Incident #	Date	Start Time	Total Time (hours)	Corrective downtime (hours)	Preventive downtime (hours)	Delay time (hours)	Problem	Cause	Device (Location)
A6	12/7/2004	?	2.75		2.75		Hardware replacement (SAN disk array controller)		SAN
A7	12/14/2004	?	2.42		2.42		Hardware replacement (SAN CPU and cache modules)		SAN

The counts of incidents by problem type are listed in Table 4-2. Table 4-2 shows that some incidents had multiple types of problems.

Table 4-2. Counts of Incidents by Problem Type

Problem Type	Count
Software	9
Human or procedure	7
Hardware	6
Interfaces (workstation-server communications or networking)	3
Database configuration	1

A more detailed classification of incident types can be found in Table 4-3.

Table 4-3. Classification of Incident Types

Incident #	Downtime (Hour)	CAD				RMS (H/W)	SAN (H/W)	SNA Gateway (H/W)	Workstation-Server communication	Human or procedure error
		Software	Hardware	DB Config	Admin					
(System went live and the acceptance test period started on 23 Sep 2003.)										
B1	0.23	x								Northrop Grumman
B2	0.08	x								
B3	0.28							x		Northrop Grumman
B4	0.45	x								
B5	0.25	x							x	
B6	0.12	x							x	
B7	0.62	x								
B8	0.25		x							
B9	4.38	x				x				
B10	0.98	x				x				
(System was accepted on 1/2/2004)										
A1	3.18			x						Northrop Grumman
A2	0.90	x								
A3	12.00				x					Northrop Grumman
A4	5.00						x			HP

Incident #	Downtime (Hour)	CAD				RMS (H/W)	SAN (H/W)	SNA Gateway (H/W)	Workstation-Server communication	Human or procedure error
		Software	Hardware	DB Config	Admin					
A5	8.00						x			HP
A6	2.75						x			
A7	2.42						x			

Incident B1 (9/24/2003). The outage was caused by an incompatible software upgrade and is not likely to occur again if configuration requirements are carefully processed.

Incident B2 (9/30/2003). The outage was a software bug in an analysis program that is not critical to call processing and dispatching. Portions of the program were temporarily disabled and are not likely to cause future outage.

Incident B3 (10/2/2003). The outage originated from a bad network card at the backup SNA gateway. This is regarded as a single point of failure. Unless fault isolation is considered, whether in the architecture or at the application level, this kind of outage may happen again.

Incident B4 (10/8/2003). The outage was caused by a system deadlock for database transactions. This was fixed with a code change.

Incidents B5 (11/5/2003) and B6 (11/7/2003) are the same kind of outage. CAD had more than 800,000 TCP packets pending transmission/retransmission from CAD to a remote workstation at 61 Riesner. This large amount of communications backlog caused CAD to go down. The resolution was to limit the amount of data that could be requested at one time from each workstation. Users needing large amounts of data would have to do queries outside of CAD; e.g., using SQL on database server. This problem should not occur again, but the root cause of CAD ability to operate when large communications backlog happens may still be a problem. A better understanding of capacity limits will help develop fault detection and performance monitoring capabilities.

Incident B7 (11/10/2003). The outage was caused by an archive logging process error. This problem should not occur again if the correct procedures are followed.

Incident B8 (11/16/2003). This was the only CAD hardware (memory module) failure. Reoccurrence is dependent on the hardware reliability.

Incidents B9 (11/28/2003) and B10 (12/3/2003). Both outages had the same symptom: incomplete transactions between RMS and CAD or the failure of RMS to report completed transactions caused the integrated database locked. Manual unlock was done by support contractors. Transaction process functions were examined and reengineered by Northrop Grumman in conjunction with Oracle. The root cause, bad memory modules in RMS, was identified, and all memory modules in RMS were replaced by HP. Reoccurrence of the problem is dependent on the hardware reliability.

Incident A1 (4/10/2004). The outage was caused by insufficiently allocated space in database, and it was compounded by an inexperienced DBA on site. Database space was expanded and a more experienced DBA is on site. This kind of problem is unlikely to happen again.

Incident A2 (4/25/2004). The outage was caused by a software bug (memory leak) in the CAD application. This problem was fixed.

Incident A3 (5/10/2004). Improper system administration (database backup) caused system outage for 12 hours. The contract system administrator has been replaced. This kind of problem is unlikely to happen again.

Incidents A4 (8/8/2004) and A5 (12/1/2004) were both SAN hardware problems, causing downtime 5 and 8 hours, respectively. This signified single-point-of-failure in the system architecture.

The last two outages on 12/7/2004 and 12/14/2004 were both preventive maintenance.

Table 4-3 also shows that the primary CAD (CADB), as the central component interfacing many devices, was vulnerable and thus its unavailability status caused some of the outages. Some incidents did not start from CAD directly, but they still caused CAD to also be unavailable. The system design should isolate CAD from being impacted by failures in other systems.

In addition to the outages, two other categories of problem resolutions were identified. This included problems identified as minor and new requirements. These 61 problems were documented in an SIRT (Software Incident Report Tracking) list and a change order list covering the period from October 31, 2003 to December 17, 2004. The SIRT list provides a description of each problem, estimated completion date, and resolution status. None of these problems were serious enough to cause system downtime on the primary system. Most of them require only system patches, documentation, or demonstration, while a few may need additional design. A summary of SIRT problem types is shown in Table 4-4.

4.2 System Availability Calculations

The scope of service covering the Northrop Grumman agreement specifies a requirement of 99.9% system availability for the CAD and RMS systems. Hardware failures are excluded from Northrop Grumman's availability calculations. MITRE recommends that all major systems meet or exceed 99.99 system availability. MITRE independently assessed the system availability based on universally accepted definition.

Table 4-4. Summary of SIRT Problems

Problem Type	Count
Data entry/recording/display	24
Communication or data transmission	10
Address/location verification	5
Application error	5
Database configuration or management	3
System startup or switchover	3
Data edit check	2
GUI bug	2
Additional data required	2
Documentation	2
Data error	1
Erroneous messages	1
OS update	1

System availability is defined as a system (consisting of hardware and software) is operating at any point in time, when subject to a sequence of "up" and "down" cycles. It addresses the question of "How likely will the system be available in a working condition when it is needed?" In this analysis, availability was evaluated by two standard measurements, operational availability and inherent availability. The availability of the overall system will be discussed first, followed by the computation for CAD, RMS, and SANs. There are two sets of availability calculations based on two alternative views of the starting point of the system life cycle: (1) starting from the system go-live date September 23, 2003; (2) starting from January 3, 2004. All availability calculation results are summarized in Table 4-5. The upper limits of availability for the 95% confidence level are shown in Table 4-6. The purpose is to provide an objective basis for setting reasonable expectations of the system availability. The percentages of uptime for individual months and days are presented in Tables 4-7 and 4-8. Some relevant concepts and definitions can be found in Appendix C.

4.2.1 Availability of the Overall System

The operational availability[5] of the overall system starting from go-live is:

[5] This is similar to the availability defined in Section J of Scope of Services: *CAD & RMS Acceptance Test Plans*, Page 9. But the downtime considered in this report is plain and general: whenever the system is not operational, caused by either hardware or software failure, users are experiencing downtime.

A_o = Total Uptime / Assessment Period = 1- Total Downtime / Assessment Period = 0.9965

The total downtime includes all corrective repair times, preventive maintenance times, and delay times caused by administrative and logistics processes.

The inherent availability of the overall system is:

A_i = MTBF / (MTBF + MTTR) = 0.9970, where MTBF is Mean-Time-Between-Failure and MTTR is Mean-Time-To-Repair.

Also known as Intrinsic Availability, the Inherent Availability A_i does not consider delay times and preventive maintenance times.

4.2.2 Availability of CAD/RMS

The CAD/RMS availability is derived from incidents caused by problems with CAD/RMS. Outages caused by other components of the system (e.g., SAN failures) are not included.

The calculation for the operational availability of CAD/RMS includes all outages except the last four that were caused by SAN problems. The operational availability of CAD/RMS since go-live is A_o = 0.9980

In computing the inherent availability of CAD/RMS, incident A3 is not included, since it was initiated by a system administration error that subsequently caused CAD to go down. The inherent availability of CAD/RMS since go-live is A_i = 0.9991

4.2.3 Availability Since Acceptance

If the system life cycle is considered to start from the day after the system acceptance date, as opposed to the system go-live date, then the start time of the assessment period is shifted to January 3, 2004, and the first 10 items in Table 4.1 are not counted against the availability calculation.

The operational availability of the overall system since acceptance is A_o = 0.9964

The operational availability after the acceptance is slightly worse than the operational availability previously calculated for the entire period since the go-live date. The analysis of the outages prior to the acceptance show that even though they were more frequent but were also much shorter (less than an hour), than those that occurred after the acceptance period. One explanation for the apparent difference in the recovery time is that both the system developer and the technical staff might have been more expeditious for problem resolution during the Acceptance Testing phase.

After the system was accepted, the system formally moved from testing to maintenance. The maintenance and support might be less agile than in the testing period. The records show certain degree of failure to meet contingency, which was also compounded by the deficiency in the skill set of the contractors. As a matter of fact, the majority of failures after the acceptance were either caused directly or aggravated by human errors. Based on the interviews, the MITRE team believes the maintenance team is now more experienced. It is reasonable to believe that the worst time is over; it is also a fair expectation to see reduced downtime in future outages.

There were only five outages after the acceptance. (A6 and A7 are outages for preventive maintenance.) The inherent availability of the overall system since acceptance is $A_i = 0.9970$

It is a coincidence that the inherent availability of the overall system after the acceptance has exactly the same four digits as the inherent availability before the acceptance.

Next, to calculate the availability for CAD/RMS, after the acceptance date, the three incidents (A1 to A3) are used to determine the operational availability calculation and the two incidents (A1 to A2) are used for the inherent availability calculation. Therefore, the estimated availability values for CAD/RMS after acceptance are $A_o = 0.9983$ and $A_i = 0.9997$

The CAD/RMS availability after acceptance has improved largely because all of the outages except one before the acceptance involved CAD/RMS, whereas after the acceptance, only less than half were related to CAD/RMS.

4.2.4 Result Summary of System Availability

A summary of all availability numbers computed in Sections 4.2.1 through 4.2.3 is presented in Table 4-5. As mentioned earlier, these statistical estimates are meant to provide a forward-looking view of the likelihood that the system will be available at any point in time. The percentages of uptime for individual months and days are shown in Tables 4-6 and 4-7.

Table 4-5. Results of System Availability With Different Assessment Periods

	Suppose the system life cycle started from the go-live date (23 Sep 2003)	
	Operational Availability	Inherent Availability
Overall system	0.9965	0.9970
CAD/RMS	0.9980	0.9991
	Suppose the system life cycle started after the acceptance date (3 Jan 2004)	
	Operational Availability	Inherent Availability
Overall system	0.9964	0.9970
CAD/RMS	0.9983	0.9997

These estimates indicate that inherently the CAD/RMS system looks promising for keeping up with the required 99.9% availability level, while the overall system may not achieve the same level of performance. Other parts of the overall system other than CAD/RMS have negatively impacted the overall availability. In order to provide uninterrupted services to end-users, a highly-available CAD/RMS system by itself is not enough, since the past incidents have already shown that it is susceptible to failures of other parts. Thus, it is recommended that efforts be focused on raising the availability of other parts of the overall system, in particular the SANs, and in making CAD/RMS more resilient to failures passed from these interfaces.

4.2.5 Confidence Level and Confidence Limit for Availability Estimates

Various expectations or industry norms for system availability may exist. This analysis calculates availability based entirely on empirical data associated with true events. Furthermore, standard statistical methods can also use the same set of empirical data to calculate the upper limit for system availability given a desirable confidence level.

Details of confidence limit calculation are provided in Appendix C. The results for a 95% confidence level are shown in Table 4-6, where A_o and A_i denote Operational Availability and Inherent Availability, respectively.

Table 4-6. Confidence Limits of Availability With 95% Confidence Level

	Assuming the system life cycle started from the go-live date (23 Sep 2003), the 95% confidence limits are:	
Overall system	$A_o < 0.9983$	$A_i < 0.9984$
CAD/RMS	$A_o < 0.9995$	$A_i < 0.9995$
	Assuming the system life cycle started after the acceptance date (3 Jan 2004), the 95% confidence limits are:	
Overall system	$A_o < 0.9984$	$A_i < 0.9990$
CAD/RMS	$A_o < 0.9998$	$A_i < 0.9999$

The entry "Overall system $A_o < 0.9983$" means the following: If no major improvement is to be made, we can predict with 95% confidence the operational availability of the HEC overall system will be less than 0.9983. That means the overall HEC system will suffer at least 14.8

hours of total operational downtime (both planned and unplanned) per year. Other entries have similar meaning.

These confidence limits indicate that it would not be realistic to expect the availability of the overall system to reach the 0.999 level. The CAD/RMS should achieve higher availability calculations but will probably not reach the recommended 99.99.

4.2.6 Monthly and Daily Availability

The concept of monthly and daily availability has been used by some organizations for checking against service level agreement. It no longer serves as an indication of the probability that the system is in a working condition but reports the percentage of system uptime during a calendar month/day. Dividing the continuous system operation into months and days will inevitably change the calculated results[6].

All monthly operational availability numbers after the acceptance date (3 January 2004) are shown in Table 4-7.

Table 4-7. Monthly Availability (Percentage of Uptime) After Acceptance

Month	Overall System	CAD/RMS	SAN
2004-Jan	100	100	100
2004-Feb	100	100	100
2004-Mar	100	100	100
2004-Apr	99.43	99.43	100
2004-May	98.39	98.39	100
2004-Jun	100	100	100
2004-Jul	100	100	100
2004-Aug	99.33	100	99.33
2004-Sep	100	100	100
2004-Oct	100	100	100
2004-Nov	100	100	100
2004-Dec	98.23	100	98.23
2005-Jan	100	100	100

[6] This is similar to the availability defined in Section J of Scope of Services: CAD & RMS Acceptance Test Plans, Plan 9. But the downtime considered in this report is plain and general: whenever the system is not operational, caused by either hardware or software failure, users are experiencing downtime.

After acceptance, all daily operational availability numbers are 100%, except for the following days in Table 4-8:

Table 4-8. Daily Availability (Percentage of Uptime) After Acceptance

Day	Overall System	CAD/RMS	SAN
2004-04-10	86.74	86.74	100
2004-04-25	96.25	96.25	100
2004-05-10	50.00	50.00	100
2004-08-08	79.17	100	79.17
2004-12-01	66.67	100	66.67
2004-12-07	88.54	100	88.54
2004-12-14	89.93	100	89.93

4.2.7 System Availability Enhancement

MITRE assessed methods to improve system availability given the less than desirable availability results of the overall system. In addition, while the CAD/RMS is close to meeting the 99.9 availability requirements, MITRE recommends an availability of 99.99. Thus, two methods to increase system availability were assessed. They included:

- Increasing reliability by acquiring more reliable components and also make service delivery more reliable. This method increases the MTBF.
- Increasing maintainability by performing repairs and maintenance work more efficiently and effectively. This method shortens MTTR.

In general, improving MTTR has better leverage than improving MTBF for increasing the system availability. The assumptions and formulas for this analysis are contained in Appendix C. Figure 4-1 shows the progressively estimated MTBF for the system calculated after each incident cycle. This chart indicates that the MTBF is getting better (longer) but is not yet reaching a steady state, thus, implying that the integrated HEC system has not passed the so-called “infant mortality” stage. As long as the MTBF continues to get longer, then the system reliability will continue to improve. A steady state will be achieved as the system matures.