Measuring System and Software Reliability using an Automated Data Collection Process

Brendan Murphy
Digital Equipment Scotland Ltd
Mosshill Industrial Estate
Ayr KA6 6BE
Scotland
bmurphy@dppsys.enet.dec.com

Ted Gent
Digital Equipment Corporation
129 Parker Street
Maynard
MA 01754-2198 USA
gent@poboxa.enet.dec.com

Summary

The factors which impact the behaviour of the customers computing environment, which is undergoing a revolution away from a server or timeshare centric model to a client/server or distributed model, can no longer be identified solely through using traditional methods of data collection. Digital Equipment Corporation has developed an automated data collection process, collecting on-system data logging information from customer sites that has yielded consistent, quantitative, high integrity information. This information has been used to pro-actively focus on direct product and process improvements. This paper describes the on-system data logging process and analysis methodology used by Digital to measure system, product and operating system reliability, proving examples of the application of the techniques and provides insight into the causes of failures.

Key Words: System Reliability, Software Reliability, Automated Data Collection, Customer Survey, Event Logging, Operating System Reliability.

1. Introduction

Digital has been monitoring systems in the field, using a variety of data collection techniques, for over 15 years. During that period significant changes have occurred in the reliability profile of systems and in the operating and development environments. Examples of changes in the reliability profile of VAX systems are described in this paper. In addition, increasing competitive pressures are reducing development cycle times and development budgets which makes it imperative to focus on product quality improvements in areas with the greatest impact on customer satisfaction.

It is essential that a data collection and analysis system provide information contributing to design direction and trade-offs. A number of traditional methods are available for providing performance feedback and the strength and weaknesses of each method is examined in this paper. Due to the limitations of these methodologies, Digital has developed an on-line data capture process which provides data to product design, manufacturing and services organisations to continuously improve the reliability of Digital products and systems.

This paper provides a detailed description of the on-line data capture process and the techniques applied to analyse the data. Digital has successfully used this process for a number of years resulting in a substantial amount of behavioural information being available for analysis. This paper provides examples of some of the information captured through this process and describes how the process measures the reliability of systems and versions of operating systems.

2. Changes in System Reliability

In the 1970's and early 1980's, the majority of customer systems were stand alone servers driving non-intelligent terminals. The reliability of the hardware and the operating systems were the significant factors impacting the performance of these systems. Customer sites were mainly homogeneous with systems managed by MIS departments specialising in particular product sets. Significant changes have occurred to both the environment into which the systems are configured and the reliability profiles of individual products. Digital has been monitoring these changes and measuring their impact on its product sets. This section describes these changes using, as an example, information collected from VAX systems on customer sites.

2.1 Changes in Product Reliability

During the late 1970's and early 1980's, the reliability of hardware and the operating system were the major contributors to system outages. In 1985 these factors accounted for 70% of all system crashes occurring on Digitals' VAX systems on customer sites. Other factors causing system crashes were not measured at that time. They were not viewed as being significant (Figure 1. estimates the impact of other factors causing system crashes in 1985).

Over the last 10 years, the reliability of both hardware and operating systems has dramatically increased. Improvements to hardware reliability is primarily due to the use of large scale integration and the wide scale adoption of Computer Aided Design processes. Improvements in the reliability of operating systems are due to a shift in focus from adding pure functionality to balancing any added functionality with reliability and recoverability attributes.

Customers and product suppliers continue to view the hardware failure rate as an important measure of system reliability, in spite of the changes in the profile of system crashes. This may be due to (a) hardware failures having a significant impact on the Customer operation, (b) the lack of an acceptable industry standard to measure reliability in terms of the rate of system interruptions, or (c) hardware failures resulting in a significant cost to the service providers through both part replacement and the cost of the service engineers visiting the customer site.

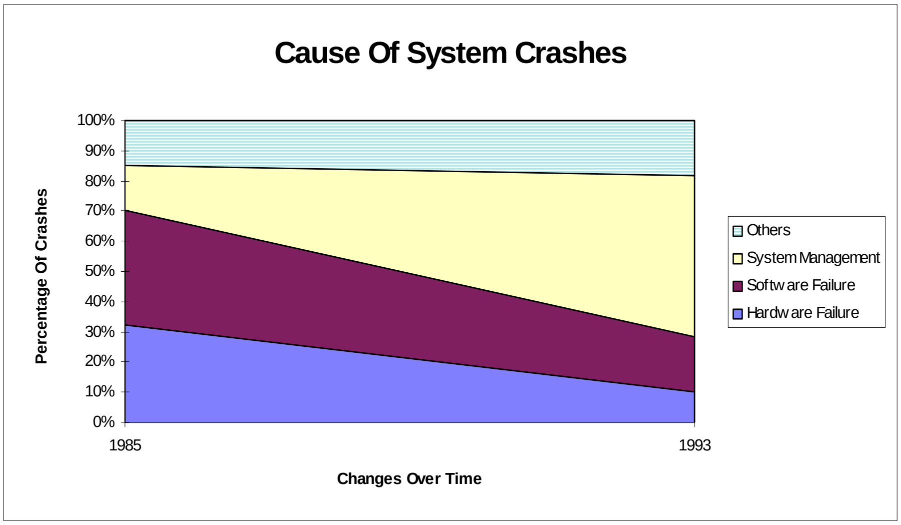


Figure 1: Cause of System Crashes

Root causal analysis of system crashes performed on VAX systems has identified the increasing impact that system management problems have on the system crash rate. In 1993, over 50% of system crashes on VAX systems were due to system management problems. The crash types classified within this category are: (a) Crashes resulting from system management actions. Examples are the incorrect setting of system parameters, the incorrect installation of applications, the incorrect configuration of systems etc., and (b) Multiple crashes resulting from one problem. Crashes increasingly occur after a disruption to the system. The system manager may ignore the initial crash and only address the problem as a result of subsequent crashes. Multiple crashes also occur due to the complexity in diagnosing the cause of system problem. The more complex the environment the greater the difficulty in diagnosing system problems.

The types of crashes that fall into the 'other' category, as shown in Figure 1, are due to: (1) Applications failures resulting in system crashes, (2) Configuration/networking problems, or (3) Power Outages. The increasing use of Uninterruptable Power Supply (UPS) systems to protect servers is decreasing the number of crashes due to power failures.

The changes in the profile of the cause of system crashes, as shown in Figure 1, specifically the increase in the number of system management related crashes, are impacted by the changes in the operating environment.

2.2 Operating Environment

The Customer computing environment is undergoing a revolution moving away from a server or timeshare centric model to a client / server or distributed model. Decentralisation of computing resources and the simplified client system profile have resulted in less management of the central system and on increased management and maintenance burden on the end-users. The tools available to manage the networks and the client server environments have not kept pace with the rapid changes away from the server centric model which results in the systems management directly impacting the overall reliability of the system. The changing nature of the operating environment is impacting the cause of system crashes and the rate and profile of system interruptions. Problems occurring in Customers computing environments are increasingly resolved through the action of an operator shutdown on one or more of the systems within that environment.

Analysis of the cause of system interruptions, on VAX systems, has shown that only 10% of system interruptions are a result of system crashes, as shown in Figure 2. Efforts to improve hardware reliability and decrease the rate of operating systems failure will only address 2% to 3% of all system interruptions.

Attempts have been made to analyse the cause of system interruptions through requesting information, during the reboot process, from the system manager to classify their cause. This process has been successful for small targeted pilot users, but has not been successful when applied to a generic group of users for practical reasons. During the reboot process, the system managers focus is on restoring the total system, which is a complex and error prone process. Placing extra requirements, which are voluntary, produces very little useful data (i.e., the system manager does not need to enter a reason but may just hit the return key).

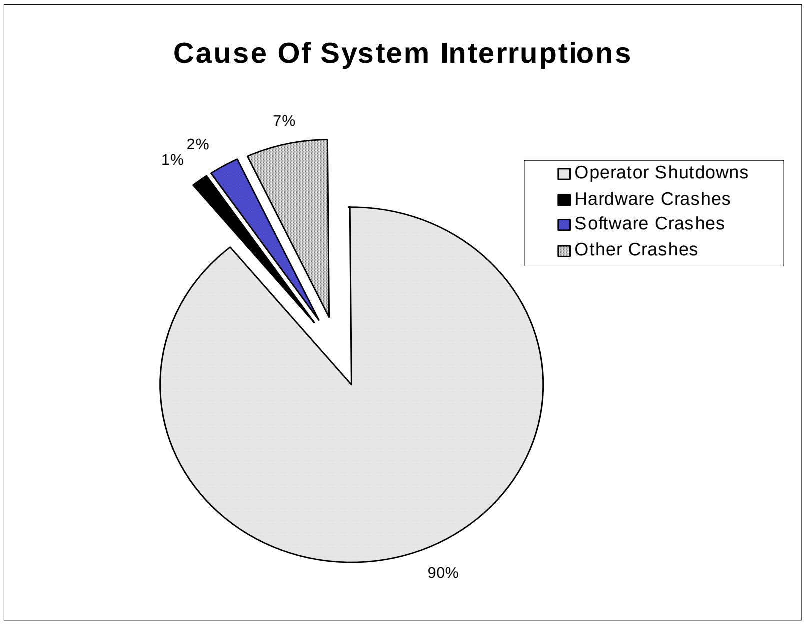


Figure 2: Cause Of System Interruptions

Changes in the profile of systems failures are also reflected in the variation in the behaviour of products on customer sites. The behaviour of a typical Digital VAX server running the same version of OpenVMS VAX, on over 130 customer sites was captured and analysed. This showed a wide variation in the behaviour of these systems, ranging from one system suffering a system interruption every 8.5 hours (the majority of the system interruption occurred within a three month period and was due to a major re-configuration of all systems on the site) to a number of systems that have had no system interruptions since installation, for over one and a half years.

The performance of products on the customer site is increasingly being impacted by system issues. To address the root cause of these system issues, Digital Equipment Corporation has developed an automatic data collection process which, when combined with the traditional methods of data collection, provides a complete picture of system behaviour.

3. Traditional Data Collection Methodologies

Digital Equipment Corporation uses a number of traditional methods to obtain product feedback including customer surveys, examining service activity, and reviewing direct customer feedback. The strengths and weaknesses of each are discussed in this section.

3.1 Customer Surveys

Customer surveys address all aspects of product behaviour, providing a method of comparing equivalent products and services across companies with a focus on the priority that the customer places on different aspects of these products and services. Surveys also provide the only method of capturing the customers perception of the look and the feel of products. Customer surveys present a subjective view of system behaviour and can be biased by experience with the immediate past product performance. They can also be biased due to pre-selling of the survey, where a customer is aware that a survey is about to arrive and this raises their level of awareness of problems. In addition, the accuracy of the survey is dependent upon the phraseology of the questionnaire.

3.2 Service Activity

A corrective maintenance service call to a customer site will usually represent a customer satisfaction issue. Comparing the average rate of service calls for each product provides a method of comparing the reliability of those products. Measuring the failed hardware replacement rate provides an indicator of hardware reliability and, with analysis can provide a measure of the individual component failure rates. Measuring the average time for field service to correct the problem provides a measure of the effectiveness of field service processes. The service call rate is a reactive measure and primarily focused on hardware failures. It is a measure of those factors that impact the customer to the point that they contact the servicing organisation.

3.3 Customer Feedback

Customer Feedback addresses all aspects of product behaviour from documentation errors or omissions to major software reliability issues. Customer feedback for software issues can be classified as a bug report with a specific level of severity. Engineering and services groups have a guaranteed level of response which is dependent upon the severity of the bug. Comparison between rate of bug reports and response speed of the correction of bugs provides a measure of product and process performance.

The frequency of problem reports does not provide a measure of the impact of a problem. Customer feedback addresses individual problems but does not identify systemic problems. This method of measuring performance is also impacted by the customers "threshold of pain" where customers accept a finite problem rate without reacting in a complaints mode. Table 1 presents the limitations of traditional methodologies which highlight the need for an additional source of data.

3.4 Overview of methodologies

Table 1: Overview of traditional methodologies

Strong Points	Weak Points
Measuring hardware reliability (Service Activity)	Diagnosis (Root Cause Analysis).
Measuring factors seriously impacting customers (Service Activity & Customer Surveys)	Capturing the actual system behaviour on the customer site.
Measuring the rate of software bug reports (Customer Feedback)	Non-Systemic.

4. Digitals' Automated Closed-loop Process.

4.1 Introduction

Digitals' automatic closed loop process, Digital Product Performance (DPP) Programme, is a generic process capturing the performance behaviour of Digital products on the customer site. All data captured from the customer site is automatically copied back to the DPP group based at the Digital manufacturing site in Ayr, Scotland. The DPP process was initially developed to monitor the behaviour of OpenVMS VAX systems and has been extended to monitor ULTRIX, OSF/1 and OpenVMS AXP systems.

The DPP process can be applied to any product containing an on-system event logging process, which automatically maintains a history of the major event occurrence. A description of the on-system event logging process is contained within this section. The DPP process consists of four items; mainly, (1) Data collection on the customer site, (2) Data transportation from the customer site to the DPP group in Ayr, Scotland, (3) Management and storing of the data, and (4) Use of analysis software.

4.2 On-System Event Logging

On-System event logging is an integral part of the fault management of the majority of operating systems. Data logging has been developed to assist in the ongoing management and repair of such systems by providing a history of the major events occurring on each system. Hardware faults occurring on a system are managed by the hardware fault management system, which provides information to the operating system regarding each failure. The operating system dictates how these events are logged. The quality of information captured through the on-system event logging is dependent upon both the fault management of the operating system and the fault management of the hardware.

Event logging has traditionally been product based, with each product separately logging information. The amount and quality of information, stored in the event log, differs between products, operating systems, and versions of operating systems. The traditional method of storing information is through an event log(s), which is a continuous record of events occurring on the product. An overview of the OpenVMS event log is presented in Figure 3.

Certain operating systems record a special event, a time stamp, to indicate that the system is operational, as shown in Figure 3. A time stamp is written to the end of the event log and is continually overwritten, with the current time, until a new event occurring on the system, is written to the event log. Occasionally a system may suffer an interruption which is not captured in the event log. This can occur on all operating systems. The time stamp identifies the period of time that the system was unavailable due to such an interruption.

A product based method of logging information does not completely account for the operations of complex systems. Each product logs its own information; however, there is no logging mechanism for system information. A greater complexity of the total system results in a greater risk that the system events will not be logged.

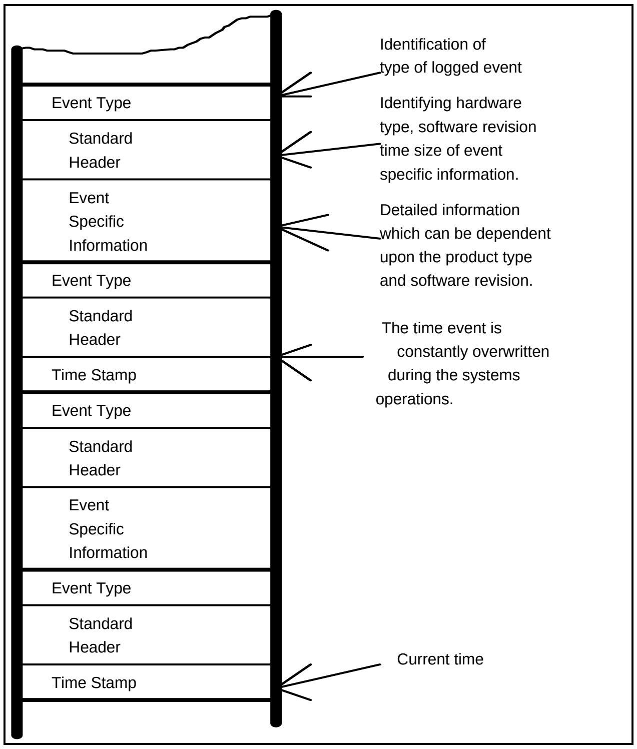

Event Type
Standard Header
Event Specific Information
Event Type
Standard Header
Time Stamp
Event Type
Standard Header
Event Specific Information
Event Type
Standard Header
Time Stamp
Identification of type of logged event
Identifying hardware type, software revision time size of event specific information.
Detailed information which can be dependent upon the product type and software revision.
The time event is constantly overwritten during the systems operations.
Current time

Figure 3: OpenVMS Error Log

4.3 Data Collection Process

A generic data collection process has been developed to monitor all operating systems used on Digitals' products. The implementation of the process differs between operating systems due to the difference in their on-system event logging processes. Data collection software captures the behaviour of individual systems on the customer site, primarily using the information captured through the on-system event log. The data collection software: (1) requires no special system privileges, (2) has a minimal system resources requirements, (3) can be de-installed, (4) collects no information that could be used against the customer (e.g., no verification of software licences), (5) is not required to be installed on the system disk, and (6) is simple to install. The data collection process consists of an installation process and an ongoing monitoring process.

The installation process captures information from the installer who identifies (1) the system(s) to be monitored, (2) the serial number of the system(s), (3) the site name, (4) the customer name, and (5) the customers site address. The installation process initiates a background monitoring process and inserts a start-up command file, into the system start-up command file, to restart the monitoring process upon a system reboot. A separate mailing process, described in section 4.4, is installed on the system. The information collected during the installation process is automatically mailed to Digital Equipment Corporation in Ayr, Scotland. After installation, the DPP process runs automatically requiring no additional system management intervention.

The monitoring process, running daily at an off peak time defined at installation, collects (1) All new events written to the event log: The event log is continuously updated by the on-system event logging process. DPP software retains a pointer to the last piece of information copied from the event log. The next time the process is invoked, it copies all new information from the event log and subsequently updates the pointer, (2) Configuration information: The configuration and the rate of change of configuration has been identified as potential causes of system problems. This process captures the number of connected servers, clients and peripheral devices (disks, tape drives, etc.) once a week, and (3) Profile of crash dump: A separate process extracts a profile of the crash dump to assist in identifying the cause of the system crash.

All data that is to be collected from the customer site is stored in a central area, designated by the system manager during the installation of the DPP software, in preparation for transportation.

4.4 Data Transportation

The DPP process uses whatever transport mechanism already exists on the customer site. A background process, compresses the collected DPP data and automatically mails the data, once a week, using (1) DSNLink Mail: DSNLink is a secure electronic link between the customer and the Digital Equipment Corporation Customer Service Centre (CSC), used to provide remote services to the customer. The data is sent from the customer site to their local CSC using DSNLink mail and is automatically re-routed to the DPP group, and (2) Internet Mail. Through using the mail utilities, the DPP program does not have any impact on the security of the customer systems.

4.5. Data Management Techniques

The DPP process currently monitors approximately 2,000 installed systems in the field. The process, as shown in Figure 4, manages the data collected from these systems and is described in this section. This is a fully automated process which ensures the correctness of the data used for analysis.

DPP (Digital Product Performance)

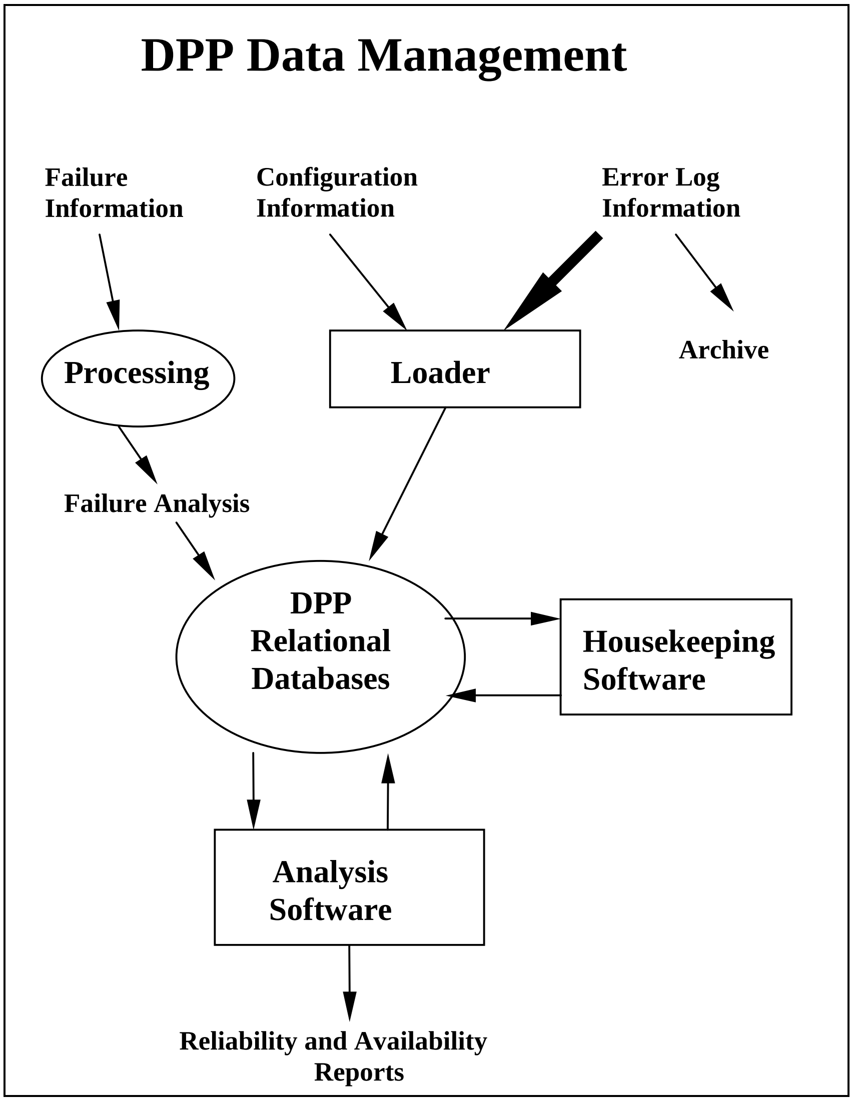


Figure 4: DPP Data Management

4.5.1. Data Processing

Data is collected continuously from the customer site arriving at the DPP site via (1) Internet mail: A continuous process automatically extracts the information from the mail utility in preparation for processing, and (2) Directly copied from other Digital sites: An area with appropriate protection allows the CSC sites to copy behavioural data collected using DSNLink.

Not all of the data collected from the customer system is directly relevant to the current analysis. The data is processed and a sub set is loaded into the relational database(s). The analysis process is being permanently reviewed and continually improved. Data captured from customer systems may not be relevant to the current analysis but may be required for future analysis. All information collected from the customer site is compressed and archived at the DPP site.

4.5.2. Data Storage

The core of the DPP processes consists of a number of RDB/VMS multi file databases, which are designed to allow maximum flexibility in analysis of the data. All monitored systems are uniquely identified through the information collected from the customer during the installation process. When DPP software is installed on the customer site, a registration file is mailed to DPP and loaded into the database identifying (1) Customer name, (2) Customer site, (3) Relevant field service group, (4) Country, and (5) The serial number of the system which uniquely identifies the manufacturing site and date of manufacture.

The ongoing monitoring process continuously sends the captured information from the monitored systems, a subset of this information is loaded into the database. Event data captured by the error log identifies the following systems characteristics; namely, (1) Product type, (2) Version of the operating system, and (3) Time of the event. Through continually monitoring this information, changes in the configuration of the product (e.g., date and time of changes to the version of the operating system) can be identified.

4.5.3. Data Housekeeping

A separate housekeeping function has been developed to ensure the correctness of the data loaded into the database. Problems with the data management process, not identified by the database loader, are corrected through this process.

4.6. Analysis

Storing the data in a relational database allows greater analysis flexibility. Analysis can be performed on any combination of data stored within the database (e.g., it is possible to analyse the behaviour of all DEC7600 OpenVMS AXP systems running in the country of Germany during May of 1994). The section describes some of the techniques used to measure the reliability and availability of Digitals' products. Further examples of the analysis techniques are documented in section 5.

The two types of analysis performed on the behavioural information are (1) Availability - The percentage of time that a product is available for use to the customer, and (2) Reliability - A measure of the time between events which disrupted the system. This provides a measure of the dependability of the system.

Reliability and availability analysis is performed on the behavioural information captured from the monitored systems. While the DPP monitoring process is a continuous, automated, data collection process, occasionally gaps appear in the data received from the customer site. In practice gaps occur due to (1) Network interruptions resulting in data not being available at the time of analysis. A network interruption will not result in lost data. Data is retained, at the customer site, until the network link recovers, (2) Internet mail problems: Data lost through internet mail problems cannot be recovered, (3) Monitored nodes being removed from the configuration, and (4) Fault management information being deleted. Individual system manager may decide to delete product fault management information due to space problems. This will result in a certain amount of data being lost.

Analysis has shown that gaps in the data occur randomly and are not a symptom of particular availability or reliability issue. Reliability and availability analysis is based upon those periods of time for which data has been collected from the systems on the customer site, no assumptions are made regarding the behaviour of the systems during the periods that data was not collected.

4.6.1. Availability Analysis

System availability, including planned and unplanned events, is proportional to the Mean Time Between System Interruptions (MTBSI). System interruptions result in a period of time where the system is not available to the end user. The two factors impacting this time is Mean Time To Repair (MTTR), which is the time required for the system to correctly shut down, and the Mean Time To Recover (MTTRc) which is a measure of the time for a system to become available for use. Availability can be expressed as the proportion of time that the system is available

$$Availability = \frac{MTBSI}{MTBSI + MTTR + MTTRc} \qquad (1)$$

The data collected through the DPP process identifies (1) the periods of time that each system is available for use (System Up Time), and (2) the periods of time between system interruptions and subsequent re-boots when systems are unavailable (System Down Time). DPP calculates product availability through the formula

$$Availability = \frac{\sum System\,Up\,Time}{\sum System\,Up\,Time + \sum System\,Down\,Time} \qquad (2)$$

Low availability may be due to inadequate system maintenance or be a result of product problems. The availability is analysed during four time periods to help identify product trends; namely, (1) Peak Periods (9am - 5pm Monday to Friday). The period that the 'average' customer requires total availability of their system, (2) Off peak periods (5pm - 9am Monday to Friday). The period that the 'average' customer will resolve major problems occurring during the day, (3) Weekend Periods (5pm Friday to 9am Monday). The period that the 'average' customer performs system maintenance, and (4) Total availability (24 hours per day, every day of the year).

Low availability during peak periods will generally be as a result of reliability problems. Low availability during off peak periods will usually be due to problems associated with system management.

Analysis of the collected data has shown that the MTTRec is becoming a significant factor affecting availability. An interruption, planned or unplanned, on a complex configuration can take a considerable time for the total configuration to become available for the end user. As system reliability trends, that were discussed previously in section 2, continue, it is the authors' opinion that system management and configuration will become the dominant factor affecting system availability.

4.5.2 Reliability Analysis

The reliability of products can be measured as the frequency of specific events (e.g., system crashes or system interruptions) occurring on a system. The assumption, widely made in reliability analysis, that events are independent of time, leads to the typical exponential distribution for "time between events" during the normal operating life of a machine.

However not all system events are independent of time. Analysis of the data collected through the DPP process has shown that events of a similar type (crashes, machine checks) occur in groups as shown in Figure 5. In a number of cases, as proved by DPP, a system problem may result in a number of related events which are not independent of time.

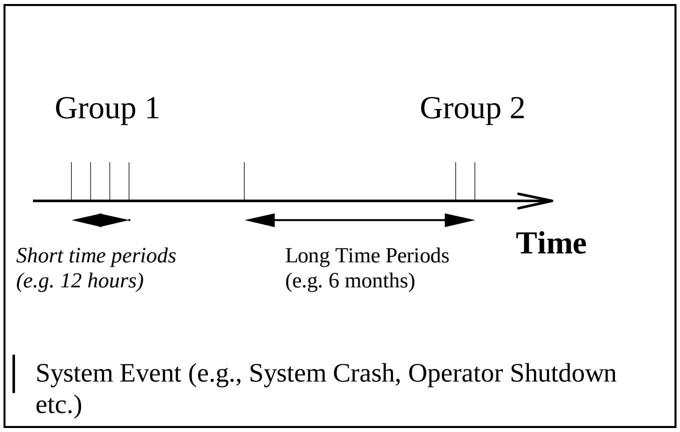


Figure 5: System Time Line

Analysis performed 4 years ago on data captured through the DPP process, formally known as the Digital ESRI process, identified that using a one hour window captures a large number of related events[1]. In similar studies[2], a second crash occurring within one hour after system restart was ignored (i.e., filtered). This study[2] demonstrated that without such "filtering" an exponential distribution could not be fitted to the crash data. Thereby validating filtering as a necessary technique in providing meaningful measurements of unique events.

The changing nature of the profile of system behaviour is increasing the difficulty of choosing a time window to ensure that only unique events are captured. The one hour time window was initially developed based upon the behaviour of stand alone nodes suffering hardware and software failures. A one hour time window provides sufficient time for the stand alone node to fail, re-boot and possibly re-encounter the problem.

Problems affecting a complex environment may result in connected events on individual nodes occurring outside of the one hour time window. (e.g., problems occurring on a client server application, requiring system re-boots on all nodes in the configuration.). The configuration may take over an hour to rebuild its total environment, during this time individual nodes may be available but not accessible, for over an hour, before the re-occurrence of the problem requiring all nodes in the configuration be re-booted.

System management procedures are also complicating the identification of this time window. A number of customer sites install or upgrade complex applications during off-peak periods, usually during the weekend. Problems with the installation may not be identified until the users of the application arrive into work on the following Monday. If the solution to the problem requires a complete reboot of the configuration, the system manager may wait until the following weekend before again addressing the specific problem.

Expanding the time window to a week to allow for the above behaviour may result in independent events occurring on stand alone nodes being ignored. The resultant DPP program uses different time windows depending on the type of analysis that is being performed to ensure maximum significant data capture.

The analysis and reporting of product reliability continues to use a one hour time window. Maintaining a consistent time window allows the identification of comparable reliability trends.

The metric used to calculate product reliability is Annual Rate of Events[3].

$$ARE = \left[\frac{1}{Steady\,State\,MTBSI}\ or\ \frac{1}{Steady\,State\,MTBF}\right] * \left(Hours\,in\,a\,year\right) * Duty\,Cycle \qquad (3)$$

which is a count of the number of specified events occurring on the system in a given year. The two specific events which are currently used to track the reliability of systems are (a) system interruptions, and (b) system crashes. DPP calculates AREs based upon (a) System Up time - The number of monitored hours that the system has been available, and (b) The number of unique events. DPP calculates ARE through the formula

$$ARE = \frac{\sum Unique\,Events}{\sum System\,Up\,Time} * \left(Hours\,in\,a\,year\right) \qquad (4)$$

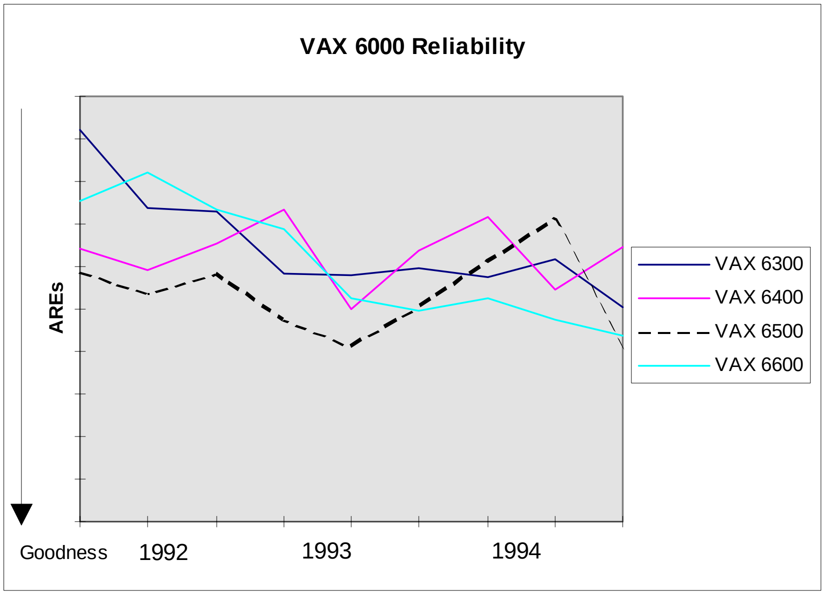


Figure 6: VAX6000 Reliability Trends

A DPP monthly report tracks the reliability of each of monitored hardware product type, as shown in Figure 6, allowing Digital Equipment Corporation to (1) Ensure that no problems are occurring to current released products, and (2) Provide a set of behavioural goals for future products.

5. Analysis Results

5.1 Product Reliability

The analysis of reliability trends of mature product sets shows similarity in the rate of system interruptions. Analysing the rate of system interruptions, using the metric ARE on VAX 6000 computers as shown in Figure 6, indicates little difference in the behaviour of the different product types, in spite of there being differences in the rate of crashes on these systems (as measured through the DPP process) and the hardware failure rate of these products (as measured through hardware return rates).

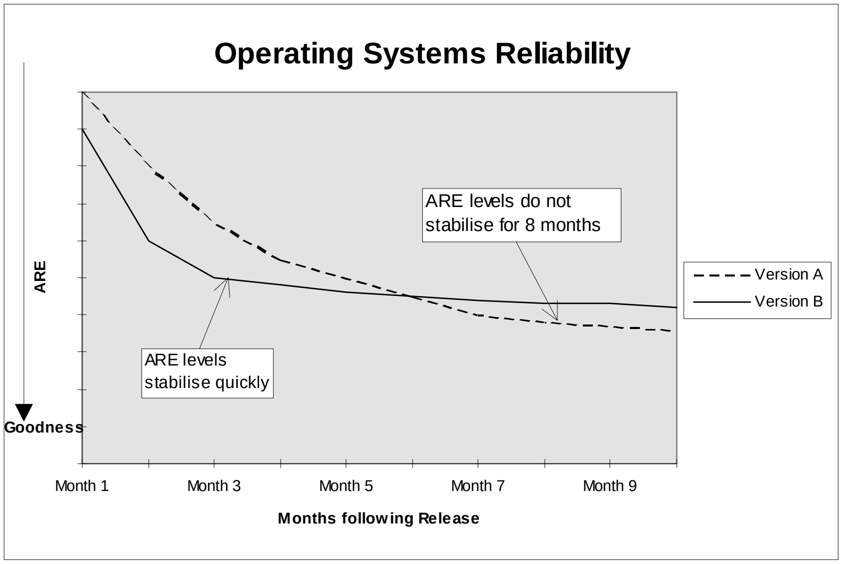
The reliability trend show that hardware is not the dominant factor affecting system interruptions. Further analysis of the reliability trends appeared to show that the version of the operating system has a greater impact on system reliability. This is due to the versions of the operating systems affecting (1) Operating system reliability, (2) System management, (3) Configuration/networking, and (4) Application software.

As previously discussed in section 4, the DPP database contains the system behaviour of products over a number of years. In addition the analysis software provides the capability to measure the performance of product by versions of the operating systems. For example product behaviour can be measured in terms of system interruptions per version of OpenVMS VAX per calendar month as shown in Figure 7. This data was graphed for all relevant versions of the operating systems and normalised to the date of shipment of the version of the operating system as shown in Figure 7.

The following points arise from the graphs in figure 7. (1) These graphs would appear to indicate that it can take up to 8 months for the reliability of systems running a new version of the operating system to stabilise, contradicting the known situation, (2) There is no direct correlation between the rate of improvement in reliability in the period after FRS and the resulting reliability for the different versions of the operating systems, and (3) Figure 6. does not show the dramatic changes in system performance as shown in figure 7. This would indicate that only a small proportion of systems upgrade their versions of the operating system in the months following its release.

This Analysis identified a large proportion of users are running older versions of the operating systems, with some users running operating systems released 4 or 5 years ago. The data trends captured through this analysis were further analysed identifying:

1) The process of installing a new version of an operating system has an impact on the behaviour of that system.
   The impact appears to be proportional to the complexity of the customer configuration.
2) Delays exist between the release of a new version of an operating system and customers installing that version.
   The more mature an operating system is, the smaller the proportion of users that will install the new version immediately following its release. A mature operating system will already be providing the features that the majority of users currently require. New versions of the operating system will contain additional features which are essential for some users but not necessarily for all.


Operating Systems Reliability

ARE levels do not stabilise for 8 months

ARE levels stabilise quickly

Figure 7: Reliability of Operating Systems in the Months following its Release.

The behaviour of newly released products differs from that of mature products. The impact of the installation of a product into a configuration is related to the complexity of the product (i.e., a large server has a greater impact than a PC will have). Hardware and design problems will usually impact customer systems shortly after its installation. For products currently being manufactured DPP measures (1) The system crash rate - If possible the cause of all system crashes are diagnosed, and (2) The non fatal machine bug check rate - Hardware failures may be occur on the system but may not result in a system crash. These failures are caught through the hardware fault management (e.g., memory errors being caught by Error Correction Code (ECC)) systems and logged into the event log. The data collected from products currently under manufacture are compared against mature products at similar stages in their product life cycle.

5.2 Measuring the Reliability of Operating Systems.

Analysis of product behaviour has identified that any method of measuring the reliability of operating systems must address both (1) The period immediately after installation, and (2) The steady state level of reliability as separate entities. The ease of installing a product is not necessarily an indication of the steady state reliability of that product.

Certain problems which occur during the installation process, are unique to that process and can be permanently corrected. Example of these problems are (1) Inter-operability - Certain application may be incompatible with the current version of the operating system. This can be resolved through upgrading the application or through obtaining relevant operating systems or application patches which will correct the problem, (2) Bugs, existing within the operating system, identified shortly after installation. Failures can be corrected through patches or, where possible, by not repeating the procedure which resulted in the system failure, and (3) System management - New features for that version of the operating system may result in system management problem. These may be corrected through learning through mistakes.

For complex configurations, the installation of a new version of an operating system may be combined with a number of other error prone activities (e.g., re-configuration of the system, upgrading of complex client/server, database applications) resulting in a greater level of disruption to the system. Upgrading versions of the operating system is planned well in advance and occurs at a time most convenient to the customers' operations. From a product suppliers perspective, the problems occurring during the installation process are one measure of the effectiveness of its' supply chain process. When the behaviour of a system stabilises the factors impacting its behaviour are (1) The reliability/ ease of management of the operating system, and (2) The support processes capability to manage the distribution of patches.

5.2.1 Analysis Methodology

The reliability of operating systems is measured as the differences in the behaviour of servers running different versions of the operating systems. As products can be configured as a client or as a server, DPP differentiates between clients and servers by how they are configured.

The behaviour of products acting as clients are not combined with those acting as servers due to the different factors affecting their behaviour. A product acting as a client is impacted by the same factors affecting a server but may also be impacted by (a) The physical network, (b) The networking software, (c) The behaviour of the server acting as the boot node for the client and (d) Power outages. Customers are less likely to protect clients using UPS (Uninterruptable Power Supply) systems.

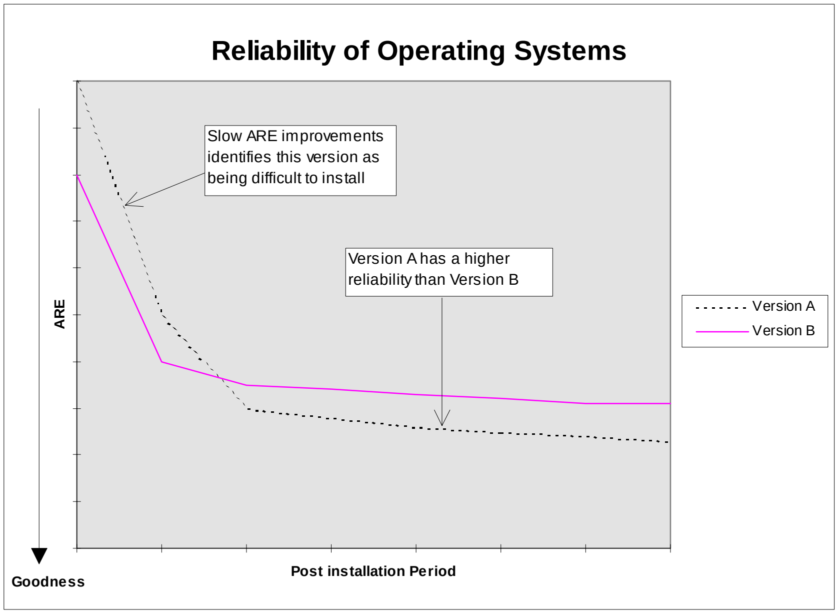


Figure 8: Reliability of Operating Systems.

Analysis has identified that the reliability of operating systems should be measured through changes in its behaviour in the period following its installation. The reliability of each version of the operating systems is analysed by (1) Normalising its behaviour based upon its date of installation on monitored product sets, and (2) Measuring the average behaviour, on the monitored product set, for defined periods after its installation. The monitored periods are adjusted based upon the type of analysis performed. This is implemented through (a) Identifying the target systems for analysis, (b) Identifying the date that versions of the operating system are installed on each systems. If the exact date of installation of a version of the operating system is unknown then the data, for that system, is ignored, (c) Identifying the observed time and the number and type of events occurring on each system, for each version of the operating system for each period. Gaps in the monitored data are excluded from the observed time, and (d) Combining the data collected from each systems for each version of the operating system for each time period.

The results of this analysis are shown in Figure 8. The behavioural trends identify: (1) The impact of the installation process for each version of the operating system, and (2) The steady state behaviour of the operating system.

All versions of operating systems are measured using this technique. Analysis showed, as expected, that different versions of the operating system affects the behavioural graph in different ways. A release may suffer installation problems but be a very reliable release after it stabilises. Other releases may be simple to install but may have long term reliability problems.

This techniques has been applied to a number of operating systems identifying (1) That all operating systems exhibit a similar reliability profile, and (2) The complexity of the operating systems and the complexity of the customer solution both have an impact on system reliability during the period following the installation of the operating systems.

This analysis has shown that the technique, developed by DPP, is applicable to all operating systems. It has also identified that it is inadvisable to compare the reliability profiles for the different operating system unless configuration is factored into the analysis.

6. Conclusions.

Developing a model that captures the performance of products on the customer site is essential for any company to meet the rapidly changing needs of the customer. The movement away from homogeneous server centric sites to decentralised heterogeneous environments is impacting the behaviour of systems and the cost of servicing of the system in ways which are not yet fully understood. The DPP model provides a mechanism to study the actual behaviour of systems and to identify the factors which impact this behaviour. DPP is continually being developed to capture the continually evolving behaviour of the customer environment.

DPP provides a flexible model which measures and provide root cause analysis of the reliability and availability of a wide range of products and operating systems. The results of this analysis, examples of which appear in this paper, have been used within the Digital Equipment Corporation to focus cost effective investment in those areas which drive customer satisfaction.

7. Acknowledgements

The authors would like to thank David R. Jones for his help and motivation in completing this paper. The authors are also please to thank the past and current members of

the DPP team for their hard work and dedication to the program.

REFERENCES

1. P. Moran, P. Gaffney, J. Melody, M. Condon and M. Hayden, “System availability monitoring”, IEEE Transactions on Reliability, Volume 39, Number 4, October 1990.
2. F.A.Nassar and D.M. Andrews, “A methodology for Analysis of Failure Predicition.”,CRC Technical Report, No 85-20, Standford University, September 1985
3. D. R. Jones, V. Murthy and J. Blanchard, "Quality and Reliability assessment of hardware and software during the total Product Life cycle".
   Quality and Reliability Engineering International. VOL 8, 477-483 (1992).
4. J. Gray, “Why do computers stop and what can be done about it”’Proc, Fifth Symposium on Reliability in Distributed Software and Database Systems, pp3-12 Los Angeles, California, 1986.

LIST OF ACRONYMS

ARE	Annual Rate of Events.
CSC	Customer Service Centre.
DPP	Digital Product Performance.
ECC	Error Correction Code
ESRI	European System Reliability Information
MIS	Management Information Services
MTBF	Mean Time Between Failures
MTBSI	Mean Time Between System interruptions
MTTR	Mean Time To Repair
MTTRec	Mean Time To Recover
OpenVMS	Open Virtual Memory System.
RDB/VMS	Relational data base running on OpenVMS.
ULTRIX	Digital’s format of UNIX Operating System
OSF/1	Open System Foundation.
UPS	Uninterruptable Power Supply
VAX	Virtual Architecture eXtension.

AXP, DEC, Rdb/VMS, OpenVMS, ULTRIX & VAX are all trademarks of Digital Equipment Corporation. OSF/1 are registered trademarks of the Open System Foundation, UNIX is a registered trademark of UNIX System Laboratories, Inc., a wholly owned subsidiary of Novell, Inc.