

# SAFETY-CRITICAL WIDE AREA NETWORK PERFORMANCE EVALUATION

**Tuncay Bayrak**

Rensselaer Polytechnic Institute, Troy, New York, USA

**Martha R. Grabowski**

Rensselaer Polytechnic Institute, Troy, New York, USA

and

LeMoyne College

Syracuse, New York, USA

## ABSTRACT

*The growing importance of real-time computing in numerous applications poses problems for network architectures, especially safety-critical Wide Area Networks (WANs). Assessing network performance in safety-critical real-time systems is difficult, and suggests the use of both human and technical performance criteria because of the importance of both dimensions in safety-critical settings. This research proposes a model that considers both technical and human performance in network evaluation.*

## 1. INTRODUCTION

Wide Area Networks (WANs) are important components in safety-critical environments where reliable data acquisition and distribution are essential. In such systems, network equipment and functions must be closely monitored and controlled to ensure safe operation and prevent costly consequences. As networks become more complex, the probability of system failure increases, particularly for real-time WANs which contain hundreds of nodes. Examples of such safety-critical wide area networks include distributed battle management systems (Mosher, 1997), intelligent transportation systems (Andrisano et al., 2000), distributed health care networks (Yamamoto et al., 2000), global oil and gas exploration and research networks (MacIntyre, 1999), and aviation traffic monitoring systems (National Research Council, 1997; Cheng et al., 2000).

Most large-scale networks depend on hardware, software, human operators and other network elements to function correctly. Failure of any of the network elements can bring the entire network down and in safety-critical settings, the consequences can be disastrous. A well-known example of such failure is the 1990 nationwide AT&T network failure (Kuhn, 1997). This example is not an isolated one: according to the Federal Communication Commission (FCC), network failures in the United States with impact on more than 30,000 customers happen on the order of one every two days and the mean time to repair them is on the order of five to 10 hours (Demeester, et al., 1999).

In safety-critical settings, the human, environmental, and economic consequences of network failures can be staggering. Network reliability is critical in these settings, as failure of a real-time system could cause an economic disaster or the loss of human lives (Shin, 1993). Since survivability and reliability are crucial in safety-critical systems, careful evaluation of these systems is important. However, few evaluation models of real-time safety-critical wide area networks have been developed, a need that motivates this research.

## 2. THEROTICAL FOUNDATIONS

Over the years, networks have been evaluated by different disciplines from different perspectives. Mathematical models based on queuing theory, Markov analysis and using well-defined metrics such as throughput, response time, and utilization have been used in many network performance evaluations (Haverkort, 1998; Bolch, et al., 1998; Higginbottom, 1998). Other metrics utilized include network traffic performance (Adie, et al., 1998; Banerjee, et al., 1997), circuit overhead of switches (Niehaus, et al., 1997; Da Silva, et al., 1997), and equipment used and network conditions (Da Silva, et al., 1997).

Statisticians frequently use statistical distributions to evaluate communication networks as distributions allow prediction of system performance measures to a reasonable degree of accuracy (Akar, et al., 1998). Technical communication models often consider network traffic over switches, routers, bridges and repeaters (Khalil, et al., 1995). Social and organizational communication models consider networks of organizations, their patterns of behavior and communication strategies, and organizational structures (Monge, et al., 1998; Orlikowski, et al., 1995).

Large-scale system models evaluate networks in terms of two important concepts, reliability and survivability. Survivability is defined as the percentage of total traffic surviving some network failure in the worst case (Myung, et al., 1999). Reliability is a measure of the system's ability to provide deterministic and accurate delivery of information (McCabe, 1998). In other words, reliability is the likelihood that a system will remain operational (potentially despite failures) for the duration of a mission (Somani and Nittin, 1997).

Technical metrics used to evaluate wide area networks include system reliability, availability, system usage, Mean Time Between Failures (MTBF) and Mean Time to Repair (MTTR). These measures have been suggested by mathematical models, large-scale system models, statistical and technical communication models, and are often interdependent. For example, increased use (e.g., an increased number of users) of a real-time WAN may result in decreased reliability, decreased network accuracy, decreased response time, and increased network workload. Similarly, increased use of a network may cause degradation or wearout in the network components, thus leading to lower reliability, erroneous data, and perhaps erroneous information. Hence, frequency of network use may not always be a good success measure; in fact, it may trigger undesired reliability problems such as increased network workload, and decreased response time, which is critically important for safety-critical applications.

Other important performance measures for networks include message delay, variance of delay, message loss and overflow (Adie et al., 1998; Bolch, et al., 1998; Higginbottom, 1998). Delay is the amount of time from the moment a message arrives until its service is complete. Loss ratio is the ratio between the total number of messages lost and the total number of messages arriving. Overflow probability is the probability that the number of messages in a network buffer exceeds a certain threshold. Two other message characteristics--peak and mean traffic--have also been utilized as metrics. Mean traffic is simply the average number of messages generated in a unit of time. Peak traffic represents the highest rate of traffic generated. If the service rate is equal to or higher than the peak rate, no messages will be buffered, and the loss ratio will be equal to zero. If the service rate is lower than the mean arrival rate, the queue will lead to unacceptable levels of delay, loss ratio, and overflow probability.

Metrics such as the processing delay at each node, the capacity of each link, and round trip propagation delay--the time it takes a bit to travel along the media at the speed of light--have also been used as performance criteria (Bournas, 1995), as have the average queue length of messages awaiting service, utilization, throughput, node delay, and end-to-end delay metrics (Chirchir and Kamal, 1995). Finally, redundancy is a metric that is often considered in network evaluations. A network should be able to accomplish its objectives satisfactorily despite failures of individual network components. For

real-time networks to survive failures, designers must often incorporate redundancy, which can involve assigning more hardware and software components (Berman and Kumar, 1999). However, redundancy may increase network complexity and increase network usage, especially in applications where network survivability is crucial.

In engineering and business models, WANs have been evaluated from the customer point of view, using such criteria as cost, connectivity, bandwidth/speed, data integrity, availability, reliability, and security (Hemrick, 1992). Business models consider network performance as a means of enhancing the performance of an enterprise (Yang et al., 2001) because network managers are interested in fully functional, high performance, and secure networks that provide resilient services (Rudd, 2000). High-performance enterprise networks can help an enterprise operate more efficiently and improve its competitive capability. Thus, economic aspects are always important (Yang et al., 2001).

From an organization's point of view, however, networks are seen as an investment. Jurison (1996) argues that success measures of interest to managers are those that can be measured and expressed quantitatively, especially in monetary terms, because such measures can be used for justifying information technology investments. Thus, organizations are usually interested in knowing cost savings, reliability, accuracy, flexibility, timeliness of data, decision support applications, isolation, integration, user involvement, security, and back-up requirements (Axelrod, 1982). Finally, psychological and sociological models of network performance assess optimal communication structures, improvement of decision making, the impact of communication networks on organizations and their performance, and distribution of decision making rights over the network using such metrics as the time taken to correctly solve a problem, the number of messages used for each problem, and the number of errors (Jehiel, 1999; Mackenzie, 2000).

Thus, network evaluation has been considered in different ways by different disciplines over the past forty years. Many of these evaluations focus on network technical performance, or an organization's performance when using a network, or individual users' performance when using a network. Few evaluations, however, consider both social and technical impacts of network performance, both of which are key in safety-critical large-scale systems. Because humans and technology cooperatively perform tasks in network-centered safety-critical large-scale systems, the model proposed in the next section for performance evaluation of safety-critical WANs in real-time settings encompasses both social and technical dimensions. We now describe the model and its theoretical underpinnings.

### **3. THEORETICAL MODEL**

The different literatures surveyed illustrate that network evaluation has been considered from several perspectives --technical, social, organizational, psychological and commercial. In safety-critical settings, where network failures can have catastrophic effects and networks provide an important social and technical infrastructure, utilizing performance criteria that reflect the differing requirements that such networks must meet is important [So and Durfee 1996]. For instance, real-time safety-critical WAN's must meet stringent response, availability, reliability, survivability, accuracy and redundancy requirements; thus, use of technical performance criteria can provide some measure of the network's ability to meet those requirements. Similarly, real-time WAN's in safety-critical settings must also meet critical communication, decision-making, problem-solving and organizational effectiveness requirements; as a result, social, psychological and organizational network performance criteria can also be used to measure the social and organizational effectiveness of the network infrastructure. Finally, in many cases, real-time WAN's in safety-critical settings must also satisfy demanding commercial and economic requirements, as befitting their industrial hosts. Thus, commercial and economic performance criteria can provide measures of the network's ability to satisfy its economic and resource requirements. These requirements suggest important performance

criteria for use in evaluating real-time WAN's in safety-critical settings. In such evaluations, technical, social, organizational, psychological, commercial and economic evaluation criteria provide a means of measuring the performance of the network, and of addressing the social, technical and economic challenges faced by real-time WAN's.

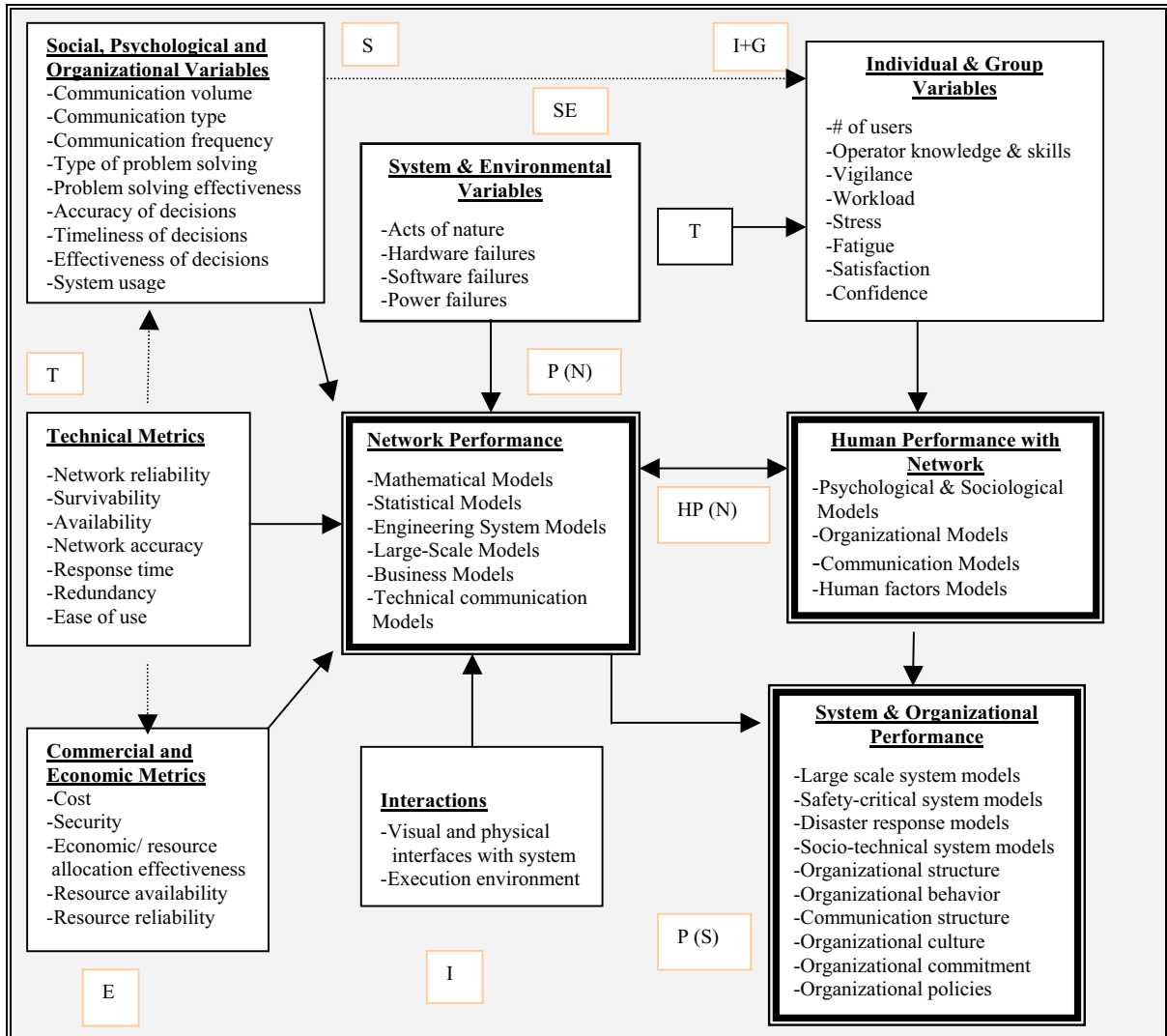


Figure 1. Proposed Model

Figure 1 illustrates the proposed evaluation approach. Three types of performance are of interest in evaluating WANs in real-time safety-critical settings: the performance of the network P (N); the human performance of those using the network --both operator and user-- HP (N); and the performance of the system and organization P (S), as seen in Figure (1). Real-time networks interact with humans, the environment, and other technologies, and interactions between these different elements may contribute to network failures. Hence, in addition to traditional technical performance considerations, the Figure 1 WAN evaluation model considers human factors and environmental considerations. This is because human error and acts of nature are among the major sources of failures in networks (Kuhn, 1997).

As discussed earlier, technical variables (T), such as network reliability, accuracy, response time and utilization, certainly impact network performance P (N), as do social, psychological and organizational variables (S), commercial and economic variables (E), human performance with the network HP (N), or system and environmental variables (SE) such as hardware failures, software

failures and acts of nature, and interactions (I) between the network and its working environment (Figure 1). Network performance, therefore, is a function of technical variables such as reliability, accuracy, response time and utilization; social, psychological and organizational variables such as communication type, frequency, and timeliness of decisions; commercial and economic variables such as cost and security; human performance with the network; as well as a function of system and environmental variables, such as acts of nature or power failures; and interactions between the network and its physical and execution environment. Network performance can be assessed using a variety of mathematical, statistical, engineering system, large-scale system, and business models, as explained in Section 2, and the relationships between the network performance factors can be expressed in the following way:

$$P(N) = f(T + S + E + SE + I + HP(N))$$

where P(N) = performance of the network,  
 T = technical network variables,  
 S = social, psychological and organizational variables,  
 E = commercial and economic variables,  
 SE = system and environmental variables,  
 I = interactions between the network and its working environment, and  
 HP(N) = human performance with a network.

Note that in Figure 1, technical variables (T) also influence commercial and economic variables such as cost and social, psychological and organizational variables (S), such as accuracy, communication and system usage. These are indirect effects on network performance P(N), and the impact vectors in Figure 1 for these variables are shown as dotted lines.

In turn, network performance P(N) influences human performance with the network HP(N) as well as the performance of the system that the network serves P(S). Individual (I) and group (G) variables such as user knowledge and skills, vigilance and workload, also influence human performance with the network HP(N), as seen in Figure.

Human performance with a network is thus influenced by the network's performance as well as by individual and group variables such as individual or group's knowledge or skills, workload, stress, experience with networking, and /or fatigue. Human performance with a network also influences the network's performance, and can be assessed using a variety of psychological, sociological, organizational, human factors, and communication models. These relationships can be expressed as:

$$HP(N) = f(P(N) + I + G)$$

where HP(N) = human performance with a network,  
 P(N) = performance of the network,  
 I = individual performance variables, and  
 G = group performance variables.

Similarly, in Figure 1, social, psychological and organizational variables (S) influence individual (I) and group (G) variables such as workload, stress, and fatigue.

Finally, overall system performance for the systems that host real-time WANs is influenced by the performance of a network P(N) as well as by human performance with the network HP(N), as in  $P(S) = P(N) + HP(N)$ ,

where P(S) = performance of the system,  
 P(N) = performance of the network, and  
 HP(N) = human performance with a network.

System performance can be assessed using large-scale, socio-technical and safety-critical system models, as well as by examining the system and organizational structures, policies, performance, behavior and culture.

It is obvious that systems success is both a social and technical accomplishment, but success/failure is far more complex than totaling up a number of factors. Differences in the various elements of the proposed model in order to evaluate the system as a success or failure will be accommodated in forthcoming research.

In the following section, we describe use of the Figure 1 model in evaluating an operational real-time WAN.

## **4. RESEARCH METHODOLOGY**

### **4.1 Research Vehicle**

There are two sets of subjects for this research: an operational wide area network (WAN) for the network performance evaluation, the operators who utilize the network for the human performance evaluation.

The vehicle for this study is an operational WAN known as the Continuous Operational Real-Time Monitoring System (CORMS). CORMS was designed and built by the U.S. National Oceanic and Atmospheric Administration (NOAA), and was implemented in April 1998. CORMS's purpose is to provide a 24 hour/day monitoring and quality control of water level and meteorological data from around the US to ensure the availability and accuracy of tide and water current observations that are used for navigation and safety of life and property decisions. To do this, CORMS takes input from two NOAA systems, the Physical Oceanographic Real Time System (PORTS) and the National Water Level Observation Network (NWLON). PORTS collects meteorological (wind, weather, tide current, etc.) and environmental data from San Francisco, New York, Tampa Bay, Houston/Galveston, Chesapeake Bay, and Narragansett Bay in the United States. NWLON, which collects water-level data, is comprised of 189 water level gauges located around the coastal United States, including Alaska, Hawaii, and U.S. territories in the Pacific, and Great Lakes.

The PORTS meteorological data and the NWLON water level data is gathered continuously via the CORMS network and transferred in real-time to the CORMS server at NOAA headquarters in Silver Spring, Maryland. 6 minute sample data from the real-time NWLON and PORTS data is monitored continuously in 24/7 mode by 6 watchstanding operators who monitor the CORMS data and displays and determine what actions are necessary if the accuracy of any of the measured parameters is deemed to be questionable (NOAA, 1999).

Since this paper focuses on establishing a technical baseline only, the human factors are not elaborated in the theoretical background. However, in forthcoming evaluation, the focus will be placed on the human factors. Thus, the forthcoming evaluation will focus on network monitoring watchstanders monitoring a Visual Display Terminal and responding based on the information displayed and behavioral patterns of 24x7 watchstanders. The purpose of the operator performance evaluation was to determine how technical variables such as network reliability and response time and individual and group variables such as workload and vigilance level influence operators' performance with the network under study.

## 4.2 Procedure

Hypotheses, variables, their operationalizations and measurements for evaluating a safety-critical real-time WAN were developed following the model in Figure 1, as seen in Table 1. Network performance was to be evaluated by utilizing well-defined and well-known network performance metrics such as reliability, availability, and response time. The appropriate statistical tests and mathematical analyses were run on collected data, and the results of the mathematical analyses and statistical tests were used to evaluate the hypotheses.

**Table 1. Hypotheses, Dependent Variables, and Metrics**

#	Hypotheses	Dependent Variable	Variable Operationalization	Measurement
1	<p><b>1a:</b>Increased use (# of users) of real-time WAN will result in decreased network reliability.</p> <p><b>1b:</b>Increased use (# of users) of real-time WAN will result in decreased network accuracy.</p> <p><b>1c:</b>Increased use (# of users) of real-time WAN will result in decreased network response time.</p> <p><b>1d:</b>Increased use (# of users) of real-time WAN will result in increased network workload.</p>	<p>Network reliability</p> <p>Network accuracy</p> <p>Network response time</p> <p>Network workload</p>	<p>1) Type, and time of breakdowns.</p> <p>2) Correctness of data.</p> <p>3) Time taken to obtain response.</p> <p>4) How much traffic is flowing from a given source to a given destination network.</p>	<p>1) MTBF, MTTR, Availability (%).</p> <p>2) Probability of detecting error.</p> <p>3) Mean response time.</p> <p>4) Flow volume in bytes.</p>
2	<p><b>2a:</b>In safety-critical and real-time settings, increased network redundancy will result in increased network workload</p> <p><b>2b:</b>In safety-critical and real-time settings, increased network redundancy will result in increased cost,</p> <p><b>2c:</b>In safety-critical and real-time settings, increased network redundancy will result in increased usage,</p> <p><b>2d:</b>In safety-critical and real-time settings, increased network redundancy will result in increased network reliability.</p>	<p>Network workload</p> <p>Network cost of spare resources</p> <p>Network usage</p> <p>Network reliability</p>	<p>1) How much traffic is flowing from a given source to a given destination network.</p> <p>2) Spare resource units utilized.</p> <p>3) Level of system use.</p> <p>4) Type, and time of breakdowns.</p>	<p>1) Flow volume in bytes.</p> <p>2) Capital expenditure.</p> <p>3) Frequency of network use by an operator.</p> <p>4) MTBF, MTTR, Availability (%)</p>

### **Hypothesis 1: Increased use of a real-time WAN will negatively impact on WAN performance.**

Large-scale network performance usually deteriorates as the number of users and operators who utilize the network increase. For instance, network response time slows when an increased numbers of users want to get response from the system simultaneously. WAN usage was measured in this research by the number of users and operators who utilize the network.

In this study, WAN performance was measured using reliability, accuracy, response time, and network workload metrics. For the CORMS system, reliability was measured by type and time of communication breakdowns, using Mean Time Between Failures (MTBF), Mean Time To Repair (MTTR), and availability as the three main metrics. Correctness of network data is especially

important for safety-critical networks. Network accuracy was measured by correctness of data using the probability of detecting inaccurate data by the network. Network response time, crucial in safety-critical applications, was measured by collecting mean response time data. Finally, network workload was measured by assessing how much data traffic was flowing from a data source to a destination using data registered every six minutes by the equipment in each station. Thus, these hypotheses assess the impact of network usage on network performance.

**Hypothesis 2: In safety-critical and real-time settings, increased network redundancy will result in decreased network performance, and increased cost.**

In safety-critical settings, redundancy may be employed to decrease mean time to repair, to ensure continuous flow of data, and to increase availability percentage; however, it may have some drawbacks such as increased equipment cost. Redundancy in this study was measured by the amount of monetary resources allocated and utilized for the redundant equipment (etc., hardware, software).

In this study, network performance was measured using network workload, cost, usage, and reliability. For the CORMS systems, network workload was operationalized by calculating the amount of data traffic in bytes flowing from each station to the CORMS system. Network cost was measured by the capital expenditure spent for redundant resource units. Similarly, network usage was measured by assessing how often the network was used by operators and users. Finally, network reliability was measured using MTBF, MTTR, and availability, as in hypothesis 1.

### 4.3 Current status

The literature review is concluded, and the proposed model, hypotheses, dependent variables, and their operationalizations to evaluate subjects have been defined. Currently, data collection and survey administration are in progress, as is analysis of the collected data. Results will be available for conference presentation.

## REFERENCES

- Adie, R. G., Zukerman, I., and Neame, T. D (1998). Broadband Traffic Modeling: Simple Solutions to Hard Problems, *IEEE Communications Magazine*, 36 (8), 88-95.
- Akar, N., Oguz, N.C., and Sohrawy, K (1998). TELPAC: An Advanced Teletraffic Analysis Package, *IEEE Communications Magazine*, 36(8), 84-87.
- Andrisano, O., Verdone, R., and Nakagawa, M (2000). Intelligent Transportation Systems: The Role of Third-Generation Mobile Radio Networks, *IEEE Communications Magazine*, 38(9), 144-151.
- Axelrod, C. W (1982). *Computer Productivity: A Planning Guide for Cost Effective Management*, John Wiley, New York.
- Banerjee, S., Tipper, D., Weiss, B.H.M., and Kahlil, A (1997). Traffic Experiments on the vBNS Wide Area ATM Network, *IEEE Communications Magazine*, 35(8), 126-133.
- Berman, O., Kumar, U. D (1999). Optimization Models for Complex Recovery Block Schemes, *Computers & Operations Research*, 26(5), 525-544.
- Bolch, G., Greiner, S., de Meer, H., and Triverdi, K.S (1998). *Queuing Networks and Markov Chains, Modeling and Performance Evaluation with Computer Science Application*, John Wiley & Sons, New York.
- Bournas, M. R (1995). Bounds for Optimal Flow Control Window Size Design with Application to High-Speed Networks, *Journal of the Franklin Institute*, 332B (1), 77-89.
- Cheng, A. Y., Liu, R. Y., and Luxhoj, J. T (2000). Monitoring Multivariate Aviation Safety Data by Data Depth: Control Charts and Threshold Systems. *IIE Transactions*, 32(9), 861-872.
- Chirchir, K.C., and Kamal, E.K (1995). The Leaky Window: A Flow and Congestion Control Technique for HS-WANs, *Computer Communications*, 18 (2), 103-112.

- DaSilva, L. A., Evans, J.B., Niehaus, D., Frost, V. S., Jonkman, R., Beng O. L., Lazarou, G.Y (1997). ATM WAN Performance Tools, Experiments, and Results, *IEEE Communications Magazine*, 35(8), 118-124.
- Demeester, P., Wu, T., and Yoshikai, N (1999). Survivable Communication Networks, *IEEE Communications Magazine*, 37 (8), 40-42.
- Haverkort, B, R (1998). *Performance of Communication Systems, A Model-Based Approach*, Wiley, Chichester, NY.
- Hemrick, C (1992). Building Today's Global Computer Inter-networks, *IEEE Communication Magazine*, 30 (10), 44-49.
- Higginbottom, G.N (1998). *Performance Evaluation of Communication Networks*, Artech House, Boston.
- Jehiel, P (1999). Information Aggregation and Communication in Organizations, *Management Science*, 45(5), 659-669.
- Jurison, J (1996). The Temporal Nature of IS Benefits: A Longitudinal Study, *Information & Management*, 30 (2), 75-79.
- Khalil, K., Hand, J.C., Mariswamy, M., and Obaidat, M.S (1995). Methodologies for Characterizing Traffic in Wide Area Networks: Analysis and Benchmarking, *International Journal of Communication Systems*, 8(2), 117-127.
- Kuhn, R. D (1997). Sources of Failure in the Public Switched Telephone Network, *IEEE Computer*, 30 (8), 31-36.
- MacIntyre, H (1999). Linking Up to the Behemoth." *Mobile Radio Technology*, 17(3), 37-9.
- Mackenzie, K.D (2000). Processes and Their Frameworks, *Management Science*, 46(1), 110-125.
- McCabe, J.D (1998). *Practical Computer Network Analysis and Design*, Morgan Kaufmann Publishers, San Francisco.
- Monge, P. R., Fulk, J., Kalman, M. E., and Flanagan, A. J (1998). Production of Collective Action in Alliance-Based Interorganizational Communication and Information Systems. *Organization Science*, 9(3), 411-433.
- Mosher, D. E (1997). Grand Plans, *IEEE Spectrum*, 34(9), 28-39.
- Myung, Y., Kim, H., and Tcha, D (1999). Design of Communication Networks with Survivability Constraints, *Management Science*, 45(2), 238-252.
- National Oceanographic and Atmospheric Administration (NOAA) (1999). Center for Operational Oceanic Products and Services (CO-OPS), Silver Spring, Maryland.
- National Research Council (NCR) (1997), *Flight To The Future, Human Factors in Air Traffic Control*, National Academy Press, Washington D.C.
- Niehaus, D., Battou, A., McFarland, A., Decina, B., Dardy, H., Sirkay, V., and Edwards, B (1997). Performance Benchmarking of Signaling in ATM Networks, *IEEE Communications Magazine*, 35(8),134-142.
- Orlikowski, W. J., Yates, J., Okamura, K., and Fujimoto, M (1995). "Shaping Electronic Communication: The Metastructuring of Technology in the Context of Use." *Organization Science* 6(4), 423-444.
- Rudd, C (2000). Network management of an e-business environment, based on the new CCTA, ITIL Model," *Proceedings of the 2nd International Network Conference*, p.99-106.
- Shin, K.G (1993). *HARTS: A Distributed Real-Time Architecture*, *Advances in Real Time Systems*, IEEE Press.
- So, Y., and Durfee, H. E (1996). Designing Tree-Structured Organizations for Computational Agents, *Computational and Mathematical Organization Theory*, 2(3), 219-245.
- Somani, K. A., and Nitin, H. V (1997). "Understanding Fault Tolerance and Reliability," *Computer*, 30(4), 45-50.
- Yamamoto, K., Takada, T., Nakai, K., and Nagaoka, H (2000). Structuring WAN and LAN for EPR Use in Community Health Care, *International Journal of Medical Informatics*, 60(2), 219-226.
- Yang, C., Chen-Hua, F., and Yueh-Heng, T (2001). Enterprise Traffic with a Differentiated Service Mechanism, *International Journal of Network Management*, 11(2), 113-28.