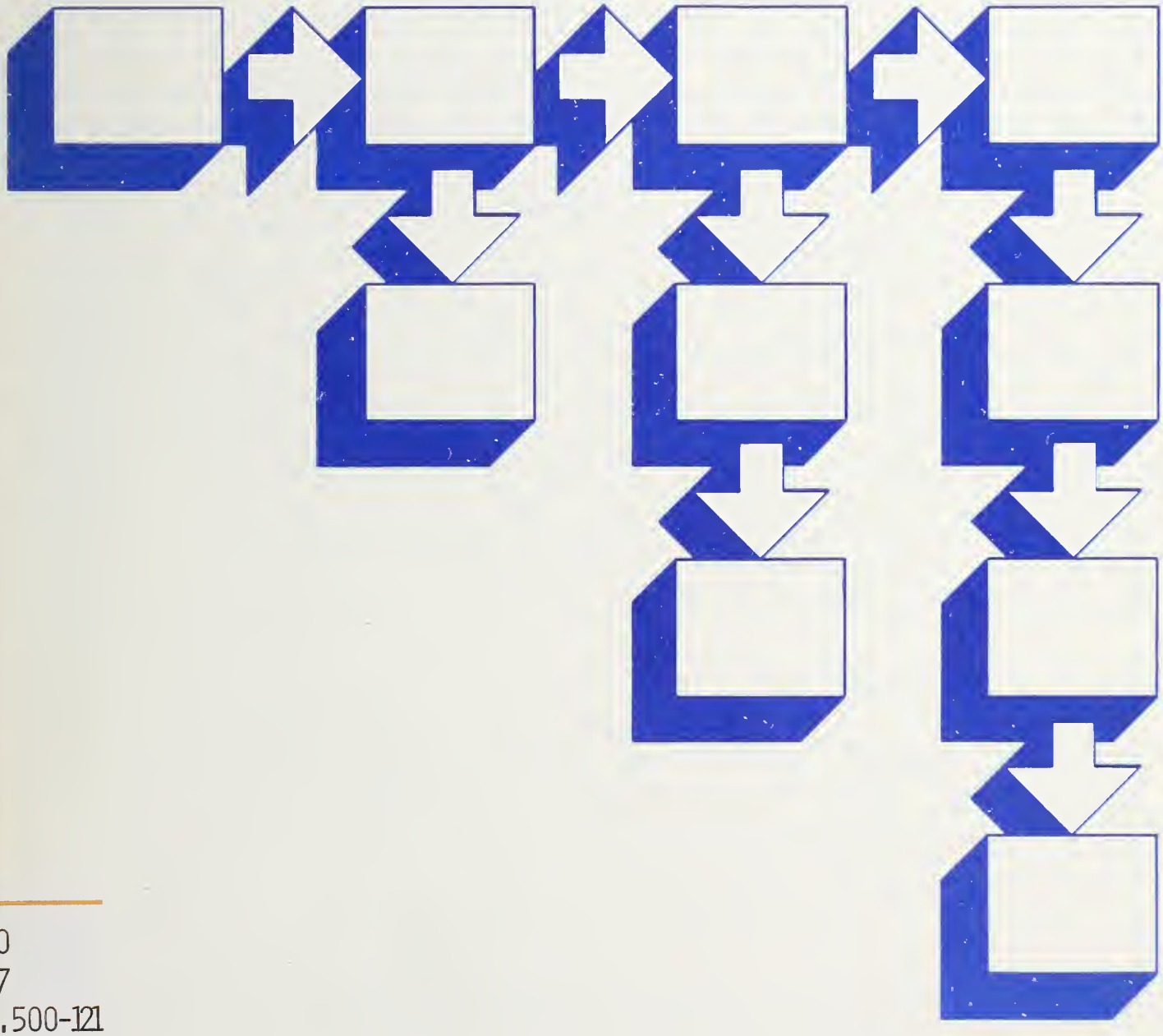
Computer Science and Technology

NBS Special Publication 500-121

Guidance on Planning and Implementing Computer System Reliability



The National Bureau of Standards[1] was established by an act of Congress on March 3, 1901. The Bureau's overall goal is to strengthen and advance the nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, the Institute for Computer Sciences and Technology, and the Center for Materials Science.

The National Measurement Laboratory

Provides the national system of physical and chemical measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; provides advisory and research services to other Government agencies; conducts physical and chemical research; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

- Basic Standards[2]
- Radiation Research
- Chemical Physics
- Analytical Chemistry

The National Engineering Laboratory

Provides technology and technical services to the public and private sectors to address national needs and to solve national problems; conducts research in engineering and applied science in support of these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the ultimate user. The Laboratory consists of the following centers:

- Applied Mathematics
- Electronics and Electrical Engineering[2]
- Manufacturing Engineering
- Building Technology
- Fire Research
- Chemical Engineering[2]

The Institute for Computer Sciences and Technology

Conducts research and provides scientific and technical services to aid Federal agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following centers:

- Programming Science and Technology
- Computer Systems Engineering

The Center for Materials Science

Conducts research and provides measurements, data, standards, reference materials, quantitative understanding and other technical information fundamental to the processing, structure, properties and performance of materials; addresses the scientific basis for new advanced materials technologies; plans research around cross-country scientific themes such as nondestructive evaluation and phase diagram development; oversees Bureau-wide technical programs in nuclear reactor radiation research and nondestructive evaluation; and broadly disseminates generic technical information resulting from its programs. The Center consists of the following Divisions:

- Inorganic Materials
- Fracture and Deformation[3]
- Polymers
- Metallurgy
- Reactor Radiation

[1]Headquarters and Laboratories at Gaithersburg, MD, unless otherwise noted; mailing address Gaithersburg, MD 20899.
[2]Some divisions within the center are located at Boulder, CO 80303.
[3]Located at Boulder, CO, with some elements at Gaithersburg, MD.

Computer Science and Technology

NBS Special Publication 500-121

Guidance on Planning and Implementing Computer System Reliability

Lynne S. Rosenthal

Institute for Computer Sciences and Technology
National Bureau of Standards
Gaithersburg, MD 20899

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Bureau of Standards
Ernest Ambler, Director

Issued January 1985

Reports on Computer Science and Technology

The National Bureau of Standards has a special responsibility within the Federal Government for computer science and technology activities. The programs of the NBS Institute for Computer Sciences and Technology are designed to provide ADP standards, guidelines, and technical advisory services to improve the effectiveness of computer utilization in the Federal sector, and to perform appropriate research and development efforts as foundation for such activities and programs. This publication series will report these NBS efforts to the Federal computer community as well as to interested specialists in the academic and private sectors. Those wishing to receive notices of publications in this series should complete and return the form at the end of this publication.

Library of Congress Catalog Card Number: 84-601159
National Bureau of Standards Special Publication 500-121
Natl. Bur. Stand. (U.S.), Spec. Publ. 500-121, 48 pages (Jan. 1985)
CODEN: XNBSAV

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1985

For sale by the Superintendent of Documents, U S Government Printing Office, Washington, DC 20402

Guidance on Planning and Implementing Computer System Reliability

Lynne S. Rosenthal

Computer systems have become an integral part of most organizations. The need to provide continuous, correct service is becoming more critical. However, decentralization of computing, inexperienced users, and larger more complex systems make for operational environments that make it difficult to provide continuous, correct service. This document is intended for the computer system manager (or user) responsible for the specification, measurement, evaluation, selection or management of a computer system.

This report addressess the concepts and concerns associated with computer system reliability. Its main purpose is to assist system managers in acquiring a basic understanding of computer system reliability and to suggest actions and procedures which can help them establish and maintain a reliability program. The report presents discussions on quantifying reliability and assessing the quality of the computer system. Design and implementation techniques that may be used to improve the reliability of the system are also discussed. Emphasis is placed on understanding the need for reliability and the elements and activities that are involved in implementing a reliability program.

Key words: computer system reliability; quantification of reliability; recovery strategy; reliability; reliability program; reliability requirements; reliability techniques.

ACKNOWLEDGEMENT

The Institute for Computer Sciences and Technology wishes to thank the representatives from the following agencies and departments for their contributions in the preparation of this report.

Johnson and Kennedy Space Centers, National Aeronautics and Space Administration
National Bureau of Standards, Department of Commerce
National Library of Medicine, Department of Health and Human Services
Rome Air Development Ceneter, Department of Defense
Social Security Administration, Department of Health and Human Services
Washington Computer Center, Department of Agriculture

TABLE OF CONTENTS

1.0 INTRODUCTION

Computer systems have become an integral part of most organizations. As the computer system and its services become more essential to the success of these organizations, the ability of the system to process information correctly and to provide continuous service becomes even more critical. However, recent trends such as decentralization of computing, inexperienced users, and larger, more complex systems have produced operational environments which thwart the attainment of these goals. Rising repair costs also make the reliability of the general purpose computer system an important issue.

Historically, computing has been dominated by the large general purpose mainframe. Associated with this type of computer system is a certain set of reliability questions and answers. Although this environment is changing with the advent of microcomputers, distributed data processing, and distributed data bases, many of the reliability concerns remain the same. Whatever the system configuration, reliability continues to be an important aspect of the computer system.

1.1 Purpose

This report is intended to assist users in acquiring a basic understanding of computer system reliability and to identify areas for further examination. It presents an overview of the fundamental concepts and concerns associated with system reliability, and identifies elements and activities involved in planning and implementing a reliability program. The report provides general guidance and as such does not present an in-depth methodology for creating or maintaining a reliable computer system. The underlying theme is that a knowledge of reliability is important in the development of new system specifications as well as in the continual assessment of existing computer system.

1.2 Scope

We are concerned here with the services and facilities of a general purpose computer system in a multiuser environment. In general, this report can be applied to other types of computer systems (i.e. minicomputers, microcomputers, distributed systems, etc.). The size of the system, its complexity, and mission within an organization will dictate the set of applicable guidance. The system planner (see definition, section 1.3) should analyze and evaluate this report with respect to the system and apply the appropriate subset.

The reliability of a computer system will be discussed in terms

of a system's three major components: hardware, software, and hyman/machine interface, the human component. An in-depth discussion of any one component is beyond the scope of this document. Sources of additional reliability information in these areas can be obtained from Appendix A: Bibliography and Related Readings. Included, are several publications prepared by the National Bureau of Standards', Institute for Computer Sciences and Technology in the areas of hardware and software reliability, verification and validation techniques, and risk assessment. References to these and other documents will help to ensure a complete understanding of reliability and its related areas.

1.3 Audience

The report is designed primarily for use by those who are responsible for the management, specification, measurement, evaluation, or selection, of a computer system. The information presented may also be relevant to individuals who are associated with the data processing center as system programmmers, analysts, technicians, and/or users. These employees may require knowledge of reliability issues to facilitate the management of their areas of responsibility and the performance of their assigned tasks. The term "system planner" will be used as a convenient title for any of the person(s) described above.

1.4 Document Overview

The document is divided into several sections and an appendix. It builds upon the concepts set forth in the early sections and concludes with a discussion of a system reliability program.

Section 2 defines reliability and related terms, and addresses the importance of reliability to the system.

Section 3 describes procedures for the quantification of the system reliability, in particular, sources of data, metrics, and the assessment of the metrics.

Section 4 presents a general discussion of the reliability techniques that can be designed into computer systems or can be implemented by the system planner.

Section 5 deals with the recovery of a computer system after it has failed or produced erroneous output.

Section 6 summarizes the important management tasks, activities, and costs involved in implementing a reliability program.

Appendix A contains a list of references and selected readings.

2.0 FUNDAMENTAL CONCEPTS

2.1 Terminology

The term system is used to denote a collection of interconnected components designed to perform a set of particular functions. In applying this definition, any component of the system may itself be regarded as a system; e.g. the central processing unit, memory, communications to and from the system, software programs, and computer system users [McDE80].

For purposes of this report, the failure of the computer system will refer to the termination, disruption, corruption, or incorrect outcome of system components (e.g. hardware, software).

The reliability of a computer system is defined as the probability that the system will be able to process work correctly and completely without its being aborted or corrupted. Note, a system does not have to fail (crash) for it to be unreliable. The computer configuration, the individual components comprising the system, and the system's usage are all contributing factors to the overall system reliability. As the reliability of these elements varies, so will the level of system reliability.

The availability of a system is a measure of the amount of time that the system is actually capable of accepting and performing a user's work. The terms reliability and availability are closely related and often used (although incorrectly) as synonyms. For example, a system which fails frequently, but is restarted quickly would have high availability, even though its reliability is low [McDE80]. To distinguish between the two, reliability can be thought of as the quality of service and availability as the quantity of service. Throughout this guideline, availability will be viewed as a component of reliability.

2.2 Reliability Distinctions

A computer system consists of a combination of hardware, software, and human components, each of which can cease functioning correctly, cause another component to malfunction, or help to increase the reliability of the other components. The reliability of this combination of components can be thought of as computer system reliability. Traditionally, the reliability of the computer system has concentrated on the hardware aspect of the system. This approach leads to the assumption that the software is 100 percent reliable. Since this is unlikely, it is necessary to include software in the system reliability calculations [ONEI83]. Finally, the need for human interaction with a computer system (to detect and correct problems,

restart the system, input key information, etc.) makes its inclusion in the determination of system reliability a necessity.

2.3 Reliability Requirements

The computer system (application) objectives and the environment in which it operates are major considerations in determining the level of reliability required of the system. An important question to be answered is: "How reliable must the system be?"

This section outlines several factors that contribute to answering this question. The discussion is in two parts: operational criteria - factors associated with the operational setting of the system and which are affected by the reliability of the system; and risk analysis - a method for balancing the degree of system reliability against acceptable levels of loss due to a less reliable system.

2.3.1 Operational Criteria -

Operational criteria are those characteristics of the system that make reliability more or less important. The identification of these factors and their relationship to reliability is necessary in evaluating the system reliability. At least the following factors should be evaluated.

1. Safety. Reliability is critical to a system where there is a potential for loss of life, health, destruction of an property, or damage to the environment.

   Examples: health care systems, scheduling of safety inspections, power system controls, air traffic control.

2. Security. Reliability is a fundamental element to the security of computer systems. A failure can decrease or destroy the security of the system. Undesirable events such as denial of information, unauthorized disclosure of information, or loss of money and resources can result from lack of reliability [FIPS31, FIPS73, HOPK80, PATR78, SHAW81].

   Examples: military command and control, eletronic funds transfer, management of classified information, inventory control.

3. Access. Reliability becomes a major concern to systems when it is unusually costly or impossible to access that system. Reliability techniques are used to minimize the potential failures that may render the system useless. These systems are usually very expensive with reliability costs a fraction of the overall system costs.

Examples: remotely operated/controlled systems (space shuttle, missiles, satellites)

4. Mode of operation. Reliability has varying levels of importance depending on the mode of operation. Failures affect real time, on-line, and batch applications differently. Real time applications are immediately affected by a failure. Similarily, a system which fails while supporting an on-line application will demonstrate a deviation from expected conditions sooner than the same system operating in exclusively batch mode.

   Examples: data management systems, centralized information systems, air traffic control, computer service bureau

5. Organizational dependency. The importance of reliability increases as the organization's dependence on the computer system and its services becomes more critical to the success of that organization. A failure can directly affect the organization by creating delays or disruptions to production schedules, administrative activities, management decision making, etc..

   Examples: all systems

2.3.2 Risk Analysis -

A balance between the application reliability requirements and the cost of designing and implementing a system needs to be evaluated. The system planner should be aware that for some applications a failure and its recovery may cost less than achieving an increase in the system reliability (prevention of a failure). A risk analysis approach should be used to determine the affect of a failure and its recovery, and the level of reliability sufficient for that system. The three key elements to be considered in such an analysis are:

- The amount of damage which can result from a failure,
- The likelihood of such an event occurring,
- The cost-effectiveness of existing or potential safeguards.

Previous guidelines (NBS FIPS PUBS 31 and 65,) deal extensively with risk analysis for automatic data processing systems. Although, these guidelines pertain to computer security, the risk analysis techniques and procedures can be applied in the case of reliability as well.

3.0 EVALUATING RELIABILITY

In order to assure that the computer system meets or exceeds performance requirements, the system planner must be able to assess or specify the reliability of the computer system. This section describes system reliability data gathering, analysis, and assessment results. The data (section 3.1) obtained about the system are used as input to the reliability metrics (section 3.2), which in turn, are used to derive policies and performance criteria about the system's reliability (section 3.3).

3.1 Sources Of Reliability Data

Reliability information can be obtained from a number of sources. The data can be derived from job accounting, system performance, error routines that are part of the operating system, diagnostic routines, operator logs, hardware and software monitors, and system users. A host of computer performance evaluation tools and capacity planning tools can also be used to acquire data about the system. Information about specific performance tools can be obtained from CPEUG, the Computer Performance Evaluation Users Group*; CMG, the Computer Measurement Group*; and publications such as "Management Control of EDP Performance" and "EDP Performance Review", both published by Applied Computer Research. Information about capacity planning tools can be found in reference [KELL83].

Whatever the source of reliability data, it is important to keep accurate, timely, and complete records. These records form the basis for assessing the reliability of the system. Typical data elements that should be recorded are:

- the date and time of any event evincing a reliability problem,
- type of event,
- amount of time lost (if any),
- the system and responsible component,
- average service and response time for a job,
- number of jobs and job mix at a given time, and
- the system resources used for these jobs.

* For information write: CPEUG, at B266 Technology, National Bureau of Standards, Gaithersburg MD 20899; CMG, 11242 North 19th Avenue, Phoenix, AZ 85029

This list is not meant to be all-inclusive nor does every record need to contain each of the above elements. (Examples of these data elements and their derivations can be found in the following sections).

A continuous record of system performance and activity provides the system planner with historical data for evaluating system reliability. This information will enable the planner to base future acquisition, current operation procedures, and maintenance decisions on past system reliability and performance.

The responsibility for recording and reporting the system reliability information should be clearly delineated. The recording and reporting procedure should be reviewed periodically for duplication and/or missing elements. It is suggested that records be maintained for at least six months. Actual time frames for maintaining these records should be determined by the system planner based on the system's reliability and performance, as well as on the usage of the records within the organization (e.g. to take contractual action against a vendor).

3.1.1 Accounting Logs -

Accounting logs provide performance information along with billing information. Accounting logs usually contain data about individual programs as well as system usage [BOUH79]. The type of data and depth of detail can vary among computer systems. A few examples of possible data elements are listed below:

- o program data: initiation and termination time, total service time for each used resource (e.g. CPU, disk), memory used, I/O counts, and user identification,
- o system data: system configuration, software parameters, checkpoint records, and device errors.

Analysis programs use accounting logs to produce system performance reports. These reports enable the system planner to recognize deviations from normal system usage and to evaluate the impact of a failure by observing and contrasting the system performance prior to, and subsequent to, a failure. Figure 1 is an example of one type of report that can be generated.

Device Type: Magnetic Tape
Unit-Serial Model Vendor Date
Tape 05-189 3420 Jul 10

NUMBER AND TYPE OF FAILURES

	This month to date			Previous Performance					
				Prior 5 Days				Prior month	
	Today	high	total	-1	-2	-3	-4	high	total
Hard Fails	3	3	13	0	1	0	0	3	18
Soft Fails	37	29	203	6	15	0	13	819	3505

- DAILY THRESHOLD LEVELS EXCEEDED: Hard = 0 Soft = 0

- DAILY FAILURE LOG:

Unit-serial	Jobname	Volser	MM/DD	HH.MM-HH.MM	Cpu	Failure	Record#	Density
Tape -05-1189	LMSPUOTO	000000	07/10	12.55-12.55	189	Eqpt-rd	003	6250
	00000000	MITSTP	07/10	11.42-11.42	189	Eqir-27	7098	1600
	Landumpe	010758	07/10	00.13-00.13	189	Data-wr	3914	6250

NOTE: - Used to identify devices exceeding threshold values.
Includes device hard failure log for total picture

(a) DAILY DEVICE FAILURE REPORT

	Device type	Model	FAILURE TYPE #Hard fails	FAILURE TYPE #Soft fails	USAGE DATA Total	RATIOS Use/ Hardfail	RATIOS Use/ Softfail
this month	Disk storage	332	4	147	6400	1600.0	40.0
prior month			13	273	6200	407.6	22.8
this mo.	Disk storage	335	2	88	4104	2052.0	46.7
prior mo.			4	321	3625	906.2	11.3
this mo.	Magnetic tape	189	18	3505	4526	251.4	1.2
prior mo.			13	1597	5616	432.0	3.5

(b) MONTHLY DEVICE SUMMARY

Figure 1: System performance reports — Examples of one type of accounting information analysis

3.1.2 System Incident Reports (SIR) -

System incident reports are generated by the operations staff whenever a problem occurs. The SIR calls for full information including time of day, system status, tasks and jobs in the system, possible cause (relating it to the hardware, software or unknown), diagnostic messages, availability of core, etc. (figure 2). The final disposition of the incident and routing information (if any) is also provided. The completed SIR and any supporting documentation is then circulated to appropriate technical personnel[FIPS31].

3.1.3 Console Operator Logs -

Console operator logs are an operator maintained account of the system's daily activity (figure 3). Typical information recorded in these logs includes: operator actions (e.g. boots, mounts, backups), system configuration, outages, crashes, downtime, malfunction of peripherals, and routine and corrective maintenance repairs. The system planner, operator, and maintenance personnel can use the information contained in the operator logs to analyze daily activity, identify problem areas, track reliability control procedures, and evaluate reliability metrics. Summary reports, such as weekly log reports (figure 4), efficiency reports (figure 5), utilization statistics (figure 6), and failure categorization reports (figure 7) can be derived from the logs and used to evaluate the system.

3.1.4 System Error Messages -

System error messages are automatically generated by the system and often provide clues to the source of an error. Relevant information pertaining to the error(s) is recorded. Such information may include: time of day, error type, control limits exceeded (exception reports), consistency checks, timers, and selected traces and dumps. Many systems automatically log the error messages and related data. Analysis programs, available from system vendors or other commercial sources are used to extract the relevant reliability information.

3.1.5 Diagnostic Routines -

Diagnostic routines provide information on the integrity of the system by identifying failures or indicating (by the absence of failures) that the system is operating correctly. The routines can be run periodically or subsequent to the occurrence of a problem. The routines can provide information about the problem type and location.

SYSTEM INCIDENT REPORT

Date _____ Down Time _____ CPU _____

Unit _____

Reference Mar _____ System Avail _____ System Return _____ Subassembly _____

Module _____

Description of Problem _____

_____

_____

_____

_____

_____

_____

_____

Corrective Action _____

_____

_____

_____

_____

_____

_____

_____

Diagnostic Routine _____

Diag. Fail: Yes No

Service Person _____

Figure 2: System Incident Report

```
%OOPCOM, 29-Jun-1984 12:28:28.72, message from user NETAAP
NET shutting down
%OPCOM, 29-Jun-1984 12:33:54.07, operator disabled
%OPCOM, 29-Jun-1984 12:35:15.47, operator enabled
%OPCOM, 29-Jun-1984 12:45:05.22, operator status
PRINTER, TAPES, DISKS, DEVices
%OPCOM, 29-JUN-1984 12:48:53.21, request from user PUBLIC
Please mount volume KLAT in device MTA0:
%OPCOM, 29-JUN-1984 12:50:02.11, request satisfied
%OPCOM, 29-JUN-1984 12:50:03.54, message from user SYSTEM
Volume KLAT mounted, on physical device MTA0:
%OPCOM, 29-Jun-1984  13:01:26.91, device LP0 is offline
%OPCOM, 29-Jun-1984  13:31:15.63, request from user PUBLIC
Mount new relative volume 2 () on MTA:
%OPCOM, 29-Jun-1984 13:33:45.05, message from user SYSTEM
MTA: in use, try later, mount aborted
%OPCOM, 29-Jun-1984 13:46:21.67, message from user SYSTEM
Current system parameters modified by process ID 001f003C
%OPCOM, 29-Jun-1984 13:46:21:97, device DSK4 is offline
Problems with DSK04
Problems with DSK04
%OPCOM, 29-Jun-1984 13:47:01:30, message from user SYSTEM
DSK04 hass been remounted - back online
```

Figure 3: Sample operator log

SYSTEM LOG REPORT FOR THE WEEK ENDING July 9

1. System Utilization for the week

Time sharing with operator coverage	128:18
Time sharing without operator coverage	16:35
Regular field service PM's	13:55
Extra field service	4:05
Computer operation - stand alone	3:09
Lost time	1:58
TOTAL HOURS	168:00

2. Equipment

Hardware problems contributing to system downtime:
MF10 - down, memory parity errors
DF110-TM10 - problems occurred when using MTA drive

Hardware problems not causing system downtime:
LPA1 - replaced hammer module col. 35
TU56 - tightened hub

3. Reruns and Lost time

-246-
Estimated lost time on system 246 was 20 hours and 25 min.

-541-
Estimated lost time on system 541 waass 7 hours and 45 min.

4. Monitor problems

Name	Date	Problem	Detail
RF6B9N	5Jul	Hung	One job running, most other jobs swapped in Run state
RF6B9N	5Jul	Loop	On PI 4 interrupt chain
RF6B9N	7Jul	Loop	Dubious crash data
RF6B9N	8Jul	Hung	Most jobs waiting for disk monitor buffer or disk I/O wait

5. Job Distribution (average number of jobs)

	0700-0900	0900-1300	11300-1700	1700-1900	1900-2400
7/4	22.70	48.05	50.20	40.00	32.00
7/5	30.90	53.89	54.95	40.70	34.43
7/6	27.15	47.50	55.50	39.40	34.50
7/7	30.15	44.15	49.90	32.50	19.14
7/8	31.80	49.89	46.12	22.90	23.20

Figure 4: Weekly log report

WEEKLY EFFICIENCY REPORT (July)

Date	System turned off	Performance during scheduled hours				Scheduled time (hours)	Actual system up time (hours)	Efficiency/wk (% uptime)
		down due to software	down due to hardware	other	total num. of hours down			
1-3	7	1:51	1:03	:40	3:34	43:30	39:56	92%
5-10	7	1:48	:43	--	2:31	95:30	92:59	97.3%
12-17	8	:48	8:53	:57	10:38	95:30	84:52	88.8%
19-24	3	:27	6:25	--	6:52	95:30	88:38	92.8%
26-31	6	:30	:29	2:52	3:51	98:30	94:39	94.6%
total	31	5:24	17:33	4:29	27:26	428:30	401:04	94%

system operational 6 days/week, 24 hours/day

Figure 5: Efficiency report

SYSTEM UTILIZATION STATISTICS (July)

Reloading System for Processing	Hours	Pecent of Total
Down - System not operational due to hardware or software failure	22:57	4.2%
Site - Down due to electrical, air conditioning, water damage, etc.	3:31	.6%
P.M. - Scheduled preventive maintenance	20:00	3.6%
Unscheduled maintenance	0:00	0%
Off - System shut off	31:00	5.6%
Idle - Work to be processed, but no one available to process it	30:00	5.4%
Development - System up, but not for "public use"	42:34	7.7%
Public use - System operational for all users	401:04	72.8%

Figure 6: Utilization statistics

NUMBER OF FAILURES BY CAUSE (July)

	Total Number	% of Total
Hardware	212	44
Software	106	22
Application	10	2
Operations	29	6
Environmental	77	16
Unknown	10	2
Reconfiguration	39	8
	483	100

Figure 7: Failure categorization

3.1.6 Hardware And Software Monitors -

Hardware monitors are electronic devices that are physically attached to the computer system and software monitors are software programs residing in and utilizing some or all of the host's resources. Both types of monitors make measurements on the system by recording, analyzing and/or presenting data under real time operation. The performance level of system resources as well as any problems that might occur can be pinpointed and tracked with respect to their cause and effect within the system. For example, in the data communications area, measurements of response time and communication line utilization can be used to identify and locate potential problem areas [JACO81].

3.1.7 User's Level Of Satisfaction -

User's level of satisfaction with the system's performance can provide an indication of the system reliability. User complaints and questions can aid the system planner (analyst, operator, etc.) in the identification of problem areas. Interaction with users may be a formal or informal procedure. Interaction may include: joint system staff/user meetings, surveys of the user community (e.g. ask about possible problem areas), or user requests for refund of purchased computer services (an indication of possible system problem areas).

3.1.8 System Performance Meetings -

System performance meetings provide the opportunity for appropriate personnel (system managers, operators, analysts, technicians) to meet, discuss, and analyze the system performance. The reports and information obtained from the above sources, as well as any additional data, form the basis of the system reviews. The members of the meeting try to identify the system components which fail most frequently, the cause of the failures, and solutions to minimize or eliminate future occurrences of such events.

3.2 Reliability Metrics

Reliability metrics provide a quantitative basis for the assessment of the computer system reliability. The actual measurement is accomplished by applying data gathered about the system as input to the reliability metrics. The data can be obtained from the system planner's in-depth knowledge of the system's capacity and activity and formal inspection of the system components, as well as the sources cited in section 3.1.

Numerous quantitative methods exist to measure the reliability of the computer system. Most metrics for system reliability are derived from a combination of hardware and software measurements [BESH83, CAST81, HERN83, SRIV83, THOM83]. Due to the complexity of computer

systems, a variety of reliability metrics should be chosen to describe the system adequately. The development and/or identification of appropriate metrics is not an easy task. Often it is necessary for a reliability expert to identify a set of metrics and/or to develop mathematical models (algorithms) to describe the system in terms of probabilities.

A quantitative value or threshold level consisting of either a number, range, or percent should be established for each measurement. These values/levels can be established in accordance with:

- Department of Defense standards [MIL217, MIL757, MIL781]
- comparison to similar systems
- system specifications by vendors and/or GSA schedule
- specific application requirements

Comparisons of these pre-stated values with the actual derived measurement values will be helpful in assessing the reliability of the computer system. Note, it is not always possible to establish a mathematical value for all measurements. In these cases, it is advisable to develop a relative importance rating (priority factor) to indicate its value [PERS83].

The remainder of this section presents an overview of system reliability metrics. The discussion will be divided into three categories: hardware, software, and human measurements. The objective is to identify the basic concepts and underlying attributes associated with the metrics of the various categories. Detailed analysis of specific metrics are beyond the scope of this document, but additional references are given for each category.

3.2.1 Hardware Measurements -

There has been an abundance of information written on hardware reliability metrics. It is these metrics that are the most familiar and are thought of as the 'traditional' measurements. The metrics are used to assess the mechanical or electrical elements of the computer system and have been used as the original tools for the evaluation of total computer system reliability.

The hardware metrics are a means of evaluating the amount of processing time lost due to the failure of the computer system or a specific component. The calculation of the reliability measurement will vary with the complexity of the system configuration. Although the basic concepts will remain the same, the hardware reliability measurements for a single, non-redundant, non-repairable system will differ from that of redundant, repairable, and/or distributed system configurations. Metrics for the latter must compensate for the special properties (e.g. replication of components) of the system.

There are two approaches to estimating hardware reliability; one is based on statistical probability distribution models, and the other is based on actual system performance. The probability model is the analytical basis for making reliability predictions. The determination of an appropriate model is necessary to achieve realistic predictions, and should be developed by an expert. Prediction tables [MIL217, MIL757, MIL781] and other literature sources [BEAU79, LAWL82, SIEW82] can provide background and general reliability models.

Quantitative metrics based on an operational system can provide information on the processing time lost due to the failure of the computer system or a specific component. Among the measurements of interest are: the number of times the hardware ceases to function in a given time period (Failure rate), the average length of time the hardware is functional (mean time between failures, MTBF), the amount of time it takes to resume normal operation (mean time to repair, MTTR), and the quantity of service (availability). Although a simplistic model, figure 8 depicts some of these measurements.

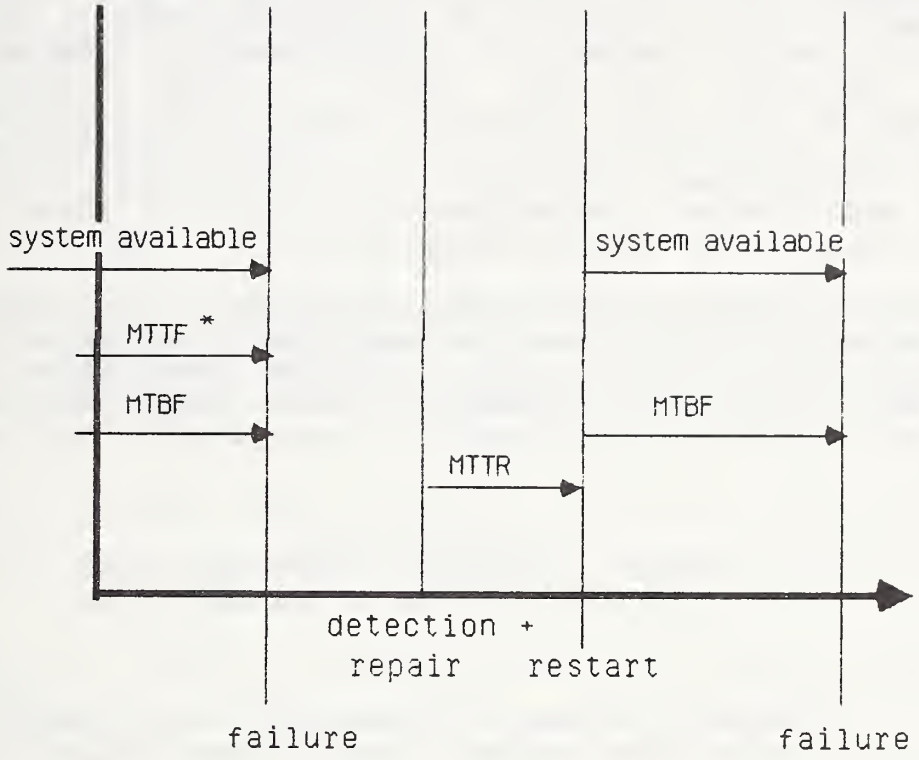
Other measurement algorithms and analysis techniques might include calculations to determine:

- a level of confidence in the system's ability to survive a failure,
- the number of intructions that could be processed before a failure,
- the amount of time the system will be inoperable,
- the response time delay.

Comprehensive descriptions of hardware metrics and analysis techniques can be found in [BEAU79], [LAWL82], [ODA81], and [SIEW82]

3.2.2 Software Measurements -

There is a tendency to use hardware metrics to evaluate the software component of the computer system. Although use of these metrics may be appropriate in a few cases, it can limit the scope of the software evaluation. This limitation is due to differences between hardware and software failure origination and repeatability. For example, hardware failures are either transient or repeatable, and result from either design, development, and component fault, whereas software failures are almost always repeatable and originate in design and development [ONEI83].



$$MTTF^{*} = \text{the number of time units the system is operable before the first failure occurs}$$

$$MTBF = \frac{\text{sum of the number of time units the system is operable}}{\text{number of failures during the time period}}$$

$$MTTR = \frac{\text{sum of the number of time units required to perform system repair}}{\text{number of repairs during the time period}}$$

* MTTF applies to non-repairable systems and is not applicable after the first failure. (Some experts consider MTTF to be a special case of the MTBF measure)

Figure 8: Measures of MTTF, MTBF, MTTR and availablity

The time line illustrates the various measurements with respect to the recognition and repair of failure occurrences

Software reliability calculations can be performed throughout the system life cycle to quantify the expected or the actual reliability. In the early phases of development, the measurements can be applied to the documentation on the system concepts and design; and in later phases, to both the documentation and the source code. The metrics should measure errors caused by deficiencies or the inclusion of extraneous functions in the system design specification, documentation, or source code. The metrics should be limited to errors caused by software deviating from its specifications while the hardware is functioning correctly. Any error that occurs several times before detection and correction, should be charged as a single error in the reliability calculations.

Software measurements are used to predict or quantify the software quality of the system. The measurement can be calculated by either the evaluation of past success or the prediction of future failure. One method of calculating software reliability might be to count the number of errors that occur in the source code [PRES83], e.g.

$$1 - \frac{\text{Number Of Errors}}{\text{Total Number of Lines of Executable Code}}$$

(Rating is in terms of the expected or actual number of errors that occur in a specified time interval).

Other measurement algorithms or analysis techniques might include calculations to:

- o define the levels of error occurrence and tolerance,
- o determine the percentage of errors during a time interval,
- o identify error types and the modules in which they occur,
- o identify deficiencies in the documentation or code.

Comprehensive descriptions of software metrics and descriptions can be found in [ONEI83], [PRES83],

3.2.3 Human Measurements -

Human reliability and its influence on the computer system is a developing discipline. Although new models are being developed, there has been a shortage of human reliability metrics, a general lack of understanding of the analysis required, and an absence of pertinent reliablity data [LASA83].

Human reliability metrics differ from those for hardware or

software. Differences stem from the ability of the human to make decisions, to learn from one's experiences, and to continue functioning in spite of a mistake ('failure') [SRIV83]. Thus, the metrics need to model a human's ability to work under different situations. A recognized approach [SRIV83] is to divide the metrics into two fundamental components:

- o the 'performance' element - a task is completed, with no decision required, and
- o the 'control' element - a task consists of several parameters and requires a decision to be made.

The failures that should be assessed by the metrics include: incorrect diagnosis, misinterpretation of instructions, inadequate support or environmental conditions, and insufficient attention or caution. Algorithms and analysis techniques might include calculations to:

- o determine improper human (operator, user, etc.) performance,
- o determine the amount of downtime caused by human/machine interaction,
- o identify the number of errors that were manually detected and corrected,
- o identify the errors caused by human alteration to the system,
- o evaluate human/machine interaction - amount required, cost to implement, and time to accomplish.

Comprehensive descriptions of human reliability metrics can be found in [LASA83] and [SRIV83].

3.3 Assessing The Quality Of The Computer System

The analyses of the information obtained from the reliability metrics (in the previous section) can be used individually, in combination, or in conjuction with historical system data to evaluate, estimate, or predict the reliability of the computer system. Specifically, the system planner can utilize this derived information to:

- o establish performance and acceptance criteria,
- o formulate policies to reflect or achieve reliability goals,
- o gather information on the effect of a failure on the system and organization [CAST81], and

- clarify and identify specific failures or problems.

Examples of assessments that can be made about the system are given in the following paragraphs. A chart listing these examples and the measurement class that might be associated with them is given by figure 9.

EXAMPLES OF ASSSESSMENTS	1	2	3	4	5	6	7
	INDEX TO CLASSES OF MEASUREMENTS						
Performance/Acceptance Criteria							
Threshold Value Assessments	X	X	X	X	X	X	X
General System Stability		X	X				X
General System Availability		X	X	X			X
Policy Formulation							
Maintenance Policies	X	X	X		X	X	X
Acquisition Policies	X		X		X	X	
Information on the Impact of a Failure							
Productivity and Workload Scheduling	X	X	X	X			
Amount of 'Backup' Work	X	X	X	X			
Clarification/Identification of Failures	X			X	X	X	

Classes of Measurements: he following are possible classes of measurements that can be used in the evaluation of the Assessment Examples.

1 Measurements that indicate QUALITY of Service
2 Measurements that indicate FREQUENCY of Failure
3 Measurements that indicate LENGTH OF TIME the System is Inaccessible
4 Measurements that indicate System ACTIVITY
5 Measurements that indicate The IDENTITY of the Failure
6 Measurements that indicate The LOCATION of the Failure
7 Measurements that indicate the RESULTING ACTION following a Failure

Figure 9: Assessment Examples and Classes of Measurements

3.3.1 Performance/Acceptance Criteria -

The following criteria can be used in the specification and evaluation of performance levels in contracts with vendors and/or in-house system policies.

Threshold Value Assessment is the comparison of pre-established metric values with the actual derived value of the metric. The technique is used to indicate if the measurement exceeds, meets, or falls short of expected levels of performance. It is a method that can be applied to all measurement types and provides a means of specifying the minimum performance level that is to be achieved.

General System Availability is the amount of time the overall computer system is operational and usable. The achievement of a predetermined availability threshold can be used to indicate acceptable, substandard, or unacceptable performance levels. A chart should be developed to indicate the limits for acceptable, substandard, and unacceptable performance of the system with values based on availability requirements. For existing systems, the derived metric values should be compared against the required levels listed in the chart. The chart below is a hardware oriented example of system availability performance limits (using hours of downtime in a computer system).

HOURS OF SYSTEM DOWNTIME

Subsystem causing downtime	Acceptable	Substandard	Unacceptable
CPU + Memory	0-16.9	17.0-33.9	>34.0
Disk Storage	0-16.9	17.0-33.9	>34.0
Magnetic Tape	0-8.4	8.5-16.9	>17.0
Printer	0-8.4	8.5-16.9	>17.0

Note: The ranges listed are for illustration purposes and are not meant to be recommended values for any particular system).

General System Stability is the average amount of time the system is operational before user services are interrupted, loss of work results, or a system reboot is required. The determination of system stability can be derived from the number of system interruptions (e.g. measurements that indicate the number and length of time the system is unavailable for use). A malfunction or failure that does not result in system interruption is ignored for system stability determination. A chart should be developed to indicate the limits of performance of the system with values based on stability requirements. For existing systems, the derived metric values should be compared against the required levels listed in the chart. For example, the chart below illustrates acceptable, substandard, and uanacceptable performance

levels for several subsystems during a 30 day period.

NUMBER OF SYSTEM INTERRUPTIONS

Subsystem causing downtime	Acceptable	Substandard	Unacceptable
CPU + Memory	0-9	10-19	>20
Disk Storage	0-9	10-19	>20
Magnetic Tape	0-4	5-9	>10
Software Module 1	0-12	13-24	>25
Software Module 2	0-12	13-24	>25

(Note: The ranges listed are for illustration purposes and are not meant to be recommended values for any particular system).

General Survivability is the probability that the system will continue to perform after a portion of the system becomes inoperable. A numerical value or importance level should be established and used to indicate the acceptable and/or required levels of survivability. Survivability can be derived from measurements that relate to the number of failures (both hard and soft failures) and the ability of the system to recover from the failure. In addition, measurements that indicate the system usage and the amount of damage that could result from a failure can influence the survivability rating and should also be considered. The following list is an example of levels of importance associated with various types (hardware, software, etc.) of subsystems. (Note: documentation survivability encompasses the scope, clarity, completeness, and correctness of the documentation that will enable a user to read, understand, and perform the activity described correctly).

IMPORTANCE LEVELS

Subsystem	Level of Importance	Comments
CPU	high	Level depends on the functional importance and usage of the device
Tape Drive 1	low	
Software Module 1	high	Level depends on the functional importance and usage of the module
Software Module 2	moderate	
Documentation	moderate	Level depends on the subject importance and the usage of the documentation

(Note: The levels listed are for illustration purposes and are not meant to be recommended values for any particular system).

3.3.2 Policy Formulation -

The values obtained from the reliability measurements are used in the formulation and adjustment of reliability policies.

Maintenance Policies and Procedures should be examined and evaluated to reflect the reliability requirements of the computer system. The system planner can use the metrics to assess the effectiveness of the current maintenance policies and to adjust them accordingly. Almost all the reliability measurements can be used to indicate system problems and are helpful tools in the identification of potential and actual subsystem failures. In addition, the logistic delay, delays encountered while waiting for parts and/or service personnel should be included in the considerations.

Acquisition Policies should be examined and evaluated to reflect the reliability requirements of the computer system. The reliability measurements are indications of the system activity and quality, and can be used as supporting factors in the justification and specification of new system acquisitions and/or system reconfiguration.

3.3.3 Information On The Impact Of A Failure -

The more dependent an organization is on the computer system, the greater the impact a failure would have on that organization. Reliability measurements that provide information about the frequency and identity of system failures and the performance level of the computer system are used in the assessment of the impact.

Productivity and Workload Scheduling is the scheduling of the amount of work that consumes computer resources. A system not functioning to its full capacity may delay or prevent the processing of user and system jobs. This interruption can affect productivity and product schedules and as such, translates into a cost. With knowledge of the computer systems reliability, a system planner can adjust and forecast current and future workload requirements accordingly.

Amount of 'Backup' Work is the amount of work performed in anticipation of a failure. This would include multiple copies of system and user programs and data, checkpoints for easy restart, and multiple runs of identical jobs. Efforts such as these are used to circumvent the effects of a failure or to facilitate recovery. In general, the less reliable a computer system is, the greater the amount of 'backup' work performed.

3.3.4 Clarification And Identification Of Failures -

The combined analysis of reliability measurement results and accumulated historical system data is a means of identifying the occurrence of specific failures/problems or of obtaining early warning indicators of potential failures/problems. This knowledge enables the system planner to take the appropriate corrective action in a timely manner. Of particular value in pinpointing the cause of the failure/problem is the correlation of measurements that pertain to the type, location, and frequency of the failure/problem with the system's resultant action (e.g. crash, recovery - retry or warm start).

4.0 BASIC TECHNIQUES

Reliability techniques are incorporated into a computer system to reduce the errors and effects resulting from the corruption of data or malfunction of the hardware during system operation. The techniques are implemented to prevent, offset, or correct the occurrence of one of the following fundamental categories of faults.

1. Physical faults. The disruption of the information processing function due to a hardware malfunction of the computer and/or its peripherals [AVIZ79]. These failures occur due to the weakening and breakage of the components over a period of time and usage.

2. Design faults. The imperfections in the system due to mistakes and deficiencies during the initiation, planning, development, programming, or maintenance of the computer system [AVIZ79].

3. Interaction faults. The malfunctions or alterations of programs and data caused by human/machine interactions during system operations.

The remainder of this section presents a general discussion of basic reliability techniques. The selection of techniques that are applicable for a given system will depend on the system objectives and configuration, and the feasibility of implementing the technique. The discussion is divided into two parts: design features and implementation techniques. Design features are the reliability techniques designed into the hardware configuration or software source code by the system developer. Implementation techniques are those the system planner can adopt to improve the reliability of the system.

4.1 Design Features

A large range of reliability techniques is available to the designers of computer systems. The goal of these techniques is to keep the system operational either by eliminating faults or in spite of the presence of faults. A combination of reliability-enhancing features may be used within a single system. The specific techniques used may vary among systems due to cost, performance, and reliability trade-offs.

Typically, the system planner does not designate which design techniques are to be incorporated into the computer system. (The development of custom designed system software may be an exception to this rule). Despite this inability, the system planner should be familiar with reliability design techniques in order to better specify and understand the reliability capabilities of the system. A list of techniques is shown in Figure 10. A brief explanation of several of

FUNCTION	TECHNIQUE
Fault Avoidance	Environment modification
	Quality components
	Component integration level
	Verification and validation
Fault Tolerance	
Fault detection:	Duplication
	Error detection codes
	M-of-N codes
	Parity
	Checksums
	Arithmetic codes
	Cyclic codes
	Self-checking and fail-safe logic
	Watch-dog timers and timeouts
	Consistency and capability checks
Masking redundancy:	NMR/voting
	Error correcting codes
	Hamming SEC/DED
	Masking logic
	Interwoven logic
	Coded-state machines
	Recovery Block
	N-version Programming
Dynamic redundancy:	Reconfigurable duplication
	Reconfigurable NMR
	Backup sparing
	Reconfiguration
	Recovery

Figure 10: Classification of reliability techniques [SIEW82]

these follows. More thorough discussion can be found in [CART79], [DENN76], [McDE80], and [SIEW82].

4.1.1 Fault Avoidance -

The goal of a fault avoidance approach is to reduce or eliminate the possibility of a fault through design practices such as component burn-in, testing and validation of hardware and software, and careful signal path routing. The approach assumes an a priori perfectibility of the system. To achieve fault avoidance, all components of the system (hardware and software) must function correctly at all times.

Fault Avoidance Techniques:

- Environmental factors. The elimination of faults caused by heat produced by the system's circuitry.
- Quality components. The acquisition and use of extremely reliable components.
- Component and system integration. The careful assembly and interconnection, and extensive testing and verification of individual modules, subsystems, and the entire system.
- Verification, validation, and testing. The process of review, analysis, and testing employed throughout the software development life cycle to insure the correctness, completeness, and consistency of the final product [BRAN81, POWE82].

4.1.2 Fault Tolerance -

The goal of a fault tolerance approach is to preserve the continued correct execution of functions after the occurrence of a selected set of faults. This is achieved through redundancy: the addition of hardware, software or repetition of operations beyond those minimally required for normal system operation.

Fault Tolerance Techniques:

- Watchdog timers and timeouts. A process must reset a timer or complete processing within a set time period. Inability to accomplish this task is an indication of possible failure. Neither timers or timeouts can be used to check data for errors.
- N Module Redundancy (NMR)/voting. The outputs of N identical modules are compared. By the use of majority voting, a fault can be detected, the correct output selected, and processing continues. The most common NMR technique is Triple Modular Redundancy (TMR) (figure 11).

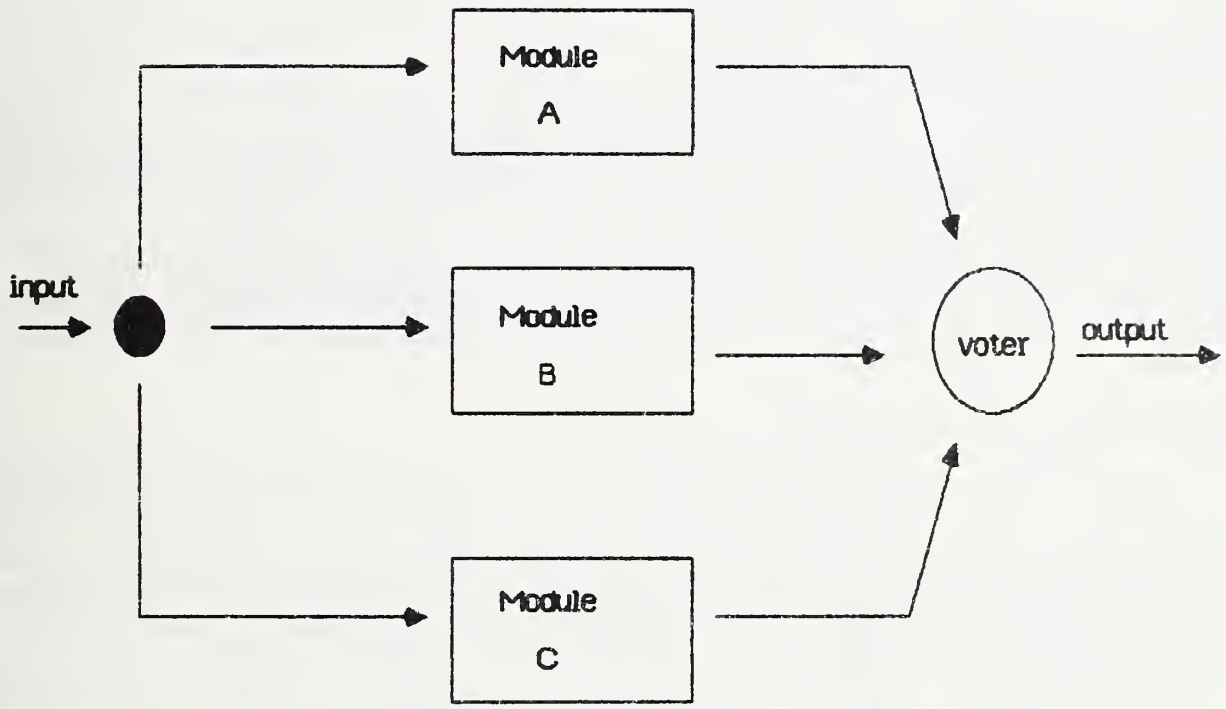


Figure 11: Triple Modular Redundancy (TMR) System with voting

- Error correction codes (ECC). The representation of information by code sequences that will enable the extraction of original information despite its corruption.

- Recovery block method. Several independent programs are developed to perform a specific task. If a fault is detected in one program, an alternate program is selected to execute the task.

- N-version Programming. The output of N independently coded and executed programs are compared. By majority voting or a predetermined stategy, a 'correct' result is identified. Since the programs are developed independently, it is assumed that the probability of a common error is close to zero.

4.2 Implementation Techniques

A variety of reliability related techniques can be implemented by the system planner. Several of these implementation techniques are simply variations of design features described above. The implementation techniques may require adjustments to current procedures, the system configuration, or management policies. The following are examples of several techniques a system planner can implement with the addition of hardware, or software, or through the management of the facility.

Implementation Techniques:

The first four techniques are based on principles of redundancy.

- Duplication of systems. The replication of the computer system, subsystem, software or peripherals to provide a replacement capability should a failure occur. The ability to switch to an alternative system (subsystem, etc.) enables usage of the system to continue as repairs (corrections) are made to the failed unit.

- Environmental backup. The ability to use alternative sources of power, air conditioning, and communication lines in case an outage should occur. Battery backup, uninterrupted power supply (UPS), and frequency interference filters are examples of techniques to counter environmental interferences.

- Reconfiguration. The removal or disenabling of a faulty module from the rest of the system. The system continues to function (without the faulty module) but at a degraded level, e.g. with limited capacity or capabilities.

- Software archive. The duplication of software to replace corrupted or inaccessible data or programs. These redundant copies of software should be kept current, on alternate storage media, and available for use should a failure occur. Due to the possible threat of damage or theft, consideration should be given to storing the software archive at an alternate location

(off-site).

- Maintenance policy. The establishment of a preventive maintenance (PM) program periodically to check the system and correct potential faults. PM is a means of locating and correcting problems before they propagate through the system and cause major damage. A corrective maintenance program should also be provided. This activity normally occurs after the system ceases to function as originally intended. The system is returned to an error-free state.

- Personnel. Support personnel (operators, analysts, technicians) should be available while the system is operational and able to intercede if a problem occurs. For example, if an operator is required to boot a system, a provision should be made for having an operator on duty any time the system is operational. Staff schedules should be adjusted in order to curtail delay due to human unavailability. Proper training and complete documentation are aids to help personnel act quickly when a problem occurs and prevent or minimize loss of information or loss of the system (i.e. crash).

- Supplies. Directly related to the hardware or software of the system, the use of quality printer ribbon, disks, tapes, etc. can eliminate many of the peripheral-related failures.

5.0 RECOVERY STRATEGY

The purpose of recovery is to restore the system to a correctly functioning state from an erroneous one. The reliability objectives, the effects of a fault, and the system's tolerance of the resulting errors must be understood and considered in the determination of a recovery strategy.

5.1 Recovery Procedures

It is necessary for the system planner to establish procedures to recover from a failure and restart the system quickly. While many of the error recovery procedures pertain to methods imbedded in the system architecture (both hardware and software), others are a result of facility management practices or site implemented techniques. Imbedded procedures are limited by the vendor design and need to be specified during the planning and acquisition of the system. Facility management and site implemented techniques can be established at system initiation as well as during the operational stage. Section 4 gives details of possible imbedded (design) techniques and implemenation techniques.

In choosing recovery techniques, the system planner needs to evaluate the system requirements with respect to:

1. the amount of time between the occurrence of a failure and the start of the recovery process

2. the amount of time between the initiation of recovery and the restoration of the system.

3. the amount of human interaction (maintenance) required to restore the system. (Manual recovery techniques generally require more time than do automatic recovery techniques.)

5.2 Recovery Levels

The level of computing achieved through recovery procedures can be grouped into 3 classes.

- Full recovery returns the system to the set of conditions existing prior to the failure. Hardware and software possess the same computing capability as before. Typically, failed components are replaced by spare equipment (hardware) or duplicate software modules. Data and information are returned to their pre-failure

state.

- Degraded recovery means the system is returned to an operational state, but with a reduced computing capacity. Malfunctioning hardware and software, and corrupted data and information are identified and excluded from the system.

- Safe shutdown occurs when the system cannot maintain a minimum level of computing capacity. The system is shut down with as little damage and as much warning as possible. Diagnostic information and warning messages are given. Attempts to reduce the amount of damage to the remaining hardware, software, and data are made.

The objective of these recovery levels is to avoid a hard, complete crash of the system. If full recovery cannot be achieved, the alternatives are to continue processing in a degraded mode or to shut down the system. To determine the appropriate recovery level, the system planner must answer the following questions:

1. System application requirements:
   Can the application tolerate a shut down or graceful degradation?

2. Extent of damage to hardware and software:
   Can critical operations continue to be processed despite damage to system components?

3. Speed with which the operation must be recovered:
   Is there sufficient time for the recovery process to complete without violating system operational (speed, safety, etc.) requirements?

4. Technical capability to implement the recovery techniques:
   Is it possible to design or implement techniques to identify, locate, correct, and record a fault to the system or its components?

5. Cost to implement the recovery process:
   Is the recovery level cost beneficial?

6. Amount of external assistance (manual intervention):
   How much maintenance is required and will be available for recovery efforts after a failure occurs?

All the above questions should be examined with respect to the system as a whole and any critical and/or self-contained subsystems or components.

6.0 THE RELIABILITY PROGRAM

6.1 Implementing A Reliability Program

A reliability program should be initiated with the conception of the system, continue through daily operation, and end only when the system is retired from use. The reliability program should be incorporated into the system life cycle as early as possible in order to maintain consistency with overall system objectives, as well as to minimize the difficulty and cost of implementation.

The tasks involved in implementing a successful reliability program require the participation of personnel from a variety of organizations (e.g. system planner, technical specialists, users, procuring personnel, vendors). To ensure the success of the program, the system planner needs to understand the reliability engineering and management tasks and coordinate the efforts of the people required to perform the tasks. The system planner must be able to:

- understand reliability engineering terminology
- specify reliability performance tasks
- schedule when the tasks are to be performed
- identify personnel to perform the tasks
- understand the consequences of eliminating or curtailing the tasks
- identify major alternatives with respect to cost and risks
- locate additional information/consultants if needed.

6.2 Financial Considerations

There are several fundamental costs associated with the implementation of a reliability program. Calculating the costs vs. benefits of such a program is not an easy task [FIPS64]. The analysis should provide the system planner with the information needed to evaluate alternative approaches and to make decisions about initiating, procuring, continuing, or modifying the reliability program.

The system planner should view the costs of a reliability program as an investment that is amortized over the life cycle of the system. It is important that the system planner consider not only the cost to implement a reliability program, but also the cost of not implementing the program. A knowledge of these considerations can aid the system

planner in accessing the effects of reliability on the costs of ownership [SIEW82].

6.2.1 Cost Of Not Implementing A Reliability Program -

As the organization becomes increasingly dependent on its computer systems, the impact of a failure on the organization needs to be examined and evaluated. Interruption of service by any fully utilized system will eventually lead to a loss of money or time. It is not possible to generalize the cost of failing to implement a reliability program since it is dependent on the system applications and the frequency with which the system fails. However, the greater the application's dependence on the computer system, the higher the cost of downtime. These costs are reflected by:

- o a disruption or delay in production, development, and schedules,
- o loss or corruption of information (data and programs),
- o an increase in maintenance costs,
- o an increase in aquisition costs of spare (replacement) parts,
- o a decrease in user productivity and confidence in the system.

6.2.2 Cost Of Implementing A Reliability Program -

Associated with the elements of a reliability program is the cost to implement and maintain those elements. The costs may be either one-time expenses or recur over the operational life of the system. Despite these costs, the deployment of a reliability program and its resulting reliability improvements will yield reductions in future operation and maintenance expenditures. The costs are reflected in the following reliability program elements and activities. (Further explanation of these elements can be found in previous sections of this guide.)

- o reliability specifications in RFP (design techniques, reliability measures, controls, and thresholds),
- o site preparation (alternate power sources and communication lines),
- o redundancy of critical subsystems,
- o hardware and software monitors,
- o auditing and analysis software,

- o auditing, analysis, and refinement of the reliability program,
- o routine maintenance program (preventive maintenance),
- o spare parts inventory,
- o trained support personnel (operators, analysts, technicians),
- o duplication and storage of software (programs and data).

6.3 Activities For Establishing And Maintaining Reliability

The successful evolution of a reliable computing system requires several important management decisions and actions. Outlined below are the major activities in the establishment and maintenance of a reliability program.

1. Establish Reliability Goals:
   - o Determine the probability of a failure and its impact on the system.
   - o Determine how much should be spent on reliability concerns (remember, reliability affects other life cycle costs, e.g. maintenance).
   - o Determine and integrate reliability concerns with overall system objectives.

2. Consider alternate ways of achieving reliability goals:
   - o Consider the various design and implementation techniques.
   - o Determine the feasibilty of implementing the targeted reliability techniques.
   - o Consider all options. For example, to provide backup computing ability, weigh the advantages of implementing a redundant computer system vs. buying time-sharing services.

3. Select Appropriate Measures and Controls:
   - o Include controls that provide early warning of reliability problems.
   - o Incorporate measures that can provide information on the performance objectives of the system.

- Include several complementary and overlappling measures in order to achieve realistic and complete reliability information.
- establish an appropriate schedule (frequency) for collecting and assessing reliability data.

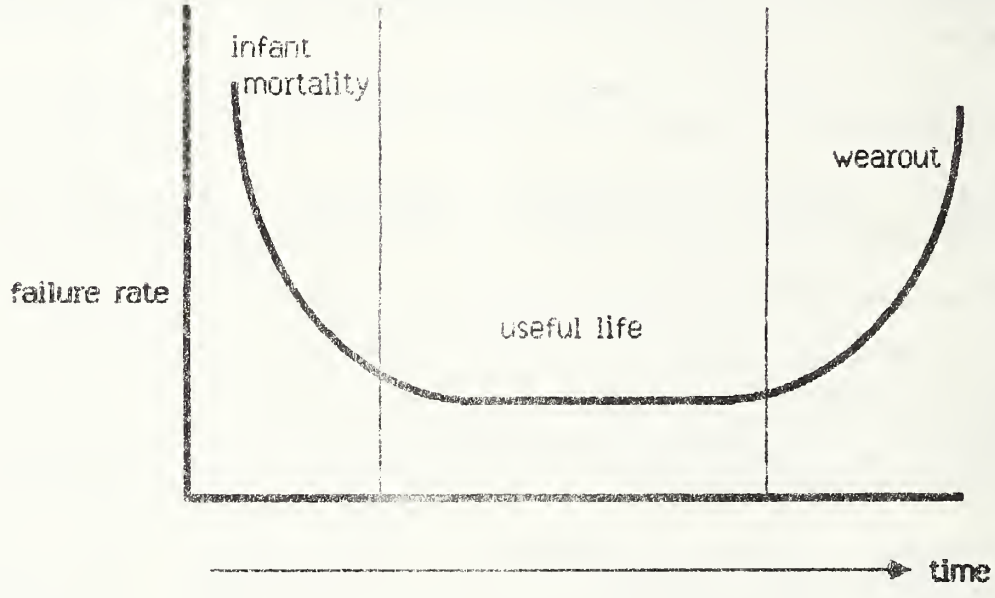
4. Establish clear contracts with system vendors:
   - Alert internal procurement personnel to reliability needs.
   - Define reliability requirements clearly and in detail.
   - Amplify requirements and tasks in RFP statement of work, technical specifications, data requirements list, data item descriptions, etc.
   - Identify the responsible agent for each requirement and/or product (including groups or personnel internal to your organization).
   - Specify penalties and contingency plans for failure to meet performance standards.

5. Define acceptance criteria:
   - Specify levels of acceptable computing performance for the system and its subsystems.
   - Define threshold levels and criteria for reliability measures.

6. Develop maintenance strategy:
   - Provide for remedial maintenance to correct any problems on a timely basis.
   - Determine optimum schedule and scope of preventive maintenance.
   - Determine if stockpiling of spare parts is cost-beneficial. If so, determine the type and quantity of equipment to store.

7. Monitor the system:
   - Implement quality control techniques to retard the deterioration of the system.
   - Process and evaluate the reliability information.

- Plan and conduct periodic reviews of the system and adjust accordingly. Account for system aging and wear out (figure 12) and initiate change when more reliable system components are available and cost-effective.



As the system gets older, more failures occur due to wear-out of the components. The time to repair increases because of the difficulty in obtaining replacement parts and knowledgable repair personnel.

Figure 12: Bathtub curve - Failure rate as a function of time

7.0 BIBLIOGRAPHY AND RELATED READING

[AVIZ79] Avizienis, Algirdas, "Toward a Discipline of Reliable Computing", Euro IFIP 79 Proceedings, 1979.

[BEAU79] Beaudry, M. Danielle, "Performance-Related Reliability Measures for Computing Systems", IEEE Transactions on Computers, Volume C-27, Number 6, June 1979.

[BESH83] Besharatian, Hossein, "A Methodology for Reliability Evaluation of Distributed Processing Systems", Total Systems Reliability Symposium, IEEE Computer Society, December 1983.

[BOSS81] Bossen, D.C. and Hsiao, M.Y., "ED/FI: A Technique for Improving Computer System RAS", FTCS-11, Eleventh Annual International Symposium on Fault-Tolerant Computing, IEEE Computer Society, June 1981.

[BOUH79] Bouhana, James, "Using Accounting Data in Performance Reporting", NBS Special Publication 500-52, October 1979.

[BRAN81] Branstad, M., Cherniavsky, J., and Adrion, W., "Validation, Verification and Testing of Computer Software", NBS Special Publication 500-75, February 1981.

[CART79] Carter, W.C., "Fault Detection and Recovery Algorithms for Fault-Tolerant Systems", Euro IFIP 79 Proceedings, 1979.

[CAST81] Castillo, Xavier and Siewiorek,Daniel P., "Workload, Performance, and Reliability of Digital Computing Systems", FTCS-11, Eleventh Annual International Symposium on Fault-Tolerant Computing, IEEE Computer Society, June 1981.

[COPP79] Coppola, Anthony, Reliability and Maintainablility Management Manual, Rome Air Development Center, July 1979, NTIS AD/A-073 299.

[DANI79] Daniels, B.K., "Reliability and Protection Against Failure in Computer Systems", NTIS NCSR R17, January 1979.

[DENN76] Denning, P.J., "Fault Tolerant Operating Systems", Computing Surveys, December 1976.

[FIPS31] Guidelines for ADP Physical Security and Risk Management, NBS FIPS PUB 31, June 1974.

[FIPS64] Guidelines for Documentation of Computer Programs and Automated Data Systems for the Initiation Phase, NBS FIPS PUB 64, August 1979.

[FIPS65] Guideline for Automatic Data Processing Risk Analysis, NBS FIPS PUB 65, August 1979.

[FIPS73] Guidelines for Security of Computer Applications, NBS FIPS PUB 73, June 1980.

[GOOD79] Goodman, Geoffrey H., Methodology for Establishing a Computer Performance Management System: A Case Study, NBS Special Publication 500-52, October 1979.

[HERN83] HERNDON, M.A. and McCall, J.A., "The Requirements Management Methodology: A measurement Framework for Total Systems Reliability, Total Systems Relability Symposium", IEEE Computer Society, December 1983.

[HOPK80] Hopkins, Albert, "Fault-Tolerant System Design: Broad Brush and Fine Print", Computer, March 1980.

[HOWA81] Howard, Phillip C., "Capacity Planning and User Service

Fulfillment", EDP Performance Review, Volume 9, August 1981.
[JACO81] Jacobsen, Leroy, "What, When, How of Hardware Monitors", Data Management, March 1981.
[KELL83] Kelly, John, Capacity Planning: A State of the Art Survey, NBS-GCR-83-440, July 1983.
[LASA83] LaSala, Kenneth and Siegel, A., "Improved R & M Productivity by Design for People", 1983 Proceeding of the Annual Reliability and Maintainability Sysmposium, IEEE Computer Society, January 1983.
[LAWL82] Lawless, Jerald, Statistical Models and Methods for Lifetime Data, John Wiley and Sons, 1982.
[MART83] Martin, Roger and Osborne, W., Guidance on Software Maintenace, NBS Special Publication 500-106, December 1983.
[McDE80] McDermid, J.A. "An Overview of Reliable Computer System Design", NTIS: AD-A099 271/1, May 1980.
[MIL217] MIL-STD-217 Reliability Prediction of Electronic Equipment.
[MIL757] MIL-STD-757 Reliability Evaluation from Demonstration Data.
[MIL791] MIL-STD-791 Reliability Tests, Exponential Distribution.
[ODA81] Oda, Y., Tohma, Y. and Furuya,K., "Reliability and Performanc

Evaluation of Self-Reconfigurable Systems with Periodic Maintenance", FTCS-11, Eleventh Annual International Symposium on Fault-Tolerant Computing, IEEE Computer Society, June 1981.
[ONEI83] O'Neil, Don, "Software Reliability: A Perspective", Total Systems Reliability Symposium, IEEE Computer Society, December 1983.
[PATR79] Patrick, Robert, Performance Assurance and Data Integrity Practices, NBS Special Publication, January 1979.
[POWE82] Powell, Patricia, Planning for Software Validation, Verification, and Testing, NBS Spcial Publication 500-98, November 1982.
[PRES83] Presson, P. Edward, et. al., Software Interoperability and Reusability Guidebook for Software Quality Measurement, Rome Air Development Center, July 1983, RADC-TR-83-174, Vol II (of two).
[SHAW81] Shaw, James and Katzke, Stuart, Executive Guide to ADP Contingency Planning, NBS Special Publication, 500-95.
[SIEW82] Siewiorek, Daniel and Swarz, Robert, The Theory and Practice of Reliable System Design, Digital Press 1982.
[SRIV83] Srivastava, Rajendra and Ward, Bart, "Reliability Modeling of Information Systems with Human Elements: A New Perspective", Total Systems Reliability Symposium, IEEE Computer Society, December 1983.
[THOM83] Thompson, William, "System Hardware and Software Reliability Analysis", Annual Reliability and Maintainability Symposium, IEEE Computer Society, January 1983.
[TOMI80] Tomita,K. and Tashiro, Kitamura, "A Highly Reliable Computer System - Its Implementation and Result", FTCS-10, Tenth International Symposium on Fault-Tolerant Computing, IEEE Computer Society, October 1980.

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET** (See instructions)	1. PUBLICATION OR REPORT NO. NBS/SP-500/121	2. Performing Organ. Report No.	3. Publication Date January 1985

4. TITLE AND SUBTITLE

Computer Science and Technology:
Guidance on Planning and Implementing Computer System Reliability

5. AUTHOR(S)

Lynne S. Rosenthal

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

National Bureau of Standards
Department of Commerce
Gaithersburg, MD 20899

7. Contract/Grant No.

8. Type of Report & Period Covered

Final

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

Same as item #6

10. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 84-601159

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

Computer systems have become an integral part of most organizations. The need to provide continuous, correct service is becoming more critical. However, decentralization of computing, inexperienced users, and larger more complex systems make for operational environments that make it difficult to provide continuous, correct service. This document is intended for the computer system manager (or user) responsible for the specification, measurement, evaluation, selection or management of a computer system.

This report addressess the concepts and concerns associated with computer system reliability. Its main purpose is to assist system managers in acquiring a basic understanding of computer system reliability and to suggest actions and procedures which can help them establish and maintain a reliability program. The report presents discussions on quantifying reliability and assessing the quality of the computer system. Design and implementation techniques that may be used to improve the reliability of the system are also discussed. Emphasis is placed on understanding the need for reliability and the elements and activities that are involved in implementing a reliability program.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)

computer system reliability; quantification of reliability; recovery strategy; reliability; reliability program; reliability requirements; reliability techniques

13. AVAILABILITY

☒ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☒ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

48

15. Price

ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SCIENCE & TECHNOLOGY

Superintendent of Documents,
Government Printing Office,
Washington, DC 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Bureau of Standards Special Publication 500-.

Name _____

Company _____

Address _____

City _____ State _____ Zip Code _____

(Notification key N-503)

☆ U.S. GOVERNMENT PRINTING OFFICE: 1985-461-105/10183

NBS Technical Publications

Periodicals

Journal of Research—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent Bureau publications in both NBS and non-NBS media. Issued six times a year.

Nonperiodicals

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NBS under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the **following** *NBS publications—FIPS and NBSIR's—from the National Technical Information Service, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

U.S. Department of Commerce
National Bureau of Standards
Gaithersburg, MD 20899