

MODELING AND SIMULATION OF INTEGRATED VOICE/DATA CELLULAR COMMUNICATION WITH GENERALLY DISTRIBUTED DELAY FOR END VOICE CALLS

Pawan K. Choudhary

Department of Electrical & Computer Engineering
Duke University
Durham, NC 27705-3750, U.S.A.

Bharat Madan

Information Science and Technology
ARL, Penn State University
State College, PA 16801, U.S.A.

Kishor S. Trivedi

Department of Electrical & Computer Engineering
Duke University
Durham, NC 27705-3750, U.S.A.

ABSTRACT

Cellular networks are gradually shifting from voice only to voice and data due to increased demand for WWW, FTP and multi-media messaging. This has substantially increased the volume of cellular data traffic. Schemes have been proposed for co-existence and fair sharing of the available bandwidth between circuit switched voice and packetized data traffic. One popular scheme is based on delaying the last incoming acceptable voice call for a random amount of time and utilizing this time to service the accumulated data traffic to prevent excessive queuing delay in delivering the data traffic. However, this delay time has been usually modeled as being exponentially distributed. In this paper, this delay is modeled as a uniformly distributed random variable instead. This scheme is extended by delaying more than one call. Performance measures like voice and data blocking probability is plotted as function of last call delay.

1 INTRODUCTION

Newer generation cellular services are characterized by laying greater emphasis on providing packet data services along with the existing voice services. Circuit-switched 2G (second generation) cellular networks are giving way to 2.5G and 3G (third generation) converged (voice and data) packet-switched cellular networks. This implies that the available resources, such as the bandwidth, need to be shared between the voice traffic and the data traffic. The voice services are primarily circuit switched so as to be able to meet the real time constraints of voice traffic. Voice over IP is also getting popular, but it is nowhere near

replacing circuit switched voice telephony. Apart from this, there is always need for connection oriented services to meet QoS constraints of real time applications including critical voice calls. This suggests that significant amount of voice traffic will be carried on circuit switched network in near future. In contrast, the data traffic is typically treated as a best effort service and, at the network layer, it tends to be connection-less datagram oriented (Connection oriented data services, such as the TCP also utilize the best effort IP or network layer datagram services). In spite of it being a best effort service, many data applications need to be provided a desirable level of quality-of-service (QoS). Desired QoS for data traffic is possible only when there is a proper sharing of the bandwidth resource between the voice and the data traffic. Several voice/data co-existence schemes have been proposed by various researchers. Koutsakis and Paterakis (2001) and Mitrou et al. (1993) have proposed new medium access protocol and call admission procedure for efficient integration of voice, video and data. Ivanov et al. (2003) have evaluated the performance of mixed traffic with varying number of shared and reserved packet data channels (PDCHs) through simulation. Cesana et al. (2002) also have evaluated different traffic scenarios (shared and non-shared data, voice and video) through simulation.

The main contribution of this paper can be summarized as follows. 1) Models for sharing of voice and data traffic have been proposed for 2.5G and 3G networks, with delay of one or more end calls generally distributed. 2) A closed form formula has been derived for the case when one call is delayed and the results have been verified by simulation. 3) When more than one call are delayed, it is shown that the resulting model cannot be solved by standard analytical methods (e.g. SMP, MRGP) and simulation is used to evaluate

its performance. SMP denotes semi-Markov process and MRGP denotes Markov regenerative process. See Cinlar (1975) and Kulkarni (1995) for more details.

2 DOVE WITH GENERALLY DISTRIBUTED DELAY TIME

In an environment of mixed voice and data traffic, bandwidth is assigned on a first come first served basis. However, since the voice service is connection oriented, once a voice call is admitted into the network, the channel (or bandwidth) once assigned remains permanently associated with this call. Contrast this with the data traffic for which bandwidth is temporarily assigned and once data packet(s) is delivered to the network, the bandwidth is returned for reuse. This implies there is finite and significant probability that data traffic may get starved of bandwidth, particularly under heavy voice traffic. To alleviate the situations, Mahdavi et al. (2001) suggested that the admittance of the last incoming voice call be delayed by a small amount. The resulting algorithm is referred to as the delay of voice end (DOVE) call. This delay can be utilized to temporarily assign the channel to carry the data packets that may have queued up. This will have the beneficial effect of preventing build up of large queuing delays for data traffic. The DOVE algorithm however, makes the assumption that this delay be a random quantity with exponential distribution. The exponential assumption for this delay time simplifies the analysis since the under stochastic process is then a homogeneous continuous time Markov chain (CTMC). However, this assumption may be hard to justify on account of the fact that exponential distribution also implies a finite and significant probability of delay in admitting a call for arbitrarily long intervals of time, which may be unacceptable to the users. To alleviate this problem, we can assume this delay to be uniformly distributed, $U(T_1, T_2)$.

The resulting state diagram for the case when only the last call is delayed is shown in Figure 1. In this state diagram, first index denote the total number of channels available and these are assumed to be n to begin with. The second index denote the number of ongoing voice calls. The arrival process is assumed to be Poisson with rate parameter λ . The voice call completion time is assumed to be exponentially distributed with service rate μ . We make two simplifying assumptions, (i) servicing of data packet takes a very small amount of time, and (ii) incoming voice call has priority with pre-emption (Mahdavi et al. 2001). Both these assumptions imply that an incoming voice call will never encounter a situation where it is denied a channel. The only exception is when the incoming voice call needs to be assigned the last available channel. In which case, this call will be delayed. Under these simplifying assumptions, we no longer need to worry about data traffic having any detrimental effect on the performance with respect to the

voice calls. Consequentially, we only need to deal with the voice traffic in our models. The last voice call is subjected to a random delay, and the transition is denoted by $G(\cdot)$ in the state diagram. The amount of delay which end voice call suffers is same as time interval between two $G(\cdot)$ transitions, which is assumed to have a $U(T_1, T_2)$ distribution. Note that there are two competing transitions $G(\cdot)$ and $EXP((n-1)\mu)$ ready to occur. If some ongoing voice call gets completed before the random delay time has elapsed, transition $EXP((n-1)\mu)$ will occur to denote such an event. The resulting freeing up of a channel will ensure that we no longer need to delay any voice call. In this state diagram, there is one transition, namely $G(\cdot)$ which is non-exponentially distributed. Moreover, when $G(\cdot)$ is enabled, all other enabled transitions are exponentially distributed. Also, occurrence of transition $G(\cdot)$ disables $EXP((n-1)\mu)$ and vice-a-versa, thereby making both these transition *competitive*. The results derived in Dugan et al. (1984) suggests that under these situations, the underlying stochastic process for this model becomes an semi Markov process.

3 GENERALLY DISTRIBUTED MULTIPLE DELAY SLOTS

Data traffic frequently exhibits bursty behavior. This behavior can cause a large number of packets to queue up at the data queue, which can have adverse effects on the quality of service (QoS) for data traffic. A simple solution to this problem may be to delay more than just the end call. Generalization of the basic DOVE scheme by delaying more than end voice calls has been suggested by Okuda et al. (2002) in their extended-DOVE (E-DOVE) algorithm. The main idea being that by delaying more than one end voice call, the network can temporarily free up more bandwidth for data traffic. Once again, in the E-DOVE algorithm, it will be preferable to model delays as being uniformly distributed. E-DOVE model is shown in Figure 2. In this the first index denotes the number of available channels and the second index denotes number of channels currently being used. Sum of these two indices when subtracted by n gives number of delayed voice calls. As an example, let us look at the specific state $\langle 1, n-2 \rangle$. In this state, three transitions are enabled, one non-exponential transition $G(\cdot)$ and two exponential transitions, $EXP(\lambda)$ and $EXP((n-2)\mu)$, are enabled simultaneously. This suggests that this is not a semi Markov process. Also from state $\langle 0, n-2 \rangle$ three transitions are enabled. Arrival transition $EXP(\lambda)$ will result in blocking of new call arrival whereas with $EXP((n-2)\mu)$ transition will occur to state $\langle 1, n-2 \rangle$. Note that in state $\langle 0, n-2 \rangle$, there is competition between two general transitions. This suggests that the underlying stochastic process is not even a Markov regenerative process. From the above discussion we conclude that the system is not

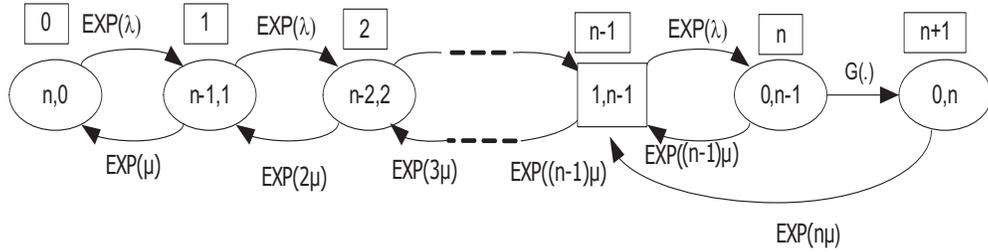


Figure 1: Semi Markov Process for DOVE

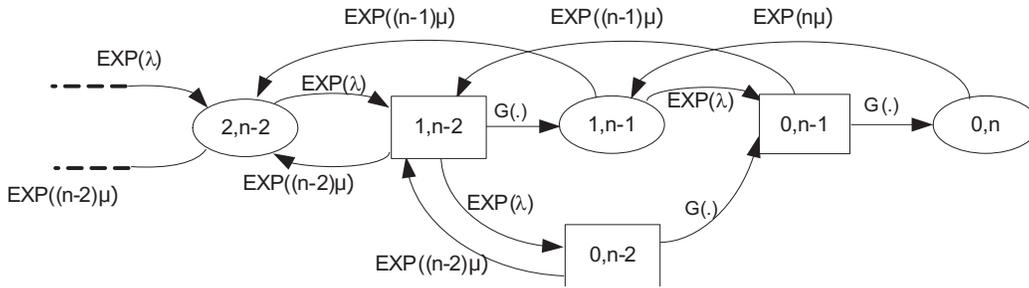


Figure 2: State Diagram for E-DOVE

a homogenous continuous Markov chain, a semi Markov process or a Markov regenerative process.

4 NUMERICAL RESULTS

This section deals with analytical closed form solution of the DOVE-SMP and simulative solution of the E-DOVE.

4.1 SMP Solution for DOVE

The SMP for the DOVE algorithm is given by the state diagram shown in Figure 1. This SMP can be solved by working its embedded DTMC which can be derived by observing this SMP only at those discrete instants of time at which state transitions occur. The resulting DTMC is characterized by its transition probability matrix P which is given in Figure 3. Here, $\rho = \lambda/\mu$ and p denotes the transition probability from state n to state $n + 1$. Because the time interval for this transition is not exponentially distributed, p can be computed by assuming two random time intervals, $X =$ time to transition from state n to state $(n - 1)$ (i.e., time to service the next voice call in state n) and $Y =$ delay time for the end voice call. Note that X is $EXP((n - 1)\mu)$ and Y is $U(T_1, T_2)$. Therefore, the transition probability p can be expressed as $P(X > Y)$,

which is given by:

$$\begin{aligned}
 p &= p_{n,n+1} = P(X > Y) & (1) \\
 &= \int_0^\infty P(X > t) f_Y(t) dt \\
 &= 0 + \int_{T_1}^{T_2} e^{-(n-1)\mu t} \frac{1}{T_2 - T_1} dt + 0 \\
 &= \frac{1}{\mu(n-1)(T_2 - T_1)} (e^{-(n-1)\mu T_1} - e^{-(n-1)\mu T_2}).
 \end{aligned}$$

To solve an SMP, we first need to solve the embedded DTMC in terms of its steady-state probabilities $v = [v_i]$, $i = 0, 1, \dots, n, n + 1$ by solving the left eigenvalue problem:

$$v = vP, \quad \sum_i v_i = 1. \quad (2)$$

The SMP steady-state probabilities π_k are then given by:

$$\pi_k = \frac{v_k h_k}{\sum_{i=0}^{n+1} v_i h_i}, \quad k = 0, 1, \dots, n, n + 1 \quad (3)$$

where, h_i denotes the mean sojourn time in state i of the SMP. Solving Eqn (2), the steady state embedded DTMC

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & \dots & n-3 & n-2 & n-1 & n & n+1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ n-2 \\ n-1 \\ n \\ n+1 \end{matrix} & \left[\begin{array}{cccccccccccc} 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{1+\rho} & 0 & \frac{\rho}{1+\rho} & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{1+\frac{\rho}{2}} & 0 & \frac{\frac{\rho}{2}}{1+\frac{\rho}{2}} & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1+\frac{\rho}{3}} & 0 & \frac{\frac{\rho}{3}}{1+\frac{\rho}{3}} & \dots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & \frac{1}{1+\frac{\rho}{n-2}} & 0 & \frac{\frac{\rho}{n-2}}{1+\frac{\rho}{n-2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & \frac{1}{1+\frac{\rho}{n-1}} & 0 & \frac{\frac{\rho}{n-1}}{1+\frac{\rho}{n-1}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 \end{array} \right] \end{matrix}$$

Figure 3: Transition Probability Matrix P

probabilities are given by:

$$\begin{aligned} v_k &= \left(1 + \frac{\rho}{k}\right) \frac{\rho^{k-1}}{(k-1)!} v_0, \quad 1 \leq k \leq (n-1) \\ v_n &= \frac{\rho^{n-1}}{(n-1)!} v_0 \\ v_{n+1} &= \frac{p\rho^{n-1}}{(n-1)!} v_0 \end{aligned} \quad (4)$$

and,

$$v_0 = \left(1 + \frac{(1+p)\rho^{n-1}}{(n-1)!} + \sum_{k=1}^{n-1} \left(1 + \frac{\rho}{k}\right) \frac{\rho^{k-1}}{(k-1)!}\right)^{-1} \quad (5)$$

Note that for all states other than state n , the sojourn time distribution is exponential with rate equal to the net rate out of that state. For state n , the sojourn time is the minimum of $EXP((n-1)\mu)$ and $U(T_1, T_2)$ random variables. From these observations, we have the sojourn time distribution as:

$$\begin{aligned} H_0(t) &= 1 - e^{-\lambda t} \\ H_k(t) &= 1 - e^{-\lambda t + k\mu t} \\ H_{n+1}(t) &= 1 - e^{-n\mu t} \end{aligned} \quad (6)$$

and H_n is given by:

$$H_n(t) = \begin{cases} 1 - e^{-(n-1)\mu t} & t \leq T_1 \\ 1 - \left(\frac{T_2-t}{T_2-T_1}\right)e^{-(n-1)\mu t} & T_1 \leq t \leq T_2 \\ 1 & t \geq T_2 \end{cases} \quad (7)$$

The mean sojourn time h_i in state i is given by:

$$h_i = \int_0^\infty [1 - H_i(t)] dt. \quad (8)$$

Hence we have:

$$\begin{aligned} h_0 &= \frac{1}{\lambda} \\ h_k &= \frac{1}{\lambda + k\mu}, \quad k = 1, \dots, n-1 \\ h_n &= \frac{e^{-(n-1)T_2\mu} - e^{-(n-1)T_1\mu} + (n-1)(T_2 - T_1)\mu}{(T_2 - T_1)((n-1)\mu)^2} \\ h_{n+1} &= \frac{1}{n\mu}. \end{aligned} \quad (9)$$

Combining, (4) and (9), the SMP steady state probabilities can be obtained from (3). However, to evaluate the performance in terms of the probability of blocking incoming calls P_b , we only need to deal with states n and $n+1$. These are only states in which an incoming call will get blocked. Therefore,

$$\begin{aligned} P_b &= \pi_n + \pi_{n+1} \\ &= \frac{v_n h_n + v_{n+1} h_{n+1}}{\sum_{i=0}^{n+1} v_i h_i}. \end{aligned}$$

Substituting for v_i 's and h_i 's gives,

$$P_b = \frac{\gamma}{\left[\frac{1}{\lambda} + \left(\sum_{k=1}^{n-1} \left(1 + \frac{\rho}{k}\right) \frac{\rho^{k-1}}{(k-1)!} \frac{1}{\lambda + k\mu}\right) + \gamma\right]} \quad (10)$$

where

$$\begin{aligned} \gamma &= \frac{\rho^{n-1}}{(n-1)!} \left[\frac{e^{-(n-1)T_2\mu} - e^{-(n-1)T_1\mu}}{(T_2 - T_1)((n-1)\mu)^2} \right] \\ &+ \frac{\rho^{n-1}}{(n-1)!} \left[\frac{1}{((n-1)\mu)} + \frac{p}{n\mu} \right]. \end{aligned} \quad (11)$$

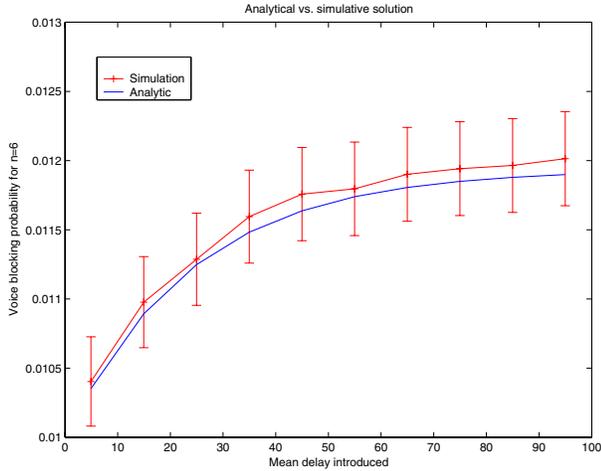


Figure 4: Comparison of Simulative and Analytical Results for SMP

4.2 Solution for Two Delay Slots (E-DOVE)

The state transition diagram is shown in Figure 2. As discussed earlier in section 3, when more than one end call is delayed, the underlying stochastic process is neither an SMP nor an MRGP. As a result, it is not feasible to obtain a closed form solution and we have to resort to simulations to analyze the behavior of such a stochastic process.

4.3 Simulation of DOVE and E-DOVE Models

For the single delay slot case, the stochastic process to be dealt with is an SMP. This SMP was solved analytically to obtain the new voice call blocking probability as given by Eqn. 10. Analytical and simulative solution were obtained and compared to prove consistency of the results. As seen in Figure 4, all analytical values lie within the 95% confidence interval of simulative result. The effect of increasing the mean delay of the single delay slot on the data traffic is shown in Figure 7 and Figure 8 by the curves labeled as “DOVE”. As is to be expected, the system delivers much better data transmission performance by dropping significantly fewer data packets. Simulation of stochastic process generated by state transition diagram shown in Figure 2 has carried out using the CSIM 19 software (See CSIM Website). It is process-oriented discrete event simulator implemented in C/C++ as a library of routines which perform necessary operations to simulate the system behavior. In addition, it also provides functions for calculating confidence intervals by batch means analysis (Banks et al. 2001) and run length control (Law and Kelton 2000) for terminating simulation when desired accuracy has been achieved. The results of this simulation study are shown in Figure 5-8 with respect to voice and data traffic, when up to two voice calls may be delayed. The delay time is assumed to be uniformly distributed $U(0, T)$. The results in Figure 5 show the

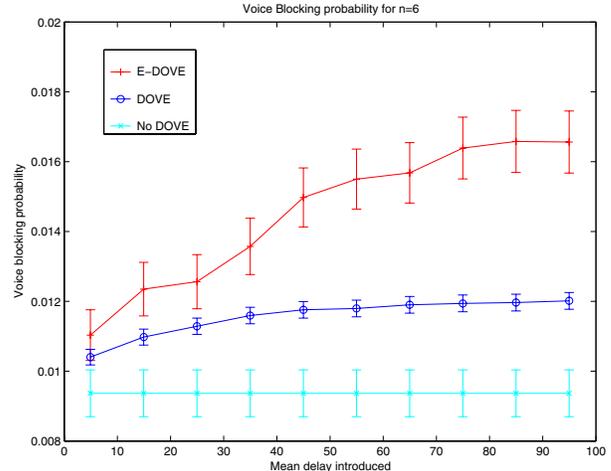


Figure 5: Voice Blocking Probability with $n=6$

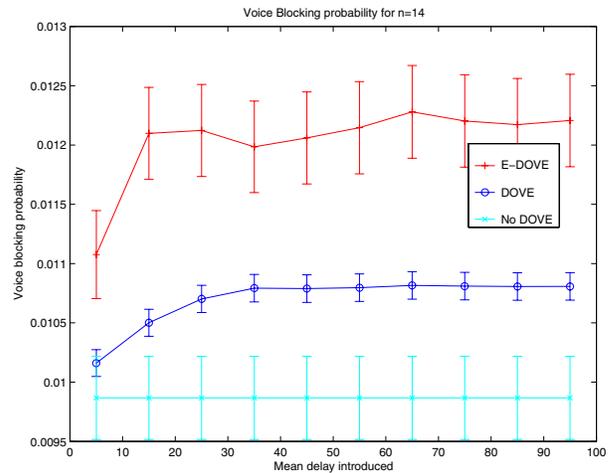


Figure 6: Voice Blocking Probability with $n=14$

effect mean delay time ($=\frac{T}{2}$) on the voice call blocking probability. For these simulation runs the system parameters were set to, $n = 6$, $\lambda^{-1} = 63$ and $\mu^{-1} = 120$. Also for each of plots, 95% confidence interval has been calculated and plotted along with mean values of observations. We used sampling from Bernoulli distribution to calculate the confidence interval for different observations (See Trivedi 2001 for more detail). As the mean delay is increased to accommodate more data traffic, the blocking probability for new incoming voice calls also increases. Figure 6 shows a similar behavior, for $n = 14$, $\lambda^{-1} = 16.33$ and $\mu^{-1} = 120$. The performance with respect to data traffic for the case of two delay slots is shown Figure 7 and Figure 8. As expected, the packet drop rate decreases with two delay slots as compared to single delay slot.

The effect of choosing uniformly distributed delay (as compared to EXP()) delay distribution) in Figures 9 and 10 for delay slots=1 and 2, respectively. The EXP() delay distribution gives better results as compared to the uniform distribution. This is due the fact that for the case of uniform

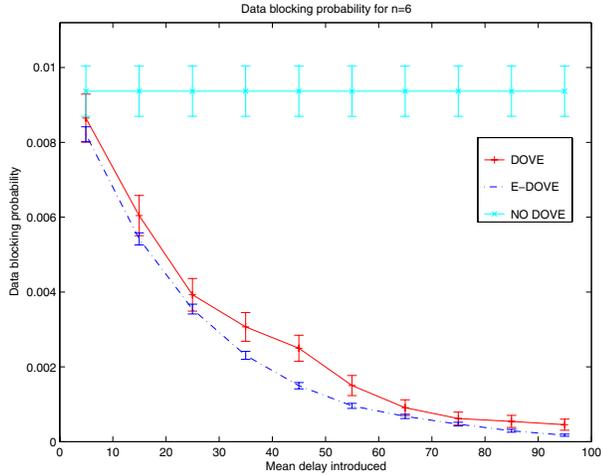


Figure 7: Data Blocking Probability with $n=6$

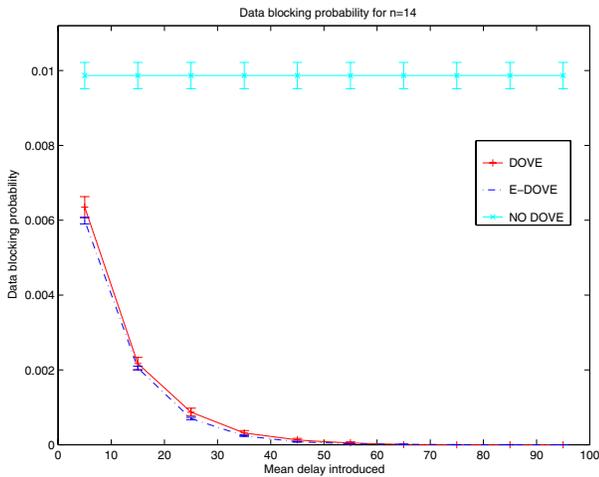


Figure 8: Data Blocking Probability with $n=14$

distribution, it is equally likely to incur large amounts delay as small delays, unlike for the $EXP()$ distribution. However, the important issue is that needs to be brought out is that for the $EXP()$ delay distribution, there is significantly higher probability that a call sitting in a delay slot can incur large amounts of delay, as shown in Figure 11. In contrast, for the uniform delay case, the maximum delay is bound to T . Voice calls sitting in the delay slots for arbitrarily long times is highly undesirable since long delays in voice communications adversely effect the quality of voice communication.

5 CONCLUSIONS

Simultaneous voice and data communication is important issue for 2.5G and 3G networks. To maintain acceptable for both one class of traffic, e.g. voice should not starve the other class of traffic, e.g. data. To ensure equitable distribution of the available bandwidth, solutions such as DOVE and its generalization E-DOVE have been proposed

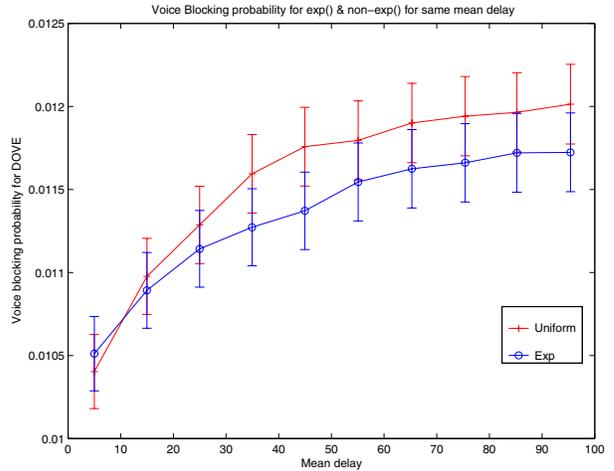


Figure 9: Voice Blocking Probability for Exp. and Uniform Delay with Same Mean for DOVE

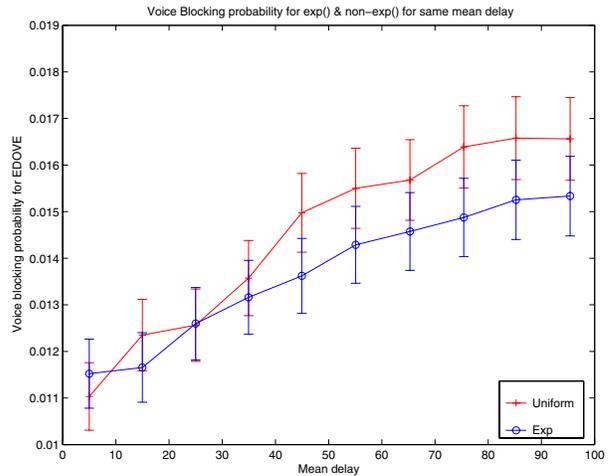


Figure 10: Voice Blocking Probability for Exp. and Uniform Delay with Same Mean for E-DOVE

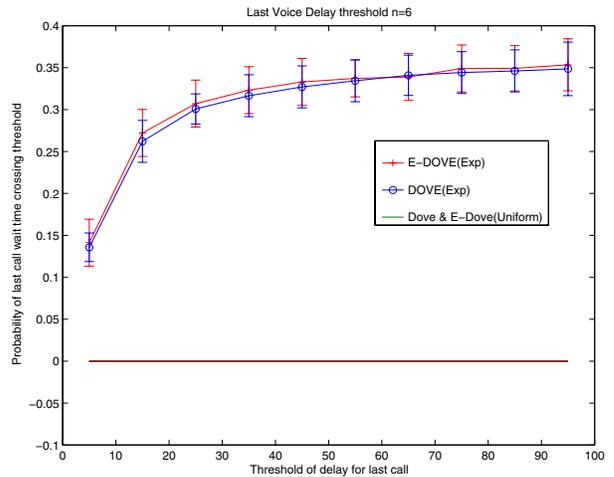


Figure 11: Probability that Last Call Suffer Delay More than Certain Threshold

in the literature. However, by assuming the delay to be $EXP()$ distributed, while resulting in easier analysis for calculating blocking probabilities, ignores the adverse effects of prolonged delays on voice calls. In order to bound the worst case delay to a fixed threshold, this paper suggests the use uniform delay distribution instead. However, using non-exponential distributions results in models that no longer exhibit Markovian properties. Consequently, such models are harder to analyze. This paper goes on to show that for a single slot, the underlying stochastic process turns out to be an SMP for which we can obtain closed form results. However, for more than one delay slot, the process is neither SMP nor MRGP. The behavior of such processes can only be analyzed via discrete event simulations. These studies indicate that while uniform delay distribution may perform poorly in terms of voice call blocking probability, it ensures that a voice call is never subjected to delay more than some fixed threshold value. The $EXP()$ delay distribution on the other hand can cause to be delayed beyond this threshold value with a much higher probability.

ACKNOWLEDGMENTS

This research was supported in part by the Air Force Office of Scientific Research under the MURI grant no. F49620-00-1-0327 and by the DARPA/Army Research Office under the MURI grant no. DAAD19-01-0504.

REFERENCES

- Banks, J., J. S. Carson, B. L. Nelson and D. M. Nicol. 2001. *Discrete-Event System Simulation*, 3rd. ed. Englewood Cliffs, New Jersey: Prentice Hall.
- Casena, M., A. Capone. 2002. Impact of mixed voice and data traffic on the UMTS-FDD performance. In *Global Telecommunications Conference, GLOBECOM '02*, 1: 758–762.
- Cinlar, E. 1975. *Introduction to Stochastic Processes*, Englewood Cliffs, New Jersey: Prentice Hall.
- CSIM 19 Simulator. Access online via <http://www.mesquite.com/> [accessed June 11, 2005].
- Dugan, J. B., K. S. Trivedi, R. M. Geist and V. F. Nicola. 1984. Extended stochastic Petri nets: Applications and analysis. In *Performance '84*, 507–519, Amsterdam, Netherlands.
- Ivanov, K., C. F. Ball and F. Trembl. 2003. GPRS/EDGE performance on reserved and shared packet data channels. In *Vehicular Technology Conference*, 2: 912–916.
- Koutsakis, P., and M. Paterakis. 2001. Integrating voice, video and email data packet traffic over wireless TDMA channels with errors. In *International Journal of Wireless Information Networks*, 8(4): 217–227.
- Kulkarni, V. G. 1995. *Modeling and Analysis of Stochastic Systems*, London: Chapman & Hall.
- Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed., McGraw Hill Higher Education.
- Mahdavi, M., R. M. Edwads and S.R. Cvetkovic. 2001. Policy for Enhancement of Traffic in TDMA Hybrid Switched Integrated Voice/Data Cellular Mobile Communication Systems. In *IEEE Communication Letters*, 5(6):242–244.
- Mitrou, N. M., G. L. Lyberopoulos and A. D. Panagopoulou. 1993. Voice and data integration in the air-interface of a microcellular mobile communication system In *IEEE Transactions on Vehicular Technology*, 42(1):1–13.
- Okuda, T., Y. Ando, T. Ideguchi and X. Tian. 2002. A Simplified Performance Evaluation for Delay of Voice End-User in TDMA Integrated Voice/Data Cellular Mobile Communication System. In *Proceedings of IEEE Globecom 2002*, 891–895, Taipei, Taiwan.
- Trivedi, K. S. 2001. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, 2nd ed., New York: John Wiley and Sons.

AUTHOR BIOGRAPHIES

PAWAN K. CHOUDHARY is graduate student at Duke University in Electrical and Computer Engineering. He did his undergraduate studies at IIT Roorkee, India. Email at pkc4@ee.duke.edu and his home page is <http://www.ee.duke.edu/~pkc4>.

BHARAT B. MADAN Currently heads the Information Sc. & Tech Div. at the ARL, Penn. State Univ. Prior to this, he was a Professor in the CSE Dept. at IIT, Delhi and a visiting scholar at Duke University. He also spent four in the industry R&D at IBM and Ericsson. He can be contacted at bbm2@psu.edu.

KISHOR S. TRIVEDI holds the Hudson Chair in the Department of Electrical and Computer Engineering at Duke University, Durham, NC. He also holds a joint appointment in the Department of Computer Science at Duke. He has been on the Duke faculty since 1975. His e-mail address is kst@ee.duke.edu, and his home page is <http://www.ee.duke.edu/~kst>.