Application-Oriented Traffic Modeling of WiFi-based Internet of Things Gateways

Atef Abdrabou*§, *Member, IEEE*, Maitha Al Darei§, Monika Prakash§, and Weihua Zhuang†, *Fellow, IEEE*
Email:{atef.abdrabou@uaeu.ac.ae, 201103856@uaeu.ac.ae, 201790521@uaeu.ac.ae, wzhuang@uwaterloo.ca}

Abstract—**Many Internet of Things (IoT) devices generate relatively small-sized data and have limited energy supply. These two factors limit their ability to connect directly to cloud servers through a wireless backbone network without imposing a burden on this network in providing efficient data transfer. In this paper, we consider an IoT network architecture where a number of different IoT devices send their data wirelessly to an IoT gateway (or a fog node) via a WiFi network. We focus on characterizing incoming traffic patterns to the gateway for three typical IoT applications with real-time and non-real-time data transfer requirements, such as video surveillance, smart city, and e-healthcare. Our study is based on generating real IoT traffic traces in a lab environment from various sensors and devices for the aforementioned applications and employing these traces to emulate a network of IoT nodes connected to a gateway via WiFi. In the conducted experiments, different homogenous and non-homogeneous traffic patterns of the selected applications are examined for synchronized and unsynchronized data sources. Based on our empirical data, the experimental results reveal that the packet inter-arrival time distribution at the gateway is close to generalized Pareto distribution for homogeneous eHealth and smart city traffic, whereas the Weibull distribution is the nearest to model the empirical packet interarrival time for the rest of the examined traffic patterns. Moreover, we show that employing the experimental findings to analyze the delay performance of connecting the gateway to the cloud, given certain backbone network resources, leads to accurate results.**

Keywords – **IoT, gateway, WiFi, fog, inter-arrival time distribution, traffic, characterization.**

I. Introduction

The basic concept of IoT is to allow the cooperation between the Internet and *Things*, which refer to objects equipped with identification, sensing, actuating, or controlling capabilities. These smart objects can interact and communicate over the Internet to accomplish specific goals, such as reducing costs and increasing optimization in a multitude of domains [1] [2]. This allows translating the physical world into a digital cyber world, providing rich information via connectivity to anyone or anything at any time in any place. Toward realizing this objective, two main architectures are introduced in the literature to support transferring data from IoT devices to their management entities or servers. In the first one, which is commonly known as cellular IoT [3], the IoT devices connect directly to the cellular infrastructure. However, some open issues need to be addressed by this architecture to achieve the global IoT vision. One of these issues is the wide-area deployment of a high density of IoT devices and the fact that many of these devices use low data rates to wirelessly transfer the sensed information through the cellular network to Internet cloud servers. This adds a burden on the backbone networks as it needs to support the access of a massive number of devices or machines in addition to the regular voice/data traffic of human users [4]. The second architecture uses gateways to aggregate the traffic of IoT devices before sending this traffic over the wireless cellular backbone [5].

Indeed, the introduction of IoT gateways [5], machine-to-machine (M2M) capillary gateways [6], and/or fog nodes can relieve the cellular network burden as the traffic can be aggregated by a gateway or locally processed, if possible, by a fog node before it is sent over the backbone to the cloud for further processing and/or storage [7]. Nevertheless, a thorough study of the incoming traffic characteristics to gateways or fog nodes is needed. This is attributed to the fact that the traffic patterns generated by different IoT devices can be different from Internet traffic generated mainly by direct human activities [2] [8]. Thus, investigating IoT traffic characteristics is essential because they are highly involved in planning and designing network infrastructure [2].

Therefore, this study focuses on the second architecture where the data from IoT devices is sent via the widely-deployed WiFi technology to IoT gateways or fog nodes, which forward it to its final destination in the Internet cloud through a wireless backbone network as shown in Fig. 1. The gateways/fog nodes aggregate traffic from different sources such as video surveillance cameras, portable eHealth devices (such as the electrocardiogram (ECG), electromyography (EMG), and blood pressure) for remote patient monitoring, and smart city sensors (air pollutant gases, temperature, luminosity, air pressure, and proximity sensors). Indeed, modeling the input traffic to these IoT gateways is of paramount importance for evaluating the packet delay performance (e.g., via queuing analysis) of the aggregated data and allocating the backbone network resources. Thus, we focus on the characterization of the incoming traffic to WiFi-based IoT gateways from IoT devices and the scaling of the WiFi network connecting the IoT devices with the gateways. The objective is to achieve an efficient allocation of the backbone radio resources and a successful realization of the second architecture.

The contribution of this research is three-fold. First, the packet inter-arrival time of the incoming IoT traffic to a WiFi-based IoT gateway is characterized for widely used IoT

*Corresponding Author.

§A. Abdrabou, M. Al Darei, and M. Prakash are with the Department of Electrical Engineering, UAE University, Al-Ain, PO 15551, UAE.

†W. Zhuang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

applications, such as video surveillance, smart city, and e-healthcare. The three applications constitute a major part of a wide smart city vision realized by IoT/M2M devices to enhance the welfare and quality of life of city residents [3] [5]. This is achieved through the integration of different services that target (i) the security (video surveillance, proximity sensing, and luminosity detection) and (ii) health (ECG, EMG, and blood pressure measurements) of human beings, in addition to essential environmental factors such as temperature, humidity, pressure, and air pollutant gases through wide-area monitoring. Different traffic patterns of these applications are constructed from traffic traces (obtained from real devices using lab experiments) and examined in a realistic setting using a virtual-machine based network emulator.

Second, we propose an analytical estimation of the number of IoT devices running similar or different applications that can connect through a WiFi network to a gateway or a fog node without causing a significant access delay variation or packet loss due to collisions. Third, based on the experimental findings, an end-to-end delay performance analysis for transferring the aggregated traffic from a gateway or fog node over a backbone communication link is provided.

The rest of the paper is organized as follows. Section II surveys the most relevant related works. Section III describes the proposed system model. In Section IV, the details of the experimental setups used to generate real IoT traffic traces are provided. Section V introduces and analyzes the emulated lab environment utilized to characterize the incoming traffic of a WiFi-based IoT gateway. The results and the major findings of the performed experiments are presented in Section VI. Using the experimental findings, Section VII introduces a mathematical analysis for the end-to-end packet delay of IoT gateways/fog nodes when connected to a backbone network. Finally, Section VIII concludes this research.

II. Related Works

The research on IoT traffic characterization is still in its infancy stages [8]. It has been shown that conventional approaches, including the standard Poisson process, are inadequate to model IoT traffic [9]. According to [10], the characteristics of aggregated periodic IoT data using the Poisson process can introduce large errors on performance metrics of interest.

The Third Generation Partnership Project (3GPP) association suggests modeling the distribution of data transfer events generated (over some period of time) by a large number of machine-type communication (MTC) devices connected directly to an LTE network by either uniform distribution for non-synchronized traffic or Beta distribution for synchronized one [11]. However, this model does not target packet inter-arrival time to IoT gateways over a WiFi network. In [12], the authors propose a model for periodic uplink reporting for cellular smart city applications (e.g., metering reports of gas/water/electric consumptions, smart agriculture, and smart environment). In this model, each device is assigned a reporting period with a Pareto-distributed packet length [12], with direct communications between IoT devices and a cellular network. Metzger et al. in [13] address the validity of using a Poisson approximation to model the arrival of aggregated periodic IoT traffic at a cloud server. The authors of [14] present fitted statistical distributions to the empirical distributions of packet interarrival times of some smart home applications such as temperature, air pressure, light, and motion sensors. In [15], the authors provide a mathematical model for the packet interarrival times of the aggregated traffic of a number of homogeneous sources representing temperature, light, or motion sensors. The works of [14] [15] focus on homogenous traffic from specific types of IoT sensors and do not include the effect of the communication technology used to aggregate the traffic.

Some research works focus on IoT traffic classification for security purposes. The authors of [16] compute the entropy value of the parameters of IoT traffic, generated by a software tool, to identify the device creating the traffic. Similarly, in [17], Sivanathan et al. propose an architecture for IoT traffic classification to autonomously detect IoT devices. Other researchers focus on studying IoT traffic characteristics for different purposes, such as investigating the effect on network storage and traffic [18] and examining the impact on long-range networks (LoRA) [19].

To the best of our knowledge, no other work in the literature addresses the characterization of the aggregated incoming traffic to WiFi-based IoT gateways for three major types of applications (video surveillance, smart city, and e-healthcare), considering inhomogeneous sources of different application types. Moreover, our work is based on traffic data captured from realistic experimental setups that mimic real-life scenarios.

III. System Model

Consider a three-tier topology in the system model. This topology divides the IoT system into three layers, namely, the perception, the network, and the application [20]. Here, the system model considers the incoming flow of data from the devices located in the perception layer and received by an IoT gateway or fog node in the network layer. The perception layer accommodates different types of devices and sensors that represent a wide range of IoT applications such as e-healthcare, smart city, and video surveillance. Furthermore, such applications can provide a mixture of traffic that is acquired in real-time or non-real time.

Our system model follows the general consensus that the traffic generated in IoT networks tends to be more uplink than downlink [19]. Thus, a gateway/fog node can be deployed in the network layer in order to collect the data coming from different IoT devices in the perception layer for further analysis and transmission to cloud management servers through a wireless cellular (4G/5G) backbone. The gateway is directly connected to each sensor node over an Internet Protocol (IP) network with a single-hop star topology using the IEEE 802.11 medium access control (MAC) protocol. Here, we assume that the data obtained by the sensors are sent wirelessly with a negligible packet loss over a WiFi network as it is currently among the most widely-deployed technologies.

It is worth noting that the gateway-based architecture outperforms the cellular IoT architecture in several aspects. First, in the cellular IoT architecture, the existence of a vast number of IoT/M2M devices side-by-side with human-based devices significantly increases the transmission collision probability when these devices compete for the random access channel (RACH) during the random access (RA) procedure. This leads either to a considerable delay in connection establishment or a failure to access the channel after exceeding some number of trials [4] [21] [22]. On the contrary, in the gateway-based architecture, the aggregate packet arrival rate is often sufficient to keep the gateway in the connected mode with the cellular backbone network for an extended period (as it receives data from different IoT devices). Thus, it does not need to send connection requests over the random access channel. This dramatically decreases the number of access requests, which in turn increases the access success probability, reduces the transmission collision probability, and decreases the data channel access delay as only the gateways connect to the cellular network [21] [23]. Second, the existence of WiFi-based gateways or fog nodes in the vicinity of IoT devices generally reduces the power consumption of the IoT devices. On the other hand, in the cellular IoT, the IoT devices need to ramp up their power (especially during the RA procedure) if their connection requests suffer from transmission collisions or are not identified by the base station due to interference [22]. Third, in the gateway-based architecture, the gateway/fog node can help perform some security and privacy functions on behalf of IoT devices (mostly resource-constrained). This includes access control, integrity, and isolation [7]. However, in the cellular IoT, these devices need to apply all the cellular network security functions, which are not optimally designed for devices with such limited resources.

Thus, the architecture under study offers efficient monitoring and response that are particularly required by any IoT nodes sending time-critical information (e.g., healthcare devices). Fig. 1 illustrates the considered system model.

IV. Experimental Setups for Collecting IoT Traffic Traces

This section details the experimental setups that are employed to generate and capture real IoT traffic traces for e-healthcare, smart city, and video surveillance applications.

These experimental setups are implemented based on two micro-controller boards, namely Waspmote [24], and Intel Galileo [25]. The data generated by both boards are transmitted as ASCII characters encapsulated in UDP packets and captured using the Wireshark packet analyzer in *pcap* format. All the setups have been set to generate a packet capture file of around a five-minute duration.

A. e-Health

In this setup, three medical sensors are selected to generate the e-health traffic, namely, electrocardiogram (ECG), electromyography (EMG), and blood pressure. The ECG and EMG sensors provide real-time measurements, whereas the blood pressure device stores its readings to be available for retrieval in a non-real-time fashion. These e-health sensors are connected to the Intel Galileo board through a health shield.

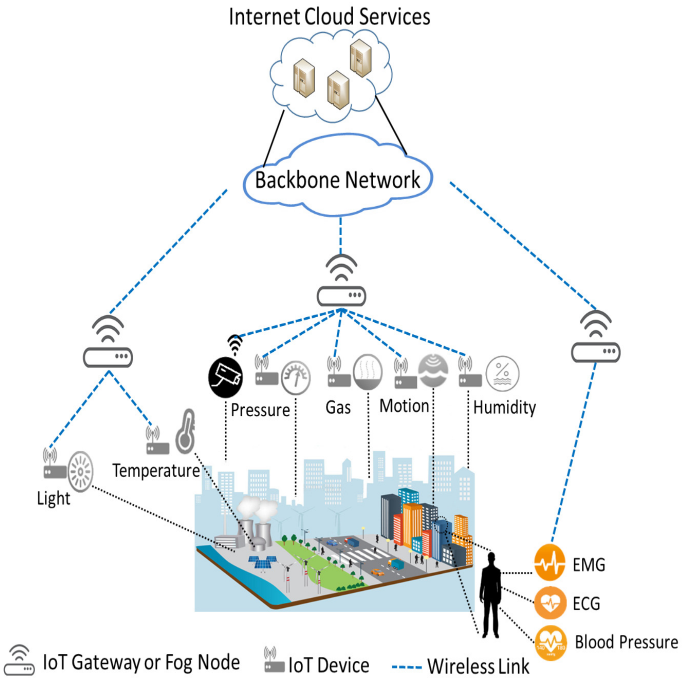


Fig. 1: An illustration of the system model.

The readings for each e-health sensor are transferred, via a WiFi network, from the Galileo board to a personal computer (PC) and captured by Wireshark, as shown in Fig. 2. The Intel Galileo board sends the data wirelessly through a bridge router.

Depending on the sensor's type, the number of samples per second is properly selected to obtain accurate results. In the case of the ECG sensor, the Intel Galileo board is programmed to obtain the data from the sensor with a sampling rate of 2000 Hz. Consequently, it sends a packet with a sufficient number of samples to ensure a smooth replaying at the destination. The sampling rate is chosen according to the recommendations of the European Society of Cardiology and North American Society of Pacing and Electrophysiology [26].

For the EMG sensor, the sampling rate is 5000 samples/second, which satisfies Nyquist criterion for this kind of signals [27]. Moreover, a sufficient number of EMG readings is placed in one UDP packet to ensure that the receiver can interpret and play the time-varying EMG graph smoothly. This is necessary to analyze EMG graphs and immediately detect any critical condition correctly.

The blood pressure device sends one packet every 20 seconds. Each packet contains 7 readings. The number of readings is selected such that the packet length of the blood pressure device matches the other medical sensors. The packet size for the eHealth traffic slightly varies around an average of 937 Bytes for ECG data, 1054 Bytes for EMG data, and 939 Bytes for blood-pressure data. The overall average data rate of the three sensors is 49 kbps.

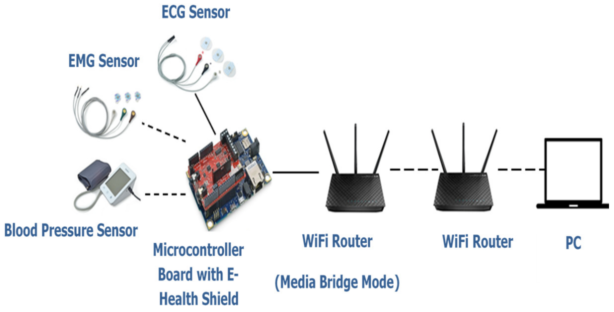


Fig. 2: Experimental setup for e-health data trace.

B. Smart City

This trace contains a variety of measured quantities, which are likely to be monitored in most smart city initiatives, as in the following.

The concentration of three gases (carbon dioxide, nitrogen dioxide, carbon monoxide) and a group of air pollutant gases (Air Pollutants II [24]) are monitored in an indoor location with normal air concentration using different sensors. Luminosity is also sensed. A reading is obtained from each one of these sensors every one second.

Moreover, temperature, humidity, and pressure are monitored with a sensor connected to the Waspmote board. A proximity sensor is also used with a sample rate of 40 Hz.

The packets generated by the aforementioned Waspmote board sensors are sent wirelessly through a WiFi router and captured by Wireshark, as depicted in Fig. 3. The packet size varies according to the sensor type. It is around 124 Bytes for gas sensors, 114 Bytes for the temperature, humidity, air pressure, and proximity sensors, whereas it is 77 Bytes for the light sensor. The average rate of the data sent from the sensor board is around 1.2 kbps.

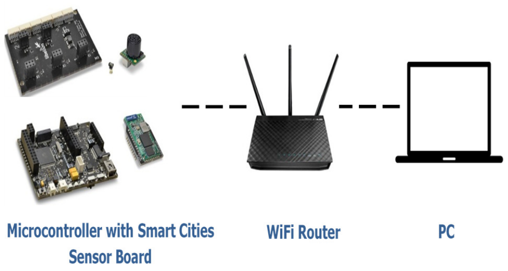


Fig. 3: Experimental setup for smart city data trace.

C. Video Surveillance

The video surveillance traffic trace is obtained using a network camera to generate a live streaming video (real-time traffic). The camera is configured to stream a UDP live video in Moving Picture Expert Group Transport Stream (MPEG TS) format using VideoLAN Client (VLC) media player via a WiFi network as illustrated in Fig. 4. The video trace is captured using Wireshark by a PC connected to the same network. The camera transmits fixed-size packets (1358 Bytes each) with a data rate of around 410 kbps.

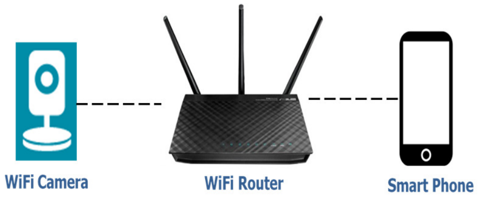


Fig. 4: Experimental setup for video streaming data trace.

V. IoT Network Experimental Setting and Gateway Load Analysis

In order to mimic the system model, two IoT network experimental setups are constructed. The first setup is emulation-based. It uses virtual machines in order to emulate the application nodes to provide large-scale experiments that cannot be performed in a lab environment. The second setup is laboratory-based. It is mainly used to validate the emulation-based scheme. The following section describes both setups, the examined traffic patterns, the analytical technique used to determine the number of application nodes employed in each experiment, and the validation of the emulation-based setup.

A. Emulated Experimental Setup

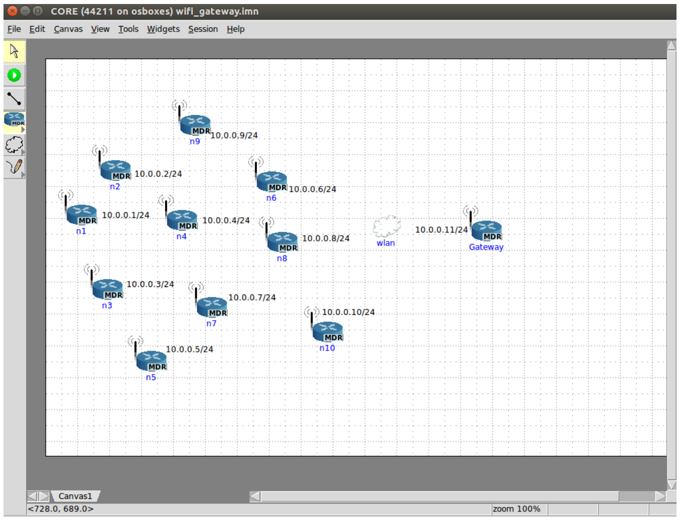


Fig. 5: The experimental emulation setup.

This setup uses the common open research emulator (CORE) tool [28] for emulating the application nodes under study. Opposite to discrete-event computer simulation, the tool builds a virtual network of virtual machines (nodes) that operate in real time. Therefore, it can run the full-fledged version of Linux applications over a real TCP/IP protocol stack. It is typically used when the required size of the actual test network is large. In the setup, a node is used to serve as an IoT gateway, or a fog node is placed in a location where it can communicate to the other nodes as shown in Fig. 5. The rest of the nodes are IoT traffic generators, where each node runs one of the aforementioned trace files obtained using the experimental setups introduced in Section IV. The number of nodes that replay a specific application trace file is determined based on the traffic pattern introduced in Subsection V-B. In the emulator, the nodes are connected with the gateway through a WiFi network that implements IEEE 802.11 medium access control (MAC) protocol. The effect of the physical channel is emulated by making the data rate r_i allocated to each node randomly assigned for every experiment sample, where $r_i \in \{6, 9, 12, 18, 24, 36, 48, 54\}$ Mbps since the wireless coverage and physical channel impairments affect the received signal-to-noise ratio (SNR) at the WiFi interface card. This accounts for changing the surroundings or even node locations to different spots, leading to different WiFi signal coverage quality, while rate adaptation is in place. Consequently, it mimics the WiFi transceiver's behavior in practice, which adaptively selects the modulation scheme and data transmission rate based on the channel condition. In addition, the starting time of sending traffic for each node is randomly determined for unsynchronized traffic. This captures the variations of the main network parameters that affect traffic characterization, targeted by this study across the samples of different experiments.

B. Application-Oriented Traffic Patterns

Seven traffic patterns are investigated. Three patterns include homogenous traffic, namely, Pattern 1 video surveillance traffic, Pattern 2 smart city traffic, and Pattern 3 e-health traffic. The inhomogeneous traffic is represented by the remaining four patterns, which include a mix of the three types of applications under study. Pattern 4 consists of equally mixed traffic of e-health, smart city, and video surveillance. The majority of Pattern 5 traffic is e-health, whereas the majority of Pattern 6 and 7 traffic come from video surveillance and smart city sources, respectively.

C. IoT Gateway Load Analysis

We want to estimate the number of IoT nodes generating homogeneous or inhomogeneous traffic connected to a WiFi-based gateway or a fog node, without causing a significant likelihood of packet loss, either due to packet collisions or large access delay variations. It is to allow all the packets generated by the IoT nodes to reach the gateway for a reliable data transfer[1]. Thus, the characterized incoming traffic includes all the packets sent to the gateway assuming the packet loss is negligible.

It is worth noting that the number of IoT nodes provided by the analysis is used to calculate the number of virtual IoT nodes in the emulation-based setup that replay the trace files and transmit them to the emulated gateway/fog node.

TABLE I: IEEE 802.11g system parameters [29].

System Parameter	Value
MAC Overhead (h_M)	208 bits
f_{ACK}	31.6μs
f_{RTS}	33.6μs
f_{CTS}	31.6μs
Slot Time σ	9μs
T_{SIFS}	10μs
T_{DIFS}	28μs
Basic Rate	6 Mbps
Data Rate	54 Mbps
CW_{min}	32
CW_{max}	1024

It has been shown in [30] and [31] that the channel access delay (service time) of an IEEE 802.11 WiFi network and packet collision probability steeply increase as the packet arrival rate of the WiFi nodes increases towards saturation. Consequently, increasing the number of nodes results in a coarse change in the network load and a faster approach towards saturation. This leads to the risk of packet loss due to packet collisions, which cause a large service time variance within this operation region. Therefore, the network should operate in a region where the standard deviation of the service time is much smaller than its average. It has been shown in [31] that this region satisfies the condition

$$g \triangleq \alpha(n-1) \ll 1 \tag{1}$$

where α is the probability of a non-empty node buffer and n is the number of IoT nodes.

Also, this region is marked by a small collision probability, p_c, which implies the probability of at least one other node simultaneously transmits with the tagged node and can be calculated as

$$p_c = 1 - (1 - \alpha p_t)^{n-1}\,. \tag{2}$$

In (2), p_t is transmission probability in any time slot of the backoff period. It can be obtained as

$$p_t = \frac{1}{E[B]} \tag{3}$$

where $E[B]$ is the average backoff period. The condition in (1) can be satisfied for a small value of α (e.g., $g \leq 0.1$). This leads to a small p_c, which can be calculated from (2) as

$$p_c \approx \frac{\alpha(n-1)}{E[B]}. \tag{4}$$

The average backoff period can be approximated, in the case of a low p_c, as [31]

$$E[B] \approx \sigma \frac{1-p_c}{1-2p_c} \frac{CW_{\min}}{2} \tag{5}$$

where σ is the slot time and W_{min} is the minimum contention window size.

In addition, α can be obtained from

$$\alpha = \lambda E[S_t] \tag{6}$$

where $E[S_t]$ is the average service time given by

$$E[S_t] = \left[E[F_t] + \frac{f_c}{2}\frac{p_c}{1-p_c}\right][1 + \alpha\,(n-1)] + E[B]. \tag{7}$$

Since the IoT nodes running different applications with differ-

[1] Packet loss due to channel impairments is assumed to be alleviated by IEEE 802.11 frame retransmissions and the adaptive transmission rate adjustment, which is implemented in almost all commercial WiFi hardware.

ent packet arrival rates, λ represents the average packet arrival rate given by

$$\lambda = \sum_{j=1}^{3} w_j \lambda_j \tag{8}$$

where w_j denotes the ratio of the number of IoT nodes running application j to the total number of nodes n, λ_j is the packet arrival rate of application j, and $j \in \{\text{Video Surveillance, eHealth, Smart City}\}$.

In (7), $E[F_t]$ is the average MAC frame transmission time and f_c is the frame collision time. The first term of (7) represents the average time spent in successful frame transmission and frame collision for a tagged node and the rest of the network nodes, whereas the second term represents the time spent in backoff [31].

The average frame transmission time can be obtained as (assuming the IEEE 802.11 handshake procedure is in place)

$$\begin{aligned} E[F_t] = f_{RTS} + f_{CTS} + 3\ T_{SIFS} \\ +f_{ACK} + (h_M + E[L])E[\tfrac{1}{R}] + T_{DIFS} \end{aligned} \tag{9}$$

where f_{RTS}, f_{CTS}, and f_{ACK} are the transmission times for the request-to-send (RTS), clear-to-send (CTS), and acknowledgment (ACK) frames, respectively; T_{DIFS} and T_{SIFS} are the distributed and short inter-frame spacing, respectively; the packet size and MAC frame overhead are denoted by L and h_M, respectively. The transmission rate is represented by R and

$$E\left[\frac{1}{R}\right] = \sum_{i=1}^{k} p_i \frac{1}{r_i} \tag{10}$$

where k is the number of available data rates and p_i is the probability of transmission at rate r_i.

The average packet size, $E[L]$ can be obtained as

$$E[L] = \sum_{i=1}^{3} w_j L_j \tag{11}$$

where L_j is the packet size for each application type.

The frame collision time can be calculated as

$$f_c = f_{RTS} + T_{DIFS}. \tag{12}$$

Solving (1) - (4) after setting $g = 0.1$ leads to p_c, which can be used to find α from (6) - (12) and to obtain n from

$$n = \left[\frac{\alpha - \lambda E[B]}{\alpha\lambda\left(E[F_t] + \frac{f_c}{2}\frac{p_c}{1-p_c}\right)} - \frac{1}{\alpha} \right]. \tag{13}$$

In fact, (13) offers an estimation of the number of IoT nodes that can be served by a WiFi-based gateway or a fog node with negligible packet loss (i.e., without significant packet collisions or considerable access delay variations).

For instance, using (13), the number of nodes for the homogenous patterns can be calculated as 6, 135, and 21 nodes, for Pattern 1, 2, and 3, respectively, when the data rate r_i of each node is chosen randomly (i.e., $p_i = \frac{1}{8}$). Similarly, for inhomogeneous patterns, the total number of nodes can be estimated for a given w_i value of each application. It is worth noting that the sampling frequency influences the traffic rate of each IoT traffic generator node for a particular application. This, in turn, impacts the estimated number of nodes (generating traffic for this application) to be connected to the WiFi-based gateway according to (8) and (13). As the traffic load per node increases, the number of nodes supported by the gateway decreases, and vice versa. Thus, reducing (or increasing) the sample rate leads to increasing (or decreasing) the number of nodes. At the same time, the gateway gets a similar traffic amount (but from a different number of nodes) with a similar impact from the WiFi network on the packet interarrival time.

D. Emulated Experimental Setup Validation

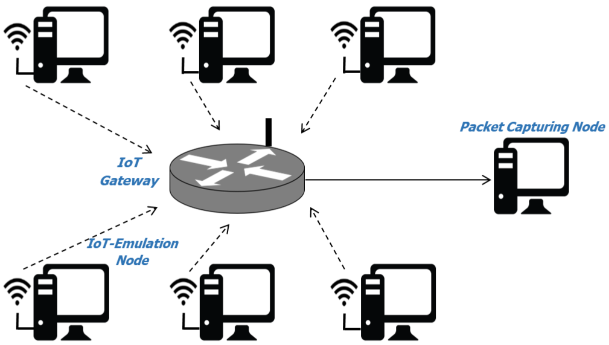


Fig. 6: The hardware experimental setup.

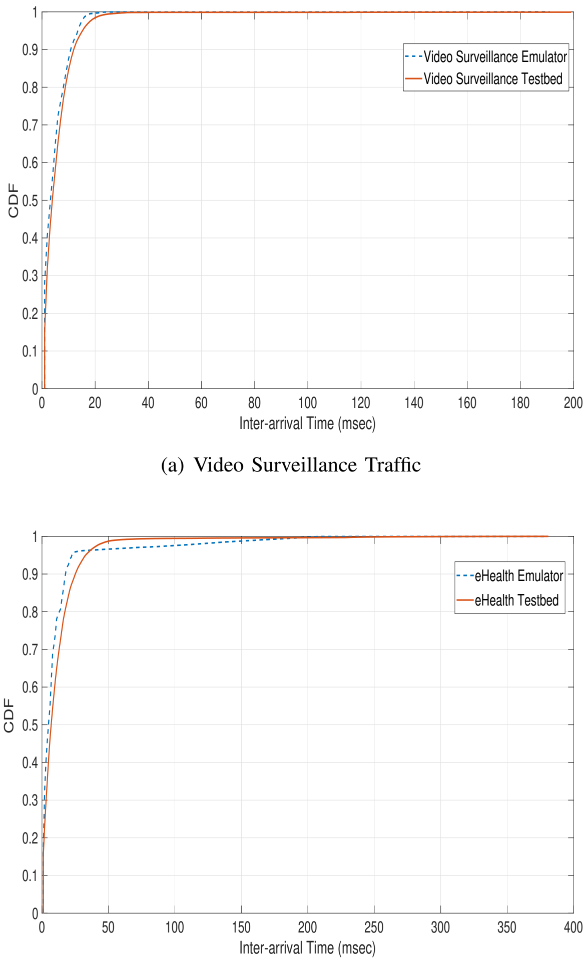
The emulated experimental setup is validated using a lab setup with real hardware. The setup consists of a WiFi router that is configured as an access point. The WiFi router emulates an IoT gateway, which wirelessly receives the IoT data traces from different IoT traffic generators and forwards these traces to a PC directly connected to it for packet capturing, as revealed in Fig. 6. An IoT traffic generator is a computer that replays a packet trace file and sends it to the packet capturing PC via the WiFi router. The trace file is obtained from one of the experimental setups mentioned in Section IV for the IoT applications under study.

To validate the emulated setup, we replicate the hardware setup mentioned above with the same configuration and number of nodes in the emulator. Besides, we maintain the same average channel rate in both setups.

Fig. 7 shows the cumulative distribution functions (CDFs) for packet inter-arrival time obtained from the emulated and hardware experimental setups using the same number of IoT traffic generators for two example traffic patterns, namely, video surveillance and e-Health. For both patterns, a close match can be observed between the CDFs of the emulated and hardware experimental setups, as Fig. 7(a) and 7(b) reveal. This implies that the emulated experimental setup can be used as a controlled experimental setting, especially for large-scale scenarios.

VI. Experimental Results

This section presents the results obtained by the emulated experimental setup for the seven traffic patterns mentioned in



(a) Video Surveillance Traffic

(b) eHealth Traffic

Fig. 7: **A sample comparison between the CDFs of the emulated and hardware experimental setups.**

Subsection V-B. We are particularly interested in checking the suitability of modeling the incoming traffic of IoT gateways by a distribution commonly used in the literature to model the packet inter-arrival time in the Internet (i.e., Exponential, Weibull, and generalized Pareto). The exponential distribution is widely used due to its mathematical tractability. However, it cannot model the packet interarrival time of Internet traffic due to the long-range dependence and heavy-tailed properties of this traffic [32]. Instead, Pareto and Weibull distributions have been shown to better model the packet interarrival time distribution for Internet traffic [33]. Moreover, different queuing models have been analyzed for Pareto [34], and Weibull [35] interarrivals.

A. Methods

1) Experiment Procedure: For each traffic pattern, a number of emulated IoT nodes (traffic generators) send their data (a traffic trace) over a WiFi network to a gateway. For inhomogeneous traffic patterns with the majority of traffic sources belonging to one type of application, around 75% of the total number of traffic sources are considered of this type, whereas the rest are divided equally among the other two types.

Each IoT node accesses the channel with data rate r_i that is chosen randomly, reflecting the quality of channel at the node location. The packet inter-arrival time at the emulated gateway is recorded, and its empirical distribution function is obtained. Each experiment is repeated at least 50 times to obtain sufficiently accurate statistics.

2) Network Configuration and Data Transmission: The emulated experiments are conducted using IEEE 802.11g/a WiFi data rates since relatively new standards such as IEEE 802.11n are shown to fall to these rates over long distances in indoor-outdoor environments [36]. Thus, these data rates are used to reflect the most anticipated rates for communication with WiFi-based gateways in the environments.

The number of IoT nodes that maintain a network operating point sufficiently away of the saturation region is calculated based on the analysis introduced in Subsection V-C to avoid getting a high packet collision probability, which may result in packet loss in the experiment samples.

Since WiFi networks use the contention-based MAC protocol, IEEE 802.11 [29], the impact of transmission synchronization of IoT nodes is studied. Thus, we examine the packet inter-arrival time distributions of the traffic patterns mentioned in Subsection V-B for two data generation scenarios, namely, synchronized and unsynchronized. The first scenario mimics a worst-case situation where the IoT nodes are simultaneously reporting data about some events. This scenario may happen when one event concurrently triggers more than one IoT node to report related information (e.g., wide area monitoring). The second scenario represents a more practical data reporting, where IoT nodes start transmission randomly over some time interval, as in [11].

3) Data Analysis: In order to statistically determine which one of the aforementioned distributions matches the empirical packet interarrival time distribution at an IoT gateway, two steps are followed. First, the empirical data is fitted to each distribution using the maximum likelihood method. Second, the goodness-of-fit is tested using Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared tests [37].

Unfortunately, the results from conducting these standard tests reveal a rejection of the null hypothesis for the studied traffic patterns for all tests. Therefore, another statistical approach is adopted to find the nearest fitted distribution to the empirical data. It is based on computing the maximum, mean, and the standard deviation of the absolute distance (AD) between the empirical CDF F_E(x) and the CDF of each fitted probability distribution F_{Fi}(x), given by

$$AD = |F_E(x) - F_{Fi}(x)|, i \in \{1, 2, 3\}. \tag{14}$$

In (14), F_{F1}(x) is represented by (15) for the exponential distribution, where μ is the mean; the generalized Pareto distribution CDF is represented by F_{F2}(x) in (16), where ξ, h, and s are the shape, threshold, and scale parameters, respectively; the Weibull distribution CDF is described by F_{F3}(x) in (17), where k and λ are the scale and shape parameters, respectively:

$$F_{F1}(x) = 1 - e^{-(x/\mu)} \tag{15}$$

$$F_{F2}(x) = 1 - (1 + \xi(x - h)/s)^{-1/\xi} \tag{16}$$

$$F_{F3}(x) = 1 - e^{(-x/k)^{\lambda}}. \tag{17}$$

The empirical distribution is then matched to the nearest distribution, i.e., the distribution with the lowest overall absolute distribution distance parameters.

B. Results

1) Traffic Pattern 1 (Only Video Surveillance Traffic): Fig. 9 shows an example of how the empirical CDF compares to the CDFs of the Exponential, Weibull, and Pareto distributions in the synchronized case. Although the three distributions look close to the empirical one for the synchronized case, the Weibull distribution has the lowest absolute distance parameters as revealed in Fig. 9. The average and standard deviation of the absolute distance of Weibull distribution are the lowest among the distributions in the unsynchronized case as depicted in Fig. 10. This implies that the Weibull distribution is the closest distribution to model the packet interarrival time for video surveillance traffic.

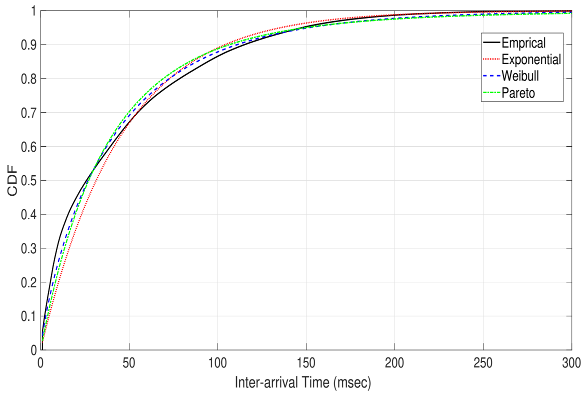


Fig. 8: **The CDF for the Pattern 1.**

2) Traffic Pattern 2 (Only Smart City Traffic): Figures 11(a) and 12(a) show the absolute distance results for the smart city traffic when IoT smart city nodes are synchronized and unsynchronized, respectively. The Pareto distribution is the lowest in absolute distance average and standard deviation, and hence considered the closest distribution to characterize this traffic pattern.

3) Traffic Pattern 3 (Only eHealth Traffic): Fig. 11(b) shows that the Pareto distribution has the lowest AD parameters among the three distributions for synchronized eHealth traffic. Similarly, for unsynchronized eHealth traffic, Pareto distribution has the lowest maximum, average, and standard deviation of AD among the examined distributions, as revealed in Fig. 12(b). This indicates that the Pareto distribution is the nearest to the empirical one.

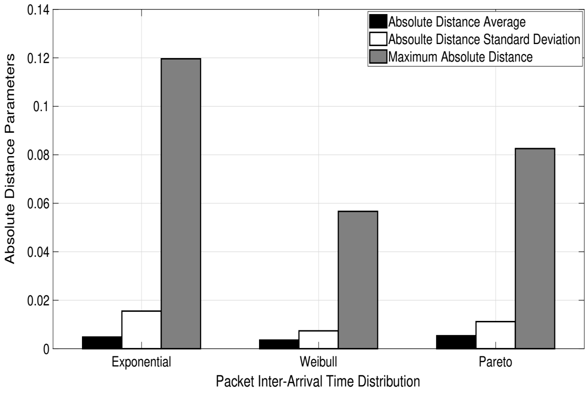


Fig. 9: **Absolute distance parameters for synchronized video surveillance traffic.**

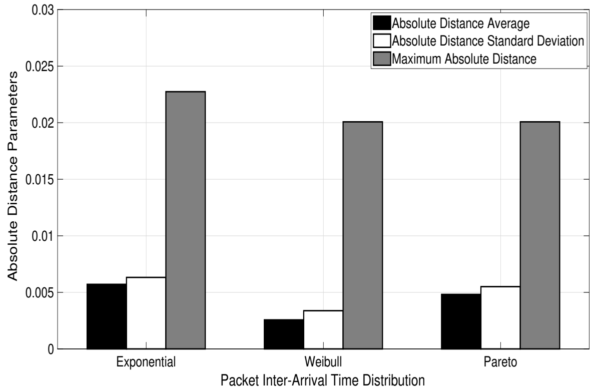


Fig. 10: **Absolute distance parameters for unsynchronized video surveillance traffic.**

4) Traffic Pattern 4 (Equally Mixed Traffic): Fig. 11(c) and Fig. 12(c) show the AD maximum value, average, and standard deviation for the synchronized and unsynchronized cases, respectively, when the traffic consists of an equal mix of video surveillance, eHealth, and smart city data (i.e., each is originated from the same number of IoT nodes). The Weibull distribution is the closest one to the empirical distribution, according to the AD parameters, for both the synchronized and unsynchronized cases.

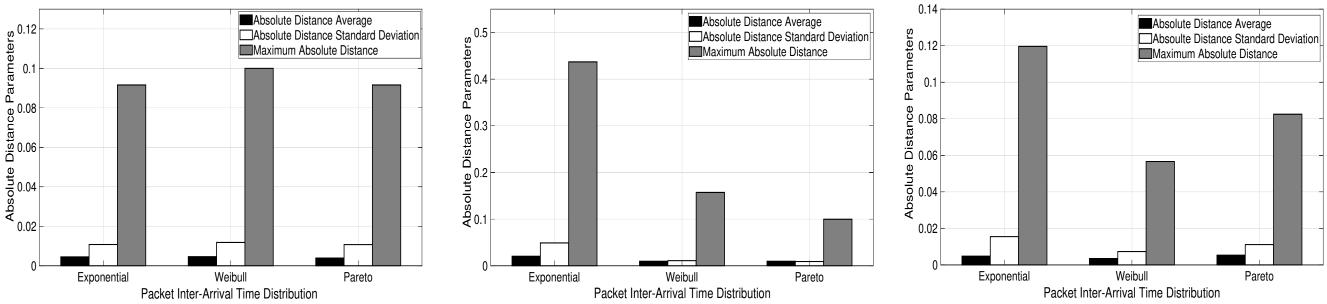
5) Traffic Pattern 5 (Majority Smart City Traffic): This traffic pattern represents a scenario in which the majority of the traffic comes from smart city sources. It is noticed from Fig. 13(a) and Fig. 14(a) that the Weibull distribution has the smallest maximum, average, and standard deviation of AD among the other distributions.

6) Traffic Pattern 6 (Majority Video Surveillance Traffic): The majority of the sources for this traffic pattern are video surveillance cameras. It is observed from Fig. 13(b) and Fig. 14(b) that the lowest AD parameters can be obtained using the Weibull distribution.

7) Traffic Pattern 7 (Majority eHealth Traffic): This traffic pattern is formed when the majority of the sending sources are eHealth sensors. Fig. 13(c) and 14(c) show that the Weibull distribution is the nearest one to match the empirical

TABLE II: Distribution parameters of different traffic patterns

Traffic Patterns		**Exponential**	**Weibull**		**Generalized Pareto**		
		μ	k	λ	ξ	s	h
Video Surveillance	Sync	45.26	41.36	0.84	0.19	36.51	0
	Unsync	48.52	49.77	1.07	-0.01	49.32	0
Smart City	Sync	53.39	51.74	0.94	0.103	47.4	0
	Unsync	55.83	53.16	0.91	0.12	48.34	0
eHealth	Sync	77.46	31.2	0.51	1.34	8.85	0
	Unsync	83.99	82.06	0.95	0.09	75.56	0
Equal Mix	Sync	45.43	40.04	0.8	0.29	32.9	0
	Unsync	47.34	48.22	1.04	0.01	46.55	0
Majority Smart City	Sync	51.46	47.94	0.87	0.14	44.21	0
	Unsync	53.77	54.98	1.06	0.02	52.72	0
Majority Video Surveillance	Sync	51.44	46.58	0.83	0.17	42.46	0
	Unsync	54.28	56.09	1.09	0.002	54.13	0
Majority eHealth	Sync	79.58	74.47	0.87	0.06	74.68	0
	Unsync	82.84	85.62	1.09	0.008	82.12	0

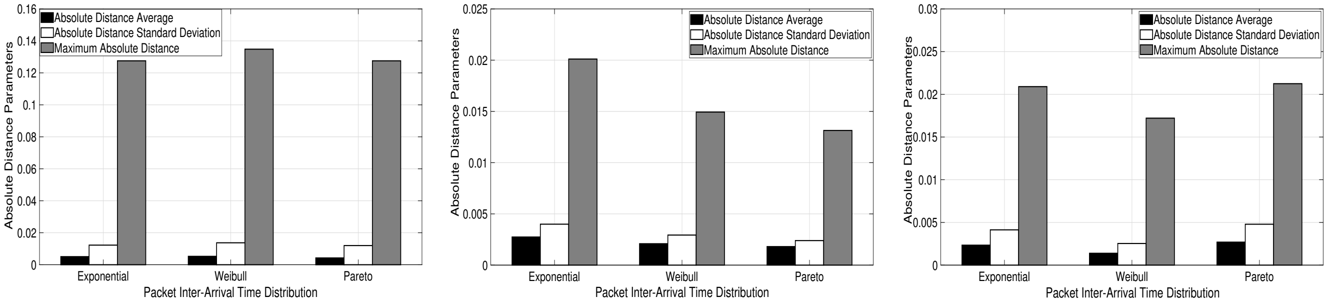


(a) Smart City Traffic

(b) eHealth Traffic

(c) Equally-mixed Traffic

Fig. 11: **Absolute distance parameters for the synchronized traffic Pattern 2, 3, and 4.**



(a) Smart City Traffic (unsync)

(b) eHealth Traffic (unsync)

(c) Equally-mixed Traffic (unsync)

Fig. 12: **Absolute distance parameters for the unsynchronized traffic Pattern 2, 3, and 4.**

distribution since it has the lowest AD maximum, average, and standard deviation irrespective of whether the traffic is synchronized or unsynchronized.

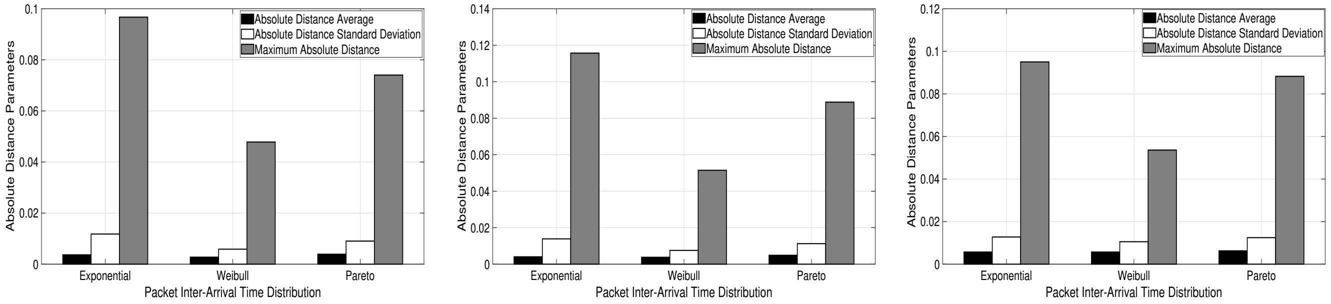
VII. Packet Interarrival Time Distribution & Queuing Analysis for IoT Gateways

Consider N IoT gateways (or fog nodes) send the received data from different sensors of the aforementioned applications to their respective management entities (e.g., cloud servers) through a wide coverage wireless backbone such as a 4G or 5G-based network. Each gateway is allocated the same amount of radio resources, which provide a fixed service time per packet S_t. In such a scenario, the gateway can be modeled as a G/D/1 queuing system.

The average end-to-end delay $E[D]$ can be obtained by the aide of Marchal approximation [38] using

$$E[D] \approx \frac{\rho S_t c_v^2}{2(1-\rho)} + S_t \tag{18}$$

where c_v is the coefficient of variation of the packet interarrival

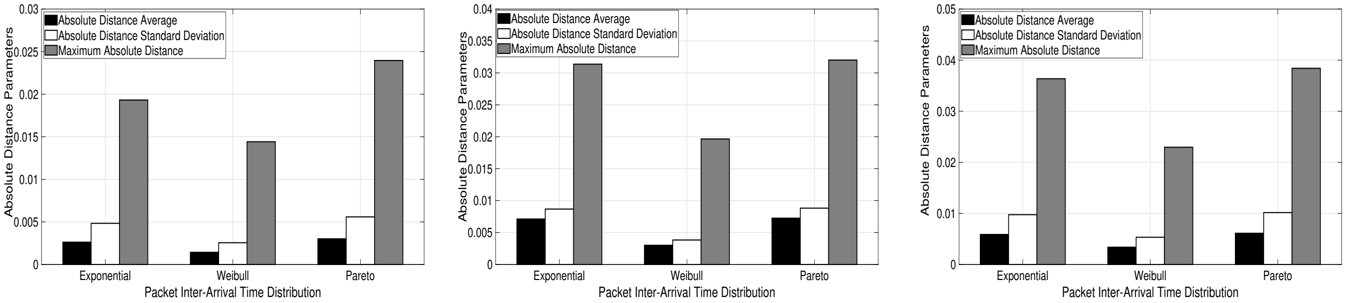


(a) Majority Smart City Traffic

(b) Majority Video Surveillance Traffic

(c) Majority eHealth Traffic

Fig. 13: **Absolute distance parameters for the synchronized traffic Pattern 5, 6, and 7.**



(a) Majority Smart City Traffic (unsync)

(b) Majority Video Surveillance Traffic (unsync)

(c) Majority eHealth Traffic (unsync)

Fig. 14: **Absolute distance parameters for the unsynchronized traffic Pattern 5, 6, and 7.**

time and ρ is the queue utilization factor.

In the sequel, we provide a numerical comparison between the measured end-to-end packet delay using CORE emulator and the analytically calculated average packet delay using (18) for three packet interarrival time distributions, namely, Exponential, Weibull, and generalized Pareto. The aim is to show the impact of the interarrival time distribution on radio resource allocation. The emulation is performed by using a packet trace of a specific traffic pattern as the input of an IoT gateway or fog node for each sample. The emulated configuration includes a gateway connected to a cloud data management server via a 4G or 5G wireless link. The wireless link rate is varied to reflect the availability of different amounts of radio resources, which depend on the number of gateways (or fog nodes), the background traffic of the backbone network, and channel condition. Consequently, the impact of channel impairments (i.e., pathloss, fading, interference) is represented by the available data rate for the wireless link.

Fig. 15 shows the average end-to-end delay that the packets from an IoT gateway/fog node experience when transferred over a link for different available link capacity (represented as different utilization factor values). The figure compares the average delay measured by the emulator for the video surveillance traffic pattern with the delay calculated by (18) for the aforementioned interarrival time distributions. Apparently, Fig. 15 reveals that the average end-to-end delay is generally closer to the one analytically obtained with Weibull distribution than with the other distributions over different utilization factor values. Similarly, Fig. 16 shows that the measured average end-to-end delay best matches the values calculated using (18) for generalized Pareto distribution. Indeed, the results depicted by Figures 15 and 16 are in agreement with the results shown in Subsection VI-B for the same patterns. The figures also depict that the assumption of the exponential distribution for the interarrival time leads to underestimating the average end-to-end delay.

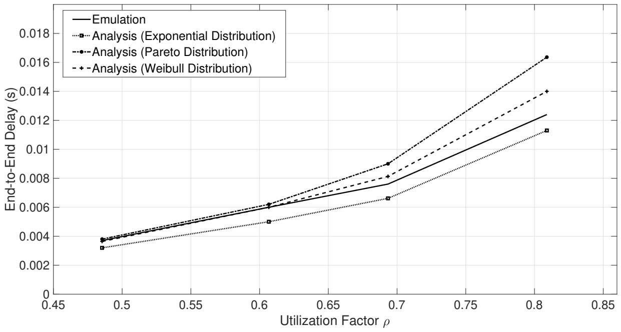


Fig. 15: **End-to-end delay from a WiFi gateway to a management server (video surveillance pattern).**

VIII. Conclusion

We have investigated the characterization of incoming traffic at WiFi-based IoT gateways or fog nodes, since it is pivotal for allocating the resources of the backbone network connecting IoT nodes to cloud servers. This characterization includes

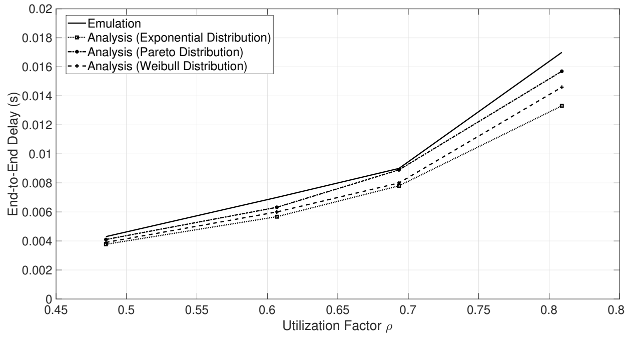


Fig. 16: **End-to-end delay from a WiFi gateway to a management server (smart city pattern).**

homogeneous and non-homogeneous traffic patterns of popular IoT applications, such as video surveillance, e-healthcare, and smart city, used in the area covered by the gateway or fog node.

Real IoT data traces are generated in a laboratory setting and used to generate the studied traffic patterns via a virtual machine-based network emulation validated by a hardware-based replica in another lab setup. The WiFi network scale in terms of the number of IoT nodes is analyzed and determined for a low probability of packet collisions and small access delay variations to achieve negligible packet loss, assuming date rate adaptation is in place to mitigate physical channel impairments.

It is observed that the generalized Pareto distribution is the closest among other common distributions to match the empirical packet inter-arrival time distribution at a WiFi-based IoT gateway or fog node for homogeneous e-healthcare and smart city traffic. On the other hand, the Weibull distribution is found to be the nearest distribution to match the empirical packet inter-arrival time for homogeneous video surveillance and any other inhomogeneous traffic pattern (a traffic mix of the three applications) irrespective of whether the IoT traffic sources are synchronized or not. Analyzing the average end-to-end delay for a backbone link connecting a gateway/fog node to a cloud entity using these packet interarrival time findings reveals accurate results. This implies that the exponential interarrival time is not suitable to model the packet interarrival time at WiFi-based IoT gateways or fog nodes. In addition, the WiFi random access scheme affects the packet interarrival time at the gateway, making it not exactly following any of the distributions commonly used for traffic modeling (such as the exponential, Weibull, and generalized Pareto), although the number of IoT nodes is set to keep access delay variations low.

IX. Acknowledgments

This research was made possible by the UAE University UPAR grant number 12N008/31N456.

References

[1] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[3] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-Things-Based Smart Cities: Recent Advances and Challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 16–24, 2017.

[4] H. G. Moussa and W. Zhuang, "RACH Performance Analysis for Large-Scale Cellular IoT Applications," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3364–3372, 2019.

[5] Y. Mehmood, N. Haider, M. Imran, A. Timm-Giel, and M. Guizani, "M2M Communications in 5G: State-of-the-Art Architecture, Recent Advances, and Research Challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 194–201, 2017.

[6] V. B. Mišić, J. Mišić, X. Lin, and D. Nerandzic, "Capillary machine-to-machine communications: the road ahead," in *International Conference on Ad-Hoc Networks and Wireless*, pp. 413–423, Springer, 2012.

[7] M. Chiang and T. Zhang, "Fog and IoT: An Overview of Research Opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, 2016.

[8] A. Sivanathan, D. Sherratt, H. H. Gharakheili, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Characterizing and classifying IoT traffic in smart cities and campuses," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 559–564, May 2017.

[9] E. Soltanmohammadi, K. Ghavami, and M. Naraghi-Pour, "A Survey of Traffic Issues in Machine-to-Machine Communications Over LTE," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 865–884, 2016.

[10] T. Hoßfeld, F. Metzger, and P. E. Heegaard, "Traffic modeling for aggregated periodic IoT data," in *2018 21st Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, pp. 1–8, Feb 2018.

[11] 3GPP, "Study on RAN Improvements for Machine-type," Technical Report 37.868, 3rd Generation Partnership Project (3GPP), 09 2011. Version 11.0.0.

[12] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," Technical Report 45.820, 3rd Generation Partnership Project (3GPP), 08 2015. Version 2.1.0.

[13] F. Metzger, T. Hoßfeld, A. Bauer, S. Kounev, and P. E. Heegaard, "Modeling of Aggregated IoT Traffic and Its Application to an IoT Cloud," *Proceedings of the IEEE*, vol. 107, no. 4, pp. 679–694, 2019.

[14] C. Majumdar, M. López-Benítez, and S. N. Merchant, "Real Smart Home Data-Assisted Statistical Traffic Modeling for the Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4761–4776, 2020.

[15] M. López-Benítez, C. Majumdar, and S. N. Merchant, "Aggregated Traffic Models for Real-World Data in the Internet of Things," *IEEE Wireless Communications Letters*, vol. 9, no. 7, pp. 1046–1050, 2020.

[16] H. Nguyen-An, T. Silverston, T. Yamazaki, and T. Miyoshi, "IoT traffic: Modeling and measurement experiments," *IoT*, vol. 2, p. 140–162, Feb 2021.

[17] A. Sivanathan, H. H. Gharakheili, and V. Sivaraman, "Detecting behavioral change of IoT devices using clustering-based network traffic modeling," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7295–7309, 2020.

[18] E. Batista, L. Andrade, R. Dias, A. Andrade, G. Figueiredo, and C. Prazeres, "Characterization and modeling of IoT data traffic in the fog of things paradigm," in *2018 IEEE 17th International Symposium on Network Computing and Applications (NCA)*, pp. 1–8, 2018.

[19] V. Gupta, S. K. Devar, N. H. Kumar, and K. P. Bagadi, "Modelling of IoT traffic and its impact on LoRaWAN," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–6, 2017.

[20] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," *IEEE Internet of Things Journal*, vol. 4, pp. 1125–1142, Oct 2017.

[21] T. Lin, C. Lee, J. Cheng, and W. Chen, "PRADA: Prioritized Random Access With Dynamic Access Barring for MTC in 3GPP LTE-A Networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2467–2472, 2014.

[22] M. Gerasimenko, V. Petrov, O. Galinina, S. Andreev, and Y. Koucheryavy, "Energy and delay analysis of LTE-Advanced RACH performance under MTC overload," in *2012 IEEE Globecom Workshops*, pp. 1632–1637, 2012.

[23] P. Zhou, H. Hu, H. Wang, and H. Chen, "An efficient random access scheme for OFDMA systems with implicit message transmission," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2790–2797, 2008.

[24] "Waspmote," in *Accessed: Mar. 26, 2021. [Online]. Available: http://www.libelium.com/products/waspmote/*.

[25] "Introduction to Intel® Galileo Boards," in *Accessed: Mar. 26, 2021. [Online]. Available: https://www.intel.com/content/www/us/en/support/articles/000005912/boards-and-kits/intel-galileo-boards.html*.

[26] M. Singh, B. Singh, and V. K. Banga, "Effect of ECG Sampling Frequency on Approximate Entropy based HRV," *International Journal of Bio-Science and Bio-Technology*, vol. 6, pp. 179–186, 08 2014.

[27] H. Chen, Y. Zhang, Z. Zhang, Y. Fang, H. Liu, and C. Yao, "Exploring the relation between EMG sampling frequency and hand motion recognition accuracy," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1139–1144, Oct 2017.

[28] J. Ahrenholz, "Comparison of CORE network emulation platforms," in *2010 - MILCOM 2010 MILITARY COMMUNICATIONS CONFERENCE*, pp. 166–171, Oct 2010.

[29] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band, *IEEE Std 802.11g/D1.1*, 2001.

[30] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of IEEE 802.11 MAC protocols in wireless LANs," *Wireless Communication and Mobile Computing*, vol. 4, pp. 917–931, Dec. 2004.

[31] A. Abdrabou and W. Zhuang, "Stochastic delay guarantees and statistical call admission control for IEEE 802.11 single-hop ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3972–3981, 2008.

[32] V. Paxson and S. Floyd, "Wide area traffic: the failure of poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.

[33] A. Feldmann, *Characteristics of TCP Connection Arrivals*, ch. 15, pp. 367–399. Wiley-Blackwell, 2002.

[34] R. Singhai, S. D. Joshi, and R. K. P. Bhatt, "Offered-load model for Pareto inter-arrival network traffic," in *2009 IEEE 34th Conference on Local Computer Networks*, pp. 364–367, 2009.

[35] A. Tamazian and M. Bogachev, "Analytical and numerical estimates of the Weibull/M/1 and Weibull/Weibull/1 queues efficiency," in *2015 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW)*, pp. 110–113, 2015.

[36] T. Paul and T. Ogunfunmi, "Wireless LAN Comes of Age: Understanding the IEEE 802.11n Amendment," *IEEE Circuits and Systems Magazine*, vol. 8, no. 1, pp. 28–54, 2008.

[37] B. Vujicic, *Modeling and characterization of traffic in public safety wireless networks*. PhD thesis, 2006.

[38] W. G. Marchal, "An approximate formula for waiting time in single server queues," *AIIE transactions*, vol. 8, no. 4, pp. 473–474, 1976.