Markov Chain-based Optimization of Multihop IEEE 802.15.4 Wireless Sensor Networks

Stefano Busanelli
Wireless Ad-hoc and Sensor Networks (WASN) Lab
Department of Information Engineering, University of Parma, Italy
busanelli@tlc.unipr.it

Marco Martalò
Wireless Ad-hoc and Sensor Networks (WASN) Lab
Department of Information Engineering, University of Parma, Italy
martalo@tlc.unipr.it

Gianluigi Ferrari
Wireless Ad-hoc and Sensor Networks (WASN) Lab
Department of Information Engineering, University of Parma, Italy
gianluigi.ferrari@unipr.it

ABSTRACT

In this work, we propose an optimization framework for the IEEE 802.15.4 medium access control (MAC) protocol. More precisely, we derive a theoretical tool providing reliable guidelines for tuning the parameters of the MAC protocol. The presented tool could be used in two different directions: (i) for fixed network topology, it might be of interest to determine the MAC protocol configuration able to guarantee the best performance according to some quality of service metric; (ii) for fixed parameters of the MAC protocol, it might be of interest to determine the optimal network topology. Both these situations appear in practical situations. In particular, the first scenario happens in all cases where one cannot decide a-priori the node positions, because of some randomness or for some physical constraints, so that the only degree of freedom is given by the MAC protocol itself. The second set of scenarios occurs in the circumstances where one can decide the nodes displacement, thus introducing more degrees of freedom, but making also the problem solution more complicated. The proposed optimization tool applies some classical operative research instruments to a recently proposed Markov-Chain based model that has shown to be suitable for the performance analysis of a generic Cluster-Tree (CT) multihop IEEE 802.15.4 network. We will also show that this tool could be effectively used in a real scenario with a low-cost low-energy hardware platform.

Categories and Subject Descriptors

G.3 [**Probability and statistics**]: Queuing theory, stochastic processes; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Performance evaluation

Keywords

Markov chains, IEEE 802.15.4, relay, buffer, medium access control (MAC) protocol, cluster tree (CT) networks.

WSNPERF 2009, October 23, Pisa, Italy

DOI 10.4108/ICST.VALUETOOLS2009.7623

1. INTRODUCTION

Despite the significant attention received by the research community and the existence of a well designed standard as the IEEE 802.15.4 [1], nowadays the market penetration of wireless sensor networks (WSNs) is still far from the expected objectives. In order to explain this relative failure, one should think that the IEEE 802.15.4 standard effectively addresses some key issues common to any WSN application, as low cost and network reliability, but at the same time this standard cannot effectively satisfy the specific requirements of any possible WSN-based application. From another perspective, the existence of a very general standard as IEEE 802.15.4, surely contributes to significantly reduce the cost of the hardware platform. On the other hand, we feel that the IEEE 802.15.4 standard is too general, thus complicating the design process and hindering a wide spread of the WSNs.

The main goal of this work is to provide some "easy-to-follow" design guidelines, to be used in the "customization" process needed by any WSN-based application. A step in this direction is offered by several analytical frameworks proposed for performance evaluation (at the MAC level) of IEEE 802.15.4 networks [7, 2, 6, 10, 8]. Until now, the majority of these models have focused on star topologies, while versions suitable to multi-hop networks are rare [4]. Moreover, to the best of our knowledge we are not aware of any attempt to define an optimization process of the IEEE 802.15.4 MAC protocol. By leveraging on the Markov chain-based general model proposed in [4], in this work we propose an innovative optimization procedure for the design of the MAC protocol of IEEE 802.15.4 networks.

This paper is structured as follows. In Section 2, we briefly describe the main characteristics of the IEEE 802.15.4 MAC protocol. In Section 3, we provide the reader with a quick description of the analytical model for performance evaluation of CT IEEE 802.15.4 networks proposed in [4]. In Section 4, the results of our performance optimization strategy are presented. Finally, concluding remarks are given in Section 5.

2. IEEE 802.15.4 STANDARD OVERVIEW

The IEEE 802.15.4 standard refers to the first two layers of the ISO/OSI stack (i.e., physical and MAC) and guarantees (in principle) a transmission data-rate equal to 250 kpbs in a wireless communication link. The IEEE 802.15.4 networks can operate in two modes: asynchronous, denoted as "unslotted," and fully synchronized, denoted as "slotted." In the slotted mode, which we focus on, the nodes share a common time reference provided by the co-

ordinator. Two consecutive synchronization packets, denoted as "beacons," delimit a so-called "superframe," which is further divided into two regions, referred to as active and inactive, respectively. While in the latter region all nodes go to the sleeping state to preserve their battery energies, in the former region all nodes can transmit their data packets. During the active region, we consider the use of Contention Access Period (CAP), where every node can transmit according to a slotted non-persistent Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) MAC protocol, with the use of a proper Binary Exponential Backoff (BEB) algorithm. A Contention Free Period (CFP), with Guaranteed Time Slot (GTS), can also be used. To limit the complexity of the model, we set to zero the duration of the CFP, so that the structure of the superframe is completely determined by the following parameters.

- The duration of the superframe, expressed in terms of numbers of symbol durations, which can assume values in the set $\{0,1,\ldots,14\}$ and is proportional to 2^{BO}, where BO is the beacon order.[1]

- The duration of the active period, which can assume values in the set $\{0,1,\ldots,14\}$ and is proportional to 2^{SO}, where SO is the superframe order.

These values are chosen by the coordinator and notified to the other nodes through the beacon packets. Unlike [6], we do not consider the presence of an inactive region (hence, BO = SO) and the use of ACK messages. Finally, in order to further reduce the complexity in both the analytical framework and in the simulations we ignore the InterFrame Space (IFS), which is a pre-defined waiting time between two consecutive transmission acts. However, our framework can be easily extended to take into account the presence of the IFS.

The used (slotted) CSMA/CA MAC protocol relies on a discretization of the time into (backoff) slots, each with a length equal to 20 times the duration of a physical symbol. Every node can attempt to get control of the channel for a maximum number of times, denoted as m. The default value of m is 4. Before the k-th attempt, the node has to wait for a random waiting time, denoted as W_k, whose length (expressed in terms of time slots) is uniformly distributed in the set $\left\{0,1,2,\ldots,2^{\mathrm{BE}_k}-1\right\}$, $k=1,\ldots,m+1$, where the variables $\{\mathrm{BE}_k\}$ are generated according to the following rule:

$$\mathrm{BE}_{k+1} = \min\left\{\mathrm{BE}^{\min}+k, \mathrm{BE}^{\max}\right\} \qquad k=0,\ldots,m.$$

In the IEEE 802.15.4 standard, the default values $\mathrm{BE}^{\min}$ and $\mathrm{BE}^{\max}$ are fixed to 3 and 5, respectively.

At the end of the k-th waiting period, the node assesses the status of the channel through a Clear Channel Assessment (CCA) operation during the first 8 time symbols of the successive slot. If the channel is busy and the k-th attempt fails, the node will start the waiting time of the $(k+1)$-th backoff cycle (if $k+1 \leq m$); otherwise, the node will perform a second CCA operation in the successive slot. We denote as the k-th backoff cycle the union of the k-th double CCA operations and the k-th waiting time. Since in our model the time unit is given by a backoff slot, we ignore the time needed by the transceiver to switch between transmission and reception modes—this time has to be shorter than 12 time symbols. Finally, observe that even if the channel access attempt succeeds, the transmission can be delayed to the successive superframe if in the current one there is not enough time to complete the transmission.

[1]Note that the value BO = 15 is used for the unslotted mode [1].

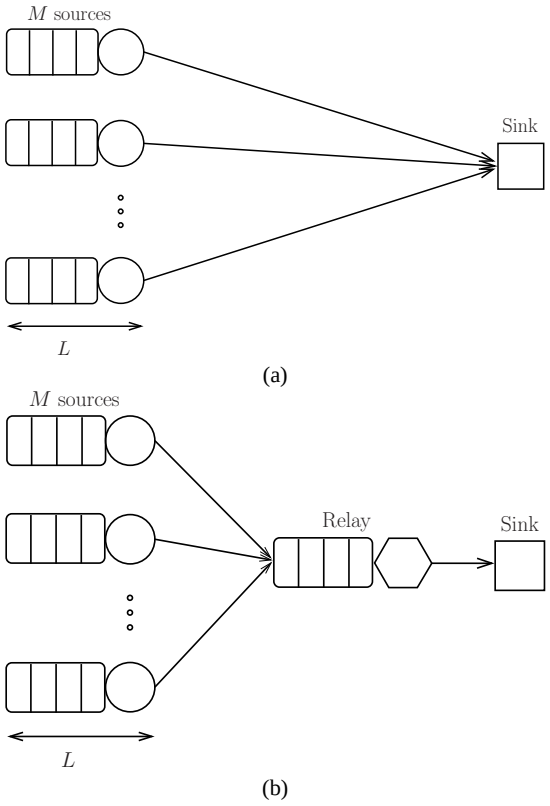

Figure 1: Network scenarios of interest for (a) star topology and (b) two-hop communications.

3. MARKOV CHAIN-BASED PERFORMANCE ANALYSIS

In this section, we briefly recall a few basics of the framework proposed in [4]. Preliminary scenarios of interest for 1-hop and 2-hop networks are shown in Figure 1 (a) and Figure 1 (b), respectively. All nodes have a finite dimension buffer (at the MAC level), with the exception of the sink, which is assumed to have an infinite buffer. The number of sources is denoted as M.

3.1 Single-hop Network Scenarios

Our analytical model is a direct extension of that proposed in [8] for "bufferless" scenarios. In this case, there is no queue for storing packets waiting to be scheduled for transmission and a generated packet can be stored at the node only if no other packet is currently stored. For the sake of clearness, we now summarize the main assumptions behind our model.

- All the sources generate traffic according to the same distribution. Therefore, it is possible to study a "tagged" node to characterize all source nodes [9].

- As stated in Section 2, if the remaining time in the current superframe is not sufficiently long for a data packet transmission to be completed, the node must defer it to the next superframe. When the duration of the superframe is sufficiently long and the traffic load is not too high, the above situation rarely appears and, therefore, will be neglected in our analytical model. Clearly, if the value of BO is too small

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

this assumption is no longer valid. In particular, the analytical model will be shown to be accurate for values of BO larger than 5.

- The probability that the channel is sensed idle in a given backoff slot is approximated by the steady state probability that the channel is idle. This assumption has already been considered in the literature [12]. We remark that the channel idleness in one slot is not independent of that in the consecutive slot, and this will be taken into account in our model.

- The probability that a node begins its transmission in a generic slot is approximated by the steady state probability that a node transmits.

On the basis of these assumptions, it is possible to identify two Semi-Markov Processes (SMPs): one for the shared radio channel and one for the tagged node. It is well known that an SMP always embeds a DTMC. Moreover, if an SMP is positive recurrent, irreducible, and aperiodic, it is sufficient to know the average sojourn time of every state of the SMP and the limiting distribution of the embedded DTMC to compute the average fraction of time that the SMP spends in each state [3]. Therefore, we can accurately model the system behavior by studying the embedded DTMCs of the channel and of the tagged node. These DTMCs are mutually coupled, as will be shown later.

3.1.1 Tagged Node DTMC

With respect to the original model proposed in [8] for bufferless scenarios, we have proposed the following modifications [4]. First, the assumption of geometric distribution for the waiting time can be relaxed, since the SMP theory states that the knowledge of the average sojourn time $\overline{W_k} \triangleq \mathbb{E}\{W_k\}$, $k \in \{1,\ldots,m+1\}$ allows to derive the long-term behavior of the parent SMP [3]. Due to this reduction of the number of states and transitions in the SMP, the embedded DTMC becomes simpler as well. Then, finite size buffers, all with length L, are inserted at the MAC level of sources and relay. As in classical queuing theory, L is the capacity of the entire node, formed by a queue of size $L-1$ and a server able to process one packet at a time. Therefore, a node can store L packets. The model proposed in [8] does not allow to handle the presence of a buffer with a finite size L. However, the key assumptions at the beginning of Subsection 3.1, namely the facts that (i) every source node has the same probability to access the channel in every time slot and (ii) the probability that the channel is idle in a certain time slot is always the same, lead to a decorrelation between the nodes' queues. Therefore, it is possible to model the queue at the tagged node as a Geo/G/1/L queue with parameter π_0^{n}. This corresponds to the probability that the queue is empty after a packet departure, either successfull or unsuccessful (because of a channel access failure).

Due to the assumptions previously made, the modifications in the node DTMC do not affect the "nature" of the process governing the behavior of a source node, i.e., this process remains an SMP. For this reason, an embedded DTMC can still be extracted. Moreover, this DTMC is ergodic, since it is irreducible, aperiodic, and finite [3]. In Figure 2, a simplified representation of the DTMC model for the buffered tagged node is shown. Note that the macrostate CSMA embeds all the workflow of the CSMA/CA MAC protocol shown in [4]. When π_0^{n}, $p_{\mathrm{i}}^{\mathrm{c}}$, and $p_{\mathrm{i|i}}^{\mathrm{c}}$ are known, using proper balance equations the limiting distribution of the tagged node DTMC, denoted as $\boldsymbol{\pi}^{\mathrm{n}}$, and of the long-term sojourn times of the SMP, denoted as $\boldsymbol{\Pi}^{\mathrm{n}}$, can be determined. This computation can be found in [4].

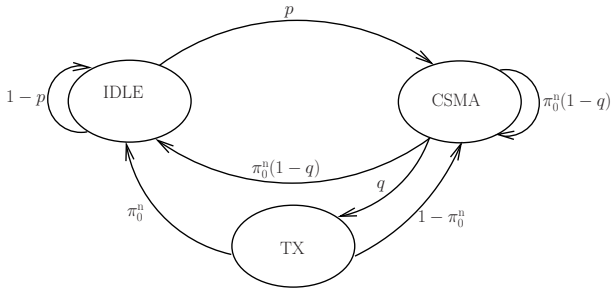

Figure 2: DTMC model for a *buffered remote node* compliant with the IEEE 802.15.4 networks.

To determine the distribution of the number of costumers in the queue of the tagged node in correspondence to a slot boundary, the expression of the packet service time is needed. This time is given by the sum of the time necessary to transmit a packet and the time spent to access the channel. Since the packet size is fixed, the distribution of the packet service time depends only on $p_{\mathrm{i}}^{\mathrm{c}}$ and $p_{\mathrm{i|i}}^{\mathrm{c}}$. The packet service time is given by two distinct components: the first, denoted as $T_{\mathrm{succ}}(z)$, accounts for the successfully transmitted packets, while the second, denoted as $T_{\mathrm{fail}}(z)$, accounts for the discarded packets because of a CSMA/CA MAC protocol failure. Using the approach proposed in [6], after a few manipulations, it can be shown that the probability generating function (PGF) of the packet service time (denoted as $T_{\mathrm{t}}(z)$) can be expressed as

$$\begin{aligned} T_{\mathrm{t}}(z) \;=\; & \underbrace{T_{\mathrm{tx}}(z)\,p_{\mathrm{i}}^{\mathrm{c}}\,p_{\mathrm{i|i}}^{\mathrm{c}} \sum_{k=1}^{m+1} \left(1-p_{\mathrm{i|i}}^{\mathrm{c}}\,p_{\mathrm{i}}^{\mathrm{c}}\right)^{k-1} \prod_{j=1}^{k} B_j(z)}_{T_{\mathrm{succ}}(z)} \\ & + \underbrace{\left(1-p_{\mathrm{i|i}}^{\mathrm{c}}\,p_{\mathrm{i}}^{\mathrm{c}}\right)^{m+1} \prod_{j=1}^{m+1} B_j(z)}_{T_{\mathrm{fail}}(z)} \end{aligned} \tag{1}$$

where $B_k(z)$ is the k-th backoff cycle duration.

From (1), using the procedure described in [11], one can derive the distribution of the number of customers that are queued at the time of a packet departure. This vector of size L is denoted by $\boldsymbol{\pi}^{\mathrm{n}} = [\pi_0^{\mathrm{n}}\pi_1^{\mathrm{n}}\ldots\pi_{L-1}^{\mathrm{n}}]$, where π_k^{n} is the probability that the queue of the tagged node contains k packets immediately after a packet departure ($k \in \{0,1,\ldots,L-1\}$).

3.1.2 Channel DTMC

In Figure 3, the DTMC model for the physical communication channel is shown. As one can see, in this case there are four states, denoted as IDLEIDLE, SUCCESS, IDLE, and FAILURE. The first state of the channel DTMC is denoted as IDLEIDLE, because before every transmission the channel has to remain idle for two consecutive slots, as stated by the IEEE 802.15.4 standard [1]. When a new packet is transmitted, the channel schedules a new transmission with probability $1-\alpha$, where $\alpha \triangleq (1-p_{\mathrm{t|ii}})^M$ is the probability that no source has transmitted given that the channel was idle during the two previous slots, $p_{\mathrm{t|ii}}$ is the probability that the tagged node transmits given that the channel was idle during the two previous slots, and M is the number of sources. Conversely, if just a single node transmits, the channel moves to the SUCCESS state: this event happens with probability $\beta \triangleq M p_{\mathrm{t|ii}}^{\mathrm{n}}(1-p_{\mathrm{t|ii}}^{\mathrm{n}})^{M-1}$. Finally, if more than one nodes transmit at the same time, the channel

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

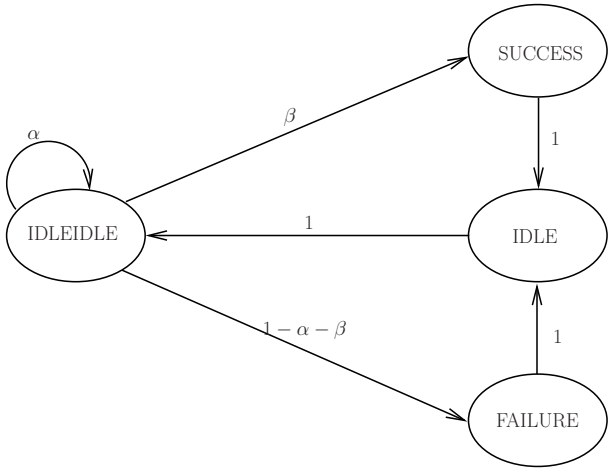

Figure 3: DTMC model for the *physical communication channel* in IEEE 802.15.4 networks.

moves to the FAILURE state: the probability of this event is clearly $1-\alpha-\beta$. After a packet transmission has been carried out, either successfully or unsuccessfully, the channel moves to the IDLE state with probability 1 and then to the IDLEIDLE state with probability 1. The presence of the IDLE state is expedient for modelling the IEEE 802.15.4 standard, where a few free time slots are inserted before a new superframe is scheduled.

Given the values of α and β (obtained by solving the tagged node DTMC), it is possible to derive the closed-form limiting distribution of the channel DTMC, denoted as $\boldsymbol{\pi}^{\mathrm{c}}$, and of the long-term sojourn times of the SMP, denoted as $\boldsymbol{\Pi}^{\mathrm{c}}$, as shown in [4].

Note that we are implicitly assuming that every node sees itself as a "competitor" for accessing the channel. Clearly, this is unrealistic, since in a real network a node cannot compete with itself. However, this assumption is expedient to allow us to extend our analytical model to scenarios with multi-hop communications or heterogeneous networks, i.e., where nodes or group of nodes may have different medium access strategies.

3.1.3 Throughput and Delay

As mentioned in the previous subsection, the tagged node and channel DTMCs are coupled. Their distributions can then be obtained by solving proper fixed-point equations, as explained in detail in [4]. Once the distributions of these DTMCs and the long-term sojourn times in the states of their parent SMPs are computed, the system performance, in terms of throughput and delay, can be evaluated as follows.

According to the formulation used in [8], the throughput coincides with the average fraction of time (over a sufficiently long time horizon) spent in the SUCCESS state by the channel SMP and is given by

$$\Pi_{\mathrm{S}}^{\mathrm{n}} = \frac{N\beta}{1+(N+1)(1-\alpha)}. \tag{2}$$

However, in order to use a common definition for both single hop and multihop networks, we define the throughout as

$$S \triangleq M p_{\mathrm{t}}^{\mathrm{n}} (1-p_{\mathrm{t}|\mathrm{ii}}^{\mathrm{n}})^{M-1} \tag{3}$$

where $p_{\mathrm{t}}^{\mathrm{n}} = \Pi_{\mathrm{TX}}^{\mathrm{n}}$ is the fraction of time spent by the tagged node SMP in the TX state. Equation (3) simply states that it is possible to obtain the aggregate network throughput by simply multiplying the per-node throughput by M, since the former is given by the fraction of time spent by a single node in the TX state multiplied by the probability that the others $M-1$ nodes do not transmit in the same slot. We remark that in 1-hop scenarios, our definition (3) for the throughput leads to the same result obtained using the definition proposed in [8] and given by (2).

In the computation of the delay, we will take into account only the packets which successfully access the channel through the CSMA/CA MAC protocol mechanism. Note that we assume that the delay experienced by a collided packet is the same of a packet successfully transmitted, since we are not considering the use of ACK messages. Once the distribution of the tagged node queue at the time of a packet departure (i.e, $\boldsymbol{\pi}^{\mathrm{n}}$) is known, using Little's law the average waiting time can be computed as

$$\overline{W} = \frac{\sum_{\ell=1}^{L-1} \ell \pi_\ell + L(\pi_o + p\overline{T}_{\mathrm{t}} - 1)}{p} - \overline{T}_{\mathrm{t}} \tag{4}$$

where $\overline{T}_{\mathrm{t}} = T_{\mathrm{t}}^{(1)}(z)\Big|_{z=1}$ is the average packet service time. The packets dropped due to CSMA/CA MAC protocol failures experience a different average *service* time ($\overline{T}_{\mathrm{fail}}$) with respect to the transmitted (successfully or not) packets ($\overline{T}_{\mathrm{succ}}$). However, the average *waiting* time is the same for all packets. Therefore, from (1) and (4) one can evaluate the average delay of a successfully transmitted packet as follows:

$$\overline{D} = \overline{W} + \overline{T}_{\mathrm{succ}}. \tag{5}$$

3.2 Two-hop Network Scenarios with Single Relay

In this subsection, we show how to modify the proposed DTMC models for 1-hop scenarios, in order to take into account the presence of an intermediate relay. Without proper approximations, almost every analytical model will become mathematically intractable. Roughly speaking, our modelling goal consists in reaching the best tradeoff between analytical complexity and accuracy of the results. The following assumptions are then considered.

- Due to the Geo/G/1/L queuing theory, it is possible to determine the output process of source nodes [11]. However, we are not aware of any work which successfully characterizes the output process of a wireless channel using the CSMA/CA MAC protocol [9]. Therefore, we arbitrarily assume that the arrival process at the relay has a Bernoulli distribution with a parameter, denoted as p_{r}, which is different with respect to the parameter p of the Bernoulli traffic distribution at each source node.
- Since the packet size is fixed and owing to the assumption that every node experiences the same long-term channel condition, the packet service time is assumed to be the same for relay and sources. Note, however, that the corresponding queues have different behaviors since their arrival processes are different.
- We assume that the numbers of customers in different queues are independent and, therefore, we independently determine their stationary distributions.
- The service time is independent of the arrivals. In fact, if a packet experiences a short channel access delay at the source nodes, no assumption about the channel access delay at the relay can be made.

Due to the above assumptions, we can study a two-hop scenario by simply adding a relay DTMC equal to a source DTMC, except for

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

a different parameter of the per slot Bernoulli traffic distribution ($p_r \neq p$). Moreover, the relay is equipped with a Geo/G/1/L queue equal to that of a source except for the arrival process parameter. Therefore, we are in the presence of three DTMCs (one for the tagged source node, one for the relay, and one for the channel) and two Geo/G/1/L queues (one for the tagged source node and one for the relay) mutually coupled. While in the one-hop scenario there is a single unknown parameter p_i^c, there are now two unknown parameters: p_i^c and p_r. In this case, a vectorial fixed point equation of the following form can be derived:

$$(p_i^c, p_r) = \Psi(p_i^c, p_r)$$

where $\Psi(p_i^c, p_r)$ depends on the structures of the tagged node block (DTMC and queue), the relay block (DTMC and queue), and the channel DTMC.

Upon the derivation of p_i^c and p_r, the long-term sojourn times of the SMPs of the tagged node and the relay, the throughput and delay performance of the considered networks can then be evaluated. As in Subsection 3.1.3, here as well we define the throughput S as

$$S = p_t^r (1 - p_{t|ii}^n)^M \tag{6}$$

where p_t^r is the long-term fraction of time spent by the relay in the TX state and $1 - p_{t|ii}^n$ is the probability of correct transmission (i.e., when all the M source nodes do not transmit). In this case, the per-node throughput of the relay, given by (8) also coincides with the aggregate network throughput, since just the packets correctly transmitted by the relay itself are received at the sink.

The average end-to-end delay is simply obtained by adding the average delay experienced by the tagged node, denoted as $\overline{D}^n$, and the average delay experienced by the relay node, denoted as $\overline{D}^r$:

$$\overline{D} = \overline{D}^r + \overline{D}^n.$$

3.3 Cluster Tree Networks

We now consider a general CT network, where the traffic generated at the sources (leafs) flows towards the sink (root), through a series of intermediate relays. In particular, each relay receives packets coming from a specific cluster of sources. At the same time, the relays may be grouped into higher level clusters, which can be associated with even higher level relays or by the sink itself. By grouping the sources and the relays at the various hierarchical levels, one obtains the CT network depth. For example, the network in Figure 4 is a particular CT network with a depth equal to 3 and the following hierarchical levels: the first level contains $M = 12$ sources, the second level is composed by the single relay, and the last level contains just the sink. We define N_{relays} as the number of relays level in the CT, e.g. in a cluster of depth equal to 3, $N_{relays} = 2$. Another example of CT network topology is shown in Figure 5, where $M = 16$ source nodes are grouped into 4 clusters composed by 4 sources each. Each cluster belongs, respectively, to one of the 4 first level relays. The network topology is then completed by 2 second level relays and by the sink. In this case, the CT is characterized by $N_{relays} = 2$ level of relays.

Following the same approach used in Subsection 3.2 for the two-hop networks, it is possible to extend the proposed performance evaluation framework to this more general CT scenario. In particular, we assume that all relays can be modeled with the same coupled DTMC-Geo/Geo/1/L queue block. In fact, we can state that all the relays "see," on average, the same channel and, therefore, the probability of finding the channel idle is the same for all relays. The only difference between the relays is given by the parameter of the generated Bernoulli traffic. In particular, the k-th relay ($k = 1, \ldots, N_{relays}$) can be modeled with the same DTMC of the tagged node but replacing the parameter p with a parameter $p_r^{(k)}$. In fact, every relay receives a different amount of traffic, which depends on the nodes in its own cluster and on the corresponding amount of generated traffic.

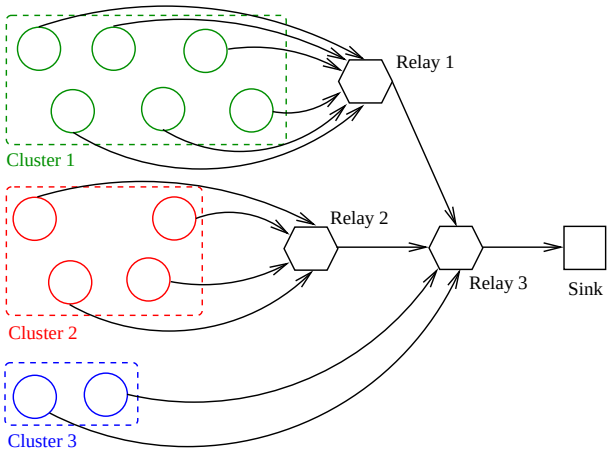

Figure 4: Cluster-tree topology with $M = 12$ sources grouped into a CT network with 3 clusters, 3 relays, and depth equal to 3.

Therefore, by using the basic node structures presented in Subsection 3.1, one can model every CT network. It is clear that the accuracy of the analytical model reduces for increasing complexity of the CT network, because of the considered simplifying assumptions. It can be easily observed that the number of unknown parameters in a CT network with increasing complexity becomes larger and larger. In a CT network with N_{relays} levels of relays, the number of unknown parameters is $N_{relays} + 1$, namely $\{p_r^{(k)}\}_{k=1}^{N_{relays}}$ and p_i^c. In this case, the following vectorial fixed-point equation needs to be solved:

$$\left(p_i^c, p_r^{(1)}, \ldots, p_r^{(N_{relays})}\right) = \Theta\left(p_i^c, p_r^{(1)}, \ldots, p_r^{(N_{relays})}\right)$$

where $\Theta(\cdot)$ depends on the structures of the tagged node block (DTMC and queue), the relays' blocks (DTMC and queue), and the channel DTMC. In a general CT network it becomes very difficult to prove the existence and the uniqueness of the solution of the fixed point equation. Since the performance results predicted by our own DTMC-based system are in a very good agreement with realistic ns-2 simulation results, we are confident of the validity of our analytical framework.

We remark that the delay is obtained by summing the delay of the tagged node of each cluster level. For example, in the case of the CT represented in Figure 4, the average end-to-end delay is

$$\overline{D} = \overline{D}^{r_2} + \overline{D}^{r_1} + \overline{D}^n \tag{7}$$

where $\overline{D}^{r_2}$ and $\overline{D}^{r_1}$ are, respectively, the average delay of the tagged node at the second and at the first relay level. However, in this case we are not interested in the aggregate throughput S, but we focus on the per-hop throughout experienced by a generic source node (e.g., the tagged node), denoted as S_n, and expressed as follows:

$$S_n = p_t^r (1 - p_{t|ii}^n)^M. \tag{8}$$

Following the approach used in Section 3.2 for the two-hop network, the aggregate throughput S could be obtained by multiplying by N_{r_2} the per-node throughput of the tagged second level relay

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

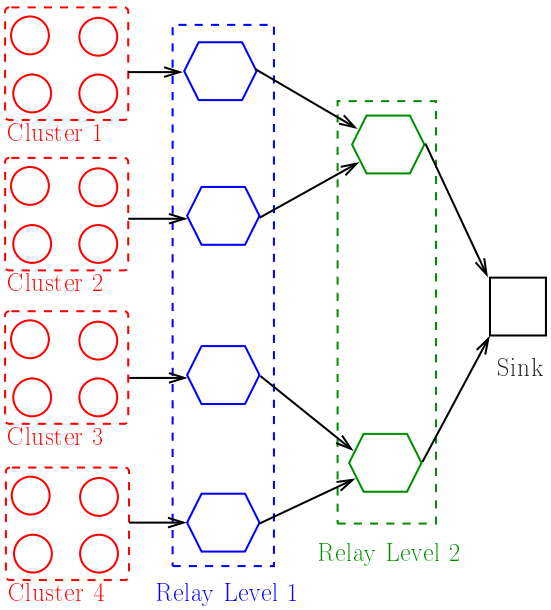

Figure 5: Cluster-tree topology with $M = 16$ sources grouped into a CT network with 4 clusters with 4 nodes each, 2 relay levels, and depth equal to 3.

(denoted as S_{r_2}):

$$S = p_t^{r_2}(1-p_{t|ii}^{n})^{M}(1-p_{t|ii}^{r_1})^{N_{r_1}}(1-p_{t|ii}^{r_2})^{N_{r_2}-1} \quad (9)$$

where the fraction of time spent by the tagged node SMP in the TX state ($p_t^{r_2} = \Pi_{TX}^{r_2}$) has been multiplied by the probability that the others M nodes, N_{r_1} first level relays and $N_{r_2}-1$ second level relayes do not transmit in the same slot. From equation (9) it is also possible to derive the ratio between the number of successfully packet transmissions and the total number of generated packets from the tagged source node, denoted as Packet Success Ratio (*PSR*),

$$PSR = \frac{S}{MNp} \quad (10)$$

where $S/(MN)$ is the average number of per-slot successfully received packets, generated by the packet node, while p is the average number of packets generated by a source in a given temporal slot. We remark that *PSR* is directly proportional to the aggregate throughput S (and, obviously, also to the per-hop throughput).

4. NUMERICAL RESULTS

4.1 Simulation Set-up

The simulations, used to verify the results predicted by our Markov chain-based framework, are performed using the built-in model, in the version 2.31 of ns-2 [13]. The traffic generation model per slot per node is Bernoulli with parameter p and each node generates packets with constant size of 100 bytes ($N = 10$ backoff slots) including the bytes of the physical and MAC layers. We evaluate the throughput and the end-to-end delay as functions of the aggregate offered load and for different values of the buffer length. In particular, we define the normalized average aggregate offered load as

$$G \triangleq M \cdot N \cdot p$$

where p, M, and N have been already introduced. Multiplying the normalized average aggregate offered load by the maximum transmission rate of the IEEE 802.15.4 standard (i.e., 250 Kb/s at 2.4 GHz) we obtain the average aggregate offered load g:

$$g = 250\ G \quad [\text{Kb/s}].$$

We recall that a portion of this aggregate offered load is dropped due to the finite length of the input buffer at the nodes.

In the analytical case, a coupled DTMC-based system is solved numerically using Matlab [5] and, then, throughput and delay are numerically evaluated.

4.2 Markov Chain-based Optimization Framework

We first evaluate the accuracy of our analytical model by varying the parameters of the CSMA/CA MAC protocol. In particular, we focus on the topologies with 1-hop and 2-hop communications presented in Figure 1, considering a finite buffer of length $L = 4$ both at the sources and relay. The CSMA/CA MAC protocol depends on four parameters: (i) BE_{min}, (ii) BE_{max}, (iii) m, and (iv) the number of CCA operations to be performed before declaring the channel idle. In [8], the authors focused on the fourth parameter, showing that in some cases the double CCA operations performed by default is a waste of time and resources. We now analyze the impact of the remaining parameters. First, the maximum number of backoff cycles is reduced to $m = 2$ (the default value is $m = 4$), in order to evaluate the performance of a network with a low delay requirement. Then, we study a network where the data reliability is more critical than the delay: in this case, we set $BE_{min} = 4$, $BE_{max} = 6$, and $m = 4$, thus increasing the average waiting time and reducing the collision probability.

In Figure 6, the delay is shown as a function of the throughput, in scenarios with (a) 1-hop and (b) 2-hop communications. In all cases, $L = 4$, $M = 12$, and various values of m, BE_{min}, and BE_{max} are considered. The curves are parametrized respect to the average normalized aggregate offered load (G). Analytical (solid lines) and simulation (dashed lines) results are shown. As one can see, the considered networks have a bimodal (stable/non-stable) behavior. We remark once more the excellent agreement between analytical and simulation results in this case as well. Therefore, the introduced approximations introduced in the analytical framework are meaningful also for varying values of the backoff parameters (m, BE_{min}, and BE_{max}).

4.3 CSMA/CA MAC Optimization

We now show how to use our analytical framework in order to optimize the CSMA/CA MAC parameters in a real case study. Let us consider a network composed by a central sink node which collects information from a $M = 16$ sensors (sources). The nodes share the same radio channel (e.g., the hidden terminal problem is not present) and the sink acts as the coordinator of the network. In this scenario, it is possible to transfer the data using one-hop communications without resorting to the use of intermediate relays. However, suppose that the network is deployed in an hostile environment, e.g., an industrial scenario, and the presence of some mobile obstacles could temporarily appear. Therefore, in order to have a robust network setup, able to operate also when some radio links are shadowed by these mobile obstacles, multi-hop communications can be established, inserting two level of relays, obtaining a CT topology, as shown in Figure 5. Clearly, the choice of using multi-hop communications degrades the network performances, since the radio resources are shared by all the nodes, but it ensures an high network reliability, also in hostile environments.

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

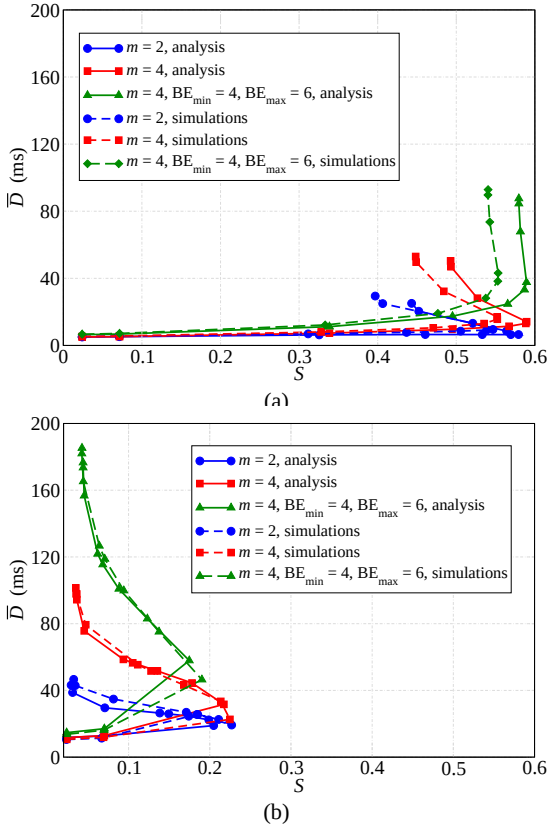

Figure 6: Delay as a function of the throughput, in scenarios with (a) 1-hop and (b) 2-hop communications. In all cases, $L = 4$, $M = 12$, and various values of m, $\mathrm{BE_{min}}$, and $\mathrm{BE_{max}}$ are considered. The curves are parametrized respect to the average normalized aggregate offered load (G). Analytical (solid lines) and simulation (dashed lines) results are shown.

In particular, we consider sources equipped with a finite buffer of length $L = 2$. The sources transmit, on average, 7 pck/s (corresponding to $p = 0.00225$). Therefore, the network is operating in a low load region, corresponding to $G = 0.32$.

We consider 54 combinations of the MAC parameters: m can assumes values in the set $\{1,2,\ldots,6\}$, $\mathrm{BE_{max}}$ in the set $\{4,5,6\}$, and $\mathrm{BE_{min}}$ in the set $\{4,5,6\}$. Therefore, for every value of m one can generate results for 9 combinations of $\mathrm{BE_{max}}$ and $\mathrm{BE_{min}}$, i.e., for the following possible combinations:

$$(\mathrm{BE_{max}}, \mathrm{BE_{min}}) \in \{(4,2),(4,3),(4,4),(5,2),(5,3),(5,4),(6,2),(6,3),(6,4)\}. \quad (11)$$

In Figure 7, the *PSR* is shown as a function of the combinations of the MAC parameters. The results are obtained both through simulations (pluses) and through the analytical model (solid bars). The bars with the same value of m are grouped together and characterized by the same color. Within every group of uniformly colored bars, i.e., for a fixed value of m, the values of the pairs $(\mathrm{BE_{max}}, \mathrm{BE_{min}})$ correspond to the values listed in (11). From Figure 7, one can observe that the analytical model tends to overestimate the *PSR*, but it captures the protocol dynamics, especially when the number of backoffs is large. It is interesting to observe that both simulations and analysis results show an increasing trend of the *PSR* as a

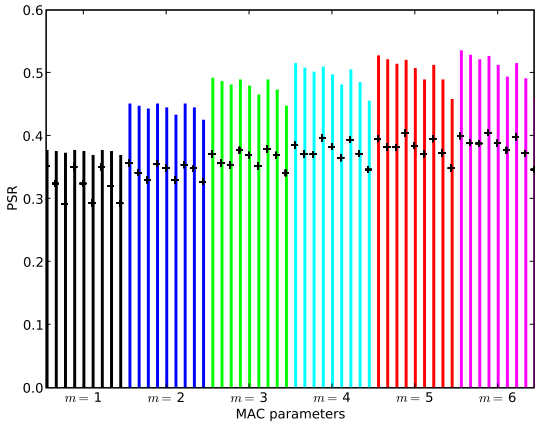

Figure 7: *PSR*, as a function of the MAC parameters m, $\mathrm{BE_{max}}$, and $\mathrm{BE_{min}}$, in a balanced CT topology. In all cases, $L = 2$, $M = 16$, and $G = 0.32$. Analytical (solid bars) and simulation (pluses) results are shown.

function of m. Therefore, one can conclude that in this low traffic load region a higher *PSR* can be achieved with a larger value of m. Moreover, according to both simulations and analytical results, the lower is the value of $\mathrm{BE_{min}}$, the higher is the *PSR*, whereas changing the $\mathrm{BE_{max}}$ does not lead to significant variations. This behavior is to be expected, since in the low traffic load region the number of retransmissions should be small.

In Figure 8, the end-to-end delay is shown, as a function of the MAC parameters, using the same previous graphical notation. In this case, the analytical model underestimates the end-to-end delay obtained by simulations. However, this discrepancy could be considered tolerable, since the model captures very well the dynamics of the CSMA/CA MAC protocol. Apart of insights obtained form this comparison, the results of Figure 8 were intuitively expected. In fact, the delay is directly proportional to the whole set of considered MAC parameters. In particular, one can observe how the dynamics of the delay are significant, both within a uniformly colored group of bars (variation of $\mathrm{BE_{max}}$ and $\mathrm{BE_{min}}$) and with respect to nodes belonging to not uniformly colored nodes (variation of m).

Finally, we arrange the presented results in "classical" delay-versus-*PSR* curves. In fact, in this way it is possible to find very quickly the best operational point. In particular, in Figure 9 and Figure 10 the end-to-end delay is shown, as a function of the PSR, on the basis of the results obtained, respectively, from the analytical model and from the simulations. Using these graphs, one can determine three "optimal" operational points according to three different perspectives, namely: (i) the operational point that guarantees the highest *PSR*; (ii) the operational point that gurantees the minimum delay; (iii) the operational point that gurantees the best tradeoff. While the first two points could be find in a deterministic way, the third one highly depends on the particular application of interest. In this paper, we select the operational point which leads to the best tradeoff between delay and *PSR*. In Figure 9, we have tagged the three optimal operational points with the corresponding MAC parameter configuration, obtained through the analytical model. In particular, it emerges that in order to get the minimum delay, it is sufficient to minimize $\mathrm{BE_{min}}$ and m (namely, $m = 1$ and $\mathrm{BE_{min}} = 2$), while the impact of $\mathrm{BE_{max}}$ is negligible. Moreover,

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

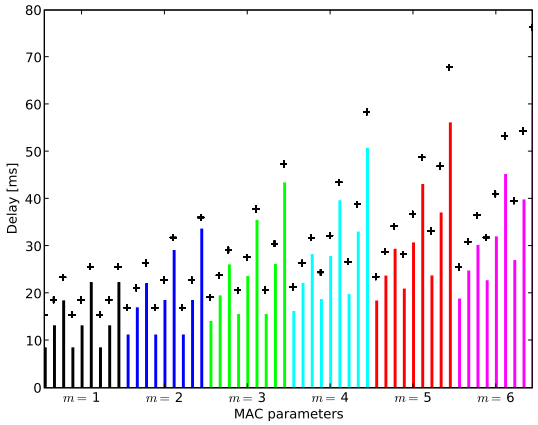

Figure 8: End-to-end delay, as a function of the MAC parameters m, BE_{max}, and BE_{min}, in a balanced CT topology. In all cases, $L = 2$, $M = 16$, and $G = 0.32$. Analytical (solid bars) and simulation (pluses) results are shown.

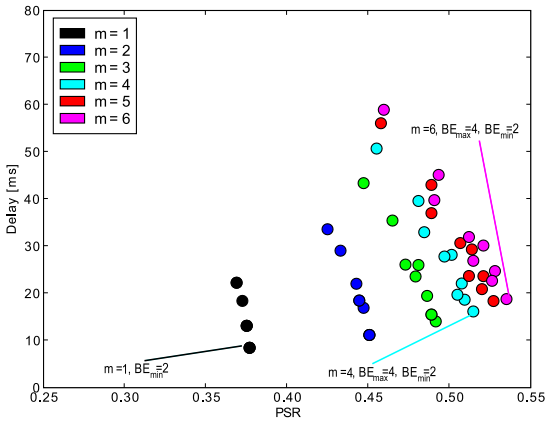

Figure 9: End-to-end delay, as function of the *PSR*, for various combinations of the MAC parameters m, BE_{max}, and BE_{min}, in a balanced CT topology. In all cases, $L = 2$, $M = 16$, and $G = 0.32$. Analytical results are considered.

one can also observe that the higher is the value of m, the higher is the achieved *PSR*. Hence, the configuration that guarantees the maximum *PSR* corresponds to $m = 6$, $BE_{min} = 2$, and $BE_{max} = 4$. Finally, from a tradeoff perspective the best configurations seem to be those characterized by a small value of BE_{min} and a high value of m. In fact, the *PSR* benefit from a large value of m is more pronounced than the penalty in terms of delay. In particular, in our opinion the configuration that offers the better tradeoff is characterized by $m = 4$, $BE_{min} = 2$, and $BE_{max} = 4$.

Finally, in order to assess the effectiveness of our analytical model, the optimal points according to the analytical model are verified through simulation results. To this regard, the three "optimal" points are tagged in Figure 10. It is immediate to see that, while the difference between the results of Figure 9 and Figure 10 is not negligible, the choices predicted the analytical model remain roughly the best also according to the simulation results. This means that our analytical model really captures the behavior of the IEEE 802.15.4 CSMA/CA MAC protocol. In particular, the configuration which minimizes the delay and the configuration with the best tradeoff are the same in both the figures. This is not true for the configuration which achieves the maximum *PSR*, but the difference with respect to the simulation-based operational point is very little.

5. CONCLUDING REMARKS

In this work, we have presented a Markov-chain based analytical framework that could be used both for analyze the performance of a certain CT IEEE 802.15.4 network and for optimize the network performance of the network itself. In particular, we have focused on a scenario in which the network topology, the nodes characteristics, and the traffic pattern, are a-priori fixed (e.g., by the application constraints). In practice, the parameters of the IEEE 802.15.4 MAC protocol, i.e., m, BE_{min}, and BE_{max}, are the only degree of freedoms. We have shown that our analytical framework is able to capture the dynamics of the network, thus allowing to tune the parameters of the MAC protocol, without resorting to lengthy and computationally expensive numerical simulations. Moreover, our framework could be potentially employed as a real-time configuration tool, for example, by the network coordinator node. As of now, the required computational time is the main obstacle to the application of our procedure in a real-time manner, but we are currently pursuing extensions of our approach with limited complexity.

6. REFERENCES

[1] IEEE 802.15.4 Std: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs). *IEEE Computer Society Press*, pages 1–679, October 2003.

[2] C. Y. Jung, H. Y. Hwang, D. K. Sung, and G. U. Hwang. Enhanced Markov chain model and throughput analysis of the slotted CSMA/CA for IEEE 802.15.4 under unsaturated traffic conditions. *IEEE Trans. Vehicular Networks*, 58(1):473–478, January 2009.

[3] V. G. Kulkarni. *Modeling and Analysis of Stochastic Systems*. Chapman and Hall, Londra, Regno Unito, 1995.

[4] M. Martalò, S. Busanelli, and G. Ferrari. Markov chain-based performance analysis of multihop IEEE 802.15.4 wireless networks. *Elsevier Performance Evaluation (PEVA)*, 2009. Special Issue on "Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks," to appear.

[5] Matlab Website. `http://www.mathworks.com`.

[6] J. Mîsic, S. Shafi, and V. B. Mîsic. Performance of a beacon enabled IEEE 802.15.4 cluster with downlink and uplink traffic. *IEEE Trans. Parallel and Distributed Syst.*, 17(4):361–376, April 2006.

[7] S. Pollin, M. Ergen, S. Ergen, B. Bougard, L. D. Perre, I. Moerman, A. Bahai, P. Varaiya, and F. Catthoor. Performance analysis of slotted carrier sense IEEE 802.15.4 medium access layer. *IEEE Trans. Wireless Commun.*, 7(9):3359–3371, September 2008.

[8] I. Ramachandran, A. K. Das, and S. Roy. Analysis of the contention access period of IEEE 802.15.4 MAC. *ACM Trans. Sensor Networks*, 3(1):29 pages, March 2007.

[9] A. Sheikh, T. Wan, and Z. Alakhdhar. A unified approach to

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623

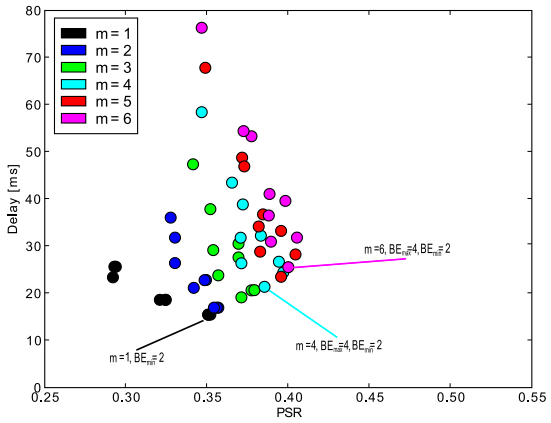

Figure 10: End-to-end delay, as function of the *PSR*, for various combinations of the MAC parameters m, $\mathrm{BE}_{\max}$, and $\mathrm{BE}_{\min}$, in a balanced CT topology. In all cases, $L = 2$, $M = 16$, and $G = 0.32$. Simulation results are considered.

analyze multiple access protocols for buffered finite users. *Elsevier Journal of Network and Computer Applications*, 27(2):49–76, January 2004.

[10] C. K. Singh, A. Kumar, and P. M. Ameer. Performance evaluation of an IEEE 802.15.4 sensor network with a star topology. *Springer Wireless Network*, 14(4):543–568, August 2007.

[11] H. Takagi. *Queueing Analysis - A Foundation of Performance Evaluation. Volume III: Discrete-Time Systems*. North-Holland, Amsterdam, Holland, 1991.

[12] H. Takagi and L. Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Trans. Commun.*, 32(3):246–257, March 1984.

[13] The Network Simulator (ns-2) Website. `http://www.isi.edu/nsnam/ns/`.

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7623
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7623