G/G/c/c Simulation Model for VoIP Traffic Engineering with non-Parametric Validation

Imad AL Ajarmeh
James Yu
School of Computing and Digital Media
DePaul University
Chicago, USA
{iajarmeh, jyu}@cdm.depaul.edu

Mohamed Amezziane
Dept. of Mathematical Sciences
DePaul University
Chicago, USA
mamezzia@depaul.edu

***Abstract*— The widespread of Voice over IP systems necessitate finding suitable and modern traffic engineering models that can help design cost-efficient systems and study their performance under various conditions. In this paper, we provide a new VoIP simulation suite that consists of a parametric simulator based on Non-homogeneous Poisson Process call arrival model, and a non-parametric simulator based on real traffic data. Our simulators are validated against real call data obtained from multiple offices of a production VoIP carrier network. The simulation results show that our simulator can provide up to 28% better resource utilization than the legacy Erlang B model. Our simulator can also help carriers dynamically allocate network bandwidth to meet various traffic demands.**

Keywords- VoIP traffic engineering; simulatin; VSIM; modeling; G/G/c/c; non-parametric.

I. INTRODUCTION

The majority of the previous work in Voice over IP (VoIP) traffic engineering and modeling is based on the exponential approximations for call arrival rate and call holding time [3][16][17][18]. The exponential approximation allows finding analytical solution for the traffic queuing model but the approximation might be too aggressive that it will result in poorly engineered systems. The Erlang B model was introduced several decades ago to solve the phone system traffic queuing problem. This model is based on the traffic intensity of the busiest hour in the busiest week of the year (Busy Season Busy Hour: BSBH). BSBH traffic is assumed constant throughout the entire year and its arrival rate is modeled as a Poisson/exponential distribution. This assumption makes traffic calculations easier but using a constant call arrival rate for the entire year causes inefficient resource utilization.

We proposed using a Non-homogeneous Poisson Process (NHPP) for the call arrival rate in our previous work [1]. In NHPP modeling call arrival rate is a non-constant function of time. Whereas in the legacy Erlang B approach calls are assumed to arrive according to a Poisson process with a constant arrival rate. Therefore, using NHPP helps avoid the approximation and assumption errors associated with a constant arrival rate over the whole engineering period. Our NHPP model development was based on real call data extracted from a production VoIP carrier network. Examining the arrival data, we constructed a model that describes the variation of call arrival rates during a week since traffic patterns were observed to be repeated weekly. It is common in statistical analysis to model the logarithm of λ(t) instead of λ(t) itself for *count* data. Such transformation would guarantee that the estimate of the intensity function is always non-negative. Our model takes into consideration the daily arrival patterns and has the time-dependent intensity function of:

$$\boldsymbol{log}[\lambda(\boldsymbol{t})] = \boldsymbol{\mu} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{k_o}} [\boldsymbol{\alpha_i} \sin(\boldsymbol{i\omega_o t}) + \boldsymbol{\beta_i} \cos(\boldsymbol{i\omega_o t})] + \sum_{\boldsymbol{j=1}}^{\boldsymbol{6}} \boldsymbol{\gamma_j I_j(t)} \qquad (\mathbf{1})$$

where: $\boldsymbol{\lambda(t)}$ is a function of time (t).

$\boldsymbol{I_j(t)}$ is day Indicator function where $\boldsymbol{j}$ is the day of the week. The value of $\boldsymbol{I_j(t)}$ is 1 if the time $\boldsymbol{t} \in \boldsymbol{j}$ and 0 otherwise. $\boldsymbol{k_o}$ is the number of harmonics in the model. $\boldsymbol{\mu}$ represents the model central tendency without daily effects. $\boldsymbol{\gamma_j}$ is the effect of day *j* and represents the difference between $\boldsymbol{\mu}$ and the mean number of calls for day *j*. $\boldsymbol{\alpha_i}$, and $\boldsymbol{\beta_i}$ are the contribution of the *ith* harmonic to the model.

We used maximum likelihood estimation to fit the proposed λ(t) to the actual call arrivals. In addition, we used likelihood ratio test and Wald's test in order to verify the significance of the model and its parameters. All the statistical test results verify that call arrivals is best fit by a NHPP rather than a constant Poisson process. We provide the detailed statistical analysis in [1]. In this paper, we introduce a comprehensive VoIP simulation suite (VSIM). VSIM consists of a G/G/c/c simulation model. According to Kendall's notation the G/G/c/c is a queuing system where calls are assumed to arrive according to a general distribution (G) and have a service time that follows another general distribution (G), the system has a limited

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-262-2

number of servers/channels (c) and no waiting queue (maximum number of calls in the system equals the number of servers c). VSIM includes a NHPP call arrival rate G/G/c/c simulator and also a non-parametric simulator based on real traffic data. The simulation models are validated against traffic data collected from an operational VoIP network.

The remaining of the paper is organized as follows: section II contains a description of the telecommunication system simulation approaches, section III gives detailed description of VSIM, section IV describes our approach in verifying the correctness of VSIM algorithms, in section V we show our process in validating VSIM results, and in section VI we present some results of VSIM simulations.

II. TELECOMMUNICATION SYSTEM SIMULATION

In telecommunication traffic engineering, it is always preferable to find analytical solutions for the queuing and traffic problems. However, the analytical solution might involve too many approximations in order to fit the data into exponential or other probability distribution functions. Such approximations will result in inaccurate engineering results. Simulation approach offers accurate and flexible model construction and validation, and can be used whenever analytical solutions are not practical [11].

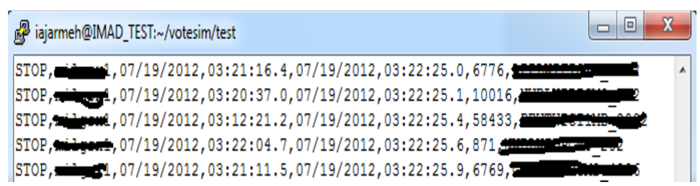
Simulation models can be discrete or continuous. Discrete event simulations are suitable for problems where variables change in discrete time fashion. On the other hand, continuous simulations are suitable for problems in which the variables might change continuously [8]. Discrete event simulations are suitable for telecommunication network queuing problems since the events happen on discrete times [9]. Using discrete event simulators has been becoming more popular during the past few years because such simulations can help solving sophisticated problems, which are impossible to be solved using analytical approaches [10][11]. In addition, the availability of low-cost powerful computers and capable simulation packages makes the simulation-based solutions more accurate, capable, and easier to implement.

With the rapid increase of VoIP residential, enterprise, and carrier deployments, researchers realized the need for modern traffic simulation models that can be used in studying and designing reliable and cost-efficient VoIP networks. In [4] VoIP traffic sources were modeled as on-off sources with exponentially distributed of on-and-off times. In [6] and [7] the authors used Markov modulated Poisson process (MMPP) traffic model to analyze VoIP performance for wired and wireless networks. In [12] [13] [14] [15] the authors provide VoIP traffic performance and evaluation simulation tools that focus on the packet performance without taking into consideration the distribution of calls arriving at the system, which will have significant impact on the packet performance and QoS design. In this work we go a further step by providing two VoIP traffic simulators based on modeling the calls arriving at the system. The first one is a parametric model and uses a NHPP to represent the time-dependent call arrival rate, and the second one is non-parametric and uses the real traffic data to simulate the system behavior

III. VOIP TRAFFIC SIMULATION MODEL (VSIM)

VSIM is part of a larger traffic engineering system that starts by collecting call data from a production network. The collected data is processed and then fed into the NHPP parameter estimation model. NHPP model parameters are passed to VSIM to be used in G/G/c/c engine. For non-parametric simulation we skip NHPP model estimation process and feed the processed call arrival and call holding time data directly into VSIM. Figure 2 illustrates a high-level design for such traffic engineering system, and Figure 1 shows a sample of the collected call data. Each raw carries information for one call with some proprietary data (trunk group name and switch name) blanked. The remaining fields are: record type (STOP means a completed call), start date and time, end date and time, and call duration in 100s of seconds.

Examining the traffic patterns in the collected data, we notice large variation in the arrival rate. For example, at one second we might receive 10 calls and at the next second we might receive no calls. This variation is smoothed if we average the traffic data over longer time intervals. Figure 3 shows the raw traffic data for 1s, 10s, and 3600s averages, along with the generated NHPP model. The figure illustrates the accuracy and significance of the generated model. This accuracy has been established through the extensive mathematical and statistical analysis we provided in [1] and [2]. The accuracy of the input NHPP model will result in accurate simulation results as proven in section VII


iajarmeh@IMAD_TEST:~/votesim/test

STOP,[redacted],07/19/2012,03:21:16.4,07/19/2012,03:22:25.0,6776,[redacted]
STOP,[redacted],07/19/2012,03:20:37.0,07/19/2012,03:22:25.1,10016,[redacted]
STOP,[redacted],07/19/2012,03:12:21.2,07/19/2012,03:22:25.4,58433,[redacted]
STOP,[redacted],07/19/2012,03:22:04.7,07/19/2012,03:22:25.6,871,[redacted]
STOP,[redacted],07/19/2012,03:21:11.5,07/19/2012,03:22:25.9,6769,[redacted]

Figure 1. Sample of collected call data

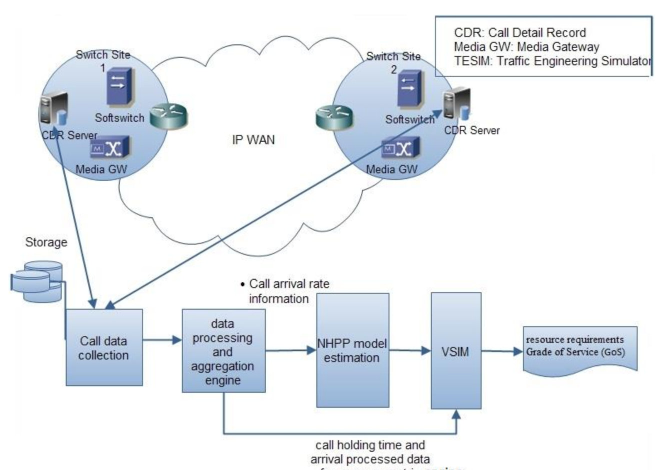


Figure 2. Traffic engineering system

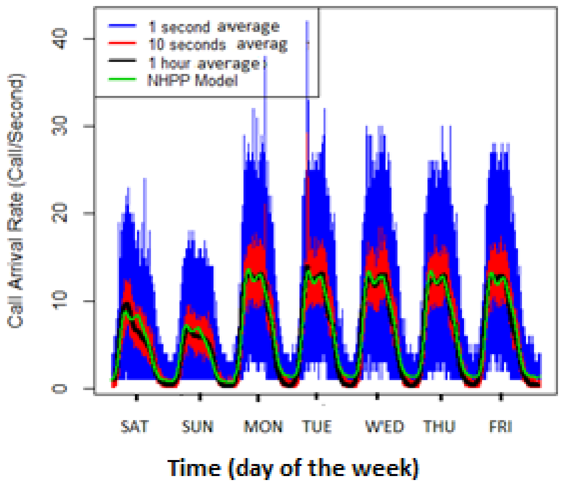


Figure 3. Call arrival data analysis and modeling

The VoIP simulation model (VSIM) estimates:

- resource requirements (IP trunk size) given a certain traffic pattern and a target blocking probability (Grade of Service: GoS), and
- GoS (blocking probability) given a certain traffic pattern and a known amount of call resources (IP trunks).

Figure 4 shows a sample of VSIM output.

VSIM is a flexible Java-based tool developed using CSIM for Java library [4], and therefore it can easily be ported to different computing platforms. VSIM can be used to estimate the trunk group size, generate GoS reports, and perform what-if analysis for VoIP networks.

VSIM is composed of two different simulation engines; the first is a parametric G/G/c/c simulator and the second is a non-parametric simulator. Both are discrete event simulators in which the VoIP system is modeled as chronological sequence of call arrivals and terminations. In the G/G/c/c engine we model the call arrival rate using the time-dependent function shown in (1), model parameters are estimated based on the collected data sample.

This call arrival function is used to generate random variables for call inter-arrival times. Once a call is generated, the simulation code polls a random call holding time from a list of real holding times. The simulation engine allocates a trunk for the duration of simulated call holding time. A separate thread is created for each call so that we can collect statistics for each individual call and trunk.

```
imad@linux18:/home/VSIM
Execution time: 22.826 s
Simulated Time: 604800.0 S
Simulated Time Offset: 0.0 S
Number of IP Trunks = 200
Holding Time data file = data/ht-nycgsx17-Mar-w24
NHPP parameters file = data/nhpp-nycgsx17-mar-w24
Blocked calls= 1635
Completed calls= 233587
Blocking Probability = 0.006950880444856348

Output is saved to GGK.out
```

Figure 4. VSIM sample output

Once the simulated call time is over, the completed calls counter is incremented by one and the trunk will be released back to the trunk pool.

The same procedure is repeated for the next calls until the pool of trunks is depleted. Once all trunks are busy we increment the blocked calls counter for each call that arrives while no trunks are available. Figure 5 illustrates the internal VSIM algorithm

The non-parametric simulator follows the same algorithm with the exception that we poll the inter-arrival time variable from a real data file rather than using a NHPP function to generate it.

A. Parametric VSIM G/G/c/c simulator

Parametric simulation is done by collecting traffic data and then developing statistical models that best approximate the collected data. Model parameters are estimated based on the data sample and then these parameters are used in the simulation. We developed a G/G/c/c simulation model for VoIP traffic engineering. The model consists of a loss multi-server queuing system with waiting queue length equal to zero (blocked calls are cleared from the system). The implementation of general call arrival rate and general call holding time in the simulator allows for arbitrary distributions and that increases the flexibility and usability of our simulation model. The examples given in this paper focus on modeling call arrival rate as NHPP using a generalized linear model that captures the variability in call arrival rate with respect to time. NHPP model parameters are estimated based on the real traffic data extracted from the production VoIP network under study.

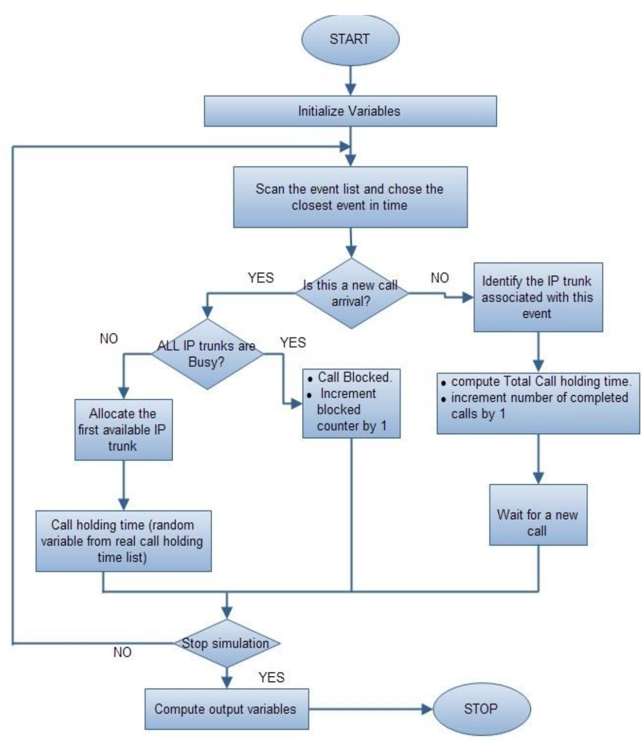


Figure 5. VSIM algorithm

B. *Non-parametric VSIM simulator*

In addition to the parametric G/G/c/c simulation engine implemented in VSIM, we also provide another non-parametric simulation engine. Non-parametric simulation is achieved throughout replaying the real traffic data without generating statistical models or estimating parameters. In other words, we use actual observations in the simulation rather than generating random variables from a statistical distribution. Therefore, non-parametric simulations are easier and tend to yield more accurate results since no modeling approximations are involved. This type of simulation is preferred when the data sample is large.

We use the non-parametric simulator in order to verify the correctness of our G/G/c/c model and to validate its results. The input of our non-parametric simulator is real call information that consists of call inter-arrival time and call holding time. The simulator will regenerate the calls based on the given data, and we can study the system and compute the required resources and GoS.

IV. VSIM MODEL VERIFICATION

It is important to verify the correctness of any simulation model before applying it to real-life problems. Simulation verification should cover the simulation engine algorithms as well as the random variables generated from the simulator statistical models. Therefore, we split VSIM model verification into two steps; the first is discussed in (*A*) and aims to verify the correctness of the NHPP random variables generated and used by the simulator. The second step is discussed in (*B*) and aims to verify the correctness of the simulation algorithms, timers and procedures

A. *Internal simulation random variables*

We instrumented VSIM and obtained the call arrival rate generated by the model based on the implemented NHPP linear model. This call rate is used as an internal input to the G/G/c/c simulation algorithm. Figure 6 shows the internally generated NHPP random variable along with the corresponding inter-arrival time against the actual traffic data

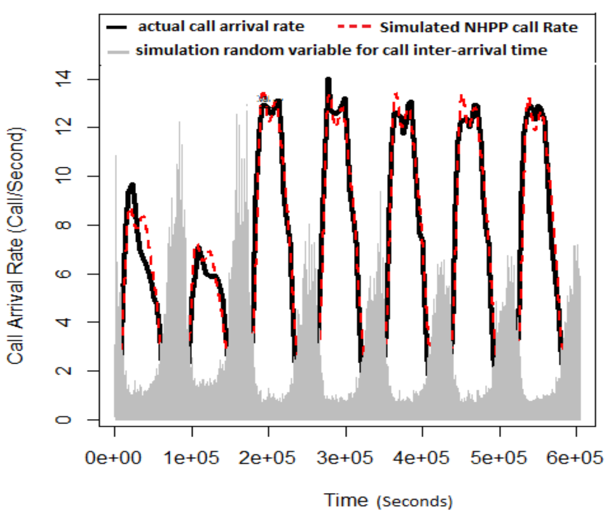


Figure 6. Simulated call arrival rate

As illustrated in Figure 6, the results of this verification process indicate that the NHPP model function used to generate call arrival random variables is correct and accurate.

B. *VSIM simulation engine algorithms*

A common approach to verify the correctness of a new simulation model is by comparing it to other well-established simulators. Unfortunately, we could not obtain any G/G/c/c model against which we could verify our work; therefore, we used the M/M/c/c special case of our model and compared the simulated result to the calculated results based on the Erlang B model. Our M/M/c/c utilizes the same simulation algorithms as our G/G/c/c and the only difference is that we use exponential distributions for both call arrival rate and call holding time. The goal is to verify the correctness of our simulation clock and algorithms. The validation of simulation results will be discussed in the next section. In this process we used different data samples each one consists of one week of traffic; an example of the results is shown below:

Mean call holding time = 185 second
Busy Hour Traffic (BHT) = 12.6 call/ second

Using the Erlang B calculator we need around 2520 trunks in order to carry this traffic without blocking (Blocking probability nearing zero). Using the M/M/c/c simulator with the same traffic parameters we found the required number of trunks to be 2515 for the same blocking probability. Comparable results were obtained for all the samples under test. These results verify the correctness of our VSIM simulation code and algorithms.

V. VSIM SIMULATION MODEL VALIDATION

The most important aspect of any simulator is that it should produce valid, correct and dependable results. The best approach to establish the validity of a simulator results is by comparison to real data. Therefore, we obtained the real resource utilization (number of simultaneous calls) from the production system for the period of time corresponding to the call arrival and holding time data used to develop the models. We used this data to validate VSIM simulation results. Different traffic samples and different simulation runs were used and all results agree with data obtained from the real network. In addition, we also used our nonparametric VSIM simulator to replay the same data samples and the results agree with those obtained from the system and those obtained from G/G/c/c simulation.

TABLE I. shows an example of the actual trunks obtained from the system compared to VSIM simulated output

TABLE I. SIMULATED VS ACTUAL IP TRUNK REQUIREMENTS

Number of Required IP Trunks	**Maximum Call Load ($Pr[B] \approx 0$)**
Actual (Observed)	1807
G/G/c/c (simulated)	1936
Non-Parametric (Simulated)	1810

It can be seen from TABLE I. that both simulation models yield satisfactory results although the non-parametric model is a little better. The reason is because we don't have any modeling or estimation approximations for the non-parametric case. We used multiple data samples and executed multiple simulation runs and all the results are similar and indicate high accuracy of our VSIM for both G/G/c/c and non-parametric while the latter shows a little better results.

VI. RESULTS AND ANALYSIS

VSIM G/G/c/c simulator is based on using a function of time to model call arrival rate, and therefore VSIM can provide the resource requirements as a function of time as well. This function is important for system design, analysis and requirement studies, especially for converged networks where voice and data ride the same IP infra-structure. This function is also available for VSIM non-parametric simulator because we have real call information that depends on the time. Figure 7 illustrates sample resource functions (number of required IP trunks Vs. Time) generated by VSIM along with the corresponding simultaneous calls observed in the actual system (real data validation). The figure shows the effectiveness of VSIM as demonstrated by its ability to compute the required system resources (IP trunks) as a function of time accurately. The resource time function provided by VSIM can be utilized for dynamic resource allocation scheme in which resources are allocated for different applications based on the actual or expected demand. Such scheme helps achieve better resource utilization and hence better engineering and cost reduction. It is important to notice that Erlang B and M/M/c/c models suggest a linear relation between blocking probability and system capacity (maximum number of simultaneous calls). However, our VSIM G/G/c/c and non-parametric simulators suggest a non-linear relation as seen in Figure 8.

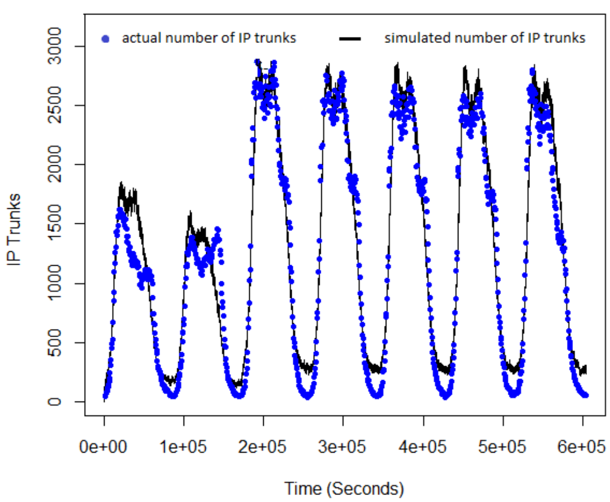


Figure 7. Number of IP trunks as a function of time (resource time function)

The figure shows identical match between the results obtained from the calculated Erlang B and the simulated M/M/c/c. In order to use Erlang B and M/M/c/c, we need to compute the average call arrival rate of the busiest hour and the average call holding time. For the results shown in Figure 8 and based on the data sample, we used 10.6 call/second for the arrival rate and 183.01 seconds for call holding time. Also, the figure shows close match between the G/G/c/c and non-parametric simulator. These results verify the correctness and validity of our procedure and modeling process. The deviation between the straight line calculated by the traditional Erlang B model and the curve generated by VSIM is significant and can affect the design and engineering decisions for the system. For example, if we want to build a switching system with 1600 maximum simultaneous calls (system capacity), the Erlang B approach suggests that the blocking probability will be 0.19 (P.19) while the VSIM G/G/c/c model results in a blocking probability of 0.02 (P.02). The difference between these two approaches is significant in the telecom world. VSIM nonparametric approach for the same data sample results in a blocking probability of 0.006 (P.006). Using the same example, we found that in order to achieve blocking probability of 0.01 (P.01), we will provision 1665 IP trunks using VSIM G/G/c/c model, or provision 1550 IP trunks using the VSIM non-parametric simulator. On the other hand we will provision 1991 trunks if we engineer the system using the Erlang B model. Therefore, we can see that using the VSIM model can save 28% of the resources over Erlang B at the P.01 blocking probability. Furthermore, Figure 8 suggests that we can achieve better than 28% resource saving if higher blocking probabilities is desired.

We conducted many simulation runs using different data samples collected from different switch offices located in different cities. Similar results were obtained throughout this study. Figure 9 shows another example where the data is collected from a different office with more trunks. It shows almost identical relation between the blocking probability and system resources.

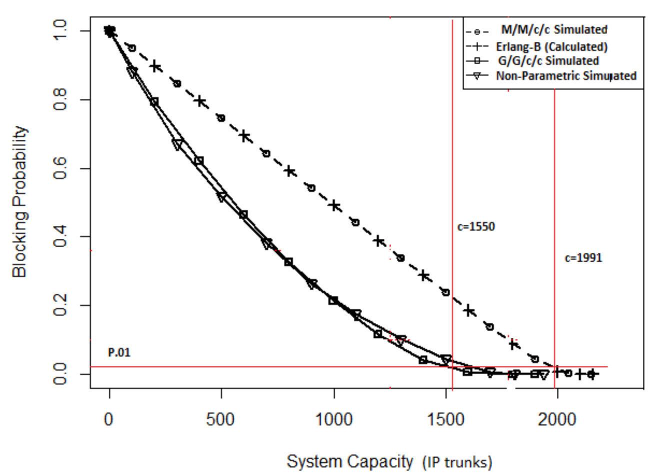


Figure 8. Blocking probability Vs system capacity

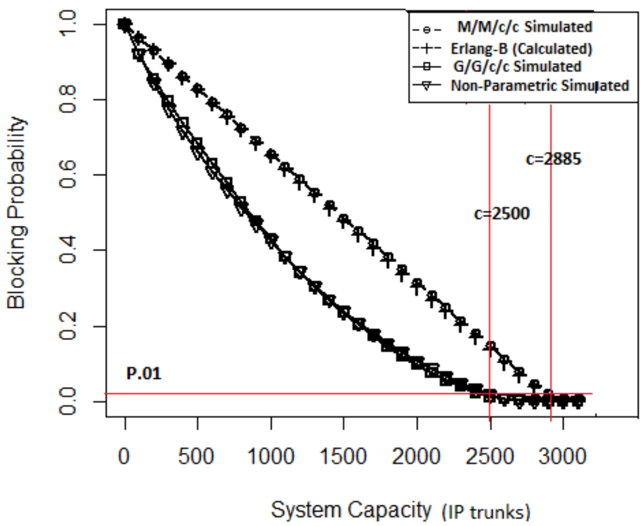


Figure 9. Blocking probability Vs system capacity example 2

For the results shown in Figure 9 and based on the data sample, we used 14.0125 call/second for the arrival rate and 204.186 seconds for call holding time

The graph suggests that we need 2500 IP trunks in order to carry the offered traffic load at blocking probability of 1%. On the other hand using the Erlang B model we will need 2885 IP trunks in order to achieve the same GoS. This sample shows that VSIM model could yield a saving of 15% of network resources in this office.

VII. CONCLUSION

One disadvantage of using complex traffic models such as NHPP is that an analytical solution is not feasible. With the availability of powerful computers, it makes the simulation approach feasible and effective, and hence, we can obtain accurate results. We provided two simulators: the first is based on NHPP call arrival rate and the second is based on non-parametric data. VSIM is capable of solving the traffic modeling problem for modern VoIP systems accurately using arbitrary and complex traffic models or by using the raw traffic information without estimation of parameters. Our results are validated against real data collected from multiple offices of a production VoIP carrier network. We observed that the non-parametric simulator results are more accurate. Real traffic data proves that using VSIM could save up to 28% of the resources over the Erlang B model or other exponential-based models

REFERENCES

[1] I. Al Ajarmeh, J. Yu, and M. Amezziane, "Framework of applying a non-homogeneous Poisson process to model VoIP traffic on tandem networks," 10th WSEAS International Conference on Informatics and Communications, Taipei, Taiwan, August 2010.

[2] I. Al Ajarmeh, J. Yu, and M. Amezziane, "Modeling VoIP traffic on converged IP networks with dynamic resource allocation," International Journal of Communications, ISSN: 1998-4480, Issue 1, Volume 4, 2010, pp. 47-55

[3] K. Salah and A. Alkhoraidly, "An OPNET-based simulation approach for deploying VoIP," International Journal of Network Management, Volume 16 Issue 3, May 2006, pp. 159-183

[4] http://www.mesquite.com/documentation/documents/CSIM_for_Java-UserGuide.pdf [retrieved: Feb, 2013]

[5] V. Rakocevic, R. Stewart and R. Flynn, "VoIP network dimensioning using delay and loss bounds for voice and data applications," Techincal Report, 2008.

[6] H. Lee and D. Cho, "VoIP capacity analysis in cognitive radio system," IEEE Communications Letters, VOL. 13, NO. 6, June 2009

[7] J-W. So, "Performance analysis of VoIP services in the IEEE 802.16e OFDMA system with inband signaling," IEEE Trans. Veh. Technol., vol.57, no. 3, May 2008, pp. 1876–1886

[8] A. M. Law and W. D. Kelton, "Simulation modeling and analysis," third ed. New York: McGraw-Hill, 2000

[9] O. Özgün and Y. Barlas, "Discrete vs. continuous simulation: when does it matter?," Proceedings of the 27th International Conference of The System Dynamics Society, July, 2009, Albuquerque, NM, USA

[10] G. Yuehong, Z. Xin, Y. Dacheng, and J. Yuming, "Unified simulation evaluation for mobile broadband technologies," IEEE Communications Magazine, vol. 47, Issue 3, March 2009, pp. 142-149

[11] N. I. Sarkar and S. A. Halim, "A review of simulation of telecommunication networks: simulators, classification, comparison, methodologies, and recommendations," Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), March Edition, 2011

[12] M. Bohge and M. Renwanz, "A realistic VoIP traffic generation and evaluation tool for OMNeT++," International Workshop on OMNeT++, Marseille, France, March 2008

[13] A. Bacioccola, C. Cicconetti, and G. Stea, "User-level performance evaluation of VoIP using ns-2," NSTools'07, October 22, 2007, Nantes, France

[14] J. D. Gupta, S. Howard, and A. Howard, "Traffic behavior of VoIP in a simulated access network," International Transactions on Engineering, Computing and Technology, 18, ISSN 1305-5313, 2006, pp. 189-194

[15] R. S. Naoum and M. Maswady, "Performance evaluation for VOIP over IP and MPLS," World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 3, 2012, pp. 110-114

[16] S. Shin and H. Schulzrinne, "Experimental measurement of the capacity for VoIP traffic in IEEE 802.11 WLANs," INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, May 2007

[17] T. D. Dang, B. Sonkoly, and S. Molnár, "Fractal analysis and modeling of VoIP traffic," 11th International Telecommunications Network Strategy and Planning Symposium, Austria 2004

[18] M. J. Fischer and D. M. Bevilacqua Masi, "Modeling overloaded voice over Internet Protocol systems," The Telecommunications Review 2006, pp. 94 - 106