Formal Probabilistic Analysis of a WSN-based Monitoring Framework for IoT Applications

Maissa Elleuch[1,3], Osman Hasan[2], Sofiène Tahar[2], and Mohamed Abid[1]

[1] CES Laboratory, National School of Engineers of Sfax, Sfax University
Soukra Street, 3052 Sfax, Tunisia
maissa.elleuch@ceslab.org
mohamed.abid@enis.rnu.tn
[2] Dept. of Electrical & Computer Engineering, Concordia University
1455 de Maisonneuve W., Montreal, Quebec, H3G 1M8, Canada
{melleuch,o_hasan,tahar}@ece.concordia.ca
[3] Digital Research Center of Sfax
Technopark of Sfax, Tunisia

Abstract. Internet of Things (IoT) has been considered as an intuitive evolution of sensing systems using Wireless Sensor Networks (WSN). In this context, energy efficiency is considered as one of the most critical requirement. For that purpose, the randomized node scheduling approach is largely applied. The randomness feature in the node scheduling together with the unpredictable deployment make probabilistic techniques much more appropriate to evaluate the coverage properties of WSNs. Classical probabilistic analysis techniques, such as simulation and model checking, do not guarantee accurate results, and thus are not suitable for analyzing mission-critical WSN applications. Based on the most recently developed probability theory, available in the HOL theorem prover, we develop the formalizations of the key coverage performance attributes: the coverage intensity of a specific point and the expected value of the network coverage intensity. The practical interest of our higher-order-logic developments is finally illustrated through formally analyzing the asymptotic coverage behavior of an hybrid monitoring framework for environmental IoT.

Keywords: Theorem proving, Wireless sensor networks, node scheduling, performance analysis, network coverage, environmental monitoring

1 Introduction

Wireless Sensor Networks (WSN) have emerged as a key enabler technology for the development of the Internet of Things (IoT) paradigm [20,24]. Deployed over a field of interest, smart sensor nodes collaborate together without any human interaction, in order to mainly achieve a monitoring or a tracking task. Such networks are covering limitless applications [28], including home automation, external environmental monitoring and object tracking, and hence integrating WSN technologies into the IoT context [20,12].

Due to their restricted size, sensors are basically battery-powered and thus have very critical energy resources. Consider the example of a WSN deployed for forest fire detection, in which the sensor nodes are randomly distributed with a high density. The network should be able to ensure the monitoring of the whole forest area while being functional for a sufficiently long period. Since a wild fire occurs only occasionally, some sensor nodes can be intuitively deactivated to save the network energy. In this context, the k-set randomized scheduling [18] is a kind of scheduling approach, suitable for a wide range of WSN applications, which mainly consists in organizing a given set of nodes by randomly partitioning them into "k" subsets, which work alternatively.

Scheduling sensor nodes for lifetime management purposes is surely a simple and intuitive approach, however it is also crucial to not compromise on the monitoring of the area. For the same forest fire application, the deployed WSN should be also able to cover, i.e., monitor, the outbreak of fires occurring at any point of the area with a high probability. Nevertheless, the coverage performance is completely probabilistic. For instance, some fire outbreaks may not be effectively covered if no nodes are deployed around the fire because of the random node deployment, or the surrounding nodes are inactive, due to random scheduling. Missing fire intrusion, can have devastating consequences.

The performance of the randomized scheduling has been generally analyzed using paper-and-pencil based probabilistic technique [18,25]. The reliability of the obtained analytical models is consolidated through simulation using the Monte Carlo method [19]. However, both paper-and-pencil proof and simulation methods cannot be regarded as completely accurate mainly due to the error proneness of the former and the in-exhaustive nature of the later.

Formal methods overcome the drawbacks of simulation by rigorously using mathematical techniques to validate the analytical model of the given system. Recently, formal methods have gained a growing interest in the context of analyzing wireless sensor networks to analyze their functional or quantitative correctness [22,3,29], but most of the existing work is focused on the validation of their functional aspects only. Nevertheless, rigorous performance evaluation of WSNs constitutes also an extremely challenging aspect.

In this paper, we are interested in providing an accurate performance analysis of WSN randomized scheduling based on the paper-and-pencil models proposed in [18,26]. In earlier work [6,7], we have presented a formalization of the k-set randomized scheduling algorithm and its coverage properties based on a probabilistic framework developed by Hasan [13] in the HOL theorem prover. While sufficient for analyzing the coverage aspects of the original WSN models [18,26], this formalization falls short to reason about other performance aspects of the same algorithm [8], like the detection metrics. In fact, the foremost requirement for reasoning about these WSN aspects in a theorem prover is the availability of the higher-order-logic formalization of probability theory and continuous random variables. In this regard, Hurd's [16] formalization of measure and probability theories is a pioneering work. Building upon this formalization, most of the commonly-used continuous random variables [14] have been formalized using

the HOL theorem prover. However, this foundational formalization of probability theory only supports the whole universe as the probability space, which limits its scope in many aspects. In particular the inability to reason about multiple continuous random variables [14] is a major obstacle for modeling and analyzing detection and lifetime properties of WSNs [9]. More recent probability theory formalizations [21,15], however, allow the use of any arbitrary probability space that is a subset of the universe and thus are more flexible than Hurd's and Hasan's formalizations of probability theory. Particularly, Mhamdi's [21] probability theory formalization which is based on extended-real numbers (real numbers including $\pm\infty$), has been included in the HOL theorem prover and thus has been chosen for our work. Therefore, in this paper we propose to use the most recent probability theory developed by Mhamdi [21] in HOL to formally reason about the coverage properties of randomly-scheduled WSN, while emphasizing on the main lessons learned through this experience. The practical interest of the new developments is illustrated through the formal analysis of the asymptotic coverage behavior of a WSN based environmental surveillance framework.

The rest of this paper is organized as follows. We review some related work on the validation of WSNs in Section 2. In Section 3, we summarize the main requirements of this work. Section 4 provides the foundational probabilistic analysis of the coverage properties. We utilize these developments to formally verify a WSN-based monitoring framework for IoT applications in Section 5. Section 6 is devoted to discuss the main results of our work. We finally conclude the paper in Section 7.

2 Related Work

Theoretical analysis, also known as paper-and-pencil based probabilistic technique, has been widely used to validate randomized scheduling algorithms for WSN [18,25,26]. Such analysis consists in constructing a theoretical model where the required random variables are determined together with the associated performance metrics. Afterwards, a probabilistic based study is achieved. For validation purposes, simulation, using the Monte Carlo method [19], is finally done.

Traditional model checking technique [2] has been successfully used to validate various aspects in the WSN context. In [22], the formal analysis of the Optimal Geographical Density Control (OGDC) algorithm, which is a kind of randomized scheduling algorithm, is done. Several other prominent works reported on the use of model checking for the analysis of WSN protocols include [10,30]. The main strength of all these methods is their formal models and automatic verification. However, they suffer from the common model checking related problem of state space explosion [2]. Hence, the analysis of the OGDC algorithm [22] has been restricted for WSN with up to 6 nodes in a region of $15m \times 15m$. Furthermore, the work of [30] has pointed out over 1 million generated states for the analysis of a simple property. Furthermore, none of the previous works has provided reliable probabilistic modelling. For example, in [22], a random function, assumed to be 'good', has been used to model the probabilistic behavior.

To cope with these major problems, probabilistic model checking [23] has also been used for the probabilistic functional analysis of wireless systems. Probabilistic model checking allows to capture the probability modelling for both the system and the property of interest. The probabilistic model checker PRISM has been applied quite frequently for the validation of Medium Access Control (MAC) protocols for WSNs [11,29]. Nevertheless, the reasoning support for statistical quantities in most of model checkers suffers from many shortcomings. Indeed, expected performance values are usually obtained through several runs on the built model [29]. The obtained results can hardly be termed as exhaustive and thus formally verified.

On the other hand, very few works based on theorem proving exist in the open literature. The work [4] reports on the use of the PVS system to build a theorem proving based framework for WSN algorithms, with some theories expressing dynamic scenarios like nodes mobility and link quality changes [4]. While the PVS framework is supposed to be extended with some "dynamic" scenarios in [4], the randomness aspect has been characterized by a pseudo-random generator. The nodes mobility, specified by the random walk pattern, has been also specified through a recursive function.

Unlike the PVS framework which is limited by the probability support of the PVS system, the work, described in this paper, provides very accurate formalizations of the randomized scheduling algorithm based on the sound probability support of the HOL theorem prover. In addition, the presented formalizations are generic and completely valid for all the parameter values.

3 Preliminaries

3.1 Probabilistic Analysis in HOL

A probability measure P is basically a measure function on the sample space Ω and an event is a measurable set within the set F of events which are subsets of Ω. By definition, a random variable is a measurable function, satisfying the condition that the inverse image of a measurable set is also measurable [21].

Definition 1. $\vdash$ ∀X p. real_random_variable X p =
prob_space p ∧
(∀x ∈ p_space p ⇒ X x ≠ NegInf ∧ X x ≠ PosInf) ∧
X ∈ measurable (p_space p,events p) Borel.

where X designates the random variable, p is a given probability space, $NegInf$ and $PosInf$ are the higher-order-logic formalizations of negative infinity or positive infinity, and $Borel$ is the HOL definition of the Borel sigma algebra.

The probability distribution of a random variable is specified as the function that accepts a random variable X and a set s and returns the probability of the event {X ∈ s}.

Definition 2. $\vdash$ ∀X p.
distribution p X = (λs. prob p (PREIMAGE X s ∩ p_space p)).

In the discrete case, the expectation of the random variable X has been formalized in HOL as follows.

Theorem 1. $\vdash$ ∀X p. (real_random_variable X p) ∧ FINITE (IMAGE X (p_space p))
⇒ (expectation p X =
$\sum_{\text{IMAGE X (p\_space p)}}$ (λr. r × Normal (distribution p X {r}))).

where (IMAGE X (p_space p)) designates the list of values taken by the random variable X over the sample space (p_space p).

3.2 The k-set Randomized Scheduling Algorithm

During the initialization stage, the k-set randomized scheduling is run in parallel on every node as follows [18]. Each node starts by randomly picking a number, denoted by i, ranging from 0 to $(k-1)$, where k is the number of subsets or partitions. A node s_j is thus assigned to the i^{th} sub-network, designated by S_i, and will activate itself only during the scheduling round of that subset. At the end of the algorithm, k disjoint sub-networks are created. These subsets will be working independently and alternatively.Fig. 1 shows a small WSN of eight sensor nodes, which is randomly portioned into two sub-networks; S_0 and S_1. Each node randomly chooses a number 0 or 1 in order to be assigned to one of these two sub-networks. Suppose that nodes 0; 2; 5, randomly choose the number 0 and thus join the subset S_0, whereas nodes 1; 3; 4; 6; 7, select the number 1 and will be in the subset S_1. These two sub-networks will work by rounds, i.e., once the nodes 1; 3; 4; 6; 7, illustrated by the dashed circles, will be active, the remaining nodes 0; 2; 5, will be at the sleep state, and vice-versa.

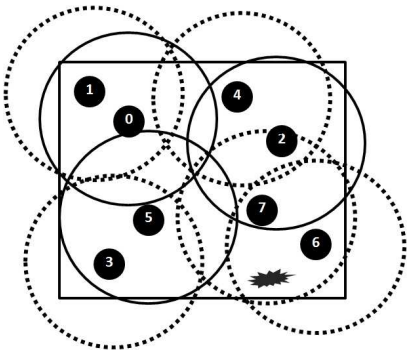

Fig. 1. The k-set randomized scheduling for $(n = 8)$ nodes and $(k = 2)$ subsets.

4 Formalization of the Network Coverage Intensity

Within a wireless sensor network, a given point is said to be covered, if any occurring event at this point, is detected by at least one active node with a given probability. According to [18], the coverage intensity of a specific point;

C_p, inside the monitored area is defined as the average time during which the point is covered in a whole scheduling cycle of length $k \times T$. A given point is covered if the current active subset contains at least one node, i.e., is not empty.

Let X be the random variable describing the total number of non-empty subsets, the coverage intensity of a given point in the monitored area, C_p, as originally specified in [18], is

$$C_p = \frac{E[X] \times T}{k \times T}\,. \tag{1}$$

where $E[X]$ denotes the expectation of X, which is described as:

$$X = \sum_{j=0}^{k-1} X_j\,. \tag{2}$$

where X_j is the Bernoulli random variable whose value is 1 in case of non-empty subset. A non-empty sub-network is described by a Bernoulli random variable with the complement probability of $\left(1 - \frac{1}{k}\right)^c$ [6], where c is the number of covering sensors for a given point.

Definition 3. $\vdash \forall$`X p k c.`
`sbst_non_empty_rv X p k c = bernoulli_distr_rv X p` $\left(1 - \left(1 - \frac{1}{\&k}\right)^{c}\right)$.

In higher-order logic, we model the coverage behavior of a specific point (Equation (1)) by the following predicate `cvrge_intsty_pt`.

Definition 4. $\vdash \forall$`p X k s c. cvrge_intsty_pt p X k s c =`
`expectation p (λx. SIGMA (λi. (X i) x) s) / (&k).`

where X: a random variable that returns an extended real number, p: the probability space, k: the number of sub-networks, s: the summation set whose cardinality is k, and c: the number of covering sensors for a given point. The operator `&` allows the conversion of the natural number m into its extended number counterpart.

The following mathematical expression for the coverage intensity of a point has been formally verified in Theorem 2.

Theorem 2. $\vdash \forall$`X p k s c. (prob_space p) ∧ (FINITE s) ∧ (1 < k)`
`∧ (CARD s = k) ∧ (∀i. i ∈ s ⇒ sbst_non_empty_rv (X i) p k c)`
`⇒ (cvrge_intsty_pt p X k s c = Normal` $\left(1 - \left(1 - \frac{1}{k}\right)^{c}\right)$`).`

- The assumption `(∀i. i ∈ s ⇒ sbst_non_empty_rv (X i) p k c)` indicates that every element of the set s is a random variable `sbst_non_empty_rv` (Definition 3).
- The HOL function `Normal` is used to convert a real value to its corresponding value in an extended real.

The proof of the above theorem is mainly based on lemmas about the linearity of the expectation property, which in turn required some reasoning on the integrability of some functions as well as operations from the Lebesgue theory. For most of these lemmas, it was a prerequisite to verify the measurability of the used events, along with some analysis on extended reals.

The whole network can be now statistically described by a single performance metric; C_n, which is the average or the expectation value of the coverage intensity over all points of the monitored area.

$$C_n = E[C_p]\,. \tag{3}$$

According to the expression of C_p, shown in Theorem 2, we can write

$$C_n = E[1 - \left(1 - \frac{1}{k}\right)^c]\,. \tag{4}$$

Based on the above equation, we notice how the value of C_n depends mainly on c which is the number of nodes covering a given point of the field. Intuitively, we can assimilate the fact of covering a point or not to a Bernoulli trial with the probability $q = \frac{r}{a}$ [18]. Considering the variable c among the n nodes of the network, it becomes a Binomial random variable (C) with the probability given in Equation (5). Thereby, the network coverage intensity C_n, shown in Equation (4), is not a simple expectation, but rather an expectation of a function of a random variable.

$$Pr(C = j) = C_n^j \times \left(\frac{r}{a}\right)^j \times \left(1 - \left(\frac{r}{a}\right)\right)^{n-j}. \tag{5}$$

where C_n^j is the binomial coefficient, r is the size of the sensing area of each sensor, a is the size of the monitored area, and $\left(\frac{r}{a}\right)$ is the probability that each sensor covers a given point. The Binomial random variable with n trials and success probability $q = \left(\frac{r}{a}\right)$ has been formalized in HOL as follows.

Definition 5. ⊢ ∀X p q n. binomial_distr_rv X p q n =
(real_random_variable X p) ∧
(IMAGE X (p_space p) = IMAGE (λx.&x) (count (SUC n))) ∧
(∀m. &m ∈ (IMAGE X (p_space p)) ⇒
(distribution p X {&m} = &(binomial n m) × q^m × $(1-q)^{(n-m)}$).

where X is a real random variable on the probability space p, and `IMAGE (λx.&x) (count (SUC n))` gives the support of the Binomial. The function `binomial`, used in the above definition, is the higher-order-logic formalization of the binomial coefficient for reals.

The coverage intensity of the whole WSN with n nodes has been formally specified by the function `cvrge_intsty_network`, shown in Definition 6. The latter takes as parameters: X: a random variable that returns an extended real number, p: the probability space, s: the summation set used in Definition 4, k: the number of sub-networks, C: the random variable describing the number of

covering nodes, n: the total number of nodes, and q: the probability that each sensor covers a given point.

Definition 6. $\vdash \forall$`X p k s C n q.`
`cvrge_intsty_network p X k s C n q =`
`expectation p (`λ`x. cvrge_intsty_pt p X k s (num (C x))).`

where the function `expectation` designates the higher-order-logic formalization of the expectation of a random variable that returns an extended real, and the values `(num(C x))`, in the above definition, are the output values of the random variable `C`. The function `num`, used here, converts an extended real; `(&m)`, to its corresponding natural value `m`, using the real function `floor`.

Based on the higher-order-logic formalizations developed so far, we have been able to formally verify the final network coverage intensity as in Theorem 3.

Theorem 3. $\vdash \forall$`p X k s C n q. (prob_space p)` $\wedge$ `(0 < q < 1)` $\wedge$
`(events p = POW (p_space p))` $\wedge$ `(1` $\leq$ `n)` $\wedge$ `(1 < k)` $\wedge$ `FINITE s` $\wedge$
`(CARD s = k)` $\wedge$ `(sn_covers_p C p q n)` $\wedge$
`(expectation p C` $\neq$ `PosInf)` $\wedge$ `(expectation p C` $\neq$ `NegInf)` $\wedge$
`(`$\forall$`i x. (i` $\in$ `s)` $\wedge$ `(x` $\in$ `p_space p)` $\Rightarrow$
`sbst_non_empty_rv (X i) p k (num(C x)))`
$\Rightarrow$ `(cvrge_intsty_network p X k s C n q = Normal` $(1-(1-\frac{\mathtt{q}}{(\&\mathtt{k})})^{\mathtt{n}}))$`.`

- The assumption `(events p = POW (p_space p))` describes the set of events to be the power set of the sample space Ω.
- The assumptions `(1` $\leq$ `n)` ensures that the WSN include at least one node, while `(0 < q < 1)` checks that the probability q lies in [0..1].
- `sn_covers_p` is the Binomial random variable (Definition 5) with a finite expectation, i.e., `(expectation p C` $\neq$ `PosInf)` $\wedge$ `(expectation p C` $\neq$ `NegInf)`. The variables `(PosInf)` and `(NegInf)` are the higher-order-logic formalizations of positive infinity and negative infinity, respectively.
- The function `(sbst_non_empty_rv (X i) p k (num(C x)))` is the function specified in Definition 3.

The proof of Theorem 3 is primarily based on Theorem 4 which verifies the expectation of a function of a random variable. Additionally, the current proof also required the application of the linearity of the expectation property. Finally, a considerable amount of real analysis associated to the Binomial theorem for reals, and to the summation function has been needed.

Theorem 4. $\vdash \forall$`C p q n k.`
`(prob_space p)` $\wedge$ `(1 < k)` $\wedge$ `(0 < q < 1)` $\wedge$
`(events p = POW (p_space p))` $\wedge$ `(1` $\leq$ `n)` $\wedge$ `(sn_covers_p C p q n)`
$\Rightarrow$ `(expectation p (`λ`x. f_fct (num (C x)) k) = Normal` $(1-\frac{\mathtt{q}}{(\&\mathtt{k})})^{\mathtt{n}})$`.`

where the function `f_fct` is defined as follows

$$\mathtt{f\_fct\ x\ k} = \mathtt{Normal}\left(1-\frac{1}{\mathtt{k}}\right)^{\mathtt{x}}. \tag{6}$$

The proof of Theorem 4 has been possible using intermediate results on the injectivity of some functions, as well as, some properties related to the random variables functions. A lot of reasoning associated with the use of extended real and the floor function, has also been required.

In this section, we presented our new higher-order-logic formalizations of the k-set randomized scheduling for wireless sensor networks, using the recently developed probability theory available in the HOL theorem prover [21]. These formalizations have been then utilized to formally reason about the coverage performance properties. The corresponding HOL code of the current formalizations is available at [5]. Due to fundamental differences in the foundations of the two probability theories in [13] and [21], the current resulting formalizations is completely different from the previous one [6]. Indeed, the new probability theory allows to cater for arbitrary probability spaces and is thus more generic and complete compared to the previous formalization in which the probability space has to be the universe of a set. Moreover, the specification of the randomized algorithm has been found to be much more intuitive with [21]. Unlike the work in [6], the developed proofs required much less reasoning about sets and lists producing thus less lengthy proofs. However, these proofs have been more laboured involving usually results from the three HOL theories: Lebesgue, measure and extended reals. A deep learning of all theoretical foundations of [21] was thus required to successfully achieve the target formalizations in the HOL theorem prover. In the next section, we will illustrate how the developed generic theorems extremely facilitate the formal analysis of real-world WSN applications.

5 Application: Formal Analysis of a WSN-based monitoring framework for IoT Applications

Numerous frameworks for environmental monitoring based on WSN have been hence proposed in the literature [1,27]. These systems can be seamlessly integrated to build an extended IoT framework for low-cost, persistent and efficient services [17,12]. Due to the new constraints of the IoT environment, deployed WSN should have a smart behavior regarding the power availability while performing a good coverage of any intrusion. The randomized node scheduling has been proposed for use to save energy in the context of an heterogeneous surveillance framework for environmental monitoring [27]. Such framework considers collaboration between sensor nodes, mobile robots and RFID tags, to ensure efficient surveillance. Using specific sensors designed for IoT [17], this framework can realize a whole IoT structure.

In this section, we focus on formally analyzing the coverage performances of the hybrid surveillance framework proposed in [27] adopted for IoT applications. The nodes can hence have any sensing area r, and are deployed into a circular region of a radius R with a total size of a, whereas the success probability q of a sensor covering a point is $q = \frac{r}{a}$. Such framework has been primarily analyzed using a paper-and-pencil model, which has been then validated through some simulation scenarios evaluating the expected coverage and the maximum number

of subsets [27]. It would be interesting to provide a more rigorous technique to validate the proposed paper-and-pencil model. Based on the formal development achieved so far, we show in this section how we are able to carry out an accurate asymptotic analysis of the probabilistic coverage according to the key design parameters: n; the total number of sensor and k; the number of subsets.

We designate the generic network coverage intensity `(cvrge_intsty_network p X s k C n q)`, shown in Definition 6, by `(Cn_wsn p X s k C n q)`, that has been checked in HOL as

$$\texttt{Normal}\left(1-\left(1-\frac{\texttt{q}}{\texttt{k}}\right)^{\texttt{n}}\right). \tag{7}$$

5.1 Formal Analysis based on the Number of Nodes

Setting the number of subsets to k and targeting a network coverage intensity Cn_wsn of at least t, we verify, in Lemma 1, the minimum number of sensors; n_{min}, that are necessary to deploy in the context of our monitoring framework.

Lemma 1. $\vdash \forall$`p X s k C n q t. (1 ≤ n) ∧ (1 < k) ∧ (0 < q < 1) ∧ (0 < t < 1) ∧ (Normal t ≤ Cn_wsn p X s k C n q)`
$\Rightarrow \left\lceil \frac{\ln(1-t)}{\ln\left(1-\frac{q}{k}\right)} \right\rceil \leq$ `&n.`

The higher-order-logic proof of the above lemma is based on some properties of transcendental functions and arithmetic reasoning.

We have been able to formally verify, in Lemma 2, that the network coverage intensity Cn_wsn is a growing function of n, i.e., a larger node number n is responding to a better coverage. For the monitoring framework, much more points of the area are expected to be covered, since it is likely that many more covering nodes are deployed in its surrounding area.

Lemma 2.

```
⊢ ∀p X s k C q. (1 < k) ∧ (0 < q < 1)
  ⇒ (mono_incr (λn. real(Cn_wsn p X k s C n q))).
```

where the function `real` is used to convert the network coverage intensity of type extended real to its corresponding real value, and `mono_incr` is the HOL definition of an increasing sequence.

While Cn_wsn increases with the increase of the number of nodes n, as verified in Lemma 2, the next lemma shows how the network coverage intensity Cn_wsn approaches 100% when n becomes infinite, independently of of the monitoring application.

Lemma 3. $\vdash \forall$`p X s k C q. (1 < k) ∧ (0 < q < 1)`
$\Rightarrow$ `(` $\lim_{\texttt{n}\to+\infty}$ `(λn. real(Cn_wsn p X s k C n q)) = 1).`

5.2 Formal Analysis based on the Number of Subsets

Targeting a network coverage intensity of at least t, we successfully verify, in Lemma 4, the upper bound on the number of disjoint subsets k for a given n.

Lemma 4. $\vdash \forall$p X s k C n q. (1 $\leq$ n) $\wedge$ (0 < t < 1) $\wedge$
(0 < q < 1) $\wedge$ (1 < k) $\wedge$ (Normal t $\leq$ (Cn_wsn p X s k C n q))
$\Rightarrow$ k $\leq \frac{\mathtt{q}}{1-\mathtt{e}^{\frac{\ln(1-t)}{(\&n)}}}$.

The above result is interesting for practical WSN applications which necessitate adjustable performance measurement quality for energy preserving purposes.

We have been able to formally check, in Lemma 5, that the network coverage intensity Cn_wsn definitely decreases when the WSN is partitioned into a quite large number of sub-networks k.

Lemma 5. $\vdash \forall$p X s C n q. (1 $\leq$ n) $\wedge$ (0 < q < 1)
$\Rightarrow$ (mono_decr (λk. real (Cn_wsn p X s k C n q))).

where the HOL function `mono_decr` defines a decreasing sequence.

We also formally confirm, in Lemma 6, that increasing the number of deployed nodes n gives smaller network coverage and hence a poor performance of the deployed application.

Lemma 6. $\vdash \forall$p X s C n q. (1 $\leq$ n) $\wedge$ (0 < q < 1)
$\Rightarrow$ ($\lim_{\mathtt{k}\to+\infty}$ (λk. real (Cn_wsn p X p s k C n q)) = 0).

The above lemma has been successfully verified in HOL using intermediate results associated to real and sequential limits.

5.3 Formal Analysis based on Uniform Partitions

We closely investigate the asymptotic coverage behavior of our monitoring framework in the case of a *uniform* split of the nodes. Here, n can be written as $k \times m$, where m is the number of nodes per subset.

In particular, as the number of sub-networks k goes infinite, the upper limit of the network coverage Cn_wsn has been formally verified in Lemma 7.

Lemma 7. $\vdash \forall$p X s C m q. (0 < q < 1)
$\Rightarrow \lim_{\mathtt{k}\to+\infty}$ (λk. real(Cn_wsn p X s k C (m $\times$ k) q)) = 1 - $\mathtt{e}^{-\mathtt{q}\times(\&\mathtt{m})}$.

The proof of the above lemma has been quite tricky requiring the important result $\lim_{\mathtt{k}\to+\infty}(1+\frac{\mathtt{x}}{\mathtt{k}})^{\mathtt{k}} = \mathtt{e}^{\mathtt{x}}$, which had to be proved in HOL beforehand.

Based on Lemma 7, we can hence verify that when m becomes very large, the uniform network coverage will surely approach 100%. Such result is considered as a second verification of Lemma 3 in the case where n and k are proportional.

Lemma 8. $\vdash$ `∀X p s C q. (0 < q < 1)`
$\Rightarrow \lim_{\mathtt{m}\to+\infty}$ `(λm.` $\lim_{\mathtt{k}\to+\infty}$ `(λk. real(Cn_wsn p X s k C (m × k) q)) = 1.`

The current analysis, presented in this section, distinctly shows how our theoretical developments, described in Section 4, match pretty well the original paper-and-pencil models of the randomized scheduling, available in the open literature [18,26].

6 Discussion

The main motivation of the current work is to provide a rigorous approach for the probabilistic performance evaluation of the k-set randomized scheduling algorithm for wireless sensor networks. The randomness in the scheduling approach and the node deployment makes the accuracy of the performance evaluation of such algorithm very critical, especially given the major limitations of classical techniques and the safety-critical of most WSN applications. In this regard, this paper describes the main formalizations of the k-set randomized scheduling and its coverage properties using the new probability theory available within the HOL4 theorem prover [21]. These higher-order-logic formalizations resulted from the porting process of our previous formalizations [6,7], developed within a precedent probabilistic framework of the HOL theorem prover [13]. The practical usefulness of our approach is shown in Section 5, where we formally analyzed the coverage performance of a general purpose surveillance framework based on WSN for IoT applications.

The higher-order-logic formalizations, presented in this paper, consumed approximatively 730 lines of code in the HOL4 theorem prover. On the other hand, the formal analysis of our application took only 200 lines of HOL code for the verification of most of the lemmas. Nevertheless, the proofs of Lemmas 7 and 8 have been quite tedious consuming in total 500 lines of HOL code, since the mathematical theorem $\lim_{\mathtt{k}\to+\infty}(1+\frac{\mathtt{x}}{\mathtt{k}})^{\mathtt{k}} = \mathtt{e}^{\mathtt{x}}$, was missing in HOL. The latter result required a lot of real analysis related to the exponential function as a power series and many other properties for the sequence convergence.

The generic nature of the theorem proving technique and the high expressibility of higher-order logic allows us a considerable amount of flexibility in several aspects. Indeed, the formalizations, presented in this paper, primarily constitutes a successful automation of the paper-and-pencil models [18,26] of the k-set randomized scheduling and its coverage performance within a higher-order-logic proof assistant. Through this work, we therefore clearly assert the complete accordance of the resulting formal developments with the mathematical models, increasing thus the confidence on the developed theory. Given the discussion, presented in Section 2, it is certain that other analysis techniques can never have this efficiency. Actually, the existing probabilistic models of the randomized scheduling are not so reliable either regarding the complete set of assumptions or the correctness of the manual mathematical analysis, which may include human

errors. In addition, while previous simulation methods usually rely on pseudo-random modelling, we have been able to provide an appropriate modelling of the inherent randomness of the algorithm of interest. Besides, unlike probabilistic model checkers where statistical properties are not so accurately specified, we have been able to achieve formal and precise analysis of the network coverage as a statistical measure of the coverage intensity for a specific point. On the other hand, the formal performance analysis of the coverage behavior of the environmental framework clearly shows the usefulness of our theoretical developments. Such verification enables reliable asymptotic reasoning of the deployed WSN. Compared to the asymptotic analysis already done in [7], we have been able to enrich our analysis with new valuable results. At the end, it is important to note that the presented application is a simple case study illustrating the practical interest of our work, but the claimed generic results can be obviously applied to any other WSN application as well.

To successfully achieve the current work, we have experienced many difficulties. Firstly, although the initial paper-and-pencil models [18,26] are depending on simple discrete random variables, the major challenge was to correctly translate these models of a real WSN algorithm into higher-order logic. These analytical modelling of real-world systems is effectively very intuitive, and the original mathematical models [18,26] are usually missing detailed explanations either when describing the probabilistic analysis or when applying the probability rules. In addition, the assumptions of the original model are never presented exhaustively. A deep investigation step was thus required in order to correctly understand all missing steps and achieve then efficiently the target higher-order-logic formalizations. For that purposes, a good background on probability coupled with a sound knowledge of the WSN context, are usually required for an effective understanding of the probabilistic reasoning.

Secondly, the choice of porting our previous higher-order-logic formalizations [6,7] into a new probability theory [21], was, at once, tough and time consuming. As previously mentioned, such choice has been primarily motivated by the fact that we were targeting more evolutive probabilistic analysis of the k-set randomized scheduling with the formalization of further performance aspects in the near future [8]. These aspects should require some probabilistic features which are not available in [13]. Moreover, while the new HOL specification seems to be more straightforward in the new probability theory, we had to get extensive understanding of all the corresponding mathematical foundations including extended reals, measure and Lebesgue theories in order to correctly conduct the probabilistic analysis. Nevertheless, the existing results from the formalized probability theory helped us to keep the amount of proof efforts reasonable.

7 Conclusions

In this paper, we presented a reliable approach for the formal analysis of the coverage performances of wireless sensor networks using the k-set randomized scheduling to save energy. This formalization enables us to formally verify the

coverage related characteristics of most WSNs using the k-set randomized scheduling. To show the practical interest of our foundational results, we apply them to perform the formal probabilistic analysis of an hybrid monitoring framework for environmental Internet of Things (IoT) applications. Such framework can be adapted for any kind of monitoring application using WSN as well.

On the other hand, the produced results are thoroughly generic, i.e., valid for all parameter values. It is clear that such results cannot be achieved in simulation or probabilistic model checking based approach. Moreover, it has been possible to provide precise formal reasoning on the statistical coverage using expectation. Finally, unlike most of the existing work that focuses on the validation of the functional aspects of WSN algorithms, our work is distinguishable by addressing the performance aspects. Finally, the proposed solution allowed us to build upon our coverage formalizations to develop our whole methodology [8] in a single coherent formalism. In particular, the current results have been very helpful for our work on the higher-order-logic formalizations of the detection properties of WSNs [9], based on the paper-and-pencil analysis of [26]. It has been useful to formally check the relationship between coverage and detection showing that coverage reflects detection [18].

References

1. Aslan, Y., Korpeoglu, I., Ulusoy, O.: A Framework for use of Wireless Sensor Networks in Forest Fire Detection and Monitoring. Computers, Environment and Urban Systems 36(6), 614–625 (2012)
2. Baier, C., Katoen, J.P.: Principles of Model Checking. The MIT Press (2008)
3. Ballarini, P., Miller, A.: Model Checking Medium Access Control for Sensor Networks. In: Proceedings of the 2nd symposium on Leveraging Applications of Formal Methods, Verification and Validation. pp. 255–262. IEEE Computer Society (2006)
4. Bernardeschi, C., Masci, P., Pfeifer, H.: Analysis of Wireless Sensor Network Protocols in Dynamic Scenarios. In: Stabilization, Safety, and Security of Distributed Systems, Lecture Notes in Computer Science, vol. 5873, pp. 105–119. Springer (2009)
5. Elleuch, M.: Formalization of the Coverage Properties of WSNs in HOL (2015), `http://hvg.ece.concordia.ca/projects/prob-it/wsn.php`
6. Elleuch, M., Hasan, O., Tahar, S., Abid, M.: Formal Analysis of a Scheduling Algorithm for Wireless Sensor Networks. In: Formal Methods and Software Engineering, Lecture Notes in Computer Science, vol. 6991, pp. 388–403. Springer (2011)
7. Elleuch, M., Hasan, O., Tahar, S., Abid, M.: Formal Probabilistic Analysis of a Wireless Sensor Network for Forest Fire Detection. In: Symbolic Computation in Software Science, Electronic Proceedings in Theoretical Computer Science, vol. 122, pp. 1–9. Open Publishing Association (2013)
8. Elleuch, M., Hasan, O., Tahar, S., Abid, M.: Towards the Formal Performance Analysis of Wireless Sensor Networks. In: Proceedings of the 22th workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises. IEEE Computer Society (2013)
9. Elleuch, M., Hasan, O., Tahar, S., Abid, M.: Formal Probabilistic Analysis of Detection Properties in Wireless Sensor Networks. Formal Aspects of Computing 27(1), 79–102 (2015)

10. Fehnker, A., Fruth, M., McIver, A.: Graphical Modelling for Simulation and Formal Analysis of Wireless Network Protocols. In: Methods, Models and Tools for Fault Tolerance, Lecture Notes in Computer Science, vol. 5454, pp. 1–24. Springer (2009)
11. Fruth, M.: Probabilistic Model Checking of Contention Resolution in the IEEE 802.15.4 Low-rate Wireless Personal Area Network Protocol. In: Proceedings of the 2nd symposium on Leveraging Applications of Formal Methods, Verification and Validation. pp. 290–297. IEEE Computer Society (2006)
12. Hart, J., Martinez, K.: Towards an Environmental Internet of Things [IoT]. Earth and Space Science 2, 1–7 (2015)
13. Hasan, O.: Formal Probabilistic Analysis using Theorem Proving. Ph.D. thesis, Concordia Univ., Montreal, QC, Canada (2008)
14. Hasan, O., Tahar, S.: Formalization of Continuous Probability Distributions. In: Automated Deduction, Lecture Notes in Computer Science, vol. 4603, pp. 3–18. Springer (2007)
15. Hölzl, J., Heller, A.: Three Chapters of Measure Theory in Isabelle/HOL. In: Interactive Theorem Proving, Lecture Notes in Computer Science, vol. 6898, pp. 135–151. Springer (2011)
16. Hurd, J.: Formal Verification of Probabilistic Algorithms. Ph.D. thesis, Univ. of Cambridge, Cambridge, UK (2002)
17. Lazarescu, M.: Design of a WSN Platform for Long-Term Environmental Monitoring for IoT Applications. IEEE Journal on Emerging and Selected Topics in Circuits and Systems 3(1), 1–6 (2013)
18. Liu, C., Wu, K., Xiao, Y., Sun, B.: Random Coverage with Guaranteed Connectivity: Joint Scheduling for Wireless Sensor Networks. IEEE Transactions on Parallel and Distributed Systems 17(6), 562–575 (2006)
19. MacKay, D.: Introduction to Monte Carlo Methods. In: Proceedings of NATO Advanced Study Institute on Learning in Graphical Models. pp. 175–204. Kluwer Academic Publishers (1998)
20. Mainetti, L., L. Patrono, L., Vilei, A.: Evolution of Wireless Sensor Networks Towards the Internet of Things: A Survey. In: Proceedings of the 19th International Conference on Software, Telecommunications and Computer Networks. pp. 1–6. IEEE (2011)
21. Mhamdi, T.: Information-Theoretic Analysis using Theorem Proving. Ph.D. thesis, Concordia Univ., Montreal, QC, Canada (December 2012)
22. Ölveczky, P., Thorvaldsen, S.: Formal Modeling and Analysis of the OGDC Wireless Sensor Network Algorithm in Real-time Maude. In: Formal Methods for Open Object-based Distributed Systems, Lecture Notes in Computer Science, vol. 4468, pp. 122–140. Springer (2007)
23. Rutten, J., Kwaiatkowska, M., Normal, G., Parker, D.: Mathematical Techniques for Analyzing Concurrent and Probabilisitc Systems. CRM Monograph Series, American Mathematical Society (2004)
24. Whitmore, A., Agarwal, A., Xu, L.: The Internet of Things–A Survey of Topics and Trends. Information Systems Frontiers 17(2), 261–274 (2015)
25. Wu, K., Gao, Y., Li, F., Xiao, Y.: Lightweight Deployment-Aware Scheduling for Wireless Sensor Networks. Mobile Networks and Applications 10(6), 837–852 (2005)
26. Xiao, Y., Chen, H., Wu, K., Sun, B., Zhang, Y., Sun, X., Liu, C.: Coverage and Detection of a Randomized Scheduling Algorithm in Wireless Sensor Networks. IEEE Transactions on Computers 59(4), 507–521 (2010)

27. Xiao, Y., Zhang, Y.: Divide-and Conquer-based Surveillance Framework using Robots, Sensor Nodes, and RFID tags. Wireless Communications and Mobile Computing 11(7) (2011)
28. Yick, J., Mukherjee, B., Ghosal, D.: Wireless Sensor Network Survey. Computer Networks 52(12), 2292–2330 (2008)
29. Zayani, H., Barkaoui, K., Ayed, R.B.: Probabilistic Verification and Evaluation of Backoff Procedure of the WSN ECo-MAC Protocol. International Journal of Wireless & Mobile Networks 12(1), 156–170 (2010)
30. Zheng, M., Sun, J., Liu, Y., Dong, J., Gu, Y.: Towards a Model Checker for NesC and Wireless Sensor Networks. In: Formal Methods and Software Engineering, Lecture Notes in Computer Science, vol. 6991, pp. 372–387. Springer (2011)