Packet loss probability estimation using Erlang B model in modern VoIP networks

Tibor Mišuth and Ivan Baroňák

Institute of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia
tibor.misuth@stuba.sk, ivan.baronak@stuba.sk

Abstract

The paper presents an approach to packet loss probability estimation in modern VoIP networks. Our proposal is based on simple Erlang B model that is widely used for classic telecommunication networks dimensioning. We analyze its potential applicability to modern convergent IP networks. We introduce a procedure to determine proper values of input variables for original Erlang B model based on characteristics of codec and network link in use. At last we prove the model outcome and applicability with simulation results using NS2 software.

1. Introduction

End of 19th and beginning of 20th century was the period of rapid growth of popularity and fast evolution of telecommunications. Together with constantly rising number of subscribers connected to the networks and therefore rising of number of executed calls also proportionally rose capacity requirements against telecommunication networks. Since the construction costs, particularly of long distance lines, were quite expensive, wrong estimation of expected traffic and overdimensioning of capacity could lead to economic problems related to return of investments. From the other point of view underdimensioning meant loss of potential profit and degradation of customers' satisfaction level as well. Therefore analysis of dependencies of telecommunication traffic and their description became important research field.

Danish mathematician A. K. Erlang focused exactly on problematic of traffic load and its relationship to available capacity of trunk lines. The result of his effort are mainly two models for call loss or call waiting probability calculations also known as Erlang B and Erlang C models or 1st and 2nd Erlangs' equations.

These equations put together offered traffic load, number of available telecommunication lines (or handling servers from the queuing systems perspective) and probability of call not being processed immediately on its arrival. Especially the first Erlang equation covers the problem of dimensioning of sufficient trunk lines capacity.

Last decades of 20th century were marked by expansion of data networks and Internet. This also caused significant changes to telecommunications. One of the most important issues became the question of convergence, sharing of common communication channel for multiple types of applications. Therefore telecommunication systems were being transformed from original circuit switched systems to packet switched systems and voice information was started to being transferred in form of samples or packets. The most widely deployed type of data networks are networks based on IP protocol, that's why voice information transferred in form of samples is also noted as Voice over Internet Protocol (VoIP).

Even though capacity and construction costs of today's packet based networks are considerably different than parameters from early telecommunication era, the problematic of proper capacity estimation is still very important. This paper proposes a possible way of simple application of original Erlangs' ideas adapted to modern convergent telecommunications environment.

Next sections of this paper are structured as follows. First the original Erlang B model together with its modification (extended Erlang B model) are presented. Then some properties of VoIP networks are analyzed in comparison to classic telecommunication networks. Next section deals with adaptation of original model to packet switched environment. Finally the results of simulation with varying input parameters and their comparison to model estimations are presented.

2. Equations

Erlang B model is the basic model which does not contain the waiting queue. Incoming calls are assigned to the idle server / line directly if there is any available, otherwise they are considered blocked or lost [1]. This implies the Erlang B model is widely used to dimension the trunk capacity between Contact center and communication networks. Today, the Voice over IP (VoIP) technology is more and more important, but the basic capacity problem is only slightly modified to available data throughput of the connection. Thus Erlang B model can be used in this case as well.

The Erlang equation uses three basic parameters:

A – the traffic load in Erlangs,

N – number of lines / trunks (requested simultaneous connections),

P_B – probability of call blocking.

The original form of the equation allows us to find the blocking probability if A and N values are known:

$$P_B(N,A)=\frac{\dfrac{A^N}{N!}}{\sum_{i=0}^{N}\dfrac{A^i}{i!}} \qquad (1)$$

If we know the rate of calls per time unit λ and the average number of served requests per the same time unit μ (so the average handling time is $1/\mu$, or average call duration time) then the traffic load can be easily evaluated as [2]

$$A = \frac{\lambda}{\mu} \tag{2}$$

If we substitute A in equation (1) we receive following form:

$$P_B(N, \lambda, \mu) = \frac{\left(\frac{\lambda}{\mu}\right)^N}{N!} \frac{1}{\sum_{i=0}^{N} \left[\left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}\right]} \tag{3}$$

We can see that it is the same formula as obtained from M/M/m/m queuing system [3], [4]. We have just analytically shown that the two mentioned models are identical in fact.

The original Erlang B model can be modified in the way that it will assume also the traffic generated by repeated attempts in the case if the previous attempt was unsuccessful (it was rejected). Then we have extended Erlang B model [5].

The new variable is recall factor r. It denotes the probability that a blocked caller will try to call again immediately. The traffic load towards the telecommunication system consists of two parts: the first-time attempts and repeated calls [6]:

$$A = A_0 + R = A_0 + A.P_B(A, N).r \tag{4}$$

where A_0 is the traffic load generated by first-time attempts (in Erlang), R represents the load of repeated calls and P_B is the blocking probability using the original Erlang B model (1). Equation (4) can be easily transformed to the form:

$$A(1 - P_B(A, N).r) = A_0 \tag{5}$$

By using iterative numerical methods we can then find out the original traffic load A_0 that the communication system can serve.

3. Model Adaptation for VoIP Environment

3.1. Differences between PSTN and VoIP traffic

As was stated in the previous section, original Erlang B model (formula) was defined for dimensioning of classical telecommunication trunks that interconnect e.g. two telecommunication switching centres. In such situation the trunk is considered as queuing system with telephone calls as the requests and set of N parallel lines (Fig. 1)in the trunk as handling nodes. It is obvious the system cannot have a waiting queue, so calls that arrive during the period of occupation of all lines cannot be put through and are therefore lost.

Requests arrival rate λ is simply defined as number of calls per time unit, whereas average request handling time $1/\mu$ is the average call duration. As each line of trunk can transfer only one call at a time, the average call duration is equal to average line utilization time by one request.

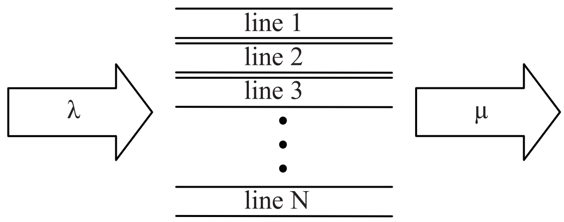

Fig. 1. Classic telecommunications trunk of N lines

To calculate the probability of call loss (blocking) the basic formula (1) can be used.

The situation is slightly different in VoIP environment. The same link between two neighbouring nodes of data network is shared among multiple data streams [7]. Thus the basic characteristic of connection is not the number of lines, but the throughput of the link, i.e. amount of data transferred per time unit. Data are transferred in form of packets of various lengths for various applications.

In this paper we focus on VoIP data only, thus we can specify more details of the data streams. Basically the principle of VoIP is to convert analog voice signal to binary representation, transfer the data in form of packets to receiving node and backward conversion to analogue form [8]. Quality of voice reproduction depends on selected codec used for voice encoding / decoding, packet loss during the transmission, total cumulative delay during the processing and transfer and several other factors. Each codec defines the set of rules for voice packetization / depacketization, sample size, packet size, number of voice samples in one packet, packetization interval, etc.

Audio streams in VoIP can be divided into two separate groups depending on time variability of generated data traffic. CBR (constant bit-rate) sources generate one packet of predefined size (defined by used codec) per one packetization interval (both defined by codec in use) while VBR (variabile bit-rate) sources can adjust amount of information in each packet to input signal complexity thus utilizing advanced features of particular codec. In this paper we analyze CBR audio stream generated by G.711 audio codec, however the results can be generalized on any CBR stream.

Based on the abovementioned codec's parameters (frame (packet) size d and packetization interval τ) the required bandwidth per one VoIP connection utilizing particular codec can be derived as

$$w = \frac{d}{\tau} \quad [bit/s] \tag{6}$$

The link is shared by multiple connections utilizing a kind of time multiplex approach, thus idea of virtual telecommunication lines can be applied (Fig. 2)

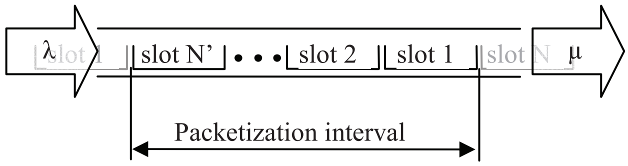

Fig. 2. VoIP channel with virtual lines

3.2. Abstraction

The original Erlang B model provides a method to estimate call loss at defined traffic load A and set of lines in trunk N. For classic telecommunication networks it means only the calls above the capacity of the trunk are lost, but all other calls intact in terms of their quality. The situation is slightly different in IP world, where the link is shared. If the current transfer demands are higher than link capacity, frames are queued or lost, if the buffers are occupied. However, real time communication can tolerate only minimum level of delay. That means, if the queue is too long, packets can be discarded as their contingent transfer to the destination would occur to late and decoder would not be able to use the data. On the other hand, some level of packet loss can be accepted in VoIP communications without significant impact on communication quality since codecs in use have features to restore information of lost packet to certain degree.

If we abstract from transport technology in use, there is no difference between classic and VoIP connections from number of parallel connections in use over the trunk / link. We assume the Erlang B model can be used to estimate call loss probability in either cases. However for VoIP traffic depending on packet arrivals from sources, call loss does not necessary mean loss of all packets of the particular stream, but packets are lost randomly from all streams instead. If certain degree of lost packets is accepted, proportional increase of handled traffic load is obtainable. Therefore Erlang B model can be used to estimate the packet loss once traffic load A and number of virtual links (link capacity) N is known.

As described above, the data channel (link between nodes) is characterized by its capacity (link speed) W. Then theoretical link capacity expressed in terms of number of parallel connections the link can carry through can be calculated as

$$N' = \left\lfloor \frac{W}{w} \right\rfloor \tag{7}$$

The traffic load A remains the same value as for classic telecommunication networks based on average call arrival rate from sources and average line occupation time (average call duration), thus

$$A' = \frac{\lambda}{\mu} \quad [Erl] \tag{8}$$

At this point, all input values to calculate probability of call loss or probability of packet loss P_B (1) are known. Since we rounded N towards minus infinity (took floor) of the fraction, the resulting value of P_B gives us only upper theoretical bound for call loss, thus the eliminated part of capacity can positively influence the overall behaviour and for observed packet loss probability P_B following can be stated

$$P_B'(N', A') < \frac{\frac{A'^{N'}}{N'!}}{\sum_{i=0}^{N'} \frac{A'^{i}}{i!}} \tag{9}$$

Furthermore the packet loss probability can be significantly influenced by buffer size that is associated to the particular data link. This phenomenon is discussed later.

4. Simulation model and results

We decided to validate our ideas using simulation created in software Network Simulator 2. The network topology was very simple (Fig. 3) consisting of two nodes and a link between them.

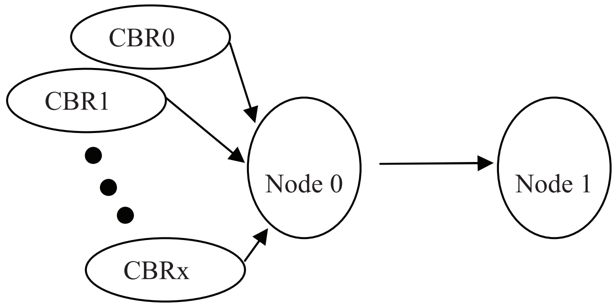

Fig. 3. Network topology

The first node worked as a source and second one was configured as a sink. The data transmission was done in one direction only, however this does not have any negative effect on results, since the link is configured as full duplex and both directions of traffic are isolated in nodes. Therefore simulation of only one direction is sufficient.

Each call was simulated as separate CBR traffic source attached to node 0. The start time of transmission of a particular source was determined using Erlang distribution with parameter λ (the transformation to exponential distribution for interarrival time was used) together with call duration time ($1/\mu$), after which the source was deactivated and stopped to send traffic. This fulfils the requirements for requests arrival distribution and average request handling time defined by Erlang B model.

We decided to simulate G.711 codec as the basic option for all VoIP devices. The codec characteristics are following [9]

Table 1. G.711 characteristics

Sampling frequency	8 kHz
Sample size	8 bits
Packetization interval	20 ms
Number of samples per packet	160
Packet payload size	160 B
Packet header size	58 B
Nett bitrate per call (payload only)	64 kbps
Gross bitrate per call (including headers)	87.2 kbps

We used following protocol stack: L2 802.3 (header + trailer length18 bytes), L3 IP (header length 20 bytes), L4 UDP + RTP (header lengths 8 +12 bytes), so total header size was 58 bytes, that is 26.6% of total frame length.

To evaluate the influence of average call duration on packet loss probability according to (8) we decided to execute two simulation rounds with different value of this parameter. The first run was with average call duration set to 30 seconds, the second one with value 60 seconds.

All simulation were executed for 5 times for each configuration with simulated 200 000 seconds. The link was configured as full-duplex with link speed 959.2 kbit/s, thus able to accommodate 11 parallel connections (7) using G.711 parameters.

Table 2. 1st round simulation results (avg. call duration 30 seconds)

Call arrival rate λ [calls/sec]	Traffic load A [Erl]	Calculated packet loss P_B' [%]	Measured packet loss P_B' [%]
0.02	0.6	0.00	0.00
0.04	1.2	0.00	0.00
0.06	1.8	0.00	0.00
0.08	2.4	0.00	0.00
0.1	3	0.02	0.01
0.12	3.6	0.09	0.05
0.14	4.2	0.27	0.16
0.16	4.8	0.65	0.45
0.18	5.4	1.30	0.89
0.2	6	2.30	1.70
0.22	6.6	3.66	2.82
0.24	7.2	5.38	4.16
0.26	7.8	7.40	6.25
0.28	8.4	9.66	8.02
0.3	9	12.08	10.36
0.32	9.6	14.61	12.65
0.34	10.2	17.18	15.10
0.36	10.8	19.76	17.82
0.38	11.4	22.30	20.47
0.4	12	24.78	23.01
0.42	12.6	27.18	25.48
0.44	13.2	29.50	27.83
0.46	13.8	31.72	30.15
0.48	14.4	33.85	32.22
0.5	15	35.88	34.45
0.52	15.6	37.82	36.57
0.54	16.2	39.66	38.39
0.56	16.8	41.41	40.28
0.58	17.4	43.09	41.97

Table I shows the results of the first simulation round with average call duration 30 seconds, whereas Table II shows the results of the second simulation round with average call duration 60 seconds. For better comparison the values are plotted to charts(Fig. 4 and Fig. 5)

Table 3. 2nd round simulation results (avg. call duration 60 seconds)

Call arrival rate λ [calls/sec]	Traffic load A [Erl]	Calculated packet loss P_B' [%]	Measured packet loss P_B' [%]
0.01	0.6	0.00	0.00
0.02	1.2	0.00	0.00
0.03	1.8	0.00	0.00
0.04	2.4	0.00	0.00
0.05	3	0.02	0.02
0.06	3.6	0.09	0.05
0.07	4.2	0.27	0.19
0.08	4.8	0.65	0.42
0.09	5.4	1.30	0.88
0.1	6	2.30	1.73
0.11	6.6	3.66	2.88
0.12	7.2	5.38	4.12
0.13	7.8	7.40	5.83
0.14	8.4	9.66	8.27
0.15	9	12.08	10.26
0.16	9.6	14.61	12.72
0.17	10.2	17.18	15.08
0.18	10.8	19.76	17.96
0.19	11.4	22.30	20.40
0.2	12	24.78	22.96
0.21	12.6	27.18	25.45
0.22	13.2	29.50	27.67
0.23	13.8	31.72	30.18
0.24	14.4	33.85	32.33
0.25	15	35.88	34.52
0.26	15.6	37.82	36.33
0.27	16.2	39.66	38.35
0.28	16.8	41.41	40.02

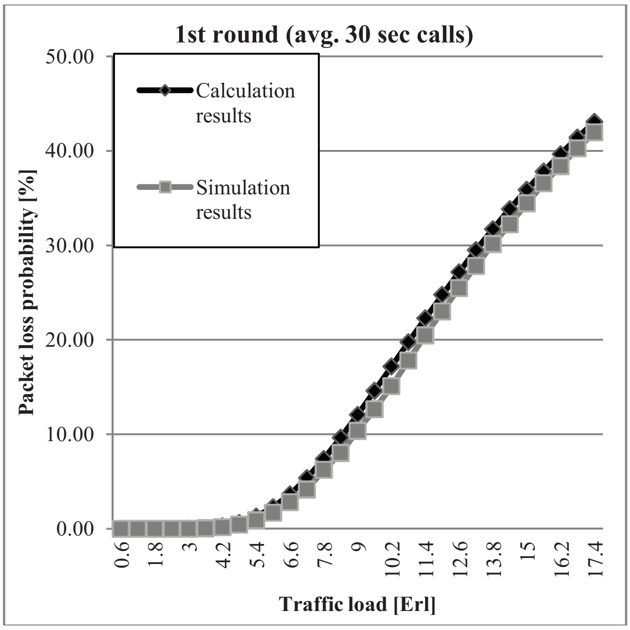

Fig. 4. Comparison of calculation and simulation results for 1st round

As expected, the calculation results provide the upper bound for packet loss probability, and for all cases the measured values remain below this bound. The difference is caused by utilization of link buffer where packets are stored until the previous packet was sent. This element smoothes the peaks of traffic thus eliminate some of the possible packet losses.

The profile of the second round remains the same as for the first case. Direct comparison between corresponding measured values for the first as second rounds does not show any difference. This adheres to the expectations that only value of traffic load A influences the packet loss probability, but not its components.

The simulation confirms validity of our ideas. The only remark concerns influence of buffer size on the results. As stated above, the buffer smoothes the traffic profile and prevent some packets from being lost. The size of the buffer influences the smoothening process, where larger buffer means more smoothening and less packet loss. However with increasing buffer size the overall transport delay of data packets increases as well that can lead to quality degradation caused by late packet delivery, thus reducing effective capacity of the link.

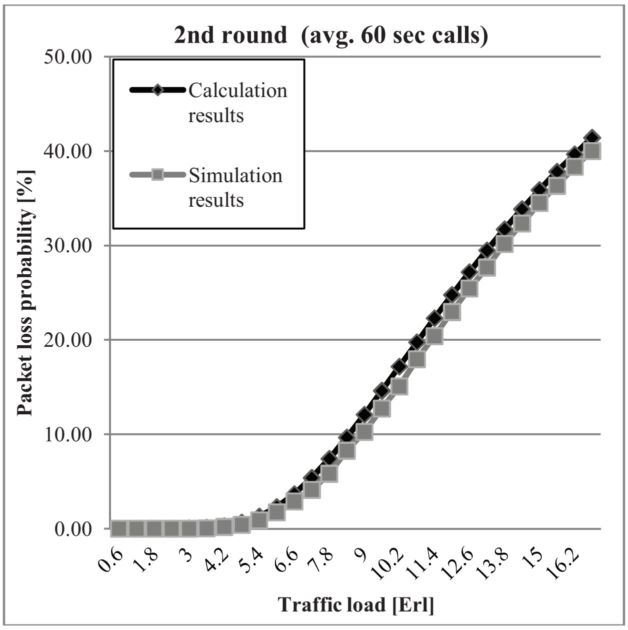

Fig. 5. Comparison of calculation and simulation results for 2nd round

5. Conclusion

The paper focuses on possible application of simple Erlang B model (1st Erlang's equation) for packet loss estimation in modern VoIP networks. An innovative approach to calculation of input variables' values for Erlang B model is presented. The calculation is based on abstraction of IP infrastructure characteristics to obtain traffic load in Erlangs and number of parallel lines. These values are then entered to original formula to obtain packet loss probability value.

We verified the expectations using simulation. We discovered that utilization of buffer in network nodes can influence the measured packet loss probability in positive manner, meaning that the resulting packet loss is lower than estimated. Therefore the Erlang B model can be used to determine the upper bound of packet loss probability in theoretical worst case scenario when no buffer is available.

6. Acknowledgement

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Institute of Telecommunications, within the scope of the projects VEGA No. 1/0565/09 "Modeling of traffic parameters in NGN telecommunication networks and services" and OPVaV - 2008/4.1/02-SORO "Support for building of Centre of Excellence for SMART technologies, systems and services" (ITMS 2624020029).

7. References

[1] Diagnostic Strategies. Traffic Modeling and Resource Allocation in Call Centers, Needham (Mass., USA) : Diagnostic Strategies, 2003.Available: http://www.fer.hr/_download/repository/A4_1Traffic_Modeling.pdf

[2] Koole, G. Call Center Mathematics: A scientific method for understanding and improving contact centers, Amsterdam, Vrije Universiteit, Available: http://www.math.vu.nl/~koole/ccmath

[3] Unčovský, L. Stochastic models of operational analysis, 1st edition, Bratislava, ALFA, 1980.

[4] Polec, J., Karlubíková, T. Stochastic models in telecommunications 1, 1st edition, Bratislava, FABER, 1999.

[5] The Extended Erlang B Traffic Model [Online], Westbay Engineers Ltd., 2008, Available: http://www.erlang.com/calculatormanual/index.html?the_extended_erlang_b_traffic_model.htm

[6] Yokota, M. Erlang, [Online], Available: http://hp.vector.co.jp/authors/VA002244/erlang.htm.

[7] Chromý, E., Kavacký, M. Asynchronous Networks and Erlang Formulas. In: International Journal of Communication Networks and Information Security. ISSN 2073-607X. Vol. 2, No. 2 (2010), p. 85-89

[8] Kľúčik, S., Tisovský, A. Queuing Systems in Multimedia Networks. In: Elektrorevue. ISSN 1213-1539. Roč. 15, 16.11.2010 (2010), art. no 99.

[9] Schulzrinne, H., RFC 3551, RTP Profile for Audio and Video Conferences with Minimal Control [Online], s.l., Columbia University, 2003. Available: http://tools.ietf.org/html/rfc3551.