Probabilistic Estimation of End-to-End Path Latency in Wireless Sensor Networks

Ramon Serna Oliver and Gerhard Fohler
Chair for Real-Time Systems
Technische Universität Kaiserslautern
Kaiserslautern, Germany
{serna_oliver, fohler}@eit.uni-kl.de

***Abstract*—The inherent properties of Wireless Sensor Networks (WSN) disqualify most classic methods targeting timeliness guarantees. Assumptions of such methods as well as a restrictive notion of timeliness borrowed from classic real-time systems clash with the indeterminism of realistic scenarios.**

In this paper, we introduce a generalized notion of timeliness which allows to provide meaningful performance metrics under unreliable conditions, common in WSN. We present a probabilistic metric to capture the level of confidence for the timeliness performance without restricting its applicability. It consists of the estimation of the end-to-end delay distribution function by using current local state information of intermediate hops, which requires low memory and computational resources.

This metric represents a hook to adaptive QoS as it is constantly updated at run-time and reflects the actual network status. Extensive simulation results underline the validity of the method and its applicability.

***Index Terms*—Wireless Sensor Networks, end-to-end, latency, estimation, probabilistic.**

I. Introduction

Timeliness guarantees are of special interest in the area of Wireless Sensor Networks (WSN) [1]. Unfortunately, the notion of timeliness borrowed from classic real-time systems [2] clashes with the unfeasibility of WSN to guarantee bounded response times for the basic network operations, which is due to, among others, ad-hoc infrastructure and strong energy constraints imposed by limited battery capacities. Nevertheless, existing methods aim at strictly satisfying predefined deadlines for individual messages.

As an example, bounded response times at the MAC layer may be achieved with periodic sensing of the medium and neighborhood synchronization. Unfortunately, these are often not affordable in terms of energy. Ad-hoc network topology forces the dynamic reorganization of routing paths, preventing end-to-end message scheduling which is also inhibited by the impossibility to store global knowledge (i.e. routing tables) in the limited memory capacity of nodes. Similarly, strategies based on resource reservations over-constrain the network capacity up to the point of loosing feasibility.

Alternative approaches try to derive offline analytical models to extrapolate the timeliness performance at runtime. However, despite the validity of their conclusions, models have to be over-constrained in order to support exact timeliness analysis. These impositions are typically related to static and regular topologies, symmetry of the radio propagation patterns [3] or absence of environmental interferences. Analysis based on deadlines for individual messages conflict with the large number of limitations preventing temporal guarantees in WSN [4].

In this paper we exploit a generalized timeliness notion introduced in [5] which provides enough flexibility to suit the characteristics of WSN. Instead of aiming at strict deadlines for individual messages, the generalized notion focuses on the timeliness capacity of a sequence of messages. The generalized notion allows to express the end-to-end timeliness requirements by means of a target time interval and a confidence level. Hence, it is possible to relax the requirements imposed by methods based on strict deadlines while still providing valid means to evaluate timeliness performance.

We introduce a probabilistic approach based on the generalized notion, which approximates the end-to-end delay distribution of a routing path by performing statistical analysis of local information gathered at intermediate hops. It allows to estimate the probability that a sequence of messages is transmitted through a network path within a time interval. The probabilistic method allows applications to exploit quality of service trade-offs based on meaningful timeliness properties and adapts well to the principles of WSN.

The rest of the paper is organized as follows: Section II explores the related work in this field. The generalized notion of timeliness is detailed in Section III, followed by the description of the probabilistic approach in Section IV. Simulation results validating the presented method are discussed in Section V. Finally, Section VI concludes the paper.

II. Related work

Ongoing research to introduce real-time guarantees in WSN is carried out at many different levels. In [6] a survey of the current state-of-the-art is presented. Additionally, an overview of the problems in combined soft and hard real-time solutions covering the whole network stack as well as open challenges are discussed.

At the routing level, work in [7] and [8] assign velocities to messages which must be kept in order to fulfill their timeliness requirements. However, both assume static networks and nodes

This work is partially financed by the European Commission under the Framework 6 IST Project ”Wirelessly Accessible Sensor Populations (WASP)”.

equipped with localization capabilities. In [9], delay guarantees are provided by means of a TDMA scheme at the expense of limiting the length of routing paths.

Traffic regulation mechanisms are also explored as means to provide end-to-end guarantees using queuing models. In [10], the combination of queuing models and message scheduler, turns into a traffic regulation mechanism that drops messages when they loose their expectations to meet predefined end-to-end deadlines. Additionally, an example is given to approximate the delay distribution of each hop in the event of instability by means of a Gaussian distribution.
Other probabilistic methods to achieve QoS have been approached by different authors. For CPU scheduling, the notion of probabilistic deadlines and execution time distribution is explored in [11]. In [12], different levels of quality of service are considered with respect to timeliness and reliability providing probabilistic multi-path forwarding to ensure end-to-end delays. Note that despite these methods apply probabilistic techniques to their algorithms, they all aim at satisfying strict deadlines for individual messages.

In [13], the authors introduce an analysis of the impact of mobility in achieving timeliness guarantees. Additionally, a prioritized event transmission protocol based on a proactive routing protocol and resource reservation is foreseen, although the authors take the assumption of a predictable medium access protocol.
A common notion of timeliness, based on the assignment of strict end-to-end deadlines to each individual message is applied in the work referred. Not surprisingly, they all present a number of assumptions with respect to the network which limit their deployment.

With respect to the MAC level, much of the existing research is based on TDMA scheduling of neighbor nodes (e.g. [14]), hence constructing a schedule of transmissions with contention free periods. However, although valid results are obtained in controlled environments, the common restriction of these methods is the assumption of error-free communications. Moreover, the complexity of such strategies, specially in mobile networks, forces the addition of global network coordinators, which discourages their use. Alternative approaches exist, such as [15] which achieves hard real-time guarantees given an hexagonal topology of static nodes. This requirement is later relaxed in [16] although it still relies on static nodes. Besides, both methods are built on the assumptions of bounded network density and optimum communication conditions.

Analytical solutions have also been studied. In this direction, [17] approaches a sufficient schedulability condition to guarantee end-to-end delays in multi-hop WSN. Nevertheless, it is based on specific assumptions on the message transmission times and channel transmission speeds, as well as network density and path lengths. Moreover, it is practically unfeasible to produce analytical models capable to capture the dynamics of a real WSN. Assumptions, again, are necessary in order to adjust reality to the models.

III. Notion of timeliness

The concept of timeliness currently exploited in WSN is greatly influenced by the one originated in general purpose networks. In particular, attention is centered around temporal guarantees of individual messages by means of fulfilling transmission deadlines: each message receives an end-to-end deadline which delimits the time to reach the destination. If the message has not been delivered after this instant, it is likely to be dropped at one of the intermediate hops, depending on the routing policy. Certain routing strategies will drop messages before the expiration of the deadline if they estimate that the deadline cannot be met.

Unlike in most general purpose networks, the ad-hoc topology and the lack of resources of WSN prevents them from being able to guarantee bounded delays for the basic network operations. This is particularly stressed in mobile networks and noisy environments where even the connectivity of nodes cannot be taken for granted.

Secondly, the network connectivity and its capacity, suffer from a high variability due to e.g. duty-cycle schemes, mobility of nodes and additional behaviors intrinsic to the protocols. The fulfillment of end-to-end deadlines for individual messages might become a useless objective as the network is generally deprived of enforcement tools.
Thus, it appears more suitable to conceive a notion which measures the progression of timeliness, rather than ineffectively enforcing individual deadlines. Moreover, this timeliness indicator should be able to express the not-always-accurate temporal performance of WSN. Statistic analysis seems more adequate to capture the timing necessities of sensor networks.

We explore the preliminary approach introduced in [5], to achieve a better alignment between the network capabilities and the desired timeliness requirements. Instead of constraining the methods to fulfill idealized timeliness properties, we propose to relax the concept of timeliness, to suit the particularities of WSN. We consider the following requirements:

1) The way in which timeliness requirements are expressed should not encourage applications to demand unfeasible degrees of performance that the network cannot provide. Hence, given the unfeasibility of WSN to guarantee single deadlines, applications should express their demands at a higher level than individual messages.
2) A notion of timeliness expressing only success or failure, i.e., deadline met or not, is of only limited value to WSN. Rather, a continuous function to embody the level of conformance with respect to the timeliness performance is more suitable to the properties of WSN.
3) The ability of WSN to enforce strict end-to-end timeliness requirements is limited and variable at run-time. Hence, a meaningful notion of timeliness should allow applications to express a level of confidence for the aimed timeliness performance.

The generalization of the notion of timeliness that we propose supports these requirements and is composed of the following parts:

1) Our notion expresses timeliness properties of a sequence of messages, which makes it possible to cope with the indeterminism of individual delivery delays in WSN and still provide meaningful values. By sequence of messages we refer to a number of consecutive messages following the same end-to-end route, without any implication on their content or additional constraint.
2) A time interval (t_i, t_j) with $t_j > t_i \geq 0$, which sets the acceptable end-to-end delay bounds for a sequence of messages.
3) The level of confidence for the required end-to-end interval, expressed by means of a probability $0 < p < 1$ of successful arrivals within the interval.
4) The end-to-end delay distribution function, used as a timeliness indicator, which allows to capture the probability density of the sequence of messages arriving within the interval. The function, obtained at run-time, provides sufficient information to determine the probability of sequences of messages arriving within the specified interval.
5) The selection of probability level and length of the interval allows the specification of strict timeliness yet providing additional levels of flexibility which suits the particularities of WSN. Thus, our notion is a generalization of the classic timeliness notion.

By analyzing a sequence instead of individual messages, it is possible to take into consideration the indeterminism of WSN and still provide meaningful values. Furthermore, the selection of the probability level and the length of the interval allows the specification of strict timeliness, yet providing additional levels of flexibility which adapt to the peculiarities of WSN. This notion is adequate to evaluate the end-to-end timeliness performance as well as to express requirements in a way that does not demand excessive levels of precision that the network cannot achieve.

IV. Estimation of end-to-end delay distribution

In this paper, we target wide-area WSN such as those related to environment monitoring (e.g. fire detection, structural monitoring of buildings, etc.) with specific timeliness sensitive data acquisition (e.g. fire or intrusion alarms, structural damage, etc.).

Our objective is to compute the probability that a sequence of messages can be transmitted through a given routing path within a bounded time interval. Unfortunately, it has been argued in Section III that the indeterminism of WSN does not allow exact analysis without introducing restrictive assumptions. However, we show that it is possible to perform estimations of such distributions with satisfying accuracy and low complexity.

The end-to-end delay experienced by a message through a routing path can be decomposed into the individual forwarding delays originating at each node.

We first analyze the forwarding delay at each intermediate hop and then proceed with the composition of an approximate end-to-end distribution for the path latency.

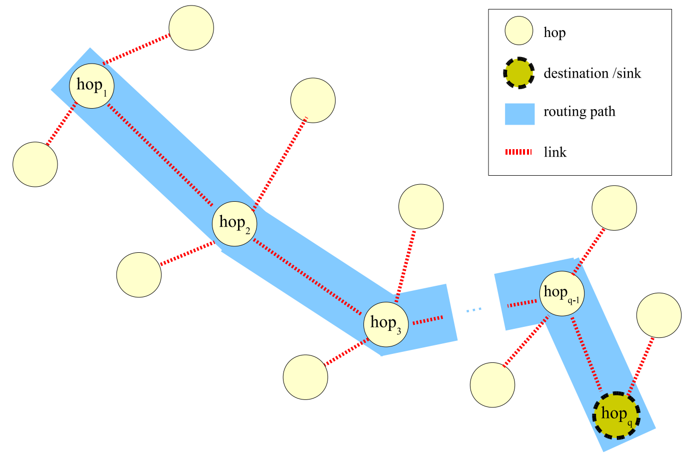

Fig. 1. Wireless Sensor Network scenario

A. Notation and definitions

A WSN, as shown in Figure 1, is represented as a graph $G(N, L)$ formed by a set of nodes N and a set of *single-hop links* L. Two nodes $hop_i, hop_j \in N$ are *directly connected* at a given time if there is a link $l \in L$, $l = (hop_i, hop_j)$ such that hop_i and hop_j can send and receive messages from each other. For the sake of simplicity, the time instance of variables that change over time are only represented when they are significant to the analysis.
$S \subset N$ is the subset of sinks. Sinks may outperform nodes with respect to resources and energy availability.

A *(routing) path* rp is a sequence of nodes $(hop_1, hop_2, ..hop_{q-1}, hop_q) \in N$ such that each pair $l_i = (hop_i, hop_{i+1}) \in L$, thus providing a *multi-hop link* between hop_1 (source) and hop_q (destination or *sink*). The length of a path is equal to its number of links ($|rp_{hop_1,hop_q}| = q$).

B. Calculation of single hop forwarding delay

Let $hop_1, hop_2 \in N$ be two nodes such that there exists one link $l \in L$, $l = (hop_1, hop_2)$. Then, we define D_{hop_1} as the *random variable* RV which characterizes the forwarding latency δ of a message from hop_1 to hop_2 and $F_{D_{hop_1}}$ as the *cumulative probability function* (cdf) such that $F_{D_{hop_1}}(\varepsilon) = P(D_{hop_1} \leq \varepsilon)$; the probability that hop_1 introduces a delay of at most ε in forwarding a message.

We define the forwarding latency δ as the time interval between the message arriving in node t_{in} (either because the application layer sets a new message to be sent, or because the MAC layer receives a message which has to be forwarded), and the reception of an acknowledgment from the receiver t_{ack}. Note that this calculation is pessimistic as it introduces the additional time to receive the acknowledgment. Depending on the MAC protocol, this may be relatively small and can be neglected, but in any case it should be possible to estimate this time as a constant φ. Hence, we adjust δ to:

$$\delta = t_{ack} - t_{in} - \varphi \quad (1)$$

The above calculation is updated at each hop every time a message goes through. Then, an increasing sequence of values $\delta_0, .., \delta_k$, which represents samples of the random variable D_{hop_i} for that node is generated. A cumulative method to avoid storing the whole sequence of values will be introduced in next section.

C. End-to-end delay distribution

We now consider two cases to extend our analysis to the end-to-end distribution. At a first step, the simple case with one single link, and later on the general case with $|rp| > 1$.

Simple case: one hop: In the simplest case, a message is forwarded by one single node. Let hop_1 be that node, and $l = (hop_1, hop_2)$ the link to the sink. Thus, $rp = (hop_1, hop_2)$ and $|rp| = 1$). In this case, the end-to-end latency is equal to the forwarding latency of the only hop δ_{hop_1} and, by extension the end-to-end delay distribution, D_{e2e} equals to its forwarding distribution D_{hop_1}.

The cdf of the distribution depends on many factors which are generally out of our control, such as the link quality, environmental noise, and most relevant, the underlaying MAC protocol. At this point, it is difficult to characterize the distribution as it is not feasible to extrapolate its shape. In any case, we can approximate the mean (μ) and variance (σ^2) which will give a rough indicator of evolution of the link performance over time.

We will use the sample mean $\bar{x}$ and sample variance s^2 as good estimators for μ and σ^2. Moreover, for the runtime calculations, we propose the *exponential weighted moving average* (*EWMA*) [18], [19] as it requires little memory utilization and has low computational overhead (Equation 2). The EWMA produces two new values, $\bar{x}_t^*$ and $s_t^{2^*}$, updated at each iteration, t being the index of the iteration (i.e. number of sample). The equations for $\bar{x}_t^*$ and $s_t^{2^*}$ are:

$$\begin{aligned} \bar{x}_t^* &= \alpha\delta_t + (1-\alpha)\bar{x}_{t-1}^* \\ s_t^{2^*} &= \frac{\alpha}{2-\alpha} s_t^2 \end{aligned} \tag{2}$$

Note parameter α ($0 \leq \alpha \leq 1$), which is set to weigh the actual measurements with respect to the previous, hence smoothing the consequences of past trends and punctual abnormalities. The discussion of the selection of its value follows in Section IV-D.

To calculate the sample variance (s_t^2) without requiring the whole set of samples to be stored in memory, we can take advantage of the following iterative Equation 3:

$$s_t^2 = \frac{t-1}{t} s_{t-1}^2 + \frac{1}{t-1}(\delta_t - \bar{x}_t)^2 \tag{3}$$

Equation 3 and 2 can be computed at each node every time a message is forwarded through it, hence updating its local state information.

General case: k hops: To compose the RV of the forwarding latency at intermediate hops, we take advantage of the *Central Limit Theorem* [20] (CLT). The classic formulation of the CLT[1] states that the sum of a number (k) of RVs with approximately the same distribution, non-negative and mutually independent tends to a *Normal* distribution $N(\mu, \sigma^2)$. The argumentation to accept the premises of similarity and independence are exposed in section IV-E.

The approximated end-to-end delay distribution D_{rp} can be characterized as:

$$D_{rp} = \sum_{\forall h \in rp} D_h \tag{4}$$

and,

$$F_{D_{rp}}(\tau) = P(D_{rp} \leq \tau) \tag{5}$$

And the parameters μ_{rp} and σ_{rp}^2 of D_{rp} are:

$$\begin{aligned} \mu_{rp} &\approx \bar{x}_{D_{rp}} = \sum_{\forall n \in rp} \bar{x_h}^*_{D_h} \\ \sigma_{rp}^2 &\approx s_{D_{rp}}^2 = \sum_{\forall n \in rp} s^{2\,*}_{h\,D_h} \end{aligned} \tag{6}$$

Applying the CLT, the probability introduced in Equation 5 converges to:

$$\begin{aligned} F_{D_{rp}}(\tau) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau} e^{\frac{y^2}{2}} dy \\ \tau &= \frac{D_{rp} - \mu_{D_{rp}}}{\sigma_{D_{rp}}} \end{aligned} \tag{7}$$

Note that the computation of Equation 7 is only performed at the sink, which in general outperforms other nodes with respect to energy and computation capacity.

D. Selection of α

The exponential weight α controls the smoothing factor in the Equations 2. Lower values of α increase the stability of the measurements as they smoothen the reaction due to small variations with respect to the averaged value. This is the desired effect to avoid insignificant imprecisions due to oscillations on the sequence of measurements. On the other hand, large values of α tune the equations to quickly adapt to changes and forget the past values with more celerity.

Figure 2 shows the evolution of $\bar{x}^*$ for different values of α during a short interval of time. Despite the benefit of the smoothing qualities of a low α for the mean sampled value, it is not the case for the calculation of the sample variance (Equation 2). In this case, the variability of values needs to be captured as it plays an important role in the estimation of the final distribution.

After performing extensive simulations [22], we observed that it is better to reflect the variability of the data in order to obtain accurate s^2, than to obtain a stable value of $\bar{x}^*$. Nevertheless, EWMA still proves effective to reduce the effect of punctual overestimations of the forwarding delay. Hence, our experiences [22] show that a large value of α produces acceptable results for most cases. For that reason, and based on our experience, we chose $\alpha = 0.9$.

[1] Although the CLT is commonly applied to large number of samples, an argumentation about good approximations for smaller sums of RVs is given in [21].

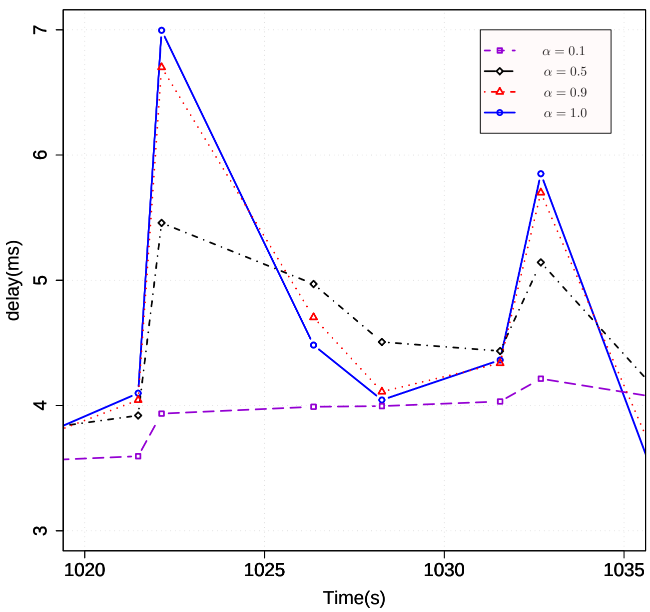

Fig. 2. Effect of different values of α

E. Assumptions on the premises of the CLT

Message latencies across a network might experience dependencies under certain circumstances forcing events to happen in a non-independent way (i.e. $E[V_a, Vb] \neq E[V_a] \cdot E[V_b]]$). For instance, dependencies may appear when a message m_1 causes additional delay to a message m_2 which message m_2 would not have experienced in isolation.

These circumstances may occur due to medium access (e.g. back-offs) or buffering constraints (e.g. messages dropped due to buffer overflows) and are likely to manifest proportionally to the network load. In networks with low traffic load, the probability of dependencies being significant are practically negligible. Contrarily, it is possible that in loaded WSN the dependencies are reflected on the estimations.

In any case, it is relevant to note that dependencies will be accounted for at runtime as the delay distribution at each hop is continuously estimated from their timeliness performance. Thus, the estimated parameters for a certain hop will vary for different network loads as well as the composed end-to-end delay distribution.

Numerous studies relax the premises of the CLT (e.g. [23]). However, the prerequisites of these reformulated versions of the theorem impose new premises which are difficult to guarantee without analytical models for the network.

To overcome this situation, we decided to take the premises as valid with the expectation that the spatial distribution of nodes and the typical low throughput of WSN minimizes the dependencies, and even in situations of high load, these are not significant to the desired accuracy of approximation.

After performing simulations, we believe that the dependencies are not of relevance in the general case and only in situations of very high network saturation they might arise. The evaluation in Section V will show whether this assumption is appropriate and if not, how significant the effects are.

F. Applicability

The timeliness performance metric described in this paper allows multiple application scenarios. Firstly, the information regarding hop forwarding latency (i.e. $\bar{x}^*$, s^{2^*}) can be concatenated to each message (or special control messages) allowing the recipient to capture the probabilistic estimations. As an example of such an application, let d be a sink of the end-to-end path rp with the requirement that approximately 80% of the messages are delivered within the time interval $(0.8s, 1.2s)$. If the accumulated parameters received by d (i.e. $\bar{x}^*$, $s_t^{2^*}$) result in a distribution that satisfies this probability for the given time interval, then there is a probabilistic guarantee that the condition will hold.

It is possible to adapt existing tree-routing protocols [24] to choose the relaying hops based on this metric, hence maximizing the value of the end-to-end parameters (e.g. achieving the best end-to-end distribution). Note that the elaboration of a routing protocol is beyond the scope of this paper.

V. Evaluation

To evaluate the timeliness performance we run extensive simulations with the simulation tool Omnet++ [25] with the Mobility Framework [26]. We chose WiseMAC [27] as an energy-efficient MAC protocol specially designed for WSN. Each run simulated a period of 10 days.

A. Evaluation criteria

Under static network conditions, the effective end-to-end delay (Φ) and the accumulated parameters estimated at each hop ($\Delta_{\bar{x}}, \Delta_{s^2}$ for simplicity) would be representative of the effective distribution and the estimated Normal distribution $N(\Delta_{\bar{x}}, \Delta_{s^2})$. Therefore, a sequence of messages large enough transmitted along the path would suffice to estimate the distributions and allow a direct comparison between them.

However, each time that a message is forwarded by a hop, it recalculates its parameters $\bar{x}^*$ and s^{2^*}. Thus, a message forwarded by any of the intermediate links of the path, will produce a change in the accumulated end-to-end parameters. In other words, the network conditions are different at every instant that a message is being forwarded.

A consequence of the above, is that for each instance t of a message going through the analyzed path we capture an effective end-to-end delay Φ^t and a set of parameters $\Delta^t_{\bar{x}}, \Delta^t_{s^2}$ which are not directly comparable to those originated by previous or following messages. This way, with only independent sets of samples it cannot be extrapolated whether their approximation to $N(\Delta_{\bar{x}}, \Delta_{s^2})$ is accurate or not.

To overcome this problem we perform two complementary tests:

Test 1: Normalize each sample of the effective end-to-end distribution to the standard Normal distribution $N(0,1)$. Given,

$$X \sim N(\mu, \sigma^2)$$

then,

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} \\ Z &\sim N(0,1). \end{aligned}$$

This way, instead of comparing each individual sample to a $N(\mu,\sigma)$ with different parameters, we can compare all samples against a $N(0,1)$. Thus, the expectation is that the distribution of the normalized samples approximates a $N(0,1)$.

Test 2: Compare the number of "hits" of each interval determined by the distance σ from the center point (μ). This is known to be around 68%, 27%, 4.2% and 0.2% respectively for the intervals $I_1 = (-\sigma,\sigma)$, $I_2 = (-2\sigma,-\sigma) \cup (\sigma, 2\sigma)$, $I_3 = (-3\sigma,-2\sigma) \cup (2\sigma,3\sigma)$ and $I_4 = (-\infty,-3\sigma) \cup (3\sigma,\infty)$. If the estimated distribution is accurate, the number of samples falling in each of these intervals should approximately follow similar proportions.

B. Scenario

We simulated traffic messages from a sender node hop_1 to a sink hop_q with the interference of cross-traffic coming from neighbor nodes as depicted in Figure 1. The motivation for the chosen scenario is partially motivated by the scenarios described in WASP [28]. A common setup for each simulation run was chosen with variation in the length of the path and cross-traffic parameters:

- path length: $|rp| = \{5, 10\}$,
- n_1 sending periodic messages to s with period $T = 30s$,
- messages aggregate the estimated parameters at each intermediate link (Equation 6).
- $\alpha = 0.9$
- effective end-to-end transmission latency experienced by each message is captured at s,
- each hop in the path has two neighbors simulating cross traffic following a Poisson distribution with parameter $\lambda = \{30s, 60s, 120s, 480s, 1200s\}$,
- radial distance between nodes following a uniform distribution with range 8 to 20 meters,
- radio interface according to the specification of the RFM TR1100 radio transceiver.

Hence, the results of ten different simulation runs with the combination of parameters λ and $|rp|$ are presented. Notice that the process of building the routing path is not considered at this moment. The purpose of the simulations is to evaluate the validity of the method to obtain the end-to-end delay distribution.

C. Simulation results

We first show two representative cases: for the experiment with $|rp| = 5$, Figure 3 shows the histogram and *probability density function (pdf)* after normalization compared to the *pdf*

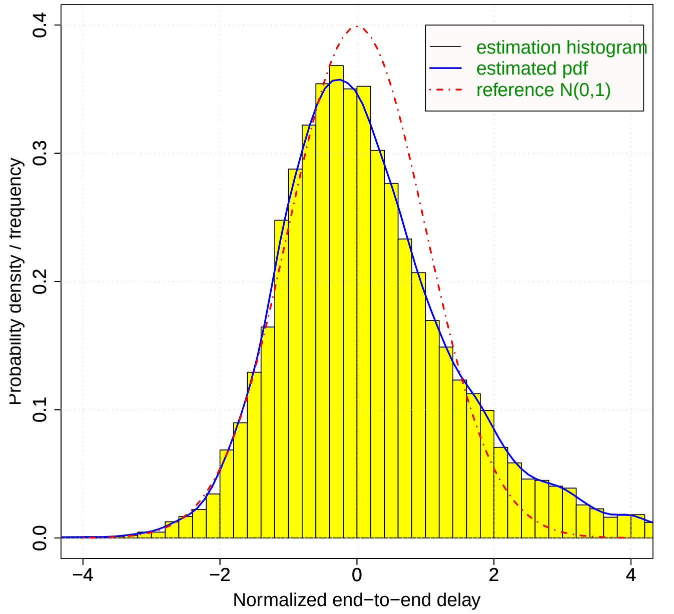

Fig. 3. Normalized histogram and estimated *pdf* vs $N(0,1)$, with $|rp| = 5$

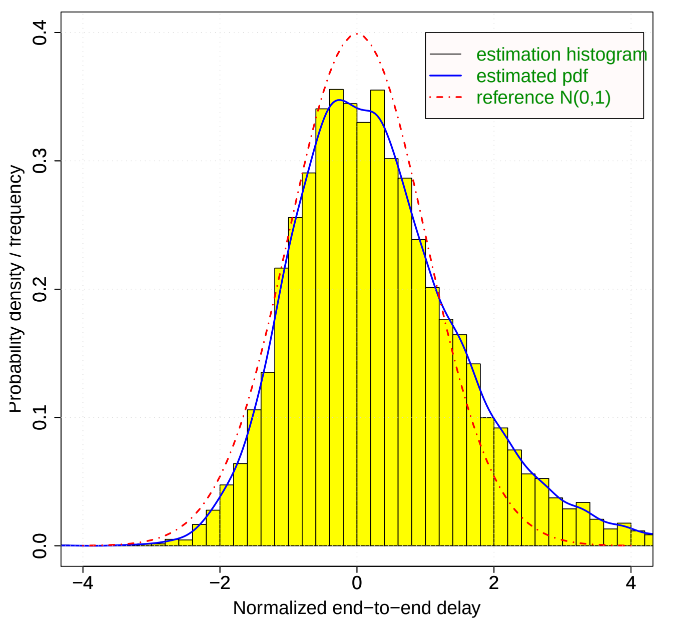

Fig. 4. Normalized histogram and estimated *pdf* vs $N(0,1)$, with $|rp| = 10$

of the standard Normal $N(0,1)$. The depicted graphics have been cropped at the interval $(-4,4)$.

At first sight, two questions arise: the difference between the two curves at the central point and the larger tail on the right side. Both effects are related to each other and can be explained by the nature of the experiment measurements. In fact, the values represented come from measured end-to-end

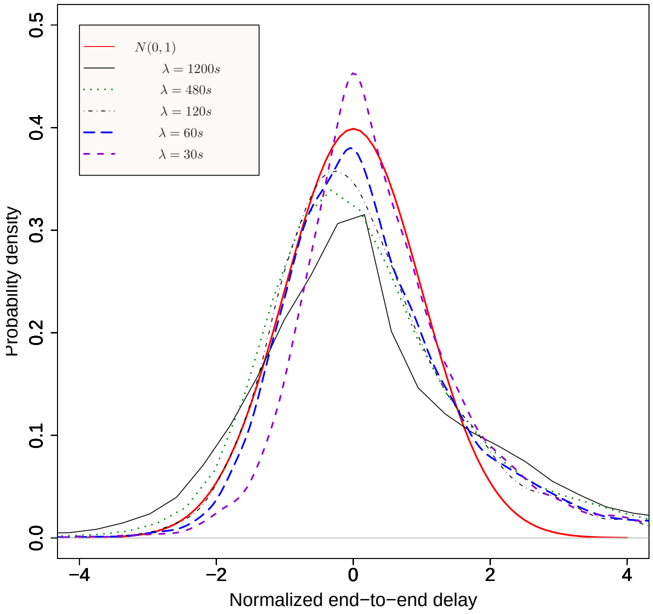

Fig. 5. Estimated pdfs vs N(0,1), with $|rp| = 5$

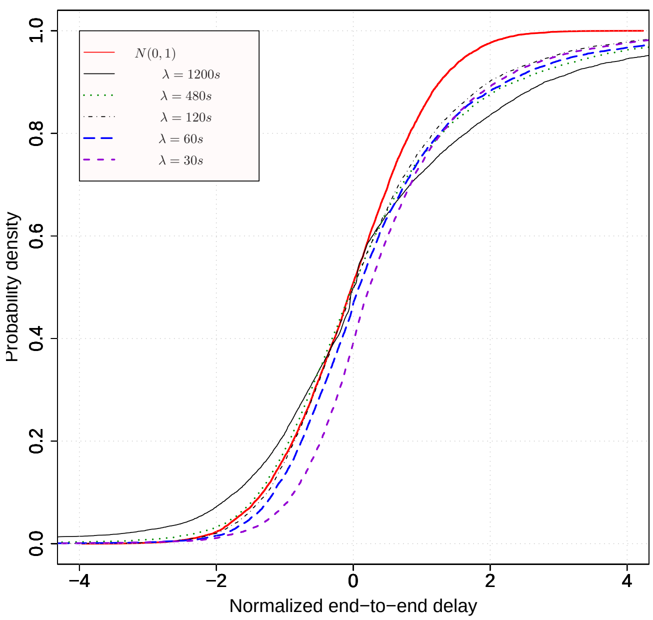

Fig. 6. Estimated cdfs vs N(0,1), with $|rp| = 5$

delays. This necessarily introduces a tail effect, as there is a clear limit on the possible values from the left side (i.e. time delays cannot be negative) but none on the right side.

With respect to the range of absolute values, having a mean sample value of $6.5ms$ very few messages achieved a delay less than or equal to $2ms$ and the distance between the minimum value and the mean is approximately $5ms$. However, on the right side, this distance goes up to around $34ms$, with

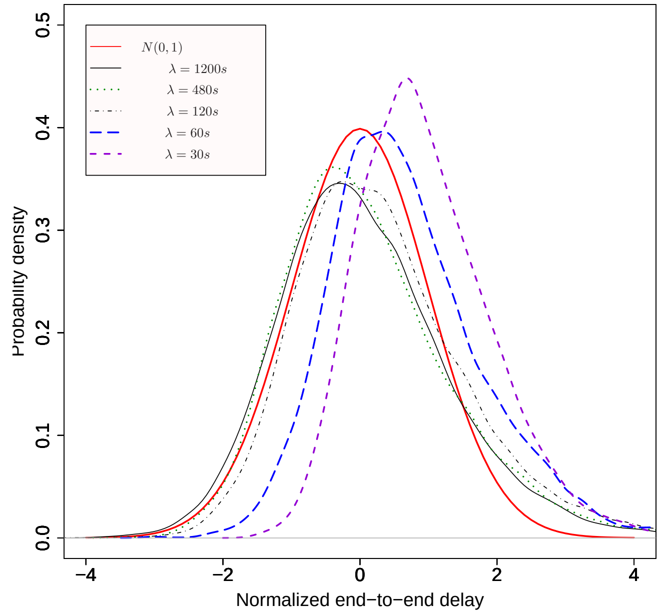

Fig. 7. Estimated pdfs vs N(0,1), with $|rp| = 10$

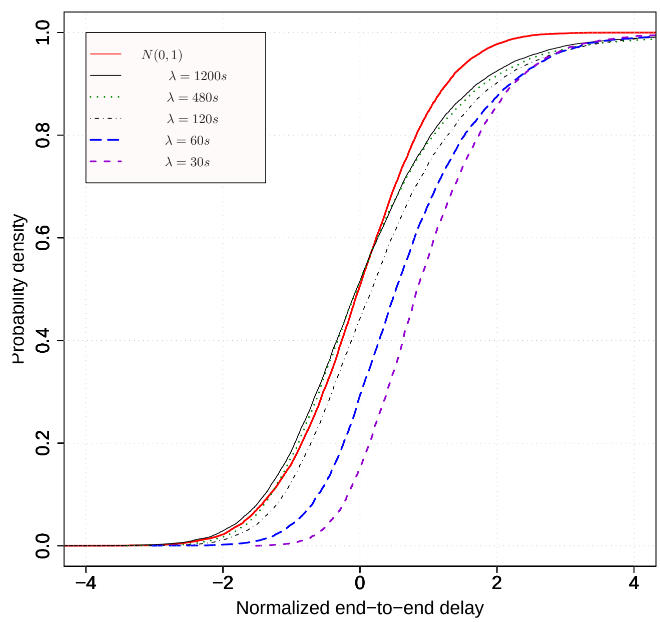

Fig. 8. Estimated cdfs vs N(0,1), with $|rp| = 10$

a maximum value close to $40ms$.

Note that the α value performing the EWMA is responsible, in a certain way, of this effect. A lower α acts as a filter for higher sampled values and hence, reduces the tail on the right side. However, this also affects the sample variance s^{2*} as the estimated values get closer to each other. Thus, low values of α introduce a distortion on the estimated distribution which results in ”thinner” curves. On the other hand, higher values of α reduce the smoothing effect of the EWMA but produce a more accurate estimation of the sample variance. This is reflected on the peak of the estimated distribution, although, at the same time, produces thicker distribution shape. Based on experience and previous simulations [22], we chose $\alpha = 0.9$, which has provide accurate estimation without introducing excessive distortions on the final distributions.

λ N(0,1)	I_1 (68%)	I_2 (27%)	I_3 (4.2%)	I_4 (0.2%)
30	66.5% (-1.5)	21.5% (-5.5)	6.8% (+2.6)	5.2% (+5)
60	62.2% (-5.8)	24.6% (-2.4)	7.1% (+2.9)	6.1% (+5.9)
120	61.1% (-6.9)	27.1% (+0.1)	7% (+2.8)	4.8% (+4.6)
480	53.3% (-14.7)	27% (=)	8% (+3.8)	7.7% (+7.5)
1200	50.8% (-17.2)	25.6%(-1.4)	11.9% (+7.7)	11.8% (+11.6)

TABLE I
PERCENTAGE OF HITS PER σ-INTERVAL WITH PATH LENGTH 5. IN BRACKETS, DEVIATION WITH RESPECT TO $N(0, 1)$.

λ N(0,1)	I_1 (68%)	I_2 (27%)	I_3 (4.2%)	I_4 (0.2%)
30	55.7% (-12.3)	30.1% (+3.1)	11% (+6.8)	3.2% (+3)
60	62.4% (-5.6)	24.9% (+2.1)	9.2% (+5)	3.5% (+3.3)
120	62% (-6)	27.1% (+0.1)	7.4% (+3.2)	3.6% (+3.4)
480	61.4% (-6.6)	28.5% (+1.5)	6.9% (+2.7)	3.2% (+3)
1200	60.7% (-7.3)	28.9% (+1.9)	7.6% (+3.4)	2.8% (+2.6)

TABLE II
PERCENTAGE OF HITS PER σ-INTERVAL WITH PATH LENGTH 10. IN BRACKETS, DEVIATION WITH RESPECT TO $N(0, 1)$.

Figure 4 shows the same results for the case of $|rp = 10|$. This results do not differ much from the previous ones, except that it is noticeable that the estimated *pdf* is slightly more centered than it was in the previous case. This again, is not an unexpected result as it was already expected that longer paths would produce better estimated distributions. However, it is remarkable that even with paths as short as five hops it is possible to obtain relatively accurate results.
Figure 5 and Figure 6 show respectively the probability density and cumulative distribution functions (i.e. *pdf* and *cdf*) of all cases with $|rp| = 5$ and variations in the cross traffic (λ parameter).

As can be observed, accuracy increases proportionally to the cross-traffic parameters. This is due to the fact that the higher the amount of messages going through the network, the more frequently intermediate nodes refresh their local estimations. In other words, if the traffic is too low, the estimated values at the arrival of a message loose accuracy by the time the next message is received.

In Figure 6, the “lower peak” described before can be appreciated from the point of view of the estimated probability. The higher part of the curve is visibly below the reference curve, which means that the estimation becomes pessimistic (i.e. the method will predict a lower probability for delays above the expected end-to-end delay). However, the same does not happen, except for the case of very low traffic, in the lowest part of the curve. This means that the estimated probability for end-to-end delays below the expected value do not over-estimate the capacity of the path.

It is important to remember at this point that the objective of this method is not to estimate the probability of individual message delays but of delays falling within a time interval. Thus, deviations with respect to the reference distribution are, in general, acceptable.

Figures 7 and 8 repeat the same experiment with a path length $|rp| = 10$. In this case, a general better fitting of the estimated curves, as suggested in Figure 4, is visible. It has been already argued that longer paths are expected to produce more accurate results. However, the curves for the experiments with higher levels of cross-traffic $\lambda = 30, 60$ draw the attention both for their accuracy with respect to the shape as well as for being shifted to the right. In Figure 8 this shift clearly shows a constant underestimation of the end-to-end delay (i.e. pessimistic predictions).

In this case, the difference with respect to the reference curve become relevant, as the estimation will be sensibly pessimistic. The explanation of this effect lies in the higher amount of missed acknowledgments for this experiment. When an acknowledgment is missed, the sender considers that the message was not received, and hence proceeds with its re-transmission. However, the message was properly delivered and the receiver is ready to forward it further. The result is that the calculated latency of the message at the sender node is notably worse than the real delay experienced by the message. Such phenomenon are expected to happen in WSN, and this result shows that measures must be taken to countermeasure its effects.

Table I and Table II present the results for the second test with the reference to the standard Normal in brackets. Again, the tail effect can be seen as the interval I_4 receives significantly more hits than expected. Similarly, interval I_1 reflects a lower percentage of hits, which agrees with the previous figures.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented a new approach to probabilistic timeliness performance metric in Wireless Sensor Networks. The paper motivates the use of probabilistic approaches instead of methods aiming at hard real-time by means of adding constraints and hence reducing its applicability.

We introduced a generalized notion of timeliness which allows capturing the level of confidence for the temporal performance and a probabilistic method which allows the estimation of end-to-end delays. It estimates the density function of the end-to-end latency of a routing path with low computational demands. The analysis of single-hop message forwarding latencies is combined into a metric which evaluates the probability of a sequence of messages achieving end-to-end latencies within a time interval.

Simulations results for a set of different scenarios underline the validity of this method.

Future work in this area includes the consideration of global energy consumption (i.e. energy-timeliness trade-offs), study of node configurations to achieve local improvements on the metric values (e.g. back-off exponents, size of preambles, etc) as well as the adaptation of existing routing protocols to take advantage of this metric. Additional possibilities to tune the presented method, such as the dynamic adaptation of α, countermeasures for missed acknowledgments, as well as the detection of congestion based on the local information are also being explored.

VII. Acknowledgment

The authors wish to express their gratitude to Jean-Dominique Decotignie, Peter van der Stok, Marc Aoun, and Jérôme Rousselot for the positive discussions and valuable feedback on the work presented in this paper. Special thanks to Jérôme for his work on the Omnet++ models used on the evaluation of this paper. Many thanks to members of the real-time systems group at TU Kaiserslautern: Raphael Guerra, Rodrigo Coelho, Alexander Neundorf and Anand Kotra, for the constructive discussions and reviewing of the paper.

References

[1] 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.

[2] G. C. Buttazzo, *Hard Real-Time Systems: Predictable Scheduling Algorithms and Applications*, second edition ed., Springer, Ed., 2005.

[3] G. Zhou, T. He, S. Krishnamurthy, and J. A. Stankovic, "Impact of radio irregularity on wireless sensor networks," in *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM, 2004, pp. 125–138.

[4] J.-D. Decotignie, "Keynote: Real-time and wireless sensor networks: Why do we need another view at it?" in *Proceedings of the 7th International Workshop on RealTime Networks (RTN) (abstract)*, July 2008, pp. 6–7.

[5] R. Serna Oliver and G. Fohler, "A proposal for a notion of timeliness in wireless sensor networks," in *Proceedings of the 8th International Workshop on Real-Time Networks RTN'09*, June 2009.

[6] Y. Li, C. S. Chen, Y.-Q. Song, and Z. Wang, "Real-time qos support in wireless sensor networks: a survey," in *Proceedings of the 7th IFAC, International Conference on Fieldbuses and Networks in Industrial and Embedded Systems - FeT'2007*, 2007.

[7] O. Chipara, Z. He, G. Xing, Q. Chen, X. Wang, C. Lu, J. Stankovic, and T. Abdelzaher, "Real-time power-aware routing in sensor networks," in *Proceedings of the 14th IEEE International Workshop on Quality of Service (IWQoS).*, June 2006.

[8] T. He, J. Stankovic, C. Lu, and T. Abdelzaher, "Speed: A stateless protocol for real-time communication in sensor networks," in *Proceedings of ICDCS'03*, May 2003.

[9] A. Sahoo and P. Baronia, "An energy efficient mac in wireless sensor networks to provide delay guarantee," in *Proceedings of the 15th IEEE Workshop on Local and Metropolitan Area Networks. LANMAN*, June 2007.

[10] K. Karenos and V. Kalogeraki, "Real-time traffic management in sensor networks," in *Proceedings of the 27th IEEE Real-Time Systems Symposium (RTSS)*, 2006.

[11] L. Abeni and G. Buttazzo, "Qos guarantee using probabilistic deadlines," in *Real-Time Systems, 1999. Proceedings of the 11th Euromicro Conference on*, 1999, pp. 242–249.

[12] E. Felemban, C.-G. Lee, E. Ekici, R. Boder, and S. Vural, "Probabilistic qos guarantee in reliability and timeliness domains in wireless sensor networks," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies. INFOCOM05*, vol. 4, March 2005, pp. 2646–2657.

[13] B. Hughes, , B. Hughes, and V. Cahill, "Achieving real-time guarantees in mobile ad hoc wireless networks," in *Proceedings of the Work-in-Progress Session 24th IEEE Real-Time Systems Symposium. RTSS04*, 2004.

[14] S. Gobriel, R. Cleric, and D. Mosse, "Adaptations of tdma scheduling for wireless sensor networks," in *Proceedings of the 7th International Workshop on Real-Time Networks (RTN)*, 2008.

[15] M. Caccamo, L. Zhang, L. Sha, and G. Buttazzo, "An implicit prioritized access protocol for wireless sensor networks," in *Proceedings of the 23rd IEEE Real-Time Systems Symposium (RTSS)*, 2002.

[16] T. Watteyne and I. Auge-Blum, "Proposition of a hard real-time mac protocol for wireless sensor networks," in *Proceedings of the 13th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*. IEEE Computer Society, 2005.

[17] T. Abdelzaher, S. Prabh, and R. Kiran, "On real-time capacity limits of multihop wireless sensor networks," in *Proceedings of the IEEE International Real-Time Symposium (RTSS)*, 2004.

[18] L. Burgstahler and M. Neubauer, "New modification of the exponential moving average algorithm for bandwidth estimation," *ITC Specialist Seminar*, 2002.

[19] E. Goldoni and G. Rossi, "Improving available bandwidth estimation using averaging filtering techniques," Università degli Studi di Pavia, Laboratorio Reti, Tech. Rep., 2008.

[20] M. G. Bulmer, *Principles of Statistics*, 1979th ed., Dover, Ed., 1967.

[21] T. Korkmaz and M. Krunz, "Bandwidth-delay constrained path selection under inaccurate state information," in *Proceedings of the IEEE/ACM Transactions on Networking*, June 2003.

[22] R. Serna Oliver, "Estimation of the probability density function of end-to-end delays in wireless sensor networks," Technische Universität Kaiserslautern, Tech. Rep., 2009.

[23] T. Birkel, "A functional central limit theorem for positively dependent random variables," *Journal of Multivariate Analysis*, pp. 314–320, 1993.

[24] J. Al-Karaki and A. Kamal, "Routing techniques in wireless sensor networks: a survey," *Wireless Communications, IEEE*, vol. 11, no. 6, pp. 6–28, Dec. 2004.

[25] "Omnet++ discrete event simulation environment," 2009. [Online]. Available: http://www.omnetpp.org

[26] A. Koepke, "Mobility framework: A framework to support simulations of wireless and mobile networks within omnet++." [Online]. Available: http://mobility-fw.sourceforge.net/

[27] A. El-Hoiydi and J.-D. Decotignie, "Wisemac: An ultra low power mac protocol for the downlink of infrastructure wireless sensor networks," in *Proceedings of the Ninth IEEE Symposium on Computers and Communication (ISCC)*, June 2004.

[28] "Wirelessly Accessible Sensor Populations (WASP)," 2009. [Online]. Available: http://www.wasp-project.org/

[29] Y. Xiao, K. Thulasiraman, and G. Xue, "Approximation and heuristic algorithms for delay constrained path selection under inaccurate state information," in *Proceedings of the International Conference on Quality of Service in Heterogeneous Wired/Wireless Netowrks (QSHINE)*, 2004.

[30] R. Serna Oliver and G. Fohler, "Probabilistic routing for wireless sensor networks," in *Proceedings of Work-in-Progress Session, 29th IEEE Real-Time Systems Symposium 2008*, December 2008.