Enabling Voice Services in Hybrid/Cross Network Framework

N. Kryvinska† and Do van Thanh‡

† University of Vienna, ‡ Norwegian University of Science and Technology

Abstract: The Internet and Intelligent Network (IN) are converging. The convergence is motivated by the fact of a global telecommunication infrastructure development. The converged communication network has to deal with a wide spectrum of traffic characteristics. And, it has to support applications with wide range of expectations and requirements. In this paper, we propose queuing system models for the analysis of the converged architecture that span two INs in different regions through the IP network. We build here service scenario framework. Also, we explore the issues related to this model and propose a platform, as well as a set of components, to facilitate its creation and deployment. And, we illustrate the strength of our model by discussing and evaluation of the mathematical analysis results.

1. Introduction

The computer communications networks, particularly the Internet, have been designed independently of PSTN, since their initial development over thirty years ago. These networks have sometimes interacted, for example, in the use of PSTN to carry Internet traffic, but attempts have just recently begun to really converge services onto single integrated platform. The convergence attempts have been motivated by the fact that one of the key requirements in deploying a pervasive and ubiquitous information superhighway is the development of global integrated telecommunications infrastructure [1, 2].

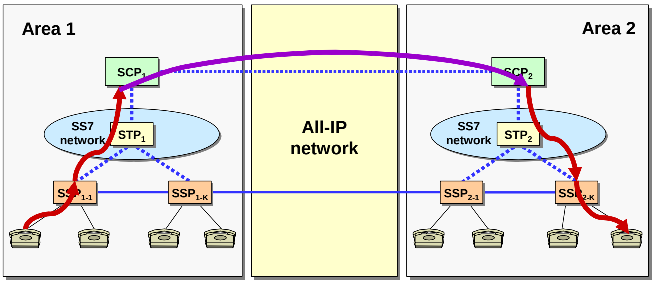

Figure 1. Connection set-up request between users of different INs

This infrastructure is built onto hybrid sub-networks. Where in reality can exist such hybrid sub-networks? For example, two small regions still have circuit-switched infrastructures (e.g., INs). They are removed from each other. And, between them lying big, high-developed region, into which network operators has already deployed high-speed packet-based infrastructure. We can find a lot of examples like this nowadays, where in high-developed countries existing since several years packet-based (e.g., all IP) networks and in less developed countries still PSTN only infrastructures. The problem become significant if a user from one small region wants to reach a user from other small remote region, he/she has to pass through that big region with packet-based backbone network. Another example is the case of operators that offer cheap international phone calls on phone cards. To avoid international tariff, they use IP network to connect two PSTN networks and transfer voice traffic in IP packets. The lack of careful dimensioning results often to quite poor performance. The task of network operators, as well as service providers is to provide the users with good QoS.

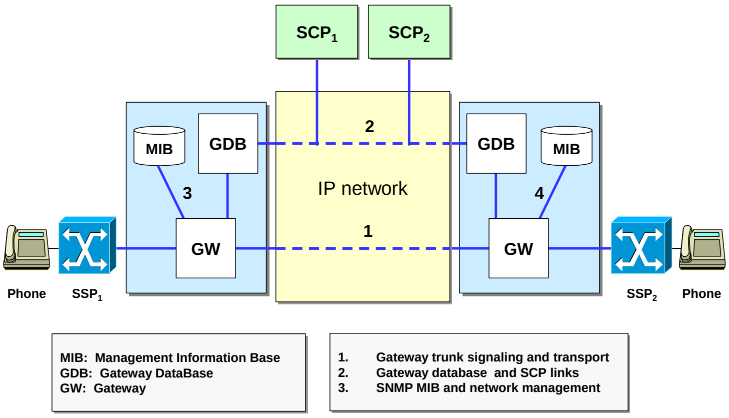

Figure 2. Functional framework for hybrid architecture

Concerning above, we build a service scenario framework (Figure 1), where the same protocols are used at the interface of each terminal, but a different protocol is used between them. In this scenario, both media mapping, media control channels, and mapping between signalling protocols are generally required. However, mapping between signalling protocols can be avoided in some configurations. In particular, when core network is all-IP, all IN/PSTN signalling information can be transferred transparently through the IP network [1, 2]. The typical functional framework [2, 3] for two different IN, when the IP network used as a backbone network, is shown in Figure 2.

2. Hybrid/Cross architecture and modelling tools

For the hybrid architecture modelling we are applying two queuing systems model using method of stages. The scenario in Figure 1 presents two separated Intelligent Networks that have connections through packet-based network, e.g. the Internet. In case, when a user from the first network (area 1) wants to get IN service from the second network (area 2), he/she has to go through two stages: first stage – call processing in SCP_1; and, second stage – a request from SCP_1 has to be processed in SCP_2.

For the first stage we are applying $M/E_2/1$ queuing system. The call processing in Intelligent Network is going through the several stages. And, in the typical SCP, it goes through the two main stages: receiving and processing a request for IN service from SSP by SCP-server, and processing of this request in SCP-database. However, a request for IN service has to be considered as a single, composite. It means, from point of view of teletraffic theory, that service processing has to be presented as a two-stage Erlangian queuing system. And, for the second (e.g., area 2 Intelligent Network), we have selected $H_2/E_2/1$ queuing system. In this system arrival processes have two-stage hyper-exponential distribution (Figure 3). In the second area IN, requests are coming from two sources (e.g., from SCP_1 and $SSP_{1\text{-}K}$, its own users). So, these requests are coming onto the first stage (λ_2) or onto the second (λ_1). From one of these two stages they have taken for the serving in SCP_2, which also is presented same like SCP_1 as Erlangian system (E_2) [6, 11].

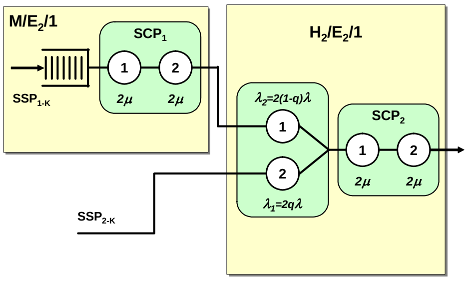

Figure 3. Queuing systems model for hybrid architecture design

3. First stage modelling by $M/E_2/1$ system.

In the $M/E_k/1$ model, customers arrive as a Poisson process, but are served by a single server, which takes a *k*-stage Erlangian distributed amount of time (Figure 4). Since the Erlang distribution is not memoryless for $k > 1$, the description of the state of queue results in a non-Markovian process.

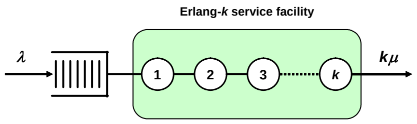

Figure 4. $M/E_k/1$ queuing system

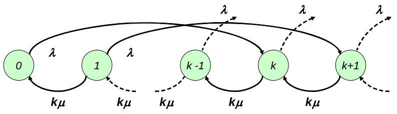

Figure 5. State transition diagram for the M/Ek/1 system

But, the Erlangian distribution is the sum of identical and independent exponential distributions which are each memoryless. So if we change the concept of state, we can describe a Markov chain. The state transition diagram is shown in Figure 5 [4, 5]. The total time that a customer spends in the service facility is the sum of *k* independent identically distributed random variables [6]. Next, the Erlang-k distribution has a coefficient of variation squared ranging from $C_S^2 = 1$, when it is an exponential distribution, to $C_S^2 = 0$ when it is a constant distribution. The coefficient of variation is chosen by selecting the value of *k*, since - formulae (1).

$$C_S^2 = \frac{1}{k} \qquad (1) \qquad\qquad k \le \frac{1}{C_S^2} \qquad (2)$$

The minimum value for *k* is 1, which gives the exponential distribution. When *k* increases, the distribution becomes more symmetrical and also more tightly concentrated around the mean. For very large *k* the distribution is effectively a constant. If C_S^2, the squared coefficient of variation of service time is given, then *k* should be chosen to be the largest value of *k*, formulae (2).

The average waiting time is derived from the M/G/1 formulae and gives the following result - formulae (3). The same actions are taken for the variance of waiting time, by putting the moments of the Erlang-k distribution into the appropriate M/GX/1 formula, we get the following expression for the standard deviation of waiting time - formulae (4).

$$T_w = \frac{\rho T_S}{2(1-\rho)}\left(1+\frac{1}{k}\right) \qquad (3) \qquad\qquad \sigma_{T_W} = \sqrt{\frac{T_S^2\rho(k+1)}{12(1-\rho)^2k^2}\left(4(k+2)-\rho(k+5)\right)} \qquad (4)$$

$$T = \frac{T_S}{1-\rho}\left(1-\frac{\rho}{2}\left(1-\frac{1}{k}\right)\right) \qquad (5) \qquad\qquad \sigma_T = \sqrt{\frac{T_S^2}{(1-\rho)^2}\left[\left(1-\frac{\rho(4-\rho)}{6}\left(1-\frac{1}{k}\right)\right)\left(1+\frac{1}{k}\right)-\left(1-\frac{\rho}{2}\left(1-\frac{1}{k}\right)\right)^2\right]} \qquad (6)$$

The formula for the average time in system for $M/E_2/1$ is also taken from the M/G/1 formula. If we substitute C_S^2 = 1/*k* and rearrange slightly, then we get the result specific to $M/E_k/1$ - formulae (5). Also, for the variance of the time in system, we know the moments of the Erlang-k distribution, and we can put these into the appropriate M/GX/1 results to obtain a formula for the variance, which are - formulae (6).

4. $H_2/E_2/1$ queuing model as a particular case of GI/G/1

For the second area, it is taken into consideration a queuing system that has the hyperexponential interarrival times (Figure 3) [7, 8]. For the $H_2/E_2/1$ queuing system, there is a single server with FIFO queuing discipline. The number of requests for IN services is unlimited. Requests arrive at an average rate λ, so that the average inter-arrival time is $T_A=1/\lambda$. The distribution of inter-arrival time is hyperexponential, with variance σ_A^2, and coefficient of variation squared for inter-arrival time is C_A^2. Service times have an Erlangian distribution, with average value T_S and coefficient of variation squared of C_S^2 [8].

4.1 GI/G/1 approximate formulae of waiting time for $H_2/E_2/1$ queuing system

In the hybrid architecture communication queuing problems can be represented by queuing systems of the type GI/G/1 (general input and general service process, single server). The mean waiting time and the probability of waiting are of the particular interest for system analysis or design. The solutions for different traffic assumptions require the numerical evaluation of roots of transcendental equations by the aid of computers. But often these solutions are not straightforward and require a lot of evaluation work. Therefore, the urgent need is to support network planners and traffic engineer with simple explicit but general approximation formulae for the mean waiting time (Figure 6).

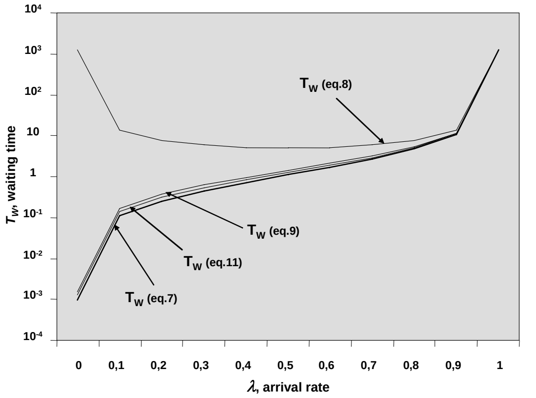

Figure 6. Average waiting time in $H_2/E_2/1$ system under different traffic conditions

$$T_W = \frac{\rho T_S}{2(1-\rho)}\left(C_A^2+C_S^2\right)\begin{cases} e^{-\frac{2(1-\rho)}{3\rho}\frac{\left(1-C_A^2\right)^2}{C_A^2+C_S^2}} & C_A^2 \le 1 \\ e^{-(1-\rho)\frac{C_A^2-1}{C_A^2+4C_S^2}} & C_A^2 \ge 1 \end{cases} \qquad (7)$$

$$T_W \approx \frac{C_A^2+\lambda^2C_S^2T_S^2}{2\lambda(1-\rho)} \qquad (8)$$

$$T_W \approx \frac{\left(1+C_S^2\right)}{\left(1/\rho^2\right)+C_S^2}\frac{C_A^2+\lambda^2C_S^2T_S^2}{2\lambda\left(1-\rho\right)} \qquad (9)$$

$$T_W \ge \frac{\rho^2C_S^2+\rho\left(\rho-2\right)}{2\lambda(1-\rho)} \qquad (10)$$

$$T_W \approx \frac{\rho T_S}{1-\rho}\left(\frac{C_A^2+C_S^2}{2}\right) \qquad (11)$$

The approximation formulae for the mean waiting time T_W in a GI/G/1 system is as following - formulae (7) [10]. Under the heavy load conditions, waiting time has an approximately exponential distribution, with mean waiting time given by formulae (8). This approximation improves as ρ gets closer and closer to 1. For low utilizations the approximation is extremely poor. The Marchal modification to the heavy-traffic approximation gives formulae (9). The Marchal, also Kleinrock vol.II –1976, lower bound on average waiting time approximate formulae is [9] – formulae (10). This bound will be non-negative only for service time coefficients of variation that satisfy $C_S^2 \ge (2-\rho)/\rho$. In our case (e.g., $H_2/E_2/1$ system, when $C_S^2 = 0.5$), it will always take negative values [8]. The formula (11) is actually a special case of the Allen-Cunneen formula for G/G/m [9].

In Figures 7, 8 we are giving comparison of the summarized time in system (e.g., in $M/E_2/1 + H_2/E_2/1$) without and with taken into consideration waiting probability p_w in $H_2/E_2/1$ system. For the second case we are doing backward calculations of $M/E_2/1$ time in system. We have λ and p_w or *q* for $H_2/E_2/1$, and, from that, we can calculate λ_2, which brings us to $T_{M/E2/1}$ ($T_{M/E2/1} = 1/\lambda_2$). From Figure 8, we can see that $T_{M/E2/1}$ and $T_{(\Sigma)}$ have form as under the heavy load conditions. It is looking logical because requests for call processing in SCP_2 are coming from two sources, e.g. from its own SSPs and from SCP_1. Of course, an arrival rate from $SCP_{2\text{-}K}$ (λ_1) will be higher than λ_2, so requests from SCP_1 will have to wait for processing longer time. And, $T_{(\Sigma)}$, because of that, is under the heavy load conditions.

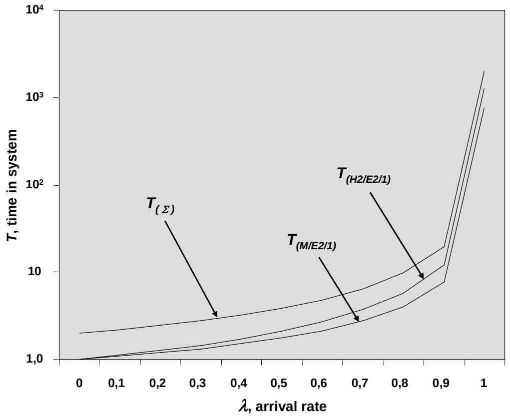

Figure 7. Time in system for the $M/E_2/1 + H_2/E_2/1$ without p_w in $H_2/E_2/1$ system

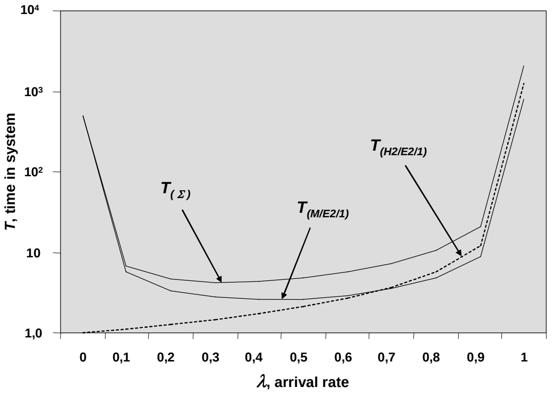

Figure 8. Time in system for the $M/E_2/1 + H_2/E_2/1$ with p_w in $H_2/E_2/1$ system

According to the definitions set up in the introduction, we present mathematical analysis results by normalized values of waiting time and time in system. We take into consideration main balance equation for the system stability, which strictly define that λ/μ always < 1. Also, we can change simultaneously one only parameter, e.g., λ (i.e., arrival process) or μ (i.e., service process). And, because of that, we set up: $\mu = 1/T_S = 1$, and change λ on the allowed interval (i.e., $0 \div 1$).

5. Conclusions

After the examination of efforts done previously in this field, we have discovered that there is a lack of studies in analyzing of network performance on 6, 7 levels of OSI model. It was a motivation to find mathematical models describing processes on these levels for services support in IN and IP networks. Of course, to show a precise picture of each network functioning separately, we have to present every small unit as a queuing model, which is going beyond of scope of this article. The model presented here has to be considered as a symbolic umbrella model for the hybrid network. It can appear as a too abstractive. But, only from 7, 6 levels of OSI the whole service scenario, supported by any network, is visible. And, we try to implement queuing models onto these levels. The importance of this study is growing with diverse of service types grow. The queuing theory techniques allow defining analytical models for applications. As result, the precise classification of existing service types as well as future ones can be done. This can greatly short time for new services modelling, simulation and implementation.

References

[1] Technical Paper, Voice Service Interworking for PSTN and IP Networks, 3ComCorporation, 2000.

[2] M. Hamdi, O. Verscheure, I. Dalgic, J.P. Hubaux, P. Wang, "Voice Service Interworking for PSTN and IP Networks", IEEE Communications Magazine, Vol. 35, No. 5, May 1999, pp.104-111.

[3] J. P. Hubaux, C. Gbaguidi, S. Koppenhoefer, J. Y. Le Boudec, "The Impact of the Internet on Telecommunication Architectures", Computer Networks and ISDN Systems, Special Issue on Internet Telephony, February 1999, Vol. 31 No. 3.

[4] M.K. Molloy, Fundamentals of performance modeling, ISBN 0-02-381910-3 1988.

[5] R.B. Cooper, Introduction to Queueing Theory – 3rd edition, ISBN 0-941893-03-0, CEEPress Books, 1990.

[6] L. Kleinrock, Queueing Systems. Vol. I: Computer Theory, ISBN 0-471-49110-1, Wiley-Interscience publication, 1975.

[7] J.A. White, J.W. Schmidt, G.K. Benett, Analysis of Queueing Systems, ISBN 0-12-746950-8, Academic Press, 1975.

[8] L. Kleinrock, and R. Gail, Solution Manual for Queueing Systems. Volume II: Computer Applications, ISBN 0-942948-01-7, Technology Transfer Institute, 1986.

[9] M. Tanner, Practical Queueing Anylysis, ISBN 0-07-709078-0, IBM McGraw-Hill Series, 1995.

[10] W. Krämer and M. Lagenbach-Belz, "Approximate Formulae for the Delay in the Queueing System GI/G/1", Proceedings of the 8th International Teletraffic Congress (ITC 8), Melbourne, Australia, 1976.

[11] K.S. Vastola, "Performance modeling and Analysis of Computer Communication Networks", Electrical Computer and Systems Engineering Dept., Rensselaer Polytechnic Institute Troy, NY.