Journal of the Nigerian Mathematical Society
Vol. 35, pp. 191-217, 2017

HMM-BASED QUALITY OF SERVICE SURVIVABILITY IN MOBILE CELLULAR NETWORKS

M. E. EKPENYONG[1] AND D. E. ASUQUO

ABSTRACT. The degree of system survivability is governed by the deployed service protocol or internal information transfer mechanism, and despite the advances recorded in wireless network research, survivability still remains an open issue. This paper investigates the impact of failures on mobile cellular networks using the Hidden Markov Model (HMM) framework. Under ideal operation conditions, an experimental 3.75G test-bed was simulated. From the simulation, it was observed that reducing the number of general channels negatively impacts on new and handoff calls. To guarantee dependability and prevent the system from severe degradation, a saturation phase was imposed to ensure self-healing, such that handoff requests do not exceed the prioritization index. To model the system dynamics, two HMM-based systems were developed using empirical data obtained from an operating carrier. An evaluation of the training showed that system failure rates can be well tolerated through the efficient utilization of available guard channels, and the best Viterbi trace obtained from path with less node failures. Further analysis of the results demonstrated that the proposed framework improved the system performance, and regardless of an increase in the arrival rate, the probability of new call blocking stabilized below the recommended threshold after 25% of the channels was utilized.

Keywords and phrases: Fault recovery, HMM, network dependability, machine learning, self-healing mechanism
2010 Mathematical Subject Classification: 97R40

1. INTRODUCTION

The increase in cellular network capacity has not only complicated mobility management, but has also threatened dependability of the network. An undependable system most likely results in low quality of service (QoS) and users' dissatisfaction. Dependability is a measure of three key parameters namely reliability, availability and survivability. Reliability is vital for estimating the network

Received by the editors February 03, 2016; Revised: July 17, 2016; Accepted: March 08, 2017
www.nigerianmathematicalsociety.org
[1]Corresponding author

failure probability; availability is a real-time measure of network usage; and survivability describes the accessible performance of the network after failure. Survivable wireless network design involves three major tasks: fault tolerant topology determination; network nodes and link dimensioning; and traffic demands, subject to QoS and survivability requirements. Several research works have offered both analytical and simulation solutions to blocking and dropping probabilities in cellular networks [1-5], but only few of them report on fault tolerant designs. Sustaining communication services in the event of network failures requires robust survivable and self-healing algorithms. Self-healing is a process whereby a system automatically detects, diagnoses and repairs localized software and hardware errors [6].

Four components on which the mobility status of a multilayered cellular system depends on are identified in [7]. These components include distance of the mobile user from the base station (BS), signal strength (SNR), mobile speed, and the probability that no handoff occurred. Survivability strategies in wireless networks are mainly developed for fault prevention and recovery. Prevention techniques target component improvement and system reliability using fault tolerant architectures in the network switches and provides backup power supplies for the network components. On the other hand, to utilize the remaining capacity after failure, recovery techniques attempt to restore failure prone connections while maintaining network stability – through the use of dynamic fault recovery algorithms or appropriate load control policy in the radio resource management (RRM) functionality. Sharma and Hellstrand [8] viewed protection and restoration as two survivability techniques that require computation of an alternative path to which the traffic is switched whenever a failure occurs. The protection technique ensured that a backup path is pre-established before any failure, and spare capacity is simultaneously reserved as the request is setup. Various factors are involved when providing survivability. The notable ones are resource utilization, request blocking ratio, recovery time, and recovery granularity.

Next-generation wireless networks would enable the transportation of higher volumes of information and ensure exact levels of reliability. This scenario is becoming crucial as faults contribute to increased rate of data loss. The life cycle of network components according to the ITU-T E800/4260 is presented in Fig. 1.

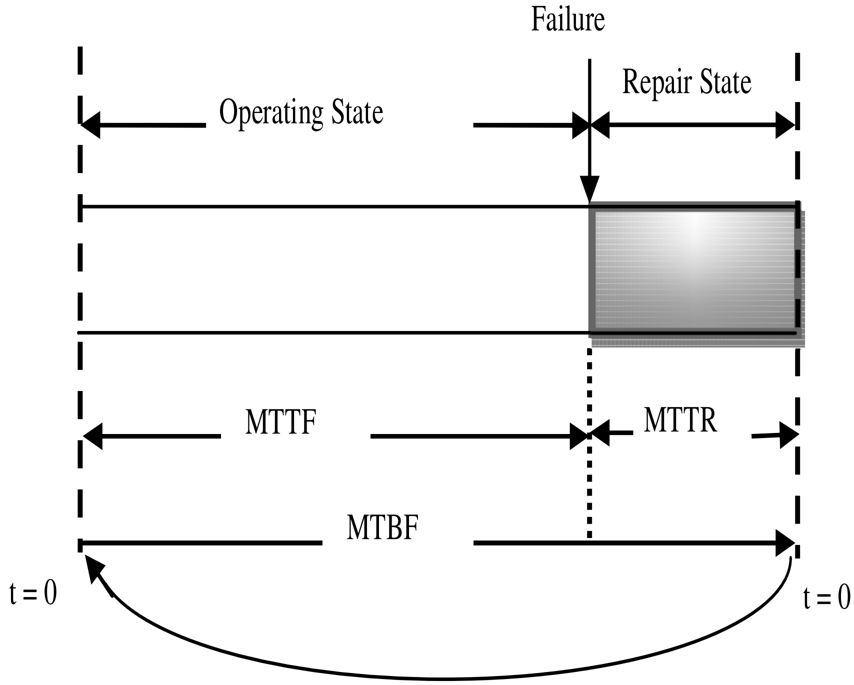


Fig. 1. Failure cycle of a repairable system

Each component generally begins at the operating state, t = 0; and when a failure occurs, the network element enters the repair state. Once the failure has been repaired, the network once more enters the operating state. The expected mean time before the first failure is the Mean Time to Failure (MTTF) and this corresponds with the operating state; the Mean Time to Repair (MTTR) is the average time spent performing all corrective maintenance repairs; and the Mean Time Between Failures (MTBF) is the MTTF including the time of repair following the last failure, i.e.; MTBF = MTTF + MTTR.. Providing survivability in cellular networks is therefore vital to minimize the recovery time while maintaining efficient resource utilization.

2. RELATED WORKS

Previous research works on the application of reliability theory to large complex systems focused on characterizing failure distributions to resolve why a system fails, the failure rate, and the possibility of mitigating such failures. In [9], the topology and capacity of physical links of specific transport networks are considered. In [10], the design of partial survivable backhaul networks for cellular systems where self-healing ring technology maintained the backbone transmission network with a diversity requirement for ensuring survivability of the network is investigated. Cox and Sanchez [11] studied the least-cost backhaul network design while meeting survivability and capacity constraints. An analytical model using Markov chains for the assessment of cellular networks is proposed in [12]. A stochastic reward net model is explored by the authors to automate the generation of Continuous Time Markov Chain (CTMC) and survivability metrics such as call blocking probabilities and latency due to failures. These works do not consider the effect of user mobility in the network design.

An analysis of survivability issues for voice services in 2G GSM mobile networks is made in [13]. The authors observed that the impact of a failure depends on a variety of factors such as the location and shape of failed area, user mobility, and user behaviour. However, the issue of coordination between layers and key concepts such as escalation, integrated and differentiated survivability were not addressed. In multilayered networks, escalation is a crucial aspect of survivability strategy. As defined in [14], it is a set of rules that describes when to start and stop, and how to coordinate the activities of the different recovery strategies. In [15], three key aspects of escalation are identified, which include activation type, escalation direction and inter-layer coordination. Chu and Lin [16] investigated the survivability of mobile wireless communication networks in the event of a BS failure and modelled the survivable network as a mathematical optimization problem that aimed at minimizing the total number of blocked traffic through the relocation of spare resources. Their results showed that the total call blocking rate is less sensitive to the call blocking probability threshold of each BS for light traffic load. A major benefit of layered structure in telecommunication networks is the simplification of hardware and network management at the topology level. Here, the physical and/or logical network topology is the most obvious way to defining layers for network survivability. This benefit was demonstrated in [17] for a three-tier internet hierarchy. The development of computational multilayered models with realistic survivability restrictions has become prominent [18-21]. Orlowski [19], for instance, proposed mathematical models for integrated optimization of two network layers with survivability constraints and described a multi-layer network design problem for various technologies, and modelled the network using mixed-integer programming (MIP) formulations. Real-world telecommunication networks consist of stack of technologically diverse networks called layers. These layers are strongly interdependent and embedded into one another. In [20], the relation of different layers of Wavelength Division Multiplexing (WDM) for multi-layer networks with service application for failure events is analyzed. In [21], a cost optimal multi-layer network design that permits technology selection at each node and incorporates traffic demand uncertainty is presented. Their model yielded full flexibility with regards to the number of layers and integrated layer-skipping and router offloading.

Previous works mainly employed analytical methods, and studied survivability metrics such as call blocking probability and call setup latency due to failure. However, these works never considered handoff probability due to frequent user mobility – a necessary metric for 3G and emerging wireless networks. This paper simulates a test-bed that validates an analytical model (a continuous time Markov chain technique) to study the performance of the system at different traffic load and channel allocation strategies. The simulated test-bed is then trained using a Hidden Markov Model (HMM) framework. The purpose of this framework is necessary to implement the system's dynamics and self-healing mechanism for efficient failure rates discovery, sufficient to guarantee dependability and prevent the network from severe degradation, as opposed to previous methods which only modelled survivability as a mathematical optimization problem without the incorporation of network dynamics. Furthermore, escalation directions for coordination between network layers for specific recovery strategies is implemented in this paper

3. SURVIVABILITY FRAMEWORK

In Fig. 2, a three-layered framework for QoS survivability evaluation is presented. These layers include the physical, transport and application layers. The physical layer consists of the transmission medium and signals responsible for realizing the network capacity for radio communication and RRM. The transport layer carries traffic and provides predefined sets of alternative routing and congestion control for managing users' mobility across the network. The application layer employs the available network services to improve end-to-end mobility management. The benefits of the proposed architecture include multi-protocol support, network survivability and efficient bandwidth allocation.

Layer/ Component	Failure Scenario	Disruption Type/ Failure Effect	QoS Metric
Application MSC, VLR, HLR, EIR, AUC, Data communication network	Loss of VLR	Degraded QoS, OS intrusion, Virus infection, Encryption fault, Loss of roaming service in MSC coverage area	Loss in user load, Packet loss rate, Throughput Database access latency
Transport BS, BSC, MSC	Loss of BSC-MSC link	Network mgt/traffic monitoring corruption, Service loss in clustered cells, Increased traffic in cells adjacent to failure	Call blocking probability, Call dropping probability, Call setup latency, Paging/location updates latency
Physical MS, BS, BSC	Loss of BS	Spectrum jamming/fading, EMI, Service loss in present cell, Increased traffic in cells adjacent to failure	Call blocking probability, Call dropping probability

Fig. 2. Survivability framework outlining failure scenario, disruption type and QoS evaluation metrics

Considering the framework in Fig. 2, a comparison of the recovery strategies is summarized in Table 1.

Table 1. Comparison of some recovery strategies (escalation directions)

Performance Criteria	Bottom Layer	Bottom-up	Top-down	Top Layer	Preferred Value
Switching granularity	Coarse	Coarse	Fine	Fine	Coarse
Failure coverage	Low	High	High	High	High
Required capacity resources	Low	High	High	Low	Low
Service differentiation	Difficult	Difficult	Average	Easy	Easy
Coordination, management	Low	High	High	Low	Low
Failure scenario	Simple	Simple	Complex	Complex	Simple
Recovery close to root	Yes	Yes	No	No	Yes
Strategy complexity	Low	Medium	Medium	Low	Low

The tradeoff between restorations at the different layers is that the physical layer has the fastest restoration. Higher layers may restore failures at lower layers but not vice-versa. One major constraint of a restoration technique is to ensure that the failure of any set of network component at any layer remains localized, thereby affecting those network sessions directly associated with such set of components.

4. SYSTEM MODEL

4.1 QoS PERFORMANCE

The system model consists of three clusters. Each cluster has four cells with each cell having same radius and number of channels, with BS at the centre. Suppose users are randomly distributed across the network and in a given cell i, the new call and handoff call arrival rates ($\lambda_{n,i}$ and $\lambda_{h,i}$) are Poisson distributed, while their mean service times ($\mu_{n,i}$ and $\mu_{h,i}$) and holding times ($1/\mu_{n,i}$ and $1/\mu_{h,i}$) are exponentially distributed. The assumption above permits the use of Erlang-B model for computing the call blocking probability [22-23]. Certain concepts are vital for explaining user's mobility and establishing relationship between call blocking and dropping probabilities in wireless network designs. The traffic performance of the network, for instance, largely depends on users' mobility, which is characterized by the cell residence time distribution [24-25]. Our tractable model is developed using the principles of queuing theory and Markov chain, and is useful for expressing the traffic characteristics, channel allocation scheme, user behaviour and mobility patterns of mobile communication systems. Next, we discuss the impact of different failure scenarios on system performance for disjoint and clustered cells, and model the effect of user mobility on channel reservation.

4.1.1 WITHOUT CHANNEL RESERVATION

In a complete resource sharing call admission control (CAC) scheme [26], no channel is reserved for call requests in any service class. A call is admitted only when the network has sufficient resources to accommodate it, otherwise it is rejected. The same CAC policy is applicable for new and handoff calls. In this policy, the cell residence time is exponentially distributed with mean $\hat{R} = 1/\gamma$, where γ, represents the degree of mobility of the user, and, γ approaches zero as mobility decreases. Let the state of a cell i, represent an instance of the number of channels occupied in that cell such that the cell states can be depicted as a CTMC (see Fig. 3), and the channel distribution in each cell corresponds to a multi server queue: $M/M/C_i$ with no buffer; where C_i represents the number of channels in cell i.

The state space of cell i is represented such that for all cells we have, $0 \leq n_i \leq C_i$, since there are C_i channels per cell. The transitions between states correspond to transitions of a CTMC. We designate $\pi(n_i, n_i')$, as the transition from state n_i to state n_i', to satisfy the

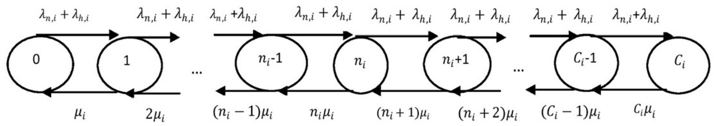


Fig. 3. Transition diagram for network state without channel reservation

normalization condition:

$$\sum_{n_1}^{n_{c_i}} \pi\left(n_i\right) = 1 \tag{1}$$

Then, the transition probabilities for adjacent states are obtained from,

$$\pi\left(n_i,\ n_i+1\right) = \ \lambda_{n,i} + \ \lambda_{h,i} \tag{2}$$

$$\pi\left(n,\ n_i-1\right) = \ n_i u_i \tag{3}$$

Equation (2) represents call arrivals at the next state, i.e., from state n_i to state (n_i+1). Call arrivals occur either when a new call arrives or during a handoff. Equation (3) represents departures from state n_i to state (n_i-1). Departures may occur either when a call exits a cell (due to handoff) or because a new call has been completed. In Fig. 3, we denote the steady state probability that the total number of ongoing calls in cell i is n_i, as $P\left(n_i\right)$, given that $n_i = 0,\ 1,\ 2,\ \ldots,\ \ C_i$. From the global balance equation, the steady state probabilities are obtained as [27]:

$$P\left(n_i\right) = \ P(0)\frac{{\rho_i}^{n_i}}{n!};0 \le n_i \le \ C_i \tag{4}$$

where

$\rho_i = \left(\frac{\lambda_i}{\mu_i}\right)$ is the traffic intensity of cell i; $\lambda_i = \lambda_{n,i} + \ \lambda_{h,i}$, $\mu_i = \mu_{n,i} + \mu_{h,i}$

$P(0)$ is the normalization factor, and is defined as $P(0) = \frac{1}{\sum_{n_i=0}^{C_i} \frac{{\rho_i}^{n_i}}{n_i!}}$

Thus,

$$P\left(n_i\right) = \ \frac{\frac{{\rho_i}^{n_i}}{n_i!}}{\sum_{n_i=0}^{C_i} \frac{{\rho_i}^{n_i}}{n_i}} \tag{5}$$

A new call requesting connection to cell i is blocked if all the C_i channels in the cell are occupied. Hence, the new call blocking probability $P\left(nb_i\right)$

in cell i, is obtained as:

$$P\left(nb_i\right) = \ P\left(C_i\right) = \ \frac{\frac{{\rho_i}^{C_i}}{C_i!}}{\sum_{c_i=0}^{C_i} \frac{{\rho_i}^{c_i}}{c_i}} \tag{6}$$

Given that no prioritization is assumed for handoff/emergency calls in the complete resource sharing scheme, the handoff dropping probability $P\left(hd_i\right)$ in cell i is the same as $P\left(nb_i\right)$. Hence,

$$P\left(hd_i\right) = \ P\left(nb_i\right) = P(C_i) \tag{7}$$

In extremely slow environments – where no priority is assigned to handoff call attempts, the new call and handoff blocking probabilities are identical [22], due to the Poisson Arrival See Time Average (PASTA) property.

4.1.2 WITH CHANNEL RESERVATION

With dynamic channel allocation, we explore the idea of guard channels to reserve some channels for handoff calls, since their failure are more sensitive to mobile subscribers than new call blocking. This allows a new call to utilize channels assigned for both new and handoff calls requests (but only handoff calls make use of the guard channels). We use the CTMC shown in Fig. 4, to evaluate these two metrics. Let $\frac{\gamma}{\mu_i+\gamma}$ denote the probability that a call can perform handoff because it has excellent signal quality. Then, the arrival rate of handoff call is thinned by $\left(\frac{\gamma}{\mu_i+\gamma}\right)\lambda_{h,i} \ = \ \ \beta\lambda_{h,i}$, where γ ranges in value from 1 to 5. It is obvious that β increases with γ.

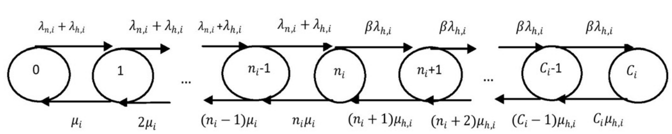


Fig. 4. Transition diagram for network state with guard channel reservation

The system is modeled with each cell having a total of C_i channels plus n_i general channels for new and handoff calls, and $(C_i - n_i)$ guard channels for handoff calls only. When the general channels are occupied and a new handoff call arrives, a channel is assigned from the guard channels and the call is accepted. However, when all channels including the guard channels are occupied, then, handoff calls are dropped and new calls are blocked. If $n_i > \frac{C_i}{2}$, then $(C_i - n_i) < n_i$. Similarly, if $n_i < \frac{C_i}{2}$, then $(C_i - n_i) > n_i$. If the general channel n_i is unoccupied at a time where only k_i channels are consumed, i.e., $k_i = n_i$; then, incoming handoff calls will share the general channels with incoming new calls. Thus, the steady state probability that k_i channels are occupied in the network is obtained as:

$$P_{k_i} = P(0)\frac{{\rho_i}^{k_i}}{k_i!} \tag{8}$$

with, $k_i = n_i$, $\rho_i = \left(\frac{\lambda_{n,i}+\lambda_{h,i}}{\mu_{n,i}+\mu_{h,i}}\right)$ and

$$P\,(0) = \frac{1}{1 + \sum_{k_i=1}^{n_i} \frac{{\rho_i}^{k_i}}{k_i!} + \sum_{k_{i=n_i+1}}^{C_i} \frac{{\rho_i}^{n_i}}{n_i!} \frac{1}{(k_i - n_{i)!}} \left(\frac{\beta\lambda_{h,i}}{\mu_{h,i}}\right)^{k_i - n_i}}$$

If the general channels n_i are occupied at a time where $k_i > n_i$, then, at that instant all $(C_i - R_i)$ general channels have been consumed and the cell will assign available R_i reserved channels to incoming handoff calls and block newly arrived calls. The steady state probability of meeting the system in this state $(k_i > n_i)$ is derived as:

$$P_{k_i} = P\,(0)\,\frac{1}{n_i!}\left(\frac{\lambda_{n,i}+\lambda_{h,i}}{\mu_{n,i}+\mu_{h,i}}\right)^{n_i}\frac{1}{(k_i - n_i)!}\left(\frac{\beta\lambda_{h,i}}{\mu_{h,i}}\right)^{k_i - n_i}; k_i > n_i \tag{9}$$

Thus, the blocking probability that a new call finds n_i general channels busy and is blocked is:

$$P\,(nb_g) = P\,(0)\,\frac{1}{n_i!}\,\left(\frac{\lambda_{n,i}+\lambda_{h,i}}{\mu_{n,i}+\mu_{h,i}}\right)^{n_i} \tag{10}$$

Similarly, the probability that an incoming handoff call finds all n_i general channels and R_i reserved (guard) channels busy and is dropped is:

$$P\,(hd_g) = P\,(0)\,\frac{1}{n_i!}\,\left(\frac{\lambda_{n,i}+\lambda_{h,i}}{\mu_{n,i}+\mu_{h,i}}\right)^{n_i}\frac{1}{R!}\frac{\beta\lambda_{h,i}}{\mu_{h,i}} \tag{11}$$

To measure the degree of prioritization achieved between new and handoff calls in the guard channels scheme, we introduce the concept of call incompletion probability which describes the probability that a call is blocked either at call initiation or new call requests or during handoff. Now, let P_{in}, be the call incompletion probability, P_{nb}, the new call blocking probability, and P_{hd}, the handoff dropping probability. We can express P_{in} mathematically as:

$$P_{in} = P\,(nb_g) + P\,(hd_g) \tag{12}$$

$$P_{in} = P\,(nb_g) + \frac{1}{R!}\frac{\beta\lambda_{h,i}}{\mu_{h,i}}P\,(nb_g) \tag{13}$$

From equations (12) and (13),

$$P(hd_g) = \frac{1}{R!}\frac{\beta\lambda_{h,i}}{\mu_{h,i}}P(nb_g) = \omega P(nb_g)$$

where, $\omega = \frac{1}{R!}\frac{\beta\lambda_{h,i}}{\mu_{h,i}}$.
Thus, the prioritization index is deduced as:

$$\frac{P(nb_g)}{P(hd_g)} = \frac{1}{\omega} \tag{14}$$

If $\frac{1}{\omega} > 1$, then the network can successfully achieve handoff prioritization, because at that instant, $P(nb_g) > P(hd_g)$.

5. SIMULATION AND EVALUATION OF EXPERIMENTAL TEST BED

A test bed (a platform for conducting rigorous, transparent, and replicable testing of scientific theories, computational tools, and new technologies) is designed in this section to provide a clear visualization of the proposed model and its feasibility within the study environment. An evaluation of the model is then carried out using two methods: simulation and machine learning.

5.1 DESCRIPTION OF THE TEST BED

Our test-bed spatially models a 3.75G network carrier. First, a survey of BSs belonging to one of the telecommunication operators in Nigeria was conducted for a period of three months. The BSs were digitized as coloured circles on a road map of the study area using the ArcGIS software. The green circles indicate low traffic; the yellow circles represent moderately high traffic; while the red circles represent very high traffic. With regards the test-bed data, it is possible to simulate Medium Access Control (MAC) and handoff process in the system. The base station controller (BSC) and the intelligence of the cell phone keep track of and allow the phone to switch from one BS to another during conversation. Fig. 5 reveals a typical case of poor service quality observed along major roads due to high traffic (within the study area). One way for minimizing this effect is to apply effective power control mechanisms and model the base stations such that users do not experience the *near-far* effect.

A random waypoint mobility model developed in [28]: where the users are randomly located in the network, with distance from the BS in a cell and movement direction (also randomly distributed) is simulated in this section to illustrate users' mobility. The users' speeds are considered as being uniformly distributed with a deterministic movement in same direction for large number of users. This case is consistent with highway travel and to portray the deterministic case, the road map is clearly demarcated with thick black lines within the test-bed in Fig. 5. Fig. 6

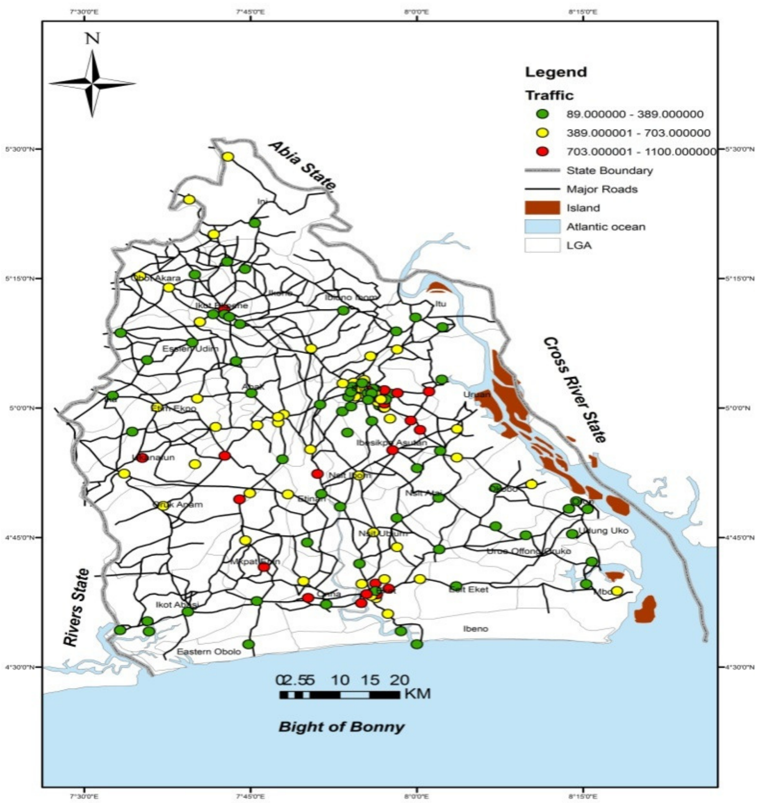


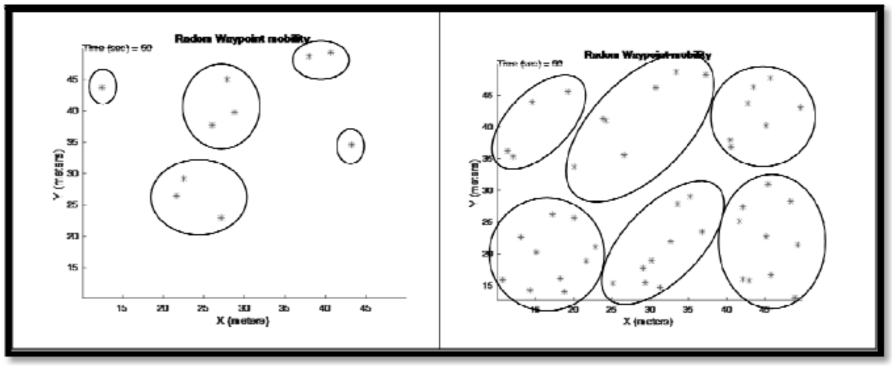
Fig. 5. Roadmap of the test-bed showing carried traffic distribution amongst various BSs in the study area.

describes a simulation of mobility patterns for 10, 50, 100 and 500 mobile users distributed within a $50 \times 50m^2$ radius. The network topology allows for movement of users across the entire network. Fig. 6(a)-(d), illustrate the movement of users within the coverage area. For lesser traffic (10 users), the traffic pattern is more defined and mobile users can communicate faster. When the traffic gets bustier (50 and 100 users), the mobile users begin to form more distinct clusters as circled in Figs. 6(b) and (c), communication becomes competitive, as more users compete for available channels. For 500 users, the traffic intensity becomes so high that mobile users begin to collide and the available resource is insufficient to service the network.

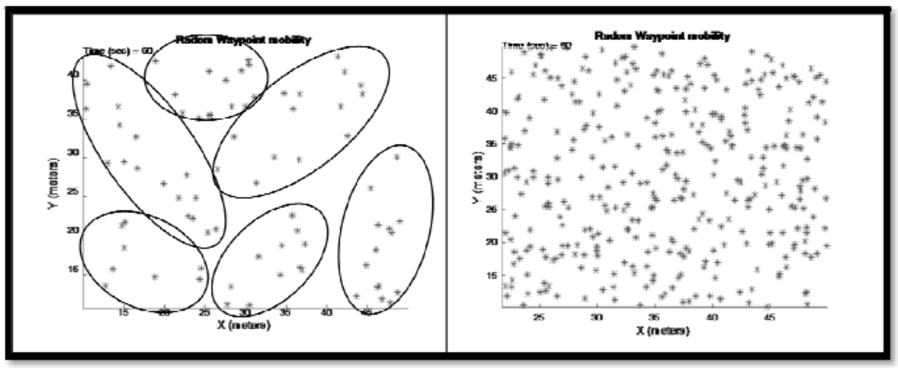
5.2 MODEL EVALUATION

5.2.1 EVALUATION USING SIMULATION

To provide knowledge of possible outcomes and their likelihood of occurrence, a discrete simulation of the survivability model was attempted to provide a microscopic view of the system. This simulation approach



(a) 10 mobile users (b) 50 mobile users



(c) 100 mobile users (d) 500 mobile users

Fig. 6. Simulation of users' mobility using the random waypoint mobility model

is valuable for approximating real life scenarios [29] and offers the investigator greater knowledge of the real system. However, it does not generate evolutionary traces for knowledge discovery and the accumulation of sufficient training data, which distribution produces the most probable estimates as well as a frame of expectations regarding variables classification. In this section, we provide a Monte-Carlo type simulation of the survivability model. Table 2 shows the various parameters and values used in the simulation.

Table 2. Simulation parameters and values

S/N	**Parameter**	**Value**
1.	Total number of channels (C_i)	20
2.	Reserved/guard channels ($C_i - n_i$)	5, 10, 16
3.	Arrival rate of new calls (λ_n)	random number
4.	Arrival rate of handover calls (λ_h)	random number
5.	Degree of user mobility (γ)	1-5
6.	Service rate of new calls (μ_n)	random number
7.	Service rate of handoff calls (μ_h)	random number

We observed the new call blocking and handoff dropping probabilities under different traffic loads, channel resources and user mobility patterns. Fig. 7 is used to evaluate the QoS metrics without call prioritization. Here, the probability of blocking a new call is the same as the probability of dropping a handoff request. The graph indicates how failure scenarios for 75%, 50% and 20% of lost channels can cause degradation in QoS delivery and impact differently on the system performance. At low traffic intensities, the blocking probabilities record very low values. But as the traffic intensity increases, the probability of new call blocking (or handoff call dropping) increases for cells with few channels. In the event of a failure resulting in the loss of 20% of the total available channels, the reliability of the network remains stable for new call blocking, $P(nb)$, and handoff dropping, $P(hd)$.

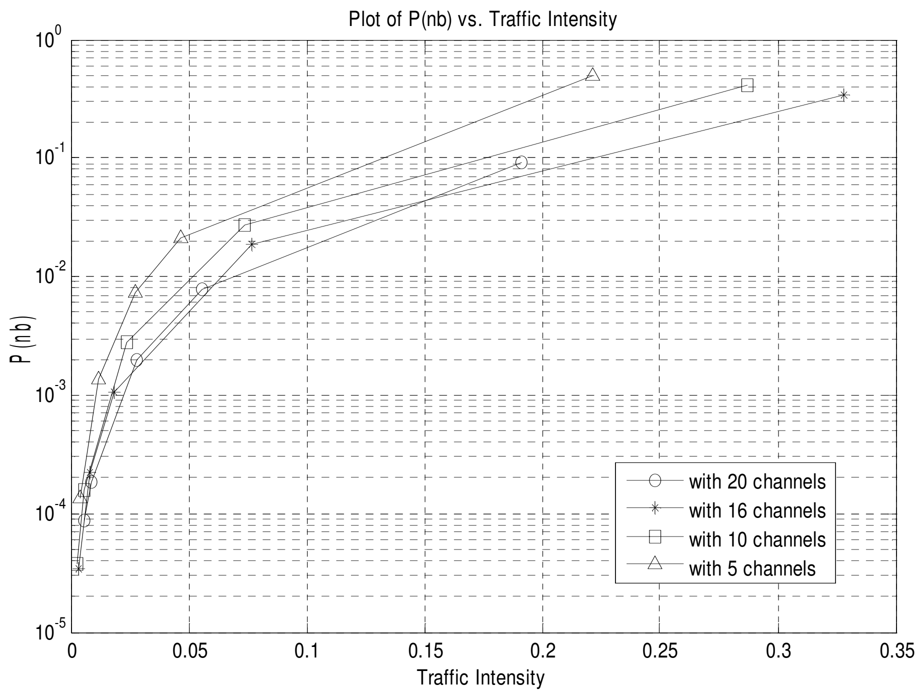


Fig. 7. A graph of probability of new call blocking vs. traffic intensity (without channel reservation)

Figs. 8 and 9 are used to investigate the impact of a number of guard channels $(c_i - n_i)$ reserved for priority handoffs. In Fig. 8, we observed that reserving more (guard) channels for handoff calls increases the rate of blocked calls and results in a high $P(nb)$, because the general channels n_i become fastly consumed by both new and handoff calls at higher traffic intensities. The result further reveals that not more than 50% of the total available channels should be assigned as guard channels in order to improve network component reliability.

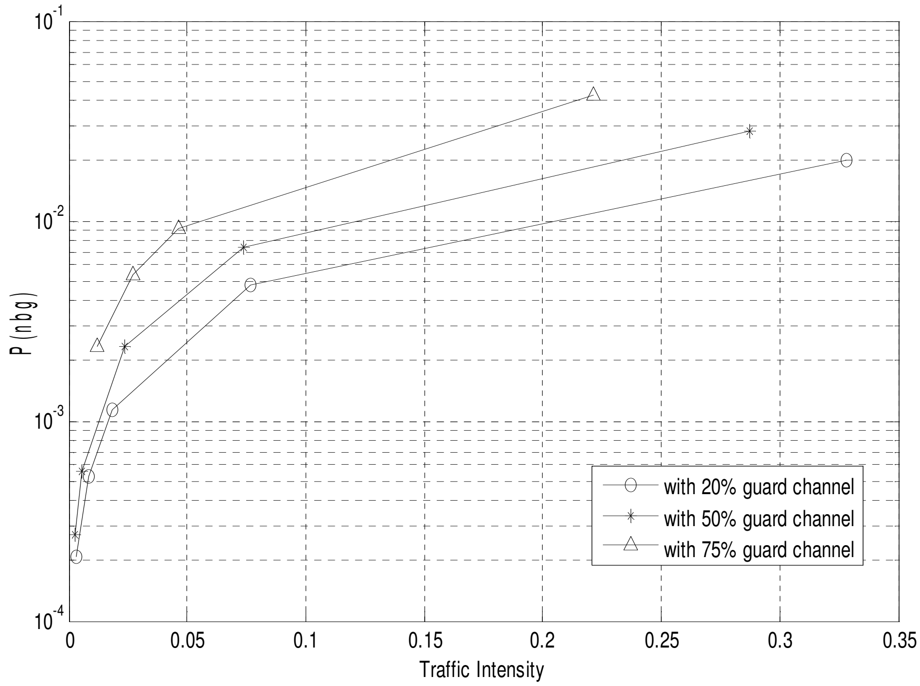


Fig. 8. A graph of probability of new call blocking vs. traffic intensity (with channel reservation)

In Fig. 9, the effect of handoff call drop probability on the system is minimized – for cells with guard channels, compared to Fig. 7 – where the cells are simulated without prioritization for a given range of traffic intensity. The result in Fig. 9 yields handoff probability values below 0.02 threshold even at high traffic intensity, which suggests that the required QoS can be maintained to sustain subscribers' loyalty and proves that the reservation schemes are useful to implement prioritization techniques for call admission in 3G wireless networks. Although it is important to allocate sufficient general channels to service increasing traffic during handoffs, it is also justified to assign enough guard channels (spare capacity) to BSs with high user mobility (in the event of adjacent BS failures), or cater for the sudden influx of priority calls at busy roads. Hence, telecommunication operators require a dynamic channel allocation strategies that learns from existing data, to ensure that guard channels do not become unnecessarily idle at very low traffic.

Fig. 10 shows that at lower degrees of mobility, all handoff call requests are granted with 75% guard channels. As the traffic gets bustier, users moving faster in an attempt to relinquish calls to other nearby cells suddenly have their calls dropped. This is so because as γ increases, ongoing calls requesting handoff may be forcefully terminated and introducing more guard channels will not benefit the system. At $\gamma = 5$, the system converges for the 50% and 75% guard channels. The

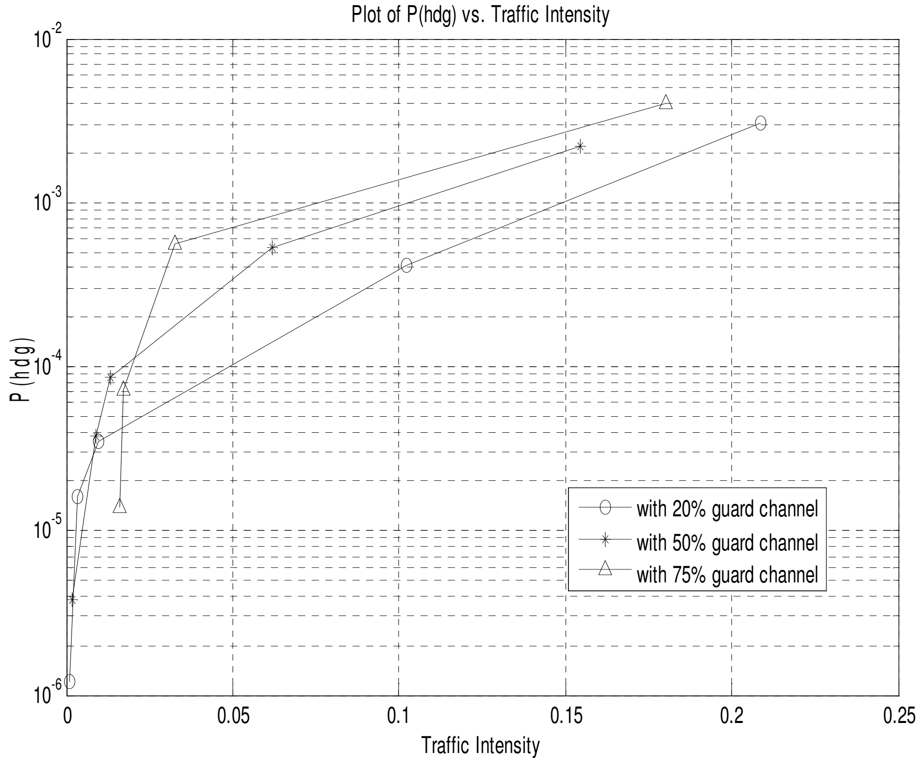


Fig. 9. Graph of probability of handoff call dropping vs. traffic intensity (with channel reservation)

convergence shows that further system degradation can be averted, since the general channels have been fully occupied by both new and handoff calls. This validates the claim that a system can remain survivable even when a failure is noticed. Put differently, the system has reached a *saturation phase*, where more calls are likely dropped due to exhaustion of the general channels. In reality, such scenario could threaten the system's stability if the mobility pattern of users is unpredictable. Certainly, users' behavior is bound to worsen the network performance should more users attempt severally to reconnect prematurely terminated calls. A possible solution to this problem is to understand the factors responsible for the failure and ensure that proper mechanisms are incorporated to mitigate the impacts of failure on the system. In practical systems, this can be achieved by invoking a survivability profile that temporally swaps unused guard channels to service new and handoff call requests, when more handoff calls are noticeably dropped or when such convergence appears.

In Fig. 11, more handoff call requests are successful when the number of reserved channels becomes high. The ratio of $P(nb_g)$ to $P(hd_g)$ gives the prioritization index which ensures that $P(hd_g)$ is reduced without

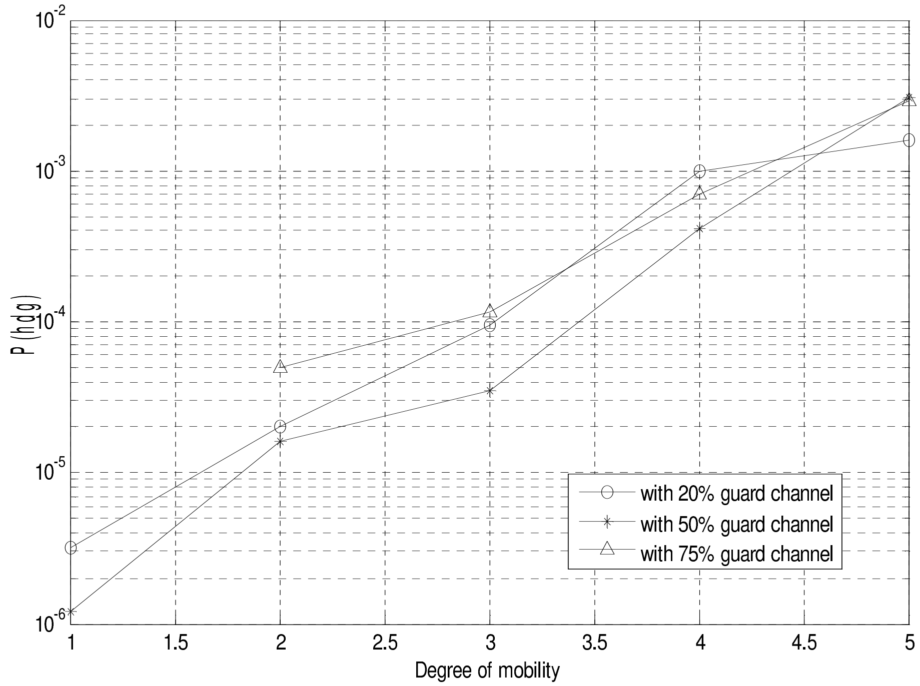


Fig. 10. Graph of probability of handoff call dropping with guard channel vs. degree of mobility

a significantly increasing $P(nb_g)$. For successful handoffs: $P(nb_g) > P(hd_g)$, and implies that $\frac{1}{w} > 1$. The end point of each plot signifies that handoff requests are not successful beyond this point to continue providing services for both classes of calls. This self-healing mechanism represents an efficient strategy for traffic management and restoration, and aims at maintaining stable and survivable systems.

5.2.2 EVALUATION USING MACHINE LEARNING

A stochastic approach is employed in this section to model the dynamic changing nature of the system. To model the CTMC property, a HMM framework (where future occurrences or states depend solely on the present state and not on the sequence of events preceding them) is implemented. The framework allows for reasoning and computation that would otherwise be intractable. In Fig. 6, we observed that the mobility pattern of mobile users can be effectively managed at low traffic intensities (Fig. 6(a)-(c)). But as the network grows, the available system resources become insufficient and must be improved to ensure proper mobility management; else, the ugly scenario in Fig. 6(d) appears. We tackle this problem by observing the sequence of emissions, without recourse to the sequence of states the model went through to generate the emissions; and then recover the states' sequence from observed data. The following steps were used to analyze the HMMs:

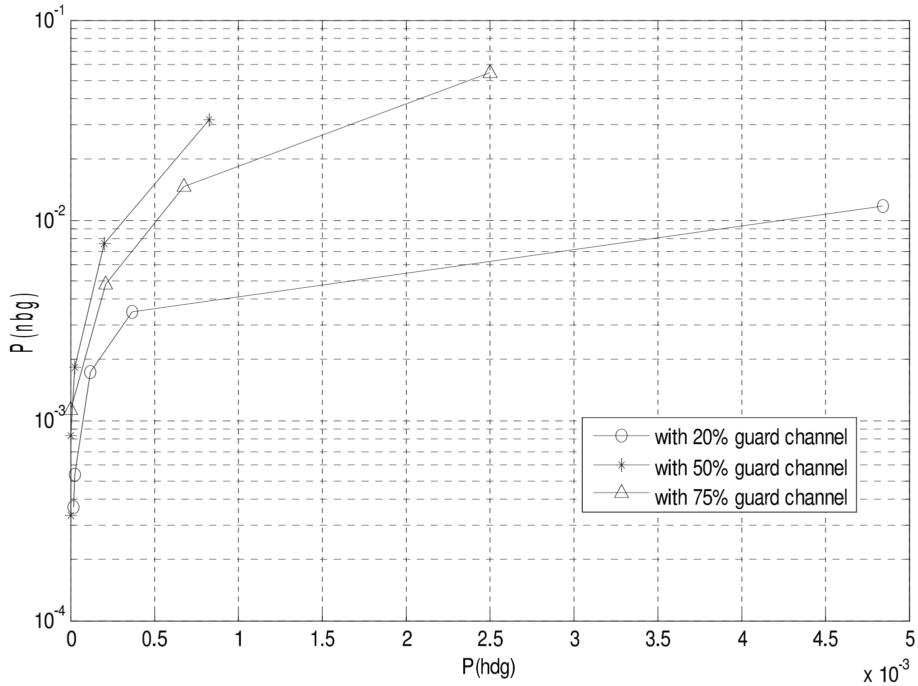


Fig. 11. Graph of probability of new call blocking vs. probability of handoff call droping (with channel reservation)

(1) *Test sequence generation*: the transition matrix, TRANS, consists of the new call blocking and handoff drop probability thresholds; while the emission matrix, EMIS, contains two dataset sequences specifying the new call blocking and handoff call drop probability for each HMM state (in this case, 20 states – indicating the number of available channels consumed at each state). The MATLAB Machine Learning (ML) toolbox *hmmgenerate* command was then used to generate 1,000 random sequence of states and emissions from the model; *[seq,states]=hmmgenerate (1000,TRANS,EMIS)*

(2) *Test sequence estimation:* Given the transition and emission matrices TRANS and EMIS, the most likely states' sequence the model went through was obtained using the MATLAB ML toolbox *hmmviterbi* command; *hmmviterbi (seq, TRANS, EMIS)*

(3) *Transmission and emission matrices estimation:* these matrices were estimated by the MATLAB ML toolbox hmmestimate and hmmtrain functions, given a sequence, *seq*, of emissions; *[estTR,estE] = hmmtrain (seq,TRANS,EMIS)*

The various input variables and respective values, used for training the HMM systems are presented in Table 3,

Table 3. Input variables and values to HMM

S/N	Variable	Value
1.	Arrival rate (λ)	2; 4
2.	Traffic	10-100; 100-1000
3.	Number of channel (C)	1-20
4.	Service rate (μ)	1-5
5.	Degree of user mobility (γ)	1-5
6.	Maximum number of HMM states	1000
7.	Recommended call blocking probability threshold	0.02
8.	Recommended handoff call success threshold	90%

5.2.2.1 MODEL TRAINING

The search for the most probable state sequence constitutes a natural issue in HMM analysis. The Viterbi algorithm was used in this paper to resolve the task of dynamically searching for the shortest path or sequence with dominant observations. Two different survivability systems (with and without channel reservation) were trained for this purpose. Table 4 and Table 5 present the results of HMM systems without channel reservation and with channel reservation, respectively. The Tables accumulate the state sequences generated along the optimal path, by the Viterbi algorithm. Generally, the best Viterbi trace was obtained from paths with less node failures. The percentage of actual sequence states (%ASS) indicates the most likely sequence of states that agreed with the random sequence used.

Table 4. State path accumulation for system without channel reservation

System 1										
Parameter	S1	S2	S3	S4	S5	S6	S7	S8	S9	Total
$\lambda(2)$; traffic (100-1000)	517	282	125	51	17	8	0	0	0	1000
$\lambda(4)$; traffic (100-1000)	471	294	149	56	22	6	1	0	1	1000
$\lambda(2)$; traffic (10-100)	901	88	11	0	0	0	0	0	0	1000

Parameter	Success	Failure	%ASS
$\lambda(2)$; traffic (100-1000)	478	522	42.5
$\lambda(4)$; traffic (100-1000)	485	515	49.4
$\lambda(2)$; traffic (10-100)	550	440	54.0

Table 5. State path accumulation for system with channel reservation

System 2										
	S1	S2	S3	S4	S5	S6	S7	S8	S9	Total
$\lambda(2)$; traffic (100-1000)	433	351	139	58	13	3	2	0	1	1000
$\lambda(4)$; traffic (100-1000)	465	307	145	45	33	4	0	1	0	1000
$\lambda(2)$; traffic (10-100)	824	171	5	0	0	0	0	0	0	1000

Parameter	success	Failure	%ASS
$\lambda(2)$; traffic (100-1000)	600	400	65.2
$\lambda(4)$; traffic (100-1000)	502	498	59.2
$\lambda(2)$; traffic (10-100)	680	320	72.7

The above state paths indicate that failure rates were properly managed through efficient utilization of the available reserved channels. Both systems converged with tolerance, $1e^{-06}$, before 500 iterations when the satisfactory optimum was reached. The idea behind the threshold is to ensure that the proposed models continue to perform optimally, even in the midst of failures.

6. RESULTS

A distribution of traffic intensity across the various HMM states, is given in Fig. 12. At low traffic (10-100 users), the traffic intensity dropped to an average of 0.2 Erlang, compared to high traffic (100-1000 users), which produced an average traffic intensity of 1.72, despite an increase in the arrival rate.

In Fig. 13, the generated emissions (new call blocking probability), as the number of states (or channels) increases for system without channel reservation (or guard channels) is presented. It was observed that the system performance improved as more channels became available. But regardless of the increase in the arrival rate, the probability of new call blocking was stabilized below the recommended threshold (i.e., 0.02), after the fourth HMM state. From this state upward, the system is bound to perform optimally without degradation.

Similar trends were observed for generated emissions (new call blocking and handoff drop call probabilities) in system with channel reservation (Figs. 14 and 15).

The estimated emissions (new call blocking probability) generated after model training, for system without channel reservation and system with channel reservation, are given in Fig. 16 and Fig. 17, respectively.

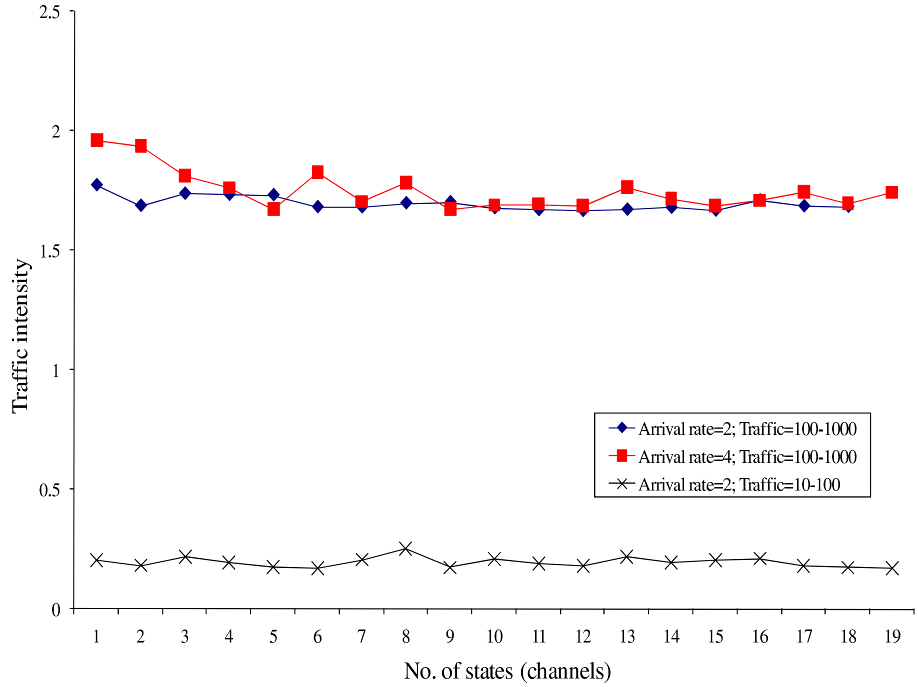


Fig. 12. Traffic intensity distribution across the HMM states

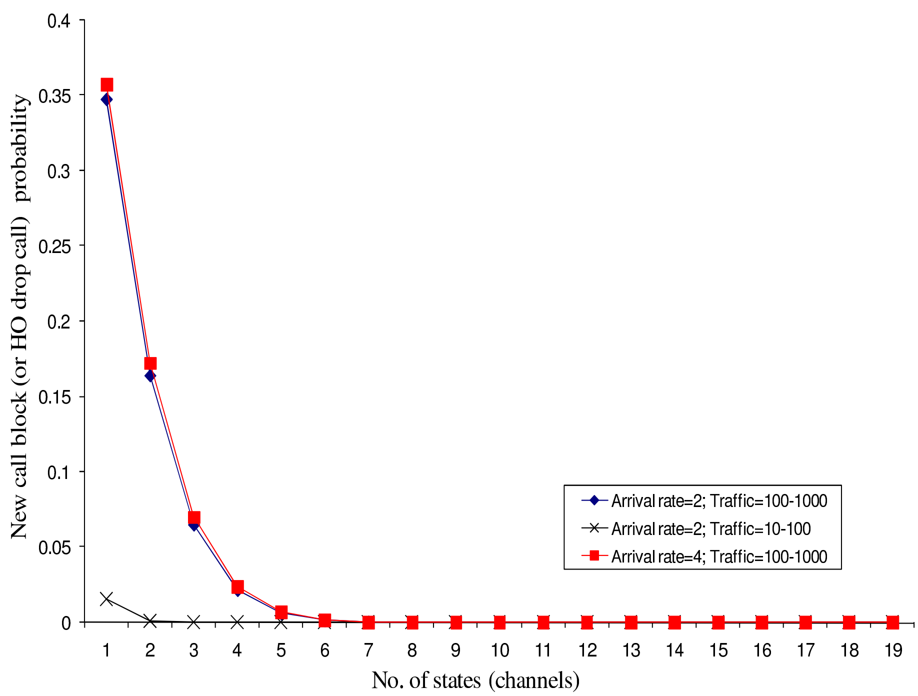


Fig. 13. New call block (or HO drop call) probability emissions across the HMM states (without reserved channel)

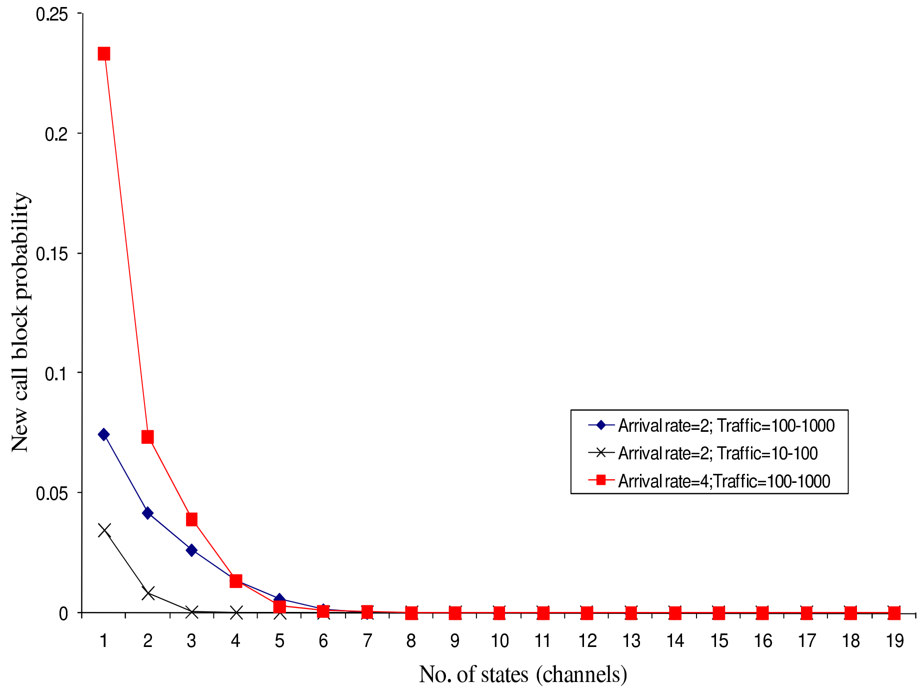


Fig. 14. New call block probability emissions across the HMM states (with reserved channel)

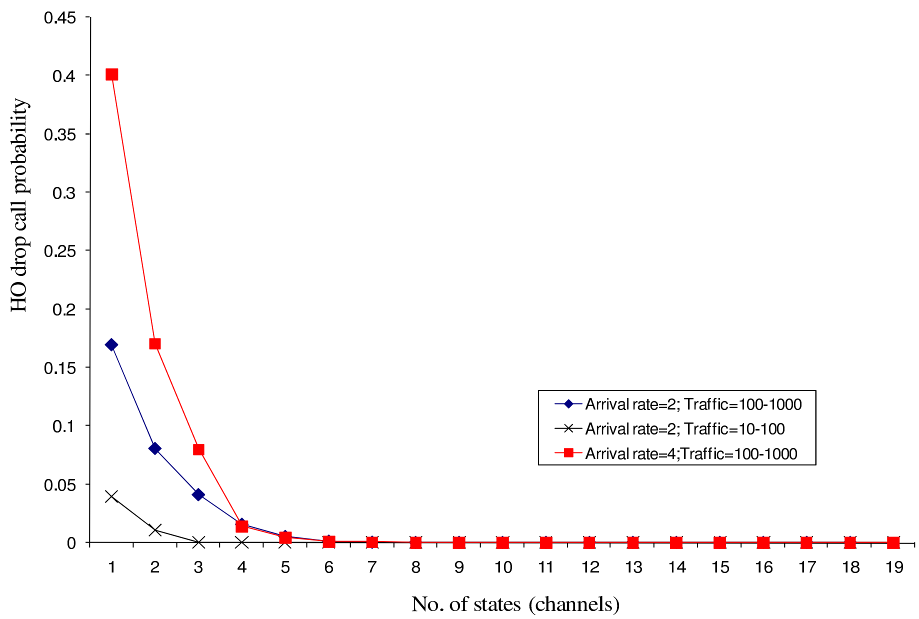


Fig. 15. Handoff drop call probability emissions across the HMM states (with reserved channel)

We observed that the generated emissions in system with channel reservation had lower blocking probabilities, compared to emissions generated from system without channel reservation.

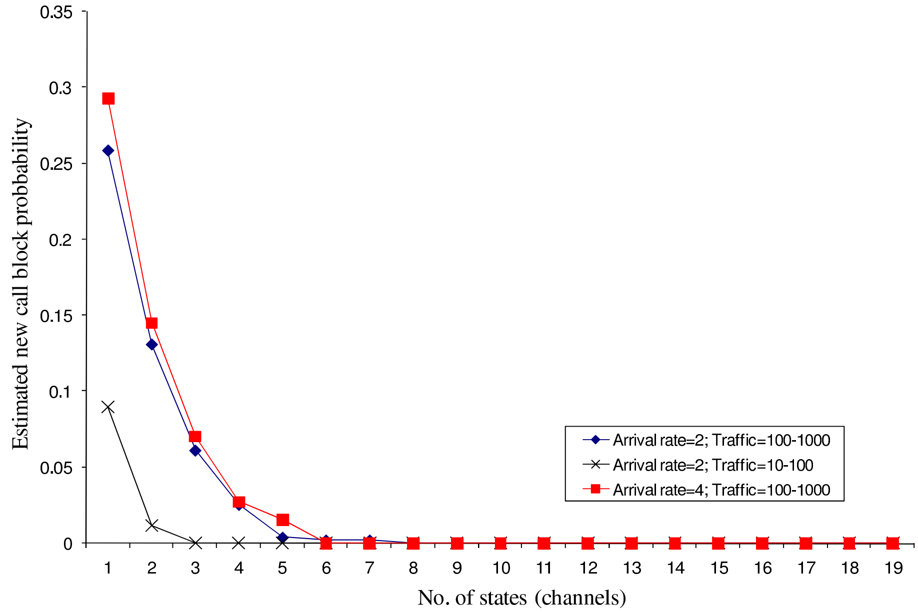


Fig. 16. Estimated new call block probability emissions across the HMM states (without reserved channel)

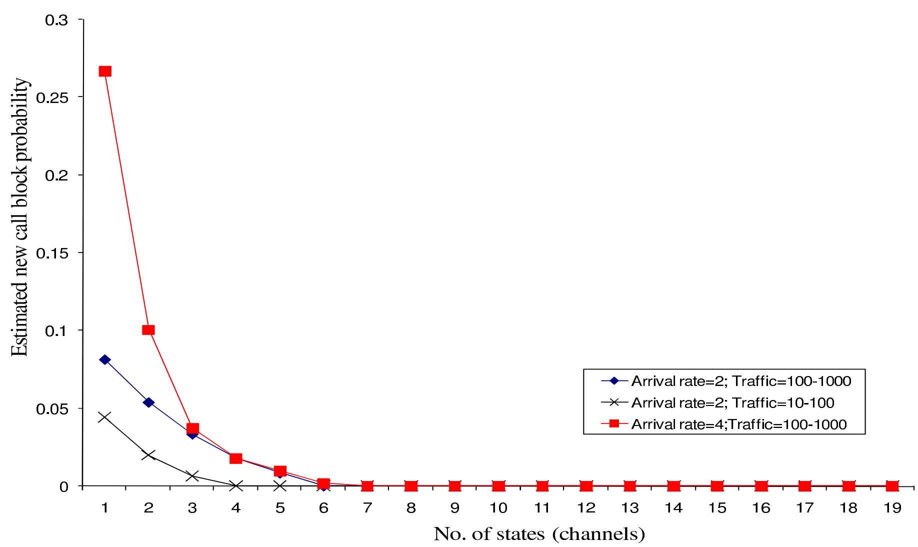


Fig. 17. Estimated new call block probability emissions across the HMM states (with reserved channel)

Further investigation revealed that there exist some correlation between the estimated emissions for new call blocking probability without channel reservation (Fig. 16.) and estimated emissions for HO drop call probability without channel reservation (Fig. 18). This result validates the relationship established in equation (7), that: *given no prioritization for handoff/emergency calls, the handoff drop call probability $P(hd_i)$ in cell i, is the same as the new call blocking probability $P(nb_i)$ in same cell*. Also, the estimated HO probability emissions outputs of systems with and without channel reservation (Fig. 18. and Fig. 19.) are almost the same, except for result with arrival rate of 4.

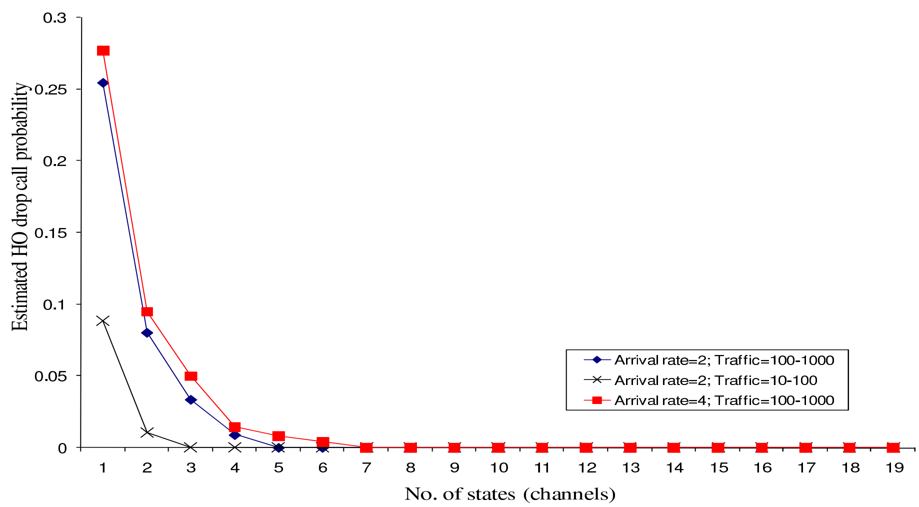


Fig. 18. Estimate handoff drop call probability emissions across the HMM states (without reserved channel)

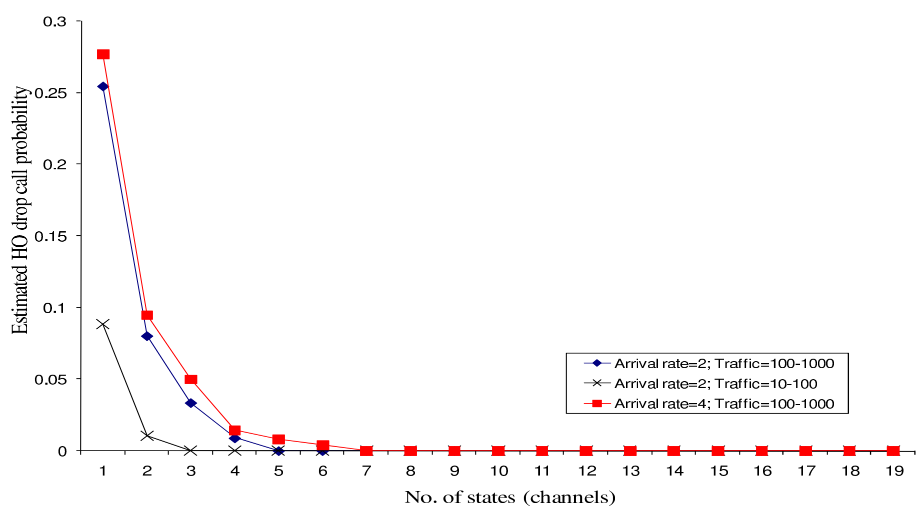


Fig. 19. Estimate handoff drop call probability emissions across the HMM states (with reserved channel)

7. CONCLUDING REMARKS

The reliability of wireless communication systems is largely determined by key dependability factors such as availability and survivability, and the major issues associated with these factors include traffic load, channel allocation, bandwidth limitation, signal propagation, quality of service, and on-demand or real time services. Hence, survivability traffic management and restoration procedures — which seek to redirect network loads such that failures impose minimal impact on their occurrence while the affected load is restore, is important. The performance of traffic restoration however depends on a combination of algorithms used for the restoration, as well as space capacity allocation in the network. In the event of failure, protection mechanisms outpace restoration when recovery the traffic, and do not have to wait to establish and restore alternative paths — as they guarantee complete availability.

This paper proposed a survivability framework, which permits the investigation of network characteristics of emerging wireless networks, and the impact of failures on the performance of the system. The metrics adopted to test the system's performance are call blocking and call dropping. Results obtained showed that the proposed framework improved the system performance and enhanced dependability of the network, to meet subscribers' demand and the required service quality.

REFERENCES

[1] J. Huang, C. Y. Huang, C. M. Chou, Soft-blocking Based Resource Dimensioning for CDMA Systems. IEEE Veh. Tech. Conf., 6, 2004, 4306–4309.

[2] G. Boggia, P. Camarda, A. D'Alconzo, Modeling of Call Dropping in Well-Established Cellular Networks, EURASIP Journal on Wireless Communications and Networking, 7 (2007) 1–11.

[3] M. Ekpenyong, J. Isabona, A. Akpaeti, Robust Modeling and Throughput Optimization in CDMA Networks. Journal of Applied Science and Engineering Technology. 3 (2009) 5–12.

[4] M. Ekpenyong, J. Isabona, Modeling the Effect of Dropped Calls on Cell Traffic in Established 3G-Based Cellular Networks, African Journal of Computing & ICT, 7(2)(2014) 157–164.

[5] E. E. Williams, D. E. Asuquo, An Efficient Model for Reducing Soft Blocking Probability in Wireless Cellular Networks, International Journal of Wireless and Mobile Networks, 6(4) (2014) 85–99.

[6] A. Avizienis, J. C. Laprie, C. E. Landwehr, Basic Concepts and Taxonomy of Dependable and Secure Computing, IEEE Transactions on Dependable and Secure Computing, 1(1) (2004): 11–33.

[7] H. Li, Modeling Multi-layer CDMA Cellular Networks Augmented with GPS using OPNET. M.Sc. Thesis. University of Manitoba, Manitoba, 2000.

[8] V. Sharma, F. Hellstrand, Framework for Multi-Protocol Label Switching (MPLS)-based Recovery. The Internet Society, IETF RFC 3469 (2003) 1–40.

[9] D. Alevras, M. Grotschel, P. Jonas, U. Paul, R. Wessaly, Survivable Mobile Phone Network Architectures: Models and Solution Methods, IEEE Comm. Magazine, 36(3) (1998) 88–93.

[10] A. Dutta, P. Kubat, Design of Partially Survivable Networks for Cellular Telecommunication Systems. European Journal of Operational Research, 118 (1999) 52–64.
[11] L. A. Cox, J. R. Sanchez, Designing Least-cost Survivable Wireless Backhaul Networks. Journal of Heuristics, (6) (2000) 525–540.
[12] V. Jindal, S. Dharmaraja, K. S. Trivedi, Markov Modeling Approach for Survivability Analysis of Cellular Networks. International Journal of Performability Engineering. 7(5)(2011) 429–440.
[13] D. W. Tipper, C. Charnsripinyo, H. Shin, T. A. Dahlberg, Survivability Analysis for Mobile Cellular Networks, In Proceedings of Communications Networks and Distributed Systems Modeling and Simulation Conference, San Antonio, Texas, 2002.
[14] P. Demeester, M. Gryseels, A. Autenrieth, C. Brianza, L. Castagna, G. Signorelli, R. Clemenfe, M. Ravera, A. Jajszczyk, D. Janukowics, K. van Doorselaere, Y. Harada, Resilience in Multilayer Networks, IEEE Commun. Mag., 37(8) (1999) 70–76.
[15] J. Zhao, L. Lei, Y. Ji, D. Xu, Integrated Multilayer Survivability Strategy with Inter-layer Signaling, International Conference on Communication Technology Proc., 1 (2003) 612–616.
[16] K-C. Chu, F. Y-S. Lin, Survivability and Performance Optimization of Mobile Wireless Communication Networks in the Event of Base Station Failure. Computers and Electrical Engineering, 32(2006) 50–64.
[17] L. Nederlof, K. Struyve, C. O'Shea, H. Misser, Y. Du, B. Tamayo, B, End-to-end Survivable Broadband Networks. IEEE Communications Magazine, 339 (1995) 63–70.
[18] A. Koster, S. Orlowski, C. Raack, G. Baier, T. Engel, Single-layer Cuts for Multi-layer Network Design Problems, In Raghavan, S., Golden, B. and Wasil, E. (Ed.). Telecommunications Modeling Policy and Technology. Operations Research/Computer Science Interface. 44: 1-23, Springer Verlag, USA, 2008.
[19] S. Orlowski, Optimal Design of Survivable Multi-layer Telecommunication Networks. Ph.D. Thesis, Technischen Universitt Berlin, 2009.
[20] L. Yong, A study of survivability with multi-layer telecom network for services application. In Proceedings of 2^{nd} IEEE Symposium on Web Society (SWS), Beijing, 2010, 733-737.
[21] U. Steglich, T. Bauschert, C. Busing, M. Kutschka, A Generic Multi-layer Network Optimization Model with Demand Uncertainty. In Bauschert, T. et. al. (Ed.). Advances in Communication Networking. Lecture Notes in Computer Science 8115, 2013, 13-24.
[22] D. Hong, S. S. Rappaport, Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures, IEEE Transactions on Vehicular Technology, 35(3) (1986) 77–92.
[23] S. Tekinay, B. Jabbari, A Measurement-based Prioritization Scheme for Handovers in Mobile Cellular Networks, IEEE Journal on Selected Areas in Communications, 10(8) (1992) 1343–1350.
[24] Y. Fang, I. Chlamtac, Y-B. Lin, Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks, IEEE Transactions on Computers, 47(6) (1998) 779–692.
[25] Y. Fang, I. Chlamtac, Y-B. Lin, Call Performance for a PCS Network, IEEE Journal in Selected Areas in Communication, 15(8) (1997) 1568–1581.
[26] F. S. Lai, J. Misic, S. T. Chanson, S. T., Complete Sharing Vs Partitioning: Quality of Service Management for Wireless multimedia Networks, In Proceedings of 7^{th} International Conference on Computer Communications and Networks: IEEE Computer Society, Lafayette, Louisiana, USA (1998) 584–593.

[27] G. I. Tsiropoulos, D. G. Stratogiannis, H-H. Chen, P.G. Cottis, Utility Based Probabilistic Call Admission Control for Complete Fairness in Wireless Networks. International Journal of Communication Systems. 27(3) (2014) 521–528.

[28] M. Ekpenyong, J. Isabona , Modeling Mobility in Mobile Ad-hoc Environments. Journal of the Nigerian Association of Mathematical Physics. 18 (2011) 407–412.

[29] R. Maidstone, Discrete Event Simulation, System Dynamics and Agent Based Simulation: Discussion and Comparison. Short Project, STOR-I, Lancaster University, UK, 2012.

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF UYO, UYO, NIGERIA
E-mail address: `mosesekpenyong@(uniuyo.edu.ng, gmail.com)`
DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF UYO, UYO, NIGERIA
E-mail addresses: `danielasuquo@uniuyo.edu.ng`