Probabilistic Analysis of Wireless Systems using Theorem Proving

Osman Hasan[1]

ECE Department
Concordia University
Montreal, Canada

Sofiène Tahar[2]

ECE Department
Concordia University
Montreal, Canada

Abstract

Probabilistic techniques play a major role in the design and analysis of wireless systems as they contain a significant amount of random or unpredictable components. Traditionally, computer simulation techniques are used to perform probabilistic analysis of wireless systems but they provide inaccurate results and usually require enormous amount of CPU time in order to attain reasonable estimates. To overcome these limitations, we propose to use a higher-order-logic theorem prover (HOL) for the analysis of wireless systems. The paper presents a concise description of the formal foundations required to conduct the analysis of a wireless system in a theorem prover, such as, the higher-order-logic modeling of random variables and the verification of their corresponding probabilistic and statistical properties in a theorem prover. In order to illustrate the utilization and effectiveness of the proposed idea for handling real-world wireless system analysis problems, we present an analysis of the automated repeat request (ARQ) mechanism at the logic link control (LLC) layer of the General Packet Radio Service (GPRS), which is a packet oriented mobile data service available to the users of Global System for Mobile Communications (GSM).

Keywords: Formal Methods, GPRS, Higher-Order-Logic, Mechanization of Proofs, Probabilistic Analysis, Theorem Proving, Wireless Networks.

1 Introduction

Wireless communication systems are increasingly being used these days in applications ranging from ubiquitous consumer electronic devices, such as cell phones and computers, to not so commonly used but safety critical domains, such as automated highways and factories, remote tele-medicine and wireless sensor networks. The correctness of operation for these wireless systems is very important due to financial or safety critical nature of their applications. Therefore, quite a significant portion

[1] Email: o_hasan@ece.concordia.ca
[2] Email: tahar@ece.concordia.ca

of the design time of a wireless system is spent on analyzing the designs so that functionality errors can be caught and reliability and performance metrics can be evaluated prior to production. Probabilistic considerations play a significant role in such analysis since wireless systems usually exhibit some random or unpredictable elements. For example, wireless channel parameters are often described in terms of their Probability Mass Functions (PMF) instead of the actual mathematical models for all reflection, diffraction and scattering processes that determine the different multi-path components of a wireless channel. Similarly, probabilistic models are used to describe the mobility of communicating stations. Randomized algorithms and probabilistic analysis are also extensively used in the area of wireless networks. A comprehensive survey in this regard is presented in [41].

Today, simulation is the most commonly used computer based probabilistic analysis technique for wireless systems, e.g., see [39,4,15,25]. Most simulation based wireless system analysis softwares provide a programming environment for defining functions that approximate random variables for probability distributions. The random elements in a given wireless system are modeled by these functions and the system is analyzed using computer simulation techniques [11], such as the Monte Carlo Method [31], where the main idea is to approximately answer a query on a probability distribution by analyzing a large number of samples. Statistical quantities, such as expectation and variance, may then be calculated, based on the data collected during the sampling process, using their mathematical relations in a computer. Due to the inherent nature of simulation coupled with the usage of computer arithmetic, the probabilistic analysis results attained by the simulation approach can never be termed as 100% accurate. Thus, simulation should not be relied upon for the analysis of wireless systems, especially when they are used in safety critical areas, such as, medicine, transportation and military, where inaccuracies in the analysis may even result in the loss of human lives.

In the past couple of decades, formal methods [16] have been successfully used for the precise analysis of a verity of hardware and software systems. The rigorous exercise of developing a mathematical model for the given system and analyzing this model using mathematical reasoning usually increases the chances for catching subtle but critical design errors that are often ignored by traditional techniques like simulation. Given the sophistication of the present age wireless systems and their extensive usage in safety critical applications there is a dire need of using formal methods in this domain. However, due to the random and unpredictable nature of wireless systems, the usage of formal methods has been quite restricted so far. Some major reasons for this include the restriction to handle random behaviors that can be modeled as a Markov chain only and the inability to precisely reason about statistical properties, such as expectation and variance, in the case of state-based approaches and the fear of huge proof efforts involved in reasoning about random components of a wireless system in the case of theorem proving.

We believe that due to the recent developments in the formalization of probability theory concepts in higher-order-logic [24,19,21,18], we are now at the stage where we can handle the analysis of a variety of wireless systems with random components in a higher-order-logic theorem prover [13] with reasonable amount of modeling and verification efforts. The main motivation of using a higher-order-logic

theorem prover for this purpose is the ability to formally analyze a broader range of wireless systems by leveraging upon the high expressiveness of the underlying logic.

The foremost requirement for conducting the probabilistic analysis of wireless systems in a higher-order-logic theorem prover is the ability to formalize commonly used random variables in higher-order logic and reason about their corresponding probabilistic and statistical properties in a theorem prover. In this paper, we present a framework illustrating the use of the existing probability theory related higher-order-logic formalizations for fulfilling this requirement and thus in turn analyzing wireless systems. The fact that we are building upon existing formalization tends to minimize the modeling and verification efforts associated with the higher-order-logic theorem proving approach.

In order to illustrate the utilization and effectiveness of theorem proving for handling real-world wireless system analysis problems, we present an analysis of the automated repeat request (ARQ) mechanism at the logic link control (LLC) layer of the General Packet Radio Service (GPRS) standard for Global System for Mobile Communications (GSM) [10]. This analysis is a good representation of a typical wireless system analysis problem that cannot be modeled as a Markov chain and thus cannot be handled by the state based formal analysis approaches. Therefore, the successful handling of this analysis problem clearly indicates the usefulness of the proposed idea. The paper provides a formalization of the ARQ mechanism at the LLC layer of the GPRS standard in higher-order-logic and the formal verification of a couple of probabilistic properties related to the number of LLC frame retransmissions required to successfully transmit a single LLC frame.

The work described in this paper is done using the HOL theorem prover [14], which is based on higher-order logic. The main motivation behind this choice is the fact that most of the work that we build upon is developed in HOL. It is important to note here that the ideas presented in this paper are not specific to the HOL theorem prover and can be adapted to any other higher-order-logic theorem prover as well, such as Isabelle [35], Coq [8] or PVS [36].

The rest of the paper is organized as follows: Section 2 provides a review of the related work. In Section 3, we present a methodology based on existing HOL formalizations of probability theory to the analysis of wireless systems. The ARQ analysis of the LLC layer of the GPRS standard is given in Section 4. Finally, Section 5 concludes the paper.

2 Related Work

Probabilistic model checking [2,38] is the most commonly used formal method in the area of probabilistic analysis of wireless systems. For example, the PRISM model checker [26] has been used to analyze a sub protocol of the IEEE 802.11 standard for wireless local area networks (WLANs) in [27], the IEEE 802.15.4 networking standard in [12] and the Medium Access Control (MAC) protocol SMAC in [3]. Similarly, the ETMCC model checker [23] has been used for the dependability analysis of a variant of the central access protocol of the IEEE 802.11 standard [32]. Just like the traditional model checking, probabilistic model checking involves the construction of a precise state-based mathematical model of the given probabilistic

system, which is then subjected to exhaustive analysis to verify if it satisfies a set of formally represented probabilistic properties. Besides the accuracy of the results, the most promising feature of probabilistic model checking is the ability to perform the analysis automatically. On the other hand, it is limited to systems that can only be expressed as probabilistic finite state machines or Markov chains. Another major limitation of the probabilistic model checking approach is state space explosion [7] as has been indicated in [27,12] that increasing the number of communicating stations is not feasible in their analysis due to this problem. Similarly, to the best of our knowledge, it has not been possible to precisely reason about statistical quantities, such as expectation and variance, using probabilistic model checking so far. The most that has been reported in this domain is the approximate evaluation of expected values in a couple of model checkers, such as PRISM [26] and VESTA [40]. For example, in the PRISM model checker, the basic idea is to augment probabilistic models with costs or rewards: real values associated with certain states or transitions of the model. The expectation properties can thus be analyzed in terms of these reward or cost values by PRISM. These expectation properties are expressed and evaluated using computer arithmetic, which introduces some degree of approximation in the results. Similarly, the meaning ascribed to expectation properties is, of course, dependent on the definitions of the costs and rewards themselves and thus there is always some risk of verifying false properties.

Besides probabilistic model checking, rewriting logic based formal tools have also been used for the probabilistic analysis of wireless systems. For example, Real-Time Maude [33], which is a language and tool supporting the formal specification and analysis of real-time and hybrid systems, has been used for the analysis of the wireless sensor network algorithm OGDC in [34]. But the probabilistic behaviors of the wireless system under analysis are not modeled in formal terms here. Instead, they are analyzed using simulation based methods. Though, formal reasoning about probabilistic specifications is listed as a potential future direction. A possible solution to this aspect would be to explore a combined approach using Real-Time Maude with methods and tools for probabilistic systems, such as PMaude [1].

The proposed higher-order-logic theorem proving based approach tends to overcome the above mentioned limitations of state based formal probabilistic analysis techniques. Due to the high expressibility of higher-order logic, it allows us to analyze a wider range of wireless systems without any modeling limitations, such as the restrictiveness to Markovian models or the state-space explosion problem, and formally verify analytically complex properties, such as expectation and variance. On the other hand, higher-order-logic is an interactive approach and thus requires more human involvement and effort than the state based probabilistic analysis techniques.

To the best of our knowledge, higher-order-logic theorem proving has never been used for the probabilistic analysis of any wireless system so far. Though, some useful research related to the foundations of probabilistic analysis is available in the open literature. The foremost criteria for implementing a theorem proving based probabilistic analysis framework is to be able to formalize and verify random variables in higher-order logic. Hurd's PhD thesis [24] can be considered a pioneering work in this regard as it presents a methodology for the formalization and verification of probabilistic algorithms in the HOL theorem prover. Random variables are basi-

cally probabilistic algorithms and thus can be formalized and verified, based on their probability distribution properties, using the methodology proposed in [24]. In fact, [24] presents the formalization of some discrete random variables along with their verification, based on the corresponding PMF properties. Building upon Hurd's formalization framework [24], we have been able to successfully verify the sampling algorithms of a few continuous random variables [19] based on their Cumulative Distribution Function (CDF) properties as well. For comparison purposes, it is frequently desirable to summarize the characteristic of the distribution of a random variable by a single number, such as its expectation or variance, rather than an entire function. For example, it is easier to compare the performance of two wireless communication protocols based on the expected values rather than the CDFs of their message transmission delays. In [21,22], we extended Hurd's formalization framework with a formal definition of expectation. This definition is then utilized to formalize and verify the expectation and variance characteristics associated with discrete random variables that attain values in positive integers only.

3 Probabilistic Analysis Framework

The framework, given in Fig. 1, outlines the main idea behind the theorem proving based probabilistic analysis approach. The shaded boxes in this figure represent the fundamental requirements of conducting probabilistic analysis in a theorem prover. Like all system analysis tools, the input to this framework, depicted by solid rectangles with curved edges, is a description about the wireless system that needs to be analyzed and a set of properties that are required to be checked for the given system. For simplicity, we have divided the system properties into two categories, i.e., system properties related to discrete random variables and system properties related to continuous random variables.

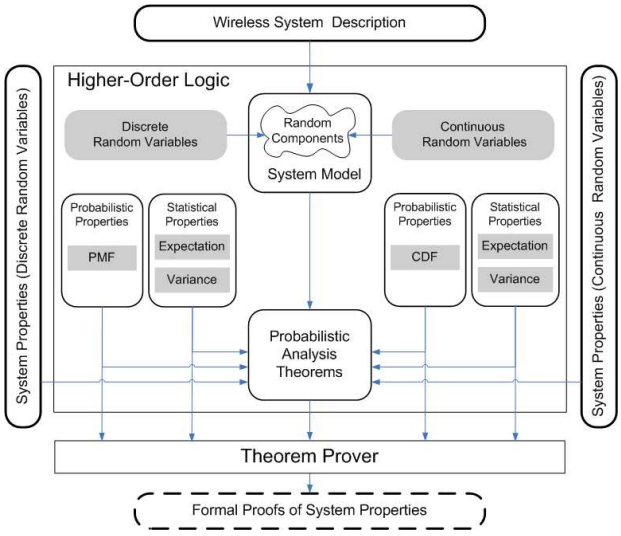

Fig. 1. Theorem Proving based Probabilistic Analysis Framework for Wireless Systems

The first step in conducting probabilistic analysis of a wireless system using a theorem prover is to construct a model of the system in higher-order-logic. For this

purpose, the foremost requirement is the availability of infrastructures that allow us to formalize all kinds of discrete and continuous random variables as higher-order-logic functions, which in turn can be used to represent random components of the given wireless system in its higher-order-logic model. The second step in theorem proving based probabilistic analysis is to utilize the formal model of the wireless system to express system properties as higher-order-logic theorems. The prerequisite for this step is the ability to express probabilistic and statistical properties related to both discrete and continuous random variables in higher-order-logic. All probabilistic properties of discrete and continuous random variables can be expressed in terms of their PMFs and CDFs, respectively. Similarly, most of the commonly used statistical properties can be expressed in terms of the expectation and variance characteristics of the corresponding random variable. Thus, we require the formalization of mathematical definitions of PMF, CDF, expectation and variance for both discrete and continuous random variables in order to be able to express the given wireless system's probabilistic and statistical properties as higher-order-logic theorems. The third step for conducting probabilistic analysis in a theorem prover is to formally verify the higher-order-logic theorems developed in the previous step using a theorem prover. For this verification, it would be quite handy to have access to a library of some pre-verified theorems corresponding to some commonly used properties regarding probability distribution functions, expectation and variance. Since, we can build upon such a library of theorems and thus speed up the verification process. Finally the output of the theorem proving based probabilistic analysis framework, depicted by the rectangle with dashed edges, is the formal proofs of system properties that ascertains that the given system properties are valid for the given wireless system.

In order to illustrate the construction details of the framework described above, we now describe the methodologies to fulfill its fundamental requirements.

3.1 Formalization of Discrete Random Variables and Verification of their PMF

A random variable is called discrete if its range, i.e., the set of values that it can attain, is finite or at most countably infinite [43]. Discrete random variables can be completely characterized by their PMFs that returns the probability that a random variable X is exactly equal to some value x, i.e., $Pr(X = x)$.

Discrete random variables are quite frequently used to model random phenomenon in the analysis of wireless systems. For example, the Bernoulli random variable is widely used to model the channel noise behavior [29], the Geometric random variable is often used to model the number of retransmission required to pass a message through a noisy wireless channel [42] and Poisson distribution is typically adopted to model message arrival patterns in wireless network analysis [44].

Discrete random variables can be formalized in higher-order-logic as deterministic functions with access to an infinite Boolean sequence $\mathbb{B}^\infty$; source of an infinite random bits with data type $(num \rightarrow bool)$ [24]. These deterministic functions make random choices based on the result of popping the top most bit in the infinite Boolean sequence and may pop as many random bits as they need for their computation. When the functions terminate, they return the result along with the remaining portion of the infinite Boolean sequence to be used by other functions.

Thus, a random variable that takes a parameter of type α and ranges over values of type β can be represented in HOL by the function

$$\mathcal{F} : \alpha \rightarrow B^{\infty} \rightarrow \beta \times B^{\infty}$$

For example, a $Bernoulli(\frac{1}{2})$ random variable that returns 1 or 0 with equal probability $\frac{1}{2}$ can be modeled as follows

```
⊢ bit = λs.  (if shd s then 1 else 0, stl s)
```

where the variable s represents the infinite Boolean sequence and the functions `shd` and `stl` are the sequence equivalents of the list operation *'head'* and *'tail'*. The function `bit` accepts the infinite Boolean sequence and returns a pair with the first element equal to either 0 or 1 and the second element equal to the unused portion of the infinite Boolean sequence, which in this case is the tail of the sequence.

Random variables can also be expressed in a more compact form using the general state-transforming monad where the states are the infinite Boolean sequences.

```
⊢ ∀ a,s.  unit a s = (a,s)
⊢ ∀ f,g,s.  bind f g s = g (fst (f s)) (snd (f s))
```

The HOL functions `fst` and `snd` above return the first and second components of a pair, respectively. The `unit` operator is used to lift values to the monad, and the `bind` is the monadic analogue of function application. All monad laws hold for this definition, and the notation allows us to write functions without explicitly mentioning the sequence that is passed around, e.g., function *bit* can be defined as

```
⊢ bit_monad = bind sdest (λb.  if b then unit 1 else unit 0)
```

where, `sdest` gives the head and tail of a sequence `s` as a pair (*shd* `s,` *stl* `s`).

In order to be able to formally reason about probabilistic properties of random variables, formalized according to the above methodology, we need to formalize a measure space of infinite Boolean sequences. Such a measure space can be used to define a probability function $\mathbb{P}$ from sets of infinite Boolean sequences to *real* numbers between 0 and 1 [24]. Thus, the domain of $\mathbb{P}$ is the set $\mathcal{E}$ of events of the probability. Both $\mathbb{P}$ and $\mathcal{E}$ can be defined using the Carathéodory's Extension theorem, which ensures that $\mathcal{E}$ is a σ-algebra: closed under complements and countable unions. Now, the formalized $\mathbb{P}$ and $\mathcal{E}$ can be used to prove probabilistic properties for random variables such as

```
⊢ ℙ {s | fst (bit s) = 1} = 1/2
```

where $\{x|C(x)\}$ represents a set of all elements x that satisfy the condition C.

The methodology described in this section is quite general and can be utilized to formalize most of the commonly used discrete random variables and formally verify their corresponding PMF relations in a theorem prover. For example, HOL definitions and PMF theorems for the Bernoulli, Uniform, Binomial and Geometric random variables can be found in [24,21,18].

3.2 Formalization of Continuous Random Variables and Verification of their CDF

A random variable is called continuous if it ranges over a continuous set of numbers [43]. A continuous set of numbers, sometimes referred to as an interval, contains all

real numbers between two limits. Continuous random variables can be completely characterized by their CDFs that return the probability that a random variable X is exactly less than or equal to some value x, i.e., $Pr(X \leq x)$.

Many wireless system models can only be constructed using continuous random variables. Examples include the modeling of inter-arrival delays between requests to a wireless host using the Exponential random variable [43] and the modeling of mobile nodes displacement by the Uniform random variable [39].

The sampling algorithms for continuous random variables are non-terminating and hence require a different formalization approach than discrete random variables, for which the sampling algorithms are either guaranteed to terminate or satisfy probabilistic termination, meaning that the probability that the algorithm terminates is 1. One approach to address this issue is to utilize the concept of the nonuniform random number generation [11], which is the process of obtaining arbitrary continuous random numbers using a Standard Uniform random number generator. The main advantage of this approach is that we only need to formalize one continuous random variable from scratch, i.e., the Standard Uniform random variable, which can be used to model other continuous random variables by formalizing the corresponding nonuniform random number generation method.

Based on the above approach, [19] presents a methodology, illustrated in Fig. 2, for the formalization of all continuous random variables for which the inverse of the CDF can be represented in a closed mathematical form. The first step in this methodology is the formal specification of the Standard Uniform random variable and the formal verification of this definition by proving the corresponding CDF property. The Standard Uniform random variable can be formalized using the methodology for the formalization of discrete random variables, described in the last section, and the formalization of the mathematical concept of limit of a *real* sequence [17] as the following sampling algorithm

$$\sum_{k=0}^{\infty} (\frac{1}{2})^{k+1} X_k \tag{1}$$

where X_k denotes the outcome of the k^{th} random bit; $True$ or $False$ represented as 1 or 0 respectively. The formalization and verification details are outlined in [20].

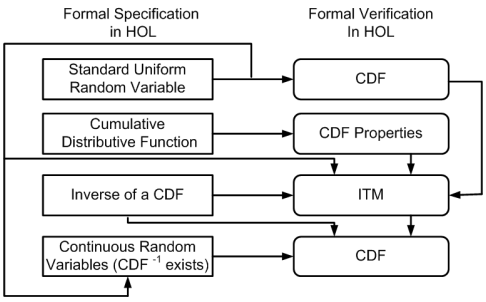

Fig. 2. Methodology for the Formalization of Continuous Random Variables

The second step in the methodology for the formalization of continuous probability distributions, given in Fig. 2, is the formalization of the CDF and the verification of its classical properties. This is followed by the formal specification of the mathematical concept of the inverse function of a CDF. This formal specifica-

tion, along with the formalization of the Standard Uniform random variable and the CDF properties, can be used to formally verify the correctness of the Inverse Transform Method (ITM) [11]. The ITM is a well known nonuniform random generation technique for generating nonuniform random variates for continuous probability distributions for which the inverse of the CDF can be represented in a closed mathematical form. Mathematically, it can be expressed for a random variable X with CDF F using the Standard Uniform random variable U as follows

$$Pr(F^{-1}(U) \leq x) = F(x) \tag{2}$$

and its formal proof can be found [20].

Based on the methodology of Fig. 2, the formalized Standard Uniform random variable can now be used to formally specify any continuous random variable for which the inverse of the CDF can be expressed in a closed mathematical form as $X = F^{-1}(U)$. Whereas, the CDF of this formally specified continuous random variable, X, can be verified, based on simple arithmetic reasoning, using the formal proof of the ITM. Using the above mentioned methodology, [19] presents the formal specification of four commonly used continuous random variables; Exponential, Uniform, Rayleigh and Triangular. The correctness of these random variables is also verified in [19] by proving their corresponding CDF properties in HOL.

3.3 Formalization and Verification of Statistical Properties

The third fundamental component of the wireless system analysis framework, given in Fig. 1, is the ability to formalize and verify statistical properties for random variables. Statistical characteristics, like expectation, play a major role in performance analysis of wireless systems as they tend to summarize the probability distribution characteristics of a random variable in a single number that are easy to compare.

The first and the foremost step towards the ability to reason about statistical properties in a theorem prover is the formalization of an expression for expectation in higher-order logic. Expectation basically provides the average of a random variable, where each of the possible outcomes of this random variable is weighted according to its probability [6]. The expectation for a function of a discrete random variable, which attains values in the positive integers only, is defined as follows [30].

$$Ex\_fn[f(X)] = \sum_{n=0}^{\infty} f(n)Pr(X = n) \tag{3}$$

where X is a discrete random variable and f represents a function of the random variable X. The expression of expectation, given in Equation (3), has been formalized in [22] as a higher-order-logic function using the formalization of the probability function, explained in Section 3.1 of this paper, and the higher-order-logic formalization of the summation of a *real* sequence, given in [17]. The expected value of a discrete random variable that attains values in positive integers can now be defined as a special case of Equation (3) when f is an identity function.

$$Ex[X] = Ex\_fn[(\lambda n.n)(X)] \tag{4}$$

Similarly, a variance function for discrete random variables can be defined in HOL using the expectation definitions given above as follows.

$$Var[X] = Ex\_fn[(\lambda n.(n - Ex[X])^2)X] \tag{5}$$

The above definitions can now be used to formally verify the classical properties of expectation and variance, given in Appendix A. The formal proofs of these properties using the HOL theorem prover can be found in [22,18]. The formal verification of these classical properties not only prove the correctness of the above definitions of expectation and variance but also facilitate the verification of expectation and variance characteristics of discrete random variables in HOL. For illustration purposes, [18] presents the formal verification of the expectation and variance relations for four discrete random variables: Bernoulli, Uniform, Binomial and Geometric.

For formally expressing and verifying statistical characteristics about continuous random variables, we require a higher-order-logic formalization of an integration function that can also handle functions with domains other than real numbers. To the best of our knowledge, a mature formalization for such an integral does not exist in the open literature so far. Thus, reasoning about statistical characteristics regarding continuous random components of a wireless system is not possible as of now. Though, the higher-order-logic formalization of some portions of the Lebesgue integration theory [37] may be extended to tackle such analysis problems.

4 Analysis of ARQ at the LLC Layer of GPRS

Due to the rapid development in mobile computing devices and emerging market of multimedia communications, the data-bearer service standard GPRS [5], which operates in packet-switched mode, was introduced as part of GSM phase 2+. GPRS uses the existing GSM infrastructure to provide high-speed (up to 270 kb/s) data communications, which is ideal for Multimedia Messaging Service (MMS) and for Internet communication services such as email and World Wide Web access. The biggest challenge in the design of GPRS is to maintain reliable data transfers without incurring too much delays under the erroneous nature of a wireless channel (due to distance losses, shadowing, and multipath fading). Higher layer protocols, such as the Transmission Control Protocol (TCP), are usually designed for wired channels that exhibit very low error rates and thus perform poorly if they are made responsible for the reliability of data transfers using a wireless channel [28]. Hence, the lower layers in GPRS stacks, e.g., the LLC and radio link control/medium access control (RLC/MAC), must be designed to address these issues of high error rates and higher layer-performance concerns.

The GPRS data transfer reliability model is as follows. A stop-and-wait ARQ [29] mechanism is implemented at the LLC to retransmit the erroneous LLC frames in order to ensure reliable transfers at the LLC peer-to-peer link. The LLC frames are passed to the RLC/MAC layer first, where they are segmented into RLC/MAC blocks of fixed size. These blocks are then transmitted through the radio channel one by one. The RLC/MAC layer also provides an ARQ mechanism and thus provides further error recovery over the radio channel. Our focus in this paper is to formally prove, using the theorem proving based probabilistic analysis approach described in Section 3, a probabilistic relation for the number of LLC frame retransmissions required for successfully transmitting a single LLC frame in terms of the probability of successful transmission of a single RLC/MAC block, say p, through the radio channel and the LLC frame size in RLC/MAC blocks, say n. Such an expression

plays a vital role in estimating the LLC frame size of the GPRS, for a given channel, to maximize performance. We also formally verify that the GPRS data reliability model ensures successful transmission of every LLC frame with probability 1. Our analysis approach is mainly inspired by a paper-and-pencil based analytical analysis of a similar problem presented in [10].

The first step according to the probabilistic analysis framework, given in Fig 1, is to describe the given system as a higher-order-logic function while representing its random components as random variables. In case of the above mentioned GPRS analysis problem, we need to develop a higher-order-logic function that describes the LLC frame transmission behavior in terms of the parameters p and n. The random component in this system is the behavior of the wireless channel, which allows data blocks to pass through with probability p. We formalized the LLC frame transmission behavior as a higher-order-logic predicate, i.e., a function that returns a Boolean value. Our predicate accepts three parameters: n, p and k, where k represents the number of transmission attempts. It returns $True$ if all n RLC/MAC blocks are successfully transmitted within k attempts and $False$ otherwise. The predicate can be expressed recursively in HOL as follows

```
Definition 1: LLC Frame Transmission Behavior
    ⊢ ∀ k p.  (llc_trans 0 k p = unit (True)) ∧
      ∀ n k p.  (llc_trans (n + 1) k p =
        bind (llc_trans n k p) (λa.  bind (prob_bino k p)
          (λb.  unit (if (b = 0) then False else a))))
```

where the function `prob_bino` represents the formalized Binomial random variable, given in [18]. A Binomial(k, p) random variable models an experiment that counts the number of successes in k independent Bernoulli(p) trials [9]. Thus, it is used in the above predicate to estimate the number of successful transmissions in k transmission attempts of an RLC/MAC block, as the behavior of a noisy wireless channel with successful transmission probability p can be modeled by a Bernoulli(p) random variable. The predicate `llc_trans` recursively checks the number of successful transmissions for each one of the n RLC/MAC blocks and returns $False$ if one or more of these blocks have no successful transmission in the k transmission attempts.

The second step according to the framework, given in Fig 1, is to utilize the formal model of the system to express the properties of interest as higher-order-logic theorems. In our case, we are interested in the probability that a single LLC frame consisting of n blocks is transmitted within k transmission attempts. The HOL theorem corresponding to this property can be expressed based on the predicate `llc_trans`, given in Definition 1, as follows

```
Theorem 1:
    ⊢ ∀ n k p k.  (0 ≤ p) ∧ (p ≤ 1) ⇒
      (ℙ {s | (fst (llc_trans n k p s))} = (1 - (1-p)^k)^n)
```

where $\mathbb{P}$ represents the formalization of the probability function, explained in Section 3. We verified the above theorem in HOL and the proof is primarily based on the PMF theorem of the Binomial random variable, verified in [18], and some arithmetic and probabilistic reasoning. For illustration purposes the proof sketch is provided in Appendix B.

Next, we formally prove the correctness of the GPRS data reliability model by verifying that the probability of successfully transmitting an LLC frame approaches 1 as the number of transmission trials becomes very very large. This property can be expressed in HOL, based on the predicate `llc_trans`, as follows

```
Theorem 2:
    ⊢ ∀ n k p k.  (0 ≤ p) ∧ (p ≤ 1) ⇒
      lim_{k→∞} (ℙ {s | (fst (llc_trans n k p s))}) = 1
```

using the HOL formalization of limit of a *real* sequence, given in [17]. The HOL proof of Theorem 2 is based on Theorem 1 and some classical properties of limit of a real sequence, verified in [17].

The above example clearly demonstrates the effectiveness of the theorem proving based wireless system analysis approach. Due to the formal nature of the model and inherent soundness of theorem proving, we have been able to verify probabilistic properties of the given system with 100% precision; a novelty which is not available in simulation. Similarly, due to the high expressibility of higher-order logic we have been able to formally reason about a problem that cannot be described as a Markov chain and thus cannot be analyzed using a probabilistic model checker. These additional benefits come at the cost of the time and effort spent, while formalizing the system and formally reasoning about its properties, by the user. But, the fact that we were building on top of already verified results in the theorem prover helped significantly in this regard as the analysis, described in this section, only consumed approximately 40 man-hours by an expert HOL user.

5 Conclusions

This paper advocates the usage of higher-order-logic theorem proving for the probabilistic analysis of wireless systems in order to be able to precisely analyze a wide range of problems. This approach can thus be of great benefit for the analysis of wireless systems used in safety critical applications, such as medicine and transportation. The paper provides a theorem proving based generic methodology for the probabilistic analysis of wireless systems. For illustration purposes, we present an analysis of ARQ mechanism at the LLC layer of the GPRS. To the best of our knowledge, this is the first time that a theorem prover has been used to conduct the probabilistic analysis of a wireless system.

There are many research directions in the field of using theorem provers for the probabilistic analysis of wireless systems that need to be explored. A couple of interesting ones include the ability to formalize and reason about statistical properties about continuous random variables and the ability to model Markov chains in higher-order-logic and reason about their probabilistic and statistical properties in a higher-order-logic theorem prover.

References

[1] G. Agha, J. Meseguer, and K. Sen. PMaude: Rewrite-based Specification Language for Probabilistic Object Systems. *Electronic Notes in Theoretical Computer Science*, 153(2):213–239, 2006.

[2] C. Baier, B. Haverkort, H. Hermanns, and J.P. Katoen. Model Checking Algorithms for Continuous time Markov Chains. *IEEE Transactions on Software Engineering*, 29(4):524–541, 2003.

[3] P. Ballarini and A. Miller. Model Checking Medium Access Control for Sensor Networks. In *International Symposium on Leveraging Applications of Formal Methods, Verification and Validation*, pages 255–262. IEEE, 2006.

[4] P. Barker, V. Vitsas, and A.C. Boucouvalas. Simulation Analysis of Advanced Infrared (alr) MAC Wireless Communications Protocol. *Circuits, Devices and Systems*, 149(3):193–197, 2002.

[5] C. Bettstetter, H. Vögel, and J. Eberspächer. GSM Phase 2+ General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface. *IEEE Communication Surveys*, 2(3), 1999.

[6] P. Billingsley. *Probability and Measure*. John Wiley, 1995.

[7] E.M. Clarke, O. Grumberg, and D.A. Peled. *Model Checking*. The MIT Press, 2000.

[8] COQ. http://pauillac.inria.fr/coq/, 2008.

[9] M. DeGroot. *Probability and Statistics*. Addison-Wesley, 1989.

[10] C. Demetrescu. LLC-MAC Analysis of General Packet Radio Service in GSM. *Bell Labs Technical Journal*, 4(3):37–50, 1999.

[11] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986.

[12] M. Fruth. Probabilistic Model Checking of Contention Resolution in the IEEE 802.15.4 Low-Rate Wireless Personal Area Network Protocol. In *International Symposium on Leveraging Applications of Formal Methods, Verification and Validation*, pages 290–297. IEEE, 2006.

[13] M.J.C. Gordon. Mechanizing Programming Logics in Higher-0rder Logic. In *Current Trends in Hardware Verification and Automated Theorem Proving*, pages 387–439. Springer, 1989.

[14] M.J.C. Gordon and T.F. Melham. *Introduction to HOL: A Theorem Proving Environment for Higher-Order Logic*. Cambridge University Press, 1993.

[15] R. Gowaikar, B. Hochwald, and B. Hassibi. Communication over a Wireless Network with Random Connections. *IEEE Transactions on Information Theory*, 52(7):2857–2871, July 2006.

[16] A. Gupta. Formal Hardware Verification Methods: A Survey. *Formal Methods in System Design*, 1(2-3):151–238, 1992.

[17] J. Harrison. *Theorem Proving with the Real Numbers*. Springer, 1998.

[18] O. Hasan and S. Tahar. Formal Verification of Tail Distribution Bounds in the HOL Theorem Prover. *Mathematical Methods in the Applied Sciences*. In-print.

[19] O. Hasan and S. Tahar. Formalization of the Continuous Probability Distributions. In *Automated Deduction*, volume 4603 of *LNAI*, pages 3–18. Springer, 2007.

[20] O. Hasan and S. Tahar. Formalization of the Standard Uniform Random Variable. *Theoretical Computer Science*, 382(1):71–83, 2007.

[21] O. Hasan and S. Tahar. Verification of Expectation Properties for Discrete Random Variables in HOL. In *Theorem Proving in Higher-Order Logics*, volume 4732 of *LNCS*, pages 119–134. Springer, 2007.

[22] O. Hasan and S. Tahar. Formal Verification of Expectation and Variance for Discrete Random Variables. Technical Report, Concordia University, Montreal, Canada, June 2007; http://hvg.ece.concordia.ca/Publications/TECH_REP/FVEVDR_TR07.

[23] H. Hermanns, J-P. Katoen, J. Meyer-Kayser, and M. Siegle. A Tool for Model-Checking Markov Chains. *Software Tools for Tech. Transfer*, 4(2):153–172, 2003.

[24] J. Hurd. *Formal Verification of Probabilistic Algorithms*. PhD Thesis, University of Cambridge, Cambridge, UK, 2002.

[25] A. Jain, S.S. Pawar, R. Upadhyay, and S.V. Charhate. Monte Carlo Simulation Based Error Performance Analysis of DS-CDMA System. In *Asia International Conference on Modelling and Simulation*, pages 230–233. IEEE Computer Society, 2008.

[26] M. Kwiatkowska, G. Norman, and D. Parker. Quantitative Analysis with the Probabilistic Model Checker PRISM. *Electronic Notes in Theoretical Computer Science*, 153(2):5–31, 2005. Elsevier.

[27] M(.) Kwiatkowska, G. Norman, and J. Sproston. Probabilistic Model Checking of the IEEE 802.11 Wireless Local Area Network Protocol. In *Joint International Workshop on Process Algebra and Probabilistic Methods, Performance Modeling and Verification*, volume 2399 of *LNCS*, pages 169–187. Springer, 2002.

[28] T.V. Lakshman and U. Madhow. The Performance of TCP/IP for Networks with High Bandwidth-delay Products and Random Loss. *IEEE/ACM Transactons on Networking*, 5(3):336–350, 1997.

[29] A. Leon Garcia and I. Widjaja. *Communication Networks: Fundamental Concepts and Key Architectures*. McGraw-Hill, 2004.

[30] A. Levine. *Theory of Probability*. Addison-Wesley series in Behavioral Science, Quantitative Methods, 1971.

[31] D.J.C. MacKay. Introduction to Monte Carlo Methods. In *Learning in Graphical Models, NATO Science Series*, pages 175–204. Kluwer Academic Press, 1998.

[32] M. Massink, D. Latella, and J-P. Katoen. Model Checking Dependability Attributes of Wireless Group Communication. In *Dependable Systems and Networks*, pages 711–720. IEEE, 2004.

[33] P.C. Olveczky and J. Meseguer1. Specification and Analysis of Real-Time Systems Using Real-Time Maude. In *Fundamental Approaches to Software Engineering*, volume 2984 of *LNCS*, pages 354–358. Springer, 2004.

[34] P.C. Ölveczky and S. Thorvaldsen. Formal Modeling and Analysis of the OGDC Wireless Sensor Network Algorithm in Real-Time Maude. In *Formal Methods for Open Object-based Distributed Systems*, volume 4468 of *LNCS*, pages 122–140, 2007.

[35] L.C. Paulson. *Isabelle: A Generic Theroem Prover*, volume 828 of *LNCS*. Springer, 1994.

[36] PVS. http://pvs.csl.sri.com, 2008.

[37] S. Richter. *Formalizing Integration Theory, with an Application to Probabilistic Algorithms*. Diploma Thesis, Technische Universitat Munchen, Department of Informatics, Germany, 2003.

[38] J. Rutten, M. Kwaiatkowska, G. Normal, and D. Parker. *Mathematical Techniques for Analyzing Concurrent and Probabilisitc Systems*, volume 23 of *CRM Monograph Series*. American Mathematical Society, 2004.

[39] P. Santi, D.M. Blough, and F. Vainstein. A Probabilistic Analysis for the Range Assignment Problem in Ad Hoc Networks. In *Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing*, pages 212–220. ACM, 2001.

[40] K. Sen, M. Viswanathan, and G. Agha. VESTA: A Statistical Model-Checker and Analyzer for Probabilistic Systems. In *Proc. IEEE International Conference on the Quantitative Evaluation of Systems*, pages 251–252, 2005.

[41] A. Srinivasan. Randomized Algorithms and Probabilistic Analysis in Wireless Networking. In *Stochastic Algorithms, Foundations, and Applications*, volume 4665 of *LNCS*, pages 54–57. Springer, 2007.

[42] P. Tran-Gia and K. Leibnitz. Teletraffic Models and Planning in Wireless IP Networks. In *Wireless Communications and Networking Conference*, volume 2, pages 598–602. IEEE, 1999.

[43] R.D. Yates and D.J. Goodman. *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*. Wiley, 2005.

[44] J. Zheng and M. J. Lee. *A Comprehensive Performance Study of IEEE 802.15.4*. IEEE Press, 2004.

Appendix A: Expectation and Variance Properties

No.	Property	Mathematical Representation
1	Linearity of Expectation 1	$Ex[\sum_{i=1}^{n} R_i] = \sum_{i=1}^{n} Ex[R_i]$
2	Linearity of Expectation 2	$Ex[a + bR] = a + bEx[R]$
3	Markov's Inequality	$Pr(X \geq a) \leq \frac{Ex[X]}{a}$
4	Alternate Definition of Variance	$Var[R] = Ex[R^2] - (Ex[R])^2$
5	Linearity of Variance	$Var[\sum_{i=1}^{n} R_i] = \sum_{i=1}^{n} Var[R_i]$
6	Chebyshev's Inequality	$Pr(\|X - Ex[X]\| \geq a) \leq \frac{Var[X]}{a^2}$

Table 1
Expectation and Variance Properties

Appendix B: HOL Proof for Theorem 1

We proceed to verify Theorem 1 by performing inductance on the variable n, which generates the following two subgoals.

$$\mathbb{P}\{\texttt{s} \mid \texttt{fst}\ (\texttt{llc\_trans}\ \texttt{0}\ \texttt{k}\ \texttt{p}\ \texttt{s})\ \}\ =\ (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{\texttt{0}}$$

$$\begin{gathered}\mathbb{P}\{\texttt{s} \mid \texttt{fst}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})\} = (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{\texttt{n}}\ \Rightarrow \\ \mathbb{P}\{\texttt{s} \mid \texttt{fst}\ (\texttt{llc\_trans}\ (\texttt{n}\ +\ \texttt{1})\ \texttt{k}\ \texttt{p}\ \texttt{s})\ \}\ =\ (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{(\texttt{n}\ +\ \texttt{1})}\end{gathered}$$

The base case, i.e., the first subgoal above, can be simply proved using the definition of the function `llc_trans`, given in Definition 1, the probability law $P(\bigcup) = 1$ and some arithmetic reasoning. Whereas, we proceed with the proof of the step case, i.e., the second subgoal above, by rewriting it using the definition of the function `llc_trans` and simplifying it using some arithmetic reasoning as follows.

$$\begin{gathered}\mathbb{P}\{\texttt{s} \mid \texttt{fst}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})\} = (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{\texttt{n}}\ \Rightarrow \\ \mathbb{P}\{\texttt{s} \mid (\texttt{fst}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})\ \wedge \\ \neg(\texttt{fst}\ (\texttt{prob\_bino}\ \texttt{k}\ \texttt{p}\ (\texttt{snd}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})))\ =\ \texttt{0})\} \\ =\ (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{\texttt{n}}(\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})\end{gathered}$$

Now, using the statistical independence between the two events in the set on the LHS of the conclusion of the above implication and the probability law $P(A \cap B) = P(A)P(B)$, the above subgoal can be simplified as follows.

$$\begin{gathered}\mathbb{P}\{\texttt{s} \mid \texttt{fst}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})\} = (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{\texttt{n}}\ \Rightarrow \\ \mathbb{P}\{\texttt{s} \mid (\texttt{fst}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})\} \\ \mathbb{P}\{\texttt{s} \mid \neg(\texttt{fst}\ (\texttt{prob\_bino}\ \texttt{k}\ \texttt{p}\ (\texttt{snd}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})))\ =\ \texttt{0})\} \\ =\ (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})^{\texttt{n}}(\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})\end{gathered}$$

Using the assumption in the above subgoal along with some arithmetic reasoning we get the following subgoal

$$\mathbb{P}\{\texttt{s} \mid \neg(\texttt{fst}\ (\texttt{prob\_bino}\ \texttt{k}\ \texttt{p}\ (\texttt{snd}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})))\ =\ \texttt{0})\} = (\texttt{1}\ -\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})$$

which can be rewritten using the complement law of the probability $P(\bar{A}) = 1 - P(A)$ as follows

$$\mathbb{P}\{\texttt{s} \mid (\texttt{fst}\ (\texttt{prob\_bino}\ \texttt{k}\ \texttt{p}\ (\texttt{snd}\ (\texttt{llc\_trans}\ \texttt{n}\ \texttt{k}\ \texttt{p}\ \texttt{s})))\ =\ \texttt{0})\}\ =\ (\texttt{1}\ -\ \texttt{p})^{\texttt{k}})$$

This subgoal can now be verified using the PMF relation of the Binomial random variable (`prob_bino`), given in [18], along with some arithmetic reasoning. This also concludes the proof of Theorem 1 in HOL.