Experimental Evaluation of the Poissoness of Real Sensor Data Traffic in the Internet of Things

Chitradeep Majumdar*, Miguel López-Benítez* and S. N. Merchant†
* Department of Electrical Engineering and Electronics, University of Liverpool
† Department of Electrical Engineering, IIT Bombay
email: Chitradeep.Majumdar@liverpool.ac.uk, M.Lopez-Benitez@liverpool.ac.uk, merchant@ee.iitb.ac.in

***Abstract*—This work proposes a novel experimental and mathematical framework to determine the statistical models for the Internet of Things (IoT) data traffic. Conventionally, it is assumed that the data packet generation for IoT based applications follows a Poisson process with exponentially distributed packet inter-arrival time. Based on such generalized premise, majority of the network related theoretical and practical analysis of the IoT platforms are carried out. Based on empirical data for a smart home application recorded for over 10 weeks duration using proposed IoT subsystem, in this paper we estimate the empirical statistical distribution of the IoT data traffic generated by temperature, light intensity and motion sensors. The inter-arrival between the data packets generated from different sensing modules of the IoT smart home application subsystem is determined. The Empirical Cumulative Distribution Function (ECDF) of the estimated time duration is fitted with few of the well-established classical statistical distributions using Method of Moments (MoM) and Maximum Likelihood (ML) estimation techniques. The goodness of fit is quantified using Kolmogrov-Smirnov (KS) test. The parameters of the fitted distributions are determined as a function of the physical input parameters. The results reveal source IoT traffic does not follow a Poisson process which is conventionally assumed in the literature, but rather depends on the type of application.**

***Index Terms*—KS-Kolmogrov Smrinov Test, IoT-Internet-of-Things, D2D-Device-to-Device, 3GPP, Traffic Modelling.**

I. Introduction

Communication networks at present and in future will not be just about connecting people, but 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. 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 realize 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 subsystem prototype is discussed in [4]. End user test cases like smart home applications, industrial equipments monitoring and healthcare are few of the classical examples where the IoT paradigm is used extensively.

While mobile networks 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. It is quite evident that the next generation cellular communication network will depend significantly on the IoT data traffic. Therefore, it is essential to model the data traffic for the IoT based applications. To this end few state of the art that has been proposed so far includes [5]-[8]. Even similar traffic models are considered by the next generation mobile communication standardization body 3GPP which considers Poisson process distribution of data traffic generation [9] and [10]. However, in terms of statistical traffic modelling all the works described above are primarily theoretical in nature. Furthermore, majority of the existing work does not take into account any empirical data for specific applications or sensor type recorded using practical IoT subsystem for over a considerable duration of time. Therefore, to the best of our knowledge there are no or too few literatures available that have addressed empirical data based IoT traffic modelling. In this work, we have determined the statistical traffic models for realistic indoor smart home end-user test case which is not available in literature. Data was obtained for ambient temperature, light intensity and motion detection using appropriate sensors and our in-house developed Raspberry Pi based IoT subsystem for over 10 weeks [4]. Most theoretical work considers that the IoT data is generated following a Poisson process but this work is based on real data and finds the distribution of the inter-arrival time of the reported data packets that fits the best. Exhaustive numerical analysis and analytical tools were then used to determine statistical distributions which could best fit different application types for a single node-multi-sensor differential reporting scheme. The rest of this work is organized as follows. Section II contains the description of the experimental setup and the methodology. Section III describes the techniques used towards statistical fitting of the inter-arrival duration of the data packets, Section IV contains the analysis of the empirical data and Section V contains the conclusion.

II. Experimental Platform and Methodology

A. Experimental Setup

Fig. 1, shows the sensor node that replicates smart home scenario [4]. Individual sensor node comprises six sensor modules (element 1 to 6). These sensor components are connected to a Raspberry Pi minicomputer which is labelled at element 7. It facilitates connection to the central processing unit through USB WiFi adapter marked as

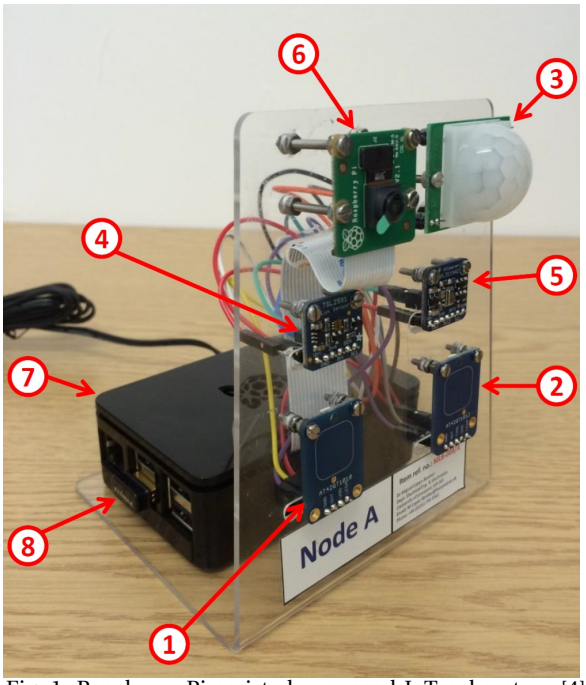


Fig. 1. Raspberry Pi assisted proposed IoT subsystem [4]

element 8. There are capacitive touch sensors (element 1 and 2), motion sensor (element 3), light sensor (element 4), temperature sensor (element 5) and an image sensor module (element 6) in the node. However, in this paper we are focussing primarily on the temperature, light intensity and the motion sensor data which are typically found on a smart home environment. In this work for the temperature, MPL3115A2 sensor (Adafruit 1893) is used. For the motion data, Passive infra-red Passive infra-red (PIR) motion sensor HC-SR501 and in case of the light data TSL2591 sensor (Adafruit 1980) are considered. The detailed sensor specifications along with the embedded system block diagram and software implementation can be referred from [4].

B. Differential Data Reporting scheme and Inter-arrival Time estimation

The amount of IoT traffic generated strongly depends on the reporting strategy. Furthermore, the reporting strategy could be periodic or differential. In periodic reporting scheme the recorded data is periodically transmitted to the gateway. Whereas, in differential reporting scheme a data packet transmission is assumed when the absolute difference in the physical parameter (ΔD) for temperature and light intensity is greater or smaller than a predetermined threshold as per the system design requirement. The analysis of the periodic reporting strategy is relatively trivial as compared to differential reporting strategy since the inter-arrival time is deterministic. In differential reporting scheme the inter-arrival time between the data packet generation will be a random variable as it is unknown when ΔD will be satisfied thus making it non-deterministic. Our motivation is to model the IoT traffic for the differential reporting scheme by estimating the distribution of the inter-arrival time of data packet generation and reporting.

In this work, simple data collection strategy is used to record the data where proposed IoT subsystem senses and samples the data at every 200 ms interval for 10 weeks. The data is then transferred and stored at the CPU back-end.

The differential input parameter denoted as ΔD for the temperature data is varied from 0.1 to 5 °C and the light intensity data is varied from 20 to 200 lux respectively. Once the parameter value which exceeds the input threshold is obtained, the initial data point gets updated and stored as that value of the physical parameter and its corresponding index. The process is continued throughout the length of the data vector.

III. Statistical fitting of the Inter-Arrival Time

The observed inter-arrival times from the captured experimental data were processed to calculate the Empirical Cumulative Distribution Function (ECDF) and then it was used to fit the distribution models. The inter-arrival times are random variable as the time duration among the parameter fluctuations are non-deterministic. Upon estimation of the ECDF, next step is to determine the statistical distribution which would provide best fit to the ECDF. The location, shaping and scaling parameters of the best fitted distribution are subsequently determined using either Method-of-Moments (MoM) or Maximum Likelihood Estimation (MLE) technique. The best fit is quantified in terms of the Kolmogrov-Smrinov distance denoted as D_{KS} which is the maximum of the absolute difference value between the empirical CDF and the fitted statistical CDF estimated at each sample point and expressed as:

$$D_{KS} = max\left\{abs\left(F_X^{ECDF}(x) - F_X^{fitted}(x)\right)\right\} \quad (1)$$

Seven classical distribution functions are considered in this work [11]. These are Exponential, Generalized Exponential, Pareto, Generalized Pareto, Gamma, Log Normal and Weibull Distributions. Numerical fitting is based on Method-of-Moments (MoM) and Maximum Likelihood (ML) [12]. In MoM technique, statistical moments of all the distributions mentioned above in terms of their location, scaling and shaping parameter are equated with the first and the second order moments (mean and variance) of the available sample population. Here, the sample population is the vector of inter-arrival times denoted as (X) where x_i is each element of the vector such that $(x_i \in X)$ for $i = \{1,2...N\}$ and N is the maximum length of the vector X or the maximum number of real and finite feasible inter-arrival durations. Subsequently, the location, shaping and the scaling parameters of the fitted distributions are determined along with their respective Cumulative Distribution Function (CDF). The CDF of the fitted distributions which best fits the ECDF based on the KS metric is selected as the best fitting distribution for that specific ΔD.

In ML estimation, the log likelihood factor is maximized in terms of the shaping and the scaling parameter of the distribution. The minimum value (μ) does not need to be included in the equations as it is estimated as the minimum observed value in the vector X. The Log-Likelihood function is determined in terms of the Probability Density Function

(PDF) of the distribution denoted as $f_X(x)$. The PDF is which is also a function of the scaling (λ) and shaping (α) and is given by the derivative of the CDF as

$$f_X(x_i, \lambda, \alpha) = \frac{\partial F_X(x_i, \lambda, \alpha)}{\partial x_i} \tag{2}$$

The Likelihood function and the Log-Likelihood function are given by

$$L(x, \lambda, \alpha) = \prod_{i=1}^{N} f_X(x_i, \lambda, \alpha) \tag{3}$$

$$\log L(x, \lambda, \alpha) = \log \prod_{i=1}^{N} f_X(x_i, \lambda, \alpha)$$

$$\log L(x, \lambda, \alpha) = \sum_{i=1}^{N} \log f_X(x_i, \lambda, \alpha)$$

The Log-Likelihood Function now needs to be maximized using other numerical optimization techniques using gradient descent, interior region method, etc. This would provide the desired scaling and shaping parameters of the fitted distribution. Upon estimation of the fitted distribution, similar to that of the Method of Moments the KS distance can be estimated to provide the best fit.

IV. Analysis of the Empirical Data

A. Temperature Data

Table I shows the D_{KS} values between the ECDF and different standard statistical distributions of the inter-arrival time duration for differentially reported temperature data ΔT. It is clearly observed from the first row of the table that the exponential distribution provides a very bad fit across all the ΔT values ranging from 0.1 to 2.9 °C. Therefore the Poisson process model, which assumes exponentially distributed inter-arrival times, is not realistic for the temperature reporting application. The threshold differential temperature fluctuation is varied from 0.1 °C to 2.9 °C. It is observed that the minimum KS distance for the temperature data using either method of moments or Maximum Likelihood estimation technique for ΔT= 0.1 °C is 0.237, ΔT= 0.5 °C is 0.0837, ΔT= 0.9 °C is 0.1178 and ΔT= 1.3 °C is 0.0435, all following Weibull Distribution marked in bold italics. While for ΔT= 1.7 °C is 0.0418 with Gamma Distribution and for ΔT= 2.9 °C is 0.0454 with Generalized-Pareto Distribution. Furthermore, it is also observed that in case of Weibull Distribution, the KS distance at ΔT= 1.7 °C is 0.0453 and for ΔT= 2.9 °C is 0.0558 marked underlined. It is clearly evident that the difference between the minimum D_{KS} values using the Gamma and Generalized-Pareto distribution as compared to Weibull Distribution at ΔT= 1.7 °C and ΔT= 2.9 °C is not widely different. Therefore, it can be concluded that the Weibull Distribution provides overall the best fit for all the ranges of the temperature difference parameter ΔT ranging from 0.1 to 2.9 °C .

Upon numerical computation of the KS distances for different temperature values for the Weibull Distribution, the location parameter (μ), the shaping parameter α and scaling parameter λ is estimated for all the values from 0.1 to 5° C with an increase of 0.1 °C. Once the values are obtained, they are interpolated with higher resolution and then conventional curve fitting technique is applied using MATLAB curve fitting toolbox to obtain a generalized analytical expression for the μ, λ and the α parameters as a function of temperature increase. From simulation, the best analytical fit turns out be Generalized Gaussian which is simply sum of weighted Gaussian exponentials. Based on these analytically calculated μ, λ and α values, the Weibull Distribution over inter-arrival times (X) is estimated for a specific temperature difference value (ΔT). The KS distance is then measured with respect to the Empirical CDF of the X for that given value of ΔT. The values of the location, scaling and the shaping parameter are expressed as

$$\lambda(\Delta D) = \sum_{i=1}^{K} a_i e^{-\left(\frac{\Delta D - b_i}{c_i}\right)^2} \tag{4}$$

$$\alpha(\Delta D) = \sum_{i=1}^{K} a_i e^{-\left(\frac{\Delta D - b_i}{c_i}\right)^2} \tag{5}$$

and

$$\mu(\Delta D) = \sum_{i=1}^{K} a_i e^{-\left(\frac{\Delta D - b_i}{c_i}\right)^2} \tag{6}$$

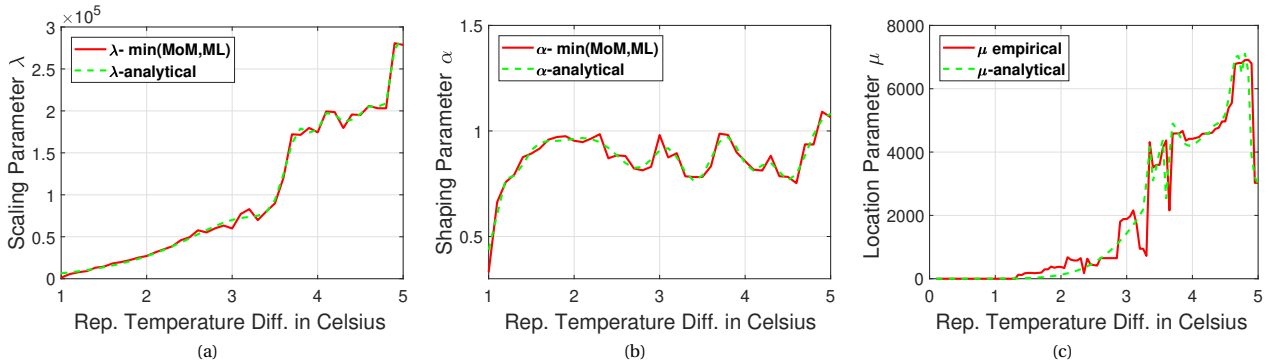
where, $D \hat{=} T$ in case of temperature data. In (4)-(6), K= 8. The coefficients a_i, b_i and c_i for the α, λ and the μ values are shown in Table II(a), (b) and (c). Fig. 2 shows the curve fitting for the scaling, the shaping parameter and the location parameters for Weibull distribution over reported differential temperature ΔD= ΔT. As evident from the figures, the data points do not follow a linear or close to linear fit and any fundamental curve fitting technique to obtain the fit for the parameters would result in wrong KS distance estimation which would become very unrealistic. Therefore, from a class of a number of fitting techniques available in the MATLAB curve-fitting toolbox that included a class of curve fitting options like Linear, Polynomial, Exponential, Fourier, Sum of Sines, Rational Power, Weibull, etc. we selected 8-th order Generalized Gaussian fit as it provided the best curve fitting for majority of the shaping, scaling and the location parameters. Fig. 3 shows the comparison between the minimum of the KS distances (D_{KS}) obtained numerically through Method-of-Moments (MoM) and Maximum Likelihood (ML) numerical fitting technique compared to the distribution whose parameters are computed through Matlab based curve fitting technique and the Analytical Fit obtained using the shaping, scaling and the location parameters in terms of ΔT using the Generalized Gaussian Fit. It could be observed that both the numerical and the analytical KS distances are less than 10%.

B. Light Intensity Data

Table III compares the KS distances for the light intensity data with ΔL ranging from 20 lux to 140 lux. Similar to Table I like the previous case of temperature, the exponential distribution fails in case of the luminous intensity data for ΔL ranging from 20 to 200 lux. The D_{KS} for exponential distribution is much higher than acceptable range

Table I. KS Distances of different statistical distributions- Differential Temperature Reporting (in °C) for 10 weeks duration

Distributions	ΔT= 0.1 ° C		0.5 ° C		0.9 ° C		1.3 ° C		1.7 ° C		2.9 ° C	
	MoM	ML	MoM	ML	MoM	ML	MoM	ML	MoM	ML	MoM	ML
Exp	0.2402	0.2402	0.1757	0.1757	0.2563	0.2563	0.1269	0.1269	0.0529	0.0529	0.0754	0.0754
G.Exp	0.1634	0.1634	0.1168	0.1168	0.1557	0.1557	0.0548	0.0548	0.0475	0.0475	0.0823	0.0823
Pareto	0.6353	0.2076	0.8793	0.4121	0.7738	0.3391	0.9925	0.4828	0.7926	0.3353	0.7120	0.3079
G.Pareto	0.1677	0.0879	0.1051	0.1050	0.2208	0.2288	0.0982	0.0706	0.0519	0.0523	***0.0454***	0.0511
Log-N	0.1882	0.0445	0.1545	0.1882	0.2844	0.1939	0.1929	0.0591	0.1522	0.0837	0.1107	0.0617
Gamma	0.1412	0.1032	0.1008	0.1355	0.1649	0.2047	0.0487	0.0543	0.0479	***0.0418***	0.0757	0.0729
Weibull	***0.0237***	0.0563	0.0877	***0.0837***	0.2002	***0.1178***	0.0515	***0.0435***	0.0495	***0.0453***	***0.0558***	0.0770



(a) (b) (c)

Fig. 2. a. Scaling Parameter (λ) vs Temperature Difference b. Shaping Parameter (α) vs Temperature Difference c. Location Parameter (μ) vs Temperature Difference, all for Weibull Distribution

Table II. Parameters obtained analytically for Temperature data

(a) λ parameter

a_i	b_i	c_i
-3.618×10^4	4.777	0.08703
1.64×10^4	4.881	0.03525
4.087×10^4	4.114	0.1401
-4.31×10^4	4.645	0.3241
0	15.87	1.709
8.102×10^4	3.77	0.2145
-6.047×10^4	3.758	0.5912
3.657×10^5	6.121	2.55

(b) α parameter

a_i	b_i	c_i
1.079	5.005	0.6403
0.7986	3.796	0.3808
0.9632	2.141	1.184
0.394	3.13	0.3048
0.1924	1.464	0.3261
0.1626	1.181	0.1159
0.7043	0.2171	0.4053
0.3563	4.285	0.2248

(c) μ parameter

a_i	b_i	c_i
2797	1.611	0.03361
2979	1.513	0.09541
653.4	1.348	0.188
-2759	0.7529	0.02518
-118.1	1.206	0.0105
1640	0.7609	0.1446
4328	1.209	0.8489
2021	0.5533	0.02146

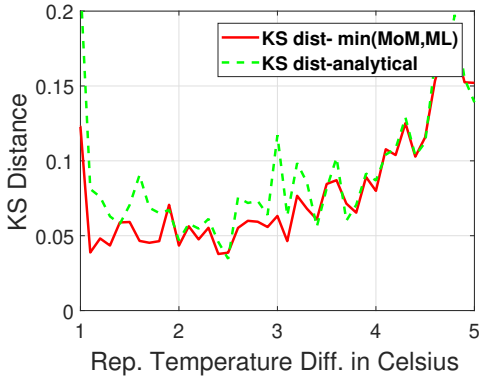


Fig. 3. Numerically and Analytically computed KS distances vs Temperature Difference, for Weibull Distribution

of 10%. Therefore, Poisson process is not a valid case in this scenario either. It is observed that for 20 and 40 lux, Generalized-Pareto distribution provides best fit with KS distance 0.0261 and 0.0279 respectively under ML fitting. For 60, 80 and 100 lux, Log-Normal distribution provides best fit with KS distances of 0.0290, 0.0275 and 0.0456 under ML fitting. For 140 lux Gamma distribution is the best fit with KS value of 0.0674. Similar to the case of temperature data, at ΔL= 20, 40 and 140 lux the KS values of the Log-Normal Distribution provides a reasonably close fit KS value 0.0660, 0.0401 and 0.0744 marked with underline. The difference between these values underlined with the minimum KS distance under Generalized-Pareto and Gamma distribution marked in bold italics is not significant. Furthermore, with Log-Normal Distribution at ΔL= 60 and 80 lux, the KS distance is less than 3% which is almost a perfect fit. Therefore, we select the Log-Normal Distribution as best fit. Fig. 4 shows the scaling parameter λ, the location parameter μ and the KS distances is shown. Similar to that of the temperature data, in case of ΔL data the analytical expressions for the distribution parameters are estimated as a function of ΔL. In this case the parameters λ and μ can be estimated as a function of ΔL from (4) and (6) with coefficients $\{a_i, b_i, c_i\}$ obtained Table V(a) and Table V(b) for corresponding λ and μ.

If we closely examine Fig. 4(c), it is observed that as we increase the value of ΔL beyond 140 lux, the value of the KS

Table III. KS Distances of different statistical distributions- Differential Light Intensity Reporting (in degree Lux) for 10 weeks duration

Distributions	ΔL= 20 lux		40 lux		60 lux		80 lux		100 lux		140 lux	
	MoM	ML	MoM	ML	MoM	ML	MoM	ML	MoM	ML	MoM	ML
Exp	0.6120	0.6120	0.5761	0.5761	0.5473	0.5473	0.4975	0.4975	0.4324	0.4324	0.3095	0.3095
G.Exp	0.4379	0.4379	0.5487	0.5487	0.5259	0.5259	0.3701	0.3701	0.1629	0.1629	0.0867	0.0867
Pareto	0.4697	0.0962	0.6898	0.2206	0.8042	0.2839	0.8332	0.2764	0.8463	0.2779	0.9236	0.3759
G.Pareto	0.5319	***0.0261***	0.4946	***0.0279***	0.4638	0.0395	0.4267	0.0587	0.3731	0.0919	0.2512	0.0750
Log-N	0.2401	***0.0660***	0.2951	***0.0401***	0.3364	***0.0290***	0.3782	***0.0275***	0.3720	***0.0456***	0.2950	***0.0744***
Gamma	0.4298	0.2397	0.5350	0.2128	0.5015	0.1899	0.3388	0.1533	0.1350	0.1057	***0.0674***	0.0727
Weibull	0.1227	0.1064	0.1352	0.0893	0.0940	0.0771	0.1138	0.0735	0.1560	0.0527	0.1094	0.0720

Table IV. KS Distances of different statistical distributions- Differential Light Intensity Reporting (in degree Lux) for 10 weeks duration

Distributions	ΔL= 150 lux		160 lux		170 lux		180 lux		190 lux		200 lux	
	MoM	ML	MoM	ML	MoM	ML	MoM	ML	MoM	ML	MoM	ML
Exp	0.2899	0.2899	0.3015	0.3015	0.2504	0.2504	0.3221	0.3221	0.2295	0.2295	0.1379	0.1379
G.Exp	0.0632	***0.0632***	0.1209	0.1209	0.3196	0.3196	0.3829	0.3829	0.1234	0.1234	0.1052	0.1052
Pareto	0.9391	0.3588	0.9671	0.3956	0.8456	0.2980	0.9264	0.3802	0.8329	0.2907	0.8278	0.3388
G.Pareto	0.2363	0.0976	0.2402	0.0889	0.1905	0.0853	0.2561	0.0911	0.1970	0.1673	0.1294	0.1311
Log-N	0.2917	0.0801	0.2813	0.0698	0.1807	0.0744	0.2123	0.0695	0.2546	0.1633	0.2040	0.1979
Gamma	0.0664	0.0717	0.1021	0.0952	0.2915	0.0854	0.3541	0.1287	0.1162	0.0970	***0.1041***	0.1260
Weibull	0.1229	***0.0756***	0.1012	***0.0638***	0.0973	***0.0513***	0.1059	***0.0631***	***0.0961***	0.1042	***0.1053***	0.1272

distance is gradually increasing. Beyond 180 lux the KS distance under Log-Normal Distribution is more than 0.1 that is almost over 10% and inappropriate to be considered as best possible fitting distribution. Therefore, the KS distances with other distributions are investigated with ΔL ranging from 150 to 200 lux in Table IV. It is observed that at ΔL =150 lux Generalized-Exponential Distribution provides best fit with KS distance of 0.0632. From ΔL= 160 to 180 lux, Weibull Distribution provides best fit under ML estimation with KS distances of 0.638, 0.0513 and 0.631 respectively whereas at ΔL= 190 lux it is Weibull Distribution as well under MoM Estimation with KS distance equals 0.0961. At ΔL=200 lux the best fit is provided by the Gamma Distribution with KS value of 0.1041. However, at ΔL =150 and 200 lux the KS values under Weibull Distribution is 0.0750 and 0.1053 marked underlined in bold italics. Therefore, it could be easily observed that these KS distances are not too far from the best fit KS value of 0.0632 and 0.1041. Hence, ΔL ranging from 150 to 200 lux, Weibull Distribution could be considered as the best possible fit. Like Fig. 2, in Fig. 5 the shaping, scaling and the location parameters are shown for ΔL ranging from 150 to 200 lux under Weibull Distribution. The analytical expressions for λ, α and μ in terms of ΔL are determined which follows again a Generalized Gaussian expression as per (4) to (6) with the coefficients a_i, b_i and c_i given in Table VI(a), VI(b) and VI(c) for Weibull distribution. Fig. 6 shows the KS distances for the ΔL values ranging from 150 to 200 lux. It is clearly observed both the empirical and the estimated analytical KS distances are 10%.

Table V. Luminous Intensity- fitting parameter coefficients for Log-Normal Distribution

(a) λ parameter

a_i	b_i	c_i
0.4656	102.4	9.473
0.3321	115.3	5.533
0.3319	88.49	10.74
0.2029	127.6	4.575
0.5653	75.34	24.64
2.272	136.3	45.32
2.304	32.71	69.04
1.68	192.9	24.24

(b) μ parameter

a_i	b_i	c_i
38.65	260.8	36.11
-0.3388	170.1	9.096
-0.404	159	1.508
-1.532	137.4	21.44
0.5592	135.5	9.772
0.4056	122	2.315
9.566	158.9	82.39
2.248	48.84	43.33

Table VI. Luminous Intensity- fitting parameter coefficients for Weibull Distribution

(a) λ parameter

a_i	b_i	c_i
$3.261 \cdot 10^4$	1.712	0.259
8702	1.473	0.07715
$2.235 \cdot 10^4$	1.078	0.3924
9349	0.5467	0.1693
7735	0.2264	0.2643
5353	1.255	0.03825
$2.06 \cdot 10^4$	-0.6246	1.632
3070	-0.1823	0.1162

(b) α parameter

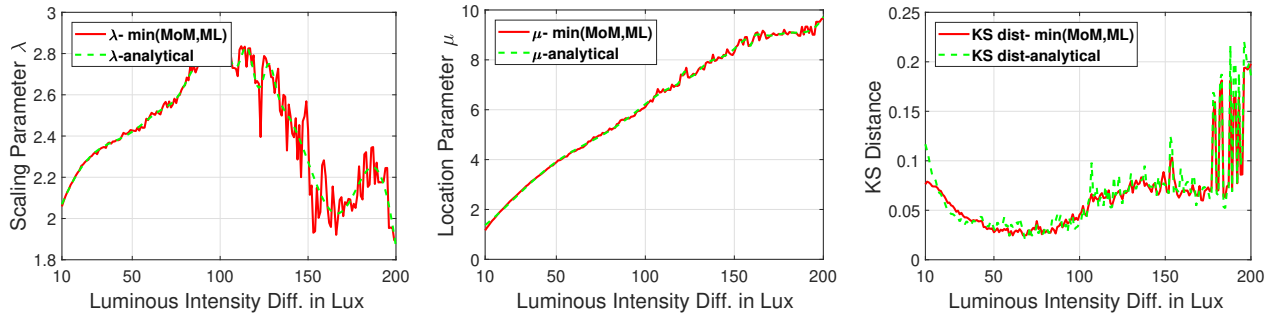
a_i	b_i	c_i
1785	1.615	0.1328
-1785	1.615	0.1328
0.5182	1.149	0.415
0.4591	-1.6	0.3643
0.5354	0.2724	0.6343
0.1521	0.526	0.07169
0.5535	-0.8251	0.6655
0	-1.001	$2.735 \cdot 10^{-5}$

(c) μ parameter

a_i	b_i	c_i
77.3	163.1	0.3403
62.54	166.6	0.587
70.8	165.1	1.036
34.99	161.6	.226
13.09	193.8	0.3571
-27.96	201.4	3.157
8.463	194.9	0.9891
$2.995 \cdot 10^{12}$	632.2	87.38

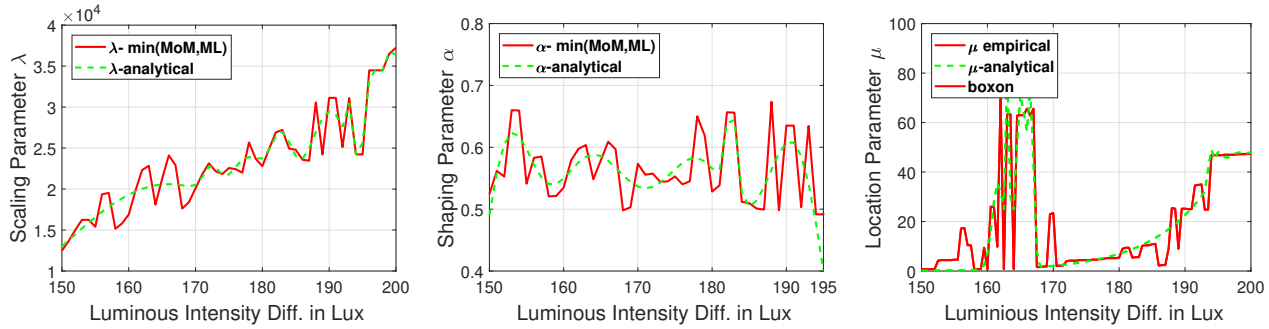
C. Motion sensors

Table No. VII, VIII and IX shows the KS distance for the involved inter-arrival times with reference to the motion sensor data. Table VII shows the inter-arrival duration between two ON- states of the motion sensor data. It follows a similar trend like the previous results. The D_{KS} for the inter-arrival time between two successive ON-states of the motion sensor does not follow exponential distribution indicating it is not a Poisson process yet again. Pareto and Generalized Pareto distributions are the best fit. Similarly, in Table VIII which is for the ON duration, the best fits are Exponential and Generalized Pareto distributions for which the KS distances are almost 5% and 6% respectively. Table IX shows the KS distance for the OFF duration of the motion sensor. It is similar to that of the two intermediate ON-



(a) (b) (c)

Fig. 4. a. Scaling Parameter (λ) vs Luminous Intensity Difference b. Location Parameter (μ) vs Luminous Intensity Difference c. Numerically and Analytically computed KS distances vs Luminous Intensity Difference, all for Log-Normal Distribution



(a) (b) (c)

Fig. 5. a. Scaling Parameter (λ) vs Luminous Intensity Difference b. Shaping Parameter (α) vs Luminous Intensity Difference c. , Location Parameter (μ) vs Luminous Intensity Difference, all for Weibull Distribution

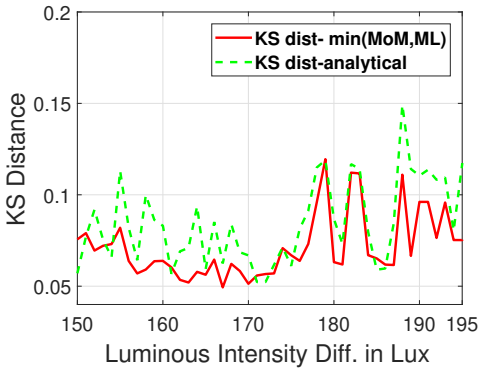


Fig. 6. Numerically and Analytically computed KS distances vs Luminous Intensity Difference, for Weibull Distribution

Table VII. KS Distances of statistical distribution of inter-arrival times between two successive ON-states of the motion sensor for 10 weeks

Distribution	MoM	ML
Exp	0.7428	0.7334
G.Exp	0.9456	0.1792
Pareto	0.1392	***0.0388***
G.Pareto	0.6807	***0.0416***
Log-N	0.2730	0.1364
Gamma	0.9404	0.3312
Weibull	0.4677	0.2296

Table VIII. KS Distances of statistical distribution of ON-durations of the motion sensor for 10 weeks

Distribution	MoM	ML
Exp	***0.0580***	***0.0580***
G.Exp	0.0951	0.0951
Pareto	0.3454	0.1792
G.Pareto	***0.0640***	***0.0634***
Log-N	0.0798	0.0817
Gamma	0.0917	0.0718
Weibull	0.0801	0.2219

states which Pareto and Generalized Pareto again proving the best fit. This is mainly due to the fact that the order of duration between two ON states and

the OFF states are similar. Furthermore, in our observation the time duration of the ON period or the period of activity is much smaller as compared to the period of inactivity. The Fig. 7 shows the fitted and empirical CDFs for the two successive ON-states, the duration of ON-states and the duration of OFF-states. In Fig. 7(a) and Fig. 7(c) since the time duration between the initiation of two intermediate ON-duration and the OFF-duration is high, therefore the x-axis is shown in the Logarithmic scale. Table X shows the parameter values of the best fitting distributions for the length between two successive ON

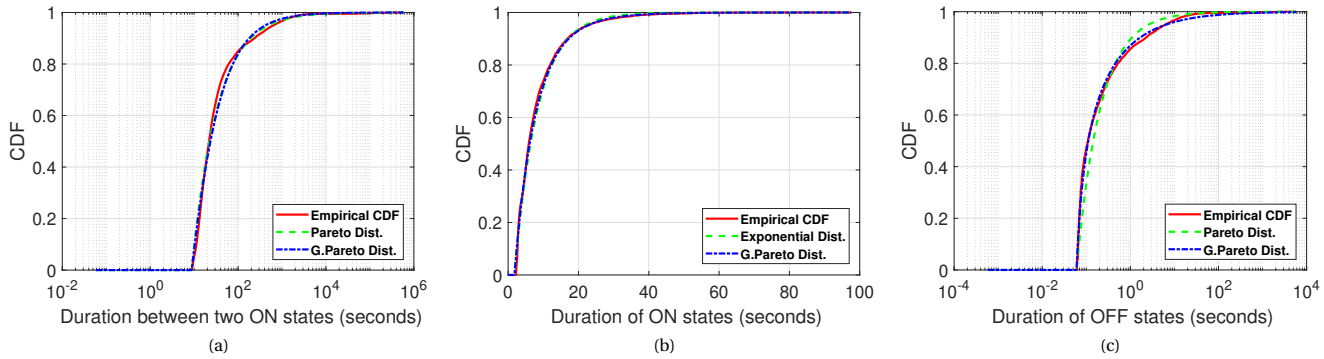


Fig. 7. Fitted and empirical CDFs for: (a) two successive ON-states, (b) duration of ON-states, (c) duration of OFF-states.

Table IX. KS Distances of statistical distribution of OFF-durations of the motion sensor for 10 weeks

Distribution	MoM	ML
Exp	0.7488	0.7454
G.Exp	0.9526	0.1631
Pareto	0.1157	***0.0456***
G.Pareto	0.6924	***0.0164***
Log-N	0.3138	0.0676
Gamma	0.9482	0.2940
Weibull	0.3953	0.1757

Table X. Best Fitting Distributions and their parameters for motion sensors data

Time Duration	**Best-fit Dist.**	λ	α	μ
Between two ON	Pareto	8.8	0.7442	−−
ON	Exponential	0.1510	−−	1.8
OFF	Gen. Pareto	0.0345	1.9692	0.0600

durations, length of the ON-duration and length of OFF duration along with their parameter values.

V. Conclusion

Based on our findings from the real time data collected for over 10 weeks we have thus found some counter intuitive results. The traditional idea of assuming any data packet generation process as Exponential or Poisson distributed is not always correct. For change in Temperature parameter Weibull Distribution provides the best fit for values ranging from 0.1 to 5 °C. In case of the change in Luminous Intensity parameter, from 10 to 140 lux Log-Normal Distribution provided the best fit while from 150 to 200 lux it is Weibull Distribution. The shaping, scaling and the location parameters of the distribution which best fits the empirical CDF are expressed as a function of the input parameter. This will be helpful to emulate a real time IoT traffic generation setup which could closely track the inter-arrival periods of data packet generation and can be extensively used for optimization of the next generation mobile network architecture where IoT and M2M communication will be an integral part.

VI. Acknowledgement

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).

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] 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.
[6] 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.
[7] 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. Dolher and C. Antón-Haro, Woodhead Publishing, 2015.
[8] 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.
[9] J. Gozalvez,"New 3GPP Standard for IoT [Mobile Radio],"*IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 14-20, Mar. 2016.
[10] A. Hoglund et al.,"Overview of 3GPP Release 14 Enhanced NB-IoT,"*IEEE Netw.*, vol. 31, no. 6, pp. 16-22, Dec. 2017.
[11] 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.
[12] 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.