Noname manuscript No.
(will be inserted by the editor)

Does the Assumption of Exponential Arrival Distributions in Wireless Sensor Networks Hold?

Krishna Doddapaneni · Ali Tasiran · Fredrick A Omondi · Enver Ever · Purav Shah · Leonardo Mostarda · Orhan Gemikonakli

Received: date / Accepted: date

Abstract Wireless Sensor Networks have seen a tremendous growth in various application areas despite prominent performance and availability challenges. One of the common configurations to prolong the lifetime and deal with the path loss phenomena is having a multi-hop set-up with clusters and cluster heads to relay the information. Although researchers continue to address these challenges, the type of distributions for arrivals at the cluster head and intermediary routing nodes is still an interesting area of investigation. The general practice in published works is to compare an empirical exponential arrival distribution of wireless sensor networks with a theoretical exponential distribution in a Q-Q plot diagram. In this paper, we show that such comparisons based on simple eye checks are not sufficient since, in many cases, incorrect conclusions may be drawn from such plots. After estimating the Maximum Likelihood parameters of empirical distributions, we generate theoretical distributions based on the estimated parameters. By conducting Kolmogorov-Smirnov test statistics for each generated inter-arrival time distributions, we find out, if it is possible to represent the traffic into the cluster head by using theoretical distribution. Empirical exponential arrival distribution assumption of wireless sensor networks holds only for a few cases. There are both theoretically known such as Gamma, Log-normal and Mixed Log-Normal of arrival distributions and theoretically unknown such as non-Exponential and Mixed cases of arrival in wireless sensor networks. The work is further extended to understand the effect of delay on inter-arrival time distributions based on the type of medium access control used in wireless sensor networks.

K.Doddapaneni, A.Tasiran, F.A.Omondi, P.Shah, L.Mostarda, O.Gemikonakli,
School of Science and Technology
Middlesex University
London, UK
E-mail: k.doddapaneni@mdx.ac.uk

E.Ever,
Computer Engineering Program
Middle East Technical University
Northern Cyprus Campus, Kalkanlı, Güzelyurt, Mersin 10, Turkey
E-mail: eever@metu.edu.tr

Keywords Wireless Sensor Networks, Performance, Maximum-Likelihood Estimates of Empirical Distributions, Exponential Distribution, Non-Exponential Empirical Distributions, Gamma, Log-normal Distributions, Mixed Log-Normal Distributions, Histograms and Empirical Densities, Q-Q Plots, P-values, Kolmogorov-Smirnov Test Statistics, Theoretical and Empirical Densities and Cumulative Distribution Functions

1 Introduction

Wireless Sensor Networks (WSNs) rely on cooperative effort of the densely deployed sensor nodes to gather information from the habitat [1–3] typically to achieve either environmental monitoring or target tracking and sensing. Depending on the area of application, information monitoring and reporting may further be classified as continuous, periodic, or event-based (driven) [1, 4]. An example may be temperature monitoring where the first case involves reading and reporting periodically irrespective of the changes involved, the second scenario may be where only variations from previous readings are reported and finally, the case where a report is sent only when a specific temperature is reached. In all these cases, data arrival delay is clearly determined by the nature of application and the chosen monitoring scheme.

Apart from the common challenges of WSNs including energy consumption, network connectivity, data aggregation, computation power, limited sensor node memory, the end to end delay of transmitted packets remains a serious concern in relation to Quality of Service (QoS) provision [3, 5]. In [4], cross layer analysis of the end to end delay distribution in WSNs was studied and the results show that inter-arrival time (time between two consecutive arrivals) mostly follow exponential distribution except for low periodic traffic. There are many studies which consider exponential arrivals to sensor nodes [6–10]. However, in other quarters there has been mixed opinions on the appropriate distribution for modelling inter arrival delay of WSN data packets [2, 4, 11]. In other works, there has been mixed opinions on the appropriate distribution for modelling inter-arrival time of WSN data packets [2]. This strongly indicates the the need for a study to identify acceptable types of distributions for inter-arrival times used in modelling WSNs.

Characterization of the end-to-end delay distribution is fundamental for real-time communication applications with probabilistic QoS guarantees. Indeed, the cumulative distribution function (CDF) of the delay for a given deadline can be used as a probabilistic metric for reliability and timeliness [4]. Researchers have also continued to develop algorithms and protocols to address some of the challenges like balancing cluster energy consumption in clustered WSNs as well as path loss effects [12, 13].

In this paper, an investigation is carried out to establish the most appropriate distribution for the inter-arrival times at Cluster Heads (CH) and relay nodes. The process is started by identifying and characterizing various applications and determining suitable data delivery models depending on application requirements. Simulation results are presented and analysed in detail to characterize end to end delay between arriving data packets. The effects caused by medium access control (MAC) protocol properties are also analysed by experimenting with well known MAC protocols.

Most existing WSN simulators assume that exponential distribution is valid for characterising arrivals of data packets at nodes within the WSN. To the best of the our knowledge, this is the first work that provides statistical substantiation of the results along with probabilistic analysis of arrival distributions at the CH or relay nodes in WSNs.

Kolmogorov-Smirnov (K-S) test statistics are used to decide whether a certain type of distribution function assumption is appropriate for inter-arrival time distribution. The rest of the paper is organised as follows: Related work in this area is summarized in Section 2. Section 3 discusses the data delivery models that are characterised based on the application requirement along with the related performance aspects. A detailed description of the communication paradigm considered in this work is presented in Section 4, followed by the system's detailed discussions on inter-arrival distributions along with various aspects of the case studies like the effects of MAC, data rates and application types are provided in Section 5. Simulation results for inter-arrival distributions at the CH are presented along with their equivalence using statistical studies and further probabilistic analysis are presented in Section 6. Finally, Section 7 concludes the paper with detailed explanation about various distributions in Appendices A and B.

2 Literature Survey

Performance modelling and analysis continues to be of great importance in supporting research as well as in the design, development and optimization of WSN and their applications. The current trend towards the use of WSNs for sensing and control now has the potential for significant advances, not only in science and engineering, but also, on a broad range of applications. This brings the need for performance modelling for the optimization of deployment of WSNs. However, the special design, characteristics of sensors and their applications separate them from the traditional networks. These characteristics pose great challenges for the architecture, protocol design, performance modelling and their implementation. It is essential to consider energy efficiency of WSNs because of their limited energy sources (most of the times batteries). In order to minimise the energy consumption, one of the effective techniques is to place sensors in sleep mode during the idle period [14]. In [15–17], a wake-up scheduling scheme at the MAC layer is proposed, which wakes up the sleeping nodes when there is a need to transmit or receive, thus avoiding a degradation in network connectivity or quality of service provisioning.

Characterising delay in distributed systems has been considered in various contexts. However, it can be observed that accurately characterizing end-to-end delay at the CH is still an open problem. Considerable amount of research on sensor networks reported recently has been ranging from network capacity and signal processing techniques, to topology management, algorithms for traffic routing and channel access control. The model presented in [2] is used to investigate system performance in terms of energy consumption, network capacity, delay in data delivery along with the trade-off's that exist between performance metrics and sensor dynamics in active/sleep modes. A Markov model is presented for WSNs, where the nodes may enter into sleep mode. Through standard Markovian techniques, a system model representing the behaviour of a single sensor has been constructed along with the dynamics of the entire network, and the channel contention among interfering sensors. The proposed solution of the system model is then obtained by means of a Fixed Point Approximation (FPA) procedure, and the model has been validated via simulation.

Due to hardware constraints for energy efficiency, optimizing node packet buffer and maximizing the performance is necessary to improve the Quality of Service(QoS) for transmission in WSNs. In [18], a packet buffer evaluation method using queuing network models is proposed where, the blocking probabilities and system performance indicators of each node are calculated using an approximate iterative algorithm. The model considered focuses on a single server model in WSNs and the method used to calculate packet buffer capacity for nodes also indicate that the sink node requires higher performance, when compared to the other nodes in the network. The Markov model of the sensor sleep/active dynamics is presented in [19], that predicts the sensor energy consumption by acquiring this information for each sensor, while a central controller constructs the network energy map representing the energy reserves available in various parts of the system. Only a single node is represented by a Markov chain, while the network energy status is derived with the help of simulation studies.

With regard to analytical studies, results on the capacity of large stationary ad-hoc networks are presented in [20]. Two network scenarios were considered; one including arbitrarily located nodes and traffic patterns, while the other one with randomly located nodes and traffic patterns. An analytical approach on network coverage and connectivity of sensor grids is presented in [21]. The sensors are considered unreliable and fail with a certain probability leading to random grid networks. Results on coverage and connectivity are derived as functions of key parameters such as the number of nodes and their transmission radius.

Several approaches based on simulations and experiments, have been proposed for performance evaluation of IEEE 802.15.4 networks [22]. In [23], an analytical framework based on a Markov chain characterization of the MAC protocol is proposed for IEEE 802.11 networks in saturation conditions. Based on this pioneering work, several approaches have been proposed for the characterization of the MAC performance in IEEE 802.15.4 networks with a star topology. In this work, a scenario with acknowledgement (ACK) messages is considered and an evaluation of the network performance in both saturation and non-saturation regimes is presented, while trying to characterize the conditions under which the network enters the saturation region [24]. A simple Markov chain theoretical model to characterize the sensors as well as the channel status is proposed in [25]. The models shows good agreement with ns-2 based simulations. This model allows to investigate throughput and energy consumption metrics within WSNs. In [26], an extended framework of the one proposed by [25] is presented for a 2-hop network scenario, i.e., networks where sensors communicate with the coordinator through an intermediate relay node, which forwards data packets from the sources (the sensors) towards the destination (the coordinator). Similar works have been presented in [27, 28], emphasising the use of a relay for interconnecting two different clusters in IEEE 802.15.4 networks and analysing the performance through a queueing theoretical analysis. However, the proposed scenario models the (simpler) cases where the relay does not content the medium access to the sensors. Hence, it is observed that accurately characterizing arrivals at the cluster head in WSNs is still an open problem. Although it is quite difficult to analyse each possible application in WSNs, it is sufficient to analyse each class of application classified by data delivery models, as most of these applications in each class have common requirements on the network [29]. A well established simulation tool Castalia which provides realistic node behaviour, wireless channel and radio models, and enables to mimic and analyse the real life scenarios for various types of applications is employed in this study.

3 Characterising Data Delivery Models

From the point of view of network QoS, the network is concerned with how to transmit the sensed data from the sensor field to the sink node, fulfilling the corresponding required QoS. The factors that characterize the application requirement are presented in Table 1. The practical realization of the current WSN applications depends on the energy-efficient, real-time and reliable communication capabilities of WSN. WSNs have distinct traffic characteristics. The primary traffic is generally a many to one type communication, i.e., from the sensor nodes to the base station, in the upstream direction. Upstream traffic delivery can be classified as: continuous, event driven, query driven and hybrid-based data delivery models. Depending on their specific applications, these data delivery models have different QoS and reliability requirements [30]. These data delivery models greatly influence the protocol design and affects the performance of WSNs. The four models and their related performance aspects are discussed below:

Table 1: Application Requirements of Data-Delivery Models

Factor	Event-Driven	Query-Driven	Continuous	Hybrid
Interactivity	✓	✓	✗	✓
End-to-End Performance	✗	✗	✗	✗
Delay Tolerance	✗	Query-specific	✓	✗
Criticality	✓	✓	✓	✓

Event-driven Delivery Model - most event-driven applications in WSNs are interactive, delay intolerant (real-time), mission critical and non-end-to-end applications. When an event occurs, the sensor node begins to report the event, and possibly an associated value, to the sink. The application needs to receive the desired data reliably and as quickly as possible. The query-driven data delivery model is very similar to the event-driven model, except that the data are pulled by the sink where as in event-driven models, the data are pushed to the sink. The application in most of the cases may not be an end to end one, i.e., one end of the application is the sink, where as in the other end, a group of sensor nodes within the area that are influenced by the event. Also, the traffic generated by a single sensor node may be of very low intensity, however, more random and unforeseeable bursty traffic may be generated by a set of sensors due to the occurring event. Therefore, the routing efficiency for this delivery model is heavily dependent on the frequency of occurrence of the events. CSMA medium access arbitration is a good fit for event-based data delivery models since the data are generated sporadically.

Query-driven Delivery Model - similar to event-driven applications, most query-driven applications in WSNs are also interactive, mission critical, query-specific, delay tolerant and non-end-to-end applications. Based on the application considered, queries can be sent on demand in order to save energy. Sometimes, the base station may be interested in a specific piece of information that has already been collected by the sensor nodes. The sensor only reports the observed data in response to an explicit request from the user. In this delivery model, the sink broadcasts the query message, a path is constructed automatically when the query arrives at the sensor nodes, and the sensor nodes report their findings according to the request in the query message. Query-driven systems store gathered information locally and communicate it on request. This type of sensor network can be useful in logistics or home applications, but is not very common in applications of environmental monitoring.

Continuous Delivery Model - each sensor reports regularly, perhaps continuously or periodically to the sink at a specified rate. Some networks apply a hybrid model using a combination of continuous, event-driven and query-driven data delivery. Time based medium access control protocols can be used to achieve significant energy savings in case of continuous data delivery models.

Hybrid Delivery Model - some networks apply a hybrid model using a combination of continuous, event-driven and/or query-driven data delivery, as the types of sensors and the data they sense may be very diverse. For example, data may be reported continuously by some nodes, and the sink may need to query information from other sensor nodes.

4 System Communication Paradigm

A system of Wireless Sensor Network with identical sensor nodes deployed in a cluster tree topology is considered. The sensor nodes used are assumed to self-configure during initial deployment and remain stationary thereafter. All the nodes in a cluster and adjacent CHs are considered directly connected to the CH. The primary focus is to study the inter-arrival distribution of packets at the CH. The total arriving data packets at the CH at any given time is therefore equal to the sum of all the independent arrivals from the cluster nodes and arrivals from adjacent CHs forwarding their data to the sink. For this case continuous monitoring of event driven systems are considered.

In this set up all nodes are considered to be equipped with an omnidirectional antenna and they also have a common maximum radio range r within which they are able sense event occurrences and also transmit information to the CH based on the

802.15.4 standards. The topology of interest is shown in Figure 1. For simplicity, all sensor nodes are shown connected directly to the CH0 in Figure 1. CH0 can forward data to the sink either through CH1 or CH4, whereas CH2 and CH3 forwards their packets to the sink passing through CH0. It is also shown that nodes N1 to N8 are directly connected to the CH0.

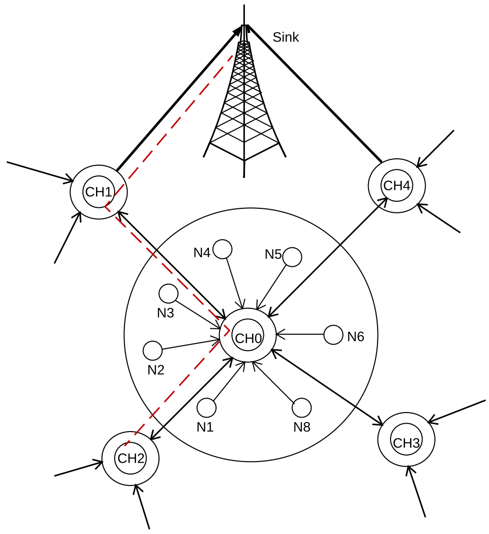

Fig. 1: Network topology of the reference scenario

Each sensor node is able to independently monitor its habitat and organise the information sensed into fixed data units storable at the sensor buffer before finally forwarding to the CH. The buffers, both at the sensor nodes and at the CH are assumed to have infinite capacity and are follows First in First out (FIFO) queuing discipline. The Cluster Head is only able to receive or transmit at one go within the assigned time slots of unit duration. Once Information sensed and aggregated at the nodes are forwarded to the CH, it finalizes cluster aggregation and transmits all the information to the sink either directly or through other intermediary CHs. It is assumed that at least one path always exists towards the sink [2].

In this study continuous monitoring applications where the nodes periodically (deterministic) sense and transmit information are considered for various MAC protocols, in order to see the effects of MAC protocols on the distribution of arrival process for the CHs. Castalia simulation environment is employed in order to analyse the inter-arrival distribution at the CH. For each experiment, packet arrival rate and number of nodes is set at desired values. Desired MAC properties; TMAC, CSMA, and no MAC(no MAC protocol is applied) are then considered for each experiment. The generated inter-arrival distribution time results are then further analysed using statistical tools to identify the actual distribution pattern.

5 Detailed Analysis of Case study and Simulations

5.1 Inter-Arrival Distributions

Providing QoS guarantees in terms of delay, jitter, and throughput has been the main focus of researchers, as the connectivity between different domains improves, also improving the overall performance of the network. End-to-end QoS guarantees are complicated by the inherent differences in the nature of the wireless media. Therefore, providing QoS guarantees in a network, in general, requires sophisticated traffic management and admission control procedures. This requirement is even more important in networks of low-power, low-data-rate sensor nodes, where network resources are scarce and dynamic. Considering the non-deterministic nature of communication due to wireless channel errors and traffic characteristics, probabilistic analysis of network performance is crucial to provide QoS guarantees.

One of the most important metric of QoS is the probability distribution of inter-arrival times of packets in WSNs. In order to characterise the distribution of packet inter-arrival times, the number of arrivals are considered from the numerical results provided by Castalia. In a typical cluster network, the inter-arrival time is characterised by the following: the resulting job arrivals at the CHs is a collection of jobs from locally generated packets and relay packets from other neighbouring CHs. Locally generated packets consists of the sensed information by the CH itself and from other cluster nodes in the clusters. We carry out investigation to establish the most appropriate distributions suitable for modelling the inter-arrival distributions of these packets at the CH. The inter-arrival time of the packets received by the CH depends on the application requirements, with which the sensor data are generated. The generated traffic mainly depends on the physical phenomenon of interest and the type of application, while the relay traffic depends on the network parameters. For evaluation purposes, a clustered network is considered where the inter-arrival distribution is found for the CH under the contention from the cluster nodes. The distribution of the inter-arrival time of the packets is recorded at the CH. Each of the nodes are interrelated according to the traffic constraints. Each cluster node transmits its generated packets to the corresponding CH, where the CH aggregates the packets received from its cluster nodes, along with its own generated packets and relay packets from other neighbouring CHs and forwards them to the next CH on route or directly to the sink. In other words, the sum of the incoming relay traffic rate at each CH is equal to the transmitted traffic rate from each of the cluster node.

5.2 Event-driven and Continuous-monitoring Applications

The arrival time of the generated packets from each sensor node sent to the CH depends on the application requirements, from which the sensor data are accordingly generated. Depending on the type of application, i.e., in case of event-based applications, the sensor node begins to report the event and possibly an associated value to the CH or to the sink (if the node is a CH itself), when an event occurs. In such cases, the data generated are often sporadic. Considering such physical events, e.g., fire alarm system, temperature sensing systems etc., the event being monitored do not occur very frequently, i.e., occurring at irregular intervals in time. Extensive work has been already carried out in estimating the distribution of inter-arrival time of the packets at each node, considering physical events that do not occur very often. In [4], it was shown that the probability of any event occurring at any time is governed by a Poisson's process, and the inter-arrival times are exponentially distributed. Query-driven applications are also very similar to event-driven applications in terms of arrival time of the generated packets from each sensor node to the CH. This is because they also depend on the application requirements.

In applications involving the source sensors sending their sensed data continuously to the sink, for example, in a temperature-sensing systems, the sensors send their data to the cluster head/sink in a continuous manner throughout the time, at a specified rate. The deployment at Great Duck Island [31] is an example of a continuous monitoring network, where the nodes are capturing the movement of Petrels once every 5 to 10 minutes. The class of continuous data delivery model can be further classified, depending on the data rate of operation. Although, WSNs are usually considered as very low data rate networks, there is a great potential to utilize the benefits of WSNs for high data rate and low delay demanding applications, such as media streaming and critical control. Examples of low data rate sensors include temperature, humidity, and peak strain captured passively whereas, examples of high data rate sensors include strain, acceleration, and vibration sensors.

5.3 Effects of MAC

Channel contention plays an important role in causing additional delay, queuing delay and wireless channel errors at the CH due to the job arrivals from cluster nodes and forwarded data from other CHs. The MAC layer is responsible for scheduling and allocation of the shared wireless channel which eventually determines the link level QoS parameters, namely MAC delay. MAC protocols provide the greatest influence over communication mechanisms and provide the most direct influence over utilization of the transceiver, as transceiver that constantly senses the channel will quickly deplete the sensor node energy resources and shorten the network lifetime to unacceptable levels. The main design goal of a typical MAC protocols is to provide high throughput and QoS. On the other hand, wireless sensor MAC protocol gives higher priority to minimize the energy consumption rather than the QoS requirements. Hence, characterization of inter-arrival distribution is fundamental and can be used as a probabilistic metric to estimate the QoS in WSNs. Channel contention is a serious problem in WSNs resulting in collisions, re-transmissions, energy depletion, and ultimately loss of event reports. MAC protocols employ a back-off algorithm to resolve contention among nodes to acquire channel access. Most common contention-based MAC protocols can be employed such as CSMA or T-MAC for transmissions to keep the energy consumption low, reducing the amount of energy wasted on idle listening, in which nodes wait for potentially incoming messages, while still maintaining a reasonable throughput [32].

Majority of the WSN MAC protocols are contention-based, wherein the contention window size setting involves an important trade-off between the collision probability and idle listening durations in contentions where both are aimed to be lowered for efficient network operation. Sensor network MAC protocols often trade performance characteristics, such as throughput and latency, for a decrease in energy consumption to lengthen a sensor node's lifetime. The key challenge of supporting real-time data transmission in CSMA-based model is the non-deterministic nature of delay for a successful transmission of a data packet. CSMA/CA is a contention-based technique where the node needs to sense whether the channel is idle before it can transmit a packet. If the channel is not idle at that time, the node needs to wait for a certain period of time before it can sense the channel again. This scheme makes the delay time for a successful transmission non-deterministic. If there is a duty cycle in place (for example, like in TMAC) then whenever the node back-off's, it also goes to sleep. When it is time to wake up there is an extra delay before the node can sense the channel. The nodes (excluding the sink) turn off their radio periodically to save energy. When any node has a packet to send, it starts to repeatedly transmit request to send (RTS) beacon packets based on CSMA/CA manner, i.e. through carrier sense and random back-off manner, and therefore causing delays. This extra delay slightly increases the probability of collisions, but on the other hand, a node can save considerable energy (especially in heavy traffic where it is backing off often) [32]. Some parts of the delay can be governed by equations, such as the transmission time of a packet by the radio which obviously depends on the data rate of the radio and the size of the packet. But most of the delay happens because of the MAC protocol. An example could be that the MAC layer is waiting for the channel to be clear or waiting for the active period to commence. Since retransmissions involve the MAC each time, then most of the delay of a retransmission packet will be due to the MAC as well. Hence, buffering the packet for retransmission, will cause unnecessary delay, affecting the performance of the network. The delay distribution models presented in the literature do not consider the uncertainties due to random back-off's because of the MAC protocols. Therefore, it is quite an important task to characterise the inter-arrival distributions at the CH especially when considering the affects of MAC protocols, causing delays.

5.4 Case Study and Simulation Parameters

Most of the applications in WSNs have common requirements on the network. Therefore, in order to analyse possible applications a typical clustered network scenario where nodes transmit sensed data to the CH is considered. Importantly, in order to characterise the distribution of inter-arrival time of packets, case studies based on typical scenarios are considered. The real contribution here is to provide statistical substantiation of the results along with probabilistic analysis of statistical distributions of the arrival distributions at the CH and relay nodes in WSNs.

Simulation results are obtained with simulation package Castalia, the WSN framework of OMNET++. It is mainly used for initial testing of protocols and/or algorithms with realistic node behaviour, wireless channel and radio models. The

OMNET++ platform is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators. Castalia is highly tunable, features an accurate radio model based on the work of the authors in [33]. It also features physical process model, considering clock drift, sensor energy consumption, CPU energy consumption, sensor bias etc. Specific details related to unpredictability of the wireless channel, energy spent in transmission/receiving packets, performance degradation experienced by duty cycles, collisions are well established in Castalia [34].

In order to characterise the distribution of packet inter-arrival times at the CH, a typical scenario in WSN applications considering constant transmissions from nodes to CH, having cluster networks of various sizes (from 10 nodes to 40 nodes) are considered. The following parameters are used throughout the simulations, unless otherwise stated. A CC2420 chip, compatible with 802.15.4, is used to provide wireless communication, operating at 2.4 GHz and providing a data rate of 250 kbps. For TMAC and CSMA, the internal MAC buffer size in packets is 32. The packet size is considered to be 105 bytes [35]. The simulation scenarios are chosen from the examples of prototyped applications for WSNs presented in the literature [36]. Although it is quite difficult to analyse each possible application in WSNs, it is sufficient to analyse each class of application classified by data delivery models, as most of these applications have common requirements [29]. R software environment for statistical computing and graphics is used for K-S test and statistical evaluation.

6 Numerical Results and Discussions

In table 14 below, we report the results of finding theoretical distributions to the empirical arrival distributions of simulated data series at the CH and the intermediary routing nodes. The first column presents the number of observations in the simulated series. The second column displays estimated Maximum Likelihood parameters of empirical distributions [1].

The well-known theoretical distributions corresponding properly to the empirical distributions of the simulated data series are Exponential, Gamma, Log-Normal and Mixed Log-Normal distributions. The detailed information about these distributions can be found in Appendix A.

Columns three and four report the K-S Test Statistics and their P-Values. Although we display Q-Q plots to compare empirical distribution to theoretical distributions whether these two population distributions are exactly the same, we also

[1] When the joint density for a set of variables is viewed as a function of the parameters alone, that function is called a *Likelihood function*. Hence the Likelihood function, $L(\theta)$, is defined as $L(\theta) = f_\theta(x)$. Here $\log f_\theta(x)$ is a scalar function of a k-dimensional variable θ and $x = (x_1, x_2, \ldots, x_n)$. A value of the parameter θ that maximizes $L(\theta)$ is called a maximum likelihood estimator (MLE), and is denoted by θ_{ML}. It is often easier to maximize the log-likelihood function, $\log L(\theta)$, and since the (natural) logarithmic function is monotonically increasing in θ, the same value of θ_{ML} maximizes both $L(\theta)$ and $\log L(\theta)$. Under quite general conditions, MLEs have a number of favourable properties. Consistency: Under mild conditions, MLEs converge to the true parameter value as the sample size increases. Asymptotic Normality: As the sample size increases, the distribution of the MLE approaches that of a (potentially) Multivariate Normal variables.

conduct a statistical test to prove it. Checking by eye, the quantiles for the first distribution versus the quantiles for the second distribution will fall on the $0-1$ line of the Q-Q plots can be insufficient. It can be both difficult and subjective to decide how differences between distributions will yield various kinds of deviations from a straight line. Appendix B presents details about the probability plots or Q-Q plots.

K-S Test Statistics belong to the goodness of fit tests which indicate whether or not it is reasonable to assume that a random sample comes from a specific distribution. They are a form of hypothesis testing where the null and alternative hypotheses are:

– H_0: the data follow a specified distribution
– H_A: the data do not follow the specified distribution

The K-S test is used to decide if a sample comes from a population with a specific distribution. It can be applied both for discrete (count) data and continuous binned and both for continuous variables. It is based on a comparison between the empirical distribution function (ECDF) and the theoretical one that is the upper extreme among absolute value differences between ECDF and the theoretical CDF.

The hypothesis regarding the distributional form is rejected if the K-S Test Statistic, $KSTS$, is greater than the critical value obtained from a table, or, which is the same, if the P-value is lower than the significance level.

For example in Table 2 for 10 nodes and employing no MAC protocol, the K-S Test Statistic, $KSTS = 0.09$, P-value = 0.13 alternative hypothesis is two sided. Also, wta represents the waiting time of arrivals, while wtf represents waiting time of first part of arrivals and wts represent waiting time of the second part of arrivals. These values are obtained as means of $KSTS$ values and P-values of 87 runs starting from the Lower Confidence Level value of the estimated rate parameter of Exponential distribution to the Upper Confidence Values. It means that we cannot reject null hypothesis that the data follow an Exponential distribution because the P-value is enough higher than significance levels usually referred in statistical literature.

Table 2: Distribution of Inter-Arrival times, for 10 nodes with no MAC protocol applied, sending 1 packet every 5 minutes; corresponding Figures 2, 3, 4, 5

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:246 Used:214	Exponential rate = 16.11 LCL = 14.03 UCL = 18.03	Average of 87 runs (from LCL to UCL):0.09	Average of 87 runs (from LCL to UCL):0.13	Exponential

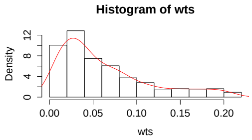

Fig. 2: Histogram of inter-arrival times

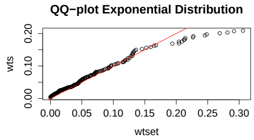

Fig. 3: QQ-plot for Exponential Distribution

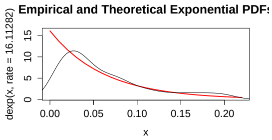

Fig. 4: Empirical and Theoretical Exponential PDF

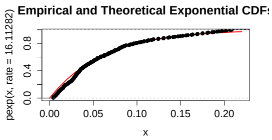

Fig. 5: Empirical and Theoretical Exponential CDF

Table 3: Distribution of Inter-Arrival times, for 10 nodes with TMAC, sending 1 packet every 5 minutes; corresponding Figures 6, 7, 8, 9, 10, 11

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:279 Used:249 First part: 224 Second part: 23	Mixed Log-Normal Meanlog1 = -5.14 Sdlog1 = 0.23 Meanlog2 = -0.52 Sdlog2 = 0.02 Mixing proportion: 0.09	Average of 100 runs :0.11	Average of 100 runs :0.15	Mixed Log-Normal

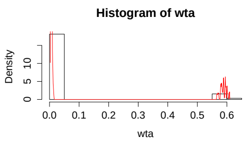

Fig. 6: Histogram of inter-arrival times

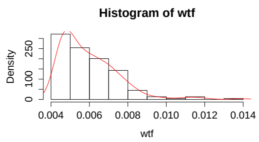

Fig. 7: Histogram of inter-arrival times of first part

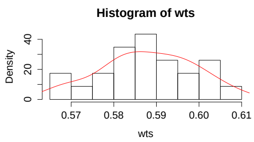

Fig. 8: Histogram of inter-arrival times of second part

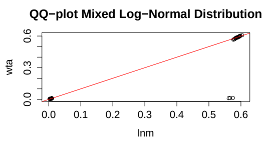

Fig. 9: QQ-plots of Mixed Log-Normal Distribution

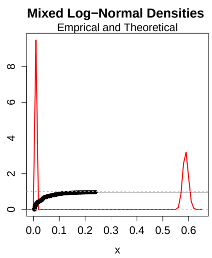

Fig. 10: Empirical and Theoretical Mixed Log-Normal PDF

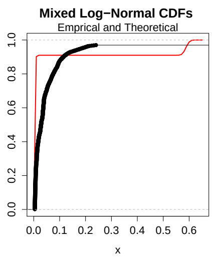

Fig. 11: Empirical and Theoretical Mixed Log-Normal CDF

Table 4: Distribution of Inter-Arrival times, for 10 nodes with CSMA, sending 1 packet/10 minutes; corresponding Figures 12, 13, 14, 15

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:255 Used:255	Exponential Rate = 20.86 LCL = 18.38 UCL = 23.50	Average of 141 runs :0.04	Average of 141 runs :0.45	Exponential

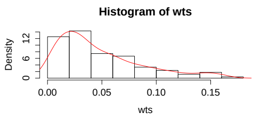

Fig. 12: Histogram of inter-arrival times

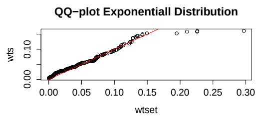

Fig. 13: QQ-plot for Exponential Distribution

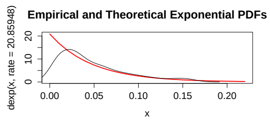

Fig. 14: Empirical and Theoretical Exponential PDF

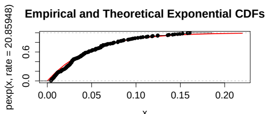

Fig. 15: Empirical and Theoretical Exponential CDF

Table 5: Distribution of Inter-Arrival times, 20 nodes without MAC, sending 1 packet every 5 minutes; corresponding Figures 16, 17, 18, 19

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:443 Used:411	Gamma shape = 1.49 scale = 0.03	Average of 100 runs :0.08	Average of 100 runs :0.24	Gamma

Table 6: Distribution of Inter-Arrival times for 20 nodes, with TMAC, sending 1 packet every 5 minutes; corresponding Figures 20, 21, 22, 23, 24, 25, 26

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:574 Used:542 First part: 508 Second part: 34	Mixed Log-Normal Meanlog1 = -5.18 Sdlog1 = 0.25 Meanlog2 = -0.58 Sdlog2 = 0.04 Mixing proportion: 0.06	Average of 100 runs :0.16	Average of 100 runs : $7.66 * e^{-06}$	An unknown Mixed distribution

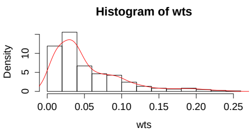

Fig. 16: Histogram of inter-arrival times

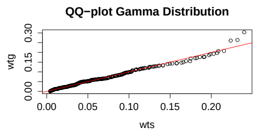

Fig. 17: QQ-plot for Gamma Distribution

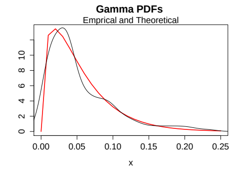

Fig. 18: Empirical and Theoretical Gamma PDF

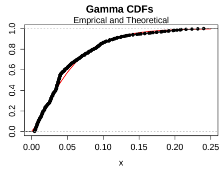

Fig. 19: Empirical and Theoretical Gamma CDF

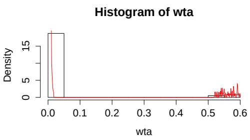

Fig. 20: Histogram of inter-arrival times

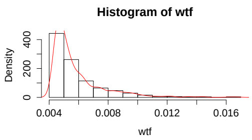

Fig. 21: Histogram of inter-arrival times first part

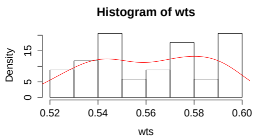

Fig. 22: Histogram of inter-arrival times second part

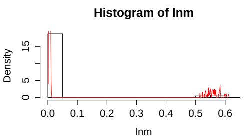

Fig. 23: Histogram of log-normal distribution

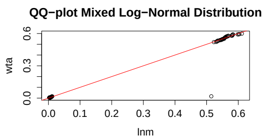

Fig. 24: QQ-plot of Mixed Log-Normal Distribution

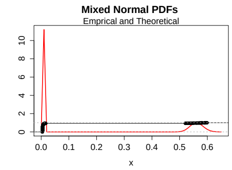

Fig. 25: Empirical and Theoretical Mixed Log-Normal Densities

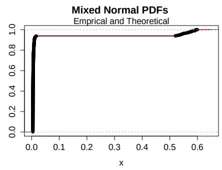

Fig. 26: Empirical and Theoretical Mixed Log-Normal CDF

Table 7: Distribution of Inter-Arrival times, for 10 nodes with CSMA, sending 1 packet every 5 seconds; corresponding Figures 27, 28, 29, 30, 31, 32, 33, 34

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:596 Used:596 First part: 578 Second part: 18	Mixed Log-Normal Meanlog1 = -3.78 Sdlog1 = 1.07 Meanlog2 = 1.59 Sdlog2 = 0.008 Mixing proportion: 0.03	Average of 100 runs :0.09	Average of 100 runs : 0.035	Mixed Log-Normal at p-values 3.5% or less. It is not Mixed Log-Normal at traditional 5 or 10% significance levels

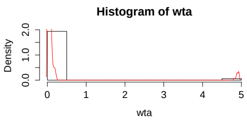

Fig. 27: Histogram of inter-arrival times

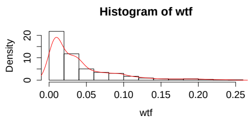

Fig. 28: Histogram of inter-arrival times first part

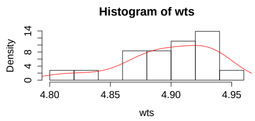

Fig. 29: Histogram of inter-arrival times second part

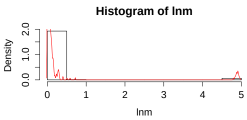

Fig. 30: Histogram of log-normal distribution

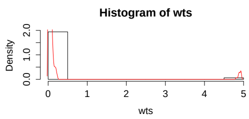

Fig. 31: QQ-plot of Mixed Log-Normal Distribution

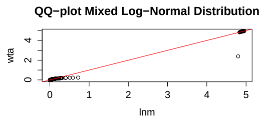

Fig. 32: QQ-plot for Mixed Log-Normal Distribution

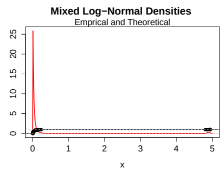

Fig. 33: Empirical and Theoretical Mixed Log-Normal Densities

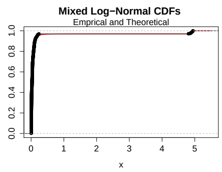

Fig. 34: Empirical and Theoretical Mixed Log-Normal CDF

Table 8: Distribution of Inter-Arrival times, for 20 nodes with CSMA, sending 1 packet every 1 second; corresponding Figures 35, 36, 37, 38

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:1889 Used:1889	Exponential rate = 16.91 LCL = 16.91 UCL = 17.69	Average of 153 runs (from LCL to UCL):0.06	Average of 153 runs (from LCL to UCL):$4.60 * e^{-05}$	An unknown non-Exponential distribution

Table 9: Distribution of Inter-Arrival times for 20 nodes, with CSMA, sending 1 packet every 5 seconds; corresponding Figures 39, 40, 41, 42, 43, 44, 45

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:2010 Used:2010 First part: 1990 Second part: 20	Mixed Log-Normal Meanlog1 = -2.94 Sdlog1 = 1.05 Meanlog2 = -0.10 Sdlog2 = $0.2 * e^{-04}$ Mixing proportion: 0.001	Average of 100 runs :0.06	Average of 100 runs : $2.62 * e^{-03}$	An unknown Mixed distribution

Table 10: Distribution of Inter-Arrival times, for 35 nodes with no MAC, sending 1 packet every 5 minutes; corresponding Figures 46, 47, 48, 49

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:644 Used:612	Log-Normal Meanlog = -3.59 Sdlog=0.93 LCL = -3.66 UCL = -3.52	Average of 148 runs (from LCL to UCL):0.08	Average of 148 runs (from LCL to UCL):0.13	Log-Normal

Table 11: Distribution of Inter-Arrival times for 35 nodes, with TMAC, sending 1 packet every 5 minutes; corresponding Figures 50, 51, 52, 53, 54, 55, 56

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:990 Used:958 First part: 914 Second part: 44	Mixed Log-Normal Meanlog1 = -5.23 Sdlog1 = 0.23 Meanlog2 = -0.62 Sdlog2 = 0.09 Mixing proportion: 0.05	Average of 100 runs :0.19	Average of 100 runs : $3.18 * e^{-14}$	An unknown Mixed distribution

Table 12: Distribution of Inter-Arrival times, for 40 nodes with no MAC, sending 1 packet every 5 minutes; corresponding Figures 57, 58, 59, 60

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:751 Used:720	Log-Normal Meanlog = -3.77 Sdlog=0.90 LCL = -3.84 UCL = -3.71	Average of 100 runs (from LCL to UCL):0.08	Average of 100 runs (from LCL to UCL):0.07	Log-Normal at p-values 6% or less. it is not Log-Normal at traditional 10% significance level

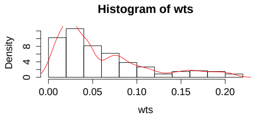

Fig. 35: Histogram of inter-arrival times

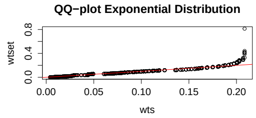

Fig. 36: QQ-plot for Exponential Distribution

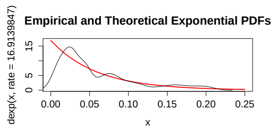

Fig. 37: Empirical and Theoretical Exponential PDF

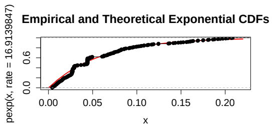

Fig. 38: Empirical and Theoretical Exponential CDF

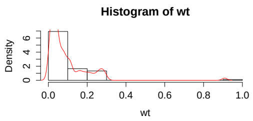

Fig. 39: Histogram of Inter arrival times

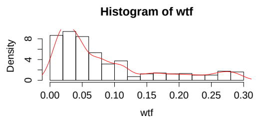

Fig. 40: Histogram of Inter arrival times first part

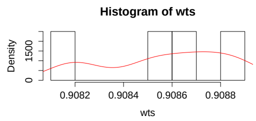

Fig. 41: Histogram of Inter arrival times second part

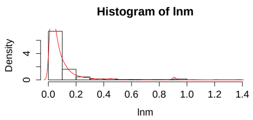

Fig. 42: Histogram of log-normal distribution

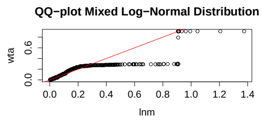

Fig. 43: QQ-plot of Mixed Log-Normal Distribution

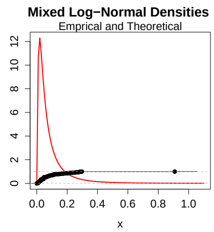

Fig. 44: Empirical and Theoretical Mixed Log-Normal Densities

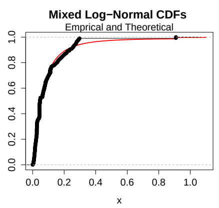

Fig. 45: Empirical and Theoretical Mixed Log-Normal CDF

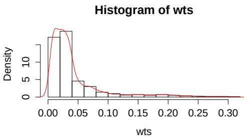

Fig. 46: Histogram of Inter arrival times

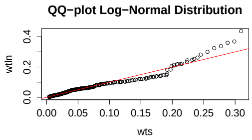

Fig. 47: QQ-plot for Log-Normal Distribution

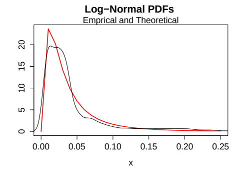

Fig. 48: Empirical and Theoretical Log-Normal PDF

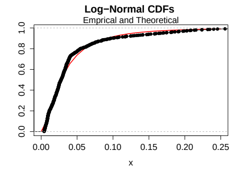

Fig. 49: Empirical and Theoretical Log-Normal CDF

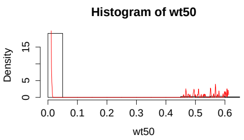

Fig. 50: Histogram of Inter arrival times

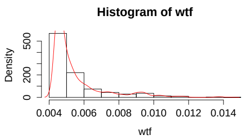

Fig. 51: Histogram of Inter arrival times first part

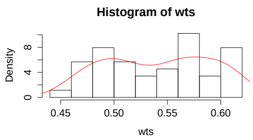

Fig. 52: Histogram of Inter arrival times second part

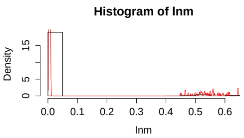

Fig. 53: Histogram of log-normal distribution

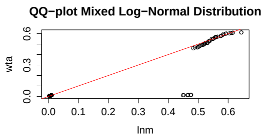

Fig. 54: QQ-plot of Mixed Log-Normal Distribution

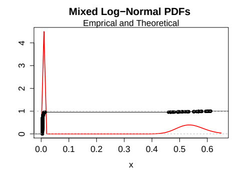

Fig. 55: Empirical and Theoretical Mixed Log-Normal Densities

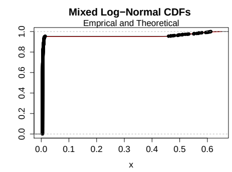

Fig. 56: Empirical and Theoretical Mixed Log-Normal CDF

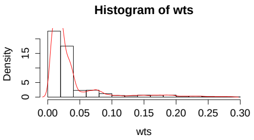

Fig. 57: Histogram of Inter arrival times

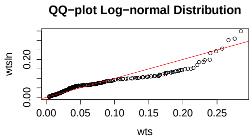

Fig. 58: QQ-plot for Log-Normal Distribution

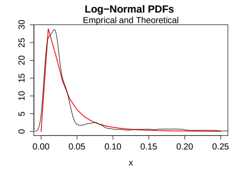

Fig. 59: Empirical and Theoretical Log-Normal PDF

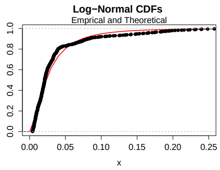

Fig. 60: Empirical and Theoretical Log-Normal CDF

Table 13: Distribution of Inter-Arrival times for 40 nodes, with TMAC, sending 1 packet every 5 minutes; corresponding Figures 61, 62, 63, 64, 65, 66

Number of observation in the simulated series	ML Estimates of the parameters of empirical distribution	Kolmogorov-Smirnov Test statistics	P-values	Corresponding theoretical distribution for the empirical one
All:1138 Used:1106 First part:1064 Second part: 38	Mixed Log-Normal Meanlog1 = -5.26 Sdlog1 = 0.22 Meanlog2 = -0.62 Sdlog2 = 0.10 Mixing proportion: 0.04	Average of 100 runs :0.22	Average of 100 runs : 0.0	An unknown Mixed distribution

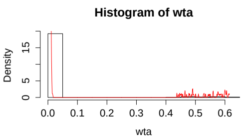

Fig. 61: Histogram of Inter arrival times

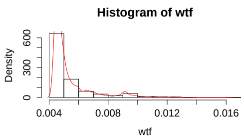

Fig. 62: Histogram of Inter arrival times first part

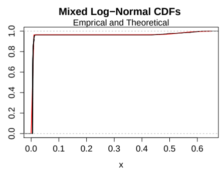

Fig. 63: Empirical and Theoretical Mixed Log-Normal CDF

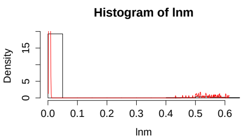

Fig. 64: Histogram of log-normal distribution

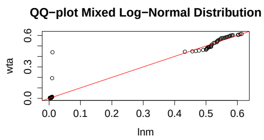

Fig. 65: QQ-plot of Mixed Log-Normal Distribution

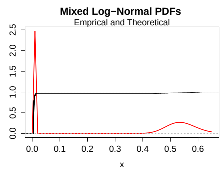

Fig. 66: Empirical and Theoretical Mixed Log-Normal Densities

As we see from the tables presented, the empirical exponential arrival distribution assumption of wireless sensor networks holds only for two cases: 10 nodes, one packet every 5 minutes without MAC and 10 nodes, one packet every 10 minutes with CSMA. There are both theoretically known such as Gamma, Log-normal and Mixed Log-Normal of arrival distributions and theoretically unknown such as non-Exponential and Mixed arrival distributions in WSNs. It seems by increasing the number of nodes, the modes of empirical distributions are getting lower values. At the same time, the right tails of the distributions are getting higher values. In other words, the empirical distributions are squeezed and pushed to the right having tails from Exponential to Gamma and then to Log-Normal distributions. If there are discontinuities of the empirical distributions then mixed theoretical distributions look more proper such as Mixed Log-Normal distribution of 10 nodes, one packet every 5 minutes with TMAC and 10 nodes, one packet every 5 seconds with CSMA. However, finite mixture models are often over-parametrized, leading to identification issues such as in 20, 35 and 40 nodes with TMAC and 20 nodes, one packet every 5 seconds of CSMA where the distributions are mixed but theoretically unknown.

When CSMA/CA is employed as the MAC protocol, for low data rates and lower number of nodes (10 nodes sending 1 packet every 10 minutes), the corresponding theoretical distribution for the empirical one is exponential (Table 4). This is due to the exponential back-off. As the number of nodes increases and as the data rate increases to 1 packet every 1 second, characterization of the arrival distributions at the CH becomes arduous due to the delay caused while the nodes try to repeatedly transmit the RTS beacon packets. Maximum Likelihood parameters of empirical distributions are estimated, theoretical distributions based on the estimated parameters are then generated. K-S Test Statistics for each generated data series are conducted in order to verify if it is possible to have a corresponding theoretical distribution. Due to the effects of CSMS/CA, arrivals at the CH follow theoretically unknown mixed distributions (Tables 7, 8, 9).

For the cases where TMAC is employed as MAC protocol, though the energy consumption is low by reducing the amount of energy wasted on idle listening by placing an adaptive duty cycle, there is extra delay incurred as the node takes an extra amount of time to wake up. Hence, characterizing arrival distribution at the cluster head accurately becomes complex. When any node has a packet to send, it starts to repeatedly transmit RTS beacon packets based on CSMA manner, i.e. through carrier sense and random back-off manner, making sure the channel is idle before it can transmit a packet, therefore causing delays. Although this can save considerable amounts of energy, this extra delay slightly increases the probability of collision. From the results presented in Tables 3, 6, 11, 13 and their corresponding figures, it is evident that the effects of TMAC on the arrival distributions are clear. Except for the case where the number of nodes is relatively smaller, where the corresponding distribution follows Mixed Log-Normal, in all the other cases the corresponding theoretical distribution for the empirical one follows an unknown mixed distribution.

7 Conclusion

To the best of our knowledge, this is the first work that provides statistical proof for finding theoretical distributions of arrivals at the CH and relay nodes in WSNs. A clustered model is considered characterised by its sending rate, inter-arrival distribution and the service process. The empirical distributions of inter-arrival times of the packets considering such physical events that do not occur frequently are generally assumed by Poisson processes, and the inter-arrival times by exponential distributions. The general practice in published works is thus to compare empirical exponential arrival distributions of wireless sensor networks with theoretical exponential distributions in Q-Q plot diagrams. In this paper, we show that such comparisons based on simple eye checks are not sufficient since in many cases incorrect conclusions may be drawn from such plots. After estimating Maximum Likelihood parameters of empirical distributions, we generate theoretical distributions based on the estimated parameters. By conducting Kolmogorov-Smirnov Test Statistics for each generated data series, we find out, if it is possible, a corresponding theoretical distribution. Empirical exponential arrival distribution assumption of wireless sensor networks holds only for a few cases. There are both theoretically known such as Gamma, Log-normal and Mixed Log-Normal of arrival distributions and theoretically unknown such as non-Exponential and Mixed arrival distributions in wireless sensor networks. The effects caused by MAC properties are also analysed by experimenting with well known MAC protocols and the summary of the inter arrival time distributions after extensive tests are presented for various application categories in 14. Therefore, these results confirm that the assumption of exponential inter-arrival distributions does not hold in all the cases. Exponential arrival distribution assumption of wireless sensor networks holds only when a fewer nodes (10-15), sending packet every 5-10 minutes with no MAC properties, as-well as when CSMA properties are considered.

Table 14: Summary for Inter arrival time distributions for various application categories

S.No	Application Type	Packet Rate (1packet/time)	Inter arrival time distribution		
			No MAC	TMAC	CSMA
1	Environment monitoring, Smart agriculture [31, 37]	5 - 10 min	Exponential for 10-15 nodes	Mixed-Log normal for 10-15 nodes	Exponential for 10-15 nodes
2	Traffic monitoring, Vehicle tracking [38, 39]	5 - 10 sec	Gamma for 20-30 nodes	Unknown Mixed distribution	Unknown mixed distribution
3	Military applications, Body Area Networks [40, 41]	1 sec or higher	Constant	Unknown mixed distribution	Unknown non-Exponential

A Some Probability Distributions

A.1 Exponential Distribution

This is a distribution of the time to an event when the probability of the event occurring in the next small time interval does not vary through time. It is also the distribution of the time between events when the number of events in any time interval has a Poisson distribution. The exponential distribution is characterized as follows:

Definition Let X be an absolutely continuous random variable. Let its support be the set of positive real numbers:

$$R_X = [0, \infty) \tag{1}$$

Let $\lambda \in R_{++}$. We say that X has an exponential distribution with parameter λ (rate parameter) if its probability density function is:

$$f_X(x) = \begin{cases} \lambda exp(-\lambda x) & \text{if } x \in R_X \\ 0 & \text{if } x \notin R_X \end{cases}$$

A.2 Gamma Distribution

The gamma distribution includes the chi-squared, Erlang, and exponential distributions as special cases, but the shape parameter of the gamma is not confined to integer values. The gamma distribution starts at the origin and has a flexible shape.

A.3 Log-normal Distribution

In probability theory, a log-normal distribution is a probability distribution of a random variable whose logarithm is normally distributed. If Y is a random variable with a normal distribution, then X = exp(Y)has a log-normal distribution; likewise, if X is log-normally distributed, then Y = log(X)is normally distributed. The log-normal distribution is applicable to random variables that are constrained by zero but have a few very large values. The resulting distribution is asymmetrical and positively skewed.

A.4 Mixture Distributions

A mixture distribution has a distribution function with a representation as a convex combination of other specific probability distribution functions. A mixture may be comprised of a finite number of base elements, where usually a relatively small number of individual distributions are combined together, or an infinite number of base elements. Often an individual base distribution is thought of as representing a unique sub population within the larger (sampled) population. In both the finite and infinite case, the probability of an outcome may be thought of as a weighted average of the conditional probabilities of that outcome given each base distribution, where the relevant mixture weight describes the relative likelihood of a draw from that distribution being obtained.

A.4.1 Finite Mixture

A *finite mixture of two distributions* having cdfs $F_1(x)$ and $F_2(x)$, respectively, has cdf $F_x = \eta F_1(x) + (1 - \eta)F_2(x)$, as long as 0 <$\eta$ <1. Extending this notion to a finite mixture of K distributions (sometimes referred to as a finite K-mixture) involves using a convex combination of distinct distribution functions. As the combination is convex, each of the mixture weights η_1, η_2, . . . ,η_k are between zero and one, and sum to unity.

Due to their ability to combine very different distributional structures, finite mixture distributions are well suited to cater for a large range of empirical distributions in practice. However, finite mixture models are often over-parametrized, leading to identification issues.

B Probability Plots or Quantile-Quantile Plots

A probability plot or quantile-quantile (Q-Q) plot is a graphical display invented by Wilk and Gnanadesikan [42], to compare a data set to a particular probability distribution or to compare it to another data set. The idea is that if two population distributions are exactly the same, then they have the same quantiles (percentiles), so a plot of the quantiles for the first distribution versus the quantiles for the second distribution will fall on the 0 – 1 line (i.e., the straight line $y = x$ with intercept 0 and slope 1). If the two distributions have the same shape and spread but different locations, then the plot of the quantiles will fall on the line $y = a + x$ (parallel to the 0 – 1 line) where a denotes the difference in locations. If the distributions have different locations and differ by a multiplicative constant b, then the plot of the quantiles will fall on the line $y = a + bx$ [43, 44]. Various kinds of differences between distributions will yield various kinds of deviations from a straight line.

Acknowledgement

The authors would like to thank the members of **SENSOLAB** for their valuable feedback.

References

1. N. Bouabdallah, M. E. Rivero-Angeles, and B. Sericola, “Continuous monitoring using event-driven reporting for cluster-based wireless sensor networks,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3460–3479, 2009.
2. C. Chiasserini and M. Garetto, “Modeling the performance of wireless sensor networks,” in *INFOCOM, Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 1, pp. 4 vol. (xxxv+2866), march 2004.
3. I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.
4. Y. Wang, M. C. Vuran, and S. Goddard, “Cross-layer analysis of the end-to-end delay distribution in wireless sensor networks,” *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 1, pp. 305–318, 2012.
5. K. Doddapaneni, *Energy aware performance evaluation of WSNs*. PhD thesis, Middlesex University, 2014.
6. I. Dbibih, O. Zytoune, and D. Aboutajdine, “On/off markov model based energy-delay aware mac protocol for wireless sensor network,” *Wireless Personal Communications*, vol. 78, no. 2, pp. 1143–1155, 2014.
7. F. Omondi, E. Ever, P. Shah, and O. Gemikonakli, “Modelling wireless sensor networks for performability evaluation,” in *Ad-hoc, Mobile, and Wireless Network* (J. CichoÅĎ, M. GÈľbala, and M. Klonowski, eds.), vol. 7960 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013.
8. Y. Zhang and W. W. Li, “Modeling and energy consumption evaluation of a stochastic wireless sensor network.,” *EURASIP J. Wireless Comm. and Networking*, p. 282, 2012.
9. Y. Zhang and W. Li, “An energy-based stochastic model for wireless sensor networks.,” *Wireless Sensor Network*, vol. 3, no. 9, pp. 322–328, 2011.
10. K. Doddapaneni, F. A. Omondi, E. Ever, P. Shah, O. Gemikonakli, and R. Gagliardi, “Deployment challenges and developments in wireless sensor networks clustering,” in *2014 28th International Conference on Advanced Information Networking and Applications(AINA)*, pp. 227–232, May 2014.
11. C. Chiasserini and M. Garetto, “An analytical model for wireless sensor networks with sleeping nodes,” *IEEE Transactions on Mobile Computing*, vol. 5, pp. 1706–1718, Dec 2006.

12. E. Ever, R. Luchmun, L. Mostarda, A. Navarra, and P. Shah, "Uheed - an unequal clustering algorithm for wireless sensor networks," in *SENSORNETS*, pp. 185–193, 2012.
13. K. Doddapaneni, E. Ever, O. Gemikonakli, I. Malavolta, L. Mostarda, and H. Muccini, "Path loss effect on energy consumption in a wsn," *International Conference on Computer Modeling and Simulation*, vol. 0, pp. 569–574, 2012.
14. S. Singh and C. S. Raghavendra, "Pamas—power aware multi-access protocol with signalling for ad hoc networks," *SIGCOMM Comput. Commun. Rev.*, vol. 28, pp. 5–26, July 1998.
15. W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient mac protocol for wireless sensor networks," in *Proceedings of the IEEE Infocom*, (New York, NY, USA), pp. 1567–1576, IEEE, June 2002.
16. C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava, "Topology management for sensor networks: Exploiting latency and density," in *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking &Amp; Computing*, MobiHoc '02, (New York, NY, USA), pp. 135–145, ACM, 2002.
17. R. Zheng, J. C. Hou, and L. Sha, "Asynchronous wakeup for ad hoc networks," in *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '03, (New York, NY, USA), pp. 35–45, ACM, 2003.
18. T. Qiu, L. Feng, F. Xia, G. Wu, and Y. Zhou, "A packet buffer evaluation method exploiting queueing theory for wireless sensor networks.," *Comput. Sci. Inf. Syst.*, vol. 8, no. 4, pp. 1028–1049, 2011.
19. A. F. Mini, B. Nath, and A. A. F. Loureiro, "A probabilistic approach to predict the energy consumption in wireless sensor networks," pp. 23–25, 2002.
20. P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.
21. S. Shakkottai, R. Srikant, and N. Shroff, "Unreliable sensor grids: coverage, connectivity and diameter," vol. 2, pp. 1073–1083 vol.2, 2003.
22. G. Ferrari, P. Medagliani, S. D. Piazza, and M. Martalò, "Wireless sensor networks: performance analysis in indoor scenarios," *EURASIP J. Wirel. Commun. Netw.*, no. 1, pp. 41–41, 2007.
23. G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE J.Sel. A. Commun.*, vol. 18, pp. 535–547, Sept. 2006.
24. J. Misic, S. Shafi, and V. Misic, "Performance of a beacon enabled ieee 802.15.4 cluster with downlink and uplink traffic," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 4, pp. 361–376, 2006.
25. I. Ramachandran, A. K. Das, and S. Roy, "Analysis of the contention access period of IEEE 802.15.4 MAC," *ACM Trans. Sen. Netw.*, vol. 3, no. 1, p. 4, 2007.
26. M. Martalò, S. Busanelli, and G. Ferrari, "Markov chain-based performance analysis of multihop ieee 802.15.4 wireless networks," vol. 66, no. 12, pp. 722–741.
27. J. Misic, J. Fung, and V. Misic, "Interconnecting 802.15.4 clusters in master-slave mode: queueing theoretic analysis," in *ISPAN 2005. Proceedings. 8th International Symposium on Parallel Architectures,Algorithms and Networks,.*, pp. 8 pp.–, 2005.
28. J. Misic and R. Udayshankar, "Slave-slave bridging in 802.15.4 beacon enabled networks," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, pp. 3890–3895.
29. I. Emary and S. Ramakrishnan, *Wireless Sensor Networks: From Theory to Applications*. Telecommunications books, Taylor & Francis, 2013.
30. S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensor network models," *ACM MOBILE COMPUTING AND COMMUNICATIONS REVIEW*, vol. 6, pp. 28–36, 2002.
31. A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, WSNA '02, (New York, NY, USA), pp. 88–97, ACM, 2002.
32. T. van Dam and K. Langendoen, "An adaptive energy-efficient mac protocol for wireless sensor networks," in *Proceedings of the 1st international conference on Embedded networked sensor systems*, SenSys '03, (New York, NY, USA), pp. 171–180, ACM, 2003.
33. M. Zuniga and B. Krishnamachari, "Analyzing the transitional region in low power wireless links," in *First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, IEEE SECON.*, pp. 517 – 526, Oct. 2009.
34. C. Newport, D. Kotz, Y. Yuan, R. S. Gray, J. Liu, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," pp. 78–82, ACM Press, 2004.
35. B. LatrÃl', P. D. Mil, I. Moerman, N. V. Dierdonck, B. Dhoedt, and P. Demeester, "Maximum throughput and minimum delay in IEEE 802.15.4.," in *MSN*, vol. 3794 of *Lecture Notes in Computer Science*, pp. 866–876, Springer, 2005.

36. M. Kuorilehto, M. Hännikäinen, and T. D. Hämäläinen, “A survey of application distribution in wireless sensor networks,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2005, pp. 774–788, Oct. 2005.
37. A. Boulis, C.-C. Han, and M. B. Srivastava, “Design and implementation of a framework for efficient and programmable sensor networks,” in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, MobiSys ’03, (New York, NY, USA), pp. 187–200, ACM, 2003.
38. W. Heinzelman, A. Murphy, H. Carvalho, and M. Perillo, “Middleware to support sensor network applications,” *Network, IEEE*, vol. 18, pp. 6–14, Jan 2004.
39. K. RÃűmer, “Tracking real-world phenomena with smart dust,” in *In EWSN 2004*, pp. 28–43, Springer-Verlag, 2004.
40. L. Schwiebert, S. K. Gupta, and J. Weinmann, “Research challenges in wireless networks of biomedical sensors,” in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, MobiCom ’01, (New York, NY, USA), pp. 151–165, ACM, 2001.
41. M. Storey, G. Blair, and A. Friday, “Mare: Resource discovery and configuration in ad hoc networks,” *Mob. Netw. Appl.*, vol. 7, pp. 377–387, Oct. 2002.
42. M. B. Wilk and R. Gnanadesikan, “Probability plotting methods for the analysis of data.,” *Biometrika*, vol. 55, pp. 1–17, Mar. 1968.
43. S. Millard, N. Neerchal, and P. Dixon, *Environmental Statistics with S-PLUS*. Chapman & Hall/CRC Applied Environmental Statistics, Taylor & Francis, 2000.
44. R. M. Hirsch, R. B. Alexander, and R. A. Smith, “Selection of methods for the detection and estimation of trends in water quality,” *Water Resources Research*, vol. 27, no. 5, pp. 803–813, 1991.