JOURNAL ON
COMMUNICATIONS

VOLUME XLIII JANUARY 1992

RELIABILITY IN ELECTRONICS

News – Events

JOURNAL ON COMMUNICATIONS

A PUBLICATION OF THE SCIENTIFIC SOCIETY FOR TELECOMMUNICATIONS, HUNGARY

Editorial office
Gábor Áron u. 65.
Budapest, P.O.Box 15.
Hungary, H–1525
Phone: (361) 135-1097
(361) 201-7471
Fax: (361) 135-5560

Subscription rates
Hungarian subscribers
1 year, 12 issues 2900 HUF, single copies 360 HUF
Individual members of Sci. Soc. for Telecomm.
1 year, 12 issues 480 HUF, single copies 60 HUF
Foreign subscribers
12 issues 90 USD, 6 English issues 60 USD, single copies 15 USD
Transfer should be made to the Hungarian Foreign Trade Bank,
Budapest, H–1821, A/C No. MKKB 203-21411

JOURNAL ON COMMUNICATIONS is published monthly, alternately in English and Hungarian by TYPOTEX Ltd. H–1015 Bp. Batthyány u. 14., phone: (361) 201-3317. Publisher: Zsuzsa Votisky. Type-setting by TYPOTEX Ltd. Printed by HUNGAPRINT, Budapest, Hungary
HUISSN 0866-5583

EDITORIAL

At present it is a basic task of the technical–scientific–economic life to put into practice the newest results of electronics, considering also the cost factors. One of the most important conditions for introducing new development results is to assure the dependability a new term for the description of reliability, availability, maintainability and maintenance support) of items (components, units, equipments, systems and services) during all phases of their life cycle.

Since 1964, the Scientific Society for Telecommunications has organized a series of international symposia on reliability in electronics. The RELECTRONIC symposia were giving a survey of reliability theory and practice in the field of electronics. At each symposium the new results and the actual problems of dependability have been reviewed. The 8th Symposium on Reliability in electronics was held in August 1991 in Budapest with sessions on: Reliability Theory; Quality of Services; Network Reliability; Software Reliability; Failure Physics; Reliability Test Methods. In addition to the 96 papers presented in these sessions, two round table discussions were held in the following topics:

- Standardization in the field of dependability
- Service quality

In this special issue we would like to give an overall review on the main topics of the conference, therefore 6 papers were selected for publication. For the sake of illustration, these papers discuss the problems of robot reliability data collection, the evaluation of service and network reliability, performance-reliability analysis (stochastic reward models), interval availability analysis of large Markovian systems, and software quality. The other papers are also valuable and interesting and they can be found in the Proceedings of the Symposium which can be ordered from the Scientific Society for Telecommunications (HTE), H-1372 Budapest, P.O.Box 451, Hungary.

From the papers and lively discussions between the experts from 26 countries, the following main conclusions can be drawn:

In reliablity theory, great attention should be paid to develop reliability-performance measures (stochastic reward models) for characterizing the dependability of complex systems and networks by using mathematical models of non-constant failure intensity, Bayesian estimations and the application of different statistical tests.

Dependability has become known as the science of predicting, optimizing and estimating the dependability measures of components and systems. Therefore, an automated data collection system should be developed and dependability data should be established for obtaining the relevant information.

The quality assurance of services is considered as a basic requirement. The users' needs should be satisfied at an economic cost level. In the future the higher quality should be paid for by the users.

In the field of electronic components, the electromigration processes leading to failures should be investigated. It is very important to establish a relationship between the production yield and reliability. The importance of failure analysis and short time tests has been emphasized.

A great attention should be paid to the dependability design of services, networks and complex systems using the reliability data of components. The required quality, and dependability of services based on users' needs can be achieved by suitable allocation of the dependability parameters. The optimum allocation policy is a shophisticated technical and economic problem. The background of this work is the calculation method combined with the performance of the network topology and the reliability of the elements. Different allocation methods were developed for different types of the networks.

In the dependability design an automated data collection system can be used advantageously for obtaining the necessary informations.

The changes in policy and economy in the Eastern Countries have a considerable effect on the field of dependability and quality. Presently, the market determines the dependability requirements for services and products. Therefore, system dependability should meet the users' need.

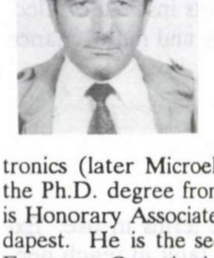
In the field of dependability, poor communication can be observed between the reliability experts and the top management of companies. These communication lines should be improved by means of training and education courses, a better standardization strategy, conferences and seminars.

A companywide dependability management and assurance procedure should be introduced.

The next RELECTRONIC Symposium will be held in 1994, and the International Advisory Board recommended that the efforts should be focused on key topics: Quality and Reliability of Components (joind to the ESRET '94 Conference); System Dependability Including Safety and Security in the Telecommunication Industry.

These conferences will be organized simultaneously by the Scientific Society for Telecommunications. Finally, I hope this Special Issue, completed with informations on the reliability of products and services, business, research and education news, will be of interest to all experts from academic and industrial environments who are actively working in the fields of reliability, maintainability, availability and maintenance support of products and services.

A. BALOGH



Albert Balogh received the B.Sc. degree in mathematics at Kossuth Lajos University, Debrecen in 1957. Since 1961 he has been dealing with the reliability testing of electronic components and reliability estimation problems. He has published about 80 papers in these topics. From 1961 to 1991 he was the senior research worker and later the head of reliability department in the Industrial Research Institute for Electronics (later Microelectronics Company). In 1985 he received the Ph.D. degree from the Hungarian Academy of Sciences. He is Honorary Associate Professor of the Technical University, Budapest. He is the secretary of the Reliablity Committee of the European Organization for Quality (EOQ) and the chairman of the Reliability Section of the Hungarian Telecommunications Society. In 1981 he received the Quality Award of EOQ.

ROBOT RELIABILITY: A REVIEW

B.S. DHILLON
DEPARTMENT OF MECHANICAL ENGINEERING
UNIVERSITY OF OTTAWA
OTTAWA, ONTARIO K1N 6N5
CANADA

The scope of reliability engineering is extremely wide, encompassing many areas of engineering technology. Reliability engineering helps ensure the success of space missions, maintains the national security, delivers a steady supply of electric power, provides reliable transportation, and so on. There has been a considerable growth of knowledge in several areas of reliability engineering and its applications. Robot reliability is one of the application areas. This paper discusses various aspects of robot reliability. Some of these are robot related failure terms, robot failures, published literature on robot reliability, general categories of robot failures and their protection, robot effectiveness and hazard detectors' fail safe design.

1. INTRODUCTION

The subject of robot reliability is very complex and there are numerous interlocking variables in evaluating and accomplishing various reliability levels. A successful robot installation has to be safe and reliable. A robot with poor reliability leads to many problmes: high maintenance cost, unsafe conditions, inconvenience, and so on. Nevertheless, the American National Standard for Industrial Robots and Robot Systems-Safety Requirements [1] specifically calls for the design and construction of robots in such a way that any single, reasonable foreseeable failure will not lead to robot's hazardous motion. There are many different types of parts which are used in robots: electrical, electronic, hydraulic, pneumatic, and mechanical. This makes the task of producing highly reliable robots rather a challenging one. Furthermore, environments in which robots have to operate may be harsh and may vary enormously from one installation to another even for identical models.

The expected useful life of robots is at least 40,000 operating hours and mean time between failures (MTBF) at least 400 hours. In addition, the mean time to repair (MTTR) of 8 hours or less is desirable [2], [3]. The best MTBF achieved for robots so far is only in the order of 2,500 hours [4], [5]. However, over the 10-year period, the Unimate robots have demonstrated the availability of 0.98 [3]. The yearly cost of maintenance associated with robots on average is approximately 11 percent of the procurement cost. The mechanisms of robots instead of electronics account for most of the downtime and maintenance costs [6].

2. ROBOT RELATED FAILURE TERMS

There are many robot related failure terms in use. Examples of these terms are erratic robot, fault in teach pendant, and graceful failure. In robot reliability studies, the clear understanding of such terms is essential. Some of these terms are described below [6], [7], [41].

Erratic robot:	This means a robot moved off its defined path.
Fuses blown:	This means fuses in controls of the robot failed due to various causes, including failures elsewhere.
Fault in teach pendant:	This means part failure in the teach pendant of a robot.
Graceful failure:	This means that the performance of the manipulator degraded in a slow manner in response to overloads instead of failing catastrophically.
Robot out of synchronization:	This means the position of the robot arm is not in line with the robot's memory where it is supposed to be.
Stiffness in robot:	This means mechanical difficulties in the robot's arm to thwart motion—leading to controls terminating the motion.
Robot in emergency stop:	This means controls brought the robot's arm to a standstill.
Fault in cabinet:	This means a recognized failure in the control cabinet's elements. One example of this type of failure is a faulty circuit board.
Overheated hydraulics:	This means power failed due to overheating of the hydraulic power pack which has frozen the motion of the robot.
Robot failed to move:	This means the robot was unable to move without any reason—possibly due to a software fault.

3. SOME ROBOT FAILURES

Two most important considerations concerning robot reliability are the danger of robot releasing the workpiece it is holding and the danger of robot carrying out sudden moves because of its control system failure. A Japanese automobile assembly plant has conducted a study of problems associated with its robots in 1982. The study revealed robot malfunctions such as follows [8], [9], [41]:

i) A robot made a motion outside its defined program.
ii) A robot's arm unexpectedly shot up (beside the fact it had been operating normally earlier) during a hot summer day.
iii) A robot's arm unexpectedly sprang up as the oil-pressure source was being cut off at the end of robot's work.

iv) A robot destroyed the welding work due to a programming instruction error when functioning alone.
v) The slewing shaft unexpectedly swung away from its previously programmed path during a training course.
vi) A robot began moving immediately after the activation of its power source even though its interlock conditions were not in a ready state.
vii) After the activation of the power switch, the arm of the robot sprang out as well as the slewing shaft rotated and stopped after the entanglement with the welding machine.

There were many causes of robot malfunctions such as unexpected starts. These include printed circuit board troubles, oil pressure valve problems, encoder-related malfunctions, servo-valve troubles, human errors, and noise. According to Refs. [4], [10], the robot problems followed the following order:

i) Control system troubles
ii) Jigs' and other tools' incompatibility
iii) Robot body troubles
iv) Errors associated with programming and operation
v) Welding gun problems and those of other tooling parts
vi) Deterioration, precision deficiency
vii) Runaway
viii) Miscellaneous

The reported [8] field meantime between failures (MTBF) in hours and corresponding frequency for robots were 100 or less (28.7%), 250–500 (19.5%), 500–1000 (14.7%), 100–250 (12.2%), 1000–1500 (10.4%), 2500 or over (8.5%), 1500–2000 (4.9%), and 2000–2500 (1.2%). This means that approximately 90% of robots studied have meantime between failures 2,000 hours or less.

Another study [8], [10] of a Japanese automobile assembly plant reported many types of failures associated with spot-welding robots: servo valve, seal, and pipe failures; welding gun failures; gear, shaft, and cover failures; electrical machine failures; power overload, interference, and other control abnormalities; damages to cables; damages to hoses; and correction of dotting position.

4. LITERATURE ON ROBOT RELIABILITY

Many publications concerned with robot reliability have appeared over the years. Fourteen publications related to robot reliability specifically are listed in Ref. [11]. Table 1 presents titles of selected references [2]–[4], [6]–[9], [12]–[38], [41] related to robot reliability. For each of the titles listed, the full reference is given at the end of this paper. Refs. [31], [41] present an extensive list of publications on robot reliability and safety (see Table 1.).

5. GENERAL CATEGORIES OF ROBOT FAILURES AND THEIR PROTECTION

Today's industrial robots make use of electronic, mechanical, hydraulic, and pneumatic parts. This leads to many sources of robot failures resulting in hazards to personnel associated with robots and surrounding equipment. There are basically four categories of failures which affect the robot reliability and its safe operation [12], [14]: random component failures, software failures, human errors, and systematic hardware faults. Failures that occur during the useful life of a component are known as random component failures because they occur unpredictably. Some of the reasons for such failures are undetectable defects, unexplainable causes, unavoidable failures, low safety factors, and so on. Software faults in industrial robots may occur due to reasons such as embedded software, the controlling software or application software. One study [39] reports that over 60% of the total software errors are made during the requirement and design phase as opposed to less than 40% during the coding phase. Robot software faults may be reduced by performing failure mode and effect analysis (FMEA), fault tree analysis, testing, etc. There are several software reliability models [39], [40] which can be utilized to evaluate reliability when the software under consideration is put into its operational use. In order to protect against software failures, the redundancy offers the best solution, even though it is an expensive venture. Data redundancy can be used to make hierarchical and multiprocessor systems more reliable. The following one or more types of redundant information should be included in messages exchanged between computers [26]:

i) Cyclic redundancy check characters
ii) Error-detecting codes
iii) Parity bits
iv) Message sequence numbers
v) Checksums
vi) Addresses of sender and receiver

Another effective failure test is timeout. This could be used in the interface hardware between an industrial robot and its associated controlling computer. For example, the interface immediately stops the robot in the event the computer is unable to dispatch to the robot interface a keep-alive signal, say, every 100 milliseconds. In addition, software can also be implemented with timeouts.

An additional method to detect software failures is status check. In this case, one computer dispatches certain data to another computer capable of sensing if the data are self-consistent.

Human errors are due to personnel who design, manufacture, test, operate, and maintain a robot. One study [42] reports that 20 to 50% of all equipment failures were due to some kind of human error. Some of the reasons for the occurrence of human error are poor equipment design, task complexity, poorly written maintenance and operating procedures, poor training of operating and maintenance people, improper tools, high temperature in the work zone, and inadequate lighting in the work area. Human errors may be broken into categories such as design errors, inspection errors, maintenance errors, assembly errors, installation errors, and operating errors. Some of the methods used for the reduction of the occurrence of human errors are as follows:

i) Man–machine systems analysis
ii) Error–cause removal program
iii) Quality-control circles
iv) Fault trees

All four methods are described in Ref. [42]. Systematic hardware faults are those failures which happen because of unrevealed mechanisms present in the robot design. Reasons such as peculiar wrist orientations and unusual joint to straight line mode transition may lead the robot to ignore a certain task or to execute only specific parts of a program. Failures to make necessary environmental provisions in the initial design could also lead to robot problems. Generally the weak link in a robot hydraulic system is the servo valve because dirt in the hydraulic fluid could case

Table 1. Titles of selected references related to robot reliability

No.	Reference Title
1	Reliability of Industrial Robots: A Safety Viewpoint [12]
2	People and Robots: Their Safety and Reliablity [7]
3	RAM of Robots: Reliability, Availability, Maintainability [13]
4	Fault Tree Analysis of Hazards Created by Robots [4]
5	Safety Control on Introduction of Industrial Robots to Factories [9]
6	Safety Measures of Industrial Robots [8]
7	Reliability Assessment of Industrial Robots [14]
8	Three Million Hours of Robot Field Experience [15]
9	Reliable Real-time Robot Operation Employing Intelligent Forward Recovery [16]
10	Error Recovery in Robot Systems [17]
11	Towards Automatic Recovery in Robot Programs [18]
12	Assessment of Industrial Programmable Electronic Systems with Particular Reference to Robotics Safety [19]
13	Planning for Robot Installation and Maintenance: A Safety Framwork [20]
14	The Robotics Revolution (Chapter 15 – Safety and Reliability) [21]
15	Robot Design Handbook (Chapter 8) [6]
16	Robotics in Practice (Chapter 5 – Reliability, Maintenance and Safety) [2]
17	Robots: Safe or hazardous [22]
18	Robotics (Chapter 16 – Robot Reliability) [23]
19	Reliability in Industrial Robots for Spray Gun Application [24]
20	Robotics Applications for Industry (Chapter 6, Section 6.3 – Safety Considerations) [25]
21	Robot safety Considerations [26]
22	Elements of Industrial Robotics (Chapter 8, Sections 8.2.2 and 8.2.3 – Hardware Failure, and malfunction, control system failure and malfunction) [27]
23	Industrial Robotics (Chapter 17 – Safety, Training, Maintenance, and Quality) [28]
24	Reducing Downtime via Warm Restarts [29]
25	Artifical Intelligence Applied to Robot Fail Safe Operations [30]
26	On Robot Reliability and Safety-Bibliography [31]
27	Structural Risk Analysis in Robot Design [32]
28	Analysis of First UTD Installation Failures [33]
29	Flexible Assembly Systems (Chapter 9 – Failure Analysis) [34]
30	Towards Developing Reliability and Safety Related Standards Using Systematic Methodologies [35]
31	The Role of Hardware, Software and People in Safeguarding Robot Production Systems [36]
32	The Impact of Robots on Product Reliability [37]
33	Introduction to Robotics (Chapter 3, Section 3 – Life Expectancy, Reliability, and Maintainabiliy) [3]
34	Reliability of Basic Robot Automated Manufacturing Systems [38]

the spool to stick in an open mode, leading to arm's uncontrolled motion. Two guidelines to deal with the servo valve problem are as follows [26]:

i) Make spool to rotate on a continuous basis (or back and forth) around its axis freely of its normal control motion along that axis. This is useful in two ways: detection of the valve clogged by dirt in the fluid becomes possible; and the static friction in the valve reduces to zero because of the rotational motion.

ii) Introduce a redundant on – off control valve in each servo valve's feed line. An advantage of this action is that in the event the servo valve fails to close, the redundant valve would stop movement of the robot arm.

Other protection against robot hardware failures is inclusion of sensors in the system for detecting loss of pneumatic pressure, line voltage, or hydraulic pressure. In addition, excessiveness of the following items:

i) Force
ii) Speed
iii) Servo errors
iv) Acceleration
v) Temperature

Several techniques useful for reducing systematic hardware failures are described in Ref. [40].

6. ROBOT EFFECTIVENESS

There are many factors which may dictate the effectiveness of a robot. Among those are [23]

i) Availability and quality of the robot repair facilities and equipment
ii) Robot meantime between failures (MTBF)
iii) Robot meantime to repair (MTTR)
iv) Availability and quality of manpower needed to keep the robot in its working state
v) The percentage of time the robot functions normally
vi) Rate of the required spare parts' availability
vii) The relative robot performance under extreme situations
viii) The percentage of time the robot is ready for operation

From the reliability point of view, the designer should keep in mind that when light alloys are used in the robot construction, it would be very wise to limit fatigue risks. This could be achieved by designing the ultimate stress to be at least eight times the actual working stress.

7. HAZARD DETECTORS' FAIL SAFE DESIGN

A safety device associated with a robot is composed of a hazard detection sensor, electrical circuits, and other parts. Such items are subject to failures. To guard against sensor failures, new equipment may be installed for simulating conditions the hazard detection sensor is supposed to detect [26]. Periodically, the newly installed equipment would challenge the system used for detection, and then would test for the detection of each and every challenge. The safety device generates a warning signal when the sensor fails to respond to a challenge or it responds when there is no challenge. The fail-safe hazard detector is made up of three subsystems:

i) Monitor subsystem: This subsystem keeps a close eye on any interruption of the challenge-and-response sequence.
ii) Sensor subsystem: This subsystem detects any hazardous conditions.
iii) Challenge subsystem: This subsystem drives the sensor.

REFERENCES

[1] *American National Standard for Industrial Robots and Robot Systems: Safety Requirements*, ANSI/RIA r15.06-1986, June 1986. American National Standards Institute, Inc., 1430 Broadway, New York, New York 10018.

[2] J.E. Engleberger, *Robotics in Practice*, Kogan Page Ltd., London, 1980.

[3] A.J. Critchlow, *Introduction to Robotics*, Macmillan Publishing Company, New York, 1985.

[4] N. Sugimoto and K. Kawaguchi, "Fault Tree Analysis of Hazards Created by Robots.", *Proceedings of the 13th International Symposium of Industrial Robots*, 1983, pp.9.13 – 9.28.

[5] B.S. Dhillon, *Mechanical Reliability: Theory, Models and Applications*, American Institute of Aeronautics and Astronautics, Inc., Washington, D.C., 1988.

[6] G.B. Andeen (Editor-in-Chief), *Robot Design Handbook*, McGraw-Hill Book Company, New York, 1988.

[7] R. Jones and S. Dawson, "People and Robots: Their Safety and Reliability", *Proceedings of the 7th British Robot Association Annual Conference*, 1984, pp.243 – 258.

[8] Research Institute of Industrial Safety, Japanese Ministry of Labour, Safety Measures of Industrial Robots — Part II, *Japan Electronic Engineering*, January 1983, pp.88 – 94.

[9] S. Shima, "Safety Control on Introduction of Industrial Robots to Factories", *Safety*, Vol. 33, 1982, pp.28.

[10] K. Sato, "Case Study of Maintenance of Spot-Welding Robots", *Plant Maintenance*, Vol. 33, 1982, pp.18.

[11] R.D. Kilmer, "Robot Safety Research: List of selected references and research organizations", *Proceedings of the Robotic Industries Association(RIA) Robot Safety Seminar*, Boston, Oct. 1985, pp.55 – 65.

[12] K. Khodabandehloo, F. Duggan and T.F. Husband, "Reliability of Industrial Robots: A Safety Viewpoint". *Proceedings of the 7th British Robot Association Annual conference*, 1984, pp.233 – 242.

[13] B.W. Pollard, "RAM for Robots: Reliability, Availability, Maintainability", *Robotics Today*, 1981, pp.51 – 53.

[14] K. Khodabandehloo, F. Duggan and T.F. Husband, "Reliability Assessment of Industrial Robots", *Proceedings of the 14th International Symposium on Industrial Robots*, 1984, pp.209 – 220.

[15] J.F. Engelberger, " Three Million Hours of Robot Field Experience", *The Industrial Robot*, 1974, pp.164 – 168.

[16] R.E. Smith and M. Gini, "Reliable Real-time Robot Operation Employing Intelligent Forward Recovery", *Journal of Robotic Systems*, Vol. 3, 1986, pp.281 – 300.

[17] S. Srinivas, *Error Recovery in Robot Systems*, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1977.

[18] M. Gini and G. Gini, "Towards automatic Recovery in Robot Programs", *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, 1983, pp.821 – 823.

[19] R. Bell, B.K. Daniels and R.I. Wright, "Assessment of Industrial Programmable Electronic Systems with Particular Reference to Robotics Safety", *Proceedings of the 4th National Reliability Conference*, U.K., 1983, pp. 3C/1/1-3C/1/11.

[20] R.J. Barrett, R. Bell and P.H. Hodson, "Planning for Robot Installation and Maintenance: A Safety Framework", *Proceedings of the 4th British Robot Association Conference*, May 1981, pp.10 – 15.

[21] P.B. Scott, *The Robotics Revolution*, Basil Blackwell Publishers Ltd., Oxford, U.K., 1984, pp.233 – 245.

[22] W.W. Stowe, "Robots: Safe or Hazardous?", *Professional Safety*, 1983, pp.32 – 35.

[23] J.F. Young, *Robotics*, Butterworth & Co. (Publishers) Ltd., London, 1973, pp.274 – 290.

[24] K.M. Haugan, "Reliability in Industrial Robots for Spray Gun Applications", *Proceedings of the 2nd Conference on Industrial Robot Technology*, 1974, pp. E7/93 – E7/98.

[25] L.L. Toepperwein, M.T. Blackmon, W.T. Park, W.R. Tanner and W.F. Adolfson, "Robotics Applications for Industry", *Noyes Data Corporation*, Park Ridge, New Jersey, 1983, pp.116 – 122.

[26] W.T. Park, "Robot Safety Suggestions", *Technical Note* No. 159, Prepared by SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025, April 1978.

[27] B. Leatham-Jones, *Elements of Industrial Robotics*, Pitman Publishing, London, 1987, pp.199 – 201.

[28] M.P. Groover, M. Weiss, R.N. Nagel and N.G. Odrey, *Industrial Robotics*, McGraw-Hill Book Company, New York, 1986, pp.469 – 485.

[29] D. Elbert, "Reducing Downtime via Warm Restarts", *Robotics Engineering*, October 1986, pp.14 – 16.

[30] C.A. Ramirez, "Artificial Intelligence Applied to Robot Fail Safe Operations", *Proceedings of the Robots-9*, June 1985, pp. 19.21 – 19.37.

[31] B.S. Dhillon, "On Robot Reliability and Safety – Bibliography", *Microelectronics and Reliability*, Vol. 27, 1987, pp.105 – 118.

[32] A. Longinow, "Structural Risk Analysis in Robot Design", *Proceedings of the 5th International Symposium on Industrial Robots*, 1975, pp.549 – 556.

[33] G.C. Macri, "Analysis of First UTD Installation Failures in Industrial Robots", edited by W.R. Tanner, Published by Robotics International Society of Manufacturing Engineers (SME), One SME Drive, P.O.Box 930, Dearnborn, Michigan, 48128, 1981, pp.74 – 84.

[34] A.E. Owen, "Flexible Assemble Systems", *Plenum Press*, New York, 1984, pp.171 – 190.

[35] F. Duggan, R.H. Jones and K. Khodabandehloo, "Towards Developing Reliability and Safety Related Standards Using Systematic Methodologies in Robot Technology and Applications", edited by K. Rathmill, P. MacConaill, S. O'Leary, J. Browne, Springer-Verlag, Berlin, 1985, pp.90 – 107.

[36] R. Jones, S. Dawson, "The Role of Hardware, Software, and People; in Safeguarding Robot Production Systems", *Proceedings of the 15th International Symposium on Industrial Robots*, 1985, pp.557 – 568.

[37] J. Mihalasky, "The Impact of Robots on Product Reliability", *Proceedings of the Annual Reliability and Maintainability*, 1985, pp.464 – 466.

[38] H. Alayan, C.A. Niznik, R.W. Newcomb, "Reliability of Basic Robot Automated Manufacturing Networks", *Proceedings of the IEEE Southeastcon Conference*, 1984, pp.291 – 294.

[39] B.S. Dhillon, *Reliability in Computer System Design*, Ablex Publishing Corporation, Norwood, New Jersey, 1987.

[40] B.S. Dhillon, *Reliability Engineering in Systems Design and Operation*, Van Nostrand Reinhold Company, New York, 1983.

[41] B.S. Dhillon, *Robot Reliability and Safety*, Springer-Verlag, New York, 1991.

[42] B.S. Dhillon, *Human Reliability*, Pergamon Press Inc., New York, 1986.

Dr. B.S. Dhillon is a full Professor of Mechanical Engineering and Director of the Engineering Management Program at the University of Ottawa. He has published over 210 articles on Reliability Engineering and related areas, and has written twelve books on various aspects of System Reliability, Maintainability and Engineering Management. His books on Reliability have been translated into several languages including Russian, Chinese and German. He has served as General Chairman of two International Conferences on Reliability and Quality Control held in Los Angeles and Paris in 1987.

AUTOMATED COLLECTION OF TELECOM SYSTEM DEPENDABILITY DATA

K. STRANDBERG
PRODUCT MANAGEMENT DIVISION
ERICSSON TELECOM AB
S-126 25 STOCKHOLM, SWEDEN

Further development of Ericsson's dependability analysis and calculation software system DEPEND in accordance with the Hayward concept will provide an almost fully automated collection and analysis of field data for telephone exchanges. Data can be collected by means of "electronic log books" followed by data storage and analysis including estimation, hypothesis testing and specially developed charting techniques. The new tools give all personnel categories, from field maintenance staff to design engineers and product managers, a powerful means for management of the operational exchanges dependability, and an effective tool for dependability improvement programmes.

1. INTRODUCTION

Field data collection, evaluation and corrective action are essential tasks in a dependability programme such as that specified by International Standard IEC 300, and are necessary complements to a quality system, e.g. to that stipulated by the ISO 9000 series of International Standards (see Refs. [9], [10], [13] and [14]).

Everyone in the dependability profession recognizes the need for reliable field data for the supervision of quality and dependability of telecommunication networks and systems. Most of us — suppliers and customers alike — have experienced great problems in actually collecting data, from convincing the individual maintenance engineer in the field that he/she has to fill in forms, to persuading unions that this data collection will not be a cover for employee efficiency supervision.

Many have also been frustrated by the fact that even if data is collected (one way or another), analysis resources may not be available or are not effectively used. The management therefore, earnest in its quest for quality and customer/user satisfaction, too often takes action based on other indications than those provided by evaluated dependability data.

Another problem is that data collection relies on many individuals feeding information into a central storage, analysis and action point. The whole idea of data collection and evaluation has to be turned "upside-down". The responsibility for the entire process of collection, evaluation and taking action should be as close as possible to the source of data, viz. the telephone exchanges and other kinds of telecom equipment in field operation. Centralized storage and analysis must be provided on an "as needed" basis.

2. DEPENDABILITY PLANNING — COMPLEMENT TO QUALITY PLANNING

Continuous improvement is inherent to the concept of total quality management. This means that the basic purpose of collection and evaluation of dependability data from field operation of our products is, beside improvements of the product itself, improvement of the operational and maintenance support given to customers and improvement of our internal operational processes.

The planning itself is a process repeated each year. Evaluated field data is one of the fundamental inputs to this process. Other inputs are stated customer requirements, customer complaints and results of customer satisfaction surveys.

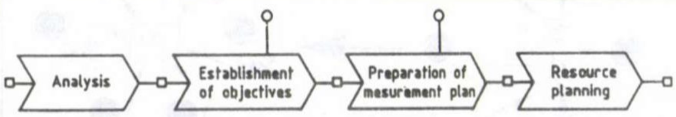


Fig. 1. Process for quality and dependability planning

3. DATA COLLECTION PROCEDURE

The basic steps of collection and evaluation of information related to failure events and other relevant events and fault states are (see Fig. 2 and Refs. [1], [3], [6] and [12]):

- Collection of all information relevant to the occurrence of the event and the related maintenance activities.
- Storage of the collected information in a form suited for evaluation of (individual or congregated) data.
- Data evaluation and preparation of output results.

The results are subsequently used for determination of action:

- Analysis of the evaluation results and other information relevant to the product or process.

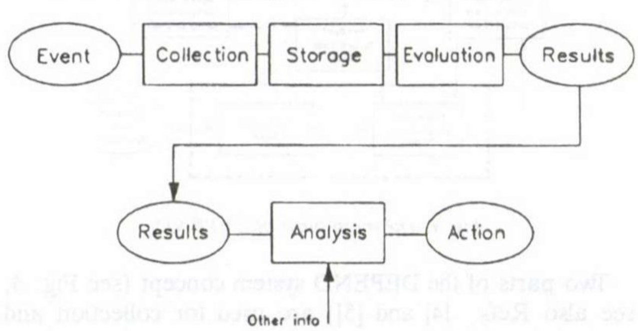


Fig. 2. Process for data collection and evaluation, and taking action

4. THE HAYWARD CONCEPT

As long as systems perform according to specifications, there should normally be no need for extensive centralized data collection and analysis. Deviations from specifications have to be detected and acted upon, rapidly and correctly, but close to field operation. Personnel having responsibility for initiation of corrective actions should be close to the exchange.

This approach will not prevent initiatives and actions for improvements beyond specifications, nor would it prevent centralized data analysis and actions, if and when required, as data can be solicited from the remote data storage locations.

We call our approach for applying these principles the "Hayward concept" (a hayward was in the older days, according to the dictionary, "one who was in charge of fences and enclosures and prevented cattle from breaking through; one who herded the common cattle of a town"). In our case, the Hayward will be that person, within the customer's or our own organization, responsible for "herding" or caring about a given set of telephone exchanges.

This means that the responsibility for data collection, storage and evaluation, supported by the VERA part of DEPEND, is moved from one central location to many decentralized locations.

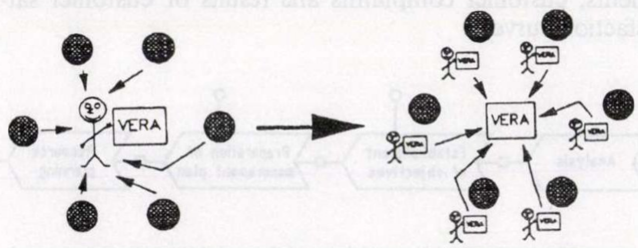


Fig. 3. The Hayward concept leads to a decentralized structure

The dependability analysis system DEPEND is used for automated data storage and calculation of various measures of reliability performance, maintainability performance, availability performance and maintenance support performance for all our products.

In order to achieve the above objectives the field oriented parts of DEPEND (Ref. [4]) have been reconsidered and revised.

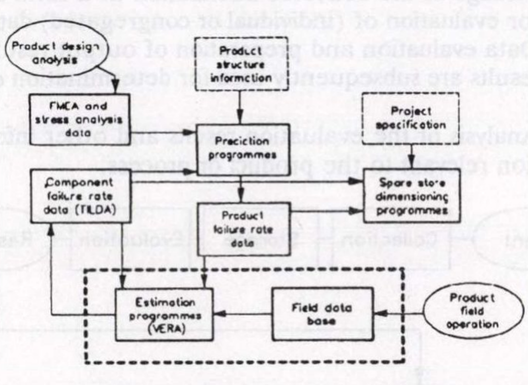


Fig. 4. The main parts of DEPEND

Two parts of the DEPEND system concept (see Fig. 4; see also Refs. [4] and [5]) are used for collection and evaluation of field data, viz.

- a set of estimation programmes (called VERA),
- a field data base containing data collected from field operation.

The fully implemented system will offer four levels of automation (see Fig. 5): PC or work station (1) at the exchange site (AXE); (2) at the customer's operation and maintenance centre (OMC); (3) at an Ericsson customer support office (ESO), and (4) at the Ericsson service support centre in Stockholm (ESSC). Any combination of fully or partially automated data collection may be utilized. Communication and data transfer is accomplished via our communication network FSN (Field Support Network) and via the world-wide Ericsson Communication Network.

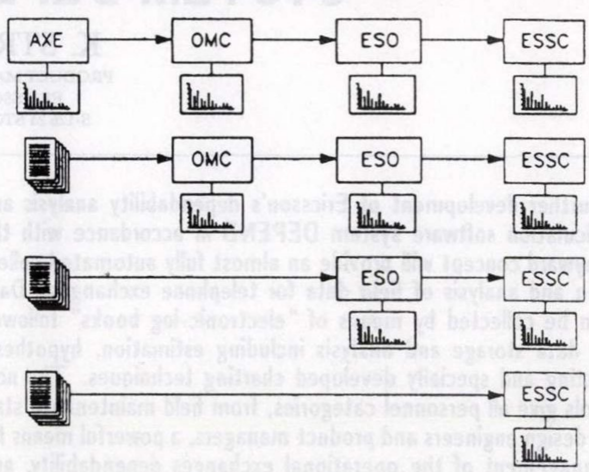


Fig. 5. DEPEND supports various stages of decentralization

5. A HIERARCHY OF MEASURES USED FOR TELECOM

Collection, evaluation and presentation of data for dependability characteristics is done for those measures most commonly specified by Administrations and Operating Companies. These measures take into account effects of software and hardware failures and faults in our exchanges (see also Ref. [5]).

Availability performance:
- Mean accumulated system down time
- Mean accumulated subscriber line down time
- Mean accumulated trunk line down time

Reliability performance:
- Intensity of compmlete system failures
- Intensity of system restarts
- Intensity of hardware failures

Maintainability performance:
- Mean down time
- Mean active repair time

Maintenance support performance:
- Mean administrative delay
- Mean logistic delay

There is a close hierarchical relationship between some of the availability performance measures, as illustrated by Fig. 6.

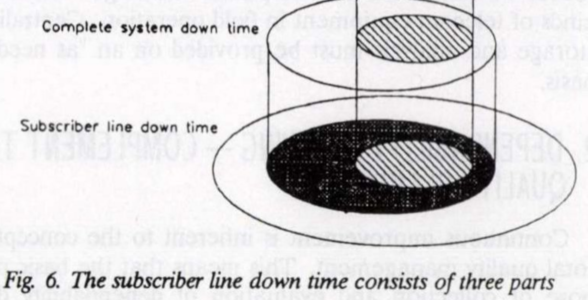


Fig. 6. The subscriber line down time consists of three parts

Complete exchange down time is counted when the exchange has a fault from which it cannot recover without manual intervention.

Complete system down time is related to a complete system fault, i.e. the situation when the entire exchange is unable to perform its function.

Subscriber line down time includes complete system down time and down times at those states when some, but not all, subscriber lines connected to the exchange are inaccessible by the subscriber due to a partial exchange fault.

Trunk line down time includes system down time, but is also related to those states when some, but not all, trunk lines (to other exchanges in the network) are inaccessible due to an exchange fault.

In order to support the analysis of these measures, the following "raw data" has to be collected at each failure event by electronic means and/or complemented and completed by the maintenance personnel via PC or work station screen forms:

1 Failure report number
2 Exchange identity
3 Date and time of failure
4 Down time
5 Corrective action taken
6 Unit affected or replaced
7 Active repair time
8 Number of subscriber lines affected
9 Number of trunk lines affected
10 Type of fault (or type of state, see Fig. 7)
11 Cause (as perceived by maintenance personnel)
12 Supplementary text information (from the maintenance personnel; useful for the subsequent fault analysis)

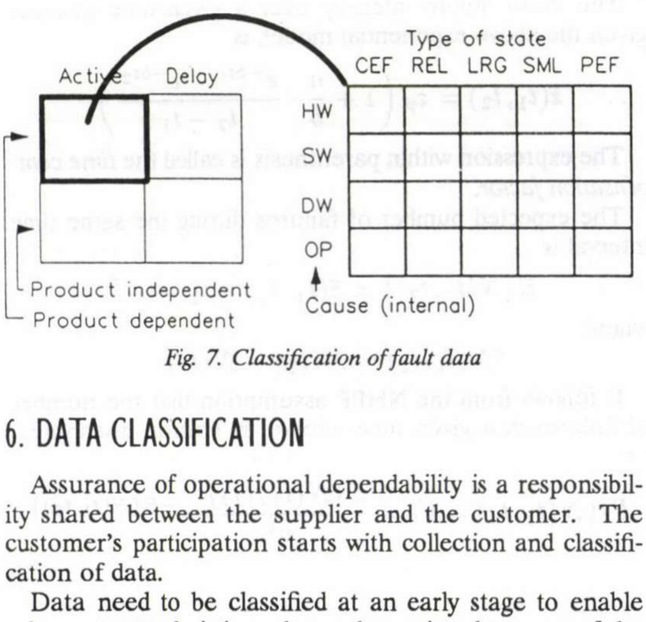


Fig. 7. Classification of fault data

6. DATA CLASSIFICATION

Assurance of operational dependability is a responsibility shared between the supplier and the customer. The customer's participation starts with collection and classification of data.

Data need to be classified at an early stage to enable subsequent analysis in order to determine the cause of the fault and to implement appropriate corrective or future preventive actions.

Fault data is first classified in terms of product dependency and active/not active (delay) parts of down time. A further subdivision is done in terms of type of state (the system AXE specific states CEF (complete exchange fault), REL (system reload), LRG (large system restart), SML (small system restart), PEF (partial exchange fault)) and location of the cause (hardware (HW), software (SW), documentation (DW), operations personnel (OP)).

Data retrieval and manipulation (data reduction and calculation of intermediate parameters) is done on a monthly basis and results are transferred to the centralized mainframe storage utilities in Stockholm.

Use of a relational data base system and a query language facilitates sorting, manipulation and evaluation of data.

7. STATISTICAL EVALUATIONS

Statistical evaluations may be performed at each of the four organizational levels indicated in Fig. 5.
These statistical evaluations include

- point and interval estimation
- compliance evaluation (hypothesis testing)
- charting with warning limits

The evaluations in VERA further take into account

- the shape of the early failure period
- population size considerations for hardware

Results are presented in tabular and graphic form. The graphic form is preferred for overview presentations and executive reports.

8. EVALUATION EXAMPLE: FAILURE INTENSITY

As an example of the evaluation methods and procedures we in this paper use failure intensity evaluation of hardware. For this evaluation we need a failure rate model, a population spread model and a failure intensity model.

The evaluation is performed at a supervisory level, with an especially developed supervision chart (see description later), and as hypothesis testing and point and interval estimations.

Time models are used to describe the time-dependency of failure intensities. These time-dependencies influence evaluation of measures of reliability performance and availability performance.

Population spread models are used to describe the effects of population non-homogeneity as regards reliability performance measures.

Hypothesis testing and supervision charting may be done with an acceptable value or a reference value (as the case may be) taken from predictions, specifications, or specifically chosen at each evaluation (e.g. to adapt to a particular customer or contract condition).

9. FAILURE RATE AND POPULATION SPREAD MODELS

We assume that the instantaneous failure rate as defined below (see Refs. [7] and [8]) is constant for each individual printed circuit board, i.e. that

$$\lambda(t) = \lambda \quad \text{for all values of} \quad t.$$

Instantaneous failure rate: The limit of the ratio of the conditional probability that the instant of time of a failure of an item falls within a given time interval, $(t, t + \Delta t)$, when the length of this interval, Δt, tends to zero and given that the item is in an up state at the beginning of the time interval, i.e.

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{Pr(t < T \leq t + \Delta t \mid T > t)}{\Delta t}$$

We assume that different items have different failure rate values. The randomness caused may for instance be due to variations in the manufacturing process. We may thus describe the population by a function with a shape as given by Fig. 8.

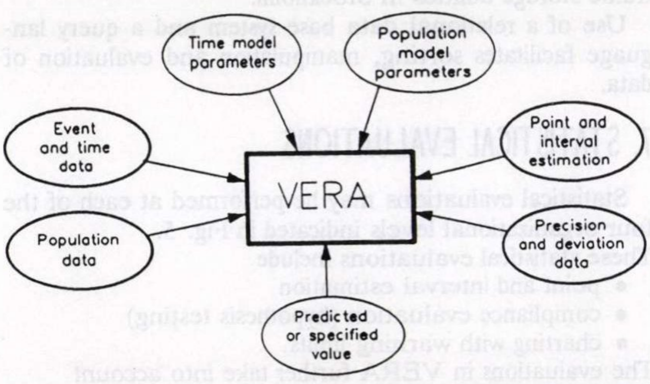


Fig. 8. The VERA program for failure intensity evaluation takes into account both the characteristics of the time model and the population model. Output is both in the form of tables and charts

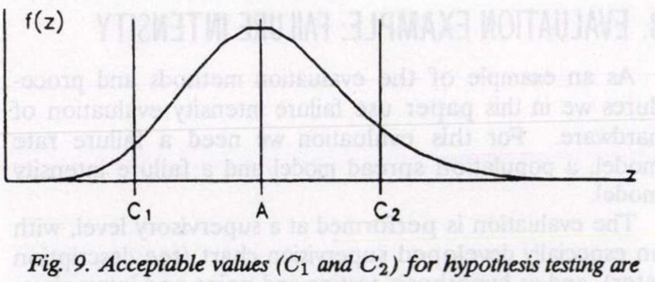


Fig. 9. Acceptable values (C_1 and C_2) for hypothesis testing are determined from the expected failure rate as $C_1 = A/\kappa(n)$ and $C_2 = A \cdot \kappa(n)$, where A is the population mean

We call the permitted relative deviation (from the mean value, which may be represented by the predicted mean value) the *population spread factor*, $\kappa(n)$, with a value dependent on population size, n, determined as $\kappa(n) = \alpha + \beta e^{-\gamma\sqrt{n}}$.

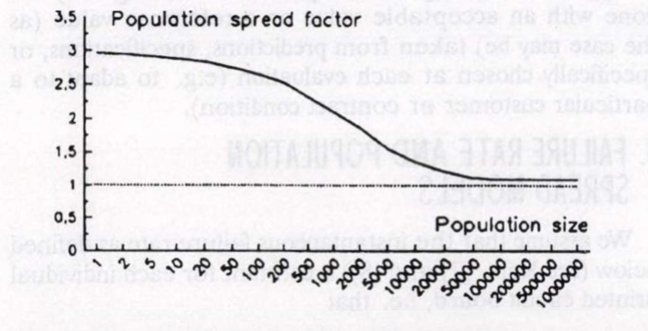


Fig. 10. The population spread factor, $\kappa(n)$, used for printed board populations (and sub-populations)

10. FAILURE INTENSITY MODEL

The failure process of a set of boards in operation is assumed to constitute a Poisson process with time dependent intensity parameter, i.e. a non-homogeneous Poisson process (NHPP). The intensity parameter of the process is the *instantaneous failure intensity*, that is the limit of the ratio of the mean number of failures of a repaired item in a time interval $(t, t + \Delta t)$, and the length of the interval Δt, when the length of the interval tends to zero, i.e.

$$z(t) = \lim_{\Delta t \to 0+} \frac{E\{N(t + \Delta t) - N(t)\}}{\Delta t}$$

where $N(t)$ is the number of failures in the time interval $(0, t)$ (see Ref. [8]).

We also need the concept of *mean failure intensity*, that is the the expected number of failures in a given time interval, (t_1, t_2), divided by the length of the time interval, i.e.

$$\overline{z}(t_1, t_2) = \frac{E\{N(t_2) - N(t_1)\}}{t_2 - t_1}$$

or, equivalently,

$$\overline{z}(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} z(t)dt$$

The time model chosen for the population failure intensity during field operation takes in a tractable manner into account a higher intensity during the early period and an approximately constant intensity during the subsequent period. The model is also used for software failure processes (see Ref. [2]).

$$z(t) = z_p(1 + ae^{-bt})$$

where a and b are non-negative real constants.

The failure intensity function starts at a value $z_p(1+a)$ and reaches z_p asymptotically at rate b.

z_p is thus the constant (aggregated) failure intensity of the set of n boards during the constant failure intensity period, predicted on the basis of the component failure rates.

The mean failure intensity over a given time interval, given the above exponential model, is

$$\overline{z}(t_1, t_2) = z_p \left(1 + \frac{a}{b} \cdot \frac{e^{-bt_1} - e^{-bt_2}}{t_2 - t_1}\right)$$

The expression within parenthesis is called the *time compensation factor*.

The expected number of failures during the same time interval is

$$E\{N(t_1, t_2)\} = \overline{z}(t_1, t_2) \cdot (t_2 - t_1)$$

where

$$N(t_1, t_2) = N(t_2) - N(t_1)$$

It follows from the NHPP assumption that the number of failures in a given time interval is Poisson distributed, i.e.

$$Pr(N(t_1, t_2) = \nu) = \frac{E\{N(t_1, t_2)\}^{\nu}}{\nu!} e^{-E\{N(t_1, t_2)\}}$$

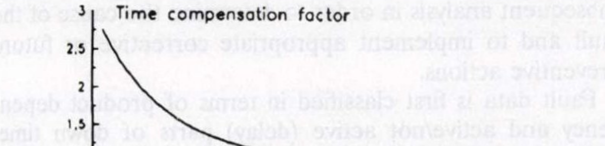


Fig. 11. The time compensation factor for printed circuit boards

11. ESTIMATION AND TESTS FOR DEVIATION

As point estimate of mean failure intensity we use

$$\overline{z}^*(t_1, t_2) = \frac{r}{t_2 - t_1} \qquad r_j > 0$$

$$\overline{z}^*(t_1, t_2) = \frac{1}{3(t_2 - t_1)} \qquad r_j = 0$$

where r is the number of failures observed.

Interval estimates are calculated with the chi square distribution at a confidence level $\alpha \cdot 100$ percent.

One-sided, upper limit

$$\left[0, \ \frac{\chi^2(2r+2, \alpha)}{2 \cdot (t_2 - t_1)}\right]$$

One-sided, lower limit

$$\left[\frac{\chi^2(2r, 1-\alpha)}{2 \cdot (t_2 - t_1)}, \infty\right]$$

Two-sided

$$\left[\frac{\chi^2\left(2r, \frac{1-\alpha}{2}\right)}{2 \cdot (t_2 - t_1)}, \ \frac{\chi^2\left(2r+2, 1-\frac{1-\alpha}{2}\right)}{2 \cdot (t_2 - t_1)}\right]$$

with $\alpha = 0.9$.

The deviation from the predicted failure intensity is determined by a two-sided hypothesis testing procedure at a significance level $= 0.1$. The precision is considered high if the confidence interval is shorter than the point estimate, which means that the number of failures must be at least 13 (see Ref. [5]).

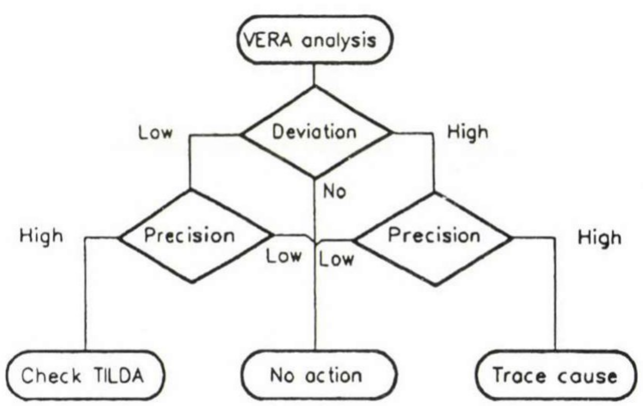


Fig. 12. Flowchart for test for deviation

12. THE SUPERVISION CHART

The purpose of the supervision chart is to serve as a quick visual aid for determination of whether the observed results are close to or far from what is required or expected. It is equipped with three curves, one each for upper and lower warning limits, and one for the point estimate of the observed failure intensity.

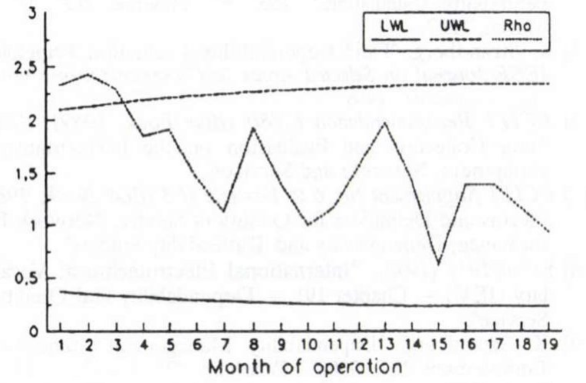


Fig. 13. The supervision chart shows for each month: Lower warning limit (LWL)$=W_U$; upper warning limit (UWL)$=W_L$; and point estimate (Rho) $= \rho$

The y-axis expresses the normalized failure intensity, i.e. the value 1 corresponds to the predicted failure intensity. The upper warning limit is calculated as

$$W_U = \frac{U}{E\{N(j)\}}$$

where U has been determined as the smallest value of U that satisfies the equation

$$\sum_{\mu=1}^{U} \frac{(E\{N(j)\} \cdot \kappa(n))^{\mu}}{\mu!} e^{-E\{N(j)\} \cdot \kappa(n)} \geq 0.95$$

Similarly, the lower warning limit is determined as

$$W_L = \frac{L}{E\{N(j)\}}$$

where L has been given by the largest value of L that satisfies the equation

$$\sum_{\mu=1}^{L} \frac{\left(\frac{E\{N(j)\}}{\kappa(n)}\right)^{\mu}}{\mu!} e^{\frac{-E\{N(j)\}}{\kappa(n)}} \leq 0.05$$

The point estimate of the normalized failure intensity is defined as

$$\rho = \frac{r_j}{E\{N(j)\}} \qquad r_j > 0$$

$$\rho = \frac{1}{3 \cdot E\{N(j)\}} \qquad r_j = 0$$

In the formulas above j denotes that $j : th$ time period (i.e. the $j : th$ month, see Fig. 13). $E\{N(j)\}$ denotes the expected number of failures in the $j : th$ period, with the time compensation factor taken into account.

REFERENCES

[1] O. Silkoset and K. Strandberg, "The Reliability and Maintainability Performance Data Collection Project for the Nordic Public Data Network", *3rd European Reliability Data Bank Seminar,* University of Bradford, 9–11 April, 1980.

[2] K. Strandberg and H. Andersson, "On a Model for Software Reliability Performance", *Microelectronics and Reliability*, Vol. 22, No. 2, pp.227–240, 1982.

[3] K. Strandberg, "Dependability Performance Guarantees — Some Specification and Evaluation Principles", *EUROCON '82*, Copenhagen, 14–18 June, 1982.

[4] K. Strandberg, "Automated Tools for Telecom Systems Dependability Calculations", *ISS '87*, Phoenix, AZ, 15–21 March, 1987.

[5] K. Strandberg, "Field Dependability Evaluation Principles", *IEEE Journal on Selected Areas in Communication*, Vol. 6, No. 8, October, 1988.

[6] *CCITT Recommendation E.880 (Blue Book, 1989)*: "Field Data Collection and Evaluation on the Performance of Equipment, Networks and Services".

[7] *CCITT Supplement No. 6 to Fascicle II.3 (Blue Book, 1989)*: "Terms and Definitions for Quality of Service, Network Performance, Dependability and Trafficability Studies".

[8] *EC50(191) (1990)*: "International Electrotechnical Vocabulary (IEV) — Chapter 191 — Dependability and Quality of Service".

[9] K. Strandberg, "Dependability Management Standard to Complement ISO 9000", *Reliability '91,* London, 10–12 June, 1991.

[10] K. Strandberg, "IEC 300: The Dependability Counterpart to ISO 9000", *Annual Reliability and Maintainability Symposium*, Orlando, F1, 29–31 January, 1991.

[11] K. Strandberg, "Automated Collection of Telecom System Dependability Data", *Relectronic '91*, Budapest, Hungary, 26–30 August, 1991.

[12] IEC 56 (Central Office) 156: "Guidelines for the Collection of Reliability, Maintainability, Availability and Maintenance Support Data from the Field" (1990-03). *International Electrotechnical Commission*, Geneva.

[13] *IEC 56 (Central Office) 188/ISO DIS 9000-4*: "IEC 300-1 — Dependability Management — Part 1: Dependability Assurance for Products"/"ISO 9000-4 — Quality Management and Quality Assurance Standards — Guidelines for Application of Dependability Management" (1991-11).

[14] *IEC 56 (Central Office) 189*: "IEC 300-2 — Dependability Management — Part 2: Dependability Programme Elements and Tasks" (1991-11).

Kjell Strandberg is since 1983 Secretary of IEC/TC 56, which he has served actively since 1974. He is chairman of ITU/IEC JCG WGY (the Committee which prepared the draft for IEV 191 and CCITT Rec. G.106, later to become E.800) and has held many Rapporteurship positions within ITU/CCITT. He is a member of IEC/ACET and is presently (1991) also the Secretary of IEC/ISO JCG QDS and a member of the ETSI ad hoc group on Quality Systems. His career at Ericsson started in 1965 in the field of teletraffic theory and teletraffic engineering, after which he has had various manager assignments for reliability, maintainability and quality within product management for public telecommunication systems.

EVALUATION OF TRAFFIC LOSS IN CASE OF UNRELIABLE CIRCUIT GROUPS

DR. F. RICINIELLO DR. R. DELLEANI

SIP – SOCIETÀ ITALIANA PER L'ESERCIZIO
DELLE TELECOMUNICAZIONI
VIA FALMINIA 189, 00196 ROME, ITALY

A major problem in telecommunications is presented in the evaluation of the traffic loss in case of unreliable circuit groups. Usually the traffic loss is calculated by using the Erlang-B formula which is valid in case of fully reliable circuits. However, the actual traffic loss depends highly on the failure rate of the circuits and on the restoring time. This paper presents a methodology allowing a more accurate evaluation of traffic loss under general assumptions. This methodology leads to a class of functions including as a particular case the Erlang-B formula, assuming fully reliable circuits and Poisson arrival of call attempts. These functions were obtained by assuming a two-dimensional Markov model; closed or recursive formulas have been found due to the reversibility characteristic of the Markov model. The same Markov model has been already proposed in the literature but only numerically approached to solve TASI related problems and to evaluate the performances of some protocols. As an application example, closed or recursive formulas and design diagrams are given for two particular cases.

1. INTRODUCTION

The evaluation of the loss probability of fully reliable circuit groups can be obtained by solving the birth–death process $M/M/m/0/\infty$ that can be represented by a Markovian model (see Fig.1 showing an example with $m = 5$). In particular, the loss probability is given by the probability of having all the group circuits busy, and leads to the well-known Erlang-B formula:

$$E(Ao, m) = \frac{\frac{Ao^m}{m!}}{\sum_{j=0}^{m} \frac{Ao^j}{j!}} \qquad (1)$$

where Ao is the traffic offered to m circuits.

The Erlang-B formula is often employed in traffic engineering for the evaluation of the loss probability of circuit groups, and is valid under the following assumptions:

a) Poisson distribution of the call attempt arrival time;
b) negative exponential distribution of circuit holding time;
c) fully reliable circuits (i.e. no circuit failures).

It has been shown in the literature [1] that assumption b) is not necessary because the loss probability does not depend on the circuit holding-time distribution. As far as assumption c) is concerned, the actual value of the loss probability will be different from the value given by the Erlang-B formula because of circuit failures. In this paper, a methodology for loss probability evaluation in case of unreliable circuits is proposed.

2. ASSUMPTIONS AND NOTATIONS

Assumptions:

1. The trunk group is fully available (in the traffic sense; for definition see reference [2]);
2. The circuits fail one at a time;
3. The failures and traffic processes are statistically independent;
4. Fails may occur only when the circuit is not busy.

Notations:

m	number of circuits;
h	number of faulty circuits ($h = 0 \ldots m$);
k	number of busy circuits ($k = 0 \ldots m-h$);
$S_{h,k}$	state corresponding to h faulty circuits and k busy circuits;
λ_k	transition rate from $S_{h,k}$ to $S_{h,k+1}$ ($k = 0 \ldots m-1$);
μ_k	transition rate from $S_{h,k}$ to $S_{h,k-1}$ ($k = 1 \ldots m$);
α_h	transition rate from $S_{h,k}$ to $S_{h+1,k}$ ($h = 1 \ldots m-1$);
β_h	transition rate from $S_{h,k}$ to $S_{h-1,k}$ ($h = 1 \ldots m$);
$P_{h,k}$	limiting state probability of $S_{h,k}$;
Ao	average traffic offered to the circuits;
W	parameter related to the unavailabilitiy (see later (4.3)).

3. THE MARKOVIAN MODEL

Considering the circuit failure effect in the evaluation of the loss probability leads to a two-dimensional Markovian model. Fig. 2 shows the state transition-rate diagram of the two-dimensional Markovian model for a circuit group with m circuits. This diagram is the result of a combination of two birth–death processes. The first is related to the traffic of the circuits with birth rate λ_k and death rate μ_k. The second is related to the failures and repairs of the circuits with birth rate α_h and death rate β_h. We have a triangular model due to the fact that we assume that

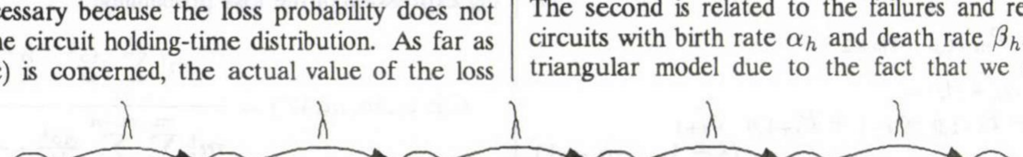


Fig. 1. One-dimensional Markovian model for an $M/M/5/0/\infty$ process

it is not possible to seize a faulty circuit. Each state $S_{h,k}$ corresponds to the probability of having h faulty circuits and k busy circuits. Typically the definition of the loss probability is given by formula (2):

$$\text{Loss Probability} = \sum_{i=0}^{m} P_{m-i,i} \qquad (2)$$

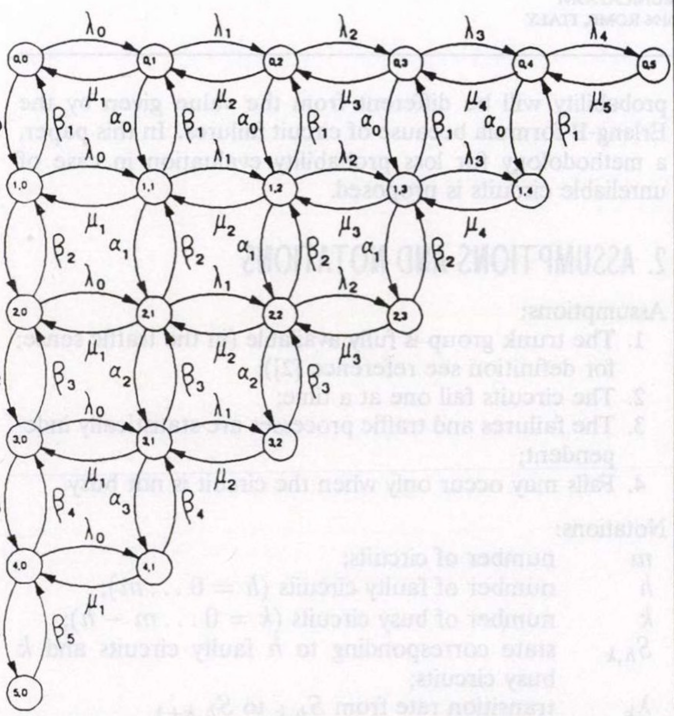


Fig. 2. Two-dimensional state transition-rate diagram, $m = 5$

From Fig. 2 it is easy to find the following equilibrium equations:

$$P_{0,0} \cdot (\lambda_0 + \alpha_0) = P_{0,1} \cdot \mu_1 + P_{1,0} \cdot \beta_1$$

$$P_{0,j} \cdot (\lambda_j + \mu_j + \alpha_j) = \\ = P_{0,j-1} \cdot \lambda_{j-1} + P_{0,j+1} \cdot \mu_{j+1} + P_{1,j} \cdot \beta_1 \\ (j = 1 \ldots m-1)$$

$$P_{i,j} \cdot (\lambda_i + \mu_i + \alpha_i + \beta_i) = \\ = P_{i,j-1} \cdot \lambda_{j-1} + P_{i,j+1} \cdot \mu_{j+1} + \\ + P_{i-1} \cdot \alpha_{i-1} + P_{i+1} \cdot \beta_{i+1} \\ (i = 1 \ldots m-2;\ j = 1 \ldots m-i-1)$$

$$P_{0,m} \cdot \mu_m = P_{0,m-1} \cdot \alpha_{m-1}$$

$$P_{i,m-i} \cdot (\mu_{m-i} + \beta_i) = \\ = P_{i,m-i-1} \cdot \alpha_{m-i-1} \cdot \lambda_{m-i-1} + P_{i-1,m-i} \cdot \alpha_{i-1} \\ (i = 1 \ldots m-1) \qquad (3)$$

$$P_{m,0} \cdot \beta_m = P_{m-1,0} \cdot \alpha_{m-1}$$

$$P_{i,0} \cdot (\lambda_0 + \alpha_i + \beta_i) = \\ = P_{i,1} \cdot \mu_1 + P_{i-1,0} \cdot \alpha_{i-1} + P_{i+1,0} \cdot \beta_{i+1} \\ (i = 1 \ldots m-1)$$

The solution of the set of equations can be found by considering the reversibility characteristics of the model. The property of reversibility (Ref. [3]) yields an easy method for finding the expressions of the limiting state probability:

$$P_{h,k} = P_{0,0} \cdot \prod_{i=0}^{k} \frac{\lambda_{i-1}}{\mu_i} \cdot \prod_{j=0}^{h} \frac{\alpha_{j-1}}{\beta_j} \qquad (4)$$

The expression in (4) has been obtained considering any closed path from $S_{0,0}$ to $S_{h,k}$ and backwards. In particular, according to the reversibility property, the ratio $P_{h,k}/P_{0,0}$ is equal to the ratio of the product of all transition rates from $S_{0,0}$ to $S_{h,k}$ to the product of all transition rates from $S_{h,k}$ to $S_{0,0}$.

A more accurate assumption would be the possibility of failures occurring also when a circuit is busy. This would require an additional transition rate $\delta_{h,k}$ from $S_{h,k}$ to $S_{h+1,k-1}$ $(h = 0 \ldots m-1;\ k = 1 \ldots m-h)$. The consideration of these transitions could give information on the cutoff phenomenon. Nevertheless, this issue is not discussed in this paper, and has been already studied in Ref. [4]. Moreover, in this case, the corresponding Markovian process is not reversible; an analytical solution is not known and therefore numerical techniques should be used. These transitions from $S_{h,k}$ to $S_{h+1,k-1}$ "move" the system state along the diagonal parallel with the one of the blocking states, not effecting, immediately, the blocking probability. Another issue of great interest would be the evaluation of the loss probability when more than one circuit simultaneously fails. This more realistic hypothesis would consider failure modes corresponding to the multiple failures due to the transmission systems, but the corresponding Markovian model is much more difficult to solve analytically. However, the method proposed in this paper gives a lower bound for the loss probability as compared with that obtained by including the failures during the circuit holding time. Nevertheless, the calculated loss probability is more accurate than the Erlang-B formula.

4. TWO APPLICATIONS

Through the application of (2) and (4), it is possible to obtain the expression of the loss probability for unreliable circuit groups. In this paper, two cases of interest are considered.

4.1. Application 1

This is the case in which the queueing system associated with the failures and repairs is of type $M/M/1/m/m$ and where

$$\alpha_k = \alpha \cdot (m-k) \quad (k = 0, 1, \ldots, m-1) \\ \beta_k = \beta \qquad\qquad (k = 1, 2, \ldots, m) \qquad (5)$$

This hypothesis and the expression (4) lead to the following expression of the loss probability:

$$GE1(Ao, m, W) = \frac{m! \sum_{k=0}^{m} \frac{Ao^k}{k!} \cdot \frac{W^{m-k}}{k!}}{m! \sum_{h=0}^{m} \sum_{k=0}^{m-h} \frac{Ao^k}{k!} \cdot \frac{W^h}{(m-h)!}} \qquad (6)$$

where:

$$W = \frac{\alpha}{\beta};\ Ao = \frac{\lambda}{\mu} \qquad (7)$$

It should be noted that Ao and W characterize the traffic process and the failure/repair process, respectively.

In addition to expression (6) there is another version of $GE1(Ao, m, W)$ given as recursive formula, useful for numerical calculations. If we define $Den[\cdot]$ and $Num[\cdot]$ as the denominator and the numerator of expression (6) we have

$$GE1(Ao,m,W) = = \frac{\frac{Ao^m}{m!} + m \cdot W \cdot Num[GE1(Ao,m-1,W)]}{\sum_{k=0}^{m} \frac{Ao_k}{k!} + m \cdot W \cdot Den[GE1(Ao,m-1,W)]} \quad (8)$$

and, remembering the expression of the Erlang-B formula in (1), we have

$$GE1(Ao,m,W) = \quad (9)$$

$$= \frac{\frac{Ao^m}{m!} + m \cdot W \cdot Num[GE1(Ao,m-1,W)]}{\frac{Ao_m}{m!E(Ao,m)} + m \cdot W \cdot Den[GE1(Ao,m-1,W)]}$$

The expressions in (8) and (9) are given for $m > 1$ and have following initial values:

$$Num[GE1(Ao,1,W)] = Ao + W$$
$$Den[GE1(Ao,1,W)] = \quad (10)$$
$$= 1 + Num[GE1(Ao,1,W)] = 1 + Ao + W$$

4.2. Application 2

This is the case in which the queueing system associated with the fails and repairs is of type $M/M/m/0/m$ and where

$$\begin{array}{ll} \alpha_k = \alpha \cdot (m-k) & (k = 0,1,\ldots,m-1) \\ \beta_k = \beta \cdot k & (k = 1,2,\ldots,m) \end{array} \quad (11)$$

This hypothesis and the expression (4) lead to the following expression of the loss probability:

$$GE2(Ao,m,W) = = \frac{m! \sum_{k=0}^{m} \frac{Ao^k}{k!} \cdot \frac{m!}{k!(m-k)!} \cdot W^{m-k}}{m! \sum_{h=0}^{m} \sum_{k=0}^{m-h} \frac{Ao^k}{k!} \cdot \frac{m!}{h!(m-h)!} \cdot W^h} \quad (12)$$

5. DIMENSIONING OF CIRCUIT GROUP

Expressions (6) and (12) lead to the Erlang-B formula when $W = 0$. It is easy to note that the two expressions are increasing function of Ao for given values of m and W. In Fig. 3, the diagram of the loss probability evaluated by the Erlang-B formula is presented, with $GE1(Ao,m,W)$ and with $GE2(Ao,m,W)$ for Ao varying from 1 to 100 Erl, with $m = 50$ and $W = 0.05$. It is to be noted how important the effect of non-reliable circuit groups is: it is more evident, obviously, in application 1, but still considerable in application 2. New considerations can be made looking to a different representation of the formulas $GE1(Ao,m,W)$ and $GE2(Ao,m,W)$. In Fig. 4 are presented the diagrams of $GE1(Ao,m,W)$ and $GE2(Ao,m,W)$ by fixing $Ao = 20$ Erl and $W = 0.05$ and varying m ($m = 1 \ldots 100$). In this representation, we note a phenomenon of "saturation" for $GE1(Ao,m,W)$. This leads to the existence of a lower limit for the loss probability in presence of values W different from zero. In other words, for a given value of Ao and W, an increase in the number of circuits m, above a certain threshold, does not effect the loss probability.

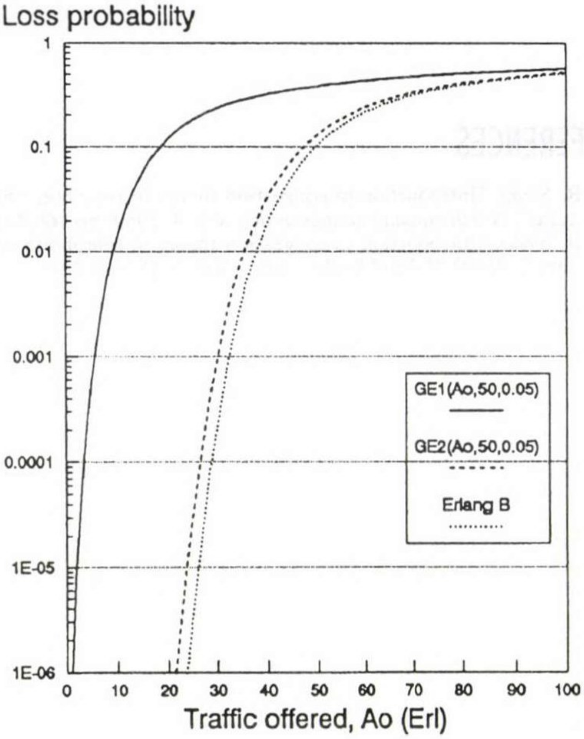


Fig. 3. Loss probability vs. offered traffic, $m = 50$ *and* $W = 0.05$

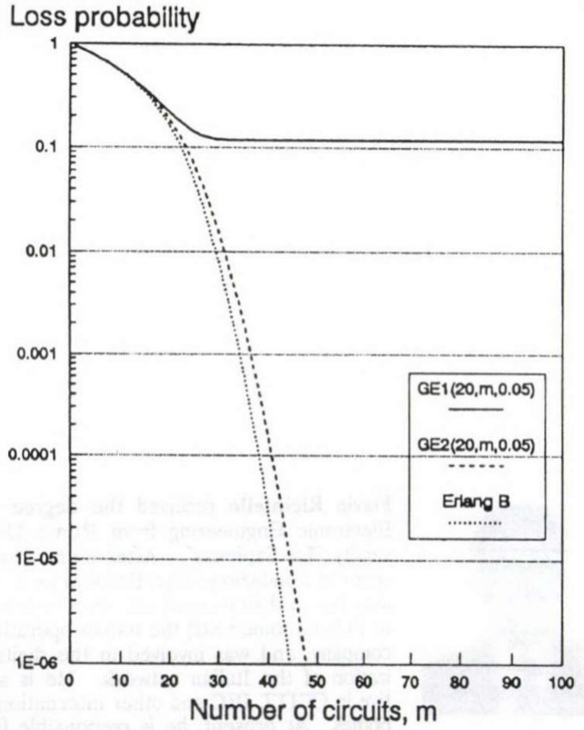


Fig. 4. Loss probability vs. number of circuits $A_0 = 20$ *Erl and* $W = 0.05$

6. CONCLUSIONS

The results presented in this paper show the effect of considering the failure/repair process in the evaluation of the loss probability. The hypothesis assumed for $GE2(Ao, m, W)$ is not realistic, in general, because it is not economical to have as many repairmen as the number of circuits. However, we note that the values of the loss probabilities calculated by using the Erlang-B formula, $GE2(Ao, m, W)$ and $GE1(Ao, m, W)$, are in an increasing order. This means that $GE2(Ao, m, W)$ gives a lower bound (but still higher then the Erlang-B formula) when failures are considered. Moreover, the decision to choose one out of three expressions should be, of course, the result of a technical and economical analysis (i.e. increasing the number of circuits or reducing the circuit unavailability).

REFERENCES

[1] R. Syski, "Introduction to congestion theory in telephone systems", *North Holland Studies in tlc.*, Vol. 4, 1959, pp.166-167.

[2] R. Syski, "Introduction to congestion theory in telephone systems", *North Holland Studies in tlc.*, Vol. 4, 1959, p.180.

[3] J.F.K. Kingman, "Markov population processes", *J. Appl. Prob.* 6, 1968, pp.1 – 18.

[4] M. Tortorella, "Cutoff Calls and Telephone Equipment Reliability", *B.S.T.J.*, Vol. 60, No. 8, October 1981, pp.1861 – 1889.

Flavio Riciniello received the degree in Electronic Engineering from Rome University "La Sapienza". After some years spent at Fondazione Ugo Bordoni as a researcher in digital communication systems, in 1974 he joined SIP, the Italian operating company, and was involved in the digitalization of the Italian network. He is active in CCITT, IEC and other international bodies. At present, he is responsible for Quality System development and is professor of Information Theory at L'Aquila University.

Riccardo Delleani received the degree in Electronic Engineering from Rome University "La Sapienza" in 1987. He joined SIP, the Italian operating company in 1988, and is working in the network quality department. He is now responsible for traffic forecasting for networks.

SOFTWARE QUALITY AND SCHEDULE CONTROL FOR A SWITCHING SYSTEM DEVELOPMENT PROJECT: A CASE STUDY

A. FUKUMOTO
NEC COMMUNICATION SYSTEMS, LTD.
1-2-1, ASAHI-CHO, KASIWA-SHI
CHIBA, 277, JAPAN

To ensure the required quality and delivery time in software development, meticulous design of the development program is required. Human power considerations, among others, is of particular importance. In conventional practice, the estimation of required human power is based on the program size, depending on the experience in similar previous projects and the intuition of estimators. It follows, therefore, that the estimate is accompanied by significant errors.
To compensate for this shortcoming, the Rayleigh Curve as a tool has been applied for the estimation of human power requirements in switching system software development. The result indicates that the application of the Rayleigh Curve helps in the estimation of the total human power required, if applied properly.

1. INTRODUCTION

With sophisticated communications network services and intensified competition among equipment suppliers and service providers, the volume of switching system software development is ever increasing, whereas users are calling for faster delivery of products.

Under such circumstances, an essential requirement for the successful implementation of a project is a skillfully-designed development program. Subsequently, the management of this development program in comparison with actual accomplishments would be of prime importance.

The formulation of a software development program is normally initiated by the human power estimation. The estimation of required human power depends largely on the estimator's experience and intuition. The consequence is that significant errors are produced in the human power estimation if the estimator has a limited experience in similar projects. It is commonly accepted that the Norden/Rayleigh Model (Fig. 1) proposed by Putnam is well adapted for general-purpose computer software design.

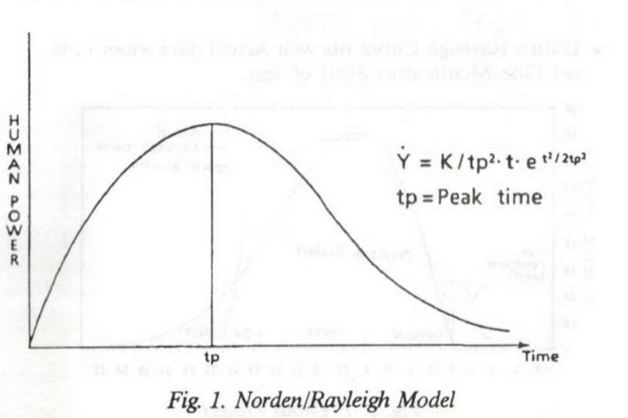


Fig. 1. Norden/Rayleigh Model

Assuming that the Norden/Rayleigh Model can be effectively applied to the area of switching software as well, the author tried to apply this model to a few previous projects for a case study and verified that the result of using this model was in agreement with the true state of affairs to a satisfiable extent.

An attempt was also made to estimate the total human power requirement for the entire case-study project by using the Rayleigh equation for software development for that project. As a result, the author is convinced that the use of the Rayleigh equation is a promising means of estimation. In addition, the method used for program quality control is also introduced.

2. CHARACTERISTICS OF THE CASE-STUDY PROJECT

The project for case study is characterized by the requirement for concurrent development of hardware and software. The author is with a software department in which application software is developed for the NEAX61 central office digital switching system. In almost all cases, software development work was associated with small amount of hardware modifications. The analysis of the case-study project requirements revealed that a large amount of hardware development was involved. Where the concurrent development is carried out for both hardware and software, there is a fear of significant problems that are likely to occur in terms of product quality and delivery time.

3. PROBLEMS OF THE CONVENTIONAL HUMAN POWER ESTIMATION [1]

The conventional human power estimation depends on skillful estimators who use their judgment for their own experience and intuition. More specifically, the estimator determines the number of lines for each program by guessing in order to obtain the entire program size L by summing up the numbers of lines for all programs involved. Next, the estimator determines the required human power E by substituting a productivity factor P per line, obtained from previous experience, into the following equation:

$$E = PLc \qquad (P > 0,\ c > 0) \qquad (1)$$

where c is a constant and empirically we assume that $c =$ $= 1$ in our department.

The conventional method is similarly effective in the human power estimation for software development in a similar preceding project. In a project with higher novelty, however, the program size L often tends to be underestimated because of lower accuracy of the estimation. Furthermore, the productivity factor P will largely vary with the differences in individual's and team's capabilities, complexity of a program, and completion factor of hardware development [2]. The human power requirement also tends to be underestimated with inaccurate productivities due to these possible variations.

With these unsteady factors, an estimated human power E is more than likely to be underestimated. Although an underestimated E will bring about a shortage of human power, directly linked with "degradation of software quality" and "a delay in delivery time", the human power shortage is something hardly recognizable quantitatively by managers.

In most cases, the human power shortage is discovered at the last moment before a program is released from the software department to the quality assurance department. It is then already too late when the manager becomes aware of the human power shortage. An urgent increase in human power often fails to save the situation and, often, the personnel increase can only contribute to a confusion. "Be quick in sensing a latent problem and settle the problem quickly" is the essence in software management.

4. TASK

Since this case-study project was different from previous projects requiring the conventional way of estimation, that is, the reuse of previous patterns of the estimation, the enhancement of accuracy in the human power estimation for software development was a large task assigned to the development management. It was therefore required that the development program be formulated and reviewed during development, and that the deployment of programmers be changed according to the reviewed development program. In our company, such an action of putting the development program into shape is called "Designing of Design Work (DDW)" activity.

5. TRIAL

Monthly review on estimated program size

To improve the accuracy of human power estimation, the estimated program size was reviewed every month to revise human power E for each month.

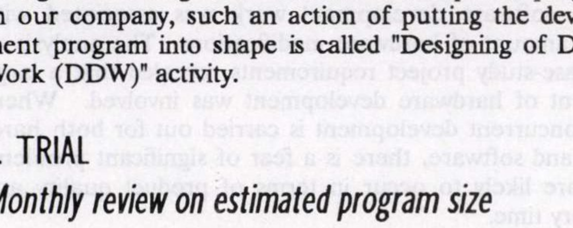


Fig. 2. Human Month estimate

Fig. 2 shows the estimated human power calculated from the estimated program size. Because the same value was used for the productivity factor P in each review, a variation in monthly human power estimation resulted from the outcome of the program size L review. Note that the value of human power relates to the application software alone, not the human power required for the entire system.

Application of Rayleigh curve model [4] to switching software

Next, an attempt was made to estimate the human power requirement by using the Rayleigh Curve, independent of the program size.

Application of the Rayleigh model to the previous project

The Rayleigh curve represents the distribution of human power requirements (precluding those for planning and functional specifications) during the life cycle after the start of design work.

The human power required for development at the development time can be obtained from the following equation:

$$dy/dt = K/t_p^2 \cdot t \cdot \exp[-t^2/2t_p^2] \ (K > 0, t_p > 0) \quad (2)$$

where

Y: total human power required for development up to development time t,

K: total human power required for development during the life cycle, and

t_p: development time at which dy/dt is maximized.

From the analysis on switching software development in a few previous projects, the author learned that the application of the Rayleigh curve would be best adapted to the actual state of affairs when the value of t_p in the model was set at one month after the start of test.

Although hardware development was partially involved in some of these projects, it appeared that the involvement had no impact on the timing of t_p. Note that the setting of t_p at one month after the start of the test does not mean that the actual human power was peaked at this time, but it means that the entire Rayleigh curve model did work well when t_p was set to this time.

On the other hand, t_p is the time when the development phase is ended in the case of the ordinary computer software and it is generally understood that the remaining human power is used for software maintenane and modifications.

APPLYING THE RAYLEIGH CURVE TO SWITCHING SOFTWARE

- $\dot{Y} = K/tp^2 \cdot t \cdot \exp[-t^2/2tp^2]$
- Entire Rayleigh Curve fits well Actual data when t_p is set One Month after Start of Test.

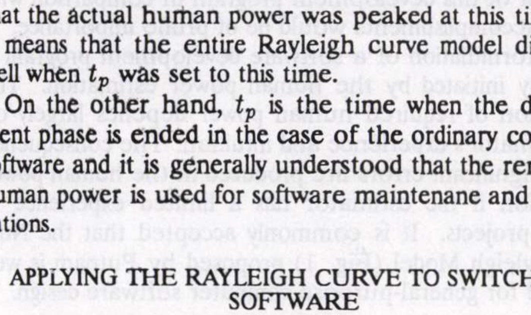


Fig. 3. Previous Project

The software has various qualitative characteristics such as portability, reliability, efficiency, testability, understandability, modifiability, and human engineering-based considerations.

Above all, the reliability is considered to be the most important factor for the switching software since a very stringent outage period of only one hour during the operation for 20 years is normally assessed to switching systems.

In the author's view, the fact that the model was best adapted to the actual human power requirement when the value of t_p was set at a specific value is the outcome of product characteristics and to peculiarities of the software development environment.

PARAMETER ESTIMATION

- K (human power) — Estimated from current project data using least-squares formula.
- t_p (peak time) — Estimated from data obtained through experience.

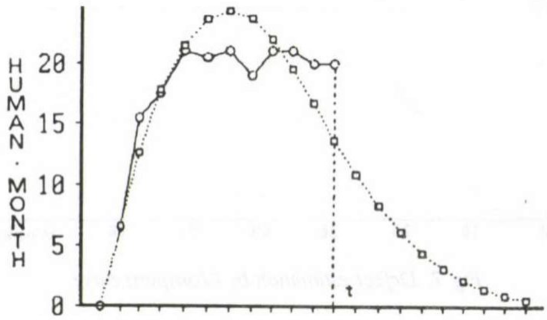


Fig. 3.a. Estimation method of human power K

Estimation by regression to Rayleigh equation

An attempt was made in this project to estimate a total human power K through a regression analysis to Rayleigh equation by using the results obtained in previous projects (Fig. 3.a).

Assuming that the human power in a certain month t is y, we obtain from equation (2):

$$y = K/t_p^2 \cdot t \cdot \exp[-t^2/2t_p^2] \quad (3)$$

The author did not use the ordinary least-squares estimation formula but used t_p that had been obtained through experience to obtain the most appropriate value for K in the following calculation.

If both parameters K and t_p are obtained by the least-squares estimation formula, the model would be forcedly applied to the actual human power determined by various factors in and around the project, with the result that the estimated value would be unstable. This is the reason why a value closely associated with the degree of work progress was used for t_p. From equation (3), we obtain

$$l_n yi = l_n k - 2l_n t_p + l_n ti - ti^2/2t_p^2 + Ei \quad (4)$$

$$\sum Ei^2 = \sum (l_n yi - l_n K + 2l_n t_p - l_n ti + ti^2/2t_p^2)^2 \quad (5)$$

$$\partial \sum Ei^2 / \partial K = \quad (6)$$
$$\sum l_n yi - n l_n K + 2n l_n t_p - \sum l_n ti + 1/2t_p^2 \sum ti^2 = 0$$

$$K = \quad (7)$$
$$= \exp[1/n(\sum l_n yi - \sum l_n ti + 1/2t_p^2 \sum ti^2) + 2l_n t_p]$$

where, Ei is an error in data i and n is the number of data. In the project, the peak time was assumed to be one month after the start of the test ($t_p = 6$). The total human power K was thus estimated by using equation (7), i.e. t_p was estimated from the data obtained through experience, while K was estimated by regression analysis.

Fig. 5 shows the value of K for each month that was estimated from the latest data. The dotted line indicates estimated values obtained from the program size. An example of the application when $t = 14$ is shown in Fig. 4.

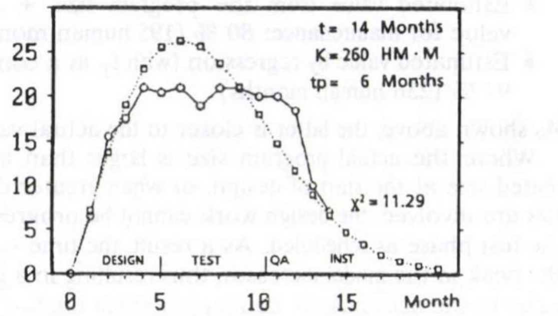


Fig. 4. Application to Rayleigh curve

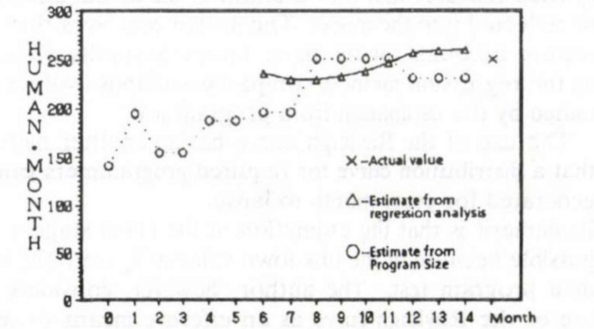


Fig. 5. Monthly estimates of required human power

Estimation of both parameters K and t_p
(For reference purpose)

An attempt was made to calculate both parameters K as a total human power and t_p as the number of months up to the peak time, regarding them as variables, although the values obtained were not used for the actual programmer deployment.

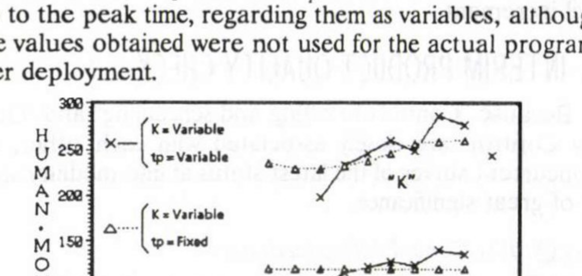


Fig. 6. Estimation of K and t_p

The estimated values obtained by this method are indicated in Fig. 6, using solid lines. Dotted lines indicate estimated values of K when the value of t_p is assumed to be a constant.

Since the values of two parameters are unstable and fluctuating to a large extent if obtained from the actual data, the author realized that the method in which K is estimated with t_p as a constant would be more practical than the former.

6. COMPARISON OF ESTIMATED VALUES AND CONSIDERATIONS

In contrast with the actual human power (244 human months) that includes the human power (5 human months) for maintenance after the program release, estimated values when $t = 6$th month obtained by each estimation method are as follows:

- Estimated value from the program size + Actual value for maintenance: 80 % (195 human months)
- Estimated value by regression (with t_p as a constant): 97 % (236 human months)

As shown above, the latter is closer to the actual value.

Where the actual program size is larger than an estimated size at the start of design, or when greater difficulties are involved, the design work cannot be progressed to the test phase as scheduled. As a result, the time t_p up to the peak in the model increases, thus resulting in a greater value of the accumulative human power required up to the present.

In other words, the method of obtaining the value with t_p fixed features that the variation in the actual state can be reflected into the model. The author conceives that this feature accounts for the more favorable results obtained by the regression method, compared with those values obtained by the estimation from program size.

The use of the Rayleigh curve has an another merit in that a distribution curve for required programmers can be generated for every month to lapse.

Its demerit is that the estimation at the initial stage is impossible because of the unknown value of t_p until the start of a program test. The author, however, considers the use of the Rayleigh curve as an effective means of security against an understaffed project, because this method will provide a means of checking the unreliable estimation from program size even in the interim phase.

It is possible at this stage to provide training for supplementary personnel, to plan gradual personnal reinforcement, and to avoid untimely personnel withdrawal, as well as to avoid a confusion in a project resulting from personnel increment.

7. INTERIM PRODUCT QUALITY CHECK

Because "Control of staffing and scheduling" and "Quality Control" are closely associated with each other, the concurrent survey of the latest status at intermediary stage is of great significance.

(1) *Quality Control during Inspection*

1) First, the estimated number of latent defects was obtained by multiplying the program size by a defect estimation coefficient for each work process.

2) Next, a lower control limit was determined against the number of latent defects to establish a defect elimination program for each work process.

3) The frequencies and methods of the inspection and test were reviewed so that the number of defects actually eliminated surpassed the lower control limit of elimination.

Thus, it became possible to eliminate defects every day while confirming the status of actual accomplishments against the number of latent defects.

This defect elimination technique was put to practice during the period from the preparation of functional specifications to the release of a program, and proved itself to be effective, particularly for inspection management in the design process.

(2) *Quality Control during Integration and Stability Tests*

The Gompertz curve is used in our department to estimate the cumulative number of defects and the number of remaining defects at every point of time, and it was also used for this study (see Fig. 7).

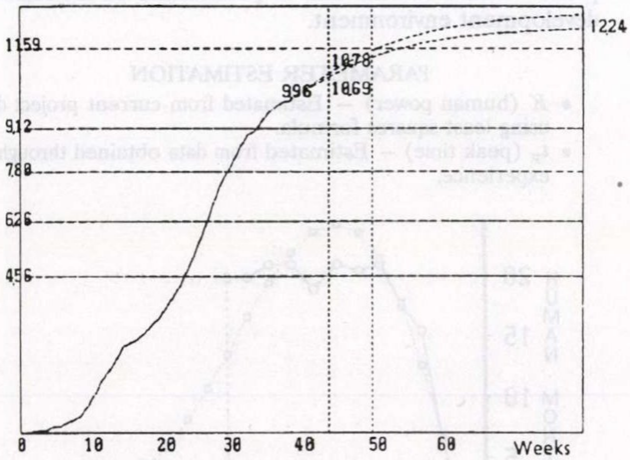


Fig. 7. Defect estimation by Gompertz curve

The Gompertz curve, however, can hardly be regarded as a specific quality criterion, since the curve contains some temporary variations depending on testing methods. Besides, a program will not be passed to the quality assurance without which the priority is eventually given to scheduling rather than to the program quality (see Fig. 8). Testing by the software department continues even after the release to the quality assurance department if it contains a defect in quality, even if the estimated number of remaining defects in the entire program should be less than the target value.

To cope with this situation, quality items of the standard were passed to the quality assurance without which the priority is eventually given to scheduling rather than to the program quality (see Fig. 8). Testing bey the software department continues even after the release to the quality assurance department debugging subgroups. Specific standards are indispensable for quality asssurance department so that the product can be shipped from the factory in flawless state (until a full mak of 100 points is achieved).

Item No.	Quality Item	Full Mark
1	Stability	20 points
2	Completion ratio of stability test items	20 points
3	The number of remaining problems	30 points
4	Customer specifications	20 points
5	Processing capability	5 points
6	Completion ratio of connection	5 points
7	Documents for external release	Y
	Acceptance threshold: 93 points	100 points

Fig. 8. Standards for release to quality assurance department

8. SUMMARY

(1) An attempt was made to apply the Rayleigh curve model to the NEAX61 switching software. In this paper, the method of estimating the total human power by expanded application of the preceding model has been proved to be effective and valid.
Although no estimation methods produce definitely accurate values, the proposed method will serve to ascertain the accuracy of an estimated value deprived from inaccurate program size so as to supplement and modify the estimated value. In the author's experiment, the accuracy obtained was higher than the human power estimation calculated from the program size.

(2) In this case study, the time t_p was first calculated from the data obtained from the past projects, and the total human power K was then obtained by the least-squares estimation formula. This method proved to be more practical in the human power estimation, compared with the method of obtaining both t_p and K concurrently by the least-squares estimation formula.

(3) It is important in software development to measure "the staffing and scheduling" and "quality control" on a concurrent basis, because both are closely associated with each other.
The author is convinced that proper measurement, accurate estimation, and appropriate management and control based on the above will be surely fruitful to the benefit of development personnel, management, and customers.

9. THE FUTURE ISSUES

(1) It will be necessary to further improve the estimation methods used in the experiment by applying them to various types of projects for estimating total human power and the peak time with higher accuracy. Also, an improvement is required so that an alarm of understaffed projects will be issued at an earlier stage.

(2) An automatic program size measurement is possible in a newly developed program "Software Modification History Management System". With this system, however, the measurement of work accomplishments for portions other than the new program such as program tests on the portion diverted from others is still experiencing difficulties. The accuracy of estimation at the initial stage of a project should be further upgraded through the advancement of automation and the enrichment of databases.

10. ACKNOWLEDGMENTS

The author wishes to express his thanks to Mr. K. Ashihara, corporate chief engineer and Mr. H. Hirayama, general manager, of NEC communication Systems, Ltd. for their guidance and numerous helpful suggestions in the course of this study.

REFERENCES

[1] W.S. Humphrey, *Managing the Software Process*, Addison-Wesley Publihshing Company, Inc., Massachusetts, pp.89–112, 1989.

[2] B.W. Boehm, *Software Engineering Economics*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.

[3] C.E. Walston and C.P. Felix, "A Method of Programming Measurement and Estimation", *IBM S.J.* Vol.16(1), pp.54–73, 1977.

[4] L.H. Putnam, "A General Empirical Solution to the Macro Software Sizing and Estimating Problem", *IEEE Trans. Software Engineering*, Vol. SE-4, No.4, pp.345–361, July 1978.

[5] A. Fukumoto, "Software Quality and Schedule Control in Switching System Development Project: A Case Study", *Proceedings of RELECTRONIC'91*, Scientific Society for Telecommunication, Hungary, pp.544–553, August 1991.

Akira Fukumoto is an engineering manager of Communication Systems, Ltd. He is currently responsible for the development of the NEAX61 digital switching system software. He received his B. E. degree in electrical engineering from Ritsumeikan University. He is a member of th IEICE of Japan.

LARGE MARKOVIAN SYSTEMS WITH STATE SPACE TRUNCATION: LOWER AND UPPER BOUNDS FOR INTERVAL AVAILABILITY

L. JEREB G. BÁNFI V. T. DO
TECHNICAL UNIVERSITY OF BUDAPEST
DEPARTMENT OF COMMUNICATION
H-1111 BUDAPEST
STOCZEK U. 2.

In the reliability analysis of complex electronic systems modelled by continuous time Markov chains, some problems arise because of the unmanageable large state space. In the paper, a method is presented yielding lower and upper bounds for the expected interval availability. The method is based on the truncation of the Markovian model, and the efficiency of the technique is illustrated by a numerical example generated by an IBM 3090 computer.

1. INTRODUCTION

In case of complex electronic systems modelled by large continuous time Markov chains, the reliability analysis has two problems. On the one hand, the generation and the numerical solution of the large model are impractical. Some efficient techniques have been published for the solution of the steady-state parameters of these systems ([1], [2]).

On the other hand, the steady-state parameters are often not adequate in case of large systems because the limiting probability distribution does not describe the behaviour of systems during the relatively short life time.

Considering the second problem, we apply the cumulative availability for the reliability qualification of the system. Since the exact reliability measures are generally not required, and because of the problems originating from the large model, only approximate results are derived for the expected interval availability.

The paper is organized as follows. In the first part, the basic mathematical background is summarized. Applying this model, upper and lower bounds of the expected interval availability are given, and only a small set of the most probable states (and the transitions among these states) are taken into account.

In the second part of the paper, some improvements of the basic technique are introduced. These models result in tighter approximations of the interval availability for Markovian systems, illustrated by numerical examples. Finally, some possible extensions of the described method are mentioned.

2. MATHEMATICAL PRELIMINARIES

We suppose that the system can be described from reliability point of view by a continuous time Markov chain (CTMC) $\{X(t),\ t \geq 0\}$, with discrete state space Ω, $\Omega = 1, 2, \dots Q = [q_{ij}]$ is the 'infinitesimal generator' of the CTMC. Let $\Pi_i(t)$ denote the probability of the CTMC being in state i at time t, and the row vector $\underline{\Pi}(t)$ is called the state probability vector.

Three different kinds of reliability measures can be used for the description of the system reliability behaviour. These parameters can be based on

- the steady-state (equilibrium),
- the transient and
- the cumulative

state probabilities. We assume that the analysis aims at the determination of the long-term behaviour so we can disregard the transient parameters. Assuming a great number of states, the steady-state probabilities cannot exist during the lifetime of the system. Accepting this assumption, the cumulative reliability measures seem to be the most appropriate, and we can focus our attention only on the determination of the expected interval availability.

2.1. Model 0

In order to derive the reliability measures, we separate the set of states Ω into the set of operational or up states Ω_u and of failed or down states Ω_d.
We introduce the indicator random variable

$$I(t) = \begin{cases} 1, & \text{if } X(t) \in \Omega_u, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The expected interval availability can be defined as:

$$\mathsf{E}[IAV(t)] = \mathsf{E}\left[\frac{1}{t}\int_0^t I(y)dy\right] = \frac{1}{t}\sum_{i\in\Omega_u}\int_0^t \Pi_i(y)dy \tag{2}$$

where $IAV(t)$ is the interval availability.

Introducing $\underline{r} = [r_i]$ as a row vector (reward) of the system, where

$$r_i = \begin{cases} 1, & \text{if } i \in \Omega_u, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Expression (2) can be rewritten as

$$\mathsf{E}[IAV(t)] = \frac{1}{t}\int_0^t \Pi(y)\underline{r}^T dy = \frac{1}{t}\int_0^t \Pi(y)dy\underline{r}^T. \tag{4}$$

Using the initial probability distribution $\Pi(0)$, the well-known solution of the differential equations of the CTMC is the following:

$$\Pi(t) = \Pi(0)\exp(Qt). \tag{5}$$

The expected interval availability can be numerically computed ([3]), but in case of a large state space, both the model generation and the numerical analysis can be unmanageable.

3. BASIC BOUNDS OF EXPECTED INTERVAL AVAILABILITY

In order to solve the above problems, we define the following model of the system which makes possible the generation of a relatively small state space, and at the same time it provides lower and upper bounds for the investigated parameter.

3.1. Model 1

We separate the set of states into two subsets (independently from the up or down state of the system), Ω_1 and Ω_2, and without the loss of generality, the states are reordered so that the states in Ω_1 have the first indices. In this case, the transition matrix Q can be also partitioned as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}.$$

Now we can modify the transitions so that the transitions within Ω_1, and from Ω_1 to Ω_2 remain as earlier, but the transitions within Ω_2, and from Ω_2 to Ω_1 are neglected.

This modification results in $\Omega_{21} = 0$ and $\Omega_{22} = 0$, and a modified CTMC, $X_1(t)$, which is identical to $X(t)$, if $X(y) \in \Omega_1$, $\forall y$, $0 \leq y \leq t$, and $X_1(t) \in \Omega_2$, $\exists y$, $0 \leq y \leq t$, $X(y) \in \Omega_2$.

In this case, the row vector $\underline{\Pi}(t)$ of the state probability can be also partitioned, $\underline{\Pi}^{(1)}(t) = [\underline{\Pi}_1^{(1)}(t)\underline{\Pi}_2^{(1)}(t)]$, and from (5) the following probability distribution follows because of $Q_{21} = 0$:

$$\Pi_1^{(1)}(t) = \Pi_1^{(1)}(0)\exp(Q_{11}t), \qquad (6)$$

where we assumed that the initial state is in Ω_1.

We can now introduce two indicator functions of $X_1(y)$:

$$I_1'(y) = \begin{cases} 1, & \text{if } X_1(y) \in \Omega_u \cap \Omega_1, \\ 0, & \text{otherwise.} \end{cases} \qquad (7)$$

and

$$I_1''(y) = \begin{cases} 1, & \text{if } X_1(y) \in \Omega_u \cup \Omega_1, \\ 0, & \text{otherwise.} \end{cases} \qquad (8)$$

3.1.1. Lower bound

Consider the following indicator function of the original CTMC, $X(y)$:

$$I_b(y) = \qquad (9)$$
$$= \begin{cases} 1, & \text{if } X(y) \in \Omega_u, \text{ and } X(v) \in \Omega_1, \forall v, 0 \leq v \leq y, \\ 0, & \text{otherwise,} \end{cases}$$

for which $I_b(y) \leq I(y)$ for each y, and for each realization of $X(y)$, and consequently $\mathsf{E}[I_b(y)] \leq \mathsf{E}[I(y)]$.

The following equation is also valid:

$$\mathsf{E}\left[\frac{1}{t}\int_0^t I_b(y)dy\right]_{\text{MODEL0}} = \mathsf{E}\left[\frac{1}{t}\int_0^t I_1'(y)dy\right]_{\text{MODEL0}} \qquad (10)$$

so the lower bound of the expected interval availability can be written as

$$\mathsf{E}[IAV(t)] = \mathsf{E}\left[\frac{1}{t}\int_0^t I(y)dy\right] \geq \mathsf{E}\left[\frac{1}{t}\int_0^t I_1'(y)dy\right]. \qquad (11)$$

Using (6) the lower bound of the expected interval availability can be rewritten in the following form:

$$\mathsf{E}[IAV(t)] \geq \mathsf{E}\left[\frac{1}{t}\int_0^t I'(y)dy\right] =$$
$$= \frac{1}{t}\Pi_1^{(1)}(0)(-Q_{11})^{-1}\left[E - \exp(Q_{11}t)\right]\underline{r}'^T, \qquad (12)$$

where E is the identity matrix, $\underline{r}' = [r_1']$ is the reward structure, and

$$r_i' = \begin{cases} 1, & \text{if } i \in \Omega_u \cap \Omega_1, \\ 0, & \text{otherwise.} \end{cases}$$

3.1.2. Upper bound

Introducing another indicator function of $X(y)$,

$$I_a(y) = \qquad (13)$$
$$= \begin{cases} 1, & \text{if } X(y) \in \Omega_u, \text{ or } \exists v, 0 \leq v \leq y, X(v) \in \Omega_2, \\ 0, & \text{otherwise,} \end{cases}$$

It can be realized that $I_a(y) \geq I(y)$ for each realization of $X(y)$ and for each y, consequently $\mathsf{E}[I(y)] \geq \mathsf{E}[I(y)]$.

Since the following equation is also valid:

$$\mathsf{E}\left[\frac{1}{t}\int_0^t I_a(y)dy\right]_{\text{MODEL0}} = \mathsf{E}\left[\frac{1}{t}\int_0^t I_1''(y)dy\right]_{\text{MODEL0}} \qquad (14)$$

using the steps as in case of the determination of the lower bound, the upper bound of the expected interval availability can be written as follows:

$$\mathsf{E}[IAV(t)] = \mathsf{E}\left[\frac{1}{t}\int_0^t I(y)dy\right] \leq \mathsf{E}\left[\frac{1}{t}\int_0^t I_1''(y)dy\right]. \qquad (15)$$

Exchanging the up and down states of the system, $I_1''(y)$ can be expressed by $I_1'(y)$:

$$I_1''(y) = \left\{\begin{matrix} 0, & \text{if } X_1(y) \in \Omega_d \cap \Omega_1, \\ 1, & \text{otherwise.} \end{matrix}\right\} = \qquad (16)$$
$$= 1 - I_1'(y)\,|_{u \to d, d \to u}\,.$$

Equation (16) is valid for the expected values as well:

$$\mathsf{E}[I_1''(y)] = 1 - \mathsf{E}(I_1'(y))\,|_{u \to d, d \to u}$$

and so

$$\mathsf{E}[IAV(t)] \leq \mathsf{E}\left[\frac{1}{t}\int_0^t I_1''(y)dy\right] = \qquad (17)$$
$$= 1 - \frac{1}{t}\Pi_1^{(1)}(0)(-Q_{11})^{-1}\left[E - \exp(Q_{11}t)\right]\left[\underline{e} - \underline{r}'\right]^T,$$

where $\underline{e}$ is a vector of ones. Equations (12) and (17) provide the lower and upper bounds of the expected interval availability. The size of the model to be generated and computed is reduced, and only the matrix Q_{11} has to be managed. The states of Ω_2 can be aggregated into a single state. The difference between (17) and (12) can be expressed as follows:

$$1 - \frac{1}{t}\Pi_1^{(1)}(0)(-Q_{11})^{-1}\left[E - \exp(Q_{11}t)\right]\underline{e}^T \qquad (18)$$

however, the problem of this technique is that the achievable bounds strongly depend on the investigated time interval.

4. MODIFICATION OF MODEL 1

According to the previous section, the efficiency of the technique is determined by the exciting probability from Ω_1. Before presenting better bounds, the basic idea of a modified analysis is illustrated.

This method makes possible the return of the process from a part of Ω_2 to Ω_1, and the system "can accumulate more reward" both in the up and down states during the investigated time interval.

4.1. MODEL 2

Now, we partition the set of states into three subsets (independently from the up and down property), Ω_1, Ω_2 and Ω_3 so that Ω_2 contains only those states from the former Ω_2 which can be reached from Ω_1 by a single transition. Without the loss of generality, the states are reordered, and the states in Ω_1 have the first, the states in Ω_2 have the second, and the states in Ω_3 have the third group of indices, respectively.

In order to make clear the method, temporarily we introduce Ω_1', and define the process so that after the first exitation from Ω_1, the process cannot return from Ω_2 to Ω_1 but only to Ω_1'. The transition rates from Ω_1 and from Ω_1' to Ω_2 are identical, and the transition rates from Ω_2 to Ω_1' are the former ones from Ω_2 to Ω_1 (see Fig. 1).

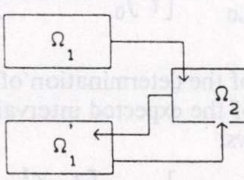


Fig. 1. MODEL 2 with state duplication of Ω_1

The transition matrix Q of MODEL 2 can be written as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix},$$

where similarly to MODEL 1, $Q_{31} = 0$, $Q_{32} = 0$ and $Q_{33} = 0$. As a consequence of this modification, Ω_3 can be aggregated into a single state.

The effect of this modification is trivial. The lower bound of the expected interval availability increases and the upper bound decreases because of the reward accumulated in Ω_1' after the first exitation from Ω_1. This improvement is true even if the reward in Ω_2 is not taken into account.

However, the consequence of the above modification is the increased size of the model, since practically MODEL 1 is generated, but instead of Ω_1, the detailed state space is $(\Omega_1 \cup \Omega_2)$.

Nevertheless, using this model and some assumptions, better bounds can be achieved than in case of MODEL 1, while the cardinality of the model remains similar to MODEL 1.

4.2. MODEL 3

The modeling assumptions are as follows:

- $Q_{13} = 0$, which can be the situation in which Ω_2 is a set of system states, which contains exactly a given number of failed components,
- $\exists i \in \Omega_1$, $q_{ji} > 0$, for $\forall\ j \in \Omega_2$. Let one of these states be numbered as 1.

In order to reduce the size of the model we modify the elements of the transition matrix in the following manner. For $\forall\ j \in \Omega_2$, the elements of Q_{21} are replaced by:

- $q_{ji}' = 0$, if $i \in \Omega_1$ and $i > 1$,
- $q_1' = \min_k q_{k1}$, where $k \in \Omega_2$,

and taking into consideration the aggregation of states in Q_3, the elements of Q_{23} are also modified for $\forall\ j \in \Omega_2$:

- $q_0' = \max_j \sum_k q_{jk}$, where $k \in \Omega_3$.

Since the transition rates from Ω_2 to Ω_1 equal q_1' and from Ω_2 to Ω_3 they equal q_0' for $\forall\ j \in \Omega_2$, the states of Ω_2 can be also aggregated into a single state $n-1$, and for $\forall\ i \in \Omega_1$:

- $q_{i,n-1}' = \sum_j q_{ij}$, $j \in \Omega_2$.

A modified CTMC, $X_3(y)$ with the state space Ω' can be introduced as MODEL 3 for which $\Omega_1' = \Omega_1 = 1, 2, \ldots, n-2$ and $\Omega_2 = n-1$, $\Omega_3 = n$ (see Fig. 2), where $q_{n-1,1}' = q_1'$, $q_{n-1,n}' = q_0'$ and $Q_{12}' = [q_{i,n-1}]$, $i \in \Omega_1$ is a column vector.

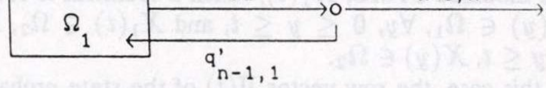


Fig. 2. MODEL 3 with the aggregation of states in Ω_2 *and* Ω_3

The following relation can be easily proved in case of the above modifications for $\forall\ j \in \Omega_2$:

$$\Pr\{\exists u \le y,\ X(u) = 1 \mid X(0) = j,\} \ge$$
$$\ge \Pr\{\exists u \le y,\ X_3(u) = 1 \mid X_3(0) = n-1\}, \quad (19)$$

and from (19) the increment of the time spent by the process in Ω_2 before returning to state 1 follows. On the other hand it can be realized that the transitions from Ω_2 to other states than state 1 in Ω_1 are neglected in this model, and the effect of these modifications is the decrement of the returning probability to Ω_2. Finally, from the fact that the process can return to Ω_1 in contrast with MODEL 1, the following relation can be established:

$$\mathsf{E}[IAV(t)]|_{\mathrm{MODEL1}} \le \mathsf{E}[IAV(t)]|_{\mathrm{MODEL3}} \le$$
$$\le \mathsf{E}[IAV(t)]|_{\mathrm{MODEL0}}. \quad (20)$$

Based on the MODEL 3 depicted in Figure 2, the transition rate matrix is the following:

$$Q' = \begin{bmatrix} Q_{11} & Q_{12}' & 0 \\ q_{n-1,1} & q_{n-1,n-1} & q_{n-1,n} \\ 0 & 0 & 0 \end{bmatrix}$$

where $q_{n-1,n-1} = -(q_{n-1,1} + q_{n-1,n})$.

We can introduce now the notation

$$Q_{11}' = \begin{bmatrix} Q_{11} & Q_{12}' \\ q_{n-1,1} & q_{n-1,n-1} \end{bmatrix}$$

and new lower and upper bounds can be obtained for the expected interval availability using (20) and the former results of (12) and (17).

The modified lower bound is the following:

$$\mathsf{E}[IAV(t)] \geq$$

$$\geq \frac{1}{t}\Pi_1^{(3)}(0)(-Q'_{11})^{-1}[E - \exp(Q'_{11}t)]\ \underline{r}''^{T}, \quad (21)$$

where $\underline{r}'' = r'$, for $i = 1, \ldots, n-2$, and $r''_{n-1} = 0$, while the upper bound is as follows:

$$\mathsf{E}[IAV(t)] \leq 1 - \frac{1}{t}\Pi_1^{(3)}(0)(-Q'_{11})^{-1}$$

$$[E - \exp(Q'_{11}t)]\ [\underline{e} - \underline{r}', 0]^T, \quad (22)$$

where the first $n-2$ elements of the right column vector are identical to the elements in Equation (17), but the last one equals 0.

The limit of MODEL 3 is the existence of state 1 which must be reached by repair from all states of Ω_2. This assumption is usually not valid, thus the direct application of MODEL 3 is excluded in most practical cases.

4.3. MODEL 4

In order to neglect the limits of MODEL 3, Ω_2 is partitioned further into subsets $\Omega_{20}, \Omega_{21}, \ldots, \Omega_{2m}$ so that from all states in Ω_{2m}, $m = 0, 1, 2, \ldots, M$, the state $i_m \in \Omega_1$ can be reached by a single transition. It is assumed that for Ω_{20}, no such state in Ω_1 exists.

The elements of the transition matrix can be modified for $\forall\, j \in \Omega_{2m}, m = 1, 2, \ldots, M$, replacing the elements of Q_{21} by

- $q'_{ji} = 0$, if $i \in \Omega_1$ and $i \neq i_m$,
- $q'_{i_m} = \min_j q_{ji_m}$,

and similarly to MODEL 3, the elements of $Q_{2m,3}$ can also be modified for Ω_{2m}, $j = 0, 1, 2, \ldots, m$:

- $q'_{m0} = \max_j \sum_k q_{jk}$, where $k \in \Omega_3$.

Since the transition rates from Q_{2m} to Ω_1 equal q'_{i_m} and from Q_{2m} to Ω_3 they equal q'_{m0} for $\forall\, j \in \Omega_{2m}$, the states of Ω_{2m} can also be aggregated into a single state j_m, and for $\forall\, i \in \Omega_1$:

- $q'_{ij_m} = \sum_j q_{ij}$, $j \in \Omega_{2m}$.

A modified CTMC, $X_4(y)$ with the state space Ω'' can be introduced for which $Q''_2 = j_0, \ldots, j_M$, $\Omega_3 = i_0$, where $q''_{j_m,i_m} = q''_{i_m}$, $q''_{j_m} = q''_{m0}$, $Q''_{12} = [q_{ij_m}]$.

Now, the results in Equations (19) to (22) still hold, but instead of MODEL 3, the notation of MODEL 4 must be taken into consideration, and in (22) the reward in all states of Ω''_2 equals 0.

5. NUMERICAL RESULTS

The effect of the method is illustrated by a simple example. In order to obtain exact results, the investigated system consists of nine two-state independent subsystems. The block digram of the system is shown in Figure 3.

The system is studied in two cases with the same failure but with different repair rates (see Table 1.).

The numerical results for the investigated system have been generated by an IBM 3090 computer and are depicted in Figs. 4 and 5.

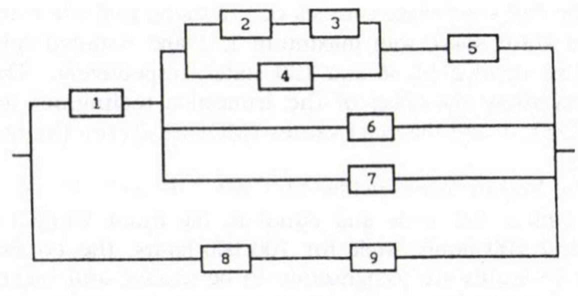


Fig. 3. The reliability block diagram of the example

Table 1. Reliability parameters of the subsystems

Index of block	Failure rates 10^{-6}/hour	Repair rates (10^{-3}/hour) CASE 1	CASE 2
1	90	1	10
2	80	2	20
3	70	3	30
4	60	4	40
5	50	5	50
6	40	6	60
7	30	7	70
8	20	8	80
9	10	9	90

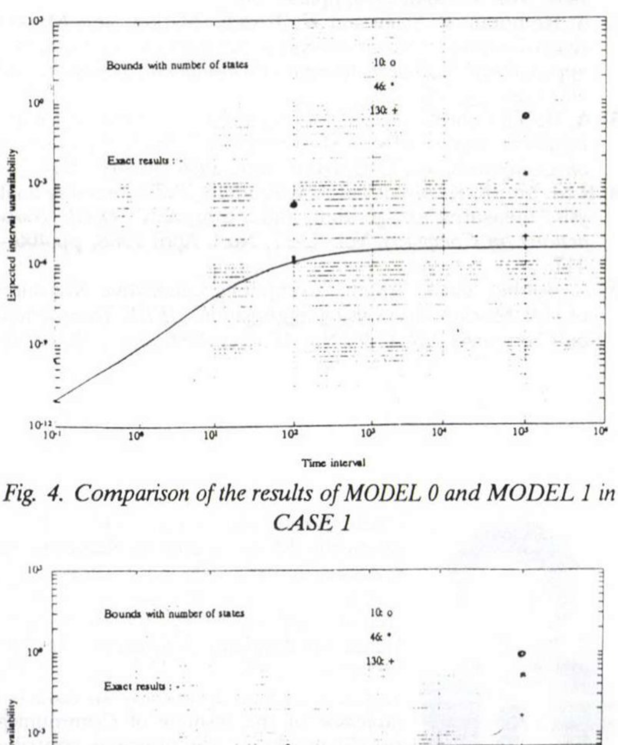


Fig. 4. Comparison of the results of MODEL 0 and MODEL 1 in CASE 1

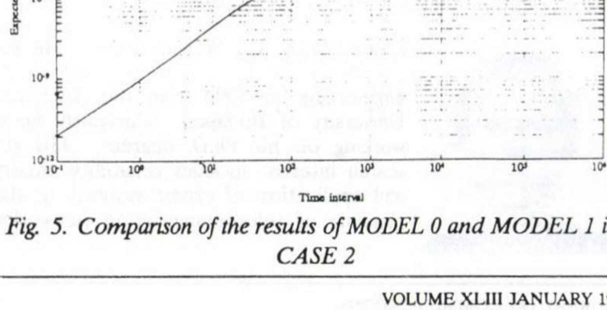


Fig. 5. Comparison of the results of MODEL 0 and MODEL 1 in CASE 2

The full state space consists of 512 states, and the truncated state space with maximum 1, 2 and 3 failed subsystems involve 10, 46 and 130 states, respectively. The curves show the effect of the truncation techniques for MODEL 1 and the exact values (MODEL 0) for the two cases.

The bounds generated by MODEL 1 46 and 130 states are almost the same and equal to the exact value for 0.1 and 100 hours, while for 100 000 hours, the bounds with 46 states are recognizable to be smaller and higher in both (CASE 1 and CASE 2) cases. It is important to mention that the lower bounds of unavailability (and the upper bounds of availability) are very close to the exact values in both cases and for all investigated time intervals. The reason for this property is the structure of the system and the high value of interval availability since the system can accumulate very few "rewards" in the down states after returning from Ω_2.

Naturally, for 10 states only the upper bound of unavailability (and the lower bound of availability) exist because all the 10 states belong to the up states of the system since the number of failed subsystems is less than 2 in these states.

6. CONCLUSIONS

A technique is proposed for the estimation of the expected interval availability using state space truncation. The greater the ratio of the repair and failure rates, the more effective the method is as shown by the example given.

The method can be applied in the reliability analysis like in RELECT, RELNET and RELCONT published in [6] – [8], and the bounds can be improved by using MODEL 3 and MODEL 4.

This limit as well as the inconvenience of the different time scales, either in the failure or in the repair rates, can be eliminated by using a similar aggregation for the states in Ω_2 like in MODEL 4, based on the group of states with similar failure or repair rates.

Finally, it can be mentioned that this technique can be applied not only for the estimation of the expected interval availability but for the transient and steady-state availability (by using the method of [1]), and the results can be extended for the case of reliability models with multilevel reward structures as well.

REFERENCES

[1] E. de Souza e Silva, R.R. Muntz and A. Goyal, "Bounding availability of repairable systems", *Performance Evaluation Review*, Vol. 17, May, 1989, pp.29 – 38.

[2] A. Reibman, R. Smith and K. Trivedi, "Markov and Markov reward model transient analysis: An overview of numerical approaches", *European Journal of Operational Research*, Vol. 40 (1989), pp.257 – 267.

[3] A. Bobbio and K. Trivedi, "An aggregation technique for the transient analysis of stiff Markov chains", *IEEE Transactions on Computers*, Vol. C-35, No.9, Sept. 1986, pp.803 – 814.

[4] R.M. Smith, K. Trivedi and A.V. Ramesh, "Performability analysis: measures, an algorithm, and a case study", *IEEE Transactions on Computers*, Vol. C-37, No.4, April 1988, pp.406 – 417.

[5] A. Bobbio and K. Trivedi, "Computing Cumulative Measures of stiff Markov chains using aggregation". *IEEE Transactions on Computers*, Vol. C-39, No. 10, Oct. 1990, pp.1291 – 1298.

[6] L. Szandi, L. Jereb, Gy. Farkas and T. Izsó, "RELECT: A program package for computer aided reliability analysis of electronic components and systems" *Proceedings of the 8th Symposium on Reliability in Electronics*, OMIKK–Technoinform, Budapest, 1991, pp.263 – 272.

[7] T. Jakab, L. Jereb and M. Telek, "RELNET: A tool for the reliability analysis of communication networks", *Proceedings of the 8th Symposium on Reliability in Electronics*, OMIKK–Technoinform, Budapest, 1991, pp.513 – 522.

[8] T. Horváth, P. Risztics and L. Szirmay-Kalos, "RELCONT: Hierarchical analysis of process control systems", *Proceedings of the 8th Symposium on Reliability in Electronics*, OMIKK–Technoinform, Budapest, 1991, pp.254 – 262.

[9] L. Jereb, T. Mohay and G. Bánfi, "State space truncation in the interval availability analysis of large Markovian systems", *Proceedings of the 8th Symposium on Reliability in Electronics*, OMIKK–Technoinform, Budapest, 1991, pp.338 – 347.

Lászlo Jereb was born in 1947. He received the M.Sc. degree in electrical engineering in 1971 from the Technical University of Budapest, and the academic degree in reliability theory in 1986 from the Hungarian Academy of Sciences. In 1971 he joined the Technical University of Budapest, where he is currently an associate professor of the Institute of Communication Electronics. His principal contributions have been in reliability theory, as well as planning and analysis of communication networks. He is a member of the IEEE Communication Society.

Gábor Bánfi was born in 1966. He received the M.Sc. degree in electrical engineering in 1990 from the Technical University of Budapest. Currently he is working on his Ph.D. degree. His research interest includes reliability theory and application of expert systems in the planning of telecommunication networks.

Van Tien Do was born in 1967. He received the M.Sc. degree in electrical engineering in 1991 from the Technical University of Budapest. Currently, he is working on his Ph.D degree. His research interest includes traffic analysis and planning of communication networks as well as the reliability modeling and analysis of complex electronic systems.

STOCHASTIC REWARD MODELS IN PERFORMANCE/RELIABILITY ANALYSIS

A. BOBBIO
ISTITUTO ELETTROTECNICO NAZIONALE GALILEO FERRARIS
STRADA DELLE CACCE 91
10135 TORINO
ITALY

This paper reviews the theory and application of Stochastic Reward Models (SRM) for the analysis of combined measures of performance and reliability in complex fault-tolerant degradable systems. Two cumulative measures associated to a SRM are introduced and discussed: the accumulated reward over a finite horizon and the completion time of a task with an assigned work requirement. The physical meaning of these two measures is illustrated and their mutual relationships are investigated. Finally, the techniques for evaluating their cumulative distribution function (Cdf) are comparatively examined.

1. INTRODUCTION

The classical reliability theory [78], [4], [84] is based on the assumption that each component, and the system as a whole, can be modelled by a binary variable representing two possible conditions: functional or non-functional. This assumption implies that the state space of the system can be univocally partitioned into two mutually exclusive subsets of states, one containing the up states, the other containing the down states. The classical reliability measures are defined on this binary partition of the state space.

For fault-tolerant or degradable systems, the binary assumption does not provide a correct description of the system behaviour versus time. The system can still provide service to the user, even in the presence of failures of some of its parts, at a possibly degraded level of operation. The occurrence of a failure causes the system to go through a sequence of degraded states characterized by reduced levels of performance. Hence, in order to quantify the reliability and performance aspects of a fault-tolerant system into a single integrated view, there is a need for a new approach and new measures [61].

With the aim of properly assessing the attributes of fault-tolerant systems, a new terminology has recently emerged from the technical literature. This terminology is based on the all-encompassing term of dependability [56], [3]. "Dependability is that property of a system that allows reliance to be justifiably placed on the service it delivers [3]". A survey on the modelling and analysis techniques for evaluating the dependability of fault-tolerant systems appeared in [6].

The basic idea of the new modelling approach consists in describing the behaviour of the system in time by means of a stochastic process, called the structure-state process, and by associating to each state of the structure-state process a non-negative real constant representing the effective working capacity or performance level of the system in that state. The real variable associated to the state is called the reward rate [41]. The structure-state process, together with the reward rates, forms the Stochastic Reward Model.

The properties of stochastic processes with a superimposed reward variable have been studied since a long time [58], [18], [50], [51], [41]. However, only recently, SRM have received attention as a modelling tool in performance/reliability evaluation. The possibility of associating a reward variable to each structure-state increases the flexibility and the descriptive power of the model, and makes it more useful in the analysis of actual systems.

Different interpretations of the structure-state process and of the associated reward structure give rise to different applications [62]. For instance, the classical reliablity theory [4] can be considered as a particular case of SRM obtained by assigning a $\{0, 1\}$ binary reward variable to the structure-state process. In applied stochastic modelling, common assignements of the reward rates are: execution rate of tasks in computing system (the computational capacity) [5], [83], number of active processors (or processing power) [7], [35], throughput [60], [29], [39], average response time [42], [52], [57] or response time distributions [85], [72], [65].

To be concrete, in the present setting, we assume that the reward variables represent the computation capacity of a degradable computing system so that the reward accumulated over a finite horizon is a measure of the total computation performed by the system. With the aim of characterizing a system modelled by a SRM, various significant measures can be defined [30], [83], [79]. Instantaneous measures reflect the system condition at a given point in time (or in steady state), while cumulative measures reflect the system's operational characteristics over a finite interval.

Two main different points of view have been assumed in the literature when dealing with SRM for degradable systems [55]. In the system oriented point of view, the most significant measure is the total amount of work done by the system in a finite interval. This measure is a random variable that corresponds to the reward accumulated by the SRM over the considered interval. The distribution function of the accumulated reward is sometimes called performability [61], and its efficient numerical evaluation is a challanging task considered in a number of papers: [60], [44], [21], [32], [79], [23]. In the task oriented point of view, the system is regarded as a server, and the emphasis of the analysis is on the ability of the system to process a task whose execution requires a known amount of computational work. In this latter case, the most significant measure is the probability that the task will be completed in a given time, taking into account the system degradation and the recovery capabilities subsequent to a failure. Even if the final target of a dependable computation is the quality of the service, the task oriented point of view is not very commonly assumed in the literature.

Gaver [28] analyzed the distribution of the completion time for a two state server with different mechanisms of interruption and recovery policies. Extensions to the above model were provided in [67], while the completion time problem for fault-tolerant computing systems was addressed in [13]. A unified framework to the system oriented and the user oriented point of view was provided by Kulkarni et al. in [55], [54], [69], where the degradable server is represented by a SRM and the task in progress interacts with the system following different preemption and recovery policies.

A formal definition of SRM is stated in Section 2, and the most significant instantaneous and cumulative measures are defined. The computational load to evaluate the instantaneous measures or the expected values of the cumulative measures is of the order of the load needed to compute the state probabilities of the underlying structure-state process. The evaluation of the distribution function of the cumulative measures over a finite time horizon is more intricate. Section 3 reviews the numerical techniques that have been proposed in the literature for evaluating the Cdf of the cumulative measures when the underlying structure-state process is restricted to be a continuous time homogeneous Markov chain (CTMC).

2. STOCHASTIC REWARD MODELS (SRM)

Let $Z(t)$, $(t \geq 0)$ denote a stochastic process defined over a discrete and finite state space Ω of cardinality n. $Z(t)$ is called the structure-state process and models the stochastic behaviour of the system in time. Let $\underline{P}(t)$ be the state probability vector such that each entry $P_i(t)$, $(i = 1, 2, \ldots, n)$ is the probability that the process is in state i at time t, and let $\underline{P}_0$ be the initial probability vector.

In the sequel of the paper, we often consider the case in which $Z(t)$ is a continuous time homogeneous Markov chain (CTMC). This choice is motivated by the observation that almost all the numerical algorithms available in the literature for determining performability measures have been built up under the CTMC restriction. On the other hand, some theoretical results are available in the more general setting of semi-markovian or non-homogeneous markovian processes, and the pertinent references will be mentioned when appropriate.

Let $Z(t)$ be a CTMC with infinitesimal generator $\mathbf{Q} = [q_{i,j}]$. The state probability vector $\underline{P}(t)$ may be computed by solving the standard Markov Equation

$$\dot{\underline{P}}(t) = \underline{P}(t)\mathbf{Q} \tag{1}$$

whose formal solution is

$$\underline{P}(t) = \underline{P}^0 \exp(\mathbf{Q}\, t).$$

A discussion on various numerical algorithms to solve efficiently Equation (1) in the transient domain has been presented in [74]. The classical instantaneous measures, like reliability, availability and maintainability are directly derived from (1) by summing up the state probabilities over suitably defined partitions of the state space [4]. Define the vector

$$\overline{B}(t) = \int_0^t \overline{P}(\tau) d\tau \tag{2}$$

The $i-th$ entry of $\overline{B}(t)$ is the expected time spent by the process $Z(t)$ in state i during the $(0, t)$ [4], [30]. By direct integration of (1), Equation (2) takes on the form [30]:

$$\dot{\underline{B}}(t) = \underline{B}(t)\mathbf{Q} + \underline{P}^0 \tag{3}$$

with initial condition $\underline{B}(0) = \underline{0}$. The system of linear differential equations (3) can be solved with the same methods used for solving (1) [75], and thus with approximately the same computational load.

Let $X(t)$ be a non-negative real-valued function, defined on Ω. $X(t)$ is the instantaneous reward function, and is assumed to have the following expression:

$$X(t) = r_i \quad \text{if} \quad Z(t) = i. \tag{4}$$

This definition implies that a reward $r_i \tau_i$ is accumulated when the sojourn time of the process $Z(t)$ in state i is τ_i. For notational convenience we introduce the following vector and matrix of reward rates:

$$\underline{r} = [r_1, r_2, \ldots, r_n]^T \text{ and } \mathbf{R} = \text{diag}\,[r_1, r_2, \ldots, r_n]. \tag{5}$$

The structure-state Markov process $Z(t)$, together with the reward rates attached to each state, form the Markov Reward Model (MRM).

2.1. Instantaneous measures

The expected instantaneous reward rate at time t is given by:

$$E[X(t)] = \sum r_i \text{Prob}\,\{Z(t) = i\} = \sum r_i P_i(t) \tag{6}$$

The distribution of $X(t)$ (i.e. Prob $\{X(t) \leq x\}$) can also be computed easily as:

$$\text{Prob}\,\{X(t) \leq x\} = \sum_{r_i \leq x, i \in \Omega} P_i(t) \tag{7}$$

An exhaustive list of appropriate performance reliability measures defined on a SRM (including various passage times and interval measures), has been presented in [30], [83], while in [65], different interpretations of the reward variables in applied problems are reviewed. Equation (7) has been proposed by different authors as a unified measure of performance and reliability [42], [64], [57].

2.2. Cumulative measures

Cumulative measures are related to the accumulation of the reward during a finite time interval.

2.2.1. The accumulated reward

The total accumulated reward up to time t is the random variable defined as:

$$Y(t) = \int_0^t X(\tau) d\tau \tag{8}$$

From (8), the expected accumulated reward is given by:

$$E[Y(t)] = E[\int_0^t X(\tau) d\tau] = \sum_i r_i \int_0^t P_i(\tau) d\tau = \sum_i r_i B_i(t) \tag{9}$$

so that (9) can be computed by solving (3). The distribution function $F_Y(t, x)$ of $Y(t)$ is defined as:

$$F_Y(x,t) = \text{Prob}\ \{Y(t) \leq x\} \qquad (10)$$

and is sometimes called the performability [61].

The double Laplace–Stieltjes transform of $F_Y(x,t)$ has been independently derived by different authors [44], [55], [83] resorting to a renewal argument. From [55] let us introduce the vector valued function $\underline{F}_{Yi}(x,t)$ whose entries $F_{Yi}(x,t)$, $(i = 1, 2, \ldots, n)$ are defined by:

$$F_{Yi}(x,t) = \text{Prob}\ \{Y(t) \leq x \mid Z(0) = i\},\ x \geq 0 \qquad (11)$$

From the above definitions, it is proved in [55] that $\underline{F}_Y^{\sim *}(u,s)$ satisfies the following matrix equation:

$$\underline{F}_Y^{\sim *}(u,s) = [s\mathbf{I} + u\mathbf{R} - \mathbf{Q}]^{-1}\underline{e} \qquad (12)$$

where $\mathbf{I}$ is the identity matrix, $\underline{e}$ a column vector of size n with all elements equal to 1, superscript $^{\sim}$ denotes the Laplace–Stieltjes transform with respect to x (being the transformed variable u) and superscript * the Laplace transform with respect to t (being the transformed variable s).

2.2.2. The accumulated reward up to system failure

If the structure-state process $Z(t)$ contains a subset of absorbing (failed) states, the accumulated reward up to absorption is a finite random variable Y_∞, given by:

$$Y_\infty = \lim_{t \to \infty} Y(t) \qquad (13)$$

The Cdf of Y_∞ was first inverstigated by Beaudry in [5] being $Z(t)$ a CTMC. Beaudry's approach is based on the observation that an exponentially distributed sojourn time in state i with parameter λ_i accumulates an exponentially distributed reward with parameter λ_i / r_i, where $r_i > 0$ is the reward rate assigned to state i. Hence, Y_∞ can be computed as the absorption probability in a modified CTMC where the original transition rates outgoing from a generic state i are scaled by means of the reward rate associated to state i. The proposed transformation from the time domain to the computational domain falls short if a reward rate $r_i = 0$ is associated to a non-absorbing state, since in the modified chain the scaled transition rate becomes undefined. If state i is transient and $r_i = 0$, no reward is gained during the sojourn of the system in state i, and in the computational domain state i behaves as an instantaneous state (a state transversed in zero time). Handling CTC with instantaneous states has become quite popular in recent years in connection with Petri nets with timed and immediate transitions [1] and with the analysis of stiff Markov chains [11].

The extension of the Beaudry's approach to the evaluation of the Cdf of Y_∞, including zero reward rates, has been investigated in [83] for CTMC and in [15] for semi-Markov processes.

2.2.3. The completion time

Let us suppose that the system must process a task whose execution requires W units of work. W is, in general, a random variable with distribution $G(x)$. The degenerate case in which W is deterministic and the distribution $G(x)$ becomes the unit step function located at $W = x$, can be considered as a particular case. Given $W = x$, the task completion time $T(x)$ is the random variable representing the time to complete a task whose work requirement is x. In order to completely characterize the problem, the interaction between the task in progress and the change of state in the structure-state process should be defined.

To this end, following [28], [55], three different interaction policies are considered:

- *prs policy* — When a change of state occurs in the structure-state process, the system keeps memory of the work already done and the task in progress is resumed in the new state. This policy is called preemptive resume (prs).
- *pri policy* — When a change of state occurs in the structure-state process, the system cannot keep track of the past, and the work already done is lost; the task in progress must be restarted from scratch in the new state and the repeated task has a work requirement identical to the one of the preempted task. This policy is called preemptive repeat identical (pri).
- *prd policy* — As in the pri case, the system cannot keep track of the past, and the work already done is lost at any change of state. The repeated task has a work requirement sampled from the same distribution of the original preempted task. This policy is called preemptive repeat different (prd).

Let $F_T(t,x)$ be the Cdf of the task completion time $T(x)$:

$$F_T(t,x) = \text{Prob}\ \{T(x) \leq t\} \qquad (14)$$

The unconditional completion time T is characterized by the following distribution:

$$F_T(t) = \text{Prob}\ \{T \leq t\} = \int_0^\infty F_T(t,x)dG(x) \qquad (15)$$

The distribution of the completion time, $F_T(t,x)$ incorporates the effect of a random variation of the execution speed consequent to a degradation and reconfiguration process, combined with the effect of the preemption and recovery policy on the execution of the task.

A unified framework for the evaluation of the Cdf of the accumulated reward and the task completion time has been proposed in [55] by resorting to a SRM. The task in progress is served at a rate r_i when the system is in the structure state i. When a transition from state i to state j occurs in $Z(t)$, the task execution resumes in state j at the new rate r_j if state i is of prs type. If state i is of pri type, the preempted task is restarted with rate r_j in the new state j and with a work requirement identical to the one of the preempted task. If state i is of prd type, the preempted task is restarted in the new state j with the reward rate r_j, and with a work requirement W resampled from the same distribution $G(x)$.

The following relationships between the different preemption policies can be easily established. If the work requirement W is an exponential random variable, the two policies prs and prd give rise to the same completion time (due to the memoryless property of the exponential distribution, the residual task requirement under the prs policy coincides with the resampled requirement under the prd policy). On the other hand, if W is deterministic, the two policies pri and prd are coincident (resampling a step function provides always the same constant value).

Moreover, assuming that the structure-states are all of prs type, so that no loss of reward occurs, the distribution of the completion time is closely related to the distribution of the accumulated reward by means of the following relation:

$$\text{Prob}\ \{Y(t) \leq x\} = \text{Prob}\ \{T(x) > t\} \qquad (16)$$

so that:

$$F_Y(x,t) = 1 - F_T(t,x) \qquad (17)$$

Kulkarni et al. [55] derived the closed form Laplace transform equations of $F_T(x,t)$ when $Z(t)$ is a CTMC and all the states belong to the same preemption class. The extension to a semi-Markov $Z(t)$ process whose state space is partitioned in the three preemption classes has been considered in [54]. Bobbio and Trivedi [12] studied the case where $Z(t)$ is a CTMC, the work requirement W is a PH random variable [66] and the task execution policy is a probabilistic mixture of prs and prd policies. The combination of prs and pri policies has been investigated in [14] having as an object the evaluation of the completion time of a program on a gracefully degradable computing system.

An alternative interpretation of the completion time problem can be given in terms of the hitting time of an appropriate cumulative functional [18], defined on $Z(t)$, against an absorbing barrier. The idea of a cumulative functional was first suggested by Kulkarni et al. [55] and then explicitly stated and exploited in [10]. Let $S(t)$ be a functional that depends on $Z(x)$ for $x \leq t$ [18] and that represents the accumulation of the reward in time according to the preemption policies assigned to each state. Hence, if all the states are prs, the functional $S(t)$ coincides with the accumulated reward $Y(t)$ in (8), while in the presence of a mixture of policies, $S(t)$ is the reward accumulated since the last passage in a pri or prd state. Furthermore, let the work requirement W act as an absorbing barrier for the functional $S(t)$. Then, the completion time $T(x)$ can be interpreted as the time at which the functional $S(t)$ hits an absorbing barrier located at $W = x$ for the first time. The above interpretation of the completion time problem has proved to be useful, for instance, in association with the use of Stochastic Petri nets [9] and with the extension to multi-reward models for the representation of the simultaneous execution of parallel tasks with different rates [10], [8].

2.3. Cumulative measures with a binary reward variable

When the reward rates take only integer binary values in the set $\{0, 1\}$, the SRM reduces to a standard reliability model in which each structure-state can represent either an up or a down condition. The SRM model provides, in this case, a unified theoretical framework for a variety of problems that have been separately considered in the reliability literature. Two typical examples are illustrated.

CASE 1 $\quad r_i = 1;\ i \in UP\ STATES$ and
$\quad r_i = 0;\ DOWN\ STATES$

A reward rate equal to 1 is assigned to the up states, so that the accumulated reward provides the total up time. The interval availability $A_I(t)$ becomes:

$$A_I(t) = \frac{Y(t)}{t} \qquad (18)$$

A closed form expression for the Cdf of the interval availability has been derived in [22] for a two state system; this expression involves an infinite series of Bessel functions. The numerical computation of the interval availability has been addressed in [33] by using a finite difference equation and in [24] by resorting to the randomization technique (see Paragraph 3.2.4).

CASE 2 $\quad r_i = 0;\ i \in UP\ STATES$ and
$\quad r_i = 1;\ DOWN\ STATES$

A subtle and interesting reliability problem concerns the determination of the system lifetime when the downtime exceeds a fixed threshold. In [25], a solution has been formulated for a two state exponential system; the case with general distributions has been analyzed in [76] and generalized in [71], [77]. In [31], the lifetime of a system under various critical constraints, including single and cumulative downtime thresholds, has been examined. In the framework of the present theory, the system lifetime can be formulated in terms of the completion time of a task whose work requirement is equal to the assigned downtime threshold [68]. In fact, due to the reward rate assignments, "useful" work is consumed only when the system is in the down states, so that the completion time is related to the sojourn time in the down states. If the down states are either of pri or prd type, a fatal failure occurs when a single sojourn in the down states exceeds the threshold, while if the down states are all of prs type, the fatal failure arises when the total accumulated downtime reaches the threshold level. Nicola et al. [68] have calculated the completion time under fairly more general conditions, and have directly derived several related measures from the knowledge of the completion time distribution.

3. NUMERICAL ALGORITHMS

Even if there is an intimate connection between the accumulated reward and the completion time, the numerical evaluation of the respective Cdf's has usually been considered separately. Almost all the algorithms have been worked out for the Cdf of the accumulated reward; then, by virtue of (17), they can be adapted to the computation of the distribution of the completion time if all the states are prs. The incorporation of preemption policies other than prs does not seem always easy or even compatible with the developed algorithms. A few efforts have been devoted to the problem of computing the distribution of the completion time taking into account execution policies which entails the loss of work. A common assumption to the numerical procedures presented in the literature is that the underlying structure-state process $Z(t)$ is a CTMC. This assumption is tacitly implied in the sequel, unless otherwise stated.

3.1. Acyclic CTMC

Meyer [60] elaborated a closed form solution of $F_Y(x,t)$ for a non-repairable two-processor system with input buffer. Meyer's methodology has been generalized in [26] to any acyclic-CTMC. The method is based on the enumeration of all possible connected paths (i.e. sequences of visited states) in the Markov graph from the initial good state to an absorbing failed state, and in the computation of the Cdf of the accumulated reward conditioned to a single path. The complexity of the algorithm grows with the number of paths, and therefore, is exponential in the number of states.

A restricted class of SRM with acyclic CTMC has been considered in [32]. The restriction implies a progressive reduction in the computation capacity going from the good state to the absorbing state so that the reward rates are decreasing over all the trajectories (i.e. $r_i \geq r_j$ if a transition is possible from state i to state j). A time-domain closed form solution of the problem has been determined under the conditions:

$$\begin{cases} 1) & \frac{\lambda_i-\lambda_k}{\lambda_j-\lambda_k} \neq \frac{r_i-r_k}{r_j-r_k} \\ \text{and} & \\ 2) & \text{either } \lambda_i \neq \lambda_j \\ 3) & \text{or } r_i \neq r_j \end{cases} \tag{19}$$

Condition 2) prevents the CTMC to have equal eigenvalues (distinct states must have distinct holding times), while condition 1) is violated for models in which the reward rates are proportional to the holding times. The proposed recursive algorithm has a complexity of the order of $O(n^3)$, being n the cardinality of $Z(t)$.

The restriction on the decreasing ordering of the reward rates is relaxed in [21] and [17], but conditions 1 and 2 in (19) are still maintained. The double Laplace transform of (10) is evaluated and then inverted by means of a partial fraction expansion (with different eigenvalues) to get a closed form for the residues. The proposed recursive algorithm has a complexity of $O(n^3)$ for the computation of the Cdf and a complexity of $O(n^2)$ when the procedure is limited to the computation of the moments.

3.2. General CTMC

For general CTMC, the solution procedures are much more intricate [73]. For a two-state CTMC with binary reward rates the closed form expression of the distribution of the interval availability involves an infinite series of Bessel functions [22]. Therefore, no closed form solution can be practically invoked as a basis for computational procedures. The evaluation of the Cdf should be based on numerical algorithms. Distinct methodologies have been proposed in the literature.

3.2.1. Evaluation of the moments of the Cdf

From the matrix equation (12), Iyer et al. [44] have derived a recursive expression for generating the $(m+1)-th$ moment of $F_Y(x,t)$ knowing the $m-th$ moment. The recursion involves the exponential of matrix **Q**, so that any algorithm able to solve a matrix exponential [63] can be applied to this problem. The authors developed an algorithm based on the spectral representation of the transition rate matrix **Q** of $Z(t)$ that requires the prior computation of the eigenvalues of **Q**. In the general case, the algorithm to compute the first m moments has a complexity of the order of $O(n^4 m^2)$, while in the acyclic case the algorithm reduces to the one described in [22].

Pattipati and Shah [70] extend the moment recursion to a non-homogeneous Markov process and propose a cascade system representation. Moreover, since the eigenvalue-eigenvector method to compute the exponential of the Markov generator is known to be unstable especially when **Q** has a set of nearly equal eigenvalues, they propose two alternative algorithms. The first one is based on the representation of the exponential of the generator by a power series expansion which in turn is not very efficient for stiff matrices where an integration step very short with respect to the interval of integration is required. The second method employs diagonal Pade approximation to compute the exponential of a matrix over a small interval coupled with the concept of doubling which provides a significant saving particularly for long mission time analysis.

3.2.2. Inversion of the double Laplace Transform

Equation (12) provides a natural starting point for the development of numerical procedures based on the inversion of the double Laplace transform. A two-phase transform inversion of Equation (12) has been first suggested by Kulkarni et al. in [53], [81]; an improved algorithm along the same line has been presented in [79]. The first phase consists in determining the eigenvalues of the matrix $(u\mathbf{R} - \mathbf{Q})$ by a QR numerical technique and then in inverting analytically Equation (12) with respect to the time transform variable s by a partial fraction expansion. This last operation requires the computation of the coefficient of the expansion by solving a set of n linear equations. The second phase involves the numerical inversion of the resulting single Laplace transform, with respect to the reward transform variable u. This second inversion is performed by approximating the Laplace transform by a truncated Fourier series [45]. A substantial speed-up of the algorithm [79] has been achieved over earlier versions [53], by reducing the computational cost to the order of $O(n^3 l)$ where l is the number of terms in the truncated Fourier series.

A different approach has been proposed in [43] by resorting to the Laguerre transform method, introduced by Keilson and Nunn [48], Keilson et al. [49], and further studied by Sumita [82]. The method provides an algorithmic framework for mapping continuum operations into lattice operations, and can be applied to invert the double Laplace transform equation (12) in terms of a double infinite series of Laguerre functions. The coefficients of the Laguerre functions are computed by a recursion involving various operation (including an inversion) on the matrices **Q** and **R**. The method can be adapted to solve the Cdf of the accumulated reward, the density or the moments. The algorithm contains a criterion for estimating the error inherent in the truncation of the infinite series after l terms and its computational complexity is $O(n^3 l)$.

3.2.3. Solution of a set of Partial Differential Equations

The density of the accumulated reward $f_{Yi}(x,t)$, conditioned to the initial state of $Z(t)$ is:

$$f_{Yi}(x,t) = \frac{dF_{Yi}(x,t)}{dx} \tag{20}$$

By denoting with $\underline{f}_Y(x,t)$ the n-dimensional vector whose entries are the conditioned densities of (20), the following partial differential equation has been derived in [80]:

$$\mathbf{R}\,\frac{\partial \underline{f}_Y(x,t)}{\partial x} + \frac{\partial \underline{f}_Y(x,t)}{\partial t} = \mathbf{Q}\,\underline{f}_Y(x,t) \tag{21}$$

This linear hyperbolic system may be solved by explicit finite difference methods. The complexity of these methods depends in general on qt, being q the maximum diagonal element of **Q** in absolute value, and t the interval of integration. On the other hand, these methods can be easily implemented using sparse matrix storage, while those based on the transform inversion require full-matrix computation.

A comparison of the above technique with those based on the double transform inversion has been discussed in [73], where it is suggested that methods based on the solution of the set of partial differential equations may be more effective for relatively large problems (hundreds of states) with moderate degree of stiffness. Further research and experimentation is needed to fully understand and exploit the capabilities of the direct numerical solution of Equation (21) in the time domain.

3.2.4. Uniformization

The theoretical bases for the technique called uniformization (or randomization) have been set by Keilson [47], who showed that a CTMC can be converted into a discrete-time Markov chain subordinated to a Poisson process.

The technique has gained widespread acceptance for the numerical handling of problems involving CTMC [37], [38], [59], [74], because of the following main advantages:

a) the procedure involves a power series of matrices whose elements are all non-negative, thus avoiding the numerical troubles that arise in methods based on the infinitesimal generator of the CTMC whose diagonal elements are negative;
b) error bounds are easily computed so that the infinite series may be truncated at a level of predetermined accuracy;
c) the procedure is easily implementable;
d) the procedure can be adapted to compute cumulative measures [59], [75] and sensitivities [40].

These favourable properties have driven some researchers to explore the applicability of the uniformization technique also in the area of SRM for the numerical evaluation of the Cdf of cumulative measures. The first attempts were addressed to the computation of the first two moments either of the total time spent in a subset of states [36], or of the accumulated reward [83].

De Souza e Silva and Gail [24] have applied the uniformization procedure to the computation of the distribution of the cumulative operational time. They have then extended the approach to several other performability measures in [23]. The esssence of the method consists in marking the time intervals of a possible realization of $Z(t)$ with different colors depending on the reward rate value associated to each state (equal rate-equal color), and then in studying the statistics of the marked intervals. This procedure is intuitively very appealing but, since the enumeration of all possible combinations of k different colors (reward rates) over $n > k$ time intervals is required, the computational complexity grows exponentially with the number of different reward values. Recently, Donatiello and Grassi [20] have proposed a uniformization algorithm for the Cdf of the accumulated reward which is polinomial in the number of states of the CTMC and in the number of the different reward rates. Their method is based on the inversion of the Laplace transform of the uniformized distribution function. An application of the technique to the computation of the sensitivity of cumulative measures has been presented in [34].

3.2.5. Markovianization (The hitting time problem)

This method has been particularly addressed to the computation of the distribution of the completion time; the adjustment to the more classical problem of the computation of the distribution of the accumulated reward is discussed next. The method is based on the interpretation of $F_T(t, x)$ as the distribution of the hitting time of a properly defined functional against an absorbing barrier of height $W = x$. The construction of the functional may accomodate complex mixtures of prs and prd policies.

The method is further based on the notion of Phase Type (PH) random variable. A random variable is PH of order m [66], [19] if it is the time to absorption of a CTMC with m transient states and one absorbing state. PH distributions have found wide application as a computational valuable tool in many areas of applied stochastic modelling, since they offer a bridge for representing non-exponential random variables by means of combinations of exponential stages.

It has been proved in [12] that, if the work requirement W is a PH random variable of order m the distribution of the unconditional completion time $F_T(t)$ in (15) can be computed as the time to absorbtion of a CTMC with $(n \times m)$ transient states and one absorbing state. The completion time is thus a PH random variable of order $(n \times m)$. This result extends the reliability operations for which the closure property of PH distributions was known to hold [2].

This technique converts the continuum problem in the Laplace transform domain into a discrete problem in the time domain; the $(n \times m)$ transition rate matrix of the transformed CTMC has been derived in closed form in [12] as a function of **Q** and **R** under a probabilistic mixture of prs and prd policies. By probabilistic mixture of policies it is meant that each transition outgoing from a generic state i has a probability ρ_i of behaving as a prs transition and a complementary probability $(1 - \rho_i)$ of behaving as a prd transition.

The technique is exact if the work requirement W is actually a PH distribution, while provides approximate results if W is non-PH and is approximated by a PH distribution. The computational complexity of the method equals that of solving a transient CTMC with $(n \times m)$ transient states (see, for instance [74]). The minimal completion time in a multi-reward Markov model has been discussed in [10] following the same idea.

In order to exploit the above procedure for the evaluation of $F_Y(x, t)$, the original problem should be recast in the following way:

- the problem scale should be converted from the time domain to the computational domain following Beaudry's technique [5] (see Paragraph 2.2.2), with the extensions proposed in [83], [15] when a zero reward rate is assigned to a transient state;
- the time t at which $F_Y(x, t)$ is computed should be considered as a random variable U with distribution $\Gamma(x)$ [27]. We can thus define an unconditional accumulated reward Y as:

$$F_Y(x) = \text{Prob}\ \{Y \leq x\} = \int_0^\infty F_Y(x, t) d\Gamma(t) \qquad (22)$$

The unconditional accumulated reward Y converges to the accumulated reward $Y(t)$ when $\Gamma(t)$ converges to the step function located $U = t$.

If the observational time U is assumed to be a PH random variable of order m, the unconditional Cdf in (22) can be modelled as the time to absorption of the $(n \times m)$ CTMC generated in the computational domain.

4. CONCLUSION

The paper has surveyed the motivation for a combined evaluation of performance and reliability measures in complex degradable systems, and has shown that a fruitful approach for merging into a unified framework the fault-free characteristics of a system and the variation of these characteristics in time is the use of SRM.

The superposition of a reward rate variable to the state space of the structure-state process increases the flexibility and the modelling power of the technique. Various measures can be introduced in order to characterize distinct aspects of a SRM. The instantaneous measures and the expected values of the cumulative measures require a computational load of the same order of magnitude as the one needed for evaluating the transient state probabilities of the original structure-state process.

The computation of the Cdf of the cumulative measures is, on the other hand, a more intricate and challenging task. The paper has introduced two cumulative measures, namely, the accumulated reward and the completion time, and has discussed their physical meaning and their mutual relationships. Finally, a tentatively exhaustive list of the various methods proposed in the literature for evaluating the respective Cdf's is reported. This list suggests the following remarks:

- No technique seems, at the moment, to definitely prevail on the other ones. Further numerical work is desirable and a comparison of the different techniques on some benchmark problem could be very illuminating.
- Most of the algorithms have been thought for the accumulated reward, i.e. for systems which do not entail loss of work during their life. The extension to more flexible and realistic execution or accumulation policies is an area of interesting future research.
- The integration of the techniques discussed in this paper into more general and user friendly reliability packages [46] is crucial for the diffusion of these concepts in a more applicative environment. Petri nets have proved to be quite useful and efficient [9], [16] also from this point of view.

The analysis implied in the concepts expressed in this paper is costly; however, practical situations have been envisaged in which the simpler knowledge of the expected values is not sufficient to completely characterize the system behaviour in time, and more stringent specifications based on the Cdf or on some percentile points [57], [72] need to be fulfilled.

REFERENCES

[1] M. Ajmone Marsan, G. Balbo and G. Conte, "A class of generalized stochastic Petri nets for the performance evaluation of multiprocessor systems", *ACM Transactions on Computer Systems*, 2:93–122, 1984.

[2] D. Assaf and B. Levikson, "Closure of phase type distributions under operations arising in reliability theory", *The Annals of Probability*, 10:265–269, 1982.

[3] A. Avizienis and J.C. Laprie, "Dependable computing: from concepts to design diversity", *Proceedings IEEE*, 74:629–638, 1986.

[4] R.E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, New York, 1975.

[5] M.D. Beaudry, "Performance-related reliability measures for computing systems", *IEEE Transactions on Computers*, C-27:540–547, 1978.

[6] A. Bobbio, "Dependability analysis of fault-tolerant systems: a literature survey", *Microprocessing and Microprogramming*, 29:1–13, 1990.

[7] A. Bobbio, "The effect of an imperfect coverage on the optimum degree of redundancy of a degradable multiprocessor system", In *Proceedings RELIABILITY'87*, Paper 5B/3, Birmingham, 1987.

[8] A. Bobbio, "A multi-reward stochastic model for the completion time of parallel tasks", In *Proceedings 13-th International Teletraffic Congress, ITC-13*, Copenhagen, 1991.

[9] A. Bobbio, "Petri nets generating Markov reward models for performance/reliability analysis of degradable systems;", In R. Puigjaner and D. Poitier, editors, *Modelling Techniques and Tools for Computer Performance Evaluation*, pp.353–365, Plenum Press, 1989.

[10] A. Bobbio and L. Roberti, "The hitting time in multi-reward semi-Markov models", In *Proceedings 1st International Workshop on Performability Modelling of Computer and Communication Systems*, pp.57–76, University of Twente – Enschede (NL), 1991.

[11] A. Bobbio and K.S. Trivedi, "An aggregation technique for the transient analysis of stiff Markov chains", *IEEE Transactions on Computers*, C-35:803–814, 1986.

[12] A. Bobbio and K.S. Trivedi, "Computation of the distribution of the completion time when the work requirement is a PH random variable", *Stochastic Models*, 6:133–149, 1990.

[13] X. Castillo and D.P. Siewiorek, "A performance reliability model for computing systems", In *Proceedings 10-th International Symposium on Fault-Tolerant Computing*, pp.187–192, 1980.

[14] P.F. Chimento and K.S. Trivedi, "The performance of block structured programs on processors subject to failure and repair", In E. Gelenbe, editor, *High Performance Computer Systems*, pp.269–280, North Holland, 1988.

[15] G. Ciardo, R.A. Marie, B. Sericola and K.S. Trivedi, "Performability analysis using semi-Markov reward processes", *IEEE Transactions on Computers*, 39:1252–1264, 1990.

[16] G. Ciardo, J. Muppala and K.S. Trivedi, "On the solution of GSPN reward models", To appear on: *Performance Evaluation*, 1991.

[17] B. Ciciani and V. Grassi, "Performability evaluation of fault-tolerant satellite systems", *IEEE Transactions on Communications*, COM-35:403–409, 1987.

[18] E. Cinclar, "Markov renewal theory", *Advances in Applied Probability*, 1:123–187, 1969.

[19] A. Cumani, "Esp – A package for the evaluation of stochastic Petri nets with phase-type distributed transition times", In *Proceedings International Workshop Timed Petri Nets*, pp.144–151, IEEE Computer Society Press No. 674, Torino (Italy), 1985.

[20] L. Donatiello and V. Grassi, "On evaluating the cumulative performance distribution of fault-tolerant computer systems", To appear on: *IEEE Transactions on Computers*, 1991.

[21] L. Donatiello and B.R. Iyer, "Analysis of a composite performance reliability measure for fault-tolerant systems", *Journal of the ACM*, 34:179–199, 1987.

[22] L. Donatiello and B.R. Iyer, "Closed-form solution for system availability distribution", *IEEE Transactions on Reliability*, R-36:45 – 47, 1987.

[23] E. De Souza e Silva and H.R. Gail, "Calculating availability and performability measures of repairable computer systems using randomization", *Journal of the ACM*, 36:171 – 193, 1989.

[24] E. De Souza e Silva and H.R. Gail, "Calculating cumulative operational time distributions of repairable computer systems", *IEEE Transactions on Computers*, C-35:322 – 332, 1986.

[25] A. von Ellenrieder and A. Levine, "The probability of an excessive non-functioning interval", *Operations Research*, 14:835 – 840, 1966.

[26] D.G. Furchtgott and J.F. Meyer, "A performability solution method for degradable non-repairable systems", *IEEE Transactions on Computers*, C-33:550 – 554, 1984.

[27] D.P. Gaver, "Observing stochastic processes and approximate transform inversion", *Operations Research*, 14:444 – 459, 1966.

[28] D.P. Gaver, "A waiting line with interrupted service, including priorities", *Journal of the Royal Statistical Society*, B24:73 – 90, 1962.

[29] F.A. Gay and M.L. Ketelsen, "Performance evaluation for gracefully degrading systems", In *Proceedings 9-th International Symposium on Fault-Tolerant Computing*, pp.51 – 58, 1979.

[30] A Goyal, S. Lavenberg and K.S. Trivedi, "Probabilistic modeling of computer system availability", *Annals of Operations Research*, 8:285 – 306, 1987.

[31] A. Goyal, V.F. Nicola, A.N. Tantawi and K.S. Trivedi, "Reliability of systems with limited repair", *IEEE Transactions on Reliability*, R-36:202 – 207, 1987.

[32] A. Goyal and A.N. Tantawi, "Evaluation of performability for degradable computer systems", *IEEE Transactions on Computers*, C-36:738 – 744, 1987.

[33] A Goyal and A.N. Tantawi, "A measure of guaranteed availability and its numerical evaluation", *IEEE Transactions on Computers*, C-37:25 – 32, 1988.

[34] V. Grassi and L. Donatiello, "Sensitivity analysis of performability", In *Proceedings 1st International Workshop on Performability Modelling of Computer and Communication Systems*, pp.43 – 56, University of Twente – Enschede (NL), 1991.

[35] V. Grassi, L. Donatiello and G. Iazeolla, "Performability evaluation of multicomponent fault-tolerant systems", *IEEE Transactions on Reliability*, R-37:216 – 222, 1988.

[36] W.K. Grassmann, "Means and variances of time averages in markovian environment", *European Journal of Operational Research*, 31:132 – 139, 1987.

[37] W.K. Grassmann, "Transient solution in markovian queueing systems", *Computers and Operations Research*, 4:47 – 56, 1977.

[38] D. Gross and D. Miller, "The randomization technique as a modeling tool and solution procedure for transient Markov processes", *Operations Research*, 32:343 – 361, 1984.

[39] B.R. Haverkort, *Performability modeling, tools, evaluation techniques ana applications*, Phd thesis, Computer Science Dept, University of Twente – Enschede (NL), 1990.

[40] P. Heidelberger and A. Goyal, "Sensitivity analysis of continuous time Markov chains using uniformization", In P. Courtois G. Iazeolla and O. Boxma, editors, *Computer Performance and Reliability*, pp.93 – 104, North Holland, 1988.

[41] R.A. Howard, *Dynamic Probabilistic Systems, Volume II. semi-Markov and Decision Processes*, John Wiley and Sons, New York, 1971.

[42] R. Huslende, "A combined evaluation of performance and reliability for degradable systems", In *Proceedings ACM/SIGMETRICS Conference*, pp.157 – 164, 1981.

[43] S.M.R. Islam and H.H. Ammar, "Performability of the hypercube", *IEEE Transactions on Reliability*, 38:518 – 526, 1989.

[44] B.R. Iyer, L. Donatiello and P. Heidelberger, "Analysis of performability for stochastic models of fault-tolerant systems", *IEEE Transactions on Computers*, C-35:902 – 907, 1986.

[45] D.L. Jagerman, "An inversion technique for the Laplace transform", *The Bell System Technical Journal*, 61:1995 – 2002, October 1982.

[46] A.M. Johnson and M. Malek, "Survey of software tools for evaluating reliability, availability and serviceability", *ACM Computing Surveys*, 20:227 – 269, 1988.

[47] J. Keilson, *Markov Chain Models – Rarity and Exponentiality*, Springer-Verlag, Berlin, 1979.

[48] J. Keilson and W.R. Nunn, :Laguerre transformation as a tool for the numerical solution of integral equations of convolution type", *Applied Mathematics and Computation*, 5:313 – 359, 1979.

[49] J. Keilson, W.R. Nunn and U. Sumita, "The bilateral Laguerre transform", *Applied Mathematics and Computation*, 8:137 – 174, 1981.

[50] J. Keilson and S.S. Rao, "A process with chain dependent growth rate" *Journal of Applied Probabiality*, 7:699 – 711, 1970.

[51] J. Keilson and S.S. Rao, "A process with chain dependent growth rate. Part II: the ruin and ergodic problems", *Advances in Applied Probability* 3:315 – 338, 1971.

[52] H.M. Khelalfa and A.K. von Mayrahauser, "Models to evaluate trade-offs between performance and reliability", In *Proceedings CMG XV*, pp.24 – 29, San Francisco, 1984.

[53] V.G. Kulkarni, V.F. Nicola, R.M. Smith and K.S. Trivedi, "Numerical evaluation of performability and job completion time in repairable fault-tolerant systems", In *Proceedings of the 16-th Symposium on Fault-Tolerant Computing*, pp.252 – 257, 1986.

[54] V.G. Kulkarni, V.F. Nicola and K.S. Trivedi, "The completion time of a job on a multi-mode system", *Advances in Applied Probability*, 19:932 – 954, 1987.

[55] V.G. Kulkarni, V.F. Nicola and K.S. Trivedi, "On modeling the performance and reliability of multi-mode computer systems", *The Journal of Systems and Software*, 6:175 – 183, 1986.

[56] J.C. Laprie, "Dependable computing and fault-tolerance: concepts and terminology", In *Proceedings 15-th International Symposium on Fault-Tolerant Computing*, pp.2 – 7, 1985.

[57] Y. Levy and P.E. Wirth, "A unifying approach to performance and reliability objectives", In *Proceedings 12-th International Teletraffic Congress, ITC-12*, pp.4.2B2.1 – 4.2B2.7, Torino, 1988.

[58] R.A. McLean and M.F. Neuts, "The integral of a step function defined on a semi-Markov process", *SIAM Journal on Applied Mathematics*, 15:726 – 737, 1967.

[59] B. Melamed and M. Yadin, "Randomization procedure in the computation of cumulative-time distributions over discrete state Markov processes", *Operations Research*, 32:926 – 944, 1984.

[60] J.F. Meyer, "Closed form solution of performability", *IEEE Transactions on Computers*, C-31: 648 – 657, 1982.

[61] J.F. Meyer, "On evaluating the performability of degradable systems", *IEEE Transactions on Computers*, C-29:720 – 731, 1980.

[62] J.F. Meyer, "Performability: a retrospective and some pointers to the future", In *Proceedings 1st International Workshop on Performability Modelling of Computer and Communication Systems*, University of Twente – Enschede (NL), 1991.

[63] C. Moler and C.F. Van Loan, "Nineteen dubios ways to compute the exponential of a matrix", *SIAM Review*, 20:801 – 835, 1978.

[64] J.A. Munarin, "Dynamic workload model for performance/reliability analysis of gracefully degrading systems", In *Proceedings 13-th International Symposium on Fault-Tolerant Computing*, pp.290 – 295, 1983.

[65] J.K. Muppala, S.P. Woolet and K.S. Trivedi, "Composite performance and dependability analysis", In *Proceedings 1st International Workshop on Performability Modelling of Computer and Modelling of Computer and Communication Systems*, pp.131 – 161, University of Twente – Enschede (NL), 1991.

[66] M.F. Neuts, *Matrix Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, Baltimore, 1981.

[67] V.F. Nicola, "A single server queue with mixed types of interruptions", *Acta Informatica*, 23:465 – 486, 1986.

[68] V.F. Nicola, A. Bobbio and K.S. Trivedi, *A unified performance reliability analysis of a system with a cumulative downtime constraint*, IBM Research Report RC 15279 (# 68179) – To appear on: *Microelectronics and Reliability*, 1991.

[69] V.F. Nicola, V.G. Kulkarni and K.S. Trivedi, "Queueing analysis of fault-tolerant computer systems", *IEEE Transactions on Software Engineering*, SE-13:363–375, 1987.

[70] K.R. Pattipati and S.A. Shah, "On the computational aspects of performability models of fault-tolerant computer systems", *IEEE Transactions on Computers*, 39:832–836, 1990.

[71] K.G.. Ramamurthy and N.K. Jaiswal, "A two-dissimilar-unit cold standby system with allowed downtime", *Microelectronics and Reliability*, 22:689–691, 1982.

[72] A. Reibman, "Modelling the effect of reliability on performance", *IEEE Transactions on Reliability*, R-39:314–320, 1990.

[73] A. Reibman, R. Smith and K.S. Trivedi, "Markov and Markov reward model transient analysis: an overview of numerical approaches", *European Journal of Operational Research*, 40:257–267, 1989.

[74] A. Reibman and K.S. Trivedi, "Numerical transient analysis of Markov models", *Computers and Operations Research*, 15:19–36, 1988.

[75] A. Reibman and K.S. Trivedi, "Transient analysis of cumulative measures of Markov chain behavior", *Stochastic Models*, 5:683–710, 1989.

[76] S.M. Ross and J. Schechtman, "On the first time a separately maintained parallel system has been down for a fixed time", *Naval Research Logistic Quarterly*, 26:285–290, 1979.

[77] G. Shanthikumar, "First failure time of dependent parallel systems with safety period" *Microelectronics and Reliability*, 26:955–972, 1986.

[78] M.L. Shooman, *Probabilistic reliability: an engineering approach*, McGraw Hill, 1968.

[79] R. Smith, K.S. Trivedi and A.V. Ramesh, "Performability analysis: measures, an algorithm and a case study", *IEEE Transactions on Computers*, C-37:406–417, 1988.

[80] R.M. Smith, *Markov Reward Models: Application Domains and Solution Methods*, Phd thesis, Computer Science Dept, Duke University, Durham, NC, 1987.

[81] R.M. Smith and K.S. Trivedi, "A performability analysis of two multi-processor systems" In *Proceedings 17-th International Symposium on Fault-Tolerant Computing*, pp.224–229, 1987.

[82] U. Sumita, "The matrix Laguerre transform", *Applied Mathematics and Computation*, 15:1–28, 1984.

[83] U. Sumita, J.G. Shanthikumar and Y. Masuda, "Analysis of fault-tolerant computer systems", *Microelectronics and Reliability*, 27:65–78, 1987.

[84] K.S. Trivedi, *Probability & Statistics with Reliability, Queueing & Computer Science Applications*, Prentice-Hall, 1982.

[85] K.S. Trivedi, A.S. Sathaye, O.C. Ibe and R.C. Howe, "Should I add a processor?", In *Proceedings 23-rd Annual Hawaii International Conference on System Sciences HICSSS-23*, pp.214–221, 1990.

Andrea Bobbio graduated in nuclear engineering from Politecnico di Torino in 1969. In 1971 he joined the Istituto Elettrotecnico Nazionale Galileo Ferraris, where he was involved in researches on degradation mechanisms in electronic devices. In 1974 he was assigned to the staff for setting up an Italian Reliability Data Bank on electronic components in cooperation with Italian electronic industries. From 1978 he turned his attention to the modeling and analysis of the performance and reliability of complex systems, with particular emphasis to the areas of Markov models, stochastic reward models and Petri nets. During 1984, 1986 and 1987 he has been visiting scientist in the Department of Computer Science at the Duke University, Durham NC (USA), and in 1989 in the Institute of Electronics and Telecommunication of the Technical University of Budapest. He is author of several reliability oriented papers, and he has been principal investigator in several grants with public and private Institutions.

EUROPEAN M.SC. STUDIES IN RELIABILITY: A TEMPUS PROJECT

A Joint European Project — Extended European Masters Course Network in Safety and Reliability — has been started within TEMPUS aimed at promoting reliability engineering education at the Technical University of Budapest.

TEMPUS is the acronym for the Trans-European Mobility Scheme for University Studies, adopted by the European Community. The main goals of TEMPUS scheme are to promote the quality and support the development and renewal of the higher education systems in the countries of Central/Eastern Europe, and to encourage their growing interaction and cooperation with partners in the European Community, through joint activities and relevant mobility.

Three years ago the European Masters Course in Safety and Reliability was set up with financial support from COMETT. This is a modular course consisting of intensive one-week modules, the number currently on offer being up to 24. Some modules recently offered are:

- Statistics for Reliability
- Reliability Assessment
- Systems Reliability Calculations
- Applied Fault Tree Analysis
- Maintenance and Safety
- Human factors in Reliability
- Emergency and Crisis Management
- Hazard and Risk in the Process Industry
- Software Reliability

Candidates for the course will normally be expected to hold a first degree, or equivalent qualification in a science, engineering or mathematical subject.

Students are required to complete ten modules within a period of three years. Each module is assessed by examination, case study or project. On completion of the ten modules the student submits a major dissertation.

The European dimension is imparted by the fact that the courses are presented at different centres, mostly at universities throughout Europe by experts in each area of safety and reliability studies. Coordination of the scheme is undertaken by the University of Bradford.
The participants of the TEMPUS project are:

University of Bradford,
Department of Industrial Technology

Technical University of Eindhoven, Netherland
Delft University of Technology, Netherland
Hogeschool Zeeland, Vlissingen, Netherland

University of Würzburg, Germany
Institute für Angewandte Mathematik und Statistik

Universidad Politecnica de Madrid
Escuela Tecnica Superior de Ingenieros Navales

University of Limerick, Ireland

European Liaison Centre

Norges Tekniske Hogskole
Institute for Maskinkonstruksjon

Technische Universität Wien
Institut für Statistik und Wahrscheinlichkeitstheorie

Technical University of Budapest
Department of Precision Mechanics and Optics

Paks Nuclear Power Plant, Hungary

Each participant in the network will bring into the project its own specific expertise in the field of Safety and Reliability and related studies as well as its experience gained to date in running modular courses for the European Masters Course in Safety and Reliability. The Department of Precision Mechanics and Optics will be the coordinator on behalf of the Technical University of Budapest.

The following courses held by guest lecturers would run before August 1992 in Budapest:

Risk Management (University of Bradford),
Design of Experiments (University of Limmerick),
Product Control (University of Würzburg).

Industrial partners – Paks Nuclear Power Plant, and Digital Equipment (Hungary) Ltd — attending the meeting pointed out their interest in the TEMPUS programme and assured the participants about considering futher support.

György Barta

COOPERATION BETWEEN AN ITALIAN INSTITUTE AND A HUNGARIAN UNIVERSITY

In 1988 László Jereb, associate professor at the Technical University of Budapest (TUB) spent some months in Torino on the basis of a bilateral scientific agreement between Italy and Hungary. During that visit he got in touch with Andrea Bobbio from Instituto Electrotecnico Nazionale Galileo Ferraris (IEN GF) who was engaged in similar researches in the field of reliability theory.

In the following years the cooperation went on harder. Andrea Bobbio visited the TUB several times, he gave some lectures on conferences and as a guest professor at TUB.

In the last phase the President of the IEN GF and the Dean of the electrical engineering faculty of TUB recognized officially the cooperation, and the President of the IEN GF offered a scholarship to a Hungarian researcher for two months in the autumn of 1991.

With the agreement of Andrea Bobbio and László Jereb, I arrived in Torino for studying numerical methods for the transient analysis of stiff continuous time Markov chains and for implementing them in computer programs and comparing their behaviours. The investigated methods are the Taylor series of matrix exponential, the randomization (or uniformization) method and implicit and explicit differential equation solvers like Runge-Kutta and Heun (or trapezoid rule) associated to BDF2. The result of this visit is that the methods are implemented not only in a test program, but also in a Petri net analyzer program package (called ESP) developed at IEN in previous years.

This cooperation is very fruitful because the ESP package becomes a more powerful integrated tool for system analysis, and moreover it opens the way to examine the reliability behaviour of such kind of large systems like communication networks.

Some opportunities for further steps in the field of reliability and safety theory can be envisaged. Moreover a possible matter of future work is the implementation of theory in computer programs even as in Torino as in Budapest. The chance will be explored to take part in common projects.

Miklós Telek

■ IEEE HUNGARY SECTION

The Institute of Electrical and Electronics Engineers (IEEE), the world's largest technical professional society, was founded in 1884, and now it has about 350000 members from more than 130 countries. The organization of IEEE contains 10 Regions which are divided into Sections. The technical objectives of the IEEE focus on advancing the theory and practice of electrical, electronics and computer engineering and computer science. To realize these objectives, the IEEE sponsors technical conferences, symposia and local meetings all over the world, publishes nearly 25% of the world's technical papers in electrical, electronics and computer engineering and provides educational programs. The IEEE, through its members, provides leadership from areas ranging from aerospace, computers and communications to biomedical technology, electric power and consumer electronics.

IEEE membership is open to professionals with varying levels of academic accomplishment and work experience. Student members have substantially discounted dues and fees. IEEE membership includes a subscription to the award winning monthly magazine, Spectrum, edited for the scientific and engineering professional. The members also receive The Institute, a monthly news supplement to spectrum. The members residing in Region 8 (Europe, Africa, Middle East) receive one more newsletter, it is the Region 8 News. The members may subscribe to the Proceedings of the IEEE, a monthly publication containing papers of broad significance in all areas of electrical science and technology.

IEEE members have the option of joining one or more of 35 specialized technical societies. As the IEEE serves the breadth of some common interests, the technical societies serve the depths of specialization. The IEEE and its 35 technical Societies publish more than 70 transactions, magazines and journals.

IEEE technical activities — through sponsoring conferences, conducting educational programs and local meetings, symposia and special events, publishing nearly 300000 pages of technical journals, books, conference records, standards, etc. — offer you a world of technical research, information and opportunities in order to keep the members' knowledge and expertise state-of-the-art.

The Hungary Section (HS) was founded in June 1987 by 51 members, as the smallest Section of the IEEE. The activity of HS is assisted by the Executive Committee containing eight professionals. The section has 5 committees: Nomination Awards, Programme, Conference, Membership Development, Education and Student Activities. To hold up the members interested in a narrow professional area the first Chapter within HS (MTT/AP/SS/COM Joint Chapter) was established in 1990. The second Chapter is recently formed. The Student Branch was formed by Technical University of Budapest in the end of 1989.

A Hungarian Newsletter is produced four times in a year. It contains a short summary of the latest Executive Committee Meeting and some important information completed by advertisements of IEEE conferences and meetings.

Many technical conferences have been held in Budapest by co-sponsorship of IEEE: European Microwave Conference (EuMC), International Conference on Harmonics in Power Systems (ICHPS'90), International Workshop on Celluar Neural Networks and their Applications (CNNA-90), Mid-European Custom Circuits Conference (CCC-91), International Conference on Information Theory, etc.

From this year, it is possible to convert membership dues for every HS members. This is a great advantage in Membership Development. The newsletter subscription fees can bee also paid in HUF by members hold scientific degree.

If you interested in more detailed information on IEEE and would like to be member of IEEE, please contact the IEEE HS Treasurer (Prof. János Bitó, Tungsram TH Co. Ltd., Váci út 77., H-1340 Budapest). ■

ESREF 92

3rd EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND ANALYSIS

Schwäbisch-Gmünd, Germany
5–8 October, 1992
Organized by
Informationstechnische Gesellschaft im VDE (ITG)

Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF) is the annual event on reliability physics of electronic components in Europe.

Symposium ESREF 92 will act as a forum for activities in Microelectronics Reliability in all phases of design, technological development, manufacturing, application, evaluation, control, specification and analysis. The technical programme is dedicated but not limited to the following main areas of interest:

- **building-in reliability by design and manufacturing and integration of associated disciplines,**
- **advanced techniques for reliability evaluation of technologies and products,**
- **failure analysis case studies and advanced analysis techniques,**
- **standardization of reliability evaluation.**

The deadline for submission of abstracts is the 2nd of March, 1992.

For further information please contact:

Prof. H. Berger
Programme Chairman
Technische Universität Berlin
Institut für Mikroelektronik
Jebenstrasse 1
D-1000 Berlin, GERMANY
Tel. +49 30 314 25880
Fax +49 30 314 24597

Dr. W. Gerling
Conference Chairman
SIEMENS AG
HL QS
Balanstrasse 73
D-8000 München 80, GERMANY
Tel. +49 89 414 42825
Fax +49 89 414 43828

SEMINAR AND EXHIBITION ON TELECOMMUNICATIONS IN THE ENERGY INDUSTRY (PRIVATE NETWORKS)

will be organized by the Scientific Society for Telecommunications at Balatonaliga, Hungary
7–9 October, 1992

This traditional bi-annual event is a platform for users of technology oriented private networks (petroleum and gas industry, electric energy industry, railway-, urban-, highway-, air- and shipping transport, etc.).

There is no limitation in the subjects of papers and the objects of exhibition, however special emphasis is placed on the following topics:

- realization of intelligent and complex networks
- computer aided supervisory systems
- cooperating and/or integration of private and public networks
- value added networks
- packet switching systems
- mobile systems for data and voice

Intention for presenting a paper should be announced by submitting a half-page abstract until 30 April, 1992, application for participation at the exhibition should be made not later than 31 August, 1992, at the Secretariat of the Scientific Society for Telecommunications.

Address:
Mrs. Katalin Mitók, Organizing Assistant
Budapest, Kossuth Lajos tér 6–8. Hungary H–1055
Phone: (361)153-1027; Fax: (361)153-1027

Any further information including participation and exhibition fees, accomodation and transport possibilities.

In the year 1992 we are continuing to publish the JOURNAL ON COMMUNICATIONS alternately in English and Hungarian, focusing each issue on a selected significant topic guest edited by an outstanding expert in the field. The planned issues are the followings:

RELIABILITY IN ELECTRONICS (in English)
Guest editor: Dr. Albert Balogh
ACOUSTICS (in Hungarian)
Guest editor: Dr. Tamás Tarnóczy
COMPUTER AIDED CIRCUIT DESIGN (in English)
Guest editor: Dr. Kálmán Tarnay
IC TESTING (in Hungarian)
Guest editor: Dr. Ferenc Kovács
SPEECH PROCESSING (in English)
Guest editor: Dr. Géza Gordos
TELECOMMUNICATION SERVICES (in Hungarian)
Guest editor: Dr. Gyula Sallai
MEDICAL ELECTRONICS (in English)
Guest editor: Dr. Attila Naszlady
TELECOMMUNICATION SOFTWARES (in Hungarian)
Guest editor: Dr. Gyula Csopaky
PERSONAL COMMUNICATIONS (in English)
Guest editor: Dr. István Frigyes
DIGITAL AUDIO BROADCASTING (in Hungarian)
Guest editor: Dr. Gábor Heckenast
OPTICAL COMMUNICATIONS (in English)
Guest editor: Dr. Tibor Berceli
DIGITAL SATELLITE COMMUNICATIONS (in Hungarian)
Guest editor: Dr. István Kása

The subscription rates will not be changed in 1992.

1 year 6 English issues	USD 60.
1 year 12 English and Hungarian issues	USD 90.

Subscription orders can be sent to the publisher TypoTEX Ltd. H-1015 Budapest, Batthyány u. 14., phone: (361)201-3317. Transfers should be made to the Hungarian Foreign Trade Bank, H-1821, Budapest, Account number: 203-21411.

Information for authors

JOURNAL ON COMMUNICATIONS is published monthly, alternately in English and Hungarian. In each issue a significant topic is covered by selected comprehensive papers.

Other contributions may be included in the following sections:

- INDIVIDUAL PAPERS for contributions outside the focus of the issue,
- PRODUCTS-SERVICES for papers on manufactured devices, equipments and software products,
- BUSINESS-RESEARCH-EDUCATION for contributions dealing with economic relations, research and development trends and engineering education,
- NEWS-EVENTS for reports on events related to electronics and communications,
- VIEWS-OPINIONS for comments expressed by readers of the journal.

Manuscripts should be submitted in two copies to the Editor in chief (see inside front cover). Papers should have a length of up to 30 double-spaced typewritten pages (counting each figure as one page). Each paper must include a 100 – 200 word abstract at the head of the manuscript. Papers should be accompanied by brief biographies and clear, glossy photographs of the authors.

Contributions for the PRODUCTS-SERVICES and BUSINESS-RESEARCH-EDUCATION sections should be limmited to 16 double-spaced typewritten pages.

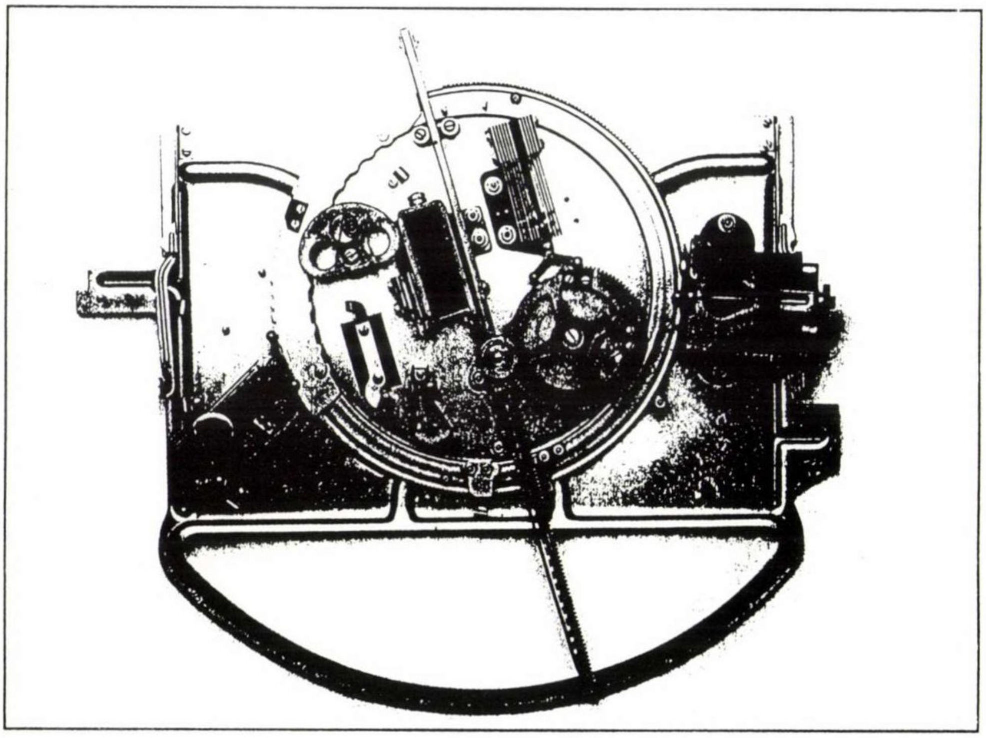
Original illustriations should be submitted along the manuscript. All line drawing should be prepared on a white background in black ink. Lettering on drawings should be large enough to be readily legible when the drawing is reduced to one- or two-column width. On figures capital lettering should be used. Photographs should be used sparingly. All photographs must be glossy prints. Figure captions should be typed on a separate sheet.

For contributions in the PRODUCTS-SERVICES section, a USD 110 page charge will be requested from the author's company or insitution.

Every machine and piece of equipment will become obsolete

some day preserved in a museum as a relic of the past.

This fifty-year-old crossbar selector went out of use long ago.



But everything has changed around us.

Our most modern technology is designed to assist effective

communication betwen people.

We are working on it.

HUNGARIAN TELECOMMUNICATIONS COMPANY