Modeling Wireless Local Loop with General Call Holding Times and Finite Number of Subscribers

Hsien-Ming Tsai, *Student Member*, *IEEE*, and Yi-Bing Lin, *Senior Member*, *IEEE*

Abstract—This paper proposes an analytic model to compute the loss probability for Wireless Local Loop (WLL) with a finite number of subscribers. The number of trunks between the WLL concentrator and the base station controller is less than the total number of radio links in the WLL. This model is validated against the simulation results. The execution of our model is efficient compared with simulation. However, its time complexity is higher than several existing analytic models that approximate the loss probability for WLL. Therefore, we design an efficient WLL network planning procedure (in terms of time complexity and accuracy) that utilizes the approximate analytic models to provide small ranges for selecting the values of system parameters. Our model is then used to accurately search the operation points of WLL within the small ranges of the system parameter values. This paper proves that the performance of WLL with limited trunk capacity and finite subscriber population is not affected by the call holding time distributions. Based on our model, we illustrate WLL design guidelines with several numerical examples.

Index Terms—Loss probability, Engset product form, supplemented generalized semi-Markov process, wireless local loop.

1 Introduction

Wireless local loop (WLL) provides two-way communication services to stationary or near-stationary users within a small service area. This technology is intended to replace the wireline local loop. In telephony, *local loop* is defined as the transmission circuits between a *Local Exchange* (LE) and *Customer Premise Equipment* (CPE). The trunks start from the LE in the local loop and are broken into several smaller bundles of circuits after some distance from the LE. These circuits are eventually separated into "drops" for individual subscribers. The cost of the local loop tends to be dominated by these drops on the end-user side, which is typically referred to as the expensive "last mile." This is particularly true for rural areas. The LE is typically the first point-of-traffic concentration in the *public switched telephone network* (PSTN), especially for older installations where, on the line side of the LE, all facilities from the line-interface card to the CPE are dedicated to a single telephone number. New installations connect residential neighborhoods or business campuses to the LE and use statistical multiplexers to concentrate traffic. However, the last few hundred yards of wiring from a residence to the statistical multiplexer in the local loop is always dedicated. Compared with the wireline local loop, WLL offers advantages such as ease of installation and deployment (installation of expensive copper cables can be avoided) and concentration of resources [15], [16].

- *The authors are with the Department of Computer Science and Information Engineering, National Chiao Tung University, Taiwan, Republic of China. E-mail: {jtsai, liny}@csie.nctu.edu.tw.*

Manuscript received 18 Dec. 2000; revised 13 Sept. 2001; accepted 27 Nov. 2001.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number 113317.

Fig. 1 illustrates a typical WLL architecture [21], [5], [6], [19], [4]. This WLL architecture consists of *Subscriber Terminals* (STs), *Base Stations* (BSs), the *Base Station Controller* (BSC), the *Concentrator*, and the *Operations, Administration, and Maintenance Center* (OA&M Center). These components are described as follows:

Subscriber Terminal: An ST is colocated with the CPE (e.g., telephone set), which is responsible for converting and delivering speech and control signals between the CPE (through the subscriber telephone line) and the corresponding BS (through the air interface).

Base Station: There are M BSs in the system. For $1 \leq i \leq M$, there are N_i STs in the radio coverage area of the *i*th BS. This BS is equipped with c_i radio channels and is connected to the BSC with c_i backhaul transmission lines.

Base Station Controller: The BSC controls the concentrator and BSs and STs to perform call setup and release between the PSTN and CPEs. The BSC connects to the concentrator with C trunks.

Concentrator: The concentrator performs concentrating and mapping functions between the subscriber lines to the LE and the trunk circuits to the BSC. The number of subscriber lines between the concentrator and the LE is equal to the number of CPE/STs in the WLL network (i.e., $\sum_{i=1}^{M} N_i$).

Operations, Administration, and Maintenance Center: The OA&M Center is responsible for operating, controlling, and monitoring the whole WLL network. An example of WLL OA&M design and implementation can be found in [9].

Based on this architecture, we have developed a commercial WLL product under the contract with Eumitcom Technology Inc. The implementation details can be

0018-9340/02/$17.00 © 2002 IEEE

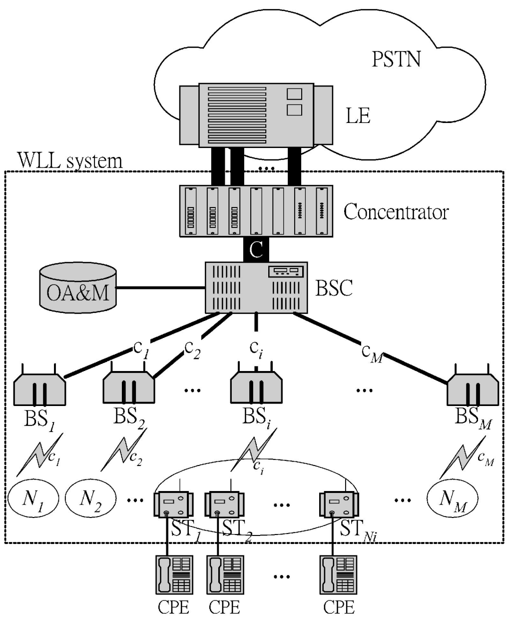


Fig. 1. A typical WLL architecture.

found in [22]. To simplify our discussion, we will ignore the concentrator and always say that the C trunks connect BSC and PSTN. The reader should bear in mind that the term "PSTN" means "concentrator plus PSTN."

The performance of a WLL network is affected by the capacities of BSC (i.e., the number C of trunks) and BSs (i.e., the number c_i of radio channels). Since the simultaneous ongoing calls in a WLL system are expected to be much smaller than the number of subscribers in the system, it is typical in network planning that

$$N_i > c_i \text{ and } C \leq \sum_{i=1}^{M} c_i.$$

To determine the C and c_i values, several models have been proposed to study WLL, including the Erlang-B formula [13], [7], [4], Engset-Syski (ES) model [20], and Erlang Product Form (EPF) model [10], [2], [11], [3], [23], [12]. These models either assume $C = \sum_{i=1}^{M} c_i$ (Erlang-B and Engset) or $N_i = \infty$ (Erlang-B and EPF). Therefore, they can only be used as approximate modeling of a general WLL system in a primary study. These models will be described in Section 2.

In Section 3, we propose an exact analytic model for general WLL systems where N_i ($1 \leq i \leq M$) are finite and $C \leq \sum_{i=1}^{M} c_i$. With this model, we prove that the performance of WLL is not affected by the call holding time distributions. This result is quite different from that for mobile networks [8]. Specifically, in a mobile network where handoff may occur, the call holding time distribution has significant impact on the output measures. In other words, the nice property we will prove for WLL cannot be found in the mobile network, which implies that we can design a WLL capacity planning procedure with much less complexity than that for a mobile network. In Section 4, we validate our analytic model with the simulation experiments and investigate the time and space complexities of the model. Section 5 provides several design guidelines for WLL resource planning. To speed up the modeling process, we also propose a procedure that utilizes the approximate analytic models to quickly identify small ranges of values for the WLL system parameters. Then, the engineered operation points of the WLL can be determined by investigating the system parameter values in these small ranges using our exact analytic model.

TABLE 1
Notations

Notation	Description
	System Parameters
M	number of BSs
N_i	number of STs served by the ith BS
N	number of STs served by a BS (when all BSs have the same numbers of STs)
C	number of telephone trunks in the BSC
c_i	number of radio channels in the ith BS
c	number of radio channels in a BS (when all BSs have the same numbers of radio channels)
	Traffic Characteristics
λ	call arrival rate per ST
$1/\mu$	mean of the call holding time
ρ	traffic intensity per ST ($\rho = \frac{\lambda}{\mu}$)
	Output Measure
P_l	call loss probability
P_b	system blocking probability

2 Approximate Analytic Models

We assume that the call arrivals to an ST (for both incoming and outgoing calls) are a Poisson stream with rate λ. The call holding time has a general distribution with mean $\frac{1}{\mu}$. Therefore, the traffic intensity per ST is $\rho \equiv \frac{\lambda}{\mu}$, which is measured in Erlang per ST. A call to an ST is lost if all radio channels in the corresponding BS are occupied or all trunks in the BSC are busy. The probability that the above event occurs is called the *(call) loss probability* [18], [13], which is denoted by P_l. Another output measure of WLL is the *(system) blocking probability* P_b, which is defined as the probability that system resources are not available in a given observation period. For a system with infinite subscriber population, P_l and P_b are the same. On the other hand, these two probabilities are different for a system with finite subscriber population [20]. The relationship between P_l and P_b will be given later (see (12)).

Based on the above assumptions, this section describes three approximate analytic models for WLL systems. These models are *Erlang*, *Engset-Syski*, and *Erlang Product Form*. The notations used in this paper are listed in Table 1.

2.1 The Erlang and Engset-Syski Models

Erlang-B formula [13], [7], [4] is often used to model blocking for telecommunication systems with infinite traffic sources. For a general WLL system with parameters C, M, N_i, and c_i, where $1 \leq i \leq M$, the Erlang model provides approximate performance measure (i.e., the blocking probability) as follows: In this model, the net call traffic to the ith BS is approximated to be

$$\rho_i^* = N_i \rho.$$

By assuming that there is no blocking in BSC, the Erlang-B formula computes the blocking probability of the *i*th BS as:

$$P_{b,Erlang}(\rho_i^*, c_i) = \frac{\rho_i^{*c_i}/c_i!}{\sum_{j=0}^{c_i}(\rho_i^{*j}/j!)}.$$

Similarly, by approximating the net call traffic ρ_{BSC} to BSC as

$$\rho_{BSC} = \left(\sum_{i=1}^{M} N_i\right)\rho$$

and, assuming that no blocking occurs in the BSs, the Erlang-B formula gives the blocking probability of BSC as

$$P_{b,Erlang}(\rho_{BSC}, C) = \frac{\rho_{BSC}^C/C!}{\sum_{j=0}^{C}(\rho_{BSC}^j/j!)}.$$

The major difference between the Engset-Syski (ES) model [20] and the Erlang model is that ES assumes finite traffic sources (i.e., limited number of STs in the WLL system). For a WLL system with finite numbers of trunks and subscribers, ES provides approximate performance measure as follows: By assuming no call loss in BSC, the ES model, which we refer to as ES-c, computes the loss probability of the *i*th BS as

$$P_{l,ES-c}(i) = \frac{\rho^{c_i}\binom{N_i-1}{c_i}}{\sum_{j=0}^{c_i}\rho^j\binom{N_i-1}{j}}.$$

Similarly, by assuming no call loss in BSs, the ES model, which we refer to as ES-C, computes the loss probability of BSC as

$$P_{l,ES-C} = \frac{\rho^C\binom{N^*-1}{C}}{\sum_{j=0}^{C}\rho^j\binom{N^*-1}{j}} \quad \text{where} \quad N^* = \sum_{i=1}^{M} N_i.$$

To consider call loss effect at both BSs and BSC, an approximation was proposed [11], [12]. If we assume that the call losses at BSs and BSC are independent (in fact, they are not), then the probability that a call is not lost (neither in BSC nor in BS) is equal to the product of individual nonloss probabilities of BS and BSC. This approximation is referred to as the *Hybrid Engset-Syski* (H-ES) model. For a call arrival to an ST in the *i*th BS, the loss probability $P_{l,H-ES}(i)$ estimated by H-ES is

$$P_{l,H-ES}(i) = 1 - (1 - P_{l,ES-c}(i))(1 - P_{l,ES-C}). \tag{1}$$

2.2 Erlang Product Form

The Erlang Product Form (EPF) model [10], [2], [11], [3], [23], [12] has been used to investigate telecommunication networks with arbitrary network topologies and infinite number of subscribers. For a WLL network with M BSs, there are M possible routes to connect calls between the PSTN and the CPEs (i.e., $\text{PSTN} \leftrightarrow \text{BSC} \leftrightarrow \text{BS}_i$, for $1 \le i \le M$). Consider a stochastic process $\mathbf{N}^*(\mathbf{t})$ with state vector $\mathbf{n}$ of size M, where $\mathbf{n} = [n_1, \ldots, n_i, \ldots, n_M]$ and n_i represents the number of outstanding calls on the *i*th BS. It is clear that the total number of radio channels and trunks occupied by the outstanding calls should be no more than the capacities of the nodes (the BS and the BSC) in the route. Thus, a legal state $\mathbf{n}$ must satisfy the following constraints:

$$0 \le n_1 + n_2 + \ldots + n_M \le C. \tag{2}$$

$$0 \le n_i \le c_i, \quad \text{for} \quad 1 \le i \le M. \tag{3}$$

Constraint (2) indicates that the number of outstanding calls in the WLL should be no more than the number C of trunk circuits in BSC. Constraint (3) states that the number of outstanding calls on the *i*th BS should be no more than the number c_i of radio channels. The equivalent traffic intensity for the *i*th route (i.e., via the *i*th BS) is $N_i\rho$. Let Γ be the state space of the legal states that satisfy (2) and (3). In EPF, the stationary probability of the state $\mathbf{n} \in \Gamma$ can be computed as

$$p(\mathbf{n}) = G \prod_{1 \le i \le M}\left[\frac{(N_i\rho)^{n_i}}{n_i!}\right], \tag{4}$$

where

$$G = \left\{\sum_{\mathbf{n}\in\Gamma}\left[\prod_{1\le i\le M}\frac{(N_i\rho)^{n_i}}{n_i!}\right]\right\}^{-1}.$$

A route that connects calls between the *i*th BS and the BSC is blocked when all radio channels in the BS are busy or when all trunks in BSC are busy. In other words, this route is blocked if the stochastic process is at state $\mathbf{n} \in \Gamma(i)$, where

$$\Gamma(i) = \left\{[n_1 \ldots n_M] \in \Gamma \,\middle|\, n_i = c_i \text{ or } \sum_{1\le j\le M} n_j = C\right\}.$$

Thus, in EPF, the blocking probability $P_{b,EPF}$ for the *i*th BS is given by

$$P_{b,EPF}(i) = \sum_{\mathbf{n}\in\Gamma(i)} p(\mathbf{n}).$$

Since the system considered in EPF has infinite subscriber population, the call loss probability is the same as the system blocking probability [20]. That is,

$$P_{l,EPF}(i) = P_{b,EPF}(i).$$

3 Exact Analytic Model

This section proposes an exact analytic model to compute the loss probability for the WLL systems. We first derive the model with the exponential call holding time distribution. The derivation is based on the concept of process reversibility [10] and the relationship between the loss probability and the blocking probability [18], [20]. Then, we prove that the exponential assumption can be relaxed for the analytic model with an arbitrary call holding time distribution. The derivation is based on the concept of supplemented generalized semi-Markov process (SGSMP) and restricted flow equation [1], [2]. We also refer the reader to [13], [18] for background knowledge of stochastic process and queueing theory required to understand the derivations.

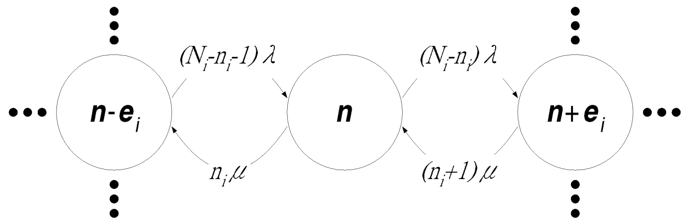


Fig. 2. Markov flow diagram for state $\mathbf{n}$ $(1 \le i \le M)$.

3.1 The Model with Exponential Call Holding Time

For the purpose of discussion, we use the term "the *i*th node" to represent the *i*th BS for $1 \le i \le M$, and the $M+1$st node represents the BSC. We propose a model similar to those in [2], [11] as follows: Define the *capacity vector* $\mathbf{B}$ of the WLL network as a vector of size $M+1$:

$$\mathbf{B} = [B_1, B_2, \ldots, B_i, \ldots, B_{M+1}],$$

where for $1 \le i \le M$, $B_i = c_i$ is the capacity of the *i*th BS, and $B_{M+1} = C$ is the capacity of the BSC. Since a call for the *i*th BS consumes one channel at the *i*th BS and one trunk at the BSC, the *occupancy matrix* A of the WLL network is defined as

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \\ 1 & 1 & \ldots & 1 & 1 \end{pmatrix}_{(M+1)\times M},$$

where A_{ij} is the number of *i*th node's resources utilized by a call connected to the *j*th BS. For the demonstration purpose, we assume that, for each call, one call party is in the WLL system and another call party is from the PSTN. To accommodate intra-WLL calls where both parties are in the WLL system, we only need to modify the occupancy matrix. Results for intra-WLL call performance with infinite traffic sources can be found in [14]. The model can also be easily extended to accommodate multimedia calls that consume multiple radio channels and trunks.

Consider a stochastic process $\mathbf{N}(t)$ with vector states of size M:

$$\mathbf{n} = [n_1, \ldots, n_i, \ldots, n_M],$$

where n_i represents the number of outstanding calls at the *i*th BS. For a legal state $\mathbf{n}$, the following constraint holds:

$$\mathbf{0} \le A\mathbf{n} \le \mathbf{B}, \tag{5}$$

where $\mathbf{0} = [0, 0, \ldots, 0]$ is the zero vector of size $M+1$. When the stochastic process is at state $\mathbf{n} = [0, \ldots, 0]$, it means that no call is in progress in the WLL network. Thus, from (5), the set Γ of the legal states is

$$\Gamma = \{\mathbf{n} | \mathbf{0} \le A\mathbf{n} \le \mathbf{B}\}.$$

Assume that the call arrivals to an ST at the *i*th BS are a Poisson stream with rate λ and the call holding time has an exponential distribution with mean $\frac{1}{\mu}$. It is clear that $\mathbf{N}(t)$ is a Markov process.

Fig. 2 shows the transition for state $\mathbf{n} = [n_1, \ldots, n_i, \ldots, n_M]$ in the Markov flow diagram. In this figure, $\mathbf{e}_i$ is the identity vector with 1 in the *i*th position. That is,

$$\mathbf{n} + \mathbf{e}_i = [n_1, \ldots, n_i + 1, \ldots, n_M].$$

For a BS i, when there are n_i calls in progress, only $N_i - n_i$ STs can generate new calls. Therefore, when the stochastic process is at state $\mathbf{n}$, the call arrival rate to the *i*th BS is $(N_i - n_i)\lambda$ and the call completion rate is $n_i\mu$, where $0 \le n_i \le \min(c_i, N_i)$. It is apparent that both call arrival and call completion rates for $\mathbf{N}(t)$ are state dependent.

We show that $\mathbf{N}(t)$ is reversible as follows: (The definition and basic concepts of reversible process can be found in p. 5 in [10].) Consider the stochastic process $\mathbf{N}^*(t)$ for EPF (see Section 2). Theorem 1 in [2] showed that $\mathbf{N}^*(t)$ is reversible. Note that $\mathbf{N}^*(t)$ is the same as $\mathbf{N}(t)$ except that the call arrival rate for $\mathbf{N}^*(t)$ is state independent. By Lemma 1.9 in [10], the change of the transition rate of a reversible Markov process does not destroy the reversibility. Thus, $\mathbf{N}(t)$ is also reversible. From Lemma 1.1 and Theorem 1.2 in [10], the process $\mathbf{N}(t)$ has stationary distribution $\pi(\mathbf{n})$ and the detailed balance equations exist. From Fig. 2, the detailed balance equations are derived as

$$(N_i - n_i)\lambda\pi(\mathbf{n}) = (n_i + 1)\mu\pi(\mathbf{n} + \mathbf{e}_i) \quad \text{for } \mathbf{n}, \mathbf{n} + \mathbf{e}_i \in \Gamma. \tag{6}$$

After rearrangement, (6) can be written as an Engset product form

$$\begin{aligned} \pi(\mathbf{n}) &= \left\{ \prod_{1 \le i \le M} \frac{[(N_i - n_i + 1)\lambda][(N_i - n_i + 2)\lambda]\ldots(N_i\lambda)}{\mu(2\mu)\ldots(n_i\mu)} \right\} \\ &\quad \times \pi(\mathbf{0}) \\ &= \left[\prod_{1 \le i \le M} \rho^{n_i} \binom{N_i}{n_i} \right] \pi(\mathbf{0}). \end{aligned} \tag{7}$$

From the probability normalization condition, we have

$$\sum_{\mathbf{n} \in \Gamma} \pi(\mathbf{n}) = 1. \tag{8}$$

Substitute (7) into (8),

$$\pi(\mathbf{0}) = \left\{ \sum_{\mathbf{n} \in \Gamma} \left[\prod_{1 \le i \le M} \rho^{n_i} \binom{N_i}{n_i} \right] \right\}^{-1}. \tag{9}$$

From (7) and (9), we have

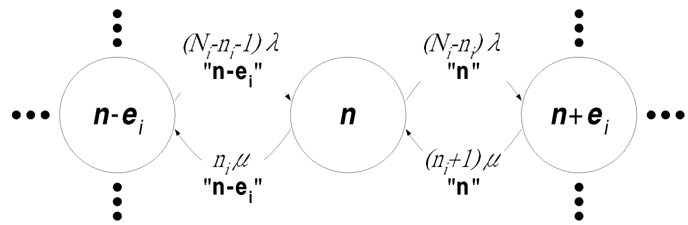


Fig. 3. Labeled flow diagram for state $\mathbf{n}$ $(1 \le i \le M)$.

$$\pi(\mathbf{n}) = \left[\prod_{1\le i\le M} \rho^{n_i}\binom{N_i}{n_i}\right] \times \left\{\sum_{\mathbf{n}\in\Gamma}\left[\prod_{1\le i\le M}\rho^{n_i}\binom{N_i}{n_i}\right]\right\}^{-1}. \quad (10)$$

A route that connects calls between the *i*th BS and the BSC is blocked when all radio channels in the BS are busy or when all trunks in the BSC are busy. In other words, this route is blocked if the stochastic process is at a state $\mathbf{n} \in \Gamma(i)$, where

$$\Gamma(i) = \left\{[n_1, \dots, n_M] \in \Gamma \middle| n_i = c_i \text{ or } \sum_{1\le j\le M} n_j = C\right\}.$$

Thus, the blocking probability $P_{b,Exact}(i)$ for the ith BS is given by

$$P_{b,Exact}(i) = \sum_{\mathbf{n}\in\Gamma(i)} \pi(\mathbf{n}). \quad (11)$$

A call to the *i*th BS is lost if the call route between the *i*th BS and the BSC is blocked when the call arrives. The loss probability can be derived from the blocking probability using the following relationship (see equation (10-38) in [18] or Section 5.2.3 in [20]):

$$P_{l,Exact}(i) = \left[\frac{\lambda_b(i)}{\lambda_a(i)}\right] \times P_{b,Exact}(i), \quad (12)$$

where

$$\lambda_b(i) = \sum_{\mathbf{n}\in\Gamma(i)}\left\{(N_i - n_i)\lambda\left[\frac{\pi(\mathbf{n})}{\sum_{\mathbf{n}\in\Gamma(i)}\pi(\mathbf{n})}\right]\right\}$$

is the call arrival rate for BS i when the route between the *i*th BS and the BSC is blocked, and

$$\lambda_a(i) = \sum_{\mathbf{n}\in\Gamma}[(N_i - n_i)\lambda\pi(\mathbf{n})]$$

is the average call arrival rate to the *i*th BS.

3.2 The Model with General Call Holding Time

Now, we prove that (10) holds for general call holding time distributions. As previously mentioned, if the call holding times are exponentially distributed, the stochastic process $\mathbf{N}(t)$ is a reversible Markov process that has the stationary distribution (10) satisfing the detailed balance equations (6). The Markov flow diagram for stationary $\mathbf{N}(t)$ is shown in Fig. 2.

Let the call holding times have a general distribution $H(u)$ with mean $\frac{1}{\mu}$. If $H(u)$ is not exponential, then the stochastic process $\mathbf{N}(t)$ is a generalized semi-Markov process (GSMP; see Section 2 in [1]). To prove the insensitivity property for $\mathbf{N}(t)$ (i.e., to prove that the stationary distribution of system states is insensitive to the call holding time distributions), the following supplemented generalized semi-Markov process (SGSMP) is introduced. For $1 \le i \le M$ and $1 \le k \le c_i$, let $[\mathbf{n}, \mathbf{u}] = [n_i, u_{i,k}]$ be an element of the state space of the SGSMP, where n_i represents the number of outstanding calls on the *i*th BS and $u_{i,k}$ represents the already processed time of the call utilizing the *k*th channel on the *i*th BS. Then, we can compute the stationary probability density $\pi^*(\mathbf{n}, \mathbf{u})$ of the SGSMP according to the algorithm in Section 3 in [1]: Starting with the Markov flow diagram in Fig. 2, we use "$\mathbf{n}$" as the label for the edge connecting $\mathbf{n}$ to $\mathbf{n} + \mathbf{e}_i$ and "$\mathbf{n} - \mathbf{e}_i$" for the edge connecting $\mathbf{n}$ to $\mathbf{n} - \mathbf{e}_i$, where $\mathbf{e}_i$ is the identity vector with 1 in the *i*th position as defined in (6). The labeled flow diagram [1], [2] is shown in Fig. 3. This diagram implies that for $\mathbf{n}, \mathbf{n} + \mathbf{e}_i \in \Gamma$ and $1 \le i \le M$, the restricted flow equations are

$$(N_i - n_i)\lambda\pi(\mathbf{n}) = (n_i + 1)\mu\pi(\mathbf{n} + \mathbf{e}_i), \quad (13)$$

which are the same as the detailed balance equations (6). Therefore, both (6) and (13) have the same solutions (10). According to Theorem in [1], the stationary probability density of the SGSMP is

$$\pi^*(\mathbf{n}, \mathbf{u}) = \pi(\mathbf{n})\mathbf{H}(\mathbf{u}),$$

where $\mathbf{H}(\mathbf{u}) = \prod_{1\le i\le M, 1\le k\le c_i}\left[\mu(1 - H(u_{i,k})\right]$. The stationary distribution of the GSMP is the marginal distribution $\pi^*(\mathbf{n})$ of the SGSMP, where

$$\begin{aligned}
\pi^*(\mathbf{n}) &= \int \pi^*(\mathbf{n}, \mathbf{u})d\mathbf{u} \\
&= \int \pi(\mathbf{n})\mathbf{H}(\mathbf{u})d\mathbf{u} \\
&= \int_0^\infty \cdots \int_0^\infty \pi(\mathbf{n}) \\
&\times \prod_{1\le i\le M, 1\le k\le c_i}\left[\mu\left(1 - H(u_{i,k})\right)\right]du_{1,1}\dots du_{M,c_M} \\
&= \pi(\mathbf{n}) \prod_{1\le i\le M, 1\le k\le c_i}\left[\mu\int_0^\infty\left(1 - H(u_{i,k})\right)du_{i,k}\right] \\
&= \pi(\mathbf{n}) \prod_{1\le i\le M, 1\le k\le c_i}\left[\mu\left(\frac{1}{\mu}\right)\right] \\
&= \pi(\mathbf{n}).
\end{aligned}$$

Equation (14) shows that the stationary distribution $\pi(\mathbf{n})$ for the Markov process is the same as the stationary

TABLE 2
Abbreviations for Performance Models

X (Abbr.)	Description
$Erlang-c$	The Erlang model with c servers
$ES-c$	The Engset-Syski model with c servers
$ES-C$	The Engst-Syski model with C servers
EPF	The Erlang Product Form model
$Exact$	The exact model
$H-ES$	The Hybrid Engset-Syski model
sim	The simulation model

distribution $\pi^*(\mathbf{n})$ of the GSMP. That is, the GSMP is insensitive to the call holding time distribution $H(u)$.

4 Accuracy and Complexities of the Analytic Models

This section investigates how accurate the analytic models can compute the loss probability P_l for a WLL system. Then, we compare the time and space complexities of these models. At the end of this section, we show, by combining these analytic models, how to provide quick and accurate modeling procedure to identify the engineered operation points (i.e., to select the values for the WLL system parameters) of a WLL system under specific workload.

4.1 Accuracy Analysis

The accuracy study of the analytic models validates the exact analytic model by the simulation experiments and investigates the errors of the approximate analytic models. For simplicity, assume that all BSs have the same numbers c of radio channels. They also have the same numbers N of STs. Our results can be easily generalized for BSs with various numbers of radio channels and STs. In the illustration examples, the values for the input parameters are $M = 6$, $1/\mu = 3$ min., and $\rho = 70m$ Erlang per ST. In simulation experiments, we consider various call holding time distributions (including exponential, uniform, and gamma distributions with different scale and shape parameters). The simulation results indicate that the loss probability is not affected by the call holding time distribution. This result is consistent with our insensitivity proof in Section 3.2. For the discussion purpose, we abbreviate the performance models as listed in Table 2. Two output measures are considered for an analytic model X:

- $P_{l,X}$: the loss probability computed based on model X.
- $\alpha_X = |\frac{P_{l,X}-P_{l,sim}}{P_{l,sim}}|$: the error of model X compared with the simulation result.

In every simulation experiment, $10^8 - 10^9$ call arrival events are executed to ensure that simulation result is stable.

Figs. 4, 5, 6, and 7 indicate that the exact model is consistent with the simulation results. These figures show that the exact model produces the same results as the simulation experiments ($\alpha_{Exact} < 1\%$ in most cases). On the other hand, the approximate models generate large errors.

In Fig. 4, $C = 72$ and $c = 12$. That is, the number of trunks is the same as the total number of the radio channels in the system (i.e., $C = cM = 72$) and no call loss will occur in BSC. In this case, ES-c becomes exact and EPF is the same as the Erlang model with c servers (i.e., Erlang-c). Fig. 4b indicates that the models with infinite traffic sources (i.e., Erlang-c or EPF) have large errors compared with the simulation results. Specifically, the error between Erlang-c (or EPF) and the simulation ranges from 40 percent to 350 percent. As N increases, this error decreases because the number of finite sources approaches infinite. As shown in Fig. 4, when investigating WLL with finite number of STs (especially when N is not very large), Erlang-c has large error in most cases. Thus, this model is not appropriate for modeling WLL.

The error of ES-C is around 100 percent, which implies that this model is very inaccurate for this particular example. The error is due to the fact that the example in Fig. 4 is only affected by c, but ES-C can only capture the effect of C. This result is independent of the N value.

In Fig. 5, $C = 42$ and $c = 12$; i.e., the number of trunks is about 40 percent less than the number of radio channels. The figure indicates that ES-C is accurate while the error of ES-c approaches 100 percent. Since the trunks in the BSC are bottleneck resources in this case, ES-c cannot capture the WLL performance. This result is independent of the N values except for the cases when N is small. When N is small, neither radio channels nor trunks in the BSC are bottleneck resources and both ES-C and ES-c partially capture the WLL performance. Both Figs. 4 and 5 indicate that the error of H-ES is only insignificantly affected by N. As we will explain later, both H-ES and EPF overestimate P_l. Although EPF assumes infinite traffic sources, it captures the blocking effects in both BSC and BS levels. Thus, when N is large, EPF provides better WLL performance insight than H-ES does.

Fig. 6 shows the effect of c on the accuracy of the approximate analytic models. In this figure, $C = 42$ and $N = 100$. In Fig. 6a, P_l of the ES-C model is constant because ES-C is independent of c. Fig. 6b shows that, when c is small (e.g., $c < 8$), the results of ES-c are almost identical to the exact model. In this case, call loss in BSC seldom occurs (i.e., radio channels are bottleneck resources) and ES-c captures the call loss characteristics of BS in the WLL. On the other hand, when c is large (e.g., $c > 12$), the results of ES-C are almost identical to the exact model. In this case, call loss seldom occurs in BSs (i.e., the trunks in the BSC are the bottleneck resources) and ES-C captures the call loss characteristics of BSC in the WLL. The error of H-ES is determined by the errors of both ES-C and ES-c. It turns out that the error of H-ES is large when c ranges from 6 to 12.

Fig. 7 shows the effect of C on the accuracy of the approximate analytic models. In this figure, $c = 12$ and $N = 100$. ES-C becomes accurate when C is small and ES-c becomes accurate when C is large. This phenomenon is the same as what we observed in Fig. 6.

To conclude, when c is the bottleneck resource in the WLL system, ES-c is a better approximation. On the other hand, when C is the bottleneck resource, ES-C is better.

Figs. 4a, 5a, 6a, and 7a indicate that ES-c and ES-C provide lower bounds for P_l and EPF and H-ES give upper bounds for P_l. The above statements can be formally proven. The intuition is given below. ES-c (ES-C) only

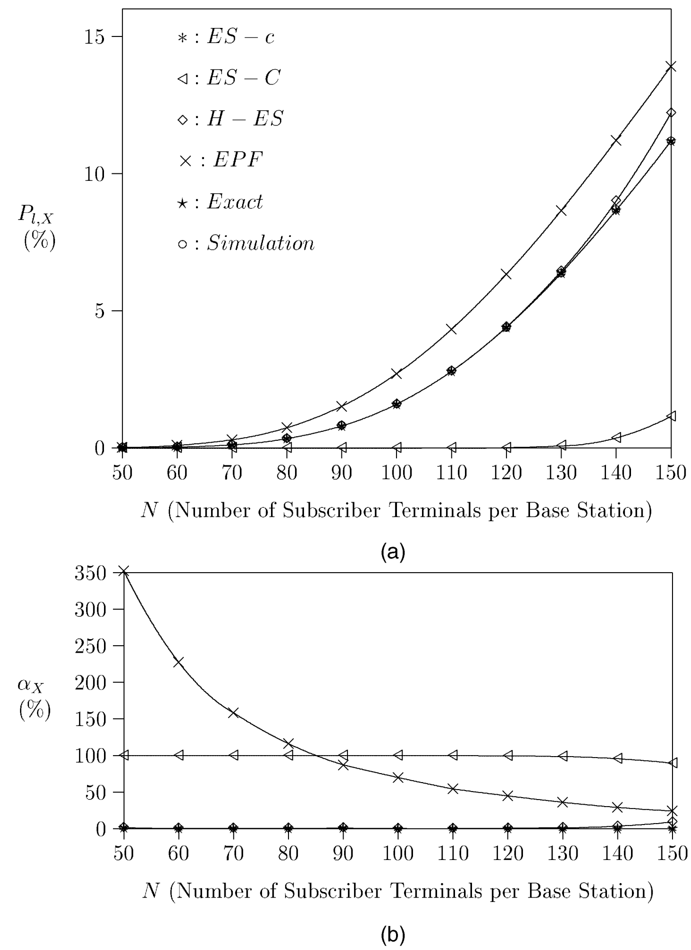


Fig. 4. Accuracy analysis of the analytic models ($C = 72,\ c = 12,\ M = 6,\ 1/\mu = 3\ \text{min.}$, and $\rho = 70m$ Erlang per ST). (a) Loss probability and (b) discrepancy of analytic models.

considers call loss in BS (BSC) and they ignore call loss in BSC (BS). Thus, they provide lower bounds for P_l. On the other hand, because the ES-c and ES-C approximations for P_l are lower bounds, from (1), H-ES provides an upper bound for P_l. For EPF, the assumption of infinite traffic sources always results in a larger offered load to the system compared with the finite traffic sources. Thus, P_l derived from EPF is larger than the true P_l value. Based on the above discussion, we have

$$\max\{P_{l,ES-c}, P_{l,ES-C}\} \leq P_{l,Exact} \leq \min\{P_{l,EPF}, P_{l,H-ES}\}. \tag{15}$$

4.2 Complexity Analysis

This section discusses the time and the space complexities of the analytic models. From [17], the time complexities of ES-c and ES-C are $O(c)$ and $O(C)$, respectively. According to (1), the time complexity of H-ES is $O(\max(C, c))$. Based on the recurrence equation in [17], the space complexities of ES-C and ES-c are both $O(1)$. From (1), the space complexity of H-ES is also $O(1)$.

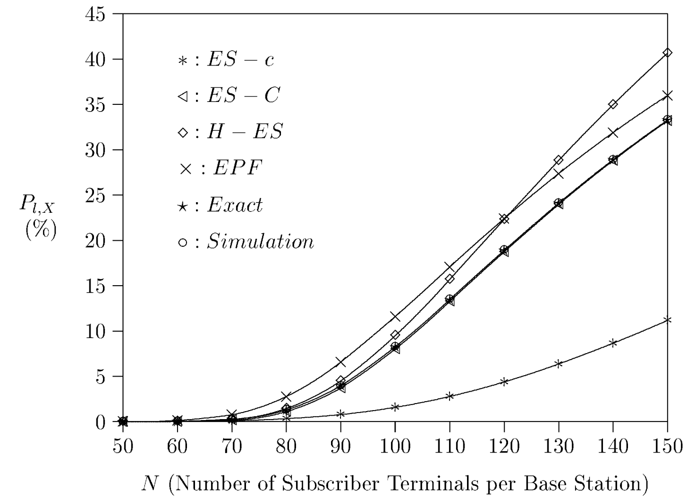
The time complexity of the exact model is derived as follows: Let $K(M, C, c)$ be the number of operations executed in the exact model for a WLL system with M BSs, C trunks, and c radio channels per BS. This number is determined by both the number of the legal states $\mathbf{n} \in \Gamma$ and the number of operations calculating the Engset product forms in (7). Since the number of legal states (see (5)) is less than $(c+1)^M$, an obvious upper bound for the time complexity of the exact model is $M(c+1)^M$, where $C \leq Mc$.

The number of legal states can be computed as follows: First of all, the trunks in the BSC are either idle or connected to BSs. Thus, C trunks are distributed into $M+1$ groups. For any group i in the first M groups, the trunks in this group, if any, are occupied by the calls to the ith BS. The trunks in the last group are idle. If there are $\Omega(M, C)$ ways to distribute the trunks, then

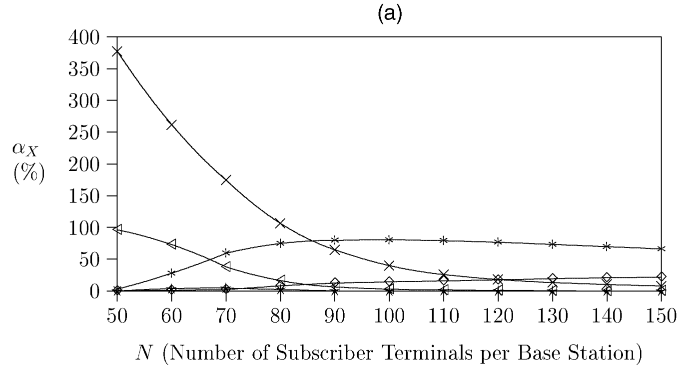
$$\Omega(M, C) = \binom{C+M}{M}.$$

However, each BS cannot consume more than c trunks (due to the limitation of the radio channel capacity). Thus, we must eliminate the combinations that any of the first M groups contain more than c trunks. We refer to a group with more than c trunks as an *overflow* group. For $j \geq 0$, let $\Omega_j(M, C, c)$ be the number of combinations for distributing the C trunks to the $M+1$ groups where j groups of the first M groups are the overflow groups. It is clear that

$$\Omega_0(M, C, c) = \Omega(M, C) - \Omega_1(M, C, c).$$



(a)



(b)

Fig. 5. Accuracy analysis of the analytic models ($C = 42,\ c = 12,\ M = 6, 1/\mu = 3\,\mathrm{min}$, and $\rho = 70m$ Erlang per ST). (a) Loss probability and (b) discrepancy of analytic models.

Consider the case where there are at least one overflow group. There are $\binom{M}{1}$ alternatives to select an overflow group from the first M groups. This group will have at least $c+1$ trunks. If the $M+1$ groups (including the overflow group) arbitrarily share the remaining $C-(c+1)$ trunks, then the number of combinations is $\binom{M}{1}\binom{C-(c+1)+M}{M}$. The above enumeration doubly counts the combinations for the case where there are at least two overflow groups. Thus, we have

$$\Omega_1(M,C,c) = \binom{M}{1}\binom{C-(c+1)+M}{M} - \Omega_2(M,C,c).$$

From an inductive proof, we can show that, for $0 \le j < M$ and $j(c+1) < C+M$,

$$\Omega_j(M,C,c) = \binom{M}{j}\binom{C-j(c+1)+M}{M} - \Omega_{j+1}(M,C,c).$$

For other j values, $\Omega_j(M,C,c) = 0$. Therefore,

$$\Omega_0(M,C,c) = \sum_{j=0}^{\lfloor (C+M)/(c+1)\rfloor} (-1)^j \binom{M}{j}\binom{C-j(c+1)+M}{M}.$$

In every combination, the exact model computes the Engset product form (see (7)) with M product operations. Thus,

$$\begin{aligned} &K(M,C,c)\\ &= M \times \Omega_0(M,C,c)\\ &= M\left[\sum_{j=0}^{\lfloor (C+M)/(c+1)\rfloor} (-1)^j \binom{M}{j}\binom{C-j(c+1)+M}{M}\right]. \end{aligned}$$

Fig. 8 plots $K(M,C,c)$ and the upper bound $M(c+1)^M$ in the log scale. Clearly, the exact model has nonpolynomial (NP) complexity. Although this complexity is very high, the execution of the exact model is still much faster than that of simulation. The execution time of the simulation is typically about 100 times longer than that of the exact model (based on measurement of our C++ programs).

The exact model requires M units of memory to represent the state of the WLL system and $c+1$ units of memory to record the Engset Forms (i.e., $\rho^j\binom{N}{j}$ for $0 \le j \le c$). Thus, its space complexity is $O(\max(M,c))$.

Since EPF [12] and the exact model have the same number of legal states and the same number of operations for calculating product forms, EPF has the same time and space complexities as that for the exact model. Table 3 lists the time and space complexities of the analytic models. The table indicates that the space complexities for these models are reasonably small. On the other hand, the time complexities for ES-C, ES-c, and H-ES are much lower than that for the exact model. By combining the exact model (that

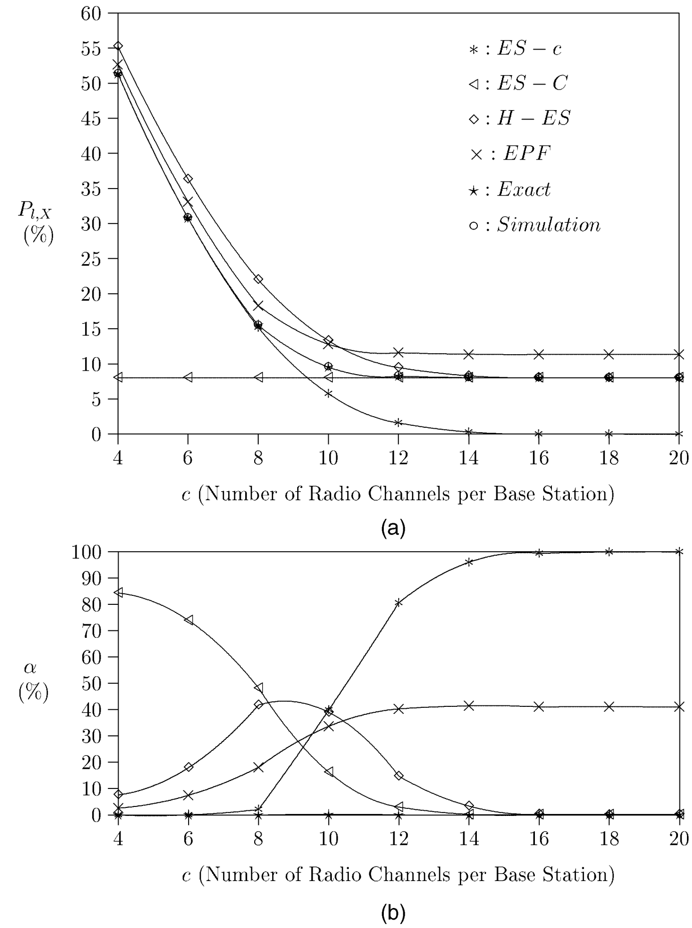


Fig. 6. Accuracy analysis of the analytic models ($C = 42,\ M = 6,\ N = 100,\ 1/\mu = 3\ \text{min}$, and $\rho = 70m$ Erlang per ST). (a) Loss probablity and (b) discrepancy of analytic models.

produces exact results) and the approximate models (that produce quick results), a WLL network planning strategy is described as follows: The WLL network planner first estimates the call traffic to be accommodated in the system and then uses performance tools to select the optimal values for system parameters such as C and c. The exact model can serve as the performance tool for parameter selection. This selection procedure repeatedly executes the exact model with various C and c values. Then, based on the computed P_l values, appropriate C and c are chosen. When the number of BSs and the call traffic to the system are large, the computation cost may be too high as indicated in Table 3. Since ES-C and ES-c provide lower bounds for P_l and H-ES provides an upper bound for P_l, an appropriate procedure for WLL network planning is to use these approximate models to quickly find the "bounds" (i.e., small ranges of C and c) that we are interested in. Then, the exact model is executed with these small numbers of c and C values to accurately find out the engineered operation points of the WLL system. EPF is not considered for bound analysis because its execution time is about the same as that for the exact model.

5 Design Guidelines for WLL Resource Planning

Based on the exact model, we derive some WLL design guidelines using the numerical examples illustrated in Figs. 4, 5, 6, and 7. We consider the effects of N (and, thus, the call traffic), c and C on the WLL performance. Figs. 4 and 5 indicate apparent results that the loss probability P_l increases as the number N of subscriber terminals (STs) per BS increases. On the other hand, some nontrivial phenomena are observed in these figures. In a WLL system with fixed C and c values, there is an operating point such that by adding extra STs per BS (i.e., increasing the value N), the loss probability P_l significantly increases. In Fig. 4, this point occurs around $N = 100$. That is, when $N < 100$, the slope of the P_l curve is less than 0.797 (per 10 STs). On the other hand, when $N > 100$, the slope of the P_l curve is more than 1.194 (per 10 STs). Similarly, in Fig. 5, this point occurs around $N = 80$. That is, when $N < 80$, the slope of the P_l curve is less than 1.047 (per 10 STs). On the other hand, when $N > 80$, the slope of the P_l curve is more than 2.661 (per 10 STs). When one selects the operation range for a WLL, he/she may not want to include the operation points beyond this "threshold" to avoid dramatic call loss performance change in the operation range.

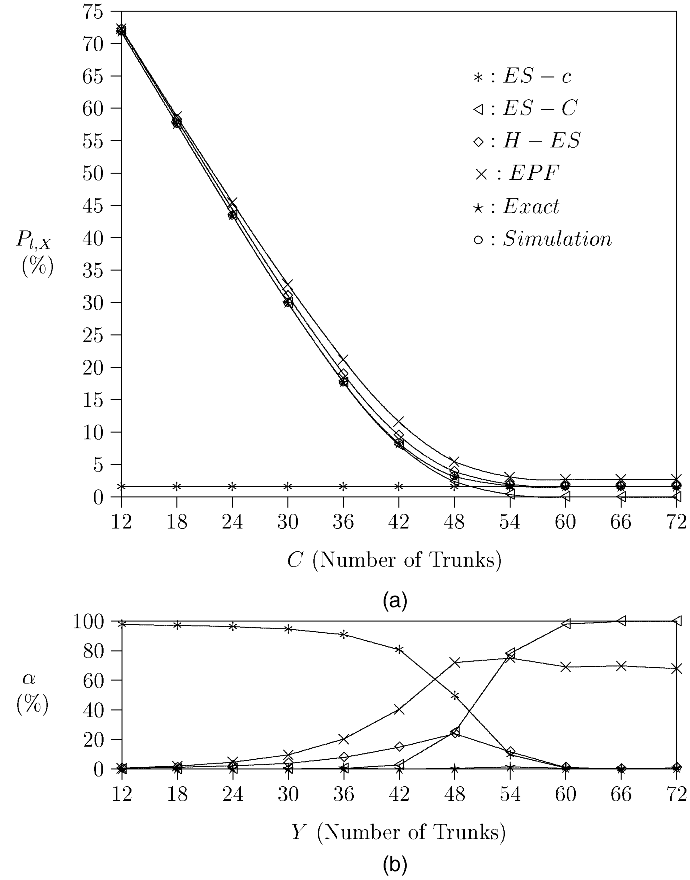


Fig. 7. Accuracy analysis of the analytic models ($c = 12, M = 6, N = 100, 1/\mu = 3$ min., and $\rho = 70m$ Erlang per ST). (a) Loss probability and (b) discrepancy of analytic models.

Fig. 6 shows how c affects the performance of WLL, where $C = 42$, $M = 6$, $N = 100$, $1/\mu = 3$ min., and $\rho = 70m$ Erlang per ST. The figure indicates that when c is small ($c < 8$ in our example), adding more radio channels in a BS significantly reduces P_l. When $c > 12$, adding extra radio channels does not improve the WLL performance at all. The reason is that, when c is small, the radio channels are bottleneck resources in the system and adding more radio channels reduces call loss in WLL. When c is large, most lost calls are due to the lack of trunks at BSC, not the radio channels in BSs. The contribution of our study is that it specifically identifies the point (for example, $c = 12$ in Fig. 6) beyond which the radio channels are no longer bottleneck resources.

Fig. 7 shows how C affects the performance of WLL, where $c = 12$, $M = 6$, $N = 100$, $1/\mu = 3$ min., and $\rho = 70m$

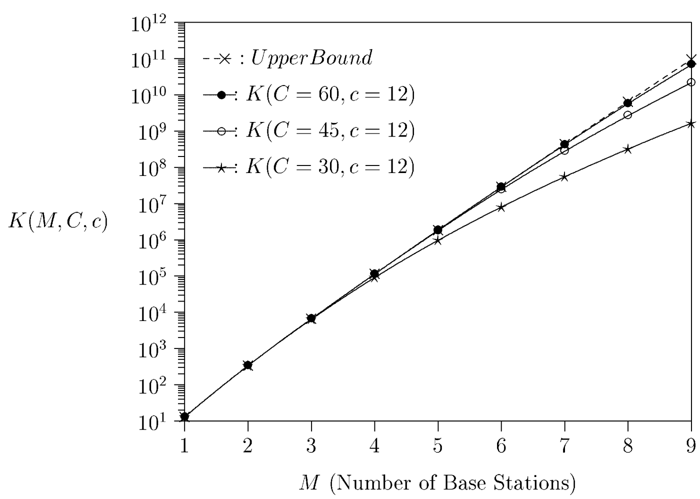


Fig. 8. The number of product operations executed in the exact model ($c = 12, N = 100$).

TABLE 3
Time and Space Complexities of the Analytic Models

Model	Time Complexity	Space Complexity
ES-c	$O(c)$	$O(1)$
ES-C	$O(C)$	$O(1)$
H-ES	$O(\max(C,c))$	$O(1)$
EPF	$O(K(M,C,c))$	$O(\max(M,c))$
Exact	$O(K(M,C,c))$	$O(\max(M,c))$

Erlang per ST. The figure indicates that, when C is small ($C < 48$ in our example), adding more BSC trunks significantly reduces P_l. When $C > 54$, adding extra BSC trunks only insignificantly improve the WLL performance. Similar to the phenomenon observed in Fig. 6, when C is small, the BSC trunks are bottleneck resources in the system and adding more trunks reduces call loss in WLL. Specifically, our model points out when BSC trunks are no longer bottleneck resources in a WLL system (i.e., when $C > 54$ in this example).

6 Conclusion

This paper studied the performance of wireless local loop (WLL) systems with a finite number of subscribers and a finite number of trunks in the Base Station Controller (BSC). We investigated several approximate analytic models and proposed an exact analytic model to compute the loss probability of WLL. In deriving the stationary distribution of the system states for the exact model, we also proved that the loss probability for the WLL is insensitive to the call holding time distributions and is only dependent on the mean of the call holding time. The exact model was validated against the simulation experiments. We observed that the time complexity of simulation is much higher than the exact analytic model. On the other hand, the executions of the approximate analytic models are much faster than that of the exact analytic model. We designed an efficient procedure (in terms of time complexity and accuracy) to identify the operation points of a WLL system. In this procedure, the approximate models are utilized to quickly compute upper and lower bounds for the engineered operation points of the WLL system parameters. Then, the exact model is used to accurately compute the performance results for the values of input parameters in the ranges identified by the approximate models. The network planner then selects the appropriate values for WLL system parameters based on the outputs of the exact analytic model.

According to the exact analytic model, we illustrated some WLL design guidelines by numerical examples. We showed, for an arbitrary call traffic, how to identify the bottleneck resources of WLL and how to appropriately increase the bottleneck resources to improve the WLL performance. Our guidelines are general enough to accommodate all kinds of call holding time distributions.

Acknowledgments

The authors would like to thank the three anonymous reviewers. Their valuable comments have significantly improved the quality of this paper. Yi-Bing Lin's work was sponsored in part by MOE Program of Excellence Research under contract 90-E-FA04-4, TAHOE Network, Ericsson, InterVideo, FarEastone, National Science Council under contract NSC 90-2213-E-009-156, and the Lee and MTI Center for Networking Research, NCTU.

References

[1] D. Burman, "Insensitivity in Queueing Systems," *J. Applied Probability*, vol. 13, pp. 846–859, 1981.
[2] D.Y. Burman, J.P. Lehoczky, and Y. Lim, "Insensitivity of Blocking Probabilities in a Circuit-Switching Network," *J. Applied Probability*, vol. 21, pp. 850–859, 1984.
[3] Z. Dziong and J.W. Roberts, "Congestion Probabilities in a Circuit-Switched Integrated Services Network," *Performance Eval.*, vol. 7, pp. 267–284, 1987.
[4] Ericsson, "DECT Access Node -DAN," *Ericsson Business Networks*, Apr. 1997.
[5] ETSI, "Radio Equipment and System (RES): Radio in the Local Loop," Technical Report ETR 139, ETSI, Nov. 1994.
[6] ETSI, "Radio Equipment and System (RES): Digital Enhanced Cordless Telecommunications (DECT): Services, Facilities and Configurations for DECT in the Local Loop," Technical Report ETR 308, ETSI, Aug. 1996.
[7] ETSI, "Radio Equipment and System (RES): Digital Enhanced Cordless Telecommunications (DECT): Traffic Capacity and Spectrum Requirements for Multi-System and Multi-Service DECT Applications Co-Existing in Common Frequency Band," Technical Report ETR 310, ETSI, Aug. 1996.
[8] Y. Fang and I. Chlamtac, "Teletraffic Analysis and Mobility Modeling for PCS Network," *IEEE Trans. Comm.*, vol. 47, no. 7, pp. 1062–1072, 1999.
[9] J.-Y. Huang, H.-M. Tsai, Y.-B. Lin, and C.C. Tseng, "Design and Implementation of an OA&M System for WLL Network," *IEEE/KICS J. Comm. and Networks*, vol. 2, no. 3, pp. 266–276, 2000.
[10] F.P. Kelly, *Reversibility and Stochastic Networks.* John Wiley & Sons Ltd., 1979.
[11] F.P. Kelly, "Blocking Probabilities in Large Circuit-Switched Networks," *Adv. Applied Probability*, vol. 18, pp. 473–505, 1986.
[12] F.P. Kelly, "Loss Networks," *The Annals of Applied Probability*, vol. 1, no. 3, pp. 319–378, 1991.
[13] Kleinrock, *Queueing Systems; Volume I: Theory.* Wiley, 1975.
[14] W.-R. Lai and Y.-B. Lin, "Resource Planning for Wireless PBX Systems," *Int'l J. Wireless Information Networks*, vol. 5, no. 4, pp. 351–357, 1998.
[15] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures.* John Wiley & Sons, 2000.
[16] A.R. Noerpel and Y.-B. Lin, "Wireless Local Loop: Architecture, Technologies and Services," *IEEE Personal Comm.*, vol. 5, no. 3, pp. 74–80, 1998.
[17] E. Pinsky, A. Conway, and W. Liu, "Blocking Formulae for the Engset Model," *IEEE Trans. Comm.*, vol. 42, no. 6, pp. 2213–2214, June 1994.
[18] M. Schwartz, *Telecommunication Networks. Protocol, Modeling and Analysis.* Addison-Wesley, 1987.
[19] Siemens, "DECTlink: Radio Connection of Fixed Subscriber Terminals and Mobile Handhelds," technical report, production catalog, 1998.
[20] R. Syski, *Introduction to Congestion Theory in Telephone Systems.* Oliver and Boyd, 1960.
[21] P. Trotter and A. May, "Wireless Local Loop: Market Strategies," technical report, Ovum, 1996.
[22] H.-M. Tsai et al. "The System Design for a DECT Wireless Local Loop: Final Report," technical report, Eumitcom Technology Inc., 1998.
[23] S. Zachary, "On Blocking in Loss Networks," *Adv. Applied Probability*, vol. 23, pp. 355–372, 1991.

Hsien-Ming Tsai received the double BS degrees in computer science and information engineering (CSIE) and communication engineering, and the MS degree in CSIE from National Chiao-Tung University (NCTU), Taiwan, in 1996 and 1997, respectively. He is currently a PhD candidate in CSIE at NCTU. His research interests are in the areas of personal communications, wireless local loop, embedded systems, and network device drivers. He is a student member of IEEE.

Yi-Bing Lin received the BSEE degree from National Cheng Kung University in 1983 and the PhD degree in computer science from the University of Washington in 1990. From 1990 to 1995, he was with the Applied Research Area at Bell Communications Research (Bellcore), Morristown, NJ. In 1995, he was appointed as a professor in the Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU). In 1996, he was appointed as deputy director of the Microelectronics and Information Systems Research Center, NCTU. During 1997-1999, he was elected as chairman of CSIE at NCTU. His current research interests include design and analysis of personal communications services network, mobile computing, distributed simulation, and performance modeling. Dr. Lin is an associate editor of the *IEEE Network*, an editor of *IEEE J-SAC: Wireless Series*, an editor of *IEEE Personal Communications Magazine*, an editor of *Computer Networks*, an area editor of *ACM Mobile Computing and Communication Review*, a columnist of *ACM Simulation Digest*, an editor of *International Journal of Communications Systems*, an editor of *ACM/Baltzer Wireless Networks*, an editor of *Computer Simulation Modeling and Analysis*, and an editor of the *Journal of Information Science and Engineering*. He has been program chair for the Eighth Workshop on Distributed and Parallel Simulation, general chair for the Ninth Workshop on Distributed and Parallel Simulation, program chair for the Second International Mobile Computing Conference, guest editor for the *ACM/Baltzer MONET* special issue on personal communications, a guest editor for *IEEE Transactions on Computers* special issue on mobile computing, and a guest editor for the *IEEE Communications Magazine* special issue on active, programmable, and mobile code networking. Dr. Lin is the author of the book *Wireless and Mobile Network Architecture* (coauthor with Imrich Chlamtac; published by John Wiley & Sons). He received the 1998 and 2000 Outstanding Research Awards from National Science Council, Republic of China, and the 1998 Outstanding Youth Electrical Engineer Award from CIEE, Republic of China. He is an adjunct research fellow of Academia Sinica. He is a senior member of the IEEE.