Queuing System Models for Performance Analysis of Intelligent Networks

Natalia Kryvinska {Natalia.Kryvinska@tuwien.ac.at}, Harmen R. van As {Harmen.R.van-As@tuwien.ac.at}
Institute of Communication Networks, Vienna University of Technology, Vienna, Austria

Abstract: With the wide deployment of intelligent network (IN) services, there is an urgent need to understand and solve teletraffic performance issues of the evolving network intelligence platform. This paper discusses a queuing system model for the performance analysis of IN call processing. The intelligent network is presented as a network of queues where the total number of customers (e.g., SSPs) is fixed, thus forming a closed queuing network. The IN distributed architecture is modeled as a finite source queuing model – M/M/1/K/K. The expected response time for that model is analyzed and computed. The numerical results and the corresponding curves are provided. And, related to open questions, future work is summarized.

1 Introduction.

With increasing deployment of intelligent network services, design and engineering of network intelligence platforms to accommodate the ever-changing and growing demands of customers, presents a rich market of opportunities and challenges, although tempered by concerns arising from the problematic experiences of similar system and network developments. As the telecommunications industry evolves, customers are increasingly coming expecting instantaneous access to service providers, together with transparency to network failures. System performance dictates that response times need to be minimized, sufficient redundant capacity to be installed in case of failure and controls embedded within the design to manage the exceptional situations (such as media stimulated events) that continually threaten network integrity. The service scenario mixes service demand, physical network topology, signaling message flows, the mapping of functional entities to physical components, and routing as part of the network design process to ensure that performance requirements are met [1].

For the system performance aspect, the most significant changes due to intelligent network (IN) are the distribution of network intelligence and the new services made possible by this distributed architecture. Whereas traditionally a call is processed within the switch, in the IN environment, a call involves the cooperative processing of several network elements connected by a signaling network. This fundamental change possesses some new challenges to teletraffic experts to ensure that IN networks are designed to provide customers' services with good performance [2].

2. Configuration of the IN architecture

The schematic representation of the IN structure is presented in Figure 1.

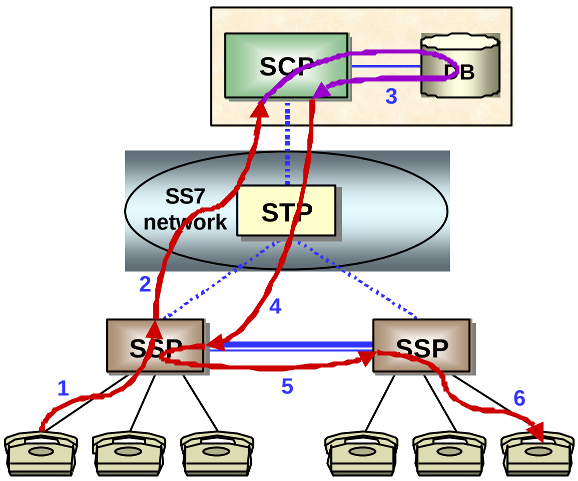

Figure 1. The IN architecture

As shown in the figure, the IN consists of service switching points (SSP) that accommodate terminals, service control point (SCP) that store databases (DB) and perform advanced service control, and the signaling system No.7 (SS7) network that transfers messages between SCPs and SSPs via signal transfer points (STP).

An IN service typically involves interactions between the SSPs and some of the other IN nodes to perform call and connection control, interact with the user, and monitor events on signaling interfaces. For example, the SSP may temporarily suspend call processing to send a query to the SCP where SCP-based services, such as user authentication, number translation, route selection, alternate billing, etc., are executed. The SSP may pass control to the intelligent peripheral (IP) to get more information from the user [2; 3].

3. Closed Queuing Network as a model for the Intelligent Network

The IN can be presented as a network of queues where the total number of customers (e.g., SSPs) is fixed because no customers are allowed to arrive or depart. Networks of this type are called closed. Closed networks can be analyzed using Markov chains. And, the steady-state occupancy distribution has a product form under assumptions similar to those used for open networks.

It is taken into the consideration a network of ***K*** queues, where:

- There are no external sources for customers and there no external destinations for customers;
- There are ***K*** **<** ¥ customers in the network;
- There is a single server at each queue. The service times at the ***i***-th queue are independent exponential random variables with rate m_i. Service times at different queues are independent;
- The routing through the network is random. The probability of a customer leaving queue ***i*** for queue ***j*** is $\boldsymbol{r_{i,j}}$.

This network is called a *Jackson closed network*. The state of a Jackson closed network is described by

$$\underline{n} = (n_1, n_2, \ldots, n_i, \ldots, n_K), \tag{1}$$

where $\boldsymbol{n_i}$ is the number of customers at queue ***i***. The steady-state probability of being in state $\underline{\boldsymbol{n}}$ is once again denoted by $\boldsymbol{p(\underline{n})}$.

The global balance equation of the Jackson closed network can be set up according to the balance between the net flow (total rate) into the state and the net flow (total rate) out of the state.

$$\sum_{i=1}^{K} m_i p\left(\underline{n}\right) = \sum_{i=1}^{K} \sum_{j=1}^{K} m_i r_{i,j}\, p\left(\underline{n} - \underline{1}_j + \underline{1}_i\right), \tag{2}$$

which is called the global balance equation at state $\underline{\boldsymbol{n}}$.

In terms of the traffic equations, we can solve the global balance equations to obtain the following product form solution

$$p\left(\underline{n}\right) = p\left(\underline{0}\right) \prod_{i=1}^{K} \left(\frac{q_i}{m_i} \right)^{n_i}, \tag{3}$$

where

$$p\left(\underline{0}\right) = \left\{ \sum_{\underline{n}} \left[\prod_{i=1}^{K} \left(\frac{q_i}{m_i} \right)^{n_i} \right] \right\}^{-1}, \tag{4}$$

Let q_i be the average throughput through queue ***i***. Then,

$$q_i = \sum_{j=1}^{K} r_{j,i} q_j, \qquad i = 1,2,\dots,K, \tag{5}$$

which are the traffic equations of the network [4].

4. M/M/1/K/K queueing system application to the IN architecture

Traditional PSTN services were provided by the service logic and data resident within the local switching machine. The capacity for these services is very much determined by the architecture and component capacities within the service node. IN has a distributed architecture in which the service logic is executed cooperatively by different network elements that can be geographically dispersed [2].

Figure 2 shows an IN network distributed architecture as a finite source model – M/M/1/K/K, also known as the machine repair model, or the cyclic queue model. In that context, there are ***K*** requests cycling in a system consisting of ***K*** SSPs and a central processor unit (CPU) with a work queue representing the SCP. A request for IN service is sent from the SSP to the SCP after an exponentially distributed think time and after being processed by the SCP the response returns to the SSP which enters in another think phase. The input and output messages of a transaction are treated as a single composite service. Also, the think time and the SCP processing time are considered as an average operating time [5].

The rate at which requests enter the SCP is equal the rate at which responses leave, because the system is assumed to be in equilibrium. The period of time that a particular SSP is in the thinking state is merely the ratio of the average thinking time ***1/****l* to the average time ***T*** + ***1/****l* which it spends in making a complete round trip. Each of the ***K*** SSPs generates requests at a rate *l* per second, provided it is in the thinking state. On the other side, the period of time when the SCP is busy is $\boldsymbol{1-p_0}$, during which time the output rate is *m*. Thus the average output rate of jobs is $m\boldsymbol{(1-p_0)}$. Equating the input and output rates we have:

$$K\,l\,\frac{1/\,l}{T+1/\,l} = m\left(1-p_0\right). \tag{6}$$

And, from that equation, we obtain the equation for ***T*** [6; 7]:

$$T = \frac{K/m}{1-p_0} - \frac{1}{l}. \tag{7}$$

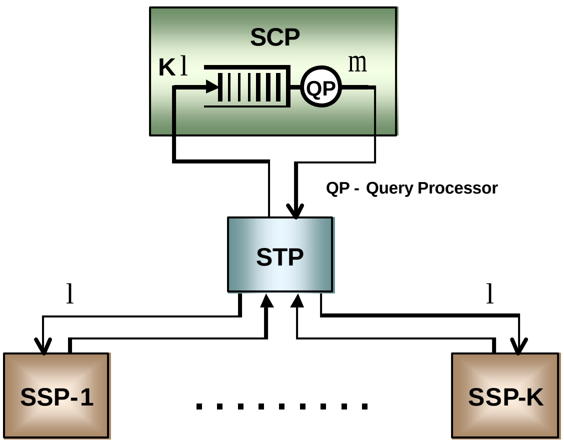

Figure 2. The IN distributed architecture

as a M/M/1/K/K model

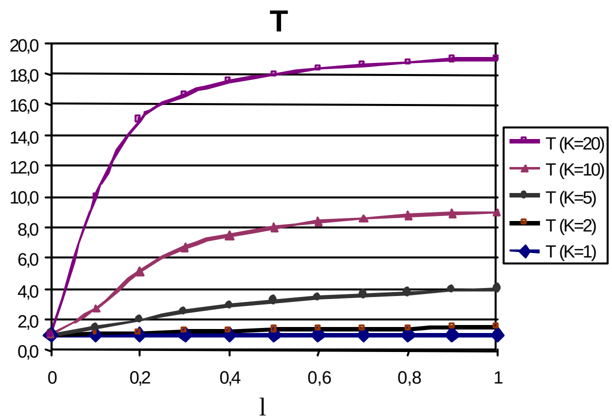

Figure 3. The average time in the system T

The numerical results for the expected response time, which is the sum of the waiting time in the queue and the service time (e.g., average time in system - ***T***), are given in Table 4 and Figure 3. In

Figure 3, we have presented ***T*** for several values of ***K***. Here the service time is exponentially distributed. [5].

Table 4. The average time in system ***T***

l	T (K=1)	T (K=2)	T (K=5)	T (K=10)	T (K=20)
0.001	1.0	1.00100	1.00401	1.00906	1.01933
0.1	1.0	1.09091	1.46630	2.73208	10.0375
0.2	1.0	1.16670	1.99171	5.18729	15.0000
0,3	1.0	1.23077	2.47526	6.68335	16.6670
0.4	1.0	1.28571	2.87479	7.50216	17.5000
0.5	1.0	1.33333	3.19048	8.00038	18.0000
0.6	1.0	1.37500	3.43741	8.33342	18.3333
0.7	1.0	1.41176	3.63177	8.57145	18.5714
0.8	1.0	1.44444	3.78677	8.75001	18.7500
0.9	1.0	1.47368	3.91225	8.88889	18,8889
0.999	1.0	1.49975	4.01445	8.99900	18,9990

5. Conclusions

The analysis and modeling of system performance issues are essential tools in the development and engineering processes that can be used at all stages of the lifecycle of IN services. Simple, approximate models have a high value in the early stages to uncover major performance problems which affect the design of the architecture before the cost of rectification is too high. The design tools support rapid prototyping, allowing users to go through the three important stages: predict, design, and comparison. The questions of the development of new modeling methods for rapid analysis, and some others, like new IN performance standards, and closer connections between performance analysis and service design are the most interesting for the future wide deployment of IN and converged networks services.

References.

[1] R. Ackeley, A. Elvidge, T. Ingham, and J. Shepherdson, "*Network Intelligence – Performance by Design*", IEICE Transactions on Communication, vol. E80-B, No.2, February 1997, pp.219-229.

[2] J. Yan and D. MacDonald, "*Teletraffic Performance in Intelligent Network Services*", Proceeding of the 14th International Teletraffic Congress, Antibes, Juan-les-Pins, France, pp.357-366, June 1994.

[3] E. Masuda, T. Mishima, N. Takaya, K. Nakai, and M. Hirano, "*A Large-Capacity Service Control Node Architecture Using Multicasting Access to Decentralized Databases in the Advanced Intelligent Network*", IEICE Transactions on Communication, vol. E84-B, No.10, October 2001, pp.2768-2780.

[4] D. Bertsekas, R. Gallager, "*Data Networks*" – Second edition, ISBN 0-13-200916-1, Prentice-Hall, 1992.

[5] M. H. van Hoorn, "*Algorithms and Approximations for Queueing Systems*", PhD thesis, Vrije Universiteit te Amsterdam, Mathematical Center, Amsterdam, 1983.

[6] L. Kleinrock, "*Queueing Systems. Volume II: Computer Applications*", ISBN 0-471-49111-X (v.2), Wiley-Interscience publication, 1986.

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