Mathematical Formulation of the Red Storm 100Hr SOW Criteria

"MTBI for the full system, as determined by the need to reboot the system, shall be greater than 100 hours of continuous operation. This means that the system will be continuously operational for 100 hours with at least 99% of the system resources available and all disk storage accessible."

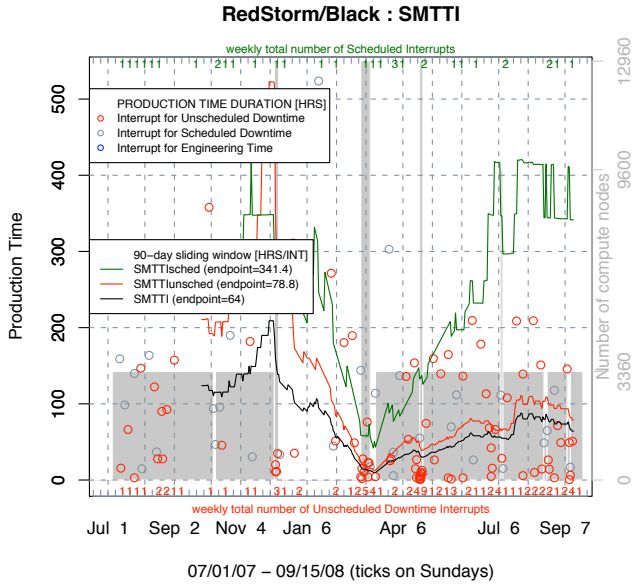
Based on incident report emails from the operations staff, Sandia tracks the Red and Black sides of the Red Storm system as being in exactly one of three states at all times:

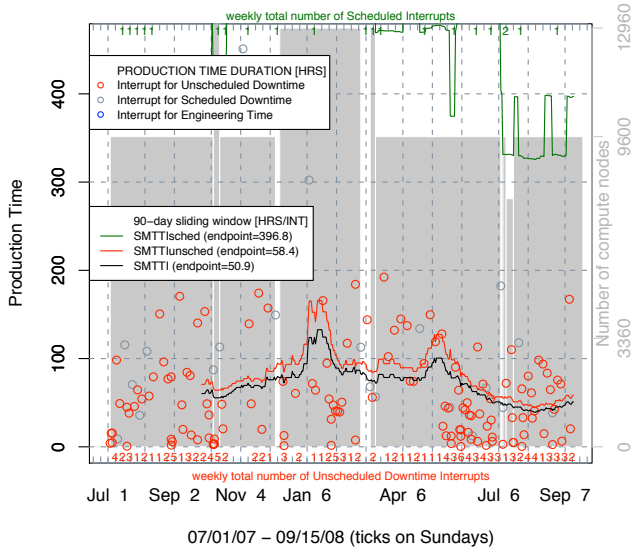
Unscheduled Downtime	**Scheduled Downtime**	**Production Uptime**
Repair Diagnosis Corrective action Verification Facilities Related	Preventative Maintenance Facilities Related Engineering Time - System Software, Hardware, or Process experiments or qualification	Production Computing

We define a system interrupt as *any* interruption in Production Uptime. The system mean time to interrupt (SMTTI) is calculated as the total number of hours spent in Production Uptime, divided by the total number of system interrupts. All metrics described herein are calculated using a 90-day sliding window.

$$SMTTI = \frac{ProductionUptime}{SystemInterrupts} \quad [HRS/INT] \qquad (1)$$

When this metric is calculated using only scheduled interrupts, we refer to it as SMTTIsched. When it is calculated using only unscheduled interrupts, we refer to it as SMTTIunsched. These metrics are tracked for the Red and Black sides. The per-side plots at right also show the duration of Production Uptime periods (circles), and the number of compute nodes present (grey background).

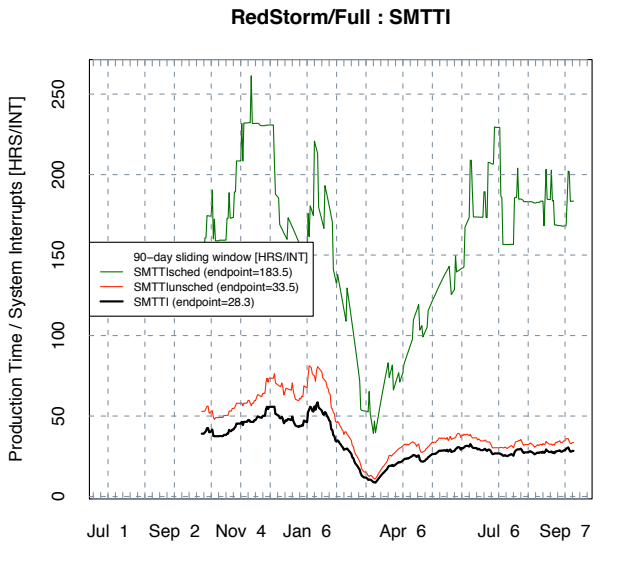




In order to measure the 100Hr criteria for the full system, the Red and Black sides are modeled as being in series. The failure rate (INT/HR) of a series system is equal to the sum of the failure rates of the components, and it follows that:

$$\frac{1}{fullSMTTI} = \frac{1}{redSMTTI} + \frac{1}{blackSMTTI} \qquad (2)$$

The above formulation for fullSMTTI is being considered for use as Sandia's official measurement of the 100Hr SOW criteria (see bold black line at right).



Mathematical Formulation of the Red Storm 40Hr SOW Criteria

"Mean time between Interrupt (MTBI) for full system shall be greater than 50 hours for continuous operation of the full system on a single application code. This means that the full system must be able to run continuously on an application using the full system for 50 hours without any hardware component failures or system software failures that cause an interrupt or failure of the application code. (In this context "full system" means the maximum configuration for the compute partition plus one service and I/O partition.)"

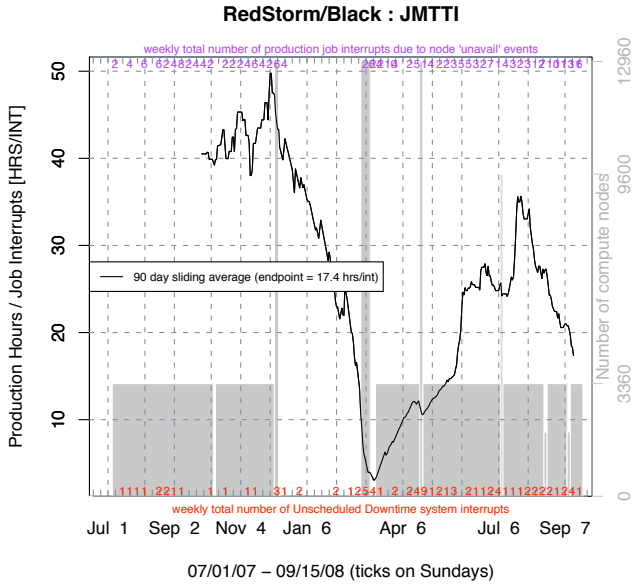
This was reduced to 40Hrs upon the addition of the 5th row of cabinets.

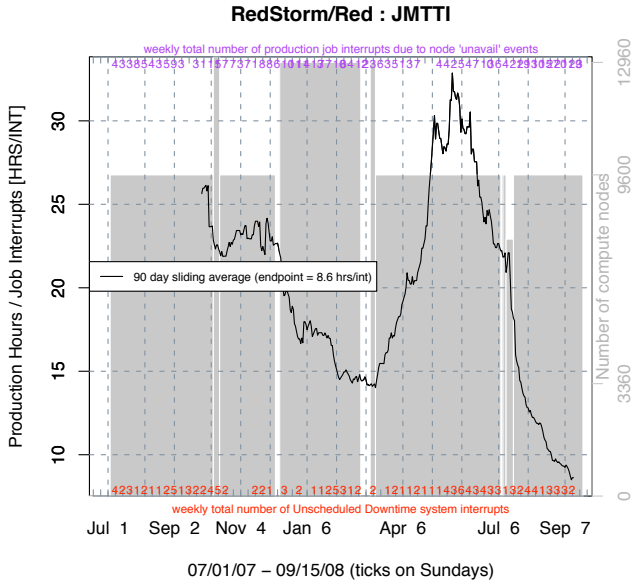
Each time the system transitions from Production Uptime to Unscheduled Downtime, a single interrupt is counted towards this metric (see "Mathematical Formulation of the Red Storm 100Hr SOW Criteria" for details). In addition, a single interrupt is counted for each job killed as a result of one or more nodes becoming unavailable while running that job. We define these to be Job Interrupts.

The job mean time to interrupt is calculated as the total number of hours spent in Production Uptime divided by the total number of job interrupts. All metrics described herein are calculated using a 90-day sliding window.

$$JMTTI = \frac{ProductionUptime}{JobInterrupts} \quad [HRS/INT] \qquad (1)$$

This is tracked for the Red and Black sides of Red Storm. The per-side plots at right also show the total number of each type of job interrupt (red and purple annotations), and the number of compute nodes present (grey background).

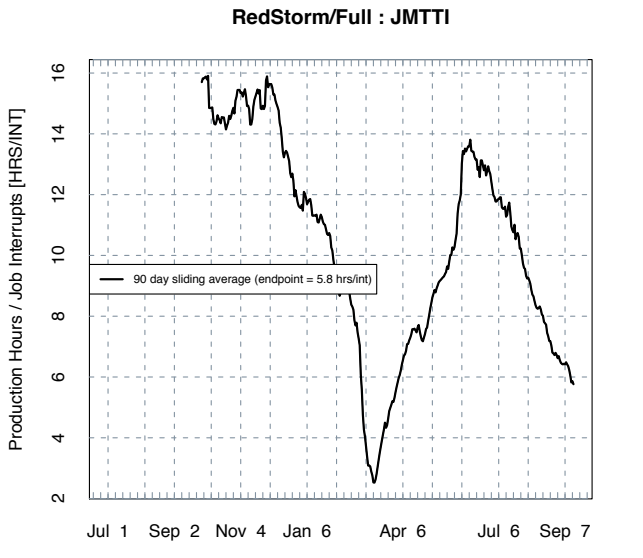


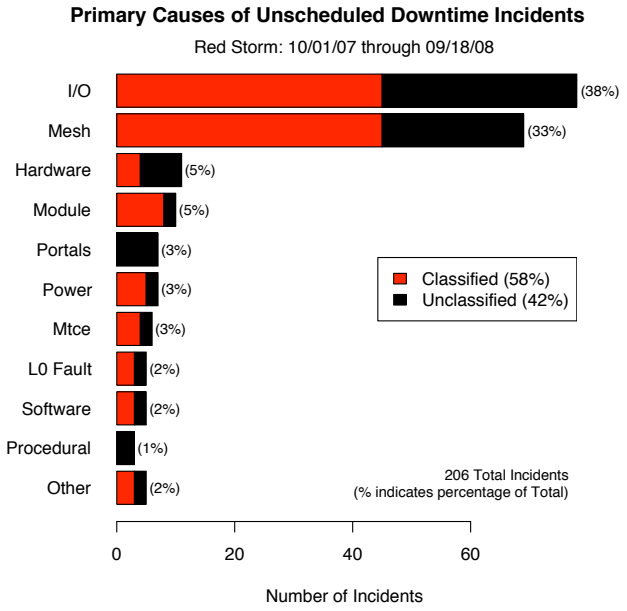


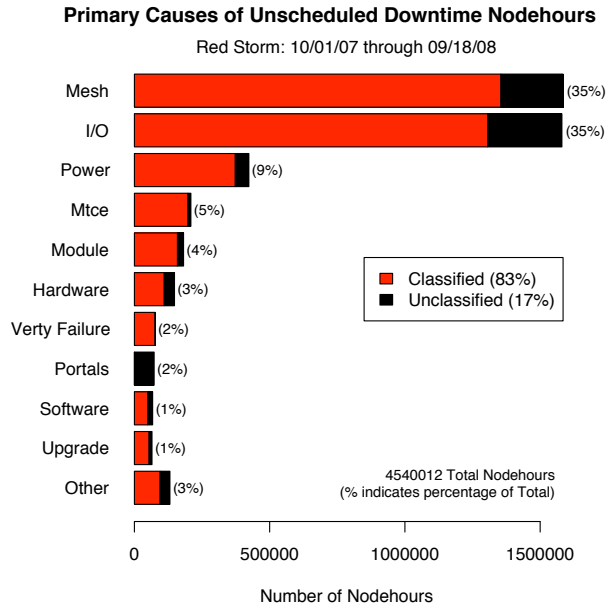
In order to measure the 40Hr criteria for the full system, the Red and Black sides are modeled as being in series. The failure rate (INT/HR) of a series system is equal to the sum of the failure rates of the components, and it follows that:

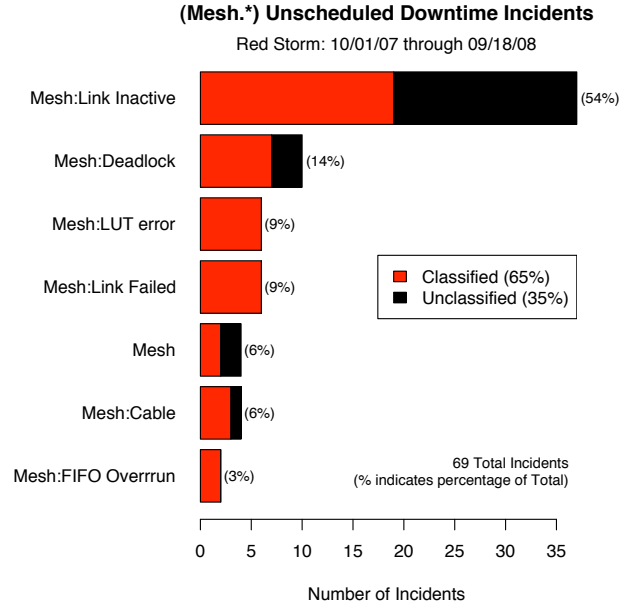
$$\frac{1}{fullJMTTI} = \frac{1}{redJMTTI} + \frac{1}{blackJMTTI} \qquad (2)$$

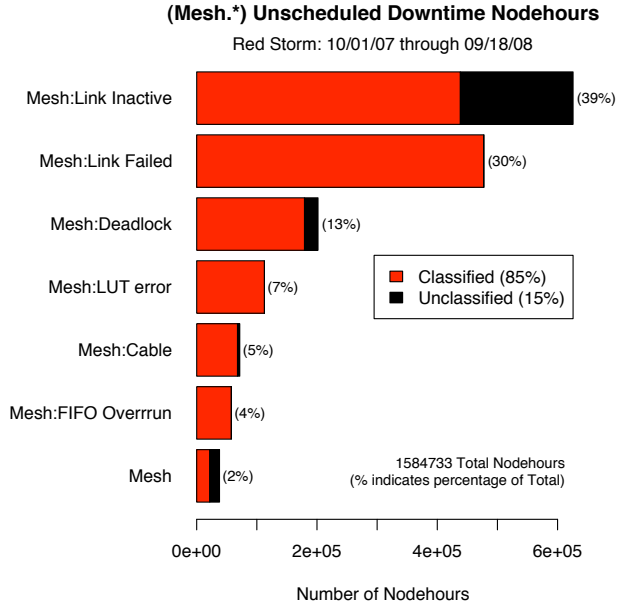
The above formulation for fullJMTTI is being considered for use as Sandia's official measurement of the 40Hr SOW criteria (see bold black line at right).

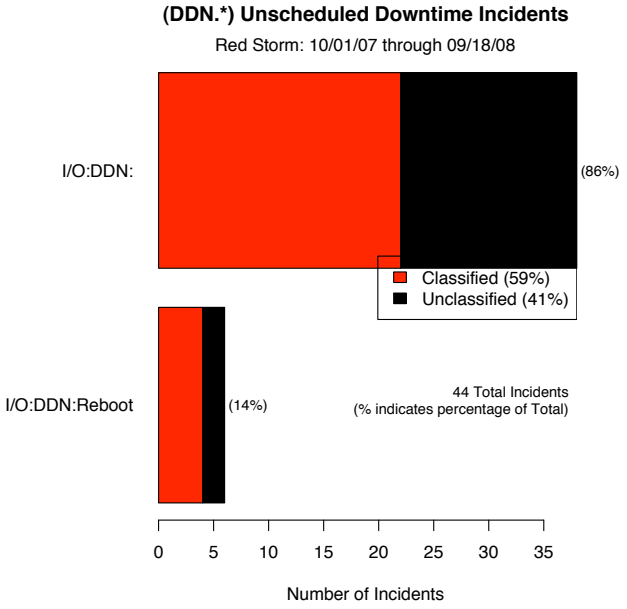


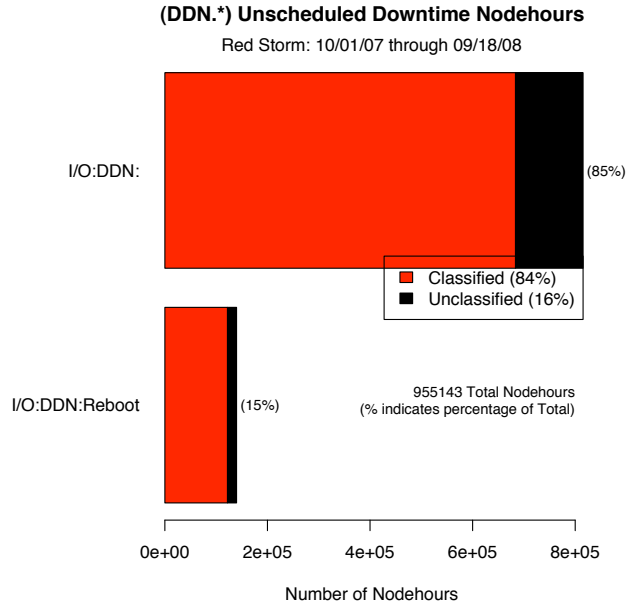

Primary Causes of Unscheduled Downtime Incidents
Red Storm: 10/01/07 through 09/18/08
I/O
Mesh
Hardware
Module
Portals
Power
Mtce
L0 Fault
Software
Procedural
Other
(38%)
(33%)
(5%)
(5%)
(3%)
(3%)
(3%)
(2%)
(2%)
(1%)
(2%)
Classified (58%)
Unclassified (42%)
206 Total Incidents
(% indicates percentage of Total)
0
20
40
60
Number of Incidents


Primary Causes of Unscheduled Downtime Nodehours
Red Storm: 10/01/07 through 09/18/08
Mesh
I/O
Power
Mtce
Module
Hardware
Verty Failure
Portals
Software
Upgrade
Other
(35%)
(35%)
(9%)
(5%)
(4%)
(3%)
(2%)
(2%)
(1%)
(1%)
(3%)
Classified (83%)
Unclassified (17%)
4540012 Total Nodehours
(% indicates percentage of Total)
0
500000
1000000
1500000
Number of Nodehours


(Mesh.*) Unscheduled Downtime Incidents
Red Storm: 10/01/07 through 09/18/08
Mesh:Link Inactive
Mesh:Deadlock
Mesh:LUT error
Mesh:Link Failed
Mesh
Mesh:Cable
Mesh:FIFO Overrrun
(54%)
(14%)
(9%)
(9%)
(6%)
(6%)
(3%)
Classified (65%)
Unclassified (35%)
69 Total Incidents
(% indicates percentage of Total)
0
5
10
15
20
25
30
35
Number of Incidents


(Mesh.*) Unscheduled Downtime Nodehours
Red Storm: 10/01/07 through 09/18/08
Mesh:Link Inactive
Mesh:Link Failed
Mesh:Deadlock
Mesh:LUT error
Mesh:Cable
Mesh:FIFO Overrrun
Mesh
(39%)
(30%)
(13%)
(7%)
(5%)
(4%)
(2%)
Classified (85%)
Unclassified (15%)
1584733 Total Nodehours
(% indicates percentage of Total)
0e+00
2e+05
4e+05
6e+05
Number of Nodehours


(DDN.*) Unscheduled Downtime Incidents
Red Storm: 10/01/07 through 09/18/08
I/O:DDN:
I/O:DDN:Reboot
(86%)
(14%)
Classified (59%)
Unclassified (41%)
44 Total Incidents
(% indicates percentage of Total)
0
5
10
15
20
25
30
35
Number of Incidents


(DDN.*) Unscheduled Downtime Nodehours
Red Storm: 10/01/07 through 09/18/08
I/O:DDN:
I/O:DDN:Reboot
(85%)
(15%)
Classified (84%)
Unclassified (16%)
955143 Total Nodehours
(% indicates percentage of Total)
0e+00
2e+05
4e+05
6e+05
8e+05
Number of Nodehours