View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Liverpool Repository

Real Smart Home Data-Assisted Statistical Traffic Modelling for the Internet of Things

Chitradeep Majumdar, Miguel López-Benítez, *Senior Member, IEEE,*
and Shabbir N. Merchant, *Member, IEEE*

***Abstract*—The majority of practical studies and analyses in the context of the Internet of Things (IoT) have been carried out assuming that data packet generation follows theoretical models (typically a Poisson process with exponentially distributed packet inter-arrival times) without previous experimental validation and supporting evidence. In contrast to this approach, this work proposes a novel experimental and mathematical framework to determine statistical models for IoT data traffic. Based on empirical data generated by common smart home devices (e.g., ambient temperature, luminous intensity, atmospheric pressure and motion sensors) recorded over a full year using an experimental IoT subsystem, this work first shows that real IoT traffic does not follow the Poisson process model conventionally assumed in the literature, but rather depends on the type of application. Consequently, we estimate the empirical statistical distribution of the inter-arrival between data packets for several smart home applications. The empirical distribution of the packet inter-arrival times is fitted with some well-established classical statistical distributions using the method of moments as well as maximum likelihood estimation techniques, and the goodness of fit is quantified using the Kolmogorov-Smirnov (KS) test. Moreover, we also carry out a regression analysis to provide mathematical relations between the distribution parameters and the considered physical input parameters (ambient temperature, luminous intensity and atmospheric pressure), which is particularly useful in practical scenarios. Furthermore, an exhaustive analysis of the variation of parameters over different time scales and the autocorrelation characteristics of the data packet generation are included as well. In summary, this work provides accurate traffic models suitable for real-life IoT scenarios that can be used for an adequate design and optimisation of future communication networks to efficiently support IoT services.**

***Index Terms*—Internet of Things, device-to-device, traffic modelling, smart homes, Poisson process.**

Manuscript received April 12, 2019; revised October 24, 2019; accepted January 9, 2020. Date of publication Month XX, 2020; date of current version Month XX, 2020. This work was jointly funded by the Royal Society of the United Kingdom and the Science and Engineering Research Board (SERB) of India under a Royal Society-SERB Newton International Fellowship (grant reference no. NF170943). (*Corresponding author: Miguel López-Benítez.*)

C. Majumdar is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, United Kingdom, and also with Samsung R&D Institute India-Bangalore, Karnataka 560037, India (email: chitradeep.majumdar@liverpool.ac.uk).

M. López-Benítez is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, United Kingdom, and also with the ARIES Research Centre, Antonio de Nebrija University, 28040 Madrid, Spain (email: m.lopez-benitez@liverpool.ac.uk).

S. N. Merchant is with the Department of Electrical Engineering, IIT Bombay, Mumbai 400076, India (e-mail: merchant@ee.iitb.ac.in).

Copyright (c) 2020 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Digital Object Identifier 10.1109/JIOT.2020.XXXXXXX

I. Introduction

COMMUNICATION networks at present and in future will not be just about connecting people, but are instead evolving into billions of interconnected smart machine-type devices that enable automatic data collection with minimal or no human intervention. This concept, known as the Internet of Things (IoT) [1], [2], is seen as the next stage of the Information Revolution. The IoT paradigm used in conjunction with social networking concepts treating physical parameters as social objects [3] is another upcoming area that incorporates the physical world with virtual cyber space. In order to realise the vision to make the concept of IoT all-pervasive, researchers have come up with several real-time IoT subsystems which are highly efficient, robust, scalable and reliable. Different designing and implementation challenges of one such efficient IoT prototypes are discussed in [4]. End-user test cases such as smart home applications, industrial equipment monitoring and healthcare are some of the classical examples where the IoT paradigm is used extensively. In [5], the acceptance towards IoT by the end-users for smart home applications is investigated and analysed mathematically based on structural equation modelling. In [6], the authors blended an interesting concept of physical IoT and cognition factor leading to a cognitive dynamic system within smart home IoT space. A smart home test case with a novel concept of mobile edge computing providing autonomous, distributed, low latency and efficient back-end computing infrastructure to support IoT for next generation wireless standards is proposed in [7].

While mobile communication networks are expected to become a key connectivity technology for IoT and will certainly support many of the anticipated new smart IoT services, they have historically been designed to support human-related services and as such are not perfectly suited to support the new machine-based IoT services, which have a different set of features and requirements [8], [9]. The work reported in [10] presents an extensive survey of the issues associated with Machine-to-Machine (M2M) applications in LTE, including the challenges posed by the traffic issues of M2M communications both at the access and data channels and congestion problems at the core network. The integration of IoT machine-type traffic in mobile communication networks along with human-driven traffic may have a significant impact on the overall network performance [11], [12]. As a result, significant thrust has been laid to come up with novel and efficient protocols to integrate the IoT paradigm with current as well as next generation mobile communication standards such as

LTE-A Pro and the future 5G New Radio (NR) [13]–[16].

The next generation of cellular communication networks will depend significantly on IoT data traffic. Therefore, it is essential to accurately and realistically model the data traffic for IoT-based applications. To this end, several models have been proposed in the context of mobile communication networks, typically based on different theoretical modelling approaches such as the Source Semi-Markov Model (SSMM) [17], the Coupled Markov Modulated Poisson Process (CMMPP) [18] and the Coupled Markovian Arrival Process (CMAP) [19] models. All of them make use of Markov chains with varying numbers of states to model the state flow of machine-type devices in a stochastic manner [20]. While aggregated traffic models allow for relatively simpler and computationally efficient models [21], [22], source traffic models allow precise traffic modelling at the expense of higher computational complexity; however the independence of sources from each other results in extra effort to model spatial or temporal correlations, for example by means of spatial Point processes [23] or spatially interacting Discrete-Time Markov Chains (DTMC) [24]. The CMMPP model introduced in [18] allows capturing spatial as well as temporal correlations with Poisson processes as source traffic models and is effectively a compromise between source and aggregated models. Different statistical traffic models generated from single source, aggregated sources and hybrid models are elaborately illustrated in [25]. The 3GPP have also suggested some simple traffic models. Some models are proposed in 3GPP TR 37.868 [26] for both uncoordinated/asynchronous and coordinated/synchronous traffic, where packet arrival times are modelled as uniform and Beta distributions, respectively. Other models are proposed in 3GPP TR 36.888 [27] for two traffic categories, namely *regular reporting* (described in terms of uplink interval time, packet size and mobility features) and *triggered reporting* (where traffic volume sizes are defined and packet inter-arrival times are exponentially distributed) – a combined analytical traffic model that takes into account the different characteristics of both periodic fixed-scheduling and random event-driven traffic is presented in [28]. Traffic models for IoT have also been proposed in the context of other cellular scenarios such as femtocells [29] and other technologies such as satellite communication systems [30] and LoRaWaN [31] (a low power, long range communication protocol developed exclusively to support IoT-based applications).

In terms of statistical traffic modelling, all previous work described above is primarily theoretical in nature. Most existing work assumes some form of (fixed or modulated) Poisson process, which is a well-known model suitable for mathematical analysis but lacks of supporting experimental evidence in the context of IoT. As a matter of fact, the majority of the existing work does not take into account any empirical data for specific applications obtained with experimental IoT platforms for a considerable duration of time. Therefore, to the best of our knowledge there are no or too few works in the available literature that have addressed the problem of traffic modelling in the context of IoT from an empirical and more realistic point of view. In our recent work in [32], [33] we presented few of our preliminary results on IoT data traffic modelling based on 10 weeks of real data. In this context, this work presents a novel statistical modelling approach to realistically model IoT traffic in a more exhaustive way, based on one full year of recorded data, with detailed description of the framework adopted. The main contributions, novelties and distinguishing features of this work are summarised below:

- As opposed to previous work, which is primarily theoretical in nature, we propose a novel methodology for traffic modelling in IoT based on a rigorous analysis of empirical data generated by real IoT sensor devices. The presented models are developed based on empirical data captured using our in-house developed experimental IoT subsystem [4], which implements real sensors typically found in indoor smart home scenarios such as ambient temperature, luminous intensity, atmospheric pressure and motion detection. This approach ensures that the obtained models are realistic and accurate.
- The period over which data are captured is one complete calendar year, from 9am on 25 September 2017 to 9am on 25 September 2018, without any gaps (complete dataset available at [34]). This long measurement period enables not only the development of highly reliable models (as opposed to models based on empirical data sets collected over much shorter time periods) but also the study of seasonal effects on the generation of data traffic related with physical parameters. To the best of our knowledge, this is the first modelling study for IoT traffic based on an extensive measurement campaign of empirical data collection for an interval as long as a complete year.
- Based on the captured empirical data, we show that the widely employed Poisson process model with exponentially distributed packet inter-arrival times is highly inaccurate at the source level and therefore an unrealistic and invalid model for IoT traffic sources in real scenarios.
- Consequently, we determine the statistical traffic models for the packet inter-arrival times in a realistic indoor smart home end-user test case, which is not available in the literature. Exhaustive numerical analysis and analytical tools are used to determine the statistical distributions that could best fit different application types for a single multi-sensor node with a differential data reporting scheme. Moreover, the parameters of the best fitted distributions are characterised as a function of the sensed physical parameters. The developed traffic models can be used to emulate realistic IoT-driven scenarios, which in turn could be used to investigate, analyse and optimise different design-related parameters of emerging mobile communication standards where IoT is an integral part.
- Finally, seasonal effects are investigated by analysing how the traffic patterns associated with physical parameters such as temperature, light and pressure are affected by the different seasons of the year at different time scales.

The rest of this work is organised as follows. First, Section II presents the employed experimental IoT platform and the relevant methodological aspects. Section III then describes the techniques used towards statistical fitting of the packet inter-arrival times, while Section IV analyses the empirical data and

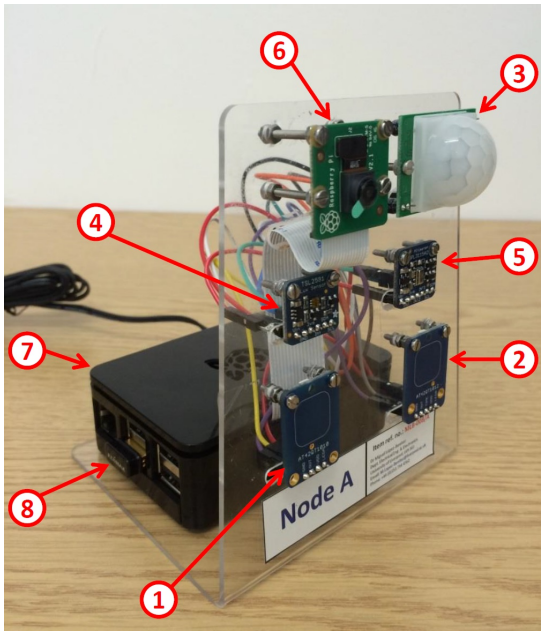


Fig. 1. Experimental IoT subsystem considered in this work [4].

characterises the proposed traffic models in the context of the considered scenario. Section V explores how several features of the empirical data (concretely, correlations and seasonal variations along the year at different time scales) affect the traffic models. Finally, Section VI concludes this work.

II. Experimental Platform and Methodology

A. Experimental Setup

Experimental data were collected for a full year using our in-house developed experimental IoT subsystem composed of several IoT nodes as the one shown in Fig. 1 [4] and a data collection node, powered from an uninterrupted power supply to guarantee continuous operation for one complete year without gaps. Each IoT sensor node comprises six sensor modules typically found in an indoor smart home scenario (elements 1 to 6), which are connected to a Raspberry Pi minicomputer (element 7) in charge of collecting the sensor data and facilitating the connection to the central processing unit through a USB WiFi adapter (element 8). The sensor set includes capacitive touch sensors (elements 1 and 2), a motion sensor (element 3), a light sensor (element 4), a pressure and temperature sensor (element 5), and an image sensor module (element 6). However, in this work we focus on the temperature, light, pressure and motion sensor data, which are typically found in smart home environments. A detailed description of each sensor and the complete IoT platform can be found in [4]. A brief description of the sensors relevant to this work is provided below:

- Passive Infra-Red (PIR) motion sensor (HC-SR501): This sensor provides a logical high output when motion is detected within a range of 7 metres and a 120-degree angle. Once triggered the output remains high for an adjustable interval, which is set to 5 seconds.

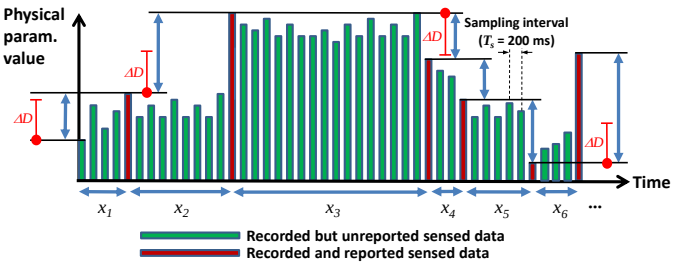


Fig. 2. Differential data reporting strategy.

- Light sensor module based on TSL2591 sensor (Adafruit 1980): This high-range luminosity sensor contains infra-red and full-spectrum diodes that can provide luminance measurements of the infra-red and full-spectrum light. Both values can be used to compute the luminance in the human-visible spectrum and the corresponding luminosity perceived by the human eye (using an empirical formula). In practice, this physical sensor integrates four logical sensors (infra-red, human-visible/full-spectrum luminance and human eye perceived luminosity).
- Pressure, altitude and temperature sensor module based on MPL3115A2 sensor (Adafruit 1893): This module consists of barometric pressure and temperature sensors. Altitude can be obtained from pressure measurements.

The detailed sensor specifications along with the embedded system block diagram and software implementation can be referred from [4].

B. Data Reporting Scheme

An IoT based smart home system is usually designed to sense and record raw data which are usually ambient physical parameters like ambient temperature, luminous intensity, atmospheric pressure, motion detectors, surveillance images, etc. However, more sophisticated IoT systems are envisaged to be integrated with next generation mobile standards that often requires to report the recorded data to a local gateway in real-time for post-processing. This reporting of the data is presumed to be carried out over the wireless medium through various short and/or long range communication protocols which is the main source of IoT traffic over the wireless channels. Therefore, the amount of IoT traffic generated strongly depends on the reporting strategy.

The reporting strategy can be periodic or differential [4]. In periodic reporting, the sensor data is periodically transmitted to the gateway. In differential reporting, as shown in Fig. 2, a data packet transmission is assumed when the absolute difference in the physical parameter for temperature, atmospheric pressure or luminous intensity with respect to the last report (shown as blue arrows) is greater than a predetermined threshold ΔD (shown in red colour) as per the system design requirement. The analysis of the periodic reporting strategy is relatively trivial as compared to differential reporting strategy since the packet inter-arrival time is deterministic. In the differential reporting scheme the inter-arrival time between data packets will be a random variable as it is unknown when ΔD will be satisfied, thus making it non-deterministic. For slowly

Algorithm 1 Estimation of packet inter-arrival times (based on emulation of differential reporting scheme)

Require: data, ΔD
Ensure: $INTARR_{time}$
1: ***Initialise***: $COUNT_{ind} = \emptyset$, $INT_{deg} = \text{data}(1)$, $i = 1$
2: **while** $i < length(\text{data})$ **do**
3: **if** $|INT_{deg} - \text{data}(i+1)| \geq \Delta D$ **then**
4: $COUNT_{ind} = COUNT_{ind} \cup \{i+1\}$
5: $INT_{deg} = \text{data}(i+1)$
6: **end if**
7: $i = i + 1$
8: **end while**
9: **if** $COUNT_{ind} = \emptyset$ **then**
10: $INTARR_{time} = \emptyset$
11: **else**
12: $INTARR_{time} = 200\,\text{ms} \times$
$\times \big(COUNT_{ind}(2:\text{end}) - COUNT_{ind}(1:\text{end}-1)\big)$
13: **end if**

changing physical parameters as the ones considered in this work, the periodic reporting scheme will in general produce a higher volume of data; therefore the differential reporting scheme is a more convenient approach in the considered scenario. Our motivation is to model the IoT traffic for the differential reporting scheme by estimating the distribution of the packet inter-arrival times. Even though under a differential reporting scheme the packet inter-arrival times are related to the variation patterns of the physical parameters, it can be shown that both are governed by different processes and are not related in a straightforward manner, which complicates the analytical treatment of the problem and motivates the empirical modelling approach considered in this work.

C. Estimation of Empirical Inter-Arrival Times

The experimental IoT subsystem was used to periodically collect sensor data every 200 ms for a total interval of one year. Data were transferred and stored at the CPU back-end. A differential reporting mechanism was mimicked based on an offline processing of the recorded data as shown in Algorithm 1, which takes the absolute difference ΔD as an input parameter along with the data and provides a set of packet inter-arrival times (denoted as $INTARR_{time}$) estimated based on the indices of the sensed data. Starting form the initial point INT_{deg}, which is the first value of the sensed parameter, $\text{data}(1)$, the algorithm subsequently computes the index of the next point at which the measured parameter (temperature, atmospheric pressure or luminous intensity) increases or decreases by at least the specified differential threshold ΔD. The indices of the points where such condition is met (i.e., the points where a report, and therefore a packet, would be sent) are added to the set $COUNT_{ind}$. Once the whole data set is processed in the *while* loop, and assuming that some reports have actually been sent (i.e., the set $COUNT_{ind}$ is not empty, denoted by $\emptyset$), then the inter-arrival times can be obtained as the differences between consecutive elements of the set $COUNT_{ind}$ (which represents the number of data samples between differential reports) multiplied by 200 ms (i.e., the time interval between successive data samples).

The proposed algorithm can be used to mimic a differential reporting scheme with an arbitrary threshold ΔD on any data set collected periodically. This approach separates the considered reporting mechanism from the actual physical data collection process and provides a convenient way to simplify both the implementation of the IoT subsystem and the processing of the collected data since a single data set can be exploited to mimic several reporting schemes and/or configurations (periodic reporting with different periods as well as differential reporting with different thresholds).

III. Modelling of the Inter-Arrival Times

As discussed in the previous section, the packet inter-arrival time is a critical factor towards the characterisation of the traffic patterns in the IoT network. The frequency of the data packet generation at individual sensor nodes strongly contributes towards the resulting overall traffic generation supported by the network. Therefore, an adequate modelling of the packet inter-arrival times at the sensor node level is essential for a complete understanding of the traffic volumes that future networks will need to support and their statistical patterns. Most existing work assumes some form of (fixed or modulated) Poisson process. The theoretical basis for this model is the Palm-Khintchine theorem [35], which demonstrates that the superposition of several independent processes converges to a Poisson process as the number of superimposed processes tends to infinity, irrespective of the statistics of the individual component processes. The Poisson process, which implies exponentially distributed inter-arrival times, is a suitable model for aggregated IoT traffic (e.g., in the main arteries of communication networks) and has been widely used by researchers, engineers and standardisation bodies such as the 3GPP [36], [37]. However, its popularity has sometimes led to questionable applications of the model, including the characterisation of the traffic generated by individual sources or a reduced number thereof. This work demonstrates the inaccuracy of this modelling approach and proposes more accurate models by fitting the distribution of packet inter-arrival times observed in the empirical data to several distribution models.

The packet inter-arrival times obtained from the empirical data as detailed in Section II-C (which can be thought of as samples of a random process X since the parameter fluctuations are non-deterministic) are processed to calculate their corresponding Empirical Cumulative Distribution Function (ECDF), $F_X^{ECDF}(x)$, which is the empirical probability that the observed packet inter-arrival time is no greater than the specified value, i.e., $F_X^{ECDF}(x) = Pr(X \leq x)$.

Upon estimation of the ECDF of the empirical inter-arrival times, the next step is to determine the distribution that would provide the best fit to the ECDF. Seven classical distribution models are considered, namely exponential (E), generalised exponential (GE), Pareto (P), generalised Pareto (GP), log-normal (LN), gamma (G) and Weibull (WB). The mathematical expression for the CDF of these models, parameter constraints as well as first and second order moments are

TABLE I. Considered distributions: exponential (E), generalised exponential (GE), Pareto (P), generalised Pareto (GP), log-normal (LN), gamma (G) and Weibull (WB). Parameters: μ (location), λ (scale) and α (shape). x represents the packet inter-arrival time. $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ are the mean and variance of the distribution, respectively. $\psi(\cdot)$ is the digamma function [39, eq. (6.3.1)], $\psi'(\cdot)$ is its derivative. $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function [39, eq. (6.5.2)], $\Gamma(\cdot)$ is the (complete) gamma function [39, eq. (6.1.1)]. Reproduced from [40].

Distribution function	Parameters	Moments
$F_E(x;\mu,\lambda) = 1 - e^{-\lambda(x-\mu)}$	$x \geq \mu > 0$ $\lambda > 0$	$\mathbb{E}\{x\} = \mu + \frac{1}{\lambda}$ $\mathbb{V}\{x\} = \frac{1}{\lambda^2}$
$F_{GE}(x;\mu,\lambda,\alpha) = \left[1 - e^{-\lambda(x-\mu)}\right]^{\alpha}$	$x \geq \mu > 0$ $\lambda > 0$ $\alpha > 0$	$\mathbb{E}\{x\} = \mu + \frac{\psi(\alpha+1)-\psi(1)}{\lambda}$ $\mathbb{V}\{x\} = \frac{\psi'(1)-\psi'(\alpha+1)}{\lambda^2}$
$F_P(x;\lambda,\alpha) = 1 - \left(\frac{\lambda}{x}\right)^{\alpha}$	$x \geq \lambda$ $\lambda > 0$ $\alpha > 2$	$\mathbb{E}\{x\} = \frac{\alpha\lambda}{\alpha-1}$ $\mathbb{V}\{x\} = \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)}$
$F_{GP}(x;\mu,\lambda,\alpha) = 1 - \left[1 + \frac{\alpha(x-\mu)}{\lambda}\right]^{-1/\alpha}$	$x \geq \mu$ $(\alpha \geq 0)$ $x \in [\mu, \mu - \frac{\lambda}{\alpha}]$ $(\alpha < 0)$ $\mu, \lambda > 0,\ \alpha < 1/2$	$\mathbb{E}\{x\} = \mu + \frac{\lambda}{1-\alpha}$ $\mathbb{V}\{x\} = \frac{\lambda^2}{(1-\alpha)^2(1-2\alpha)}$
$F_{LN}(x;\mu,\lambda) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\ln x - \mu}{\sqrt{2\lambda^2}}\right)\right]$	$x \geq 0$ $\mu \in \mathbb{R}$ $\lambda > 0$	$\mathbb{E}\{x\} = e^{\mu+\lambda^2/2}$ $\mathbb{V}\{x\} = \left(e^{\lambda^2} - 1\right) e^{2\mu+\lambda^2}$
$F_G(x;\mu,\lambda,\alpha) = \frac{\gamma\left(\alpha, \frac{x-\mu}{\lambda}\right)}{\Gamma(\alpha)}$	$x \geq \mu > 0$ $\lambda > 0$ $\alpha > 0$	$\mathbb{E}\{x\} = \mu + \lambda\alpha$ $\mathbb{V}\{x\} = \lambda^2\alpha$
$F_{WB}(x;\mu,\lambda,\alpha) = 1 - \exp\left[-\left(\frac{x-\mu}{\lambda}\right)^{\alpha}\right]$	$x \geq \mu > 0$ $\lambda > 0$ $\alpha > 0$	$\mathbb{E}\{x\} = \mu + \lambda\Gamma\left(1 + \frac{1}{\alpha}\right)$ $\mathbb{V}\{x\} = \lambda^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$

provided in Table I. Numerical fitting of these models to empirical data is based on the Method-of-Moments (MoM) and Maximum Likelihood (ML) [38]. The location parameter (μ) is always estimated as the minimum inter-arrival time observed in the empirical data, while the scale (λ) and shape (α) parameters are estimated based on the MoM and ML methods. In the MoM technique, the expressions for the population mean and population variance shown in Table I are equated to the sample mean and sample variance of the empirical inter-arrival times, respectively, and then solved for the desired distribution parameters. In the ML estimation method, the log-likelihood function is first obtained as:

$$\begin{aligned}\mathcal{L}(\mathbf{X};\lambda,\alpha) &= \log\prod_{n=1}^{N} f_X\left(x_n;\lambda,\alpha\right) \\ &= \sum_{n=1}^{N}\log f_X(x_n;\lambda,\alpha) \end{aligned} \tag{1}$$

where $f_X(x_n;\lambda,\alpha)$ is the Probability Density Function (PDF) of the considered distribution (which can be obtained by differentiating the CDFs shown in Table I) and $\mathbf{X} = \{x_n\}_{n=1}^{N}$ is the set of N empirically-observed packet inter-arrival times, x_n, estimated as discussed in Section II-C and illustrated in Fig. 2. The distribution parameters are then estimated as the values that maximise the log-likelihood function in (1). In the case of the exponential and generalised exponential models, (1) leads to analytically tractable expressions and, as a result, estimates of the scale (λ) and shape (α) parameters are obtained by solving $\partial\mathcal{L}(\mathbf{X};\lambda,\alpha)/\partial\lambda$ for λ and $\partial\mathcal{L}(\mathbf{X};\lambda,\alpha)/\partial\alpha$ for α, respectively. For the rest of considered distribution models, the maximisation of the expression obtained from (1) is non-trivial from an analytical point of view but can be accomplished with numerical optimisation techniques such as gradient descent or interior region methods. This provides an estimation of the desired scale and shape parameters for the fitted distributions.

It is worth noting that in some cases none of the candidate distributions shown in Table I may provide a satisfactory fit to the ECDF. While new candidate distributions could be introduced in the set shown in Table I, this would increase the number of distribution models to be evaluated and therefore complicate the analysis of the empirical data. An alternative is to consider a weighted sum of distributions as follows:

$$\begin{aligned}F_X^{fitted}(x) = \gamma F_X(x;\mu_1,\lambda_1,\alpha_1) \\ + (1-\gamma)\,F_X(x;\mu_2,\lambda_2,\alpha_2)\end{aligned} \tag{2}$$

where two distributions of the same type but with different sets of parameter values are weighted by the weighting parameter $0 < \gamma < 1$. As shown in our previous work reported in [33], this approach can provide significantly more accurate fits to empirical data in those cases where a single distribution function fails. Similar to the single distribution case, the parameters of the model in (2) can be estimated based on the MoM approach by equating the population moments of the model in (2) to the sample moments of the empirical inter-arrival times in order to produce a system with as many equations as unknown parameters; this approach, however, will not be used in this work as it will be discussed later on. Alternatively, the parameters of the model in (2) can also be estimated based on the ML approach as shown in (1) after replacing the PDF with the equivalent weighted version resulting from (2); a detailed description of this approach for the relevant cases of interest will provided where required (please refer to the Appendix for more details).

The fitting accuracy of the models in Table I (or a weighted version thereof) to the empirical data is quantified in terms of the Kolmogorov-Smirnov (KS) distance, which is defined as the maximum absolute difference between the empirical dis-

tribution, $F_X^{ECDF}(x)$, and the fitted distribution, $F_X^{fitted}(x)$:

$$D_{KS} = \sup_x \left| F_X^{ECDF}(x) - F_X^{fitted}(x) \right| \tag{3}$$

Since different parameter estimation methods may provide different levels of accuracy for different distributions, the models shown in Table I are fitted to the empirical data using both fitting methods (MoM and ML) and the distribution that provides the lowest KS distance (obtained with either fitting method, MoM or ML) is selected as the best model. In many practical scenarios, values of $D_{KS} \approx 0.05$ are typically considered as a reasonable/good fitting accuracy [41].

The steps described so far in this section can be followed to identify a distribution that provides a good fit for the inter-arrival times of the packets that would be generated when each of the considered physical parameters (ambient temperature, luminous intensity and atmospheric pressure) are monitored with a differential reporting method based on a certain variation threshold, ΔD. Moreover, the process would also provide a set of sample values for the parameters of the distribution (μ, λ, α), which would be useful in practical applications of the models such as simulations or numerical evaluation of analytical results obtained from these models. However, from a practical point of view, it would be convenient to express the distribution parameters (μ, λ, α) as a function of the variation threshold considered by the periodic reporting mechanism (ΔD), which reduces the total number of variables involved in the model to just one (i.e., ΔD), which in turn has a clear relation with the monitored physical parameters.

This observation motivates the introduction of a second level of modelling aimed at describing the relation between the selected variation threshold ΔD and the corresponding values of μ, λ and/or α in the resulting best-fitting distribution. To this end, the fit provided by each candidate distribution was evaluated for various values of ΔD within a predefined interval and the distribution providing the lowest average KS distance across such interval was selected as the best fitting distribution for the physical parameter under study (in cases with very similar average KS distances, the decision was made based on the lowest variance of the KS distance). The values of μ, λ and/or α that provided the best fit for each evaluated value of ΔD were then used to find an appropriate model for the distribution parameters (μ, λ and/or α) as a function of ΔD. A broad range of mathematically tractable functions were explored to this end, including linear, polynomial, exponential, Gaussian, sum of sines, Fourier and power series, and in all cases it was found out that the most accurate fit was obtained with the following generalised Gaussian model:

$$\mu\left(\Delta D\right) = \sum_{k=1}^{K} a_k \, e^{-\left(\frac{\Delta D - b_k}{c_k}\right)^2} \tag{4}$$

$$\lambda\left(\Delta D\right) = \sum_{k=1}^{K} a_k \, e^{-\left(\frac{\Delta D - b_k}{c_k}\right)^2} \tag{5}$$

$$\alpha\left(\Delta D\right) = \sum_{k=1}^{K} a_k \, e^{-\left(\frac{\Delta D - b_k}{c_k}\right)^2} \tag{6}$$

where $\{a_k\}_{k=1}^K$, $\{b_k\}_{k=1}^K$ and $\{c_k\}_{k=1}^K$ are fitting coefficients[1].

Note that with the proposed modelling approach, the end-user of the model only needs to select a valid value of the variation threshold parameter ΔD used by the differential reporting mechanism, which represents the relative difference in the observed physical magnitude (i.e., ambient temperature, luminous intensity or atmospheric pressure) between two consecutive data packets. Based on the selected ΔD, (4)–(6) are then used to determine the value of the parameters that need to be introduced in the best-fitting distribution from Table I. The introduction of the second level of modelling shown in (4)–(6) greatly simplifies the use of the models as the only variable to be selected by the end-user is the differential increment ΔD in the monitored physical parameter.

IV. Analysis and Modelling of the Empirical Data

The characterisation of the modelling approach proposed in Section III requires the determination of two aspects, first the distribution from Table I that provides the best fit to the observed packet inter-arrival times, then the values of the fitting coefficients for the expressions in (4)–(6), which are both obtained based on the captured empirical data for each physical magnitude as discussed in this section. For the first aspect of the proposed modelling framework, the accuracy of the distributions shown in Table I was assessed based on the captured empirical data. The obtained results indicated that the best-fitting distribution depends on the considered physical magnitude (i.e., ambient temperature, luminous intensity, atmospheric pressure and motion detection). For this aspect, detailed numerical results are presented, showing the accuracy attained by each distribution in Table I and discussing how the best-fitting model was selected in each case. For the second aspect, the accuracy of several models was investigated and, as pointed out in Section III, in all cases the generalised Gaussian model presented in (4)–(6) provided the best fit. Consequently, in order to avoid an excessive amount of results in this section, a comparison of the accuracy of the model presented in (4)–(6) with other considered models (linear, polynomial, exponential, Gaussian, sum of sines, Fourier and power series) is omitted and only the values for the fitting coefficients of (4)–(6) will be provided in order to enable the use of the proposed models by other researchers. Nevertheless, the validity and accuracy of this model will be evident from the presented results. In the remainder of this work, the general notation for the threshold parameter ΔD will be replaced, where appropriate, with the equivalent notation for the physical magnitude considered in each case, namely ambient temperature (ΔT), luminous intensity (ΔL) and atmospheric pressure (ΔP).

A. Ambient Temperature

The distribution models in Table I were fitted (using both estimation methods, i.e., MoM and ML) to the empirically observed inter-arrival times of the data packets generated when the ambient temperature is reported for differential

[1]Coefficients are in general different for each distribution parameter (μ, λ, α), however the same notation (a_k, b_k, c_k) is used in all cases for simplicity.

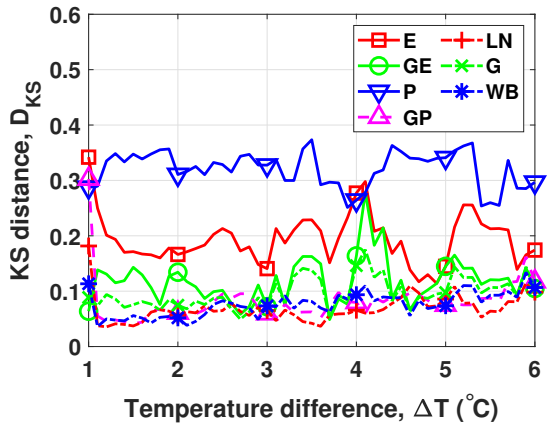


Fig. 3. Accuracy of the considered distributions in fitting the inter-arrival time of temperature-related data packets as a function of the considered temperature difference ΔT.

TABLE II. Mean and variance of the KS distances shown in Fig. 3 for temperature data across the interval $\Delta T \in [1°\text{C}, 6°\text{C}]$.

Distribution	Mean of D_{KS}	Var. of D_{KS}
E	0.1937	0.0020
GE	0.1186	0.0017
P	0.3209	0.0010
GP	0.0808	0.0015
LN	**0.0681**	**0.0006**
G	0.0966	0.0008
WB	**0.0758**	**0.0005**

variations ΔT ranging from 1 to 6 °C (degrees Celsius) in increments of 0.1 °C. This range of values here considered for the ΔT parameter is consistent with the maximum range of variation that was observed in the empirical temperature data that was recorded in the considered smart home (indoor) scenario. The KS distance obtained for the best fit of each distribution (provided by either MoM or ML) as a function of the considered ΔT is shown in Fig. 3. The mean and variance of the KS distances shown in Fig. 3 are provided in Table II.

As it can be appreciated in the obtained results, different distributions provide different fitting accuracies to the empirical data. It is worth noting that the exponential model provides in general a poor fit to the observed inter-arrival times, showing not only a high mean value of the KS distance across the considered interval (0.1937), thus indicating a poor fit, but also a high variance, which indicates high fluctuations in the fitting accuracy for different values of ΔT. This result indicates that the Poisson model, which assumes exponentially distributed inter-arrival times, is highly inaccurate as a source traffic model in the case of a temperature sensor with differential reporting. On the other hand, other models can provide a significantly better accuracy. While Fig. 3 shows that there is no single distribution that provides the best fit across the whole range of ΔT, Table II shows that the log-normal and Weibull models (shown in boldface in Table II) provide, in average, the best fit, with a KS distance close to the target accuracy of 0.05 for all the considered ΔT values. In both cases, the variance is also quite low, meaning that the observed fitting accuracy is preserved across different values of the ΔT parameter. While both distributions can be equally considered as suitable models, the log-normal distribution (underlined in Table II) provides a slightly better fit to the ECDF and therefore is selected as the best-fitting model in this case.

TABLE III. Coefficients of the model in (4)–(5) for the parameters of the log-normal fit to temperature data.

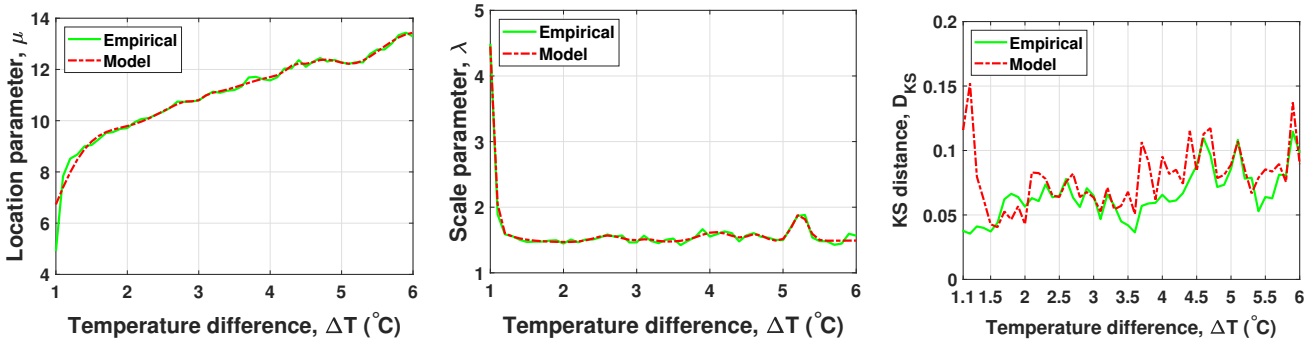
	Location, μ			Scale, λ		
k	a_k	b_k	c_k	a_k	b_k	c_k
1	13.01	6.19	1.597	2.829	0.9817	0.08471
2	0.6961	4.751	0.3582	0.4132	5.232	0.146
3	9.335	3.884	1.28	0.1441	4.075	0.3095
4	0.281	4.339	0.1527	0.1062	4.631	0.1564
5	5.128	2.638	0.8697	1.654	55.72	153.2
6	0.01908	3.182	0.1052	0.0989	2.606	0.2737
7	–0.1552	2.968	0.1084	0.03817	3.146	0.155
8	8.002	1.438	0.9688	$2.885 \cdot 10^{10}$	-13.83	2.941

Figs. 4(a)–(b) show the parameters of the log-normal distribution for temperature data as a function of the temperature difference ΔT. The figure shows the values obtained from the direct fit of the distribution to the empirical data ("Empirical") along with the corresponding fit obtained with (4)–(5) ("Model"), whose coefficients were calculated by regression with the Matlab curve-fitting toolbox and are shown in Table III. As it can be observed, the model in (4)–(5) with $K = 8$ fitting coefficients provides an excellent characterisation of the relation between the parameters of the log-normal distribution and the differential reporting threshold ΔT for temperature data. If the distribution parameters obtained by fitting the log-normal model to the temperature data are replaced with their equivalent values from (4)–(5), the resulting fitting accuracy is not affected in a noticeable manner as shown in Fig. 4(c), which corroborates the validity of the proposed modelling framework for temperature sensors with differential reporting.

B. Luminous Intensity

The accuracy of the distribution models in Table I in fitting the luminosity data is illustrated in Fig. 5 in terms of the KS distance as a function of the differential reporting threshold for the luminous intensity ΔL, which was increased from 10 to 200 lux in steps of 1 lux. This range of ΔL values is consistent with the maximum range of variation that was observed in the empirical luminosity data recorded in the considered smart home (indoor) scenario. The mean and variance of the KS distances shown in Fig. 5 are provided in Table IV.

Similar observations as for the temperature data can be made here. The exponential distribution also provides a poor fit in this case, showing an inconsistent fitting accuracy across different values of ΔL. Therefore, the Poisson model, which assumes exponentially distributed inter-arrival times, is also an inappropriate source traffic model in the case of a luminosity sensor with differential reporting. Moreover, the log-normal and Weibull models (shown in boldface in Table IV) also provide overall the most accurate fit (lowest mean and variance of the observed KS distances) across the whole range of ΔL values. Again, while both models are equally suitable, the log-normal distribution provides a slightly better fit and is therefore selected as the best-fitting model (underlined in Table IV).



(a) Location parameter, μ. (b) Scale parameter, λ. (c) KS distance, D_{KS}.

Fig. 4. (a)–(b) Value of the parameters of the log-normal fit to temperature data as a function of the temperature difference ΔT; (c) KS distance of the log-normal fit to temperature data as a function of the temperature difference ΔT.

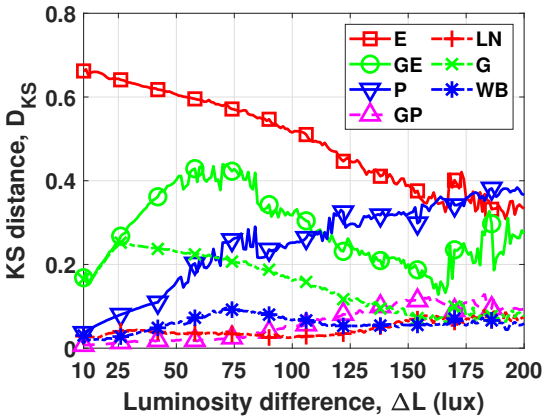


Fig. 5. Accuracy of the considered distributions in fitting the inter-arrival time of luminosity-related data packets as a function of the considered luminosity difference ΔL.

TABLE IV. Mean and variance of the KS distances shown in Fig. 5 for luminosity data across the interval $\Delta L \in [10 \text{ lux}, 200 \text{ lux}]$.

Distribution	**Mean of D_{KS}**	**Var. of D_{KS}**
E	0.4956	0.0121
GE	0.2841	0.0069
P	0.2492	0.0100
GP	0.0596	0.0015
LN	**0.0453**	**0.0003**
G	0.1537	0.0040
WB	**0.0591**	**0.0003**

Fig. 6 shows the parameters of the log-normal fit to luminosity data as a function of the luminosity difference parameter ΔL along with their counterparts obtained with the models in (4)–(5) based on the coefficients (obtained by regression) shown in Table V. As observed, there is an excellent agreement between the accuracy obtained when the models in (4)–(5) are used and that obtained from the direct fit of the log-normal distribution to the empirical data, which again corroborates the validity of the proposed modelling framework, in this case for luminosity sensors with differential data reporting.

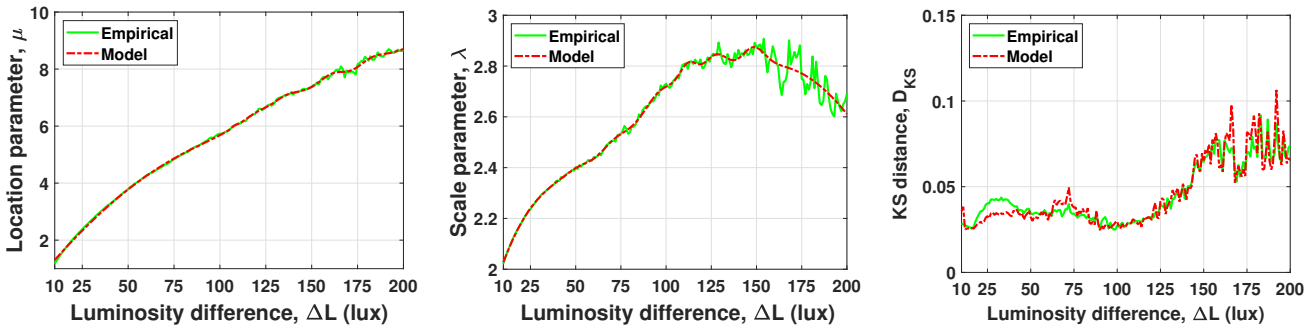
TABLE V. Coefficients of the model in (4)–(5) for the parameters of the log-normal fit to luminosity data.

	Location, μ			**Scale, λ**		
k	a_k	b_k	c_k	a_k	b_k	c_k
1	0.6164	202.6	8.931	0.08232	147.9	8.903
2	0.4146	178	4.014	0.1387	126.4	10.68
3	0	330.6	19.62	2.801	158.8	155.5
4	4.086	89.28	58.76	0.1088	112.4	6.918
5	–0.5952	154.5	15.31	0.1972	97.73	15.51
6	1.196	37.15	36.57	0.151	68.85	19.33
7	–0.8704	175.5	9.707	–0.04312	61.97	9.236
8	8.442	183.5	66.08	0.9236	17.08	53.33

C. Atmospheric Pressure

Fig. 7 shows the accuracy of the distribution models in Table I in fitting the pressure data in terms of the KS distance as a function of the differential threshold for the atmospheric pressure ΔP, which was increased from 100 to 600 Pa (Pascals) in steps of 5 Pa. This range of ΔP values is consistent with the maximum range of variation that was observed in the empirical pressure data recorded in the considered smart home (indoor) scenario. The mean and variance of the KS distances shown in Fig. 7 across this interval are provided in Table VI.

As it was observed for the temperature and luminosity data, the exponential distribution also fails to provide a reasonable fit to the empirical pressure data, which confirms once again, also in the case of pressure sensors with differential reporting, that the Possion model cannot be considered as an adequate source traffic model. The best fits in this case are provided by the gamma and the generalised exponential distributions (shown in boldface in Table VI), which provide the lowest values for the average KS distance and its variance across the considered ΔP range, thus offering the best overall fit to the empirical pressure data out of all the considered candidate distributions. However, as it can be observed in Table VI, the average KS distance for these two distributions, despite being the lowest one out of all the considered models, is greater than the reference threshold $D_{KS} \approx 0.05$ typically considered in practical scenarios as an indication of an acceptable fitting accuracy. In fact, the results shown in Fig. 7 indicate that the KS distance for these two distributions (and indeed for all the



(a) Location parameter, μ. (b) Scale parameter, λ. (c) KS distance, D_{KS}.

Fig. 6. (a)–(b) Value of the parameters of the log-normal fit to luminosity data as a function of the luminosity difference ΔL; (c) KS distance of the log-normal fit to luminosity data as a function of the luminosity difference ΔL.

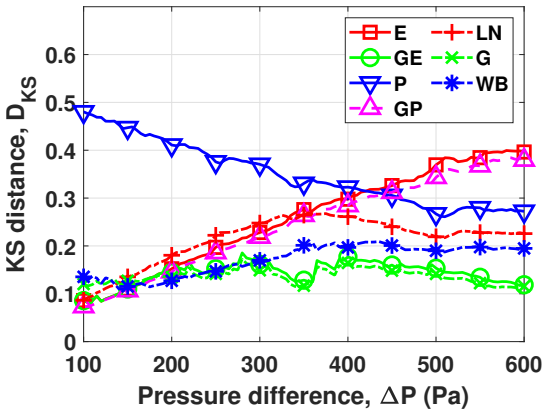


Fig. 7. Accuracy of the considered distributions in fitting the inter-arrival time of pressure-related data packets as a function of the considered pressure difference ΔP.

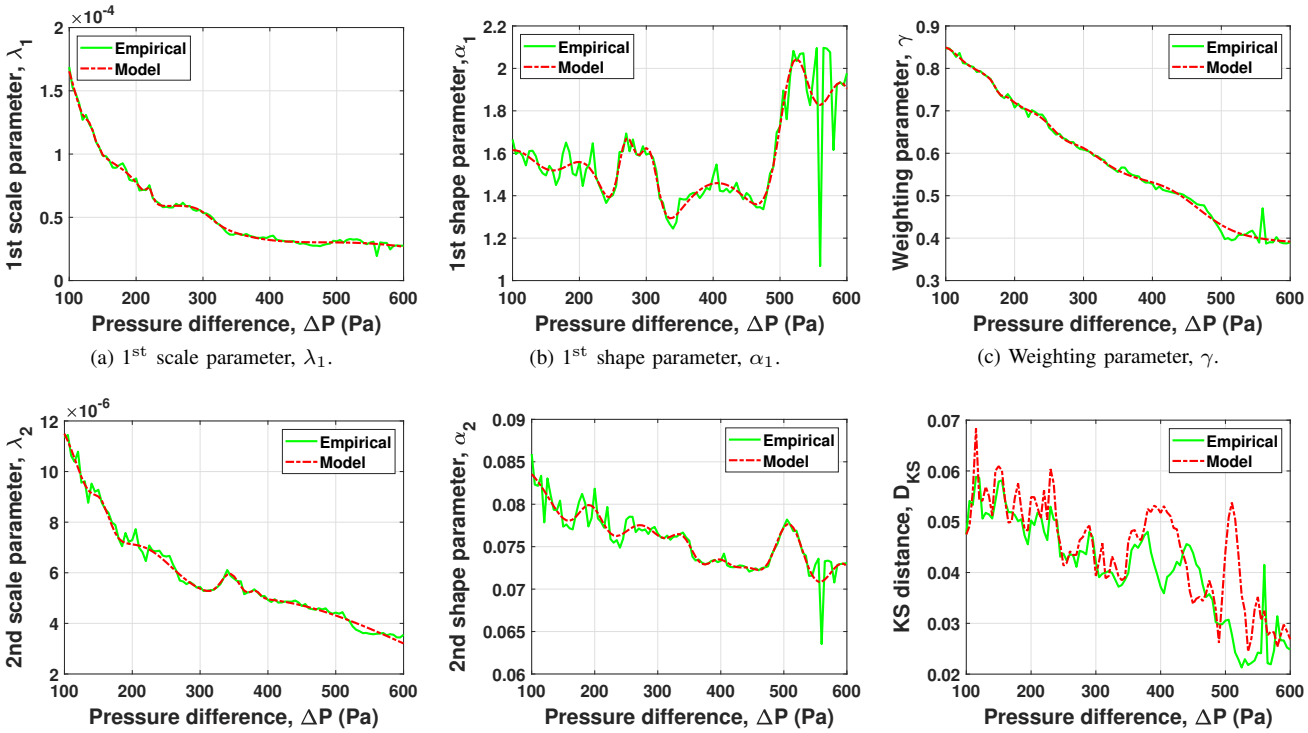
TABLE VI. Mean and variance of the KS distances shown in Fig. 7 for pressure data across the interval $\Delta P \in [100\text{ Pa}, 600\text{ Pa}]$.

Distribution	Mean of D_{KS}	Var. of D_{KS}
E	0.2418	0.0119
GE	**0.1369**	**0.0008**
P	0.3598	0.0049
GP	0.2302	0.0108
LN	0.2043	0.0032
G	**0.1323**	**0.0004**
WB	0.1654	0.0013

considered distributions) is greater than 0.10 for all the considered ΔP values. Consequently, none of these distributions can be considered as a good fit to the empirical pressure data.

To overcome this problem, a weighted sum of distributions as shown in (2) can be considered. The distributions resulting from the application of the weighting in (2) have in general more complex algebraic forms, which complicates the task of estimating the distribution parameters based on the MoM and ML methods. To facilitate the analysis, the weighting in (2) is initially considered only for the two distributions that provide the best accuracy in Table VI, namely the gamma and generalised exponential distributions. As observed in Table I, the mean and variance of the gamma distribution have simple algebraic forms and therefore the application of the MoM technique in this case would be straightforward; however, the expression for its CDF involves several types of gamma functions, which would complicate the application of the ML method. On the other hand, the mean and variance of the generalised exponential distribution involve the digamma function and its derivative, which would make it difficult the application of the MoM method; however, the relatively simple expression of its CDF would enable the analytical calculation of the expressions required for the application of the ML method. The ML method is known to provide, in general, more accurate fits than MoM, which indeed was observed in the analysis of the empirical data obtained in this work. This motivates the consideration of a model based on a weighted sum of generalised exponential distributions and the estimation of its parameters based on the ML method. While the literature on MoM- and ML-based estimation techniques for simple distributions such as those shown in Table I is abundant, the estimation of the specific form of weighted sum of generalised exponential distributions considered in this section, to the best of the authors' knowledge, has not been investigated before. The analytical expressions required for the ML estimation of such distribution are presented in the Appendix.

As shown in Table I, the generalised exponential distribution has three parameters, therefore a weighted sum of two such distributions has (including the weighting parameter) a total of seven parameters (γ, μ_1, λ_1, α_1, μ_2, λ_2, α_2), which need to be estimated based on the empirical data. In the particular case considered in this section, the number of parameters can be reduced to five by setting $\mu_1 = \mu_2 = \mu$. This simplification is possible due to the fact that the location parameter μ, which is the minimum value of the packet inter-arrival time, was observed to be the same for all the values of ΔP and equal to the data sampling interval (i.e., $\mu = 200$ ms). The analysis of the empirical pressure data showed that this was caused by the presence of abrupt variations of around 1000 Pa on the instantaneously reported pressure, which affected only some particular time instants. Further investigation demonstrated that this was caused by the interaction of the pressure sensor with elements of the indoor environment. Wind gusts caused by open doors/windows and people moving in the close vicinity of the sensor (which in the experimental location considered in this work could be as close as a few tens



(a) 1^{st} scale parameter, λ_1. (b) 1^{st} shape parameter, α_1. (c) Weighting parameter, γ.

(d) 2^{nd} scale parameter, λ_2. (e) 2^{nd} shape parameter, α_2. (f) KS distance, D_{KS}.

Fig. 8. (a)–(e) Value of the parameters of the weighted sum of generalised exponential distributions for pressure data as a function of the pressure difference ΔP; (c) KS distance of the weighted sum of generalised exponential distributions for pressure data as a function of ΔP.

TABLE VII. Coefficients of (4)–(6) for the parameters of the weighted sum of generalised exponential distributions fit to pressure data. Note: Once λ_1 and λ_2 are obtained using these fitting coefficients in (5), the obtained values need to be multiplied by 10^{-5} and 10^{-6}, respectively.

	Weighting parameter, γ			**1^{st} scale, λ_1**			**1^{st} shape, α_1**			**2^{nd} scale, λ_2**			**2^{nd} shape, α_2**		
k	a_k	b_k	c_k	a_k	b_k	c_k	a_k	b_k	c_k	a_k	b_k	c_k	a_k	b_k	c_k
1	0.04628	106.3	56.52	15.99	13.07	112.5	0.6955	570.5	6.167	7.667	86.01	67.38	0.0853	60.67	263.5
2	0.02028	159.2	14.75	1.308	157.1	6.564	1.348	535.9	45.64	0.8702	147	9.03	0.006081	194	29.56
3	0.01345	199.2	20.7	–1.335	157.8	9.694	0.4673	270.5	17.58	1.542	200.2	14.72	0.02721	506.9	40.45
4	0.02527	236.1	13.04	1.223	185	23.04	1.426	596.7	20.48	1.743	168.2	20.11	0.03634	300.3	85.42
5	0.004197	265.3	2.142	1.011	219.5	10.57	1.591	91.4	132.6	1.744	237.2	31.99	0.01775	350.5	36
6	0.001769	321.1	2.946	1.131	279.5	33.16	0.5496	217.2	55.28	0.718	343.9	16.18	0.01391	391.7	29.6
7	0.009741	359.7	4.661	$7.766{\cdot}10^{11}$	$-2.12{\cdot}10^{4}$	4226	1.456	404.4	153.1	0.1724	289.9	9.102	0.07133	596.7	106.5
8	1.758	–1127	1386	0.002402	248.9	3.528	0.4941	302	21.02	5.278	311.1	414.2	0.04843	438.3	55.51

of centimetres) can all cause momentary large variations on the instantaneous reported pressure (around 1000 Pa between consecutive pressure samples in our experiments). The consequence of these events on the empirical data is the presence of two consecutive data samples with a pressure difference of around 1000 Pa, which is greater than all the considered ΔP values and explains the observation of a fixed minimum packet inter-arrival time of $\mu = 200$ ms for all ΔP values. Notice that this short minimum packet inter-arrival time is due to the occasional events mentioned above, which are infrequent, and therefore the generation of pressure-related data packets typically occurs at a much slower rate (around one packet every few tens of seconds), following the actual variations of the real atmospheric pressure. This also explains the inability of the distributions shown in Table I to provide an accurate fit to the empirical pressure data since the packets generated by these occasional events create outliers (i.e., a few inter-arrival times that are significantly shorter than the rest of observed inter-arrival times). As the purpose of this work is to model the actual data traffic that would be generated by sensors deployed in realistic operational environments (rather than the true values of the considered physical parameters), these short inter-arrival times were left and the raw empirical data were modelled without alterations (i.e., removal of outliers).

Fig. 8 shows the parameters of the weighted sum of generalised exponential distributions estimated with the ML method as detailed in the Appendix along with the counterparts based on (4)–(6) and the fitting coefficients (obtained by regression) shown in Table VII (the same model was also applied to the weighting parameter, γ). As it can be appreciated, there is an excellent agreement in all cases and, more importantly, the overall fitting accuracy provided by this model is significantly better than that attained by any of the models shown in Table I (the final KS distance is $D_{KS} \approx 0.05$ or lower across the

TABLE VIII. Accuracy (KS distance) of the considered distributions in fitting relevant aspects of the motion detection data traffic.

(a) Interval between two ON states.

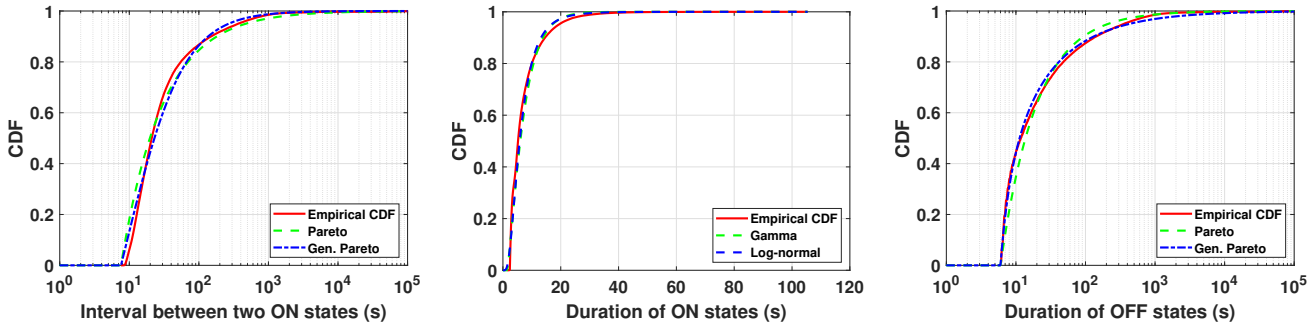
Distribution	MoM	ML
E	0.7071	0.6532
GE	0.9811	0.0955
P	0.2128	**0.0168**
GP	0.6255	**0.0154**
LN	0.3811	0.0337
G	0.9817	0.3095
WB	0.6405	0.1209

(b) Duration of ON states.

Distribution	MoM	ML
E	0.1534	0.1525
GE	0.1682	0.1672
P	0.4844	0.3128
GP	0.1577	0.1554
LN	0.1193	**0.0937**
G	0.1671	**0.0865**
WB	0.1638	0.2234

(c) Duration of OFF states.

Distribution	MoM	ML
E	0.7156	0.6517
GE	0.9678	0.0919
P	0.0868	**0.0294**
GP	0.6440	**0.0168**
LN	0.3266	0.0354
G	0.9689	0.2532
WB	0.5134	0.0816



(a) (b) (c)

Fig. 9. Empirical and fitted CDF for: (a) interval between two successive ON states, (b) duration of ON states, (c) duration of OFF states.

TABLE IX. Parameters of best-fitting distributions for motion traffic.

Time duration	Best fit	μ	λ	α
Between two ON	GP	7.60	15.76	0.93
ON state	G	1.40	4.11	1.43
OFF state	GP	5.99	3.90	1.76

considered ΔP range). The improved fitting accuracy of the proposed weighted model can be ascribed to the flexibility offered by the two weighted distributions: one distribution can be fitted to reproduce the packet inter-arrival times resulting from the actual variations of the real atmospheric pressure (which are in the order of few tens of seconds), while the other distribution can be fitted to reproduce the packet inter-arrival times caused by the infrequent events mentioned above (which are in the order of the instantaneous data sampling rate, i.e., 200 ms in this work). The overall result is a significantly more accurate fit to the empirical pressure data, which corroborates the validity of the proposed modelling framework for pressure sensors with differential data reporting.

D. Motion Detection

As opposed to the physical parameters considered in Sections IV-A, IV-B and IV-C, which have a continuous domain, the detection of motion generates a discrete binary output based on two instantaneous states, namely an ON state (i.e., motion is being detected) and an OFF state (i.e., motion is not being detected). As a result, the modelling of the data traffic generated by a motion sensor requires a different modelling approach. In this section, three relevant aspects of the motion detection data traffic are modelled, namely the time interval between the arrival (beginning) of two successive ON states, the duration of the ON states, and the duration of the OFF states. The specific way a sensor reports motion depends on the particular implementation but will in most cases be based on a combination of one or more of these three aspects.

Table VIII shows the accuracy (in terms of the KS distance) of the distributions shown in Table I in fitting the three considered aspects of the motion detection data traffic. The two-best fitting distributions are shown in boldface and the single best-fitting model (i.e., the one providing the lowest KS distance) is underlined as well. Fig. 9 compares the ECDF with these best-fitting models, whose estimated parameters (based on the ML method) are provided in Table IX. The inspection of the values shown in Table VIII(a) reveals that the time between the arrival of two successive ON states is poorly fitted by the exponential distribution. As a result, the data traffic generated by a motion sensor that reports the triggering of ON states would not follow a Poisson process and therefore the Poisson model would not be suitable in this case either, similar to the case for the rest of considered physical parameters (ambient temperature, luminous intensity and atmospheric pressure). Table VIII indicates that other distributions constitute more accurate models, not only for the inter-arrival of ON states (generalised Pareto), but also for the duration of the ON (gamma) and OFF (generalised Pareto) states. In indoor environments such as smart homes, where the overall level of motion is in general limited, the durations of the OFF states will be significantly (even orders of magnitude) longer than the ON states, which in fact can be appreciated in the axes of abscissas of Figs. 9(b)–(c). In such a case, the interval between two successive ON states will be highly

similar to the duration of the OFF states, which explains the similarities observed between Figs. 9(a) and 9(c), not only in terms of time scales but also best-fitting distributions. Overall, the selected distribution models can provide a highly accurate (nearly perfect) fit to the empirical motion data and therefore constitute adequate traffic models for this type of data traffic.

V. Analysis of the Impact of Seasonal Variations and Correlations of the Physical Parameters

The models presented in Section IV were fitted based on the complete data set obtained over the course of one full calendar year. Some of the considered physical parameters may exhibit variations on their statistical properties when analysed over shorter specific time intervals such as months, days or hours. In particular, this affects the ambient temperature, luminous intensity and atmospheric pressure. For example, temperature and luminosity tend to be higher during daytime and during the warm months of the year, while pressure may experience similar seasonal variations. Moreover, these physical parameters typically show periodic patterns over different time scales (e.g., days, months, etc.) and can therefore be expected to show certain correlations, which may also induce correlations in the associated data traffic generated by the monitoring IoT sensors. This section investigates how several features of the considered physical parameters, such as correlations and seasonal variations along the year at different time scales, affect the statistics of the associated data traffic and its models.

A. Analysis of Seasonal Variations

As mentioned above, certain physical parameters may exhibit specific variation patterns at different time scales. As a result, the measurements reported by IoT sensors for the considered physical parameters (ambient temperature, luminous intensity and atmospheric pressure) might also vary at different rates depending on the considered time scale, which would also result in different packet generation rates in sensors that employ a differential reporting mechanism as the one considered in this work. If the packet generation rate varies substantially, the underlying statistics of the data traffic might be affected, which could thus affect the validity of the models developed in this work (based on one full-year data) when applied over shorter time scales. The purpose of this section is to explore whether the models presented in Section IV are valid when time scales shorter than a year are considered. To this end, the packet inter-arrival rate can be employed as a metric to quantify the extent to which the packet generation process is affected when the monitoring of the physical parameters is analysed over shorter time intervals.

The full-year data set was divided into subsets corresponding to shorter time periods (months, days and hours) and the packet inter-arrival time observed in each subset was calculated. The results are shown in Figs. 10, 11 and 12 for the monthly, daily and hourly packet inter-arrival times, respectively, for different traffic types (physical parameters) and values of the differential reporting threshold, ΔD (ΔT for temperature, ΔL for luminosity, and ΔP for pressure). As one may expect, an increase in ΔD leads to longer packet inter-arrival times. However, for a given ΔD, the observed packet inter-arrival times remain fairly constant over the different considered periods and time scales, without significant variations (with the exception of some maximum inter-arrival times, which might be due to infrequent or exceptional events).

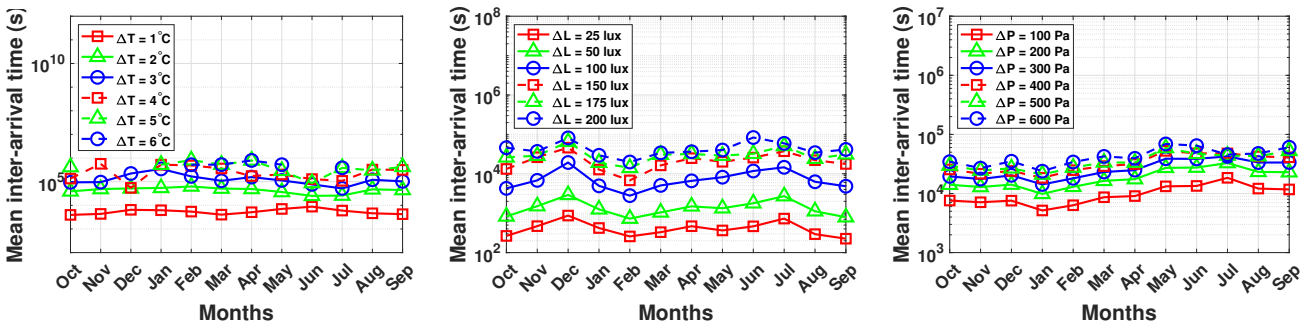
It is worth noting that Fig. 12(b) shows some missing points, which can be explained by the lack of any luminosity variations from approximately 9pm (when lights are switched off in the room where the sensor was placed) to 9am (when lights are switched on again and human activity resumes). Since the room remained in darkness overnight, no luminosity variations were observed during this time (despite the low ΔL threshold considered in this case), which explains the lack of points in this region of the figure (i.e., there are some 1-hour slots during the night where no packets are generated). Moreover, the average inter-arrival time observed in Fig. 12(b) shows a slight decrease during daytime, which might be due to more frequent variations of luminosity as a result of human activity (e.g., opening/closing of window blinds or movement of people blocking/unblocking light paths). This particular case may require some specific modelling, which is beyond the scope of this study and is suggested as future work.

Despite the particular trend observed in Fig. 12(b), which seems to disappear when longer time scales are considered (i.e., daily or monthly), the results shown in Figs. 10, 11 and 12 indicate that, overall, the data packet generation process of a sensor with differential reporting remains fairly constant over periods and time scales shorter than a year, which is in general true for the different physical parameters considered in this work. This suggests that the models presented in Section IV are valid when time scales shorter than a year are considered.

B. Analysis of Correlations

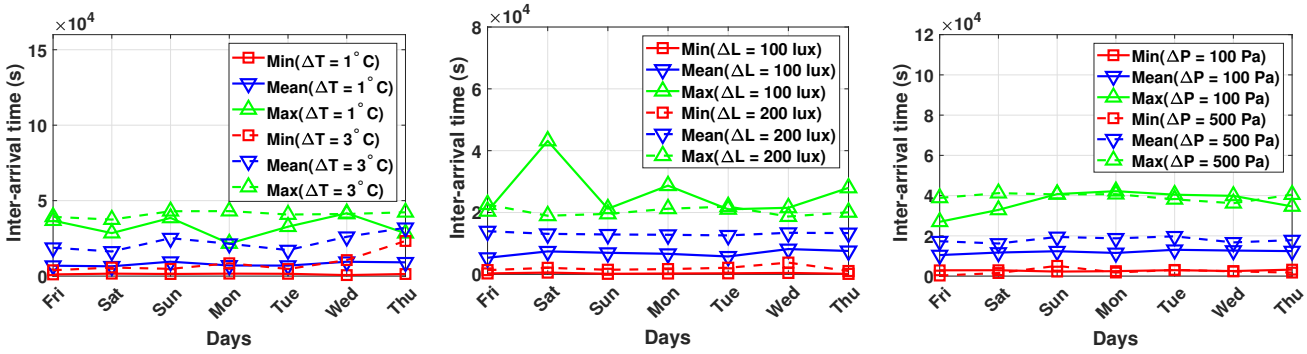
As mentioned above, certain physical parameters typically show periodic patterns over different time scales (e.g., days, months, etc.) that may result in underlying correlations in the observed values and this might also induce correlations in the data traffic that would be generated by a sensor monitoring the variations of these physical magnitudes. The purpose of this section is to determine the potential existence of correlations in the considered physical parameters (ambient temperature, luminous intensity and atmospheric pressure) and determine whether this results in correlations on the resulting packet inter-arrival times. It is worth noting that the models presented in Section IV are designed to capture the statistical distribution of the packet inter-arrival times but are unable to reproduce correlation properties. Therefore, the existence of correlations, if needed, would require specific additional modelling.

Fig. 13 shows the autocorrelation function of the physical parameters considered in this work (up to a maximum of 4 weeks lag) along with the autocorrelation of the resulting packet inter-arrival times (up to a maximum of 3 weeks lag). In most practical scenarios, an absolute value of the correlation coefficient greater than 0.5 is in general considered as an indication of strong correlation; this threshold can be used to determine whether certain data are autocorrelated. As observed in Fig. 13(a), the observed temperature values



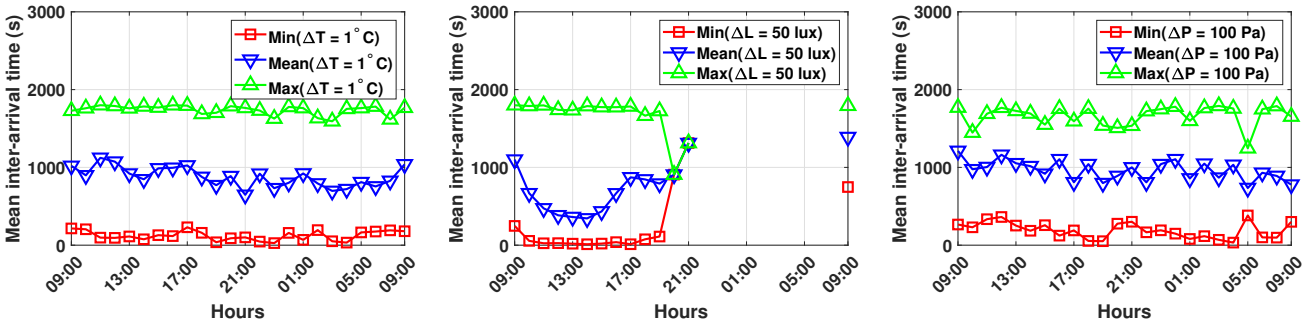
(a) Temperature-related traffic. (b) Luminosity-related traffic. (c) Pressure-related traffic.

Fig. 10. Monthly packet inter-arrival times for different traffic types.



(a) Temperature-related traffic. (b) Luminosity-related traffic. (c) Pressure-related traffic.

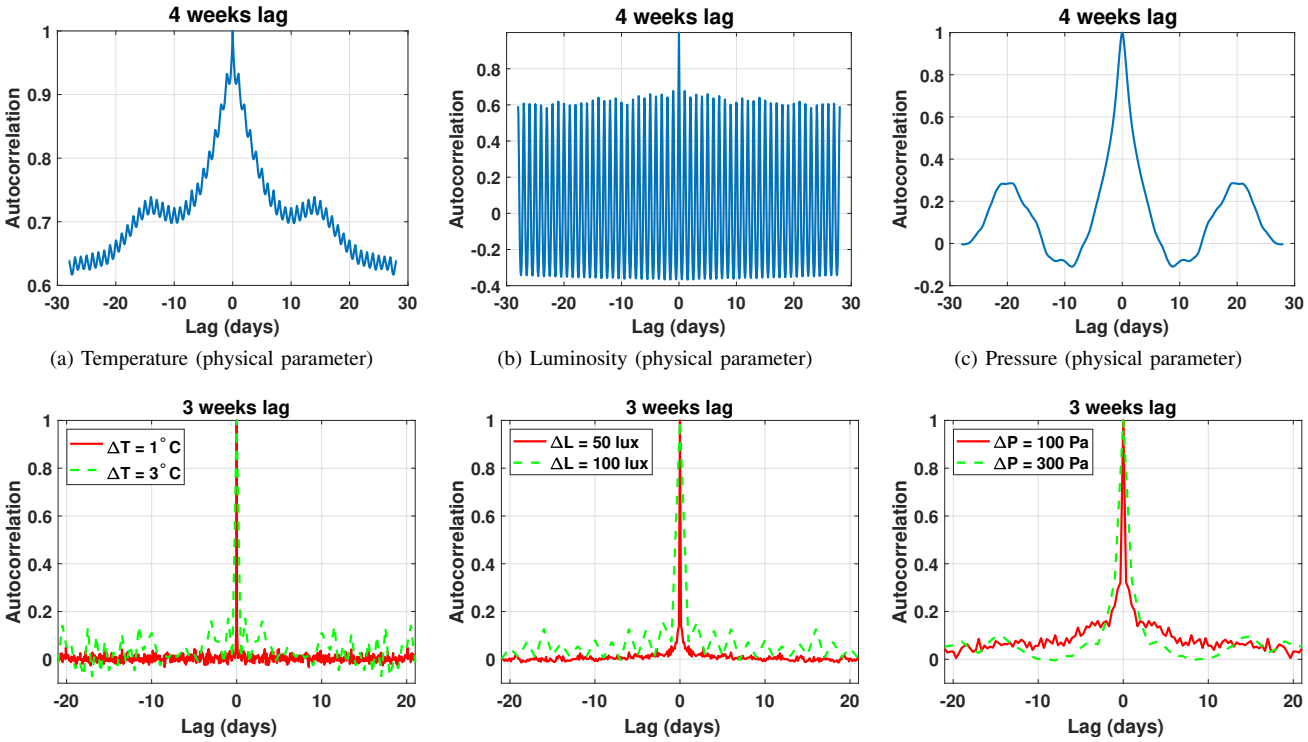
Fig. 11. Daily packet inter-arrival times for different traffic types.



(a) Temperature-related traffic. (b) Luminosity-related traffic. (c) Pressure-related traffic.

Fig. 12. Hourly packet inter-arrival times for different traffic types.

show a decaying autocorrelation as the lag increases, with all correlation values greater than 0.6, which indicates a strong correlation between temperature samples even over periods as long as around one month. This observation is consistent with the notion that significant temperature variations will in general occur between seasons (i.e., over periods longer than one month). Fig. 13(b) shows that luminosity samples also exhibit a relatively strong correlation, with periodic peaks at 0.6. In this case the autocorrelation function is periodic, which is the result of daily repetitive patterns (i.e., the separation between successive peaks is equal to one day). In the case of atmospheric pressure, Fig. 13(c) shows a looser correlation of the observed samples for this physical parameter, which is in general weak (except for a lag of very few days). Despite the existence of different levels of correlation in the physical parameters as shown in Figs. 13(a)–(c), the results shown in Figs. 13(d)–(f) indicate that the inter-arrival times of the data packets generated by a sensor with differential reporting are essentially uncorrelated. The variation of the ΔD parameter results in some minor fluctuations that do not affect significantly the overall autocorrelation function. As matter of fact, the autocorrelation is greater than the suggested threshold of 0.5 only for a lag equal to zero and lower otherwise, meaning that the time intervals elapsed between the generation of two consecutive data packets are mutually independent. This observation has important implications in some practical



(a) Temperature (physical parameter) (b) Luminosity (physical parameter) (c) Pressure (physical parameter)

(d) Temperature (packet inter-arrival times) (e) Luminosity (packet inter-arrival times) (f) Pressure (packet inter-arrival times)

Fig. 13. Autocorrelation function of the considered physical parameters (top row) and the corresponding packet inter-arrival times (bottom row) for the observed temperature (left column), luminosity (middle column) and pressure (right column).

applications of the models presented in Section IV as it is the case, for example, of software simulations, where the packet inter-arrival times could be generated as a set of independent and identically distributed random numbers drawn from the appropriate distribution model, which can be determined from the exhaustive analysis presented in Section IV.

VI. Conclusion

Current communication networks have historically been designed and optimised to support human-related services and as such are not perfectly suited to support the incoming myriad of new machine-type IoT devices, which have different requirements, features and traffic patterns. This claims for novel and realistic models that can accurately capture the relevant statistical properties of the data traffic generated by IoT devices. In this context, this work has presented a set of models that can accurately capture the statistical properties of the traffic generated by common smart home devices (e.g., ambient temperature, luminous intensity, atmospheric pressure and motion sensors). A differential data reporting mechanism has been considered in this work, which produces a significantly lower volume of data for slowly changing physical parameters such as those analysed in this study. The obtained results have demonstrated that the packet generation process cannot be modelled as a Poisson process (an aggregated traffic model frequently used in the literature also as a source traffic model). Consequently, the best-fitting distributions for each type of traffic have been determined based on the analysis of empirical data recorded over one complete calendar year. The accuracy analysis carried out in this work has shown that the proposed models can reproduce, with a high level of accuracy, the statistical distribution of the packet inter-arrival times. Moreover, these models are applicable over shorter time scales (months, days, and in most cases hours as well) and are not affected by the correlations commonly observed in slowly changing physical parameters. In summary, this work provides a set of accurate traffic models suitable for real-life IoT smart home scenarios that can be used for an adequate design and optimisation of future communication networks in order to efficiently support these and other common IoT services.

Appendix
Maximum Likelihood Estimation of Weighted Sum of Generalised Exponential Distributions

The weighted sum of generalised exponential distributions proposed in Section IV-C to fit the empirical pressure data has a CDF given by:

$$F_{WGE}(x) = \gamma \left[1 - e^{-\lambda_1 (x-\mu_1)}\right]^{\alpha_1} + (1-\gamma)\left[1 - e^{-\lambda_2 (x-\mu_2)}\right]^{\alpha_2} \quad (7)$$

where $0 < \gamma < 1$ is the weighting parameter. The corresponding PDF can be obtained as the derivative of (7), which yields:

$$f_{WGE}(x) = \gamma\alpha_1\lambda_1 e^{-\lambda_1(x-\mu_1)}\left[1-e^{-\lambda_1(x-\mu_1)}\right]^{\alpha_1-1} + (1-\gamma)\alpha_2\lambda_2 e^{-\lambda_2(x-\mu_2)}\left[1-e^{-\lambda_2(x-\mu_2)}\right]^{\alpha_2-1} \tag{8}$$

By setting $\mu_1 = \mu_2 = \mu$ as discussed in Section IV-C, which can be estimated as the minimum packet inter-arrival time observed in the empirical data, the number of variables in the model of (7)–(8) reduces to five (γ, λ_1, α_1, λ_2, α_2) and its log-likelihood function, $\mathcal{L}(\mathbf{X};\gamma,\lambda_1,\alpha_1,\lambda_2,\alpha_2)$, which for simplicity will be denoted as $\mathcal{L}(\mathbf{X})$, is then given by:

$$\mathcal{L}(\mathbf{X}) = \sum_{n=1}^{N}\log\left\{\gamma\alpha_1\lambda_1 e^{-\lambda_1(x_n-\mu)}\left[1-e^{-\lambda_1(x_n-\mu)}\right]^{\alpha_1-1} + (1-\gamma)\alpha_2\lambda_2 e^{-\lambda_2(x_n-\mu)}\left[1-e^{-\lambda_2(x_n-\mu)}\right]^{\alpha_2-1}\right\}$$

where $\mathbf{X} = \{x_n\}_{n=1}^{N}$ is the set of N empirically-observed packet inter-arrival times, x_n, and $\mu = \min_n\{x_n\}_{n=1}^{N}$. The five parameters of interest (γ, λ_1, α_1, λ_2, α_2) can be estimated as the values that maximise this log-likelihood function, which requires the calculation of the partial derivatives.

The partial derivative with respect to γ is given by:

$$\frac{\partial\mathcal{L}(\mathbf{X})}{\partial\gamma} = \sum_{n=1}^{N}\frac{1}{D_n}\left(\kappa_n^{(1)} - \kappa_n^{(2)}\right) \tag{9}$$

where

$$\kappa_n^{(1)} = \alpha_1\lambda_1 e^{-\lambda_1(x_n-\mu)}\left[1-e^{-\lambda_1(x_n-\mu)}\right]^{\alpha_1-1} \tag{10}$$

$$\kappa_n^{(2)} = \alpha_2\lambda_2 e^{-\lambda_2(x_n-\mu)}\left[1-e^{-\lambda_2(x_n-\mu)}\right]^{\alpha_2-1} \tag{11}$$

$$D_n = \gamma\kappa_n^{(1)} + (1-\gamma)\kappa_n^{(2)} \tag{12}$$

The partial derivative with respect to λ_1 is given by:

$$\frac{\partial\mathcal{L}(\mathbf{X})}{\partial\lambda_1} = \sum_{n=1}^{N}\frac{\gamma\alpha_1}{D_n}\left(\zeta_n^{(1)} + \zeta_n^{(2)}\right) \tag{13}$$

where D_n is given by (12) and

$$\zeta_n^{(1)} = [1-\lambda_1(x_n-\mu)]\times \times e^{-\lambda_1(x_n-\mu)}\left[1-e^{-\lambda_1(x_n-\mu)}\right]^{\alpha_1-1} \tag{14}$$

$$\zeta_n^{(2)} = \lambda_1(x_n-\mu)(\alpha_1-1)\times \times e^{-2\lambda_1(x_n-\mu)}\left[1-e^{-\lambda_1(x_n-\mu)}\right]^{\alpha_1-2} \tag{15}$$

The partial derivative with respect to α_1 is given by:

$$\frac{\partial\mathcal{L}(\mathbf{X})}{\partial\alpha_1} = \sum_{n=1}^{N}\frac{\gamma\lambda_1 e^{-\lambda_1(x_n-\mu)}}{D_n}\,\epsilon_n^{(1)}\left(1+\epsilon_n^{(2)}\right) \tag{16}$$

where D_n is given by (12) and

$$\epsilon_n^{(1)} = \left[1-e^{-\lambda_1(x_n-\mu)}\right]^{\alpha_1-1} \tag{17}$$

$$\epsilon_n^{(2)} = \alpha_1\log\left\{1-e^{-\lambda_1(x_n-\mu)}\right\} \tag{18}$$

The partial derivatives of $\mathcal{L}(\mathbf{X})$ with respect to λ_2 and α_2 can be obtained based on those for λ_1 and α_1, respectively, by replacing γ with $1-\gamma$, λ_1 with λ_2, and α_1 with α_2. By equating all partial derivatives to zero and solving the resulting system of equations numerically, the set of values of γ, λ_1, α_1, λ_2 and α_2 that maximises the log-likelihood function $\mathcal{L}(\mathbf{X})$ is obtained, which constitutes the set of parameter estimates.

REFERENCES

[1] L. Atzori, A. Iera and G. Morabito, "The Internet of Things: A survey," *J. Comp. Netw.*, vol. 54, no. 15, pp. 2787-2805, Oct. 2010.

[2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, "Internet of Things-A Survey on Enabling Technologies, Protocols and Applications", *IEEE Commun. Surv. & Tut.* vol. 17, no. 4, pp. 2347-2376, Forth Quarter 2015.

[3] L. Atzori, A. Iera and G. Morabito, "From smart objects to social objects: The next evolutionary step of the internet of things," *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 97-105, Jan. 2014.

[4] M. López-Benítez, T. D. Drysdale, S. Hadfield and M. I. Maricar, "Prototype for multidisciplinary research in the context of the Internet of Things," *J. Netw. and Comp. Appl.*, vol. 78, pp. 146-161, Jan. 2017.

[5] E. Park, Y. Cho, J. Han and S. J. Kwon, "Comprehensive Approaches to User Acceptance of Internet of Things in a Smart Home Environment," *IEEE Internet of Things J.*, vol. 4, no. 6, pp. 2342-2350, Dec. 2017.

[6] S. Feng, P. Setoodeh and S. Haykin, "Smart Home: Cognitive Interactive People-Centric Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 34-39, Feb. 2017.

[7] C. Vallati, A. Virdis, E. Mingozzi and G. Stea, "Mobile-Edge Computing Come Home Connecting things in future smart homes using LTE device-to-device communications," *IEEE Consumer Elec. Mag.*, vol. 5, no. 4, pp. 77-83, Oct. 2016.

[8] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang and J. Wang, "Large-Scale Measurement and Characterization of Cellular Machine-to-Machine Traffic," *IEEE/ACM Trans. Networking*, vol. 21, no. 6, pp. 1960-1973, Dec 2013.

[9] K. Smiljkovic, V. Atanasovski and L. Gavrilovska, "Machine-to-Machine Traffic Characterization: Models and Case Study on Integration in LTE," in Proc. *4th Int. Conf. Wireless Commun., Vehic. Technol., Information Theory and Aerospace & Electronic Systems (VITAE 2014)*, pp. 1-5, 2014.

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

[11] M. Centenaro, and L. Vangelista, "A Study on M2M Traffic and Its Impact on Cellular Networks," in Proc. *IEEE 2nd World Forum on Internet of Things (WF-IoT 2015)*, pp. 154-159, 2015.

[12] A. H. El-Fawal, M. Najem, A. Mansour, F. Le-Roy, D. Le-Jeune, "CTMC modelling for H2H/M2M coexistence in LTE-A/LTE-M networks," *The Journal of Engineering*, vol. 2018, no. 12, pp. 1954-1962, Dec. 2018.

[13] M. R. Palattella et al., "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE J. Sel. Areas in Commun.*, vol. 34, no. 3, pp. 510-527, Mar. 2016.

[14] D. Zhang, Z. Zhou, S. Mumtaz, J. Rodriguez and T. Sato, "One Integrated Energy Efficiency Proposal for 5G IoT Communications," *IEEE Internet of Things J.*, vol. 3, no. 6, pp. 1346-1354, Dec. 2016.

[15] S. Andreev et al., "Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 32-40, Sep. 2015.

[16] A. Rico-Alvarino et al., "An overview of 3GPP enhancements on machine to machine communications," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 14-21, Jun. 2016.

[17] N. Nikaein, M. Laner, K. Zhou, P. Svoboda, D. Drajic, M. Popovic and S. Krco, "Simple Traffic Modeling Framework for Machine Type Communication," in Proc. *International Symposium on Wireless Commun. Systems (ISWCS 2013)*, pp. 1-5, Aug. 2013.

[18] M. Laner, P. Svoboda, N. Nikaein and M. Rupp, "Traffic Models for Machine Type Communications," in Proc. *Int. Symposium on Wireless Commun. Systems (ISWCS 2013)*, pp. 1-5, Aug. 2013.

[19] E. Grigoreva, M. Laurer, M. Vilgelm, T. Gehrsitz and W. Kellerer, "Coupled Markovian arrival process for automotive machine type communication traffic modeling," in Proc. *IEEE Int. Conf. on Communications (ICC 2017)*, pp. 1-6, May 2017.

[20] M. Sansoni, G. Ravagnani, D. Zucchetto, C. Pielli, A. Zanella, K. Mahmood, "Comparison of M2M traffic models against real world data sets," in Proc. *IEEE 23rd Int. Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD 2018)*, pp. 1-6, 2018.

[21] T. Hoßfeld, F. Metzger and P. E. Heegaard, "Traffic modeling for aggregated periodic IoT data," in Proc. *Innov. in Clouds, Internet and Netw. and Workshops (ICIN)*, pp. 1-8, Dec. 2018.

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

[23] H. Thomsen, C. Navarro-Manchón, and B. H. Fleury, "A Traffic Model for Machine-Type Communications Using Spatial Point Processes," in Proc. *IEEE 28th Annual Int. Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2017)*, pp. 1-6, 2017.

[24] M. Gharbieh , H. ElSawy, H.-C. Yang, A. Bader, and M.-S. Alouini, "Spatiotemporal Model for Uplink IoT Traffic: Scheduling and Random Access Paradox," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8357-8372, Dec. 2018.

[25] M. Laner, N. Nikaein, P. Svoboda, M. Popovic, D. Drajic and S. Krco, "Traffic models for machine-to-machine (M2M) communications: types and applications," in "Machine-to-machine communications, architecture, performance and applications," edited by M. Dohler and C. Antón-Haro, Woodhead Publishing, 2015.

[26] 3GPP, "RAN Improvements for Machine-type Communications," Tech. Report, 3GPP TR 37.868, v11.0.0, Sep. 2011.

[27] 3GPP, "Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE," Tech. Report, 3GPP TR 36.888, Jun. 2013.

[28] O. Al-Khatib, W. Hardjawana and B. Vucetic, "Traffic Modeling for Machine-to-Machine (M2M) Last Mile Wireless Access Networks," in Proc. *IEEE Global Commun. Conf. (GLOBECOM 2014)*, pp. 1199-1204, 2014.

[29] F. Al-Turjman, E. Ever and H. Zahmatkesh, "Green Femtocells in the IoT Era: Traffic Modeling and Challenges – An Overview," *IEEE Netw.*, vol. 31, no. 6, pp. 48-55, Dec. 2017.

[30] M. Bacco, P. Cassarà, M. Colucci and A. Gotta, "Modeling Reliable M2M/IoT Traffic Over Random Access Satellite Links in Non-Saturated Conditions," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 5, pp. 1042-1051, May 2018.

[31] V. Gupta, S. K. Devar, N. H. Kumar and K. P. Bagadi, "Modelling of IoT Traffic and its Impact on LoRaWAN," in Proc. *IEEE Globecom Commun. Conf.*, pp. 1-6, 2017.

[32] C. Majumdar, M. López-Benítez and S. N. Merchant, "Experimental Evaluation of the Poissoness of Real Sensor Data Traffic in the Internet of Things", in Proc. *IEEE Consumer Commun. and Netw. Conf.(CCNC)*, pp. 1-7, Jan. 2019.

[33] C. Majumdar, M. López-Benítez and S. N. Merchant, "Accurate Modelling of IoT Data Traffic Based on Weighted Sum of Distributions", in Proc. *IEEE Intl. Conf. on Commun.(ICC)*, pp. 1-6, May. 2019.

[34] M. López-Benítez, "Empirical pressure, altitude, temperature, light and motion sensor data collected over a complete calendar year in an indoor environment," Data collection, Research Data Catalogue, University of Liverpool, United Kingdom, April 2019. Available: http://dx.doi.org/10.17638/datacat.liverpool.ac.uk/722.

[35] D. P. Heyman, M. J. Sobel, "Stochastic models in operations research: Stochastic processes and operating characteristics," Dover, 2003.

[36] J. Gozalvez,"New 3GPP Standard for IoT [Mobile Radio]," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 14-20, Mar. 2016.

[37] A. Hoglund et al.,"Overview of 3GPP Release 14 Enhanced NB-IoT," *IEEE Netw.*, vol. 31, no. 6, pp. 16-22, Dec. 2017.

[38] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes: The Art of Scientific Computing", 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[39] M. Abramowitz and I. A. Stegun, "Handbook of mathematical functions with formulas, graphs, and mathematical tables," 10th ed. New York:Dover, 1972.

[40] M. López-Benítez and F. Casadevall, "Time-Dimension Models of Spectrum Usage for the Analysis, Design, and Simulation of Cognitive Radio Networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2091-2104, Jun. 2013.

[41] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical recipes: The art of scientific computing," 3rd ed. Cambridge University Press, 2007.

Chitradeep Majumdar received a BTech degree in Electronics and Communication from Bengal Institute of Technology under West Bengal University of Technology in 2007, a Masters degree in Radio Frequency Communication Systems from University of Southampton in 2009, and a PhD degree from the Indian Institute of Technology Bombay in 2017. During his PhD, he worked extensively on cross-layer optimization for cognitive radio sensor networks. From 2013 to 2015, he was awarded the MEXT fellowship by the Government of Japan to work as research student at the University of Tokyo. From February 2018 to December 2019, he was a post-doctoral research associate at the Department of Electrical Engineering and Electronics, University of Liverpool, UK under a Royal Society-SERB Newton International Fellowship, where he worked on the optimization of next generation networks for device-to-device communications. Dr Majumdar is currently associated with the Samsung Research Institute Bangalore as a Senior Lead Engineer. His research interests include Internet of Things, 5G NR, beyond 5G communications (THz technologies), narrowband IoT, cognitive radio, sensor networks and other related topics.

Miguel López-Benítez (S'08, M'12, SM'17) received the BSc and MSc degrees (both with Distinction) in Telecommunication Engineering from Miguel Hernández University, Elche, Spain in 2003 and 2006, respectively, and the PhD degree (*summa cum laude*) in Telecommunication Engineering from the Technical University of Catalonia, Barcelona, Spain in 2011. From 2011 to 2013, he was a Research Fellow with the Centre for Communication Systems Research, University of Surrey, Guildford, UK. In 2013, he became a Lecturer (Assistant Professor) with the Department of Electrical Engineering and Electronics, University of Liverpool, UK, where he has been a Senior Lecturer (Associate Professor) since 2018. His research interests include the field of wireless communications and networking, with special emphasis on mobile communications and dynamic spectrum access in cognitive radio systems. He has been the Principal Investigator or Co-Investigator of research projects funded by the EPSRC, British Council, and Royal Society, and has also been involved in the European-funded projects AROMA, NEWCOM++, FARAMIR, QoSMOS, and CoRaSat. He has been a member of the Organising Committee for the IEEE WCNC International Workshop on Smart Spectrum since 2015 and is currently an Associate Editor of IEEE Access, IET Communications, and Wireless Communications and Mobile Computing. Please visit http://www.lopezbenitez.es for more details.

Shabbir N. Merchant received his BTech, MTech and PhD degrees all from the Department of Electrical Engineering, Indian Institute of Technology Bombay, India. Currently he is a Professor (Emeritus Fellow) in the Department of Electrical Engineering at IIT Bombay. He has more than 30 years of experience in teaching and research. Prof Merchant has made significant contributions in the field of signal processing and its applications. His noteworthy contributions have been in solving state-of-the-art signal and image processing problems faced by Indian defence. His broad area of research interests includes wireless communications, wireless sensor networks, signal processing, multimedia communication, and image processing, where he has published extensively in these areas. He is a co-author with his students who have won Best Paper Awards. He has served on Technical Program Committees of many IEEE premier conferences. He serves on the Editorial Board of two International Journals: International Journal of Distributed Sensor Networks, and International Journal of Ultra Wideband Communications and Systems. He has been a chief investigator for a number of sponsored and consultancy projects. He has served as a consultant to both private industries and defence organizations. He is on the Academic and Governing Advisory Boards of different engineering colleges in India. He is a Fellow of IETE. He is a recipient of the 10th IETE SVC Aiya Memorial Award for his contribution in the field of detection and tracking. He is also a recipient of the 9th IETE SVC Aiya Memorial Award for "Excellence in Telecom Education". He is a winner of the 2013 VASVIK Award in the category of Electrical & Electronic Sciences & Technology.