SRA and DRA schemes is evaluated using the mean square error (MSE) criterion. The MSE in the prediction of $\eta^{(\text{DRA})}$ is denoted by $\text{MSE}_{\eta}^{(\text{DRA})}$ and defined as:

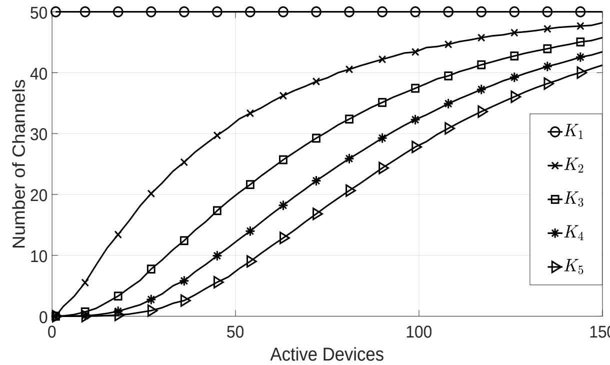

Fig. 4. Average number of channels available for delay-sensitive transmissions in each slot with $K = 50$ and $N = 5$.

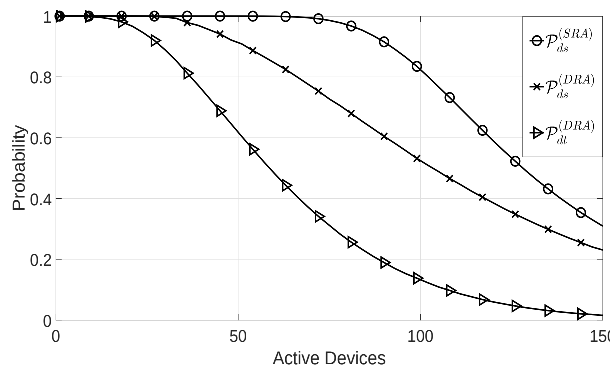

Fig. 5. Average probabilities of successful delay-sensitive and delay-tolerant transmissions per round with $K = 50$ and $N = 5$.

$$\text{MSE}_{\eta}^{(\text{DRA})} := \text{E}\left[\left(\eta^{(\text{DRA})} - \widehat{\eta}^{(\text{DRA})}\right)^2\right]. \tag{46}$$

The MSE in the prediction of $\mu_{\text{ds}}^{(\text{DRA})}$ is denoted by $\text{MSE}_{\mu}^{(\text{DRA})}$ and defined as:

$$\text{MSE}_{\mu}^{(\text{DRA})} := \text{E}\left[\left(\mu_{\text{ds}}^{(\text{DRA})} - \widehat{\mu}_{\text{ds}}^{(\text{DRA})}\right)^2\right]. \tag{47}$$

The MSE in predicting different parameters related to the SRA scheme decreases as the size of transmission history is increased [13]. In the following section we demonstrate that the MSE in the prediction of average latency and channel utilization for the DRA scheme also decreases as the size of transmission history is increased.

The complexity of Algorithm-2 is described in terms of the time required to predict desired parameters. The *Network Exploration Phase* is spanned over R rounds, where each round is composed of N slots. Therefore, for a given maximum acceptable MSE in predicting each parameter, the optimal number of rounds in the *Network Exploration Phase* can be computed by following [13]. Furthermore, Algorithm-2 allows each active device to have one successful delay-sensitive transmission per round in the *Network Exploration Phase*. In addition to that, Algorithm-2 operates in an online manner as devices use their transmission history to predict desired parameters by employing closed-form expressions (26), (28),

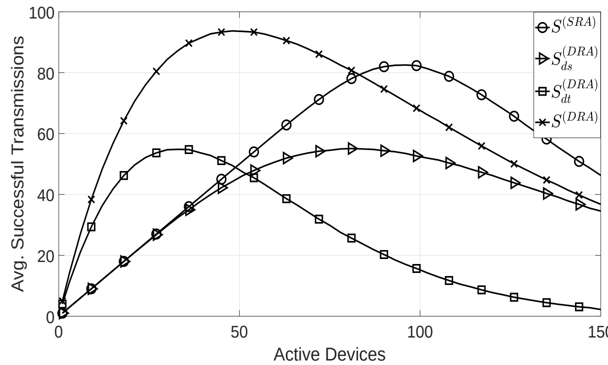

Fig. 6. Average successful delay-sensitive and delay tolerant transmissions per round with $K = 50$ and $N = 5$.

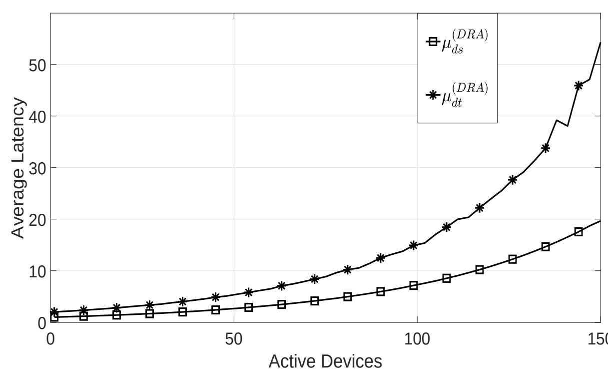

Fig. 7. Average latency for a successful delay-sensitive and delay-tolerant transmissions with $K = 50$ and $N = 5$.

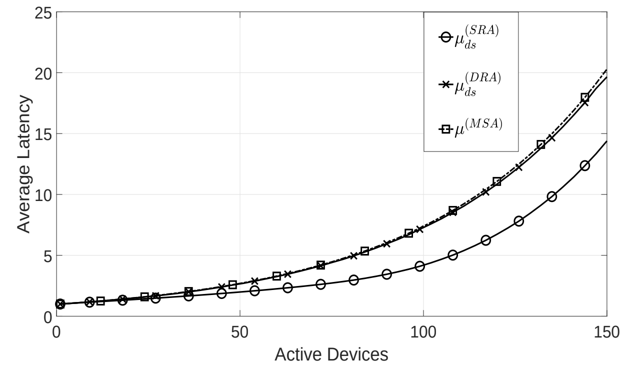

Fig. 8. Performance comparison in terms of the average latency with $K = 50$ and $N = 5$.

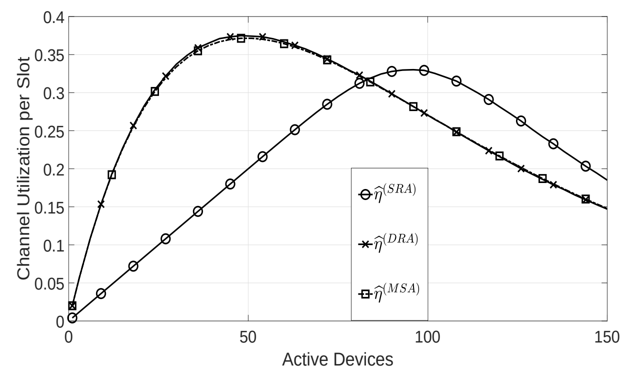

Fig. 9. Performance comparison in terms of the channel utilization with $K = 50$ and $N = 5$.

(34), and (36). On the other hand, if the current network load allows, end devices choose the DRA scheme and perform transmission in each slot of the next R round, as explained in Section-V. These features of the proposed adaptive network access make it suitable for IIoT networks where devices have limited computation power.

Moreover, in this paper, all active devices undergo network exploration in each window to predict the current network load. However, the frequency of executing the *Network Exploration Phase* depends on how frequently the network load varies. In our future work, we intend to incorporate the nature of network load variation to determine the frequency of executing the *Network Exploration Phase*.

VII. Simulation Results and Discussion

This section presents simulation results of the proposed adaptive network access mechanism. Along with analyzing different parameters related to the proposed DRA scheme, we also compare average latency and channel utilization of the DRA scheme with the SRA and conventional MSA systems. We apply the Monte Carlo simulation method with $K = 50$ channels and $N = 5$ slots per round for several active devices, while $R = 10,000$ independent rounds are used to analyze the behavior of different parameters against the varying network load. The average number of channels (K_n), computed by $K_n = \frac{1}{R}\sum_{m=1}^{R} K_{m,n}$, reserved for delay-sensitive transmissions in each slot are plotted in Fig. 4. The average number of channels allocated for delay-tolerant transmissions can be computed as: $K - K_n$. It is observed that for the given network load, K_n decreases in each slot; however, it approaches K when the number of active devices increases.

Fig. 5 plots $\mathcal{P}_{\text{ds}}^{(\text{DRA})}$ and $\mathcal{P}_{\text{dt}}^{(\text{DRA})}$ for the DRA scheme. In Fig. 5, we have also plotted the average probability of a successful delay-sensitive transmission per round ($\mathcal{P}_{\text{ds}}^{(\text{SRA})}$) in the SRA scheme defined by [13, Eq. (11)]. We can observe that $\mathcal{P}_{\text{ds}}^{(\text{SRA})}$ and $\mathcal{P}_{\text{ds}}^{(\text{DRA})}$ are similar for the lower network load. However, $\mathcal{P}_{\text{ds}}^{(\text{SRA})}$ is higher than $\mathcal{P}_{\text{ds}}^{(\text{DRA})}$ when the number of active devices becomes large. On the other hand, $\mathcal{P}_{\text{dt}}^{(\text{DRA})}$ follows $\mathcal{P}_{\text{ds}}^{(\text{SRA})}$ and $\mathcal{P}_{\text{ds}}^{(\text{DRA})}$ when the number of active devices is very small, and becomes significantly less than $\mathcal{P}_{\text{ds}}^{(\text{SRA})}$ and $\mathcal{P}_{\text{ds}}^{(\text{DRA})}$ when network load increases.

Fig. 6 plots $\mathcal{S}_{\text{ds}}^{(\text{DRA})}$, $\mathcal{S}_{\text{dt}}^{(\text{DRA})}$, and $\mathcal{S}^{(\text{DRA})}$ defined in (8), (9), and (10), respectively. Moreover, Fig. 6 also plots the average number of successful delay-sensitive transmissions ($\mathcal{S}^{(\text{SRA})}$) in the SRA scheme computed through [13, Eq. (2)]. It is observed that $\mathcal{S}^{(\text{SRA})}$ is similar to $\mathcal{S}_{\text{ds}}^{(\text{DRA})}$ for low to moderate network load; however, $\mathcal{S}^{(\text{SRA})}$ becomes higher than $\mathcal{S}_{\text{ds}}^{(\text{DRA})}$ when the number of active devices increases. On the other hand, $\mathcal{S}_{\text{dt}}^{(\text{DRA})}$ is higher than both $\mathcal{S}_{\text{ds}}^{(\text{SRA})}$ and $\mathcal{S}_{\text{ds}}^{(\text{DRA})}$ under low to moderate network load. However, $\mathcal{S}_{\text{dt}}^{(\text{DRA})}$ decreases rapidly when the number of active devices gets larger. Furthermore, $\mathcal{S}^{(\text{DRA})}$ is higher than $\mathcal{S}^{(\text{SRA})}$ for relatively lower network load; however for moderate to high network load $\mathcal{S}^{(\text{SRA})}$ is greater than $\mathcal{S}^{(\text{DRA})}$. Finally, both $\mathcal{S}_{\text{ds}}^{(\text{SRA})}$ and $\mathcal{S}^{(\text{DRA})}$ approach zero when M becomes too large.

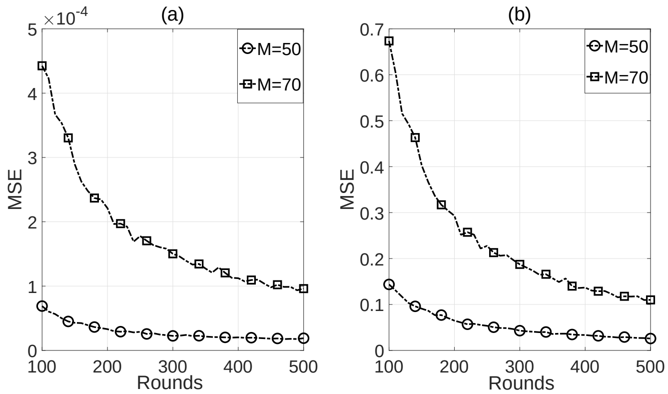

Fig. 10. Prediction accuracy against number of rounds with $K = 50$ and $N = 5$: (a) MSE in device-level prediction of $\eta^{(\text{DRA})}$. (b) MSE in device-level prediction of $\mu_{\text{ds}}^{(\text{DRA})}$.

Fig. 7 plots the average latency for a successful delay-sensitive and delay-tolerant transmissions under the DRA scheme denoted by $\mu_{\text{ds}}^{(\text{DRA})}$, and $\mu_{\text{dt}}^{(\text{DRA})}$ respectively. We compute $\mu_{\text{ds}}^{(\text{DRA})}$ from (53) while replacing $\widehat{\beta}_n$ with $\frac{1}{R}\sum_{m=1}^{R}\beta_{m,n}$. On the other hand, we compute $\mu_{\text{dt}}^{(\text{DRA})}$ from (36) while replacing $\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$ with $\mu_{\text{ds}}^{(\text{DRA})}$. Since delay-tolerant data is transmitted after successful delay-sensitive transmission, it is observed that $\mu_{\text{dt}}^{(\text{DRA})}$ is significantly higher than $\mu_{\text{ds}}^{(\text{DRA})}$ for the whole range of the number of active devices. Fig. 8 plots the average latency for a successful delay-sensitive transmission offered by the SRA and DRA schemes. We compute average latency in the SRA scheme, denoted by $\mu_{\text{ds}}^{(\text{SRA})}$, using [13, Eq. (37)]. It is observed that $\mu_{\text{ds}}^{(\text{SRA})}$ and $\mu_{\text{ds}}^{(\text{DRA})}$ show similar behavior for low network load. However, as the number of active devices increases, accommodating delay-tolerant transmissions in the grant-free access with dynamic resource allocation results in higher latency. Therefore, the SRA scheme is more suitable under higher network load for delay-sensitive transmissions.

Fig. 9 demonstrates channel utilization in the DRA and SRA schemes denoted by $\eta^{(\text{DRA})}$ and $\eta^{(\text{SRA})}$, respectively. Where $\eta^{(\text{DRA})}$ is defined in (12), and $\eta^{(\text{SRA})} := \frac{\mathcal{S}^{(\text{SRA})}}{KN}$. It is observed that $\eta^{(\text{DRA})}$ is higher than $\eta^{(\text{SRA})}$ for a relatively lower network load. While, for moderate to higher network load, $\eta^{(\text{SRA})}$ becomes larger than $\eta^{(\text{DRA})}$. Both $\eta^{(\text{DRA})}$ and $\eta^{(\text{SRA})}$ depict similar asymptotic behaviour against M. Fig. 10(a) and Fig. 10(b) plot the MSE in device-level prediction of $\eta^{(\text{DRA})}$ and $\mu_{\text{ds}}^{(\text{DRA})}$, respectively. It is observed that under a given network load, the MSEs in predicting channel utilization and average latency decrease as the number of rounds in the *Network Exploration Phase* increases.

A. Performance Comparison

We compare the performance of the proposed adaptive network access with the conventional MSA systems. In the MSA systems, every active device selects a channel randomly and independent in each slot without executing any learning strategy. The average latency in the MSA system $\left(\mu^{(\text{MSA})}\right)$ is defined as: $\mu^{(\text{MSA})} := \frac{1}{\left(1-\frac{1}{K}\right)^{M-1}}$. Since the probability of collision in each slot is almost constant under the DRA scheme, Fig. 8 shows that $\mu_{\text{ds}}^{(\text{DRA})}$ is similar to $\mu^{(\text{MSA})}$. We also plot the channel utilization of the MSA system defined as $\eta^{(\text{MSA})} := \frac{M}{K}\left(1-\frac{1}{K}\right)^{M-1}$. As demonstrated in Fig. 9, the channel utilization in the DRA scheme is similar to that of the MSA systems.

It is noteworthy that channel utilization in the DRA scheme is based upon dynamic resource allocation of the available resources in each slot for delay-sensitive and delay-tolerant transmission. On the contrary, channel utilization in the MSA systems does not incorporate dynamic resource allocation. Moreover, the proposed grant-free access scheme separates the devices transmitting delay-sensitive data from those transmitting delay-tolerant data in each slot. This separation is based on the subset of channels used for each transmission type. Therefore, allowing end devices to share $\widehat{\Gamma}$ with the BS, the DRA scheme can enable the BS to distinguish between delay-sensitive and delay-tolerant transmissions. The BS in turn can use this knowledge to optimize the number of channels in $\mathcal{C}$.

Moreover, the above discussion highlights that DRA and SRA schemes have advantages under different network loads. Thus the proposed adaptive network access enables end devices to choose the appropriate scheme for a given network load. Therefore, in contrast to the conventional MSA-based systems where devices do not employ any learning strategy, the proposed adaptive network access enables the end devices to adapt to the network dynamics.

VIII. Conclusion and Future Work

Uplink-dominant IIoT networks operate under time-varying network load and generate data with diverse QoS requirements. Therefore, adaptive network access mechanisms are essential to utilize available shared radio resources efficiently. This paper uses a statistical learning approach to enable end devices to perform delay-sensitive and delay-tolerant transmissions over dynamically partitioned resources in a grant-free manner. Consequently, the proposed DRA scheme accommodates delay-tolerant transmissions under favorable network conditions while delay-sensitive transmissions follow the prescribed latency bound. Moreover, devices perform random back-off in case of an outage providing fairness to newly active devices in accessing shared radio resources. The resultant adaptive network access mechanism enables end devices to choose an appropriate network access strategy under time-varying network load. In contrast to the existing centralized network access methods, the proposed mechanism operates in a semi-distributed manner and avoids excessive feedback overheads while adaptively managing two different types of network traffic.

This paper classified data generated in IIoT networks into delay-sensitive and delay-tolerant types. Moreover, the proposed adaptive network access considers the same latency bound for all devices in an IIoT network. Our future research aims to design an adaptive network access mechanism suitable for heterogeneous IIoT networks by accommodating multiple latency requirements.

Appendix A
Derivation of (34)

Since each active device can have only one successful delay-sensitive transmission per round in the DRA scheme, the devices can predict the average latency for their delay-sensitive transmissions as follows [13]:

$$\widehat{\mu}_{\text{ds}}^{(\text{DRA})} := \frac{1}{1-\prod_{n=1}^{N}\widehat{\beta}_n}\left(1+\sum_{\substack{z=1\\N>1}}^{N-1}\prod_{j=1}^{z}\widehat{\beta}_j\right). \tag{48}$$

By using $\widehat{\beta}_n \simeq \widehat{\alpha}_1,\ \forall n = 1, 2, ..., N$ in (48), the devices can approximate $\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$ as follows:

$$\widehat{\mu}_{\text{ds}}^{(\text{DRA})} \simeq \frac{1+\sum_{n=1}^{N-1}\widehat{\alpha}_1^n}{1-\widehat{\alpha}_1^N} \tag{49}$$

The right hand side of (49) can be expanded as:

$$\begin{aligned}\widehat{\mu}_{\text{ds}}^{(\text{DRA})} &\simeq \frac{1+\widehat{\alpha}_1^N-\widehat{\alpha}_1^N+(1-\widehat{\alpha}_1)^{-1}\sum_{n=1}^{N-1}(1-\widehat{\alpha}_1)\,\widehat{\alpha}_1^n}{1-\widehat{\alpha}_1^N}\\ &= 1+\frac{\widehat{\alpha}_1(1-\widehat{\alpha}_1)^{-1}\sum_{n=1}^{N}(1-\widehat{\alpha}_1)\,\widehat{\alpha}_1^{n-1}}{1-\widehat{\alpha}_1^N}.\end{aligned} \tag{50}$$

where

$$\sum_{n=1}^{N}(1-\widehat{\alpha}_1)\,\widehat{\alpha}_1^{n-1} = 1-\widehat{\alpha}_1+\widehat{\alpha}_1-\widehat{\alpha}_1^2+...+\widehat{\alpha}_1^{N-1}-\widehat{\alpha}_1^N. \tag{51}$$

All the terms except the 1st and the last terms at the right hand side of (51) are cancelled out and (51) is reduced to:

$$\sum_{n=1}^{N}(1-\widehat{\alpha}_1)\,\widehat{\alpha}_1^{n-1} = 1-\widehat{\alpha}_1^N. \tag{52}$$

By substituting (52) in (50), we get:

$$\widehat{\mu}_{\text{ds}}^{(\text{DRA})} \simeq \frac{1}{1-\widehat{\alpha}_1}. \tag{53}$$

By substituting $\widehat{\alpha}_1$, obtained through (25), in (53), $\widehat{\mu}_{\text{ds}}^{(\text{DRA})}$ gets the following form:

$$\widehat{\mu}_{\text{ds}}^{(\text{DRA})} \simeq \frac{R}{R-\sum_{m=1}^{R}A_{m,1}}. \tag{54}$$

This completes the derivation of (34).

Appendix B
Proof of Proposition 1

For the grant-free based restricted transmission strategy of the SRA scheme we have $0 \le \widehat{\alpha}_1 \le \widehat{\alpha}_2 \le ... \le \widehat{\alpha}_N < 1$, which satisfies the followings:

$$\prod_{i=1}^{n}\widehat{\alpha}_i \le \widehat{\alpha}_1^n,\ \forall n = 1, 2, ..., N, \tag{55}$$

From (55) we have:

$$1+\sum_{n=1}^{N-1}\prod_{i=1}^{n}\widehat{\alpha}_i \le 1+\sum_{n=1}^{N-1}\widehat{\alpha}_1^n. \tag{56}$$

Also from (55)

$$\frac{1}{1-\prod_{i=1}^{N}\widehat{\alpha}_i} \le \frac{1}{1-\widehat{\alpha}_1^N}. \tag{57}$$

Combining (56) and (57) yields the following:

$$\frac{1}{1-\prod_{n=1}^{N}\widehat{\alpha}_n}\left(1+\sum_{\substack{z=1\\N>1}}^{N-1}\prod_{j=1}^{z}\widehat{\alpha}_j\right) \le \frac{1+\sum_{j=1}^{N-1}\widehat{\alpha}_1^j}{1-\widehat{\alpha}_1^N}. \tag{58}$$

Comparing the left and right hand sides of (58) with (26) and (49), respectively, we get (35).

References

[1] I. Zhou, I. Makhdoom, N. Shariati, M. A. Raza, R. Keshavarz, J. Lipman, M. Abolhasan, and A. Jamalipour, "Internet of Things 2.0: Concepts, Applications, and Future Directions," *IEEE Access*, vol. 9, pp. 70 961–71 012, 2021.

[2] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, Opportunities, and Directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

[3] S. Vitturi, C. Zunino, and T. Sauter, "Industrial Communication Systems and Their Future Challenges: Next-Generation Ethernet, IIoT, and 5G," *Proceedings of the IEEE*, vol. 107, no. 6, pp. 944–961, 2019.

[4] Q. Wang and J. Jiang, "Comparative Examination on Architecture and Protocol of Industrial Wireless Sensor Network Standards," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 2197–2219, 2016.

[5] A. Aijaz, "Private 5G: The Future of Industrial Wireless," *IEEE Industrial Electronics Magazine*, vol. 14, no. 4, pp. 136–145, 2020.

[6] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward Computing Resource Reservation Scheduling in Industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8210–8222, 2021.

[7] M. Saki, M. Abolhasan, and J. Lipman, "A Novel Approach for Big Data Classification and Transportation in Rail Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1239–1249, 2020.

[8] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling Massive IoT Toward 6G: A Comprehensive Survey," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 11 891–11 915, 2021.

[9] M. S. Ali, E. Hossain, and D. I. Kim, "LTE/LTE-A Random Access for Massive Machine-Type Communications in Smart Cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 76–83, 2017.

[10] N. Jiang, Y. Deng, A. Nallanathan, X. Kang, and T. Q. S. Quek, "Analyzing Random Access Collisions in Massive IoT Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6853–6870, 2018.

[11] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 3, pp. 1805–1838, 2020.

[12] M. Angjelichinoski, K. F. Trillingsgaard, and P. Popovski, "A Statistical Learning Approach to Ultra-Reliable Low Latency Communication," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5153–5166, 2019.

[13] M. A. Raza, M. Abolhasan, J. Lipman, N. Shariati, W. Ni, and A. Jamalipour, "Statistical Learning-Based Grant-Free Access for Delay-Sensitive Internet of Things Applications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5492–5506, 2022.

[14] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing Grant-Free Access for URLLC Service," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 741–755, 2021.

[15] J. Gao, W. Zhuang, M. Li, X. Shen, and X. Li, "MAC for Machine-Type Communications in Industrial IoTPart I: Protocol Design and Analysis," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9945–9957, 2021.

[16] J. Gao, M. Li, W. Zhuang, X. Shen, and X. Li, "MAC for Machine-Type Communications in Industrial IoTPart II: Scheduling and Numerical Results," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9958–9969, 2021.

[17] M. Gharbieh, H. ElSawy, M. Emara, H.-C. Yang, and M.-S. Alouini, "Grant-Free Opportunistic Uplink Transmission in Wireless-Powered IoT: A Spatio-Temporal Model," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 991–1006, 2021.

[18] J. Choi, "Re-Transmission Diversity Multiple Access Based on SIC and HARQ-IR," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4695–4705, 2016.

[19] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A Novel Analytical Framework for Massive Grant-Free NOMA," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2436–2449, 2019.

[20] S. A. Tegos, P. D. Diamantoulakis, A. S. Lioumpas, P. G. Sarigiannidis, and G. K. Karagiannidis, "Slotted ALOHA With NOMA for the Next Generation IoT," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6289–6301, 2020.

[21] A. Azari, C. Stefanovic, P. Popovski, and C. Cavdar, "Energy-Efficient and Reliable IoT Access Without Radio Resource Reservation," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 908–920, 2021.

[22] N. Xia, H.-H. Chen, and C.-S. Yang, "Radio Resource Management in Machine-to-Machine CommunicationsA Survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 791–828, 2018.

[23] C. She, R. Dong, W. Hardjawana, Y. Li, and B. Vucetic, "Optimizing Resource Allocation for 5G Services with Diverse Quality-of-Service Requirements," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[24] C. Oh, D. Hwang, and T. Lee, "Joint Access Control and Resource Allocation for Concurrent and Massive Access of M2M Devices," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4182–4192, 2015.

[25] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. S. Wong, "D-ACB: Adaptive Congestion Control Algorithm for Bursty M2M Traffic in LTE Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9847–9861, 2016.

[26] A. Azari, M. Ozger, and C. Cavdar, "Risk-Aware Resource Allocation for URLLC: Challenges and Strategies with Machine Learning," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 42–48, 2019.

[27] M. A. Raza, M. Abolhasan, J. Lipman, N. Shariati, and W. Ni, "Statistical Learning-Based Dynamic Retransmission Mechanism for Mission Critical Communication: An Edge-Computing Approach," in *2020 IEEE 45th Conference on Local Computer Networks (LCN)*, 2020, pp. 393–396.