Correlation of Ground-Level Computer Installation Error Rates with Solar Activity

CSC 2232 Project (Fall 2009)

Eric LaForest, January 06 2010

Problem Statement

Is the soft error rate in ground-level computer installations correlated with the intensity of solar radiation? This problem is already significant for space-borne systems, but recent analysis of FPGA configuration failures show that radiation-induced soft errors might become a problem even at sea level as silicon device integration increases [1]. This problem would be magnified in large or widely deployed installations.

Hypothesis

It is expected that as solar activity increases, the likelyhood of radiation-based uncorrectable errors should increase, manifested either as software errors, uncorrectable memory errors, or other temporary hardware failures.

Methodology

There are failure data for a number of large computer cluster systems at Los Alamos National Laboratory (LANL) over the 1996-2005 period [6], which happens to roughly coincide with Solar Cycle 23 (1996-2008) [2].

Computer Systems

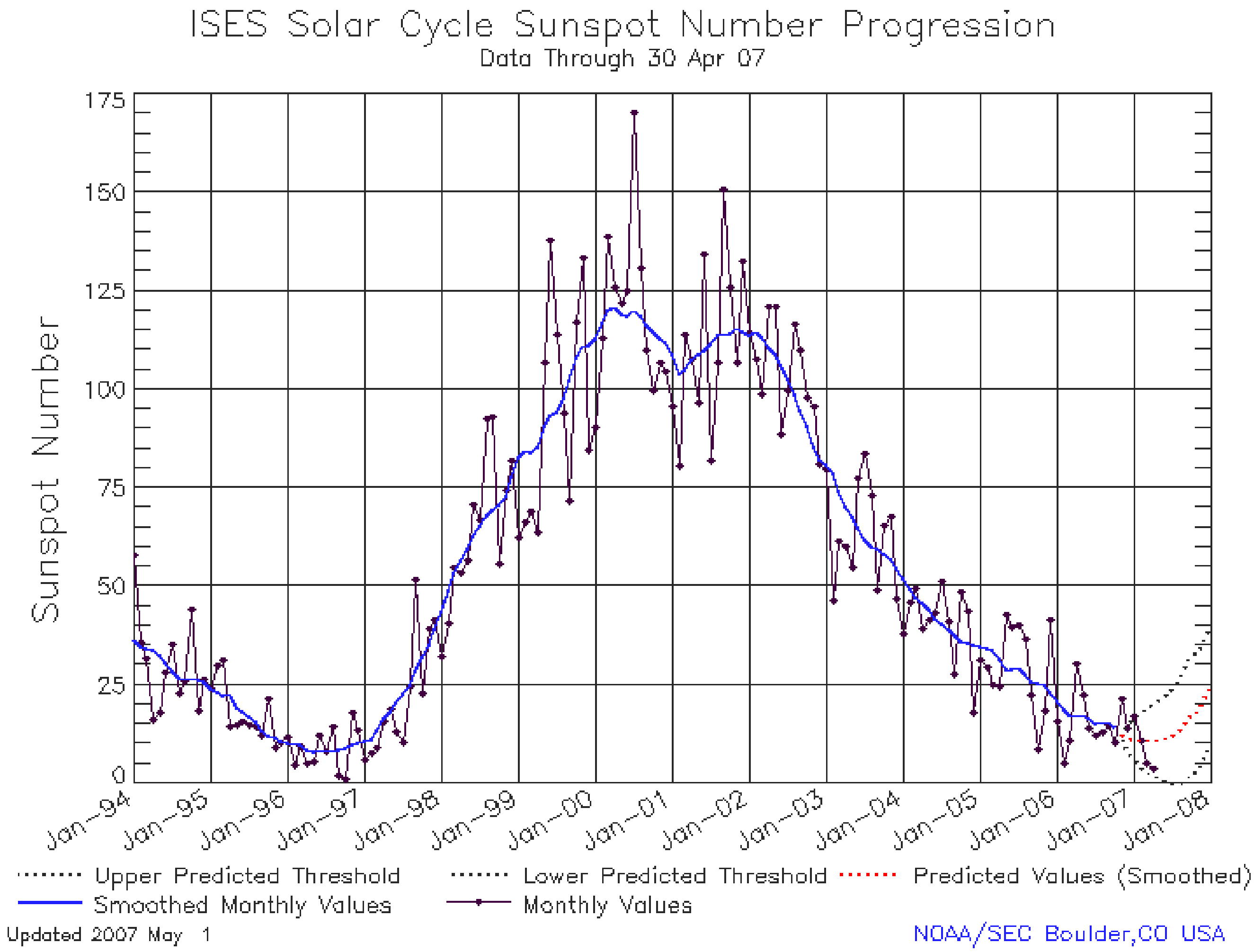
The large computers installations at LANL potentially form detectors of solar activity, as measured by the occurrence of random failures during their operation. The failures are logged separately for each system and further classified by their cause. I focused on the most prevalent causes of failure, which could also be conceivably caused by radiation:

- Hardware errors (CPU errors, uncorrectable RAM errors)
- Software errors (crashes and other malfunctions)
- Unknown errors (cause was not determined or accurately logged)

A summary of the size and structure of each installation is included in Appendix A, ordered by the machine number used in the failure logs (second column), which are the identifying numbers used throughout this report.

Solar Activity

I use the daily sunspot count as a gross measure of the level of solar activity [3]. The activity of the Sun varies from low to high and back over the span of a solar cycle. Cycles 23 approximately peaked in March 2001. I obtained the daily sunspot count from the Solar Influences Data Analysis Centre [4][5]. The following chart illustrates the change in (monthly) sunspot counts over Cycle 23 [8].



I also attempted to use a daily neutron count, but the data was scattered and spotty, difficult to relate to solar activity, and varied considerably depending on the latitude of the measuring station. There was no suitable data set of neutron measurements collected near Los Alamos or its latitude.

Difficulties

I had originally planned to simply plot the total errors per day alongside the sunspot counts, expecting to see an increase in errors as the solar cycle reached its maximum, and a decrease afterwards. Unfortunately, most of the computer systems were put into production in late 2003, long after the solar maximum had passed. Therefore, it would not be possible to separate the natural decrease in failures over time in a new system with the decrease in errors caused by a decrease in solar activity.

Additionally, there are many more days with low sunspot counts than high counts, so even though their effect should be smaller, it would be overrepresented in the data, further obscuring the impact of high solar activity.

Finally, most days log few or no failures and are thus overrepresented when related to the sunspot count since the count is not evenly distributed, but decreases slowly on average after March 2001. Thus days with little to no failures will have a higher average sunspot count than days with many failures, occulting any straightforward correlation that would have been visible if the average sunspot count had been constant over the measurement interval.

Analysis

To avoid these obscuring problems, as well as to aggregate the data in a meaningful manner, I grouped the failures by the number of sunspots that day, summed them together, then divided by the number of days which had that number of sunspots. The result is an average number of failures per day for days with a given number of sunspots. Furthermore, to avoid tainting the data with failures naturally occuring during the early life of a new computer system, the first three months of data were discarded before the analysis.

I had attempted to also include the effect of workloads on the failure rate, since it is conceivable that a system with more memory in use will be likelier to experience a failure from an uncorrectable memory error. However, the workload data is often incomplete (ie: only the first half of the nodes in System 20 have such data) and complicated to parse as workloads last anywhere from a few seconds to a few days, and may use only a fraction of all the processors in a system without indication of memory usage. It would have also been useful to separate the failures by workload, as each user ID seems to run the same workload repeatedly, based on run time. Unfortunately, I lacked the time to include these analyses.

Results

Out of 23 systems, 15 showed a positive correlation between the average rate of failure and the daily sunspot count. The remaining eight systems had very weak or slightly negative correlations, mostly due to lack of data, insufficient instrinsic reliability, or possible hardware/software artifacts.

The results are shown as charts plotting the average number of failures per day for all days with a given sunspot count, arranged in increasing order. Each bar denotes the failure rates, and is divided into the recorded causes: Hardware, Software, or Unknown. The black dotted line shows a linear regression fit to the failure rates as a further qualitative check on correlation.

Variations in Data

The number of bars in each chart varies greatly since the total number of failures experienced by a system, after filtering for cause, can vary between 51 (System 15) to 6614 (System 2). Although the analysis of some systems fails from lack of data (Systems 15 and 21), a large number of failures is not necessary to obtain a positive correlation (see Systems 6 and 7) nor a guarantee of one (System 2). The actual number of failures analyzed is actually smaller since the first three months of filtered data are discarded to exclude failures caused by 'infant mortality'.

Much of the variation in the amount and type of failures depends on logging procedures. For example, System 6 seems unusually reliable relative to other systems, which suggests underreporting. The size of the system also has an effect: Systems 7, 22, and 24 were single SMP computers, and hence are naturally more reliable than a cluster of machines.

Intrinsic System Reliability

Some systems seem to suffer so many failures on average that any external effect from solar activity is blotted out. This is apparent in Systems 2, 18, and 19, which exhibit a base daily failure rate of two to three according to linear regression.

This observation is supported in several ways:

- System 2 has been in operation throughout the entirety of Solar Cycle 23, and so any bias due to 'infant mortality' during the downward phase of the cycle should be absent.

- Systems 18 and 19 have begun operating during the downward half of Cycle 23, like most other systems, yet they do not show a positive correlation either through coincidence or genuine solar effect.

- The base daily failure rate of the systems exhibiting positive correlation is less than one, regardless of the total number of failures, only reaching a value of one in a single case (System 16). Only half of the systems without a correlation have base failure rate less than one, two of which (System 9 and 11) exhibit suspecte artifact, one (System 15) has insufficient data, and one (System 14) is unexplained, but generally seems weak.

Unexplained Differences

Systems 9, 10, and 11 appear to be identical systems put in operation at the same time, with similar numbers and types of failures recorded, yet only System 10 exhibits a positive correlation. It is unclear as to why, but I suspect that all three systems were affected by some increase in failures late in the solar cycle, which would manifest as the large cluster of high failure rates at low sunspot counts which is generally not present in any other charts. I suspect a software or hardware configuration change after installation, but lack the time to plot the failures as a time series to find out. Another way to confirm this fact would be to look at the workloads of System 13, which appears to have identical hardware to Systems 9, 10, and 11, but exhibits the expected positive correlation.

Systems 12 and 20 seem similar, with System 20 having twice as many processors and four times the RAM, but also exhibiting 10 times more errors over only about twice the period of operation. Since System 20 is two years older and has more RAM, this suggests more surface area to interact with radiation. If that is the case, Systems 18 and 19 should have exhibited similar behaviour, but their unreliability masks the effect.

Probability of Failure

Since there are many more days with lower sunspot numbers, there are more opportunities to record failures than on the rarer, high sunspot count days. This manifests on the charts as an increasing scarcity of days with any failures at all as the sunspot count increases. However, where there is a positive correlation the average rate of failure, when present, is higher on days with high sunspot counts despite being few and far between. It is this change in failure rate that the linear regression line makes explicit.

Many of the charts show a few seemingly outlying bars of high failure rates at high sunspot counts. These may appear to be artifacts, but are instead a representation of the higher probability of failure during higher solar activity. There might have been only one day with a given high sunspot count, but that day had a failure and thus shows a failure rate of one. Compare to this the many days with lower sunspot counts, most of which experience no failures, resulting in failure rate of less than one. For example, Systems 3, 4, and 5 illustrate this effect.

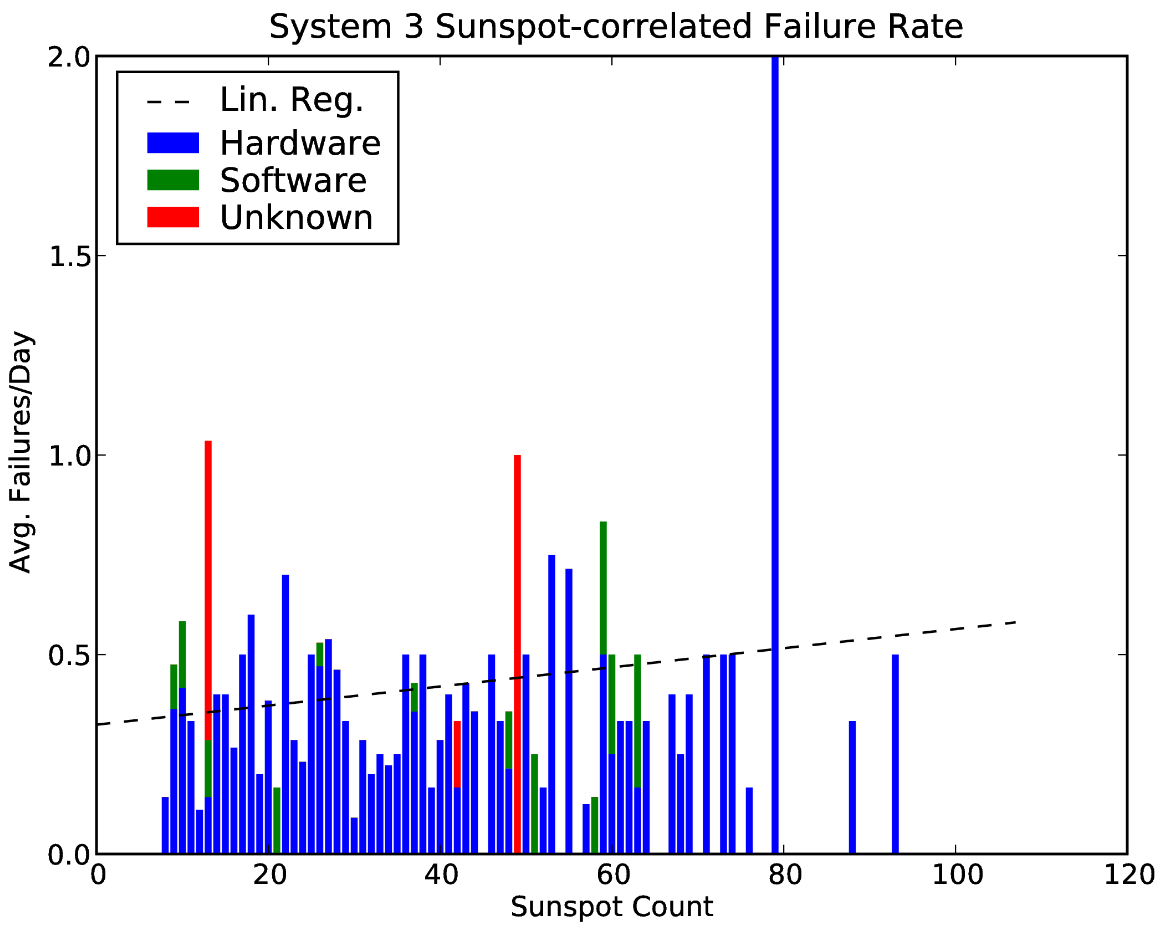
Simply put, days with high sunspot counts are rare, but a computer system is more likely to experience a failure during those days, which is the correlation I hoped to see in this analysis.

Systems with positive correlation

System 3

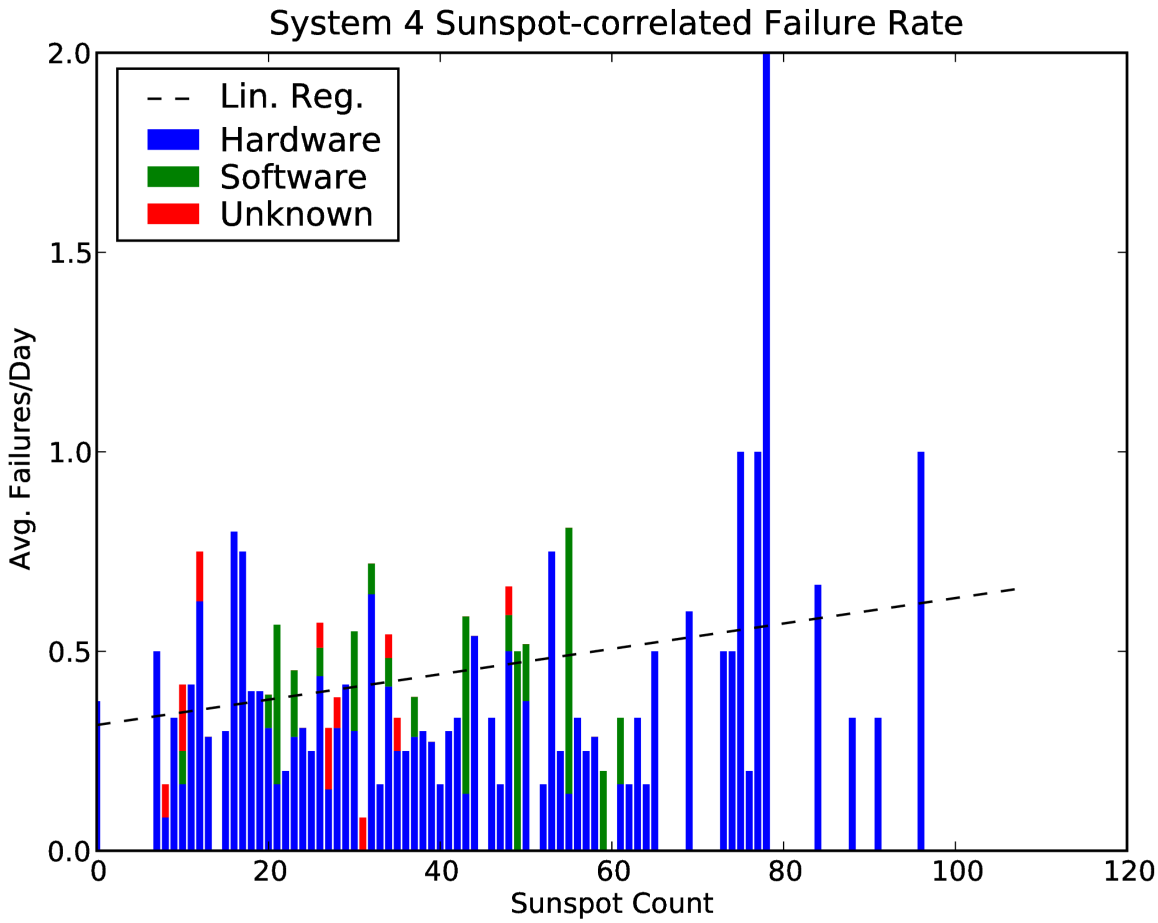
293 failures.

Systems 3, 4, and 5 have identical hardware configurations and all show the increased likelyhood of a hardware error on high sunspot count days.



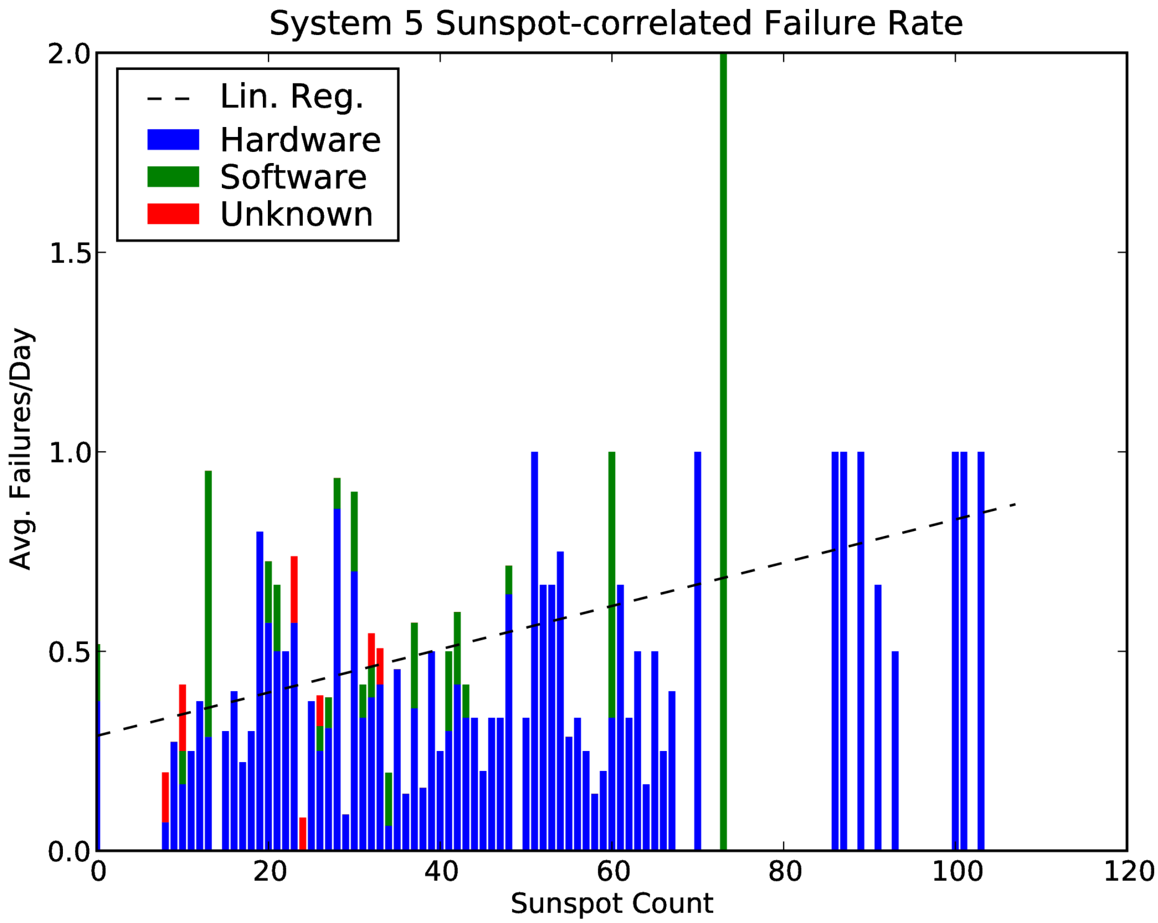
System 4

291 failures.



System 5

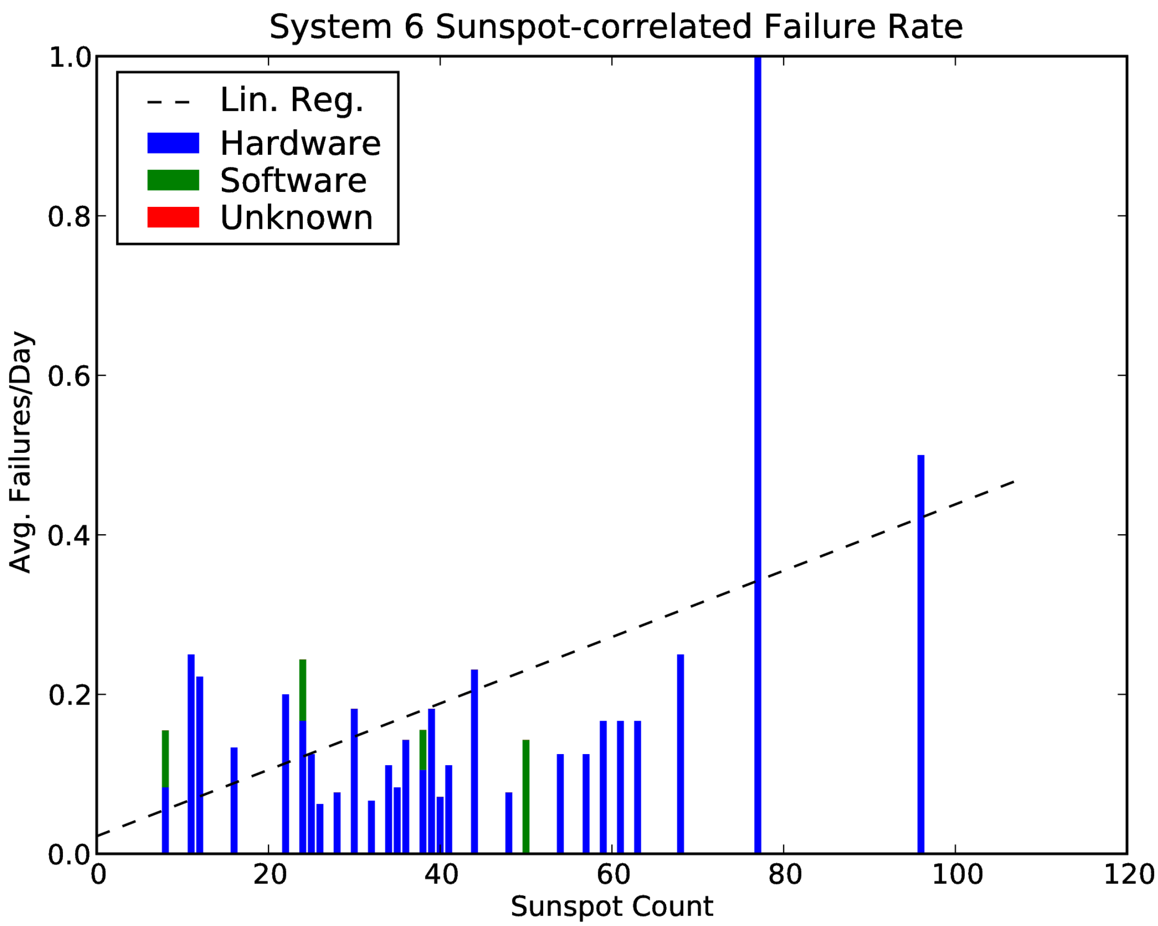
297 failures.



System 6

59 failures.

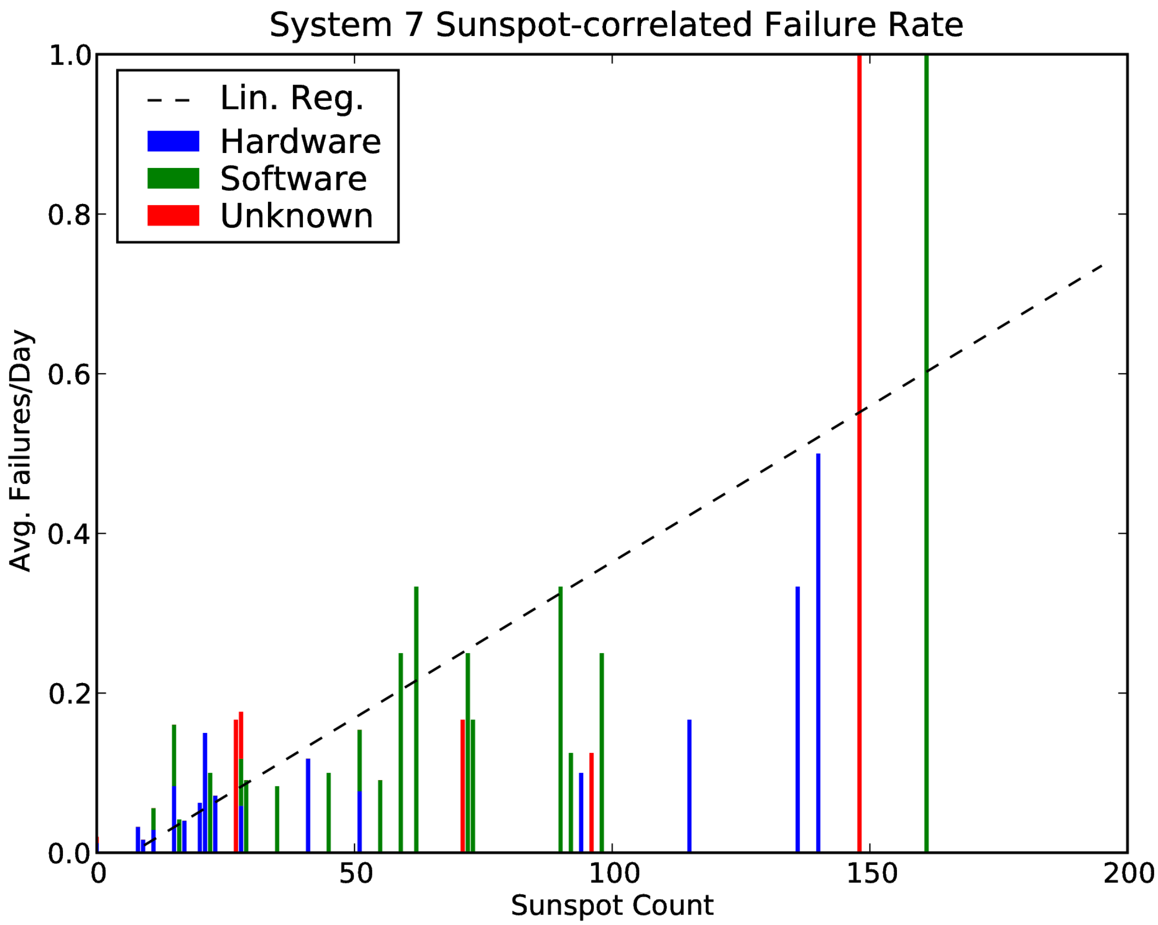
This system appears unusually reliable, likely due to how its failures were reported. Nonetheless, the increase in hardware errors is still visible.



System 7

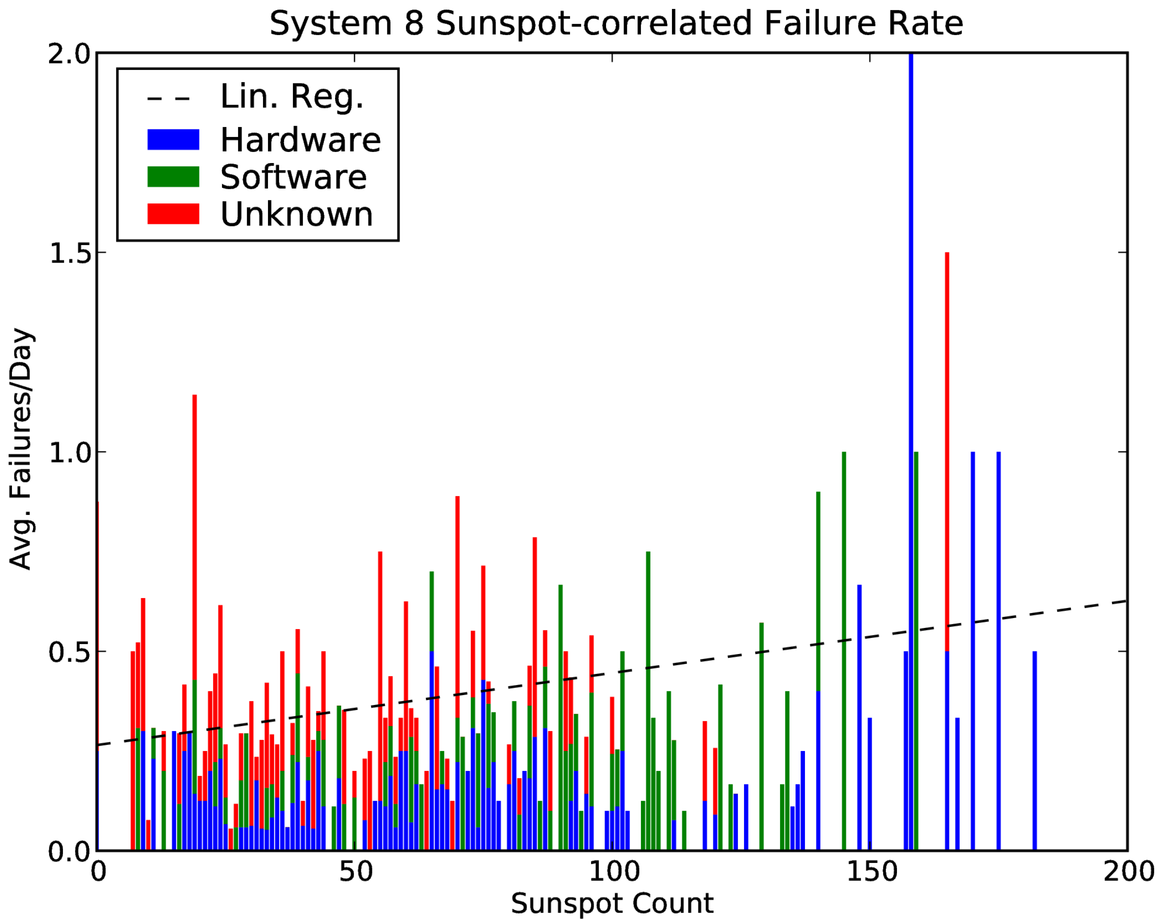
79 failures.

The low failure count is likely due to the fact that this system is a single SMP machine instead of a cluster. There are few hardware errors, and mostly software errors which do correlate with sunspot count.



System 8

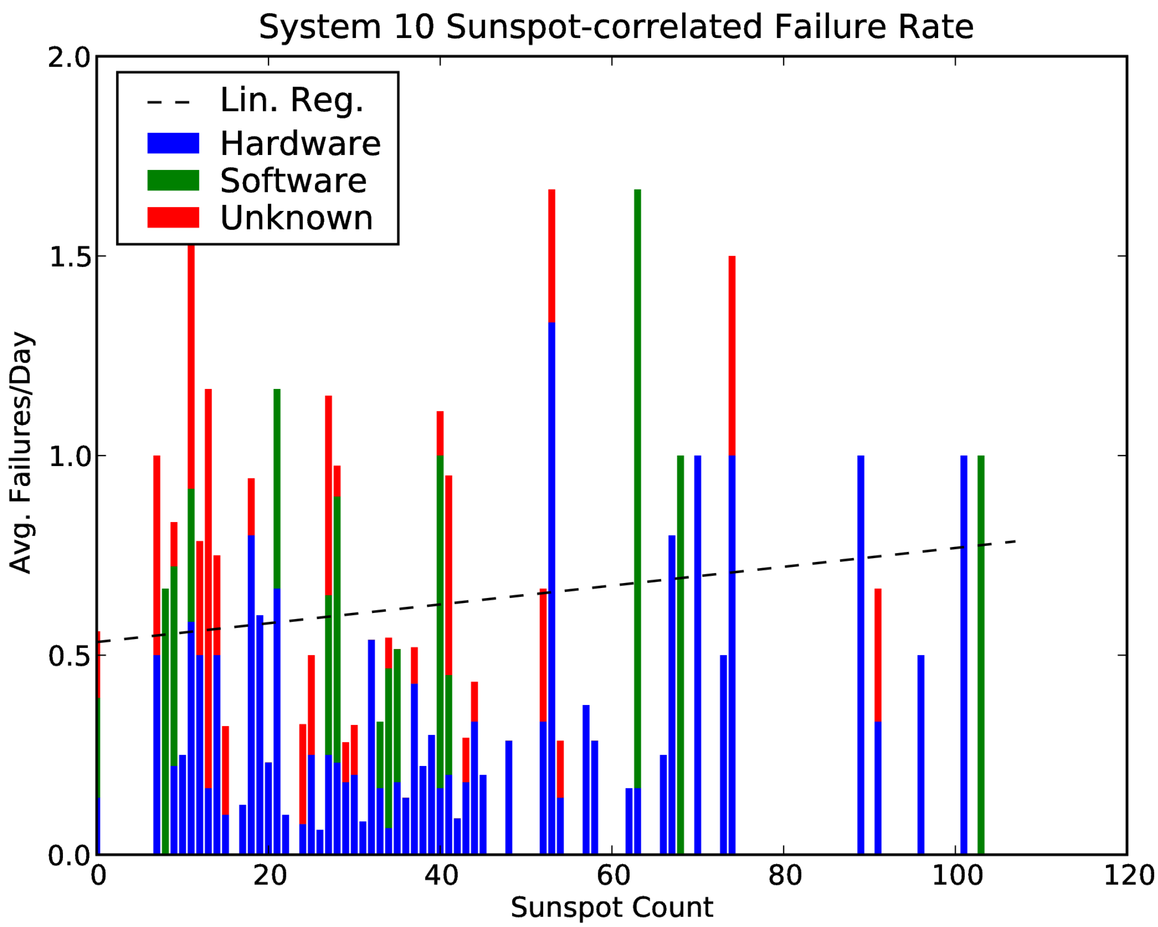
438 failures.



System 10

231 failures.

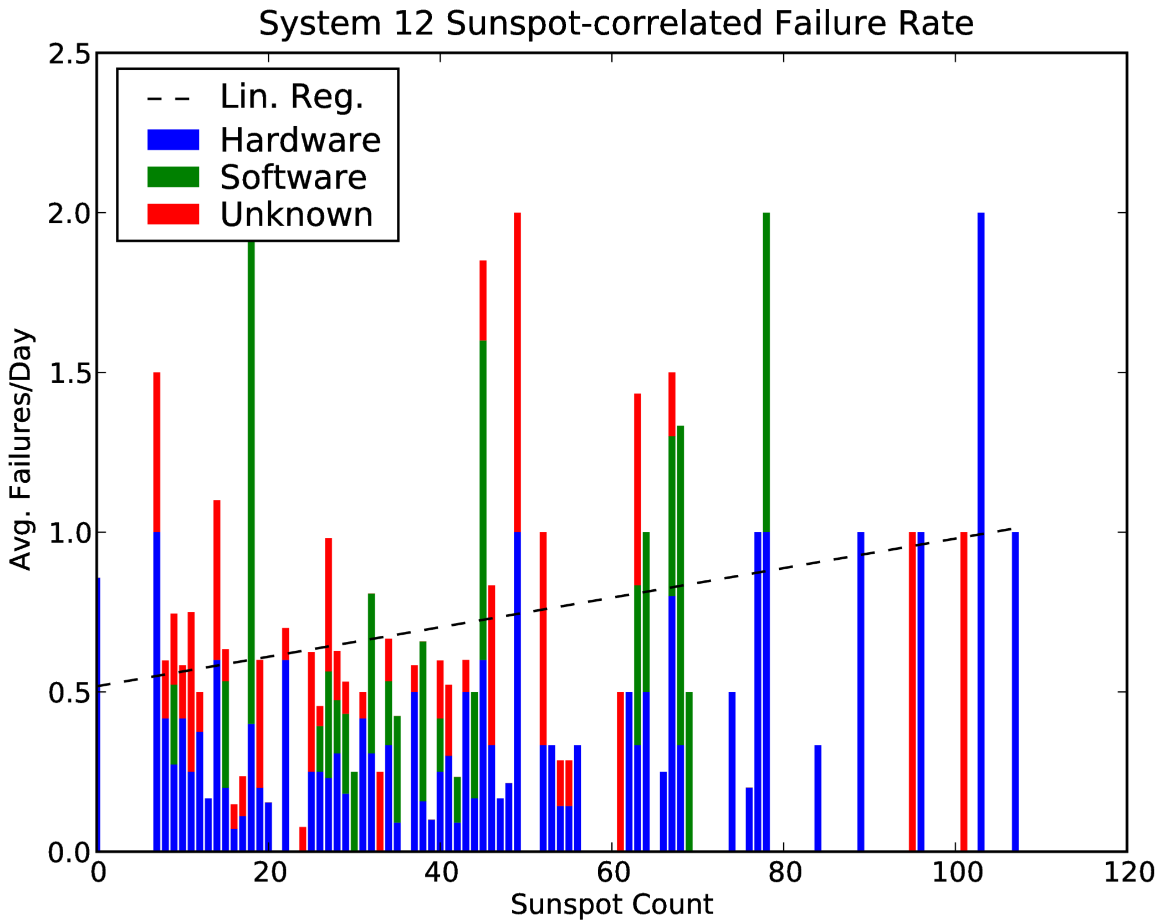
Compare to Systems 9 and 11, which do not exhibit a positive correlation. All three systems seem to have experienced a surge in failures late in the solar cycle, obscuring the effect of higher sunspot counts.



System 12

250 failures.

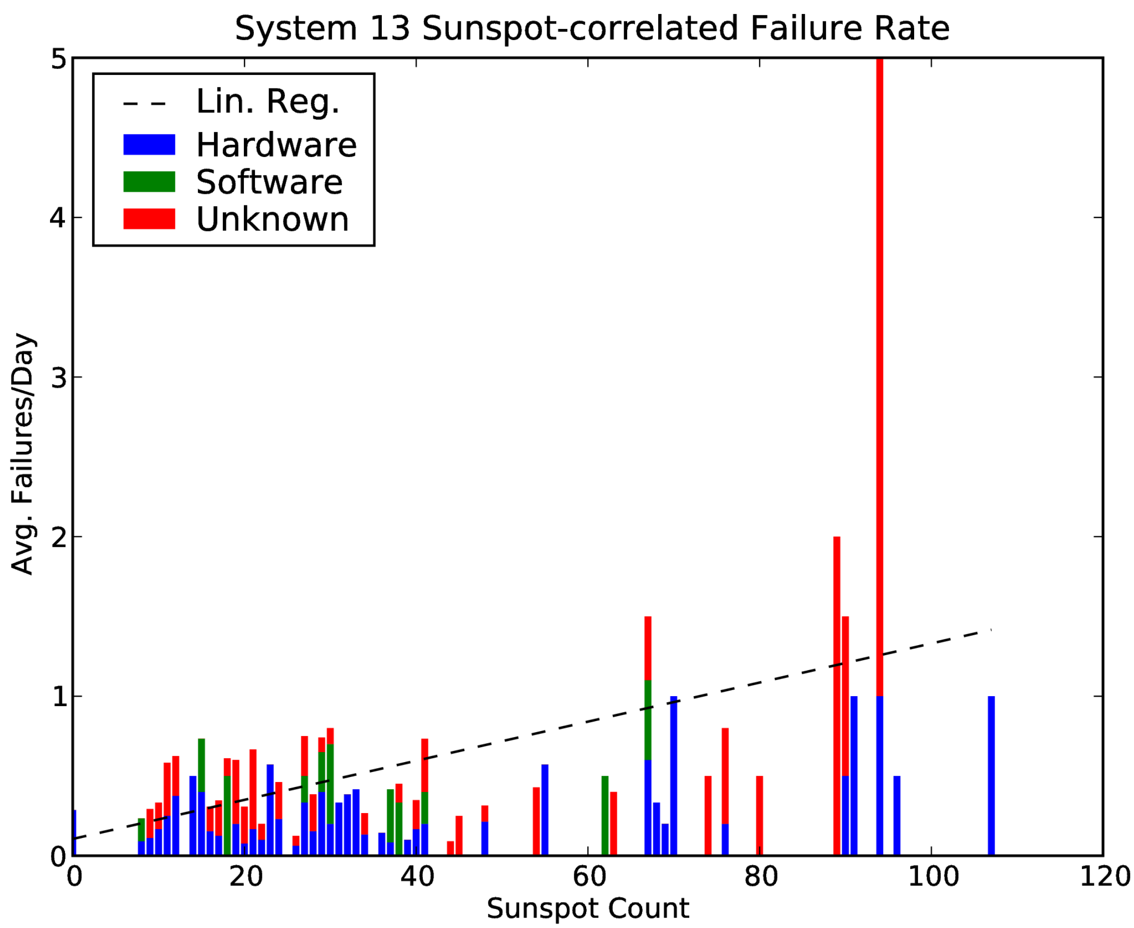
Compare with System 20, which is similar, but larger and older.



System 13

194 failures.

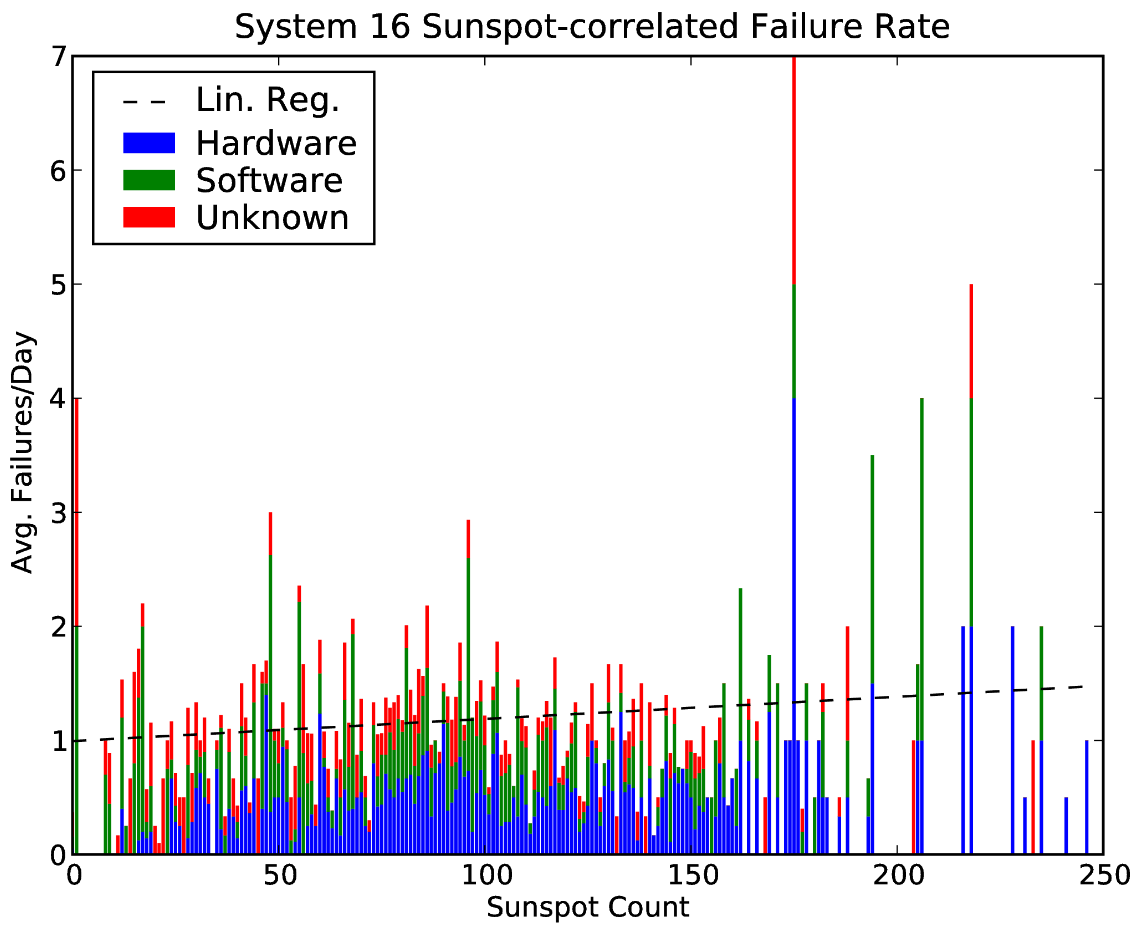
Identical hardware to Systems 9, 10, and 11, but might have run different workloads.



System 16

2262 failures.

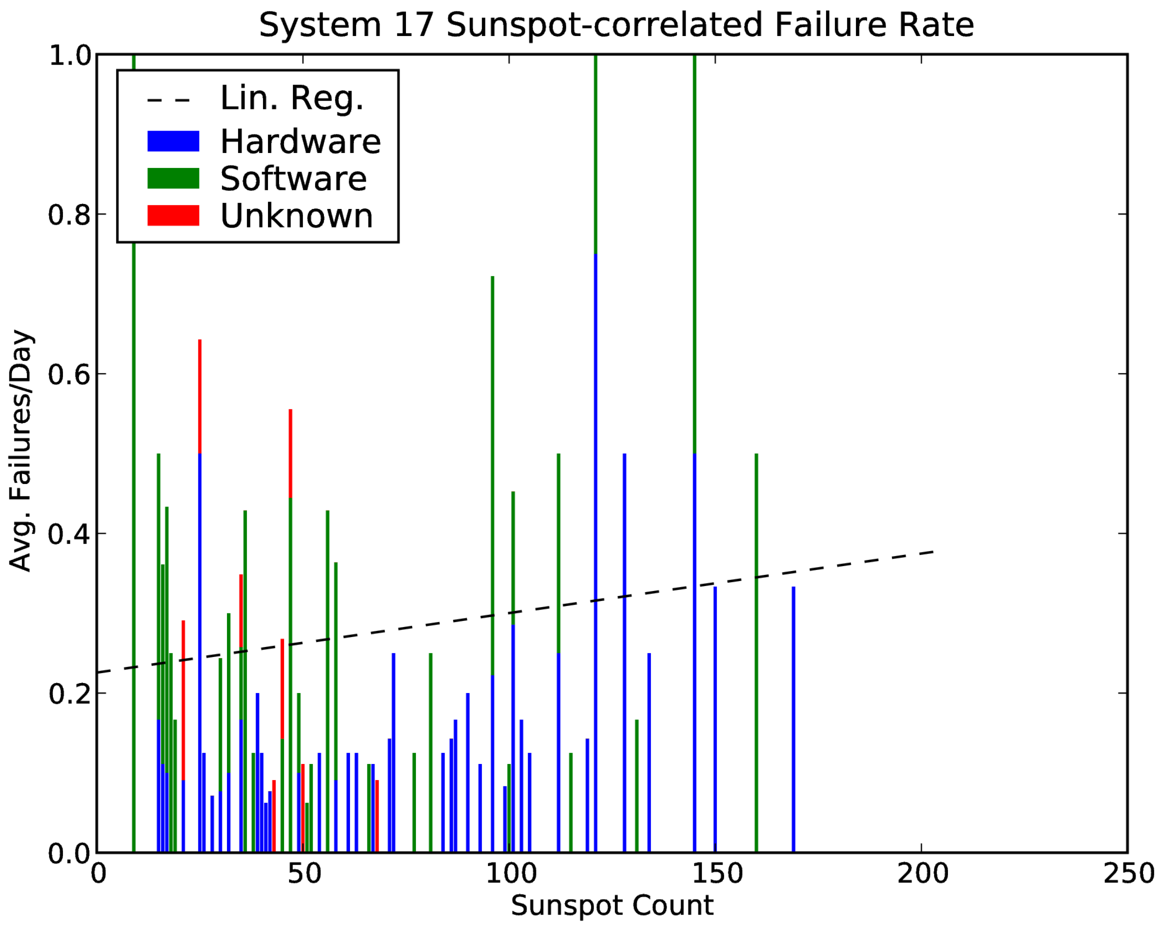
This was an older cluster system with few nodes equipped with many processors and large amounts of memory. The data suggests that even without the effect of solar activity, this system had an average of one failure per day, which is at least twice as much as the other systems.



System 17

120 failures.

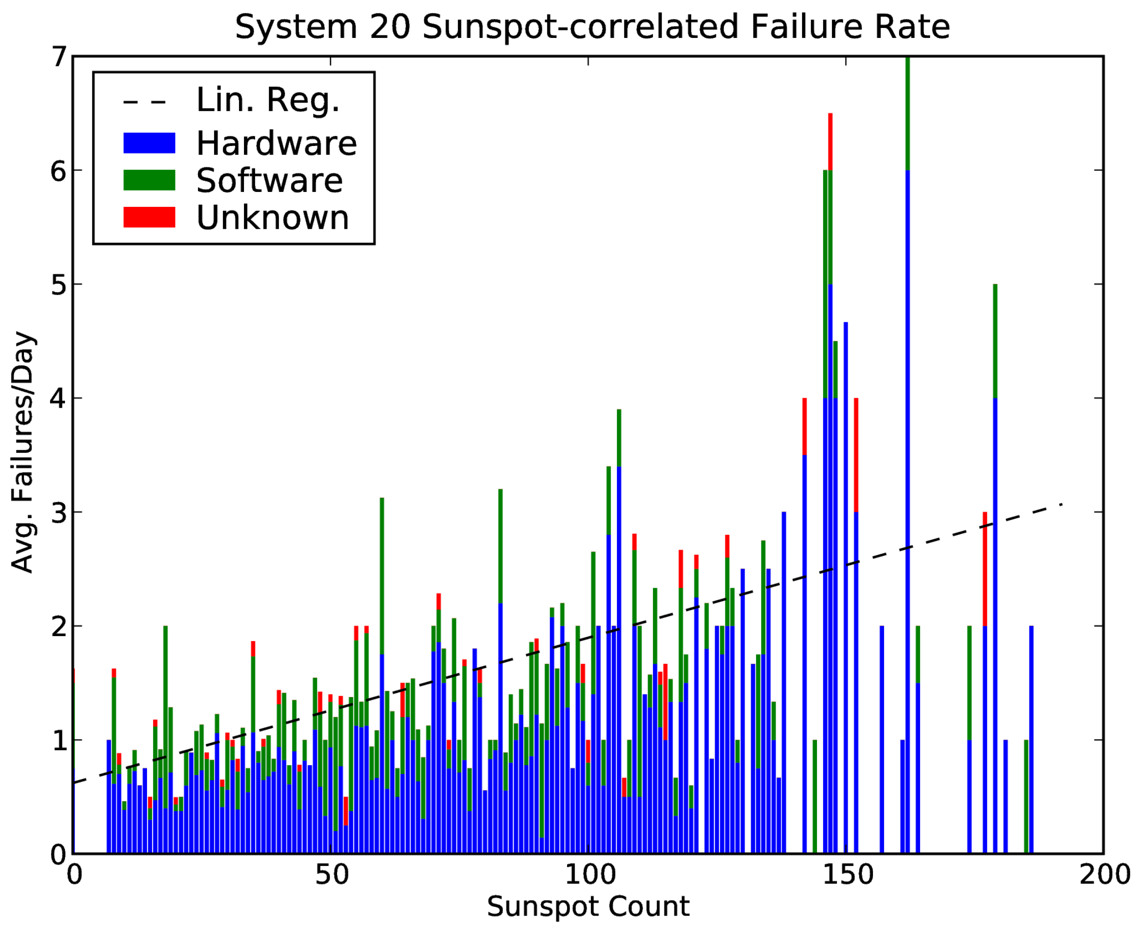
This system is not listed in Appendix A since its failure log entries were incomplete. No data are available about its configuration or service life. The reported failures span from January 1997 to April 2000.



System 20

2439 failures.

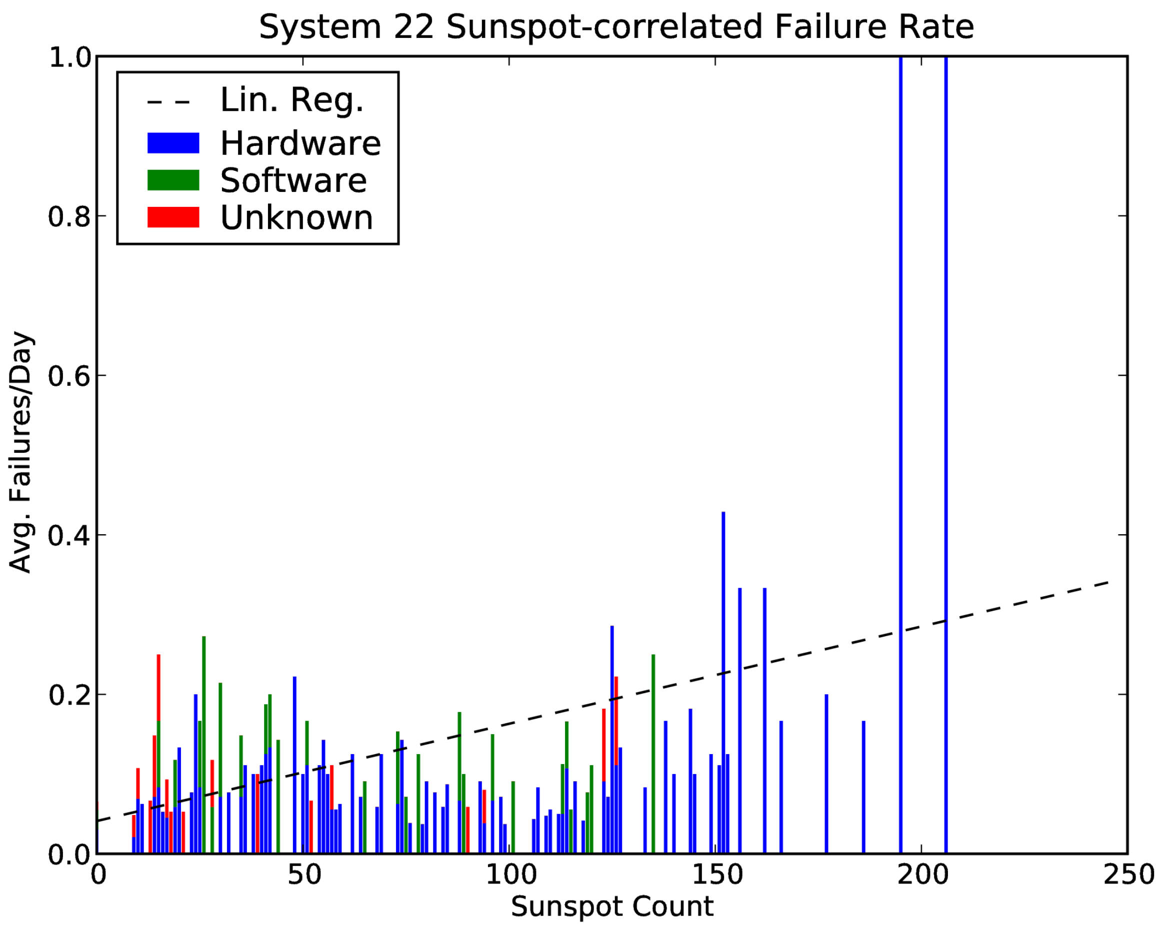
An larger amount of RAM in an older technology, as well as double the number of processors than System 12, which is similar, likley contributed to the large number of failures.



System 22

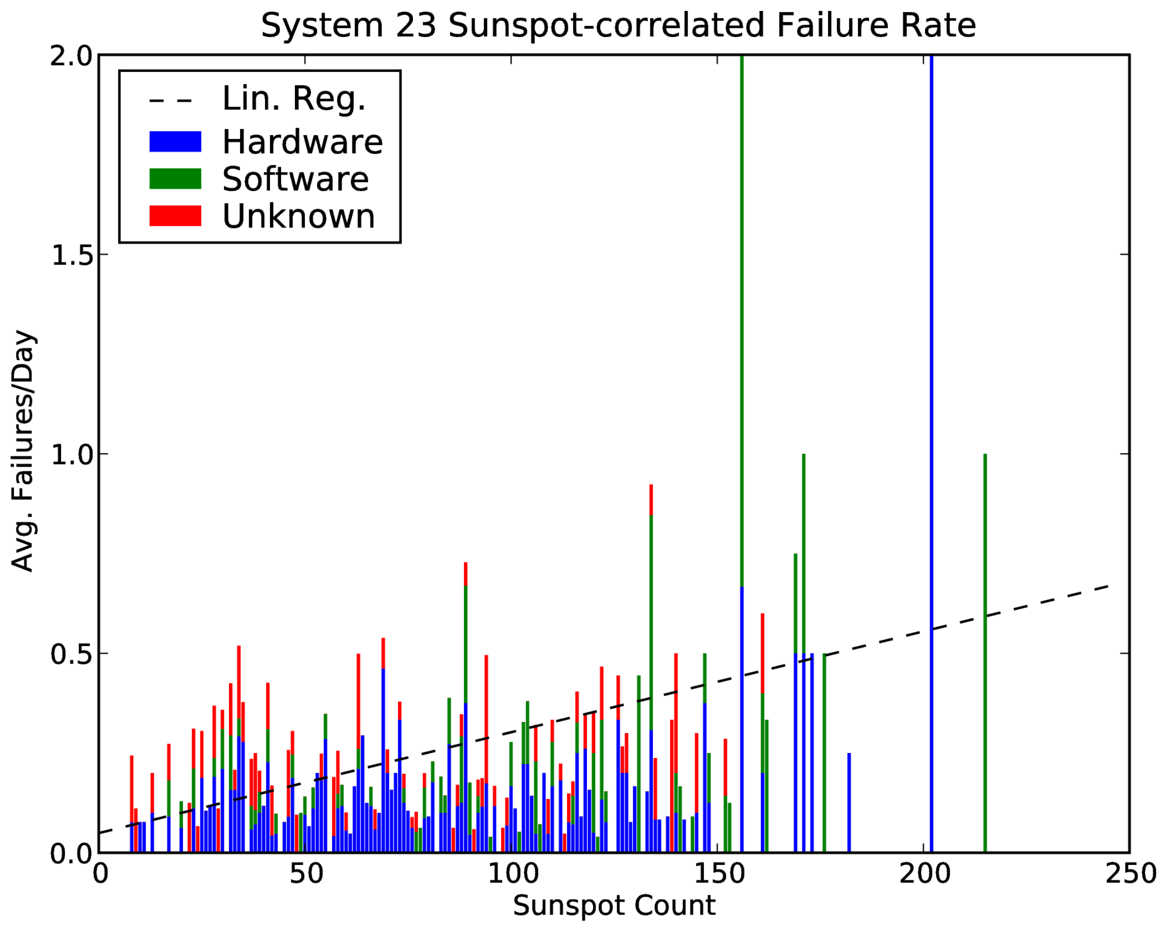
128 failures.

This system has a different processor type than all the other systems. It is also a single SMP machine instead of a cluster. Even such a 'small' system exhibits a definite positive correlation.



System 23

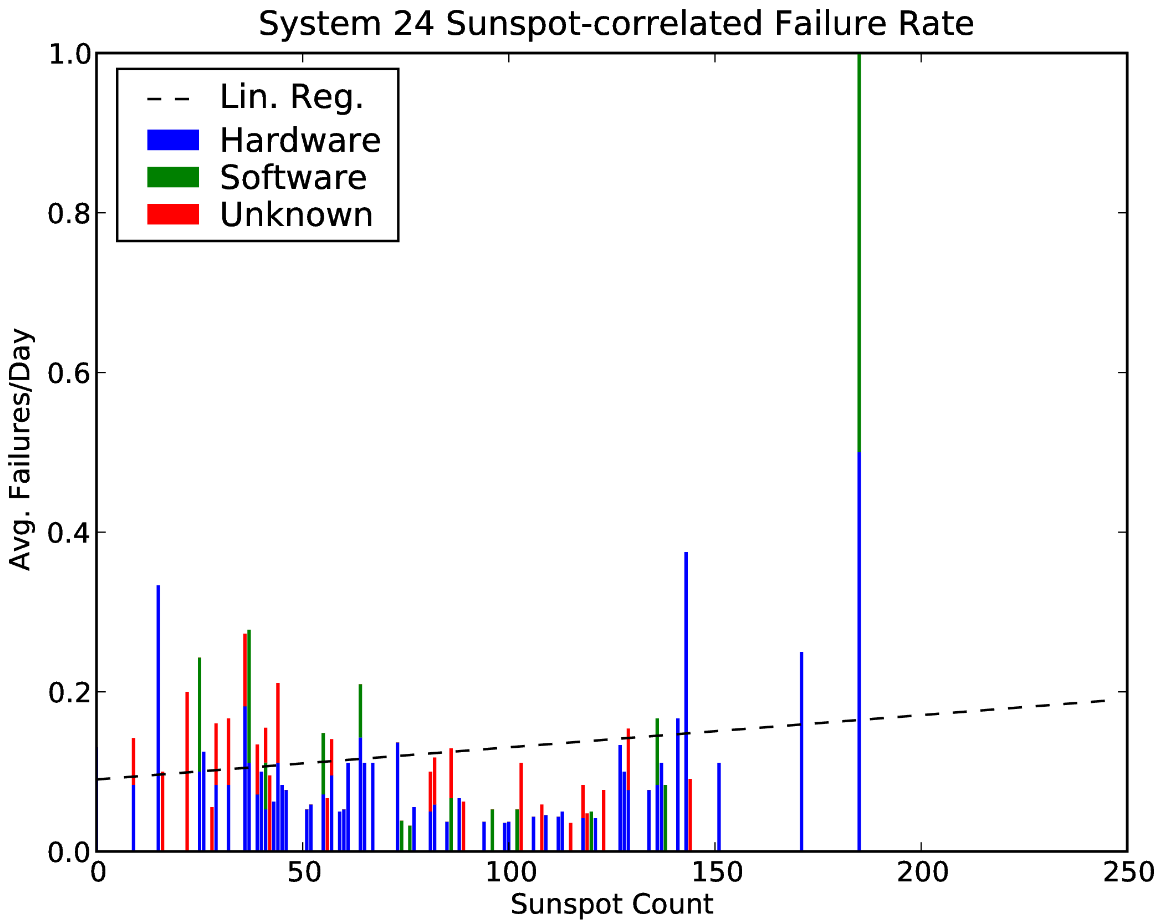
497 failures.



System 24

112 failures.

Like System 22, this was an older, unique SMP system.

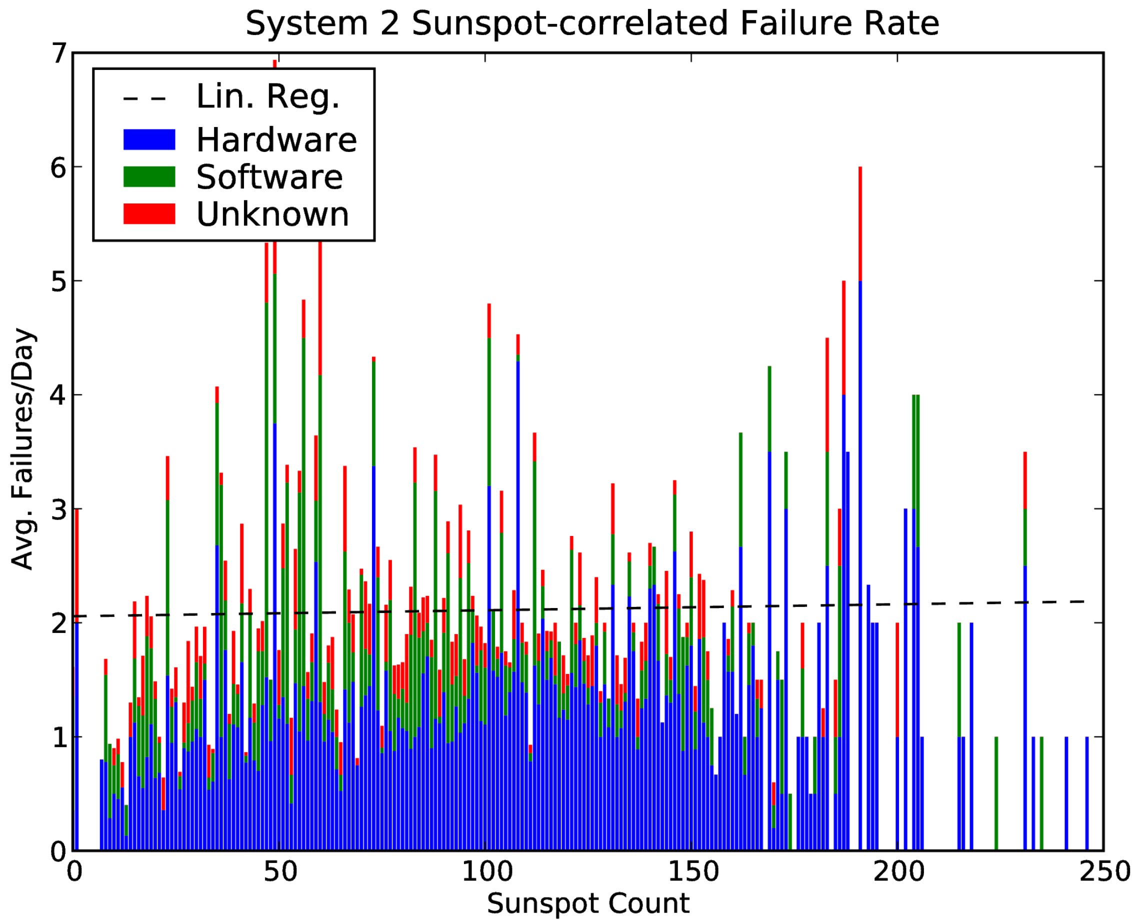


Systems with weak or negative correlation

System 2

6614 failures.

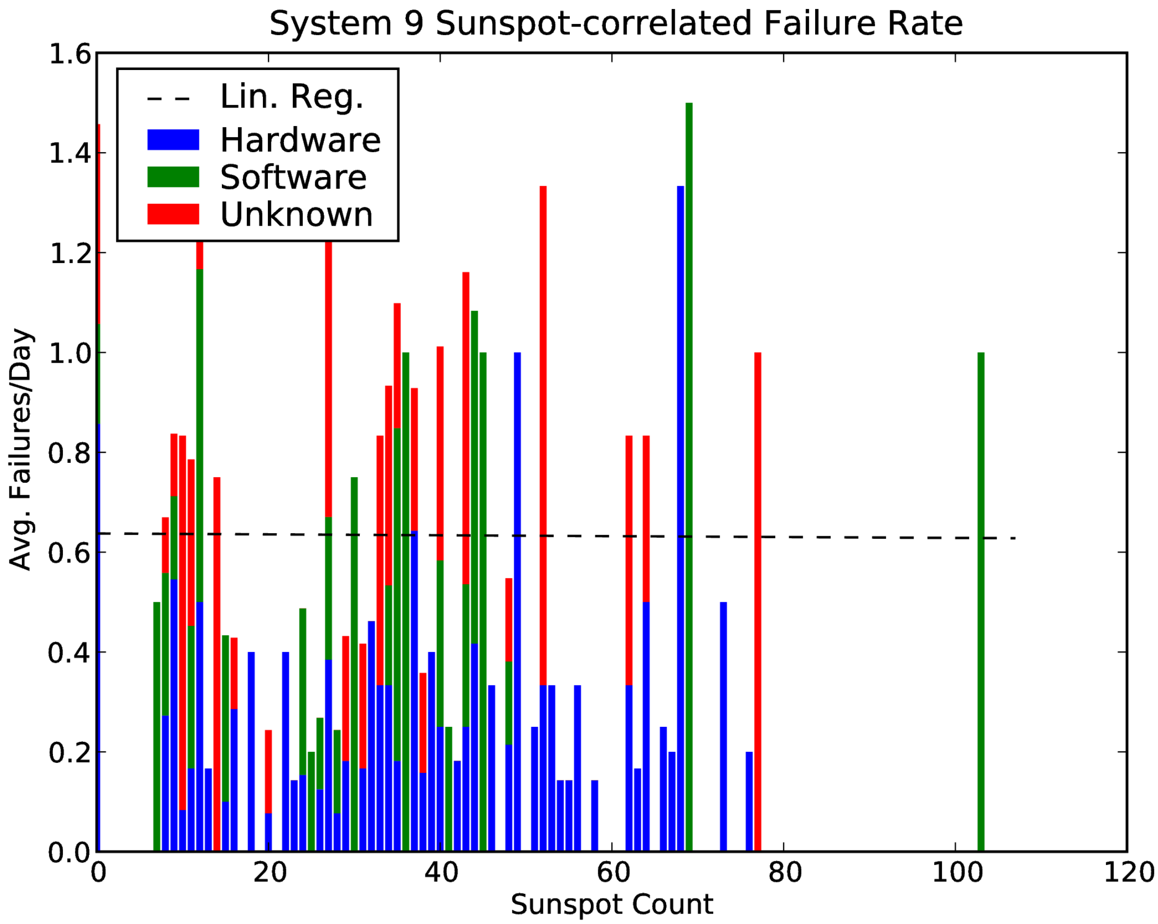
The linear regression suggests that with a nearly constant average of two failures per day, this system, although large and long-lived, is unreliable enough to mask the effect of solar activity.



System 9

272 failures.

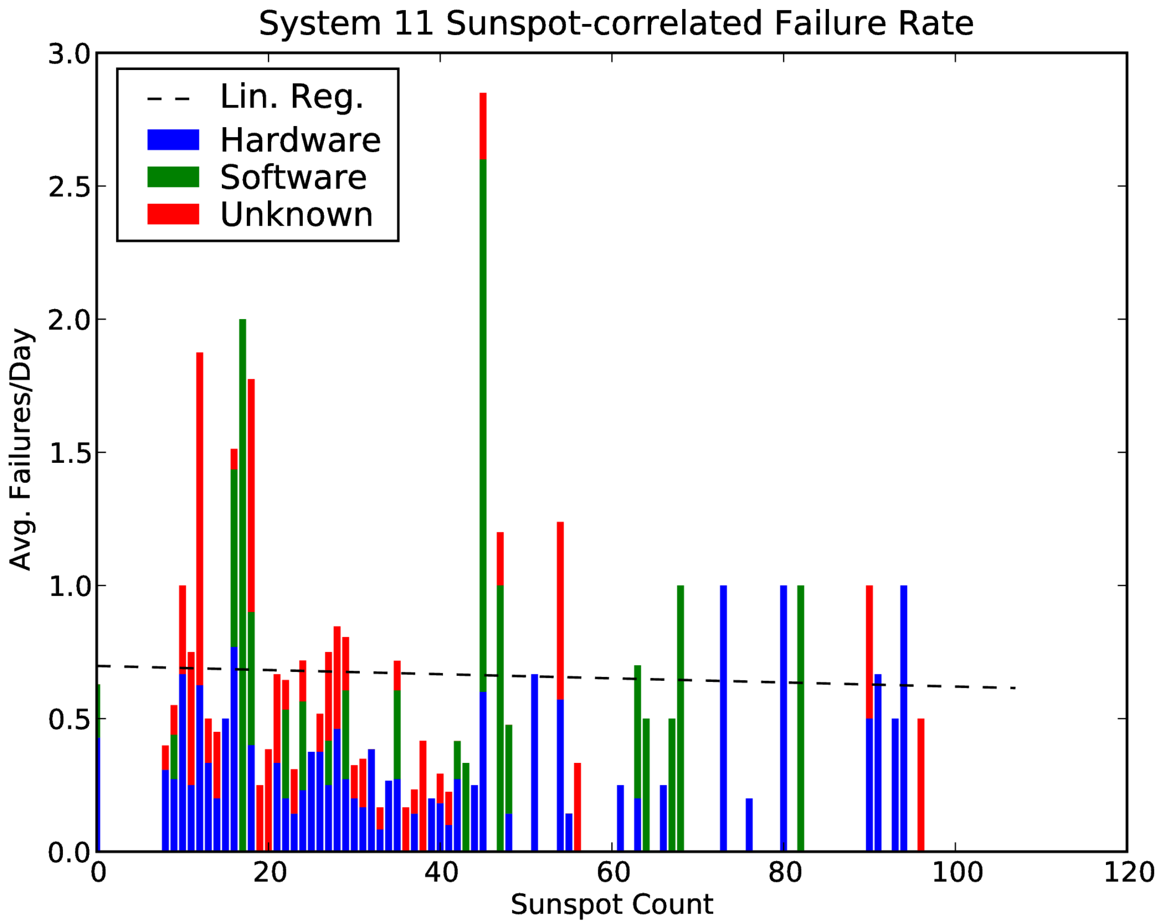
Compare to Systems 10 and 11.



System 11

263 failures.

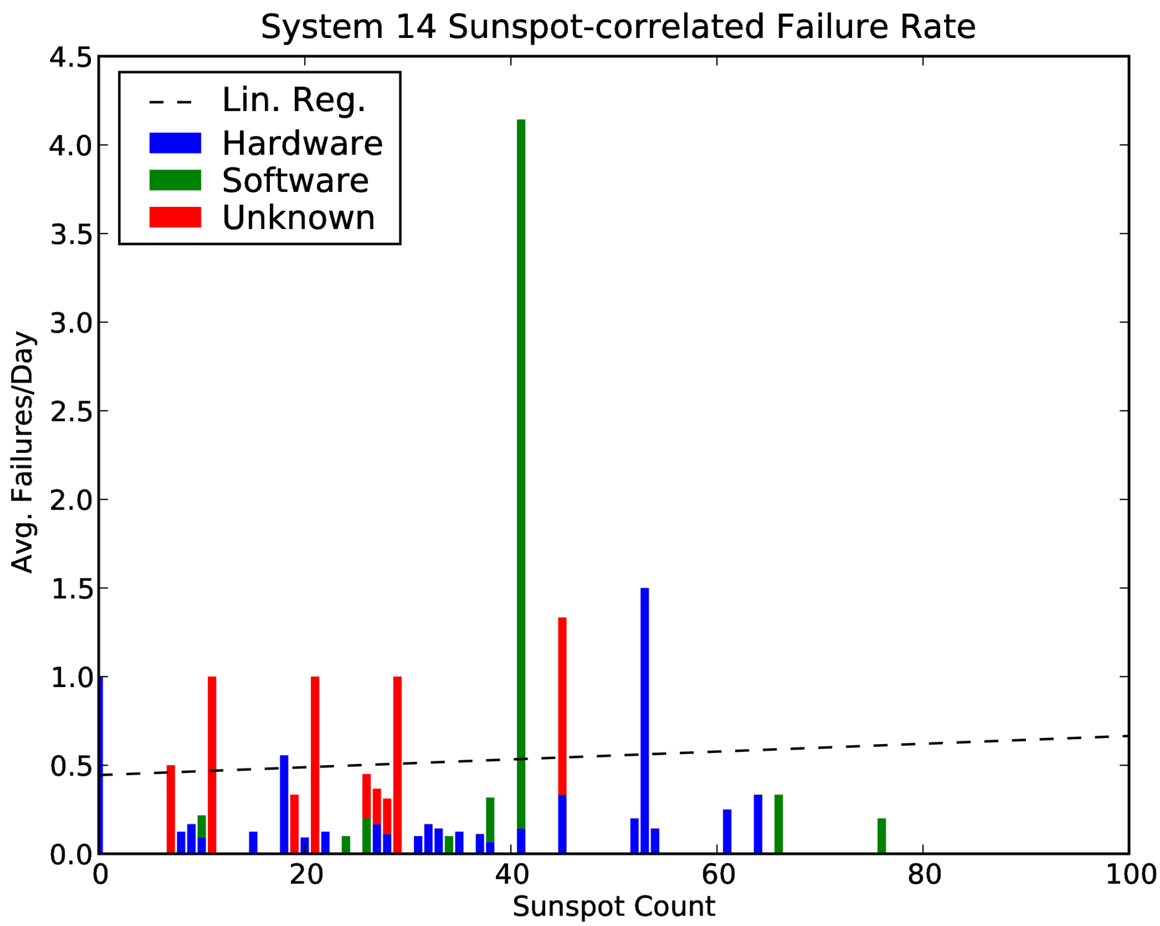
Compare to Systems 9 and 10.



System 14

122 failures.

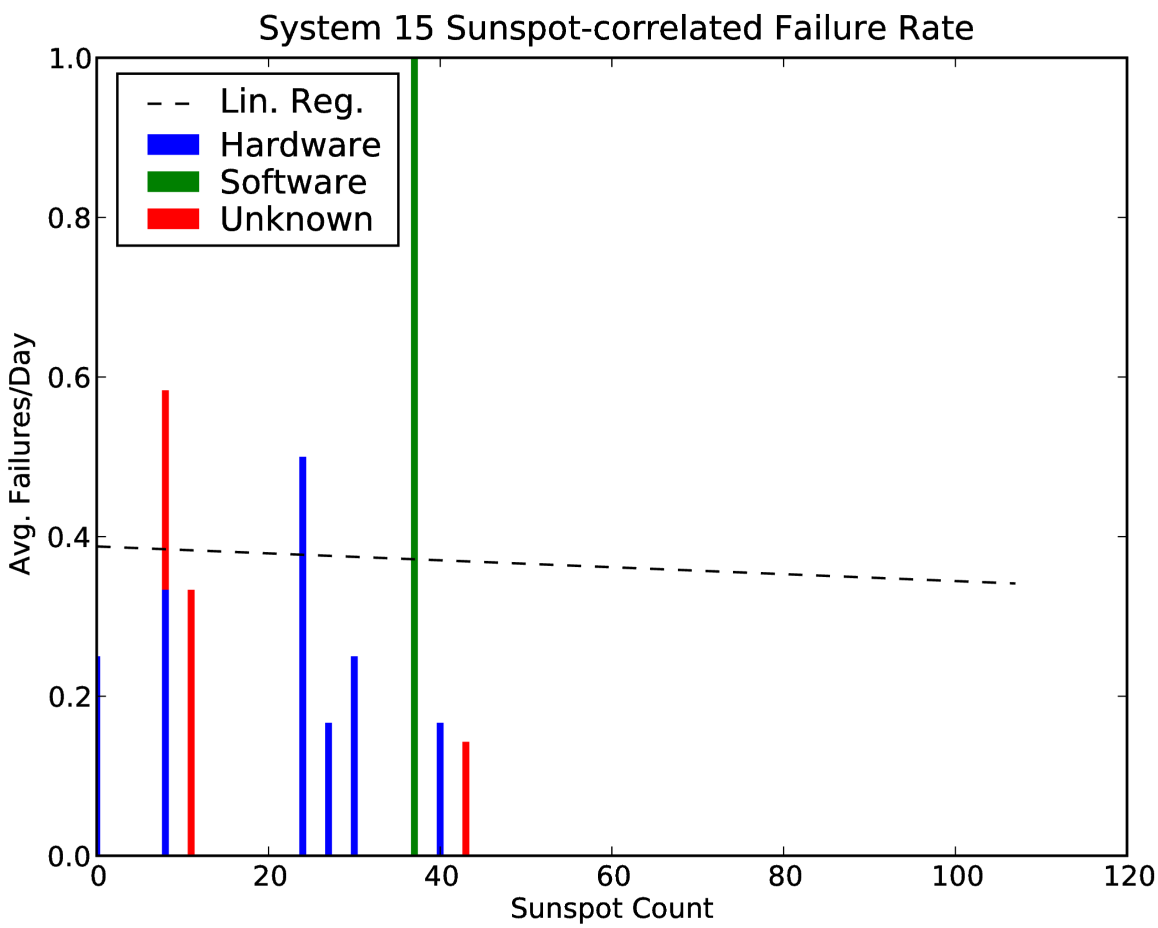
It's unclears whether this system exhibits a significant positive correlation or not. There is a large proportion of failures whose causes are undertermined.



System 15

51 failures.

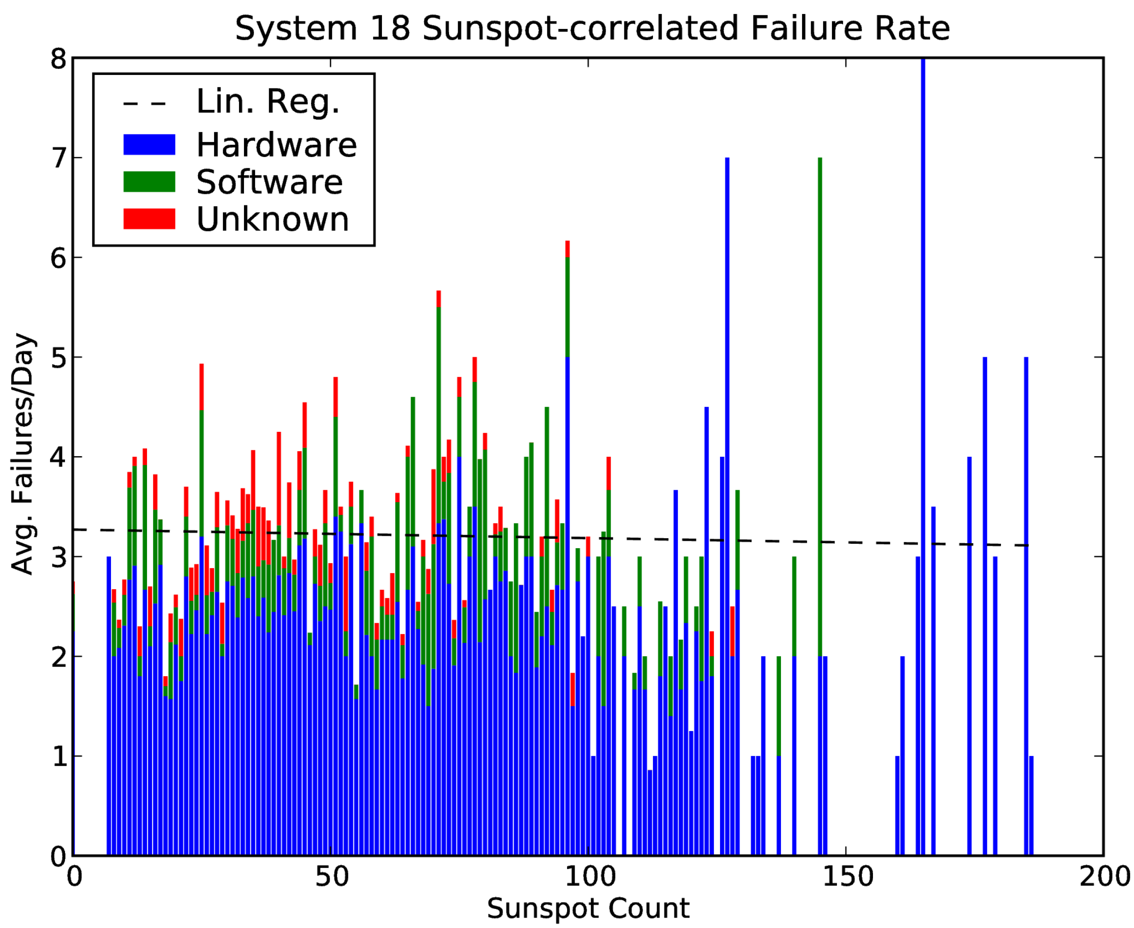
This system had been in operation for only nine months. There is not yet enough data to detect a trend.



System 18

3948 failures.

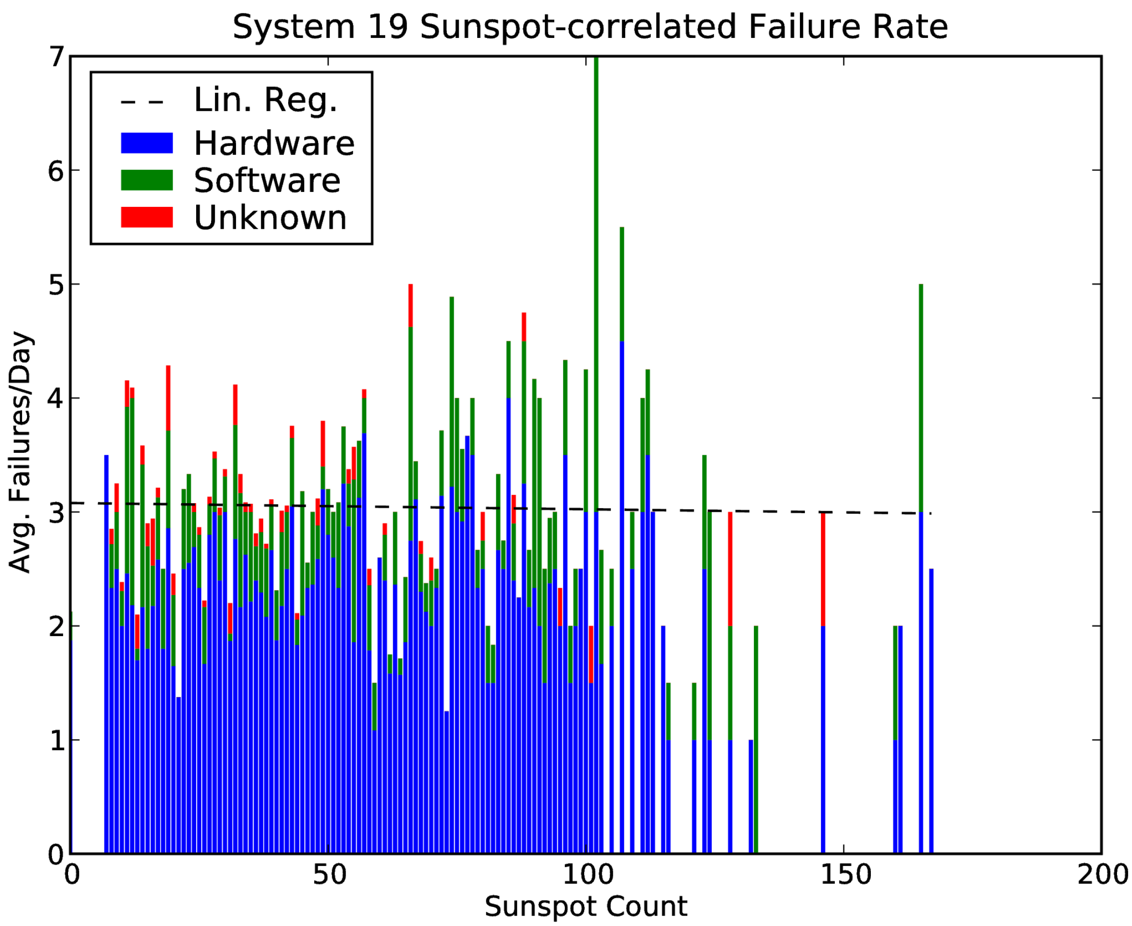
As with System 2, this system seems too unreliable to show the effect of solar activity.



System 19

3244 failures.

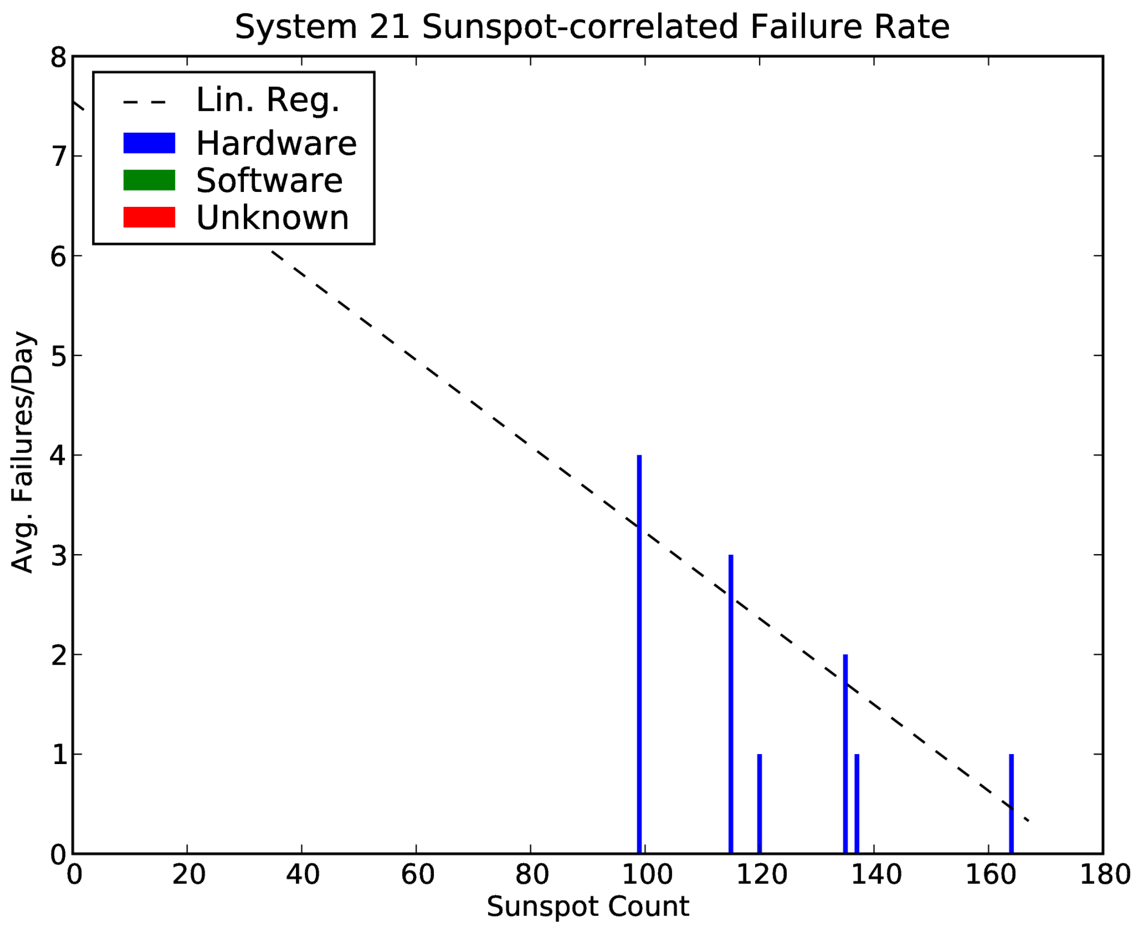
Again, too unreliable to show a correlation.



System 21

110 failures.

This system was in operation for only five months, and removing the first three months of data left too few failures to chart.



Appendix A

The systems are ordered and identified by the machine numbers used in the failure logs (2nd column). System 17 is not included since no installation or configuration data was provided.

system CMU paper number	system data machine number	system type	number nodes	number cpus	cpus/ node	install date	production date	decommision date	fru	mem per node	cpu type	number of interconnects	use type
20	2	cluster	49	6152	128	Nov-96	Jan-97	current	part	128	1	12	graphics/ compute
9	3	cluster	128	512	4	Aug-03	Sep-03	current	part	4	2	1	compute
10	4	cluster	128	512	4	Aug-03	Sep-03	current	part	4	2	1	compute
11	5	cluster	128	512	4	Aug-03	Sep-03	current	part	4	2		compute
12	6	cluster	32	128	4	Aug-03	Sep-03	current	part	16	2	1	compute
1	7	smp	1	8	8	before tracking	before tracking	Dec-99	part	16	3	0	compute
4	8	cluster	164	328	2	Mar-01	Apr-01	current	part	1	4	1	compute
14	9	cluster	256	512	2	Aug-03	Sep-03	current	node	4	2	1	compute
15	10	cluster	256	512	2	Aug-03	Sep-03	current	node	4	2	1	compute
16	11	cluster	256	512	2	Aug-03	Sep-03	current	node	4	2	1	compute
18	12	cluster	512	1024	2	Aug-03	Sep-03	current	node	4	2	1	compute
17	13	cluster	256	512	2	Aug-03	Sep-03	current	node	4	2	1	compute
13	14	cluster	128	256	2	Aug-03	Sep-03	current	node	4	2	1	compute
22	15	numa	1	265	256	Nov-04	Nov-04	current	part	1024	5	0	compute
19	16	cluster	16	2048	128	Oct-96	Dec-96	Sep-02	part	32	1	4	compute
7	18	cluster	1024	4096	4	Mar-02	May-02	current	part	16	2	2	compute
8	19	cluster	1024	4096	4	Aug-02	Oct-02	current	part	16	2	2	compute
5	20	cluster	512	2048	4	Oct-01	Dec-01	current	part	16	2	2	compute
6	21	cluster	128	512	4	Aug-01	Sep-01	Jan-02	part	16	2	2	compute
3	22	smp	1	4	4	before tracking	before tracking	Apr-03	part	1	6	0	compute
21	23	cluster	5	544	128	Oct-98	Oct-98	Dec-04	part	128	1	4	compute
2	24	smp	1	32	32	before tracking	before tracking	Dec-03	part	8	7	1	compute

Bibliography

[1] Brogley, Mike, "*FPGA Device Reliability and the Sunspot Cycle*", http://www.embedded.com/columns/technicalinsights/220301380, June 10 2009, [Accessed November 14 2009]

[2] http://en.wikipedia.org/wiki/List_of_solar_cycles, [Accessed November 14 2009]

[3] http://en.wikipedia.org/wiki/Solar_variation, [Accessed November 14 2009]

[4] "*Solar Influences Data Analysis Centre*", http://www.sidc.oma.be/, [Accessed November 14 2009]

[5] "*Sunspot Index Data Centre*", http://www.icsu-fags.org/ps11sidc.htm, [Accessed November 14 2009]

[6] "*Operational Data to Support and Enable Computer Science Research*", http://institutes.lanl.gov/data/fdata/, [Accessed November 14 2009]

[7] Alvestad, Jan, "*Solar cycles 21, 22 and 23*", http://www.solen.info/solar/solcycle.html, [Accessed November 16 2009]

[8] ISES Solar Cycle Sunspot Number Progression, http://landscheidt.auditblogs.com/2007/06/01/dr-landscheidts-solar-cycle-24-prediction/ [Accessed January 7 2010]