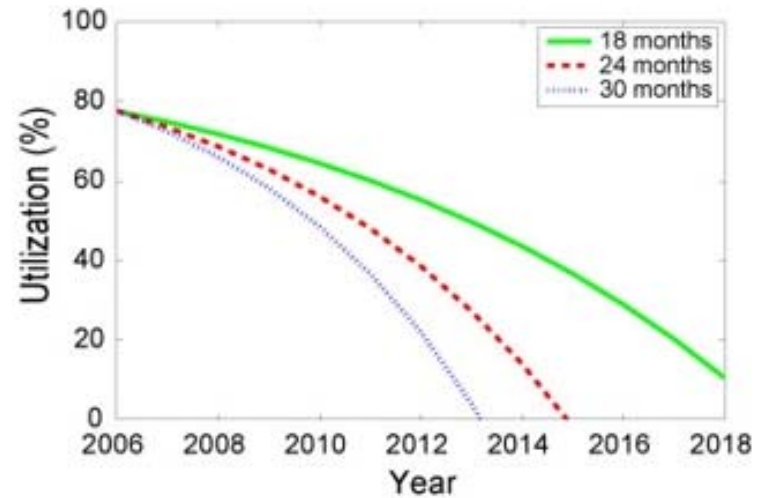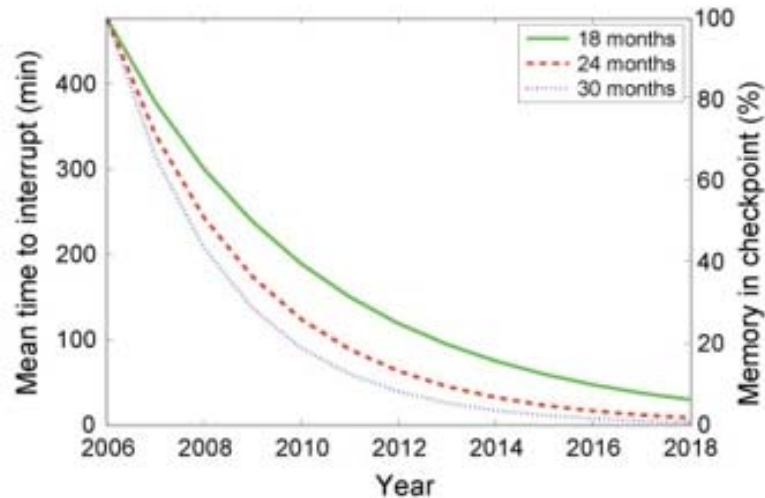# The Computer Failure Data Repository (CFDR)

Bianca Schroeder

Department of Computer Science

University of Toronto
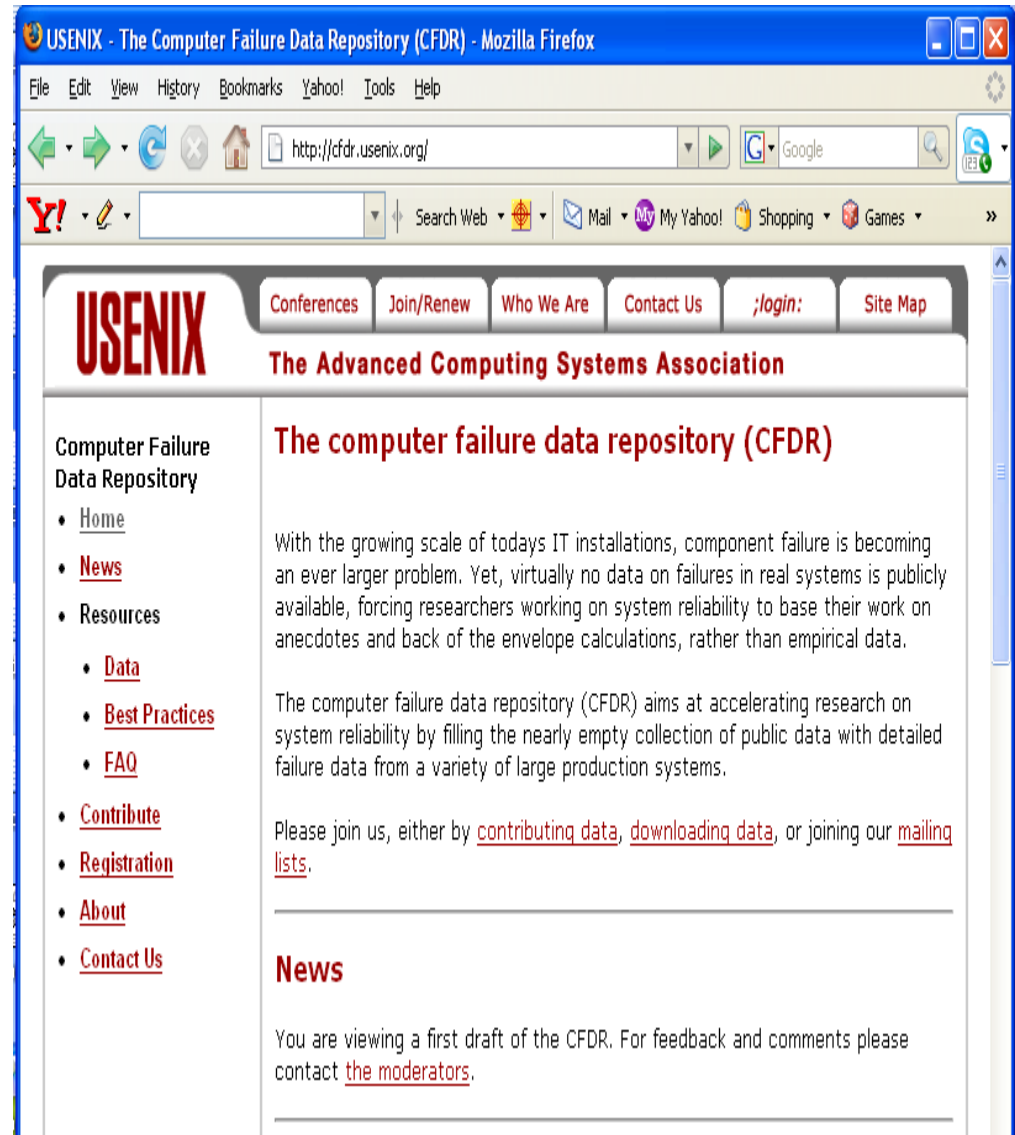
pdsi

The need for understanding failures

- Component count in systems grows
- Mean time to interrupt expected to drop sharply

The computer failure data repository (CFDR)

- Gather & publish real failure data
- Community effort
  - Usenix clearinghouse
- Data on all aspects of system failure
- Anonymized as needed

- **http://cfdr.usenix.org**

Available data

Available data

The table below provides an overview over the available data sets.

Name	Time period	System type	Type of data
LANL	Dec 96 - Nov 05	HPC clusters	The data covers node outages at 22 cluster systems at LANL, including a total of 4,750 nodes and 24,101 processors. Usage logs and error logs are available as
HPC1	Aug 01 - May 06		
HPC2	Jan 04 - Jul 06	HPC cluster	
HPC3	Dec 05 - Nov 06	HPC cluster	
HPC4	2004 - 2006	HPC cluster	
PNNL	Nov 03 - Sep 07	HPC cluster	Hardware failures recorded on the MPP2 system (a 980 node HPC cluster) at PNNL.
NERSC	2001 - 2006	HPC cluster	I/O specific failures collected at a number of production systems at NERSC.
COM1	May 2006	Internet services cluster	Hardware failures recorded by an internet service provider and drawing from multiple distributed sites.
COM2	Sep 04 - Apr 06	Internet services cluster	Warranty service log of hardware failures aggregating events in multiple distributed sites.
COM3	Jan 05 - Dec 05	Internet services cluster	Aggregate quarterly statistics of disk failures at a large external storage system.
ask.com	Dec 06 - Feb 07	Internet services cluster	Memory error data collected on a 212 node server farm at ask.com.

- Description of **systems** data comes from
- Description of **data format**
- **Papers** using/analyzing data
- Pointer to **raw data**

Available data

Available data

The table below provides an overview over the available data sets.

Name	Time period	System type	Type of data
LANL	Dec 96 - Nov 05	HPC clusters	The data covers node outages at 22 cluster systems at LANL, including a total of 4,750 nodes and 24,101 processors. Usage logs and error logs are available as well.
HPC1	Aug 01 - May 06	HPC cluster	The data covers hardware replacements at a 765 node cluster with more than 3,000 hard drives.
HPC2	Jan 04 - Jul 06	HPC cluster	Hard drive replacements in a 256 node cluster with 520 drives.
HPC3	Dec 05 - Nov 06	HPC cluster	Hard drive replacements observed in a 1,532-node HPC cluster with more than 14,000 drives.
HPC4	2004 - 2006	HPC cluster	Error logs collected at 5 supercomputing systems at SNL and LLNL, ranging from 512 to 131072 processors.
PNNL	Nov 03 - Sep 07	HPC cluster	Hardware failures recorded on the MPP2 system (a 980 node HPC cluster) at PNNL.
NERSC	2001 - 2006	HPC cluster	I/O specific failures collected at a number of production systems at NERSC.
COM1	May 2006	Internet services cluster	Hardware failures recorded by an internet service provider and drawing from multiple distributed sites.
COM2	Sep 04 - Apr 06	Internet services cluster	Warranty service log of hardware failures aggregating events in multiple distributed sites.
COM3	Jan 05 - Dec 05	Internet services cluster	Aggregate quarterly statistics of disk failures at a large external storage system.
ask.com	Dec 06 - Feb 07	Internet services cluster	Memory error data collected on a 212 node server farm at ask.com.

High-Performance Computing

Internet Services

pdsi

Available data

Cluster node outages
- Root cause
- Repair time
- Error logs
- Usage data

Hardware failures
- CPU
- Memory
- Disks
- Power adapter
- …

Storage failures
- Drives
- Tapes
- File system
- Network

Available data

The table below provides an overview over the available data sets.

Name	Time period	System type	Type of data
LANL	Dec 96 - Nov 05	HPC clusters	The data covers node outages at 22 cluster systems at LANL, including a total of 4,750 nodes and 24,101 processors. Usage logs and error logs are available as well.
HPC1	Aug 01 - May 06	HPC cluster	The data covers hardware replacements at a 765 node cluster with more than 3,000 hard drives.
HPC2	Jan 04 - Jul 06	HPC cluster	Hard drive replacements in a 256 node cluster with 520 drives.
HPC3	Dec 05 - Nov 06	HPC cluster	Hard drive replacements observed in a 1,532-node HPC cluster with more than 14,000 drives.
HPC4	2004 - 2006	HPC cluster	Error logs collected at 5 supercomputing systems at SNL and LLNL, ranging from 512 to 131072 processors.
PNNL	Nov 03 - Sep 07	HPC cluster	Hardware failures recorded on the MPP2 system (a 980 node HPC cluster) at PNNL.
NERSC	2001 - 2006	HPC cluster	I/O specific failures collected at a number of production systems at NERSC.
COM1	May 2006	Internet services cluster	Hardware failures recorded by an internet service provider and drawing from multiple distributed sites.
COM2	Sep 04 - Apr 06	Internet services cluster	Warranty service log of hardware failures aggregating events in multiple distributed sites.
COM3	Jan 05 - Dec 05	Internet services cluster	Aggregate quarterly statistics of disk failures at a large external storage system.
ask.com	Dec 06 - Feb 07	Internet services cluster	Memory error data collected on a 212 node server farm at ask.com.

pdsi

- **Thanks** to all organizations who have contributed so far!

Do you have any data
to contribute?

Contact us:
garth@cs.cmu.edu
bianca@cs.toronto.edu

pdsi