

Received October 4, 2021, accepted October 20, 2021, date of publication October 26, 2021, date of current version November 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3123054

# Towards Data-Driven Control of QoS in IoT: Unleashing the Potential of Diversified Datasets

MUHAMMAD ATEEQ<sup>1</sup>, HINA HABIB<sup>2</sup>,  
MUHAMMAD KHALIL AFZAL<sup>3</sup>, (Senior Member, IEEE),  
MUHAMMAD NAEEM<sup>4</sup>, (Senior Member, IEEE),  
AND SUNG WON KIM<sup>5</sup>

<sup>1</sup>Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur 47040, Pakistan

<sup>2</sup>Department of Computer Science and IT, The Government Sadiq College Women University, Bahawalpur 63100, Pakistan

<sup>3</sup>Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt 40740, Pakistan

<sup>4</sup>Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus, Wah Cantt 40740, Pakistan

<sup>5</sup>Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea

Corresponding author: Sung Won Kim (swon@yu.ac.kr)

This work was supported in part by the 5th Competitive Open Call of Federaion for the Fire Plus (FED4FIRE+) Project, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2021R1A6A1A03039493.

**ABSTRACT** Cognition is of paramount importance in modern communication systems for this brings the potential for adaptiveness and self-fine-tuning for dynamic reconfigurability. To achieve this feat, two primary tasks are to identify the influential configurable parameters and availability of comprehensive datasets representative of the real-world scenarios rather than simulated ones. For this article, an extensive dataset covering diverse settings of wireless sensor networks (WSNs) driven internet of things (IoT) is collected. It covers broad variations of 10 pre-configured communication parameters as well as some runtime information. In addition to legacy parameters (e.g., transmission power, and packet size, etc.), we also used two different medium access control protocols (i.e., carrier sense multiple access (CSMA) and time-slotted channel hopping (TSCH)), and routing metrics (i.e., objective function 0 (OF0), minimum rank with hysteresis (MRH), MRH with expected transmission count (ETX<sup>2</sup>)). Important quality of service (QoS) metrics like packet delivery ratio, throughput, and energy consumption against all combinations of the communication parameters are measured and recorded. A statistical analysis is carried out to identify the correlations among the communication parameters and QoS metrics. The results lay the foundation for the design of a data-driven framework for predictive QoS control in the IoT.

**INDEX TERMS** Data-driven design, quality of service, Internet of Things, wireless sensor networks, dataset.

## I. INTRODUCTION

Sensing coupled with communication gave birth to the wireless sensor networks (WSNs). WSNs are one of the major enablers for the realization of the internet of things (IoT), and thus central to realizing a futuristic smart world [1]. Internet Protocol version 6 (IPv6) as an identifier for the IoT devices can facilitate as much as 340 undecillion addresses, thus offering a humongous margin for massive extension in connectivity [2]. The number of devices connected to IoT is 8 billion, and growth is expected to reach 21 billion by 2025 [3]. The potential proliferation of connectivity

The associate editor coordinating the review of this manuscript and approving it for publication was Renato Ferrero<sup>1</sup>.

diversifies the design challenges to meet the required demands of the expanding IoT landscape. Scale, deployment scenarios, and application demands are some of the most significant factors defining the communication requirements and challenges for WSN driven IoT design [4].

Quality of service (QoS) has been a persistent challenge in the communication paradigms, and WSNs are no different. Error-prone links, limited bandwidths, constrained devices, content-rich applications, and incommensurate scaling render meeting QoS requirements more strenuous [5]. On the other hand, optimizing QoS metrics brings-forth the intractable problems that are NP-hard. The task outgrows further as applications require multiple, often conflicting metrics to be addressed simultaneously [6]. Thus, the legacy approaches

do not suffice to meet QoS requirements adequately in the prevailing evolution of WSNs and IoT at scale. Of late, artificial intelligence and machine learning (ML) have brought a data-driven approach to the research frontier. The solutions based on a data-driven approach have been proving to be very effective in different domains [7]. In addition to the potential for sufficient accuracy, the solutions based on data-driven techniques and ML are self-learning and adaptive [8] that are amongst the most important traits in the modern system design. These prospects motivate us to investigate the data-driven methods for a challenging problem like QoS in WSNs and IoT.

Before delving into the prospects of data-driven models for QoS facilitation in WSNs and IoT, it is important to understand the legacy research cycle in the context. The prevalent methods to address a challenge or solve a problem in the communication and networking domain are not inherently designed to benefit from data-driven methods. Primarily, the research focuses on the development of new protocols and approaches which are empirically and/or mathematically evaluated. Once comparison with existing proposals and/or formal modeling yields improvement, a new entry in the list of candidate solutions to that particular problem is made. A sufficient evaluation of any potential proposal is subject to extensive analysis and gauging that follows an exhausting and costly standardization cycle. Only after completing these long and hefty stages, a proposal can be put to production. This limits the potential for realizing tangible improvements in rational time and cost. Concepts like software-defined networking (SDN) and network function virtualization (NFV) have made it possible to dynamically configure many of the network properties and characteristics that were not possible before [9]. SDN separates the control plane from the data plane to achieve this feat. This opportunity demands the avenues of improvements by playing with the configurable properties available within the system. Therefore, banking merely on legacy research to identify, invent, and implement new proposals to cope with the problems must be complemented with the opening and latitude on offer in the form of SDN and NFV. Hence, we undertake to identify the potential on offer in the form of a data-driven approach for dynamic and seamless integration of possible run-time configurations and fine-tunings in the system for performance gains.

The data-driven approach paves the potential for making the best of whatever is available within a deployed and in-operation communication system. The separation of control and data planes provided by SDN enables dynamic configuration of various communication parameters (e.g., transmission power, packet size) and protocols (e.g., medium access control (MAC), routing, and modulation) [9]. Choosing a suitable configuration depends on the physical characteristics of the deployment and respective QoS demands. For example, different application-specific QoS requirements in unique physical settings can be met with particular configurations of parameters like MAC and routing protocols, modulation techniques, transmission power, traffic rates, packet

sizes, etc. Therefore, identifying the relationships of these parameters with QoS metrics can provide the basis for a system that can configure suitable values for these parameters, thus meeting the QoS goals.

To realize the data-driven solution to QoS in WSNs and IoT, performance data is a primary requirement. The legacy approach for experimental research in the domain of communication has used simulators as the main tool. The simulations based approach had been credibly accepted because of the limitations in availability and control on the required communication infrastructure. However, in the past decade, the focus shifted towards building custom testbeds and using real infrastructure for experimentation to have more realistic results [10]. For this work, we have collected a large-scale dataset against a wide range of configurations of different parameters and protocols. The experiments were performed using w-iLab.t [11], federated under FED4FIRE+ [12], a Horizon2020 project. The measurements include packet delivery ratio, throughput, and energy consumption against a wide variety of parameter configurations. Analysis is carried out to establish the relationship of parameter configurations with the QoS metrics. The results encourage the adoption of data-driven techniques for QoS in WSNs and IoT.

To summarize, the motivations and contributions of this work are the following:

- discuss the research approaches and methods used in wireless communication and networking, highlight the limitations and constraints, and identify the potential on offer in the form of a data-driven approach.
- highlights the need for comprehensive performance datasets to benefit from data-driven methods.
- collects a dataset against a wide range of values of different configurable parameters and protocols.
- analyze the relationship of communication parameters and protocols with QoS metrics.
- identify and exemplify a window of opportunity for further research based on a data-driven approach.

The rest of the paper is organized as follows. Section II highlights the research context in the form of a discussion on the legacy research cycle, its limitations, motivation to adopt data-driven methods. Section II concludes with an elaboration of the need for datasets for putting data-driven methods into use. Section III explains the experiments conducted. These details include the description of the testbed used, topologies considered, protocols variations tried, and parameters studied. A comprehensive analysis of the dataset collected is presented in Section IV. The exploration includes descriptive statistics of the QoS metrics observed, statistical correlations of the parameters with the metrics, pictorial representation of single as well as multiple parameters on the metrics. Section V concludes the article, brings out the lessons learned, and future direction of research.

## II. RESEARCH CONTEXT AND LITERATURE

In the following, we draw an account of the legacy research model and define the data-driven paradigm along with the

potential contribution in the QoS improvements in WSNs and IoT.

### A. LEGACY RESEARCH PARADIGM

For the sake of this work, the research efforts targeting QoS are classified by two method-keys. One is the types of proposals, and the second is the types of evaluation. In the following text-section, we elaborate on both aspects by highlighting some standard examples, analyzing the strengths and weaknesses, and theorizing the motivation for this study.

#### 1) TYPES OF PROPOSALS

Like the Internet, the design of the communication system for WSNs and IoT follows a layered modular structure. The role of each layer is specified according to the type of a communication system, and appropriate protocols are standardized in order to facilitate the communication needs [13]. However, as the systems evolved, and requirements diversified, the need for complex optimizations became inevitable. This led to the frequent adoption of rich mathematical techniques for specialized tasks (such as modulation) as well as at the base of various protocols (such as routing) [14]. Therefore, in the following, we categorize the types of proposals in the form of various protocols and mathematical approaches.

- **Protocols:** The intuition behind proposing new protocols that focus any QoS aspects makes sense because the communication stack has been designed in the form of layers in which each layer has a designated role offered through well-defined and standardized protocols. The concept of modular design and end-to-end arguments are still the main drive-forces behind this design philosophy of the communication stack [15]. Therefore, it is very common to find literature that proposes new solutions to facilitate QoS in the form of new protocols. The proposed protocols are found at different layers of the communication stack. However, application and MAC layers are the most common, closely followed by the routing layer [10].

Some prominent examples of standardized protocols for IETF IoT stack include constrained application protocol (CoAP) at the application layer, user datagram protocol (UDP) at the transport layer, IPv6, and routing protocol for low power and lossy networks (RPL) at the network layer, IPv6 over wireless personal area networks (6LoWPAN) at adaptation layer, and IEEE 802.15.4 as MAC and physical layers standard. The advent of SDN and NFV makes it possible to dynamically configure not just communication parameters but also protocols [16]. For example, at the MAC layer, in addition to legacy CSMA based variant, time division multiple access (TDMA) based variations (e.g., TSCH) are available. The research, aiming the improvements in various aspects as well as tailoring the solutions to serve certain scenarios continues.

Besides, cross-layer approaches (involving more than one layer) are also frequently found [14]. Although the

cross-layer approaches violate the modular design, however, various challenges paved the way for this innovation. In summary, there is a lot of literature addressing QoS in terms of protocols at the routing layer as well as at MAC and application layers [17].

- **Analytical:** The modular structure-based primary design principles for communication systems seem to assert that everything is facilitated in the form of well-defined and standardized protocols. Here the role of some of the protocols is to facilitate a negotiation between the communicating nodes (e.g., CoAP), while others required strong formal analysis for optimized operation (e.g., optimization conflicting QoS metrics [6], and modulation techniques, etc.). Initially, the formal modeling of any mathematical optimization was coupled with the protocol it was designed for. However, as the networks evolved, the need for separate analytical solutions (based in mathematics and statistics) became inevitable. These analytical solutions are being used to serve the needs of various protocols. Some of the examples requiring analytical solutions include area coverage [18] and fair rate allocation [19]. Various aspects of QoS, for example, reliability, latency, throughput, energy consumption, and lifetime maximization have been researched with the objective to optimize these metrics [20], [21].
- **Combination of Protocols and Analytical Approaches:** It is evident through literature that communication protocols (new as well as existing) are often supported by some sort of mathematical optimization in the background. Among others, energy conservation [18], [22], and network latency [22], [23] are some of the examples where the mathematical foundation is intrinsically used to design protocol.

#### 2) TYPES OF EVALUATION

For the sake of evaluation of any proposed research, three possible routes are common, namely: mathematical, simulations, and experimental evaluation on real devices [10]. The findings from a survey on performance evaluation methods in ad-hoc and WSNs reveal that there is a growing trend of experimental evaluation compared to simulations [10]. It also highlights that using custom testbeds keeps the reproducibility concern alive. This requires the development and usage of standard deployments of resources with availability ensured. In the following, we describe all three performance evaluation approaches, highlight some important literature, and present the pros and cons:

- **Mathematical:** Formal modeling is considered an important step towards evaluating any new proposal to solve problems (including QoS) in the domain of communication. Methods based on mathematics and statistics are frequently applied for this purpose. According to [10], 94% of the literature used mathematical techniques for performance evaluation. As an example, the most famous distributed coordination function in CSMA is

evaluated using Markov chain modeling [24]. Similarly, discrete-time Markov chains are used to evaluate the performance of the TSCH [25], which is an emerging MAC variant for WSNs and IoT. The use of analytical, probabilistic, optimization, evolutionary, as well as artificial intelligence-based techniques is frequent for QoS considerations in WSNs [6], [26]. The chief advantage of mathematical evaluation is that it is theoretically correct and consistent within its assumptions. However, the applicability in real situations needs separate evaluation. The major drawback comes forward in cases where problems are complex and mathematical approaches do not scale [6], [10], [26]. Multi-objective optimization (MOO) is an example of such situations [6].

- Simulations: Simulations have been arguably the most used medium for evaluation in research studies for QoS in WSNs. According to the literature surveyed in a recent article [6], well over 90% of publications used simulations for evaluation of the proposed solutions. As the WSNs got conceived and went through initial development and evolution, simulations were the primary option to demonstrate the proof-of-concept because real resources were scarce. Some of the popular simulators that have been instrumental in this regard include ns-2, ns-3, OMNeT++, Cooja, and others [10]. As simulations have been the most popular medium for evaluation because of the limitations associated with the availability and manageability of real infrastructure, the research community has made consolidated efforts to design credible and open-source simulators like ns-3. Alongside others, these have served the needs for evaluations of various solutions proposed to solve networking problems over past decades. Besides, the use of custom software tools for the evaluation of specific wireless scenarios has remained a common alternative. However, the generalization of the results from a simulations study is not considered parallel to experimental evaluation. Therefore, as the availability of resources grew, the tendency to use real infrastructure has increased [10].
- Experimental Evaluations: The trend of using real infrastructure (open or custom) for evaluation and validation of research proposals has grown significantly overtime [10]. There is also a growing trend of research funds to build open testbeds that encourages experimental evaluations. FED4FIRE+ [12] is a considerable example of a federation of such resources. It federates a large number of testbeds for experimentations in virtually all aspects of communication and networking. Not just these resources are openly available to the research community but their use is encouraged through competitive funding for innovative experiments [27]. These open calls [27] have been a cause for the use of state-of-the-art real infrastructure for experimental evaluation of scientific innovation in networking research. Also, there is a wide range of open and custom testbeds that includes MoteLab, TWIST, FIT IoT-LAB, SmartSantander,

w-iLab.t, Emulab, and ORBIT [10], [12]. Although the experimental evaluations are not parallel with the real-deployment performance, however, are still considered more realistic. This is because the evaluations are carried out using real infrastructure and more often in relatively realistic environments. The federated resource (as in the case of FED4FIRE+) facilitates reproducibility, which is considered one of the most important aspects of the credibility of results.

## B. DATA-DRIVEN PARADIGM

In a data-driven research paradigm, interactive domain data is used to drive the decision making by facilitating adaptation as the system evolves [28]. The data-driven research paradigm has been popular of late and has been applied to problems like video bitrate, and relay selection in internet telephony [29]. Data-driven techniques have been used to solve various performance and security-related problems in WSNs and IoT [30], [31]. Recent research efforts highlight the potential of data-driven paradigm in the realization of smart cities [8], and future cellular technologies like 5G [32]. The proposed solutions and improvements are based on the performance data collected mostly from the real deployments in an interactive manner. The system works in a loop where data collection and learning go side-by-side. The idea is to achieve an adaptive self-learning design that makes the best of available resources [33]. Based on the findings in existing literature [8], [29], [32], [33], it is understandable that in any data-driven system, data is of paramount importance.

Fig. 1 shows the difference between the legacy research cycle and the data-driven research cycle, and also highlights the complementing that the later offers for the former. In the legacy paradigm, an idea begins with some theoretical conception, undergoes mathematical modeling, followed by simulation and/or emulation based evaluation. If the proposal meets certain requirements, it qualifies for further analysis in the form of experimental evaluation using testbeds and ultimately on real deployments. Such a representation has also been found in other literature [26]. It is well understood that an extremely small fraction of the published research literature makes it to the final stages of experimental evaluation. There is a window of opportunity to identify and utilize the potential of any improvements within the system that has already been standardized. A data-driven paradigm is conceived to recognize these margins and enable fine-tunings in a working system interactively. Fig. 1 reveals the possible cycle for the proposed data-driven paradigm. It begins with the data collected from a WSN deployment, carries out statistical analysis to identify the relationships among the configurable parameters and QoS metrics. Having identified the suitable features, various ML models can be applied to train the data, and hence a suitable model can be chosen for configuring the parameters interactively and adaptively.

In summary, as against the legacy paradigm, the data-driven paradigm uses the existing deployments and manipulates the configurable parameters to meet the QoS

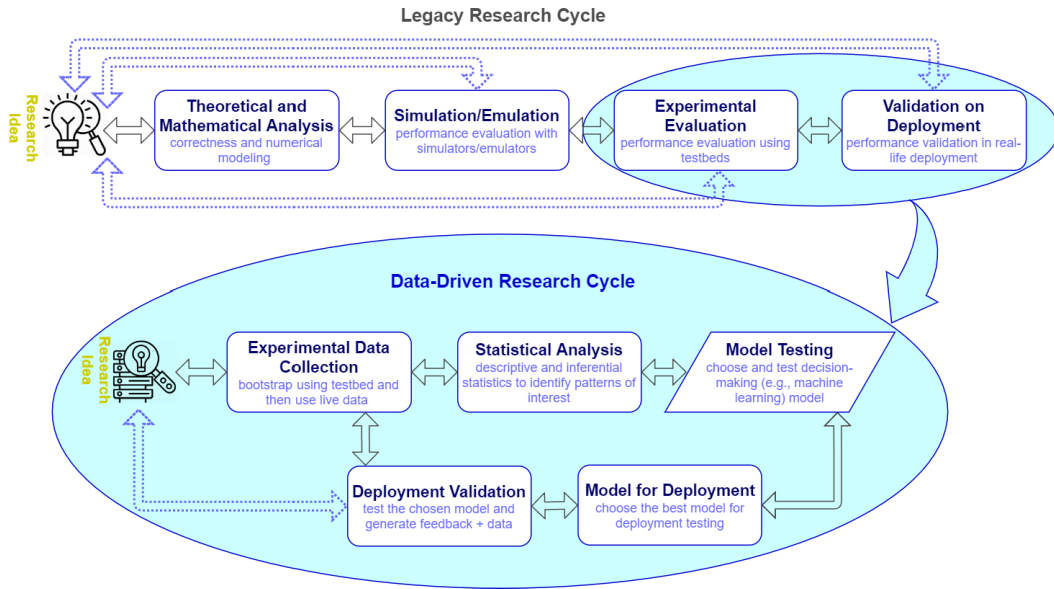


FIGURE 1. The steps in legacy research cycle vs data-driven research cycle.

goals. To achieve this feat, an experimental evaluation of the idea is carried out using any existing or custom infrastructure without the need for designing new protocols from scratch. Therefore, the idea of data-driven research can follow a short and more productive research cycle making the best of available deployments by manipulating the configurable parameters of the communication stack.

### C. NEED FOR DATASETS

To design any data-driven, intelligent, and adaptive system, the need for bootstrap data is vital. Once the system starts, it can make use of interactive data for adaptations. There is a great lacking when it comes to performance-related datasets in WSNs and IoT. Most of the available datasets are related to sensing and security [34]. Some of the prominent datasets related to sensing and security, frequently cited in research include [36], [37]. However, performance-related datasets that utilize various communication parameters to understand the variations in QoS metrics, are scarce. To the best of our knowledge [38] is the only publicly available dataset that captures the QoS information against vast variations of a wide set of communication parameters. However, the topology is limited to two nodes only, directly communicating with each other. WSNs are often deployed in situations where a large number of nodes are required, and routing is needed for nodes to relay their data to the sink. Moreover, modern implementations of operating systems tend to provide MAC protocols like TSCH, in addition to the legacy CSMA protocol to provide more synchronized and secure communication. Similarly, there are multiple options for routing and other network operations. Therefore, it is of prime importance that performance data be available for diverse settings that include complex multi-node and multi-hop scenarios, against

multiple MAC protocols and routing metrics in addition to other configurable parameters like transmission power, traffic rate, and packet size.

Considering the limitations in existing data resources and the growing need for a data-driven solution, we have carried out experiments to collect a dataset covering a wide range of parameters and settings. In the next sections, we explain the details of the experiment, dataset collected, and statistical findings.

### III. EXPERIMENT DETAILS

In the following, we describe the experiment details, including the testbed, experiment topologies, and parameter configurations.

#### A. TESTBED

The experiments were carried out using w-iLab.1 [11], located in iGent building, and is maintained by the IDLab University of Gent, Belgium. The testbed is part of the facilities federated by the FED4FIRE+ project [12] and is infrastructured for reproducibility. w-iLab.1 has a rich set of resources available for diverse experiments in the domain of wireless technologies. The testbed resources are located on the ground floor (also called data center (DC) floor), floor 9 (F9), floor 10, and floor 11 of the iGent building. We used sensor nodes (Zolertia Re-Motes) from DC and F9 floors. DC floor is almost interference-free, whereas F9 is located in an office space and generally has a lot more interference from various sources, WiFi in particular.

#### B. TOPOLOGY

The two topologies of the experiment are shown in Fig. 2 and Fig. 3, respectively. The sending nodes in both figures

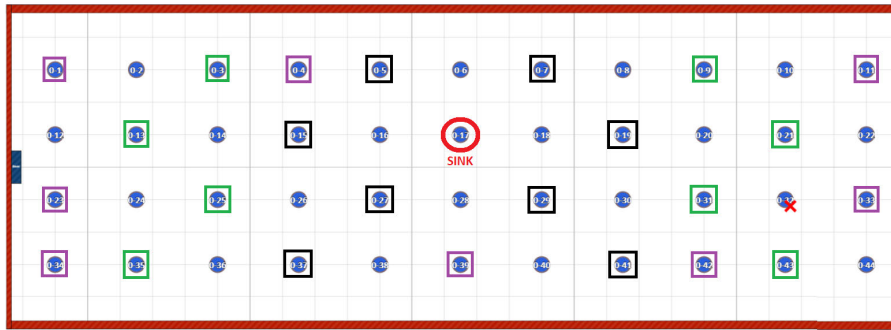


FIGURE 2. Experiment topology used on data center floor.

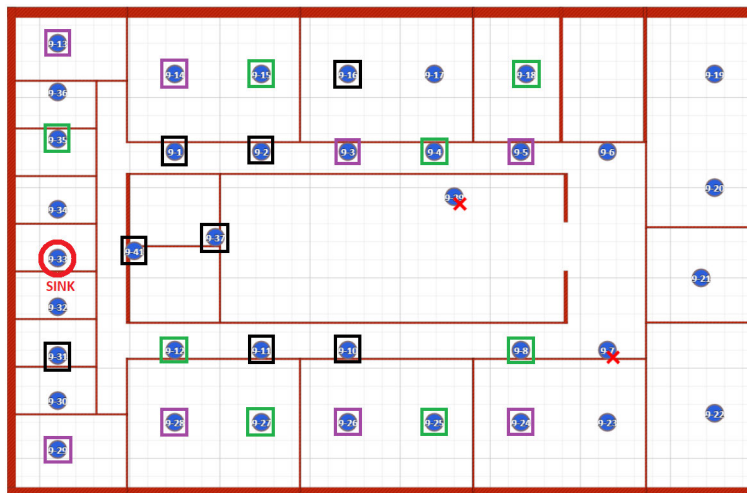


FIGURE 3. Experiment topology used on floor 9.

are indicated using squares (black, green, and purple), and receivers (sinks) are indicated using red circles. The average node-to-node distance in topology 1 (Fig. 2) is a little over 2 meters, whereas the receiver/sink is placed in the center of the network. On the other hand, in topology 2 (Fig. 3) sink is placed towards the left far-end, and the average node-to-node distance is a little over 4 meters. Thus, two topologies offer variation in average node-to-node distance and placement of the sink. Besides, the layout of topology 1 (Fig. 2) is more like a star network, whereas the layout for topology 2 (Fig. 3) is more like a bus. The numbers of sending nodes are varied between 8 (black), 16 (black+green), and 24 (black+green+purple).

It is evident from the description that the topology 2 (Fig. 3) is much more challenging compared to the topology 1 (Fig. 2) in terms of node-to-node distance, placement of the sink (and hence the layout), and interference patterns.

### C. EXPERIMENT PARAMETERS

The idea of this experiment was to carry out foundation work to stem data-driven research aiming to facilitate and improve adaptive QoS control in WSNs driven IoT. In our previous

endeavors [39]–[41] we have used the single-hop performance data [38] for statistical analysis and predictions. However, considering the limitations, we extended the idea to a broader perspective of performance by including more important metrics: i.e., throughput (THP), and detailed energy consumption (EC) in addition to packet delivery ratio (PDR). To understand the behavior of these QoS metrics, we also included diverse parameters including packet inter-arrival time (IAT), packet size (PS), maximum transmissions (MT), number of active nodes (NAN), network density (Topology), MAC protocol (MACP), RPL objective function (RPL-OF), transmission power (TP), and distance (DT). These parameters are detailed in Table 1. As the nodes are fixed, actual measurements of DT were used in the analysis.

The approximate runtime of the experiment is based on the number of combinations of parameter values. Based on the detail of distinct values for each parameter in Table 1, we present the detail of combinations (TComb) as follows:

$$TComb = TP \times NAN \times Topology \times MACP \times MT \times RPL - OF \times PS \times IAT$$

TABLE 1. Detail of parameters used in the experiments.

Protocol Layer	Parameter	Number of Distinct Values	Proposed Values	Rationale	
Physical	TP: Transmission Power (dBm)	4-5	Floor-DC: -7, -5, -3, -1, 1 Floor-9: -1, 1, 3, 5, 7	Average node-to-node distance on floor-9 is almost double compared to DC floor	
	NAN: Number of Active Nodes	3	8, 16, 24	8, 16 and 24 sender nodes are used for transmission each in 2 different topologies	
	NT: Number of Topologies	2	Data Centre and Floor-9		
MAC	MACP: MAC Protocol	2	CSMA, TSCH	To observe the behavior of both MAC protocols	
	MT: Maximum Transmissions	3	2, 4, 8	To characterize packet delivery behavior	
Network	RPL-OF	3	OF-0, MRH (ETX, ETX <sup>2</sup> )	To observe the behavior of different RPL-OF	
Application	PS: Packet Size (Bytes)	3	25, 50, 75	802.15.4 accommodates small packet sizes	
	IAT: Inter-Arrival Time (sec)	4	NAN=8	1, 2, 3, 4	The load ranges from very low to very high
			NAN=16	2, 4, 6, 8	
NAN=24			3, 6, 9, 12		

$$= 5 \times 3 \times 2 \times 2 \times 3 \times 3 \times 3 \times 4 \approx 6500 \quad (1)$$

This means that the experiment ran against these almost 6500 combinations of the parameter configurations. We ran the experiment against each configuration for 600 seconds (Time), intending to be able to observe a consistent behavior. Therefore, the total time (TotTime) required for running the experiment and overheads due to schedule, technical difficulties, erroneous runs, is as follows:

$$\begin{aligned} TotTime &= TComb \times Time + Overhead \\ &= 6500 \times 10min + 12000min \\ &= 65000min + 12000min \\ &= 77000min \approx 1280hours \\ &\approx 53days \end{aligned} \quad (2)$$

#### D. PARAMETERS AND METRICS OBSERVED

In addition to the pre-configured parameters, we noted node id (NID) for each node because each node is distinctly located to the sink as well as its respective neighbors. We also calculated the maximum number of transmittable packets (MP) according to the current configuration and the actual number of packets transmitted (PT) in each experiment run as:

$$MP = IAT \times Time \quad (3)$$

Packets received (PR) were measured, and the packet send ratio (PSR), PDR, and THP were calculated as follows:

$$PSR = \frac{PT}{MP} \quad (4)$$

$$PDR = \frac{PR}{PT} \quad (5)$$

$$THP = PR \times PS \quad (6)$$

Moreover, the Contiki Energest module [42] was used to measure CPU as well as radio utilization. CPU usage is measured distinctly for three different supported modes as regular

CPU (RCPU), low power mode (LPM), and deep low power mode (DLPM). Radio usage is measured for listen (LSTN), transmit (Tx), and off (Off) states. All these measurements are in seconds.

## IV. RESULTS AND ANALYSIS

### A. DESCRIPTIVE STATISTICS

In this section, we describe the data with the relationships among parameters and metrics. The descriptive statistics of the metrics are shown in Table 2. PDR is a ratio and ranges between 0 and 1. The mean is nearly 0.5, whereas, median (0.6) is slightly higher. The standard deviation (0.46) indicates that the values are relatively spread over the range. Throughput, in addition to PDR, caters to the variations in PS as well. The values of THP range between 0 and 48741 with a high standard deviation (5026). The median is 810 and the mean is 3014. This indicates a sparse cluster of bigger values that increases the value of the mean.

The energy consumption is separately measured for CPU and radio usage using the Energest module. CPU usage ranges between 1 and 460 seconds with a median and standard deviation of 28 each, and a mean of 34. LPM is common for both CSMA and TSCH protocols and ranges between 0 and 635 with a mean and median of 346 and 529, respectively. This indicates a larger cluster of smaller values that keeps the mean low compared to the median. DLPM is supported by TSCH only and ranges between 0 and 594. The mean is 225 and the mode is 0. This is because in the case of CSMA the value for DLPM remains 0.

The usage of radio is separately observed for LSTN, Tx, and Off periods. The values for the LSTN period range between 6 and 642 second with mean and median 363 and 599, respectively. A higher value for the median indicates the varying behavior of MAC protocols because TSCH utilizes LSTN periods smartly compared to CSMA. Values for Off periods range between 0 and 615 with a mean of 240 and a median of 0. Again, the reason for such difference in mean and median is because of MAC protocols. CSMA does not

TABLE 2. Descriptive statistics of QoS metrics.

Metric	Min	Max	Mean	Median	StDev
PDR	0	1	0.496	0.6	0.457
THP	0	48741	3014	810	5026
CPU	1	460	34	28	28
LPM	0	635	346	529	249
DLPM	0	594	225	0	250
Listen	6	642	363	599	268
Transmit	0	21	1.214	1	1.406
Off	0	615	240	0	267

turn off the radio and is always in listening mode, whereas, TSCH turns off the radio frequently. The values for transmit periods are less significant because the time is measured at the granularity of seconds, whereas, transmit times can better be measured at finer scales like nanoseconds.

B. CORRELATIONS

The statistical correlations of the parameters with QoS metrics are revealed in Fig. 4. The vertical axis list the parameters and the horizontal axis reveals the metrics. Although almost all the parameters seem to show relationships (of varying degrees) with different metrics, however, some prominent correlations can be pointed out. For example, the effect of MAC protocol on all metrics, particularly on EC (LPM, DLPM, LSTN, and Off), is visible. Other parameters that affect EC appear to be MT, MP, and DT. In the case of PSR, PT (a direct indicator of the total number of packets transmitted) comes off as the most influential parameter. Other parameters influencing PSR include DT, NID (like DT, also an indicator of placement of node), NT (representing the layout of the network), and MAC. PDR seems to have a relationship pattern similar to PSR, although, intuitively, the correlation values are comparatively lesser than those for PSR in most cases. In addition to PT, DT, NID, NT, and MAC, PS seems to have a strong correlation with PDR. This indicates that PS plays a less significant role in deciding the opportunity to transmit. However, the effect of transmission success is more significant. Fig. 4 shows the effect of individual parameters and not combinations that may present more interesting cases. In the case of THP, many parameters have correlation statistics resembling those with PDR. The noticeable differences appear in the form of MP and a much stronger relationship with PT because both these parameters deal with the number of packets. Also, understandably IAT impacts throughput more as against PS in the case of PDR.

To better understand the effect of multiple parameters (simultaneously) with QoS metrics, we use pictorial 2-dimensional (box-plots) and 3-dimensional (bar-charts).

C. 2-DIMENSIONAL RELATIONS

In the section, we show the 2-dimensional relationships among parameters and QoS metrics using boxplots. A box-plot describes the data in the form of box and whiskers where

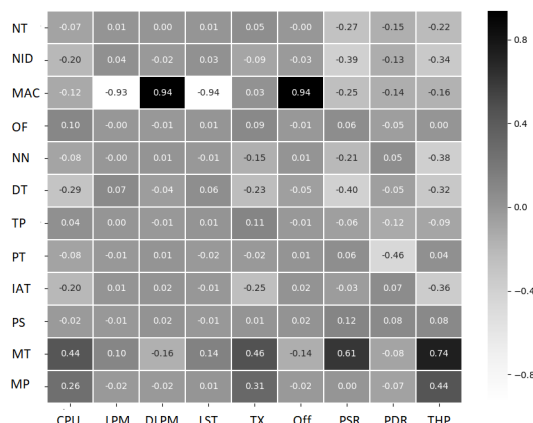


FIGURE 4. Correlation among parameters and QoS metrics.

data within a box represents 25th (bottom of the box), 50th (middle of the box), and 75th (top of the box) percentiles of data, and whiskers depict the spread on both extremes.

Fig. 5(a, b, c, d, e, f, g, h, i, and j) shows the effect of different parameters with QoS metrics. The relation of the combinations of topologies and MAC protocols on PDR is revealed in Fig. 5a. CSMA yields much better PDR than TSCH and topology 1 (SC: Short average node-to-node distance and sink located in Center) appears to provide higher PDR compared to topology 2 (LD: Large average node-to-node distance and sink located at Distance from center) because of potential suitability due to short distances, placement of receiver/sink node, and interference pattern. Fig. 5b shows the effect of combinations of topologies and PS on PDR. It is again intuitive to exhibit that smaller PS and topology 1 (SC) result in higher PDR compared to larger PS and topology 2 (LD). The drop in PDR becomes more pertinent in the case of LD with an increase in PS. Fig. 5c reveals the relation of topologies and MT with PDR. It confirms the already understood intuition that SC and larger MT values produce higher PDR. Fig. 5d shows the effect of MAC and MT on PDR. This also maintains the trend of the higher tendency of PDR with CSMA and larger MT values as against TSCH and small values for MT. The effect of MAC and RPL-OF variations on PDR is unraveled in Fig. 5e. In the case of CSMA, OF0 performs slightly better than MRH, whereas, ETX achieves lesser PDR compared to both others. In the case of TSCH, the trend seems to have reversed. The advantage of MRH over OF0 is marginal. However, ETX achieves significantly PDR compared to both others. To further understand the behavior of RPL-OF, Fig. 5f shows its relationship with PDR in combination with topology variations. In the case of topology 1 (SC), the simpler the better tendency prevails. PDR falls monotonically from OF0 to MRH and ETX. In the case of topology 2 (LD), having long distances, the performance of OFs reverses, as ETX performs much better than both MRH and OF0. The effect of TP on PDR in both topologies is revealed in Fig. 5g. More the TP, the better the PDR in both topologies. However, in the case



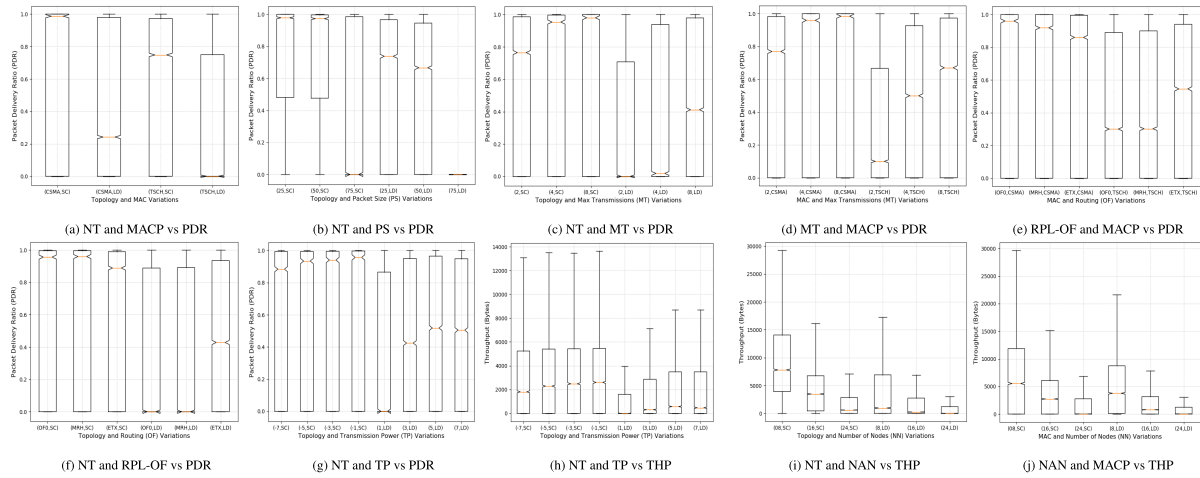


FIGURE 5. Relationship between communication parameters, and packet delivery ratio, and throughput.

of LD, TP of 7 dBm (which is the maximum), seems to cause a little performance decline. This phenomenon can occur in wireless communication systems when TP is too high.

Fig. 5h illustrates the relation of topologies and TP combinations on THP. As expected, the THP increases as TP is increased with better yielding in the case of SC compared to LD. Similarly, Fig. 5(i and j) show the effect of NAN in combinations with topologies and MAC protocols on THP, respectively. Again, as anticipated, THP increases in an inverse proportion to NAN. A positive hike in THP is witnessed when CSMA is used as MAC and SC as topology as against TSCH and LD, respectively.

### D. 3-DIMENSIONAL RELATIONS

In the following, we extend the 2-dimensional relationships exhibited previously by using 3-dimensions. Scatterplots are used for this purpose which shows all the values in the form of markers (circles in this case). These relationships are depicted in Fig. 6(a, b, c, d, e, f, and g).

In Fig. 6a, the combined effect of PS and DT from the receiver/sink on PDR is revealed. PDR keeps falling as both PS and DT increase. This is consistent with the common understanding of the increasing probability of transmission success with smaller transmission units being communicated at shorter distances. An increase in DT induces two problems in this case. One, the signal fade as DT increase. Second, it becomes more likely that a packet has to travel more number of hops, thus hampering the chances of successful transmission. Fig. 6b depicts the relationship of DT and TP on PDR. The finding remains consistent with the trends already revealed as PDR drops with smaller values of TP and larger DT. Fig. 6c zooms in on the effect that NAN casts on PDR alongside DT. PDR keeps falling as NAN and DT increase.

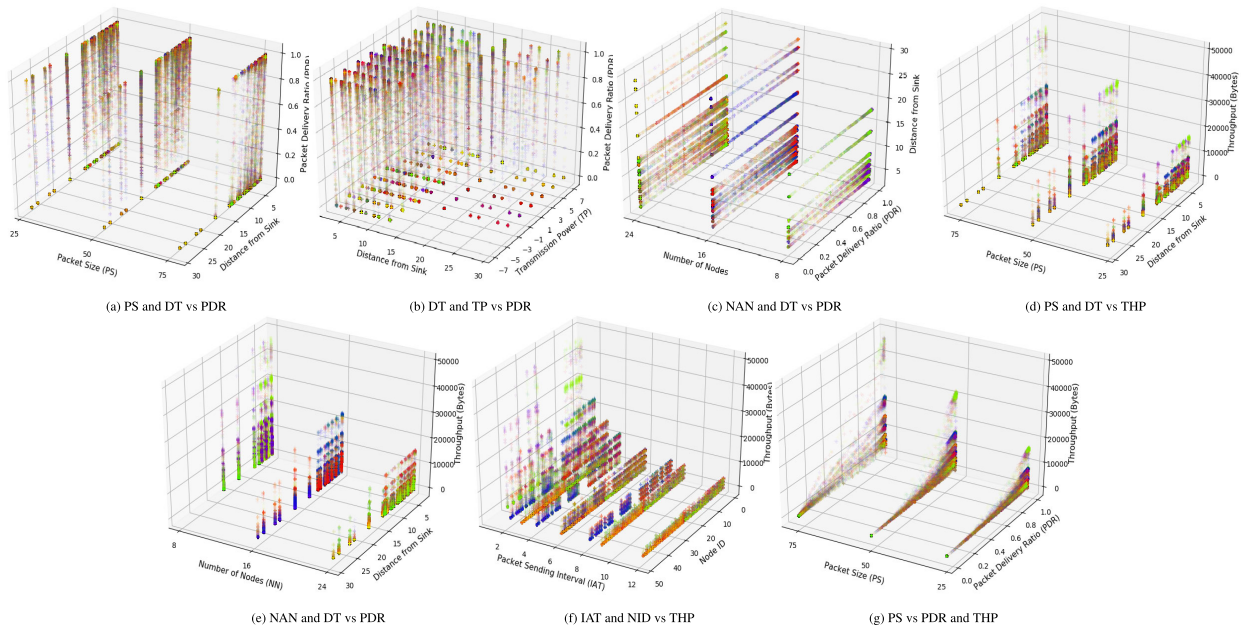
The combined effect of PS and DT on THP is revealed in Fig. 6d. Where THP decreases with increasing DT, it exhibits a relatively scattered rise with increasing PS. Although we previously witnessed that PDR falls with increasing PS,

however, PDR just counts the fraction of the number of packets as a ratio of total numbers of Packets Transmitted (PT) and does not take into account the magnitude of data. THP highlights this effect as a reverse relation with PS. This factor presents a reason to consider PDR and THP as separate metrics. Fig. 6e shows the effect of NAN and DT on THP. The resultant figure appears to resemble Fig. 6d. The trend with DT does not change, however, for smaller NAN, THP is higher. It is important to focus that the overall traffic in the network remains the same as the number of nodes varies. Therefore, the rise in THP with less NAN asserts that it better suits THP compared to accommodating more NAN even without changing the overall traffic volume. Fig. 6f describes the relationship of IAT and NID with THP. THP is higher with smaller IAT. As reasoned formerly while explaining Fig. 6d, smaller IAT does not simply mean more traffic. It means that less NAN is being used as transmitters instead of using more NAN to generate the same amount of traffic with higher IAT per node. NID on the other hand is not very evenly distributed. However, the variations do occur in THP with differing nodes.

Finally, to further stem the difference between PDR and THP, both these metrics are plotted against PS in Fig. 6g. THP rises with increasing PS, whereas PDR falls. This confirms the importance of both these metrics to be considered separately.

### E. MULTI-FACETED ANALYSIS OF METRICS

Building further on the analysis carried out previously, we focus on the PDR, THP, and EC in fine-grained detail in Fig. 7(a, b, c, d, e, f, g, and h). We use clustered bar charts for a clear and comprehensive comparison. As highlighted already, PDR just conveys the number of packets delivered as a ratio of the total transmitted packets. In addition to PDR, we focus the PSR and THP to carry out an in-depth analysis of the influence of various parameters. Fig. 7a unravels the effect of MAC and RPL-OF in both topologies (i.e., SC and LD) on PDR, PSR, and THP. The values represented on the bar are



**FIGURE 6.** Relationship of multiple communication parameters with packet delivery ratio and throughput.

the mean of all values against the respective configurations of parameters. In the case of topology 1 (SC), PSR falls as OF changes from OF0 to MRH, and then to ETX when CSMA is used as MAC. However, in the same pattern, the PSR rises when TSCH is used as MAC. However, the pattern of THP is not the same in the case of CSMA, whereas, in the case of TSCH, THP does not show any significant variation. In the case of topology 2 (LD), ETX yields a higher value for PSR, followed by OF0 and MRH for CSMA. The same sequence exists for PDR and THP. ETX is expected to do better by selecting more reliable links. This notion holds in the case of topology with more distant nodes. On the other hand, for TSCH, the values for PSR increase for OF0, MRH, and ETX. The same holds for PDR and THP. This discourse suffices to narrate that the combination of MAC, RPL-OF, and topology variations affect communication reliability in varying ways. The exact reasons for variations in PSR and PDR require deeper analysis of MAC and routing logs, which are not in the scope of this study.

Fig. 7b describes the packet delivery performance in relation to topology, NAN, and IAT variation. It is persistent that the overall traffic in the network remains the same as the NAN changes (as described in Table 1). For example, when NAN is 8 and IAT is 1, it generates the same amount of packets when NAN is 16 and IAT is 2 or NAN is 24 and IAT is 3. It is recurrent for the packet delivery performance on topology 1 (SC) is better than that of topology 2 (LD). Apart from that, the lower the traffic (IAT) and the number of active nodes (NAN), both PSR and PDR improve. THP presents an interesting case. For topology 1 (SC), the THP almost halves as the NAN keeps increasing with the same cumulative traffic. In the case of topology 2 (LD), THP suffers even more. This

analysis visible from Fig. 7b reveals the opposite trends in PDR and THP. PDR increases with less IAT, whereas, THP decreases.

Fig. 7c depicts the effect of MAC and RPL-OF for both topologies (SC and LD) on detailed radio behavior. The average time (sec) spent in listening, transmitting, and keeping the radio off is shown separately. The average total time for which the radio statistics are displayed is about 605 seconds. In the case of CSMA, the off periods are 0 because the radio is always listening, whereas, in the case of TSCH, the radio is kept off more than 70% of the time on average. The transmit periods for CSMA are smaller on SC topology compared to the LD topology. However, the trend reverses in the case of TSCH. Also, In the case of CSMA, the listen periods are slightly higher for ETX OF. Although, the same can be said about TSCH, however, the statistical difference is not very significant. The listen periods for TSCH are within 7% of the total time for topology 1 (SC). However, these times rise more than 4 times in the case of topology 2 (LD). In the case of CSMA, all the time other than transmit periods is used for listening.

The characterization of CPU usage is revealed in Fig. 7d. The details are presented at 3 levels (with descending use of power), namely, CPU, low power mode (LPM), and Deep LPM (DLPM). The parameters against which these observations are presented are topology, NAN, and IAT. The values of per-node CPU usage decrease with an increasing NAN and decreasing the traffic rate (IAT). Likely, the values for LPM and DLPM carry an opposite trend compared to CPU usage. As CPU usage decrease, both LPM and DLPM increase. The trend across topologies is that per-node CPU usage falls as the focus moves to topology 2 (LD) from topology 1 (SD),

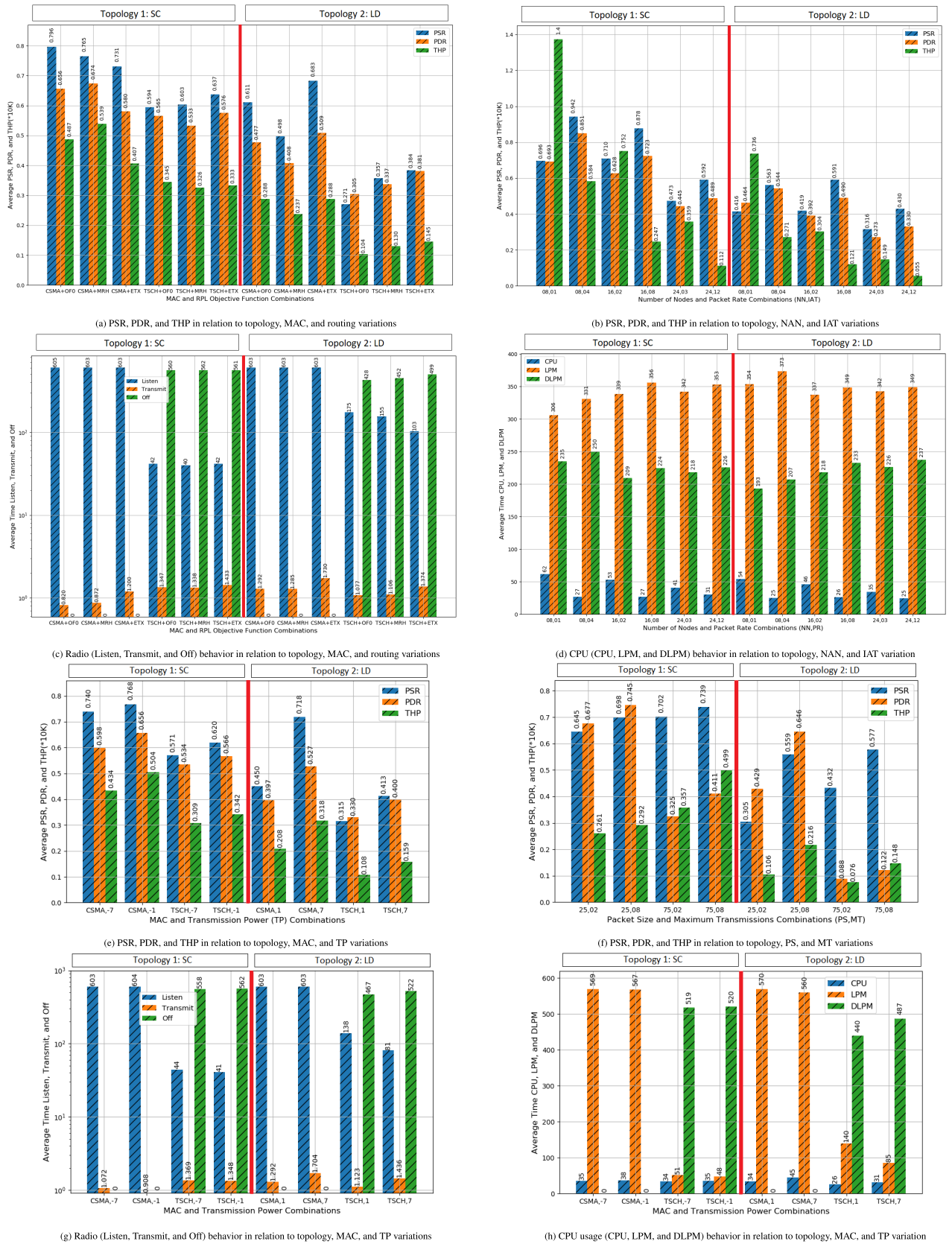


FIGURE 7. Multi-faceted analysis of QoS metrics in relation to communication parameters.

and the values of LPM and DLPM increase accordingly. This indicates that increasing NAN, as well as switching to topologies with bigger distance, tends to lessen the node activity in some sense.

In the following, we continue the analysis of the relationship between parameters and metrics. Fig. 7e highlights the packet delivery performance against MAC and TP alongside the topologies. For MAC protocols and topologies, the minimum and maximum TP levels are chosen. The trend of better performance for SC is preserved. Besides, all PSR, PDR, and THP improve with an increase in TP. The effect of PS and MT is depicted in Fig. 7f. The rise in PDR is observed with smaller PS and greater MT. The trend of PSR remains high, in general. The real difference is in THP where PDR falls with an increase in PS, THP increases. This is because fewer packets of bigger size can yield more THP compared to more packets of small size delivered. Based on the observations from Fig. 7e and Fig. 7f, it can be concluded that PDR increases when CSMA is used as MAC, and the values of TP and MT are raised. However, when TSCH is used as a MAC, and the values of PS increase, PDR falls. This fact remains consistent across topology variations.

In Fig. 7g, the effect of MAC and TP is shown on radio behavior. As per the design, there is no off period for CSMA and the nodes are almost always in listen state. Transmit time does not vary much for topology 1 (SC). However, in the case of topology 2 (LD), transmit time tends to increase with a rise in TP. In the case of topology 1 (SC), the listen and off periods for TSCH remain about 40 and 560 seconds, respectively. For topology 2 (LD) the listen periods are higher and so are the off periods. However, as TP increases from 1 dBm to 7 dBm, the listen period shrinks by almost 40% and the off period adjusts accordingly.

In Fig. 7h, the relationship of MAC and TP with the CPU brings that in the case of CSMA, there are no DLMP periods that make the task of energy conservation difficult. In general, the regulation CPU consumption is more for CSMA compared to TSCH. CPU consumption also increases slightly as TP increases. The DLMP periods also rise a bit as TP increases.

We carried out a wide spectrum analysis of different communication parameters including MAC, RPL-OF, TP, PS, MT, IAT, NAN, NT, and NID in connection with PDR, THP, and EC (divided into radio and CPU usage). The purpose of this analysis primarily was to identify whether the variation in the values of these parameters and their combinations have some certain effect on the outcome of different metrics under consideration. In seeking an answer to this question, we started with a correlation matrix (Fig. 4) to witness the correlation scores. Next, we used box plots (Fig. 5) to reveal the variation in the values of metrics against certain parameters. Following this, we used 3-D scatter plots (Fig. 6) to extend the analysis by multiple parameters simultaneously. We then used the bar-charts (Fig. 7) to show the trend in metrics including detailed radio and CPU behavior against a vast combination of parameters of interest. The analysis reveals

that the variations and combinations of different parameters do affect the outcome of QoS metrics in different ways.

## V. CONCLUSION AND FUTURE WORK

In this study, we highlighted the growing potential and significance of the data-driven research paradigm for performance improvements in the domain of WSNs in contrast to the legacy research cycle. We signified the time and cost-oriented limitations in the legacy approaches and compared it with the potential advantages offered by the data-driven paradigm to facilitate adaptive, robust, and real-time improvements interactively. The importance and rarity of datasets were accentuated in the context. We then presented the experiments conducted on a state-of-the-art real testbed (i.e., w-iLab.t) and dataset collected against a comprehensive set of parameters including inter-arrival time, packet size, maximum transmissions, number of nodes, network density/topology, MAC protocols, RPL objective functions, transmission power, and distance. In addition, the values for node-ids, maximum transmittable packets, the actual number of packets transmitted were observed for all experiment runs. For almost 6500 configurations of these parameters, we observed packet delivery ratio, throughput, and detailed energy consumption. A comprehensive statistical and pictorial analysis reveals the effective relationships among these parameters and QoS metrics. This work sets the direction for the adoption of data-driven research for solving various problems in the domain of WSNs and IoT and contributes a large scale dataset. In the future, we shall use machine learning for predicting the QoS metrics based on the parameters available in the dataset.

## REFERENCES

- [1] C. Zhu, V. C. M. Leung, L. Shu, and E. C.-H. Ngai, "Green Internet of Things for smart world," *IEEE Access*, vol. 3, pp. 2151–2162, 2015.
- [2] D. Minoli, *Building the Internet of Things With IPv6 and MIPv6: The Evolving World of M2M Communications*. Hoboken, NJ, USA: Wiley, 2013.
- [3] C. Camara, H. Martin, P. Peris-Lopez, and L. Entrena, "A true random number generator based on gait data for the Internet of you," *IEEE Access*, vol. 8, pp. 71642–71651, 2020.
- [4] F. Al-Turjman and A. Radwan, "Data delivery in wireless multimedia sensor networks: Challenging and defying in the IoT era," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 126–131, Oct. 2017.
- [5] M. Raza, N. Aslam, H. Le-Minh, S. Hussain, Y. Cao, and N. M. Khan, "A critical analysis of research potential, challenges, and future directives in industrial wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 39–95, 1st Quart., 2018.
- [6] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 550–586, 1st Quart., 2017.
- [7] T. Li, C. Zeng, Y. Jiang, W. Zhou, L. Tang, Z. Liu, and Y. Huang, "Data-driven techniques in computing system management," *ACM Comput. Surv.*, vol. 50, Oct. 2017, Art. no. 32017.
- [8] W. Tu, "Data-driven QoS and QoE management in smart cities: A tutorial study," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 126–133, Dec. 2018.
- [9] M. Miguel, E. Jamhour, M. E. Pellenz, and M. C. Penna, "SDN architecture for 6LoWPAN wireless sensor networks," *Sensors*, vol. 18, no. 11, p. 3738, 2018.

- [10] G. Z. Papadopoulos, K. Kritsis, A. Gallais, P. Chatzimisios, and T. Noel, "Performance evaluation methods in ad hoc and wireless sensor networks: A literature study," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 122–128, Jan. 2016.
- [11] *W-ILab Testbed*, iGent Building, Brussels, Belgium, 2020. Accessed: Jul. 20, 2020. [Online]. Available: <https://doc.ilabt.imec.be/ilabt/wilab/overview.html>
- [12] A H2020 Project. *Federation for FIRE Plus*. Accessed: Jul. 20, 2020. [Online]. Available: <https://www.fed4fire.eu/>
- [13] M. R. Palattella, N. Accettura, X. Vilajosana, T. Watteyne, L. A. Grieco, G. Boggia, and M. Dohler, "Standardized protocol stack for the Internet of Things," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1389–1406, 3rd Quart., 2013.
- [14] D. K. Sah and T. Amgoth, "Parametric survey on cross-layer designs for wireless sensor networks," *Comput. Sci. Rev.*, vol. 27, pp. 112–134, Feb. 2018.
- [15] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-end arguments in system design," *ACM Trans. Comput. Syst.*, vol. 2, no. 4, pp. 277–288, Nov. 1984.
- [16] H. I. Kobo, A. M. Abu-Mahfouz, and G. P. Hancke, "A survey on software-defined wireless sensor networks: Challenges and design requirements," *IEEE Access*, vol. 5, pp. 1872–1899, 2017.
- [17] M. S. Bernard, T. Pei, and K. Nasser, "QoS strategies for wireless multimedia sensor networks in the context of IoT at the MAC layer, application layer, and cross-layer algorithms," *J. Comput. Netw. Commun.*, vol. 2019, pp. 1–33, Dec. 2019.
- [18] M. Abo-Zahhad, N. Sabor, S. Sasaki, and S. M. Ahmed, "A centralized immune-Voronoi deployment algorithm for coverage maximization and energy conservation in mobile wireless sensor networks," *Inf. Fusion*, vol. 30, pp. 36–51, Jul. 2016.
- [19] T. Lu, G. Liu, W. Li, S. Chang, and W. Guo, "Distributed sampling rate allocation for data quality maximization in rechargeable sensor networks," *J. Neww. Comput. Appl.*, vol. 80, pp. 1–9, Feb. 2017.
- [20] S. B. Shah, Z. Chen, F. Yin, I. U. Khan, and N. Ahmad, "Energy and interoperable aware routing for throughput optimization in clustered IoT-wireless sensor networks," *Future Gener. Comput. Syst.*, vol. 81, pp. 372–381, Apr. 2018.
- [21] X. Lai, X. Ji, X. Zhou, and L. Chen, "Energy efficient link-delay aware routing in wireless sensor networks," *IEEE Sensors J.*, vol. 18, no. 2, pp. 837–848, Jan. 2018.
- [22] Y. Yao, Q. Cao, and A. V. Vasilakos, "EDAL: An energy-efficient, delay-aware, and lifetime-balancing data collection protocol for heterogeneous wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 810–823, Jun. 2015.
- [23] H. Li, C. Wu, D. Yu, Q.-S. Hua, and F. C. M. Lau, "Aggregation latency-energy tradeoff in wireless sensor networks with successive interference cancellation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 11, pp. 2160–2170, Nov. 2013.
- [24] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
- [25] D. De Guglielmo, B. Al Nahas, S. Duquennoy, T. Voigt, and G. Anastasi, "Analysis and experimental evaluation of IEEE 802.15.4e TSCH CSMA-CA algorithm," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1573–1588, Feb. 2017.
- [26] K. Kritsis, G. Z. Papadopoulos, A. Gallais, P. Chatzimisios, and F. Theoleyre, "A tutorial on performance evaluation and validation methodology for low-power and lossy networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1799–1825, 3rd Quart., 2018.
- [27] *FED4FIRE+ Open Calls for innovative Experiments*. Accessed: Apr. 10, 2020. [Online]. Available: <https://www.fed4fire.eu/opencalls/>
- [28] *Data Driven Research Strategy: Insights From US Research Development Professionals*. Accessed: Apr. 10, 2020. [Online]. Available: [https://www.elsevier.com/\\_data/assets/pdf\\_file/0017/53090/ResearchStrategyCycle\\_WPIr\\_1.pdf](https://www.elsevier.com/_data/assets/pdf_file/0017/53090/ResearchStrategyCycle_WPIr_1.pdf)
- [29] J. Jiang, "Unleashing the potential of data-driven networking," in *Proc. Int. Conf. Commun. Syst. Netw.* Springer, 2017, pp. 110–126.
- [30] Y. Simmhan, P. Ravindra, S. Chaturvedi, M. Hegde, and R. Ballamajalu, "Towards a data-driven IoT software architecture for smart city utilities," *Softw., Pract. Exper.*, vol. 48, no. 7, pp. 1390–1416, 2018.
- [31] H. Liu and D. Wang, "Robust state estimation for wireless sensor networks with data-driven communication," *Int. J. Robust Nonlinear Control*, vol. 27, no. 18, pp. 4622–4632, 2017.
- [32] T. Wang, S. Wang, and Z.-H. Zhou, "Machine learning for 5G and beyond: From model-based to data-driven mobile wireless networks," *China Commun.*, vol. 16, no. 1, pp. 165–175, 2019.
- [33] C. Fang, S. Guo, Z. Wang, H. Huang, H. Yao, and Y. Liu, "Data-driven intelligent future network: Architecture, use cases, and challenges," *IEEE Commun. Mag.*, vol. 57, no. 7, pp. 34–40, Jul. 2019.
- [34] CRAWDAD. *A Community Resource for Archiving Wireless Data At Dartmouth*. Accessed: Jul. 20, 2020. [Online]. Available: <http://crawdad.org/>
- [35] S. Cheng, Z. Cai, J. Li, and X. Fang, "Drawing dominant dataset from big sensory data in wireless sensor networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2015, pp. 531–539.
- [36] K. Merima, D. P. Eli, K. Tarik, and M. Ingrid. (Jan. 2017). *MAC-Level Performance Dataset for 802.15.4 WSNs*. [Online]. Available: <https://doi.org/10.5281/zenodo.228613>
- [37] M. Tang, T. Zhang, J. Liu, and J. Chen, "Cloud service QoS prediction via exploiting collaborative filtering and location-based data smoothing," *Concurrency Comput., Pract. Exper.*, vol. 27, no. 18, pp. 5826–5839, 2015.
- [38] F. Songwei, Z. Yan. (2015). *CRAWDAD Dataset Due/Packet-Delivery (2015-04-01) Downloaded*. [Online]. Available: <https://crawdad.org/due/packet-delivery/20150401>
- [39] M. Ateeq, F. Ishmanov, M. Afzal, and M. Naeem, "Multi-parametric analysis of reliability and energy consumption in IoT: A deep learning approach," *Sensors*, vol. 19, no. 2, p. 309, Jan. 2019.
- [40] M. Ateeq, F. Ishmanov, M. K. Afzal, and M. Naeem, "Predicting delay in IoT using deep learning: A multiparametric approach," *IEEE Access*, vol. 7, pp. 62022–62031, 2019.
- [41] M. Ateeq, M. K. Afzal, M. Naeem, M. Shafiq, and J.-G. Choi, "Deep learning-based multiparametric predictions for IoT," *Sustainability*, vol. 12, no. 18, p. 7752, 2020.
- [42] *Contiking Energest Module*. Accessed: Jul. 20, 2020. [Online]. Available: <https://github.com/contiki-ng/contiki-ng/wiki/Documentation:-Energest>



**MUHAMMAD ATEEQ** received the bachelor's degree in computer science from Bahauddin Zakariya University, Multan, in 2005, and the master's and Ph.D. degrees in computer science from COMSATS University Islamabad, Wah Campus, in 2007, and 2021, respectively. He has been working in academia since last 14 years. He is currently an Assistant Professor with the Faculty of Computing, The Islamia University of Bahawalpur. His research interest includes data-driven techniques to improve quality of service in wireless communications. He is also interested in the IoT and computing education. He is a Reviewer for many renowned journals, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE INTERNET OF THINGS, *IEEE Communications Magazine*, and the IEEE WIRELESS COMMUNICATIONS.



**HINA HABIB** received the B.S. degree in computer science from Bahauddin Zakariya University, Pakistan, in 2010, and the M.S. degree in computer science from The Islamia University of Bahawalpur, in 2014. From 2016 to 2019, she worked as a Lecturer with COMSATS University Islamabad, Wah Campus, and is currently associated as a Lecturer with The Government Sadiq College Women University, Bahawalpur. Her research interests include data-driven techniques to solve problems in diverse domains like communication, networking, and education. She is also looking for Ph.D. opportunities.



**MUHAMMAD KHALIL AFZAL** (Senior Member, IEEE) received the B.S. and M.S. degrees in computer science from the COMSATS Institute of Information Technology, Wah Cantonment, Pakistan, in 2004 and 2007, respectively, and the Ph.D. degree from the Department of Information and Communication Engineering, Yeungnam University, South Korea, in 2014. From 2008 to 2009, he worked as a Lecturer with Bahauddin Zakariya University, Multan, Pakistan. From 2009 to 2011,

he was a Lecturer with King Khalid University, Abha, Saudi Arabia. He is currently an Assistant Professor with the Department of Computer Science, COMSATS Institute of Information Technology, Wah Cantonment. His research interests include wireless sensor networks, ad hoc networks, smart cities, and the IoT. He is serving as a Reviewer for IEEE ACCESS, *Computers and Electrical Engineering* (Elsevier), the *Journal of Network and Computer Applications* (Elsevier), *Future Generation Computer Systems*, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and a Guest Editor for *Future Generation Computer Systems* (Elsevier), IEEE ACCESS, and the *Journal of Ambient Intelligence and Humanized Computing* (Springer).



**MUHAMMAD NAEEM** (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the University of Engineering and Technology at Taxila, Taxila, Pakistan, in 2000 and 2005, respectively, and the Ph.D. degree from Simon Fraser University, Burnaby, BC, Canada, in 2011. From 2000 to 2005, he was a Senior Design Engineer with Concept (Pvt.) Ltd., Islamabad, Pakistan, where he participated in the design and development of smart card-based GSM and

CDMA pay phones with the Department of Design. From 2012 to 2013, he was a Postdoctoral Research Associate with the Wireless Networks and Communications Research (WINCORE) Laboratory, Ryerson University,

Toronto, ON, Canada. Since 2013, he has been an Assistant Professor with the Department of Electrical Engineering, COMSATS Institute of Information Technology at Wah Cantonment, Wah Cantonment, Pakistan, and a Research Associate with the WINCORE Laboratory. He is also a Microsoft Certified Solution Developer. His research interests include the optimization of wireless communication systems, nonconvex optimization, resource allocation in cognitive radio networks, and approximation algorithms for mixed integer programming in communication systems. He was a recipient of the NSERC CGS Scholarship.



**SUNG WON KIM** received the B.S. and M.S. degrees from the Department of Control and Instrumentation Engineering, Seoul National University, South Korea, in 1990 and 1992, respectively, and the Ph.D. degree from the School of Electrical Engineering and Computer Sciences, Seoul National University, in August 2002. From January 1992 to August 2001, he was a Researcher at the Research and Development Center, LG Electronics, Korea. From August 2001 to August 2003,

he was a Researcher at the Research and Development Center, AL Tech, Korea. From August 2003 to February 2005, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. In March 2005, he joined the Department of Information and Communication Engineering, Yeungnam University, Gyeongsangbuk-do, South Korea, where he is currently a Professor. His research interests include resource management, wireless networks, mobile computing, performance evaluation, and machine learning.

...