LIBERTAS MATHEMATICA, VOL. XV (1995)

A Markov Model for Quality of Service in Public Data Networks

ILIÉ POPESCU and EUGENIE STAICUT

ABSTRACT

This paper defines the *equivalent ability* of a *public data network*, taking into account the *reliability* of network components, their *maintainability*, as well the *congestion* (a state in which the network rejects data transmission services to users). It shows that this measure reflects the *quality of service* that the users receive from the network. Then, a *Markov model* is developed to provide a means of evaluating the network equivalent availability. An expression of network equivalent availability is also derived.

Keywords: Quality of service, packet switching, public data networks, reliability, availability, congestion control, Markov model, computer networks.

1. INTRODUCTION

The aspects of *quality of service* available from *packet-switched data networks* are analyzed in many papers [1, 2, 3, 6]. The principal objective of quality of service modelling is to provide a rational foundation on which to base network planning policy decisions. To achieve this objective, the models must take into account the recent increases in network complexity. Some recent network advanced procedures (for instance, the more sophisticated routing algorithms that utilize lightly loaded parts of a network) are not generally easy to model using analytical methods. There are still difficulties in the quantification of the overal end-to-end performance of some network advanced schemes.

The quality of service available from packed-switched public data networks mainly depends on three characteristics of the network: *reliability*, *maintainability*, and *congestion*. For such networks, usually, these characteristics are examined separately. The *network* *availability* is considered a measure of quality of service for some networks [5, 12]. The congestion in packet-switched computer networks is presented in [9, 10, 11]. References [4] and [7] present a way of *combining reliability* measures with *congestion* measures to produce a measure that reflects the quality of service that the users receive from a packed-switched public data network. This measure is known as the *equivalent availability* AE(t), and it is defined as the probability that at time *t* the network is in a state of physically correct operation and also the network is not congested. More details on the congestion-reliability relationships are given in [4, 7].

This paper presents a *Markov model* which has been developed to evaluate the network equivalent availability.

2. DESCRIPTION OF THE MODEL

If we consider the communication between two users via a public data network, they are interested not only in the existence of a physical communication path which operates correctly, but also of a non-congested path. A network is called *congested* when it is in a state in which it must reject incoming data packets.

With these considerations on the availability of user services, let the *states* of a packet-switched public data network be:

i) *state 0*, the state of good operation; the state in which the network physically operates correctly; that is, there is at least one communication path among all the network users, and also the network is not congested;

ii) *state 1*, the state of physical failure; that is, the state in which there is no physical communication path among users;

iii) *state 2*, the state of congestion; that is the state of physically correct operation, but the network is congested and therefore it cannot provide data transmission services, although there is at least a physical communication path among users.

The transition from the state 0 to the state 1 is caused by physical failure (hardware or software); the transition from the state 1 to the state 0 by removing the detected faults. Let λ be the network *physical failure rate* and μ the *repair rate*.

The transition from state 0 to the state 2 occurs after the network has entered a congestion state and the transition from the state 2 to the state 0 after the network operation has been recovered from a congestion state. Let λ_c be the rate of network *congestion failures* and μ_c the rate of recovery from a congestion state.

The transition from the state 2 to the state 1 is caused by a physical failure that occurs when the network is in the congestion state. Let λ' be the network *physical failure rate* in the congestion state.

Assuming an *exponential distribution* of the physical good operation times, congestion occurrence times, physical repair times and congestion recovery times λ, λ_c, μ, μ_c and λ' are constant.

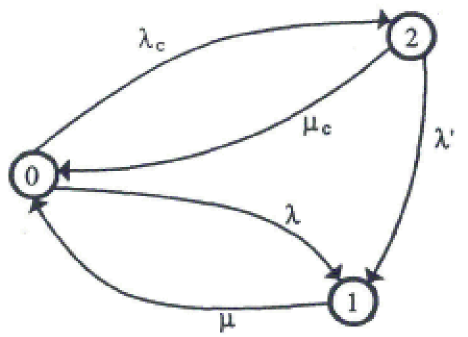

Fig. 1. A Markov model for the equivalent availability of a packet-switching computer network.

With these considerations, the network may be represented by a Markov model, as shown in Figure 1. The *equivalent availability* AE(t) represents the probability that the network will be in the state 0.

(1) $AE(t) = P_0(t)$.

3. SOLUTIONS OF THE MARKOV MODEL

Markov models are well known [8], and we omit some details here. Let $P_0(t)$, $P_1(t)$ and $P_2(t)$ be the probabilities that at time t the network will be in the state 0, 1 and 2 respectively. Also, let $p_{ij}(\Delta t)$ be the probability of transition from state i to state j, within a time interval Δt, where i,j = 0, 1, 2. Thus, the state transition probability matrix of the model is given by:

$$(2) \qquad p(\Delta t)] = \begin{bmatrix} 1-(\lambda+\lambda_c)\Delta t & \lambda\Delta t & \lambda_c\Delta t \\ \mu\Delta t & 1-\Delta t & 0 \\ \mu_c\Delta t & \lambda'\Delta t & 1-(\mu_c+\lambda')\Delta t \end{bmatrix}$$

Using the total probability formula, the probability $P_j(t+\Delta t)$ of being in state j at time $t+\Delta t$ is:

$$(3) \qquad P_j(t+\Delta t) = \sum_{i=0}^{2} P_j(t)p_{ij}(\Delta t)\ ; \text{ for } j=0,1,2.$$

From (2) and (3), we get the following system of equations:

$$(4) \qquad \begin{cases} P_0(t+\Delta t) = P_0(t)\,[1-(\lambda+\lambda_c)\Delta t] + P_1(t)\,\mu\Delta t + P_2(t)\mu_c\Delta t \\ P_1(t+\Delta t) = P_0(t)\lambda\Delta t + P_1(t)\,(1-\mu\Delta t) + P_2(t)\lambda'\Delta t \\ P_2(t+\Delta t) = P_0(t)\lambda_c\Delta t + P_2(t)\,[1-(\mu_c+\lambda')\Delta t] \end{cases}$$

Rearrangement of equations (4) yields:

$$(5) \qquad \begin{cases} \dfrac{P_0(t+\Delta t)-P_0(t)}{\Delta t} = -(\lambda+\lambda_c)\,P_0(t) + \mu P_1(t) + \mu_c P_2(t) \\ \dfrac{P_1(t+\Delta t)-P_1(t)}{\Delta t} = \lambda P_0(t) - \mu P_1(t) + \lambda' P_2(t) \\ \dfrac{P_2(t+\Delta t)-P_2(t)}{\Delta t} = \lambda c P_0(t) - (\mu_c+\lambda')P_2(t) \end{cases}$$

Passing to a limit as Δt becomes small ($\Delta t \to 0$), we get the following set of ordinary, first-order differential equations:

$$(6) \qquad \begin{cases} \dfrac{dP_0(t)}{d} = -(\lambda+\lambda_c)\,P_0(t) + \mu P_1(t) + \mu_c P_2(t) \\ \dfrac{dP_1(t)}{dt} = \lambda P_0(t) - \mu P_1(t) + \lambda' P_2(t) \\ \dfrac{dP_2(t)}{dt} = \lambda_c P_0(t) - (\mu_c+\lambda')P_2(t) \end{cases}$$

The network state occupancy probabilities can be determined from the system of differential equations represented by (6). Assuming that initially (at time t=0) the network is in the state 0, the following initial conditions hold:

(7) $$\begin{cases} P_0(0) = 1 \\ P_1(0) = 0 \\ P_2(0) = 0. \end{cases}$$

The sum of all state occupancy probabilities must be 1:

(8) $$\sum_{i=0}^{2} P_j(t) = 1.$$

The system of equations represented by (6) can be solved using Laplace transforms. Let $P_j(s)$ be the Laplace transform of $P_j(t)$:

(9) $$P_j(s) = L\{P_j(t)\}; \qquad \text{for } j = 0,1,2.$$

The Laplace transform of the derivatice of $P_j(t)$ is:

(10) $$L\left\{\frac{dP_j(t)}{dt}\right\} = sP_j(s)\text{-}P_j(0)\ ; \qquad \text{for } j = 0,\ 1,\ 2.$$

Using (9) and (10), from (6) we get the following system of linear equations:

(11) $$\begin{cases} -(s + \lambda + \lambda_c P_0(s) + \mu P_1(s + \mu_c P_2(s) = -1 \\ \lambda P_0(s) - (s + \mu)P_1(s) + \lambda' P_2(s) = 0 \\ \lambda_c P_0(s) - (s + \mu_c + \lambda')P_2(s) = 0 \end{cases}$$

The probability $P_0(s)$ of state 0 is derived by solving the system of equations (11) using Cramer's rule. We get:

(12) $$P_0(s) = \frac{(s+\mu)\,(s+\mu_c+\lambda')}{s\left(s+\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'+\sqrt{D}}{2}\right)\left(s+\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'-\sqrt{D}}{2}\right)},$$

where

(13) $$D = [(\mu+\lambda)-(\mu_c+\lambda_c+\lambda')]^2 + 4\lambda_c(\lambda-\lambda').$$

The roots of the denominator of $P_0(s)$ are:

(14) $$s_{1,2} = -\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'\pm\sqrt{D}}{2}.$$

Note that inspection of relations (14) shows that s_1 and s_2 are always negative real numbers. Expanding the expression of $P_0(s)$ given by (12) in partial fractions yields:

(15) $$P_0(s) = \frac{A}{s} + \frac{B}{s+\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'+\sqrt{D}}{2}} + \frac{C}{s+\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'-\sqrt{D}}{2}}.$$

where:

(16) $$A = \frac{\mu(\mu_c+\lambda')}{(\mu+\lambda)(\mu_c+\lambda_c+\lambda')-\lambda_c(\lambda-\lambda')};$$

(17) $$B = \frac{(-\mu+\lambda+\mu_c+\lambda_c+\lambda'+\sqrt{D})(\mu+\lambda-\mu_c+\lambda_c-\lambda'+\sqrt{D})}{2\sqrt{D}(\mu+\lambda+\mu_c+\lambda_c+\lambda'+\sqrt{D})};$$

(18) $$C = \frac{(-\mu+\lambda+\mu_c+\lambda_c+\lambda'-\sqrt{D})(\mu+\lambda-\mu_c+\lambda_c-\lambda'-\sqrt{D})}{2\sqrt{D}(\sqrt{D}-\mu-\lambda-\mu_c-\lambda_c-\lambda')}.$$

Using the inverse Laplace transform, from the relation (15) we get the probability $P_0(t)$ of state 0, which is equal to the equivalent availability AE(t) of the network:

(19) $$AE(t) = P_0(t)$$

$$= A+B.e^{\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'+\sqrt{D}}{2}} + C.e^{-\frac{\mu+\lambda+\mu_c+\lambda_c+\lambda'-\sqrt{D}}{2}}.$$

Since s_1 and s_2 given by (14) are negative real numbers, the time functions in the expression of AE(t) given by (19) are decaying exponentials.

Since in practice the availability is evaluated by the ratio K_A:

(20) $K_A = \mu/(\mu+\lambda)$;

in the case of the equivalent availability of the network we can use the factor K_{AE}, defined as follows:

(21) $K_{AE} = \lim_{t\to\infty} AE(t)$

Passing to a limit as t becomes ∞, from (19) we get:

(22) $K_{AE} = \dfrac{\mu(\mu_c+\lambda')}{(\mu+\lambda)(\mu_c+\lambda_c+\lambda')-\lambda_c(\lambda-\lambda')}$.

Dividing by $\mu(\mu_c+\lambda')$ both the numerator and the denominator, and after some simple rearrangements, the relation (22) becomes:

(23) $K_{AE} = K_A{\cdot}K_C/\left[1-\dfrac{\lambda-\lambda'}{\mu}\,\dfrac{\lambda_c}{\mu_c+\lambda'}\,K_AK_C\right]$;

where:

(24) $K_c = (\mu_c+\lambda')\,/\,(\mu_c+\lambda_c+\lambda')$.

We can observe that when $\lambda = \lambda'$, the result is:

(25) $K_{AE} = K_A{\cdot}K_C$

For a normal operation of the network, it is possible to consider $\mu_c > \lambda_c$, $\mu > \lambda$ and $\mu > \lambda'$ the result is:

(26) $$K_{AE} = K_A \cdot K_C$$

4. CONCLUSIONS

Markov model for the *equivalent availability* AE(t) of a *packet-switching network*, is defined here as being the probability that at time *t* the network will be in a state in which there is at least a physical communication path among all users, and also the network will not be congested. The equivalent availability is used as a measure of *quality of service* available from public data networks, being thus useful to network designers and planning policy decisions. The concept is also useful to network reliability analysts and theoreticians. It is useful to consider several possible extensions or modifications of the Markov model that we have presented. For instance the state transition rates may be considered functions of network operation time. Numerical methods can be used for solving the system of equations (6) in the more general case.

BIBLIOGRAPHY

[1] Ackerley, R.G., The Future of Quality of Service Modelling, in: Computer Networks and ISDN Systems, Vol. 20, No. 1-5, 1990 December, pp. 7-14.

[2] CCITT X.130 series Recommendations, CCITT Blue Book, Volume VIII, 1988.

[3] Kleinau, K.H., Fischer, K. and Pretzsch, W., Quality of Service and Human Factors, in: Reliability in Electrical and Electronic Components and Systems, E. Lauger and J. Moltoft (editors), North-Holland Publishing Company, 1982.

[4] Lazaroiu, D.F., and Staicut, E., Congestion-Reliability-Availability Relationship in Packet-Switching Computer Networks, in: IEEE Trans. on Reliability, Vol. R-32, No. 4, 1983 October, pp. 354-357.

[5] D'Ottavio, A., Mariorino, E. and Valenti, A., Reliability Aspects of the New Digital Point to Point and Multipoint Italian Data Network,: Relectronic '77, 4th Symposium on Reliability in Electronics, Budapest, Hungary, 1977 October, pp. 745-754.

[6] Park, J.L., Quality of Service Requirements in Packet Switching Public Data Networks, in: Computer Networks and ISDN Systems, Vol. 20, No. 1-5, December 1990, pp. 3-6.

[7] Popescu, I., Staicut, E., Probabilistic Model for Quality of Service in Packet-Switched Data Networks: 17th Conference on Local Computer Networks, IEEE Computer Society, Minneapolis, Minnesota, Sept. 13-16, 1991, pp. 698-708.

[8] Schooman, M.L., Probabilistic Reliability: An Engineering Approach McGraw-Hill Book Company, 1968.

[9] Schwartz, M., Telecommunication Networks: Protocols, Modeling and Analysis, Addison-Wesley Publishing Company, 1987.

[10] Stallings, W., Data and Computer Communications, Macmillan Publishing Company, 1985.

[11] Tanenbaum, A.S., Computer Networks, Prentice-Hall Inc., 1981.

[12] Taylor, R. and Addie, R., An Algorithm for calculating the Availability and Mean Time to Restore for Communication through a Network, in: Computer Networks and ISDN Systems, Vol. 20, No. 1-5, 1990 December, pp. 109-114.