Disk failures in the real world: Data corruption in the storage stack

Bianca Schroeder
University of Toronto

L. Bairavasundaram*†, G. Goodson†,
A. Arpaci-Dusseau*, R. Arpaci-Dusseau*
*University of Wisconsin-Madison, †Network Appliances

What do we know?

Drive replacements	Latent Sector Errors	Silent corruption
[Schroeder, Gibson FAST'07] [Pinheiro et al., FAST'07] [Jiang et al., FAST'08]	[Bairavasundaram et al., Sigmetrics'07]	?

- Silent data corruption
  - Not detected/reported by disk
  - Higher potential of leading to data loss
  - Many sources
    - Software (file system / software RAID)
    - Firmware (Disk / adapters)

[Bairavasundaram, Goodson, Schroeder, 2x Arpaci-Dusseau] FAST'08

<Bianca Schroeder> © August 08

Questions about silent data corruption

- How common?
- Factors
- Characteristics
- How detected?

- Disk class (Nearline / Enterprise)
- Disk model
- Disk age
- Disk size (capacity)

- Spatial locality
- Temporal locality

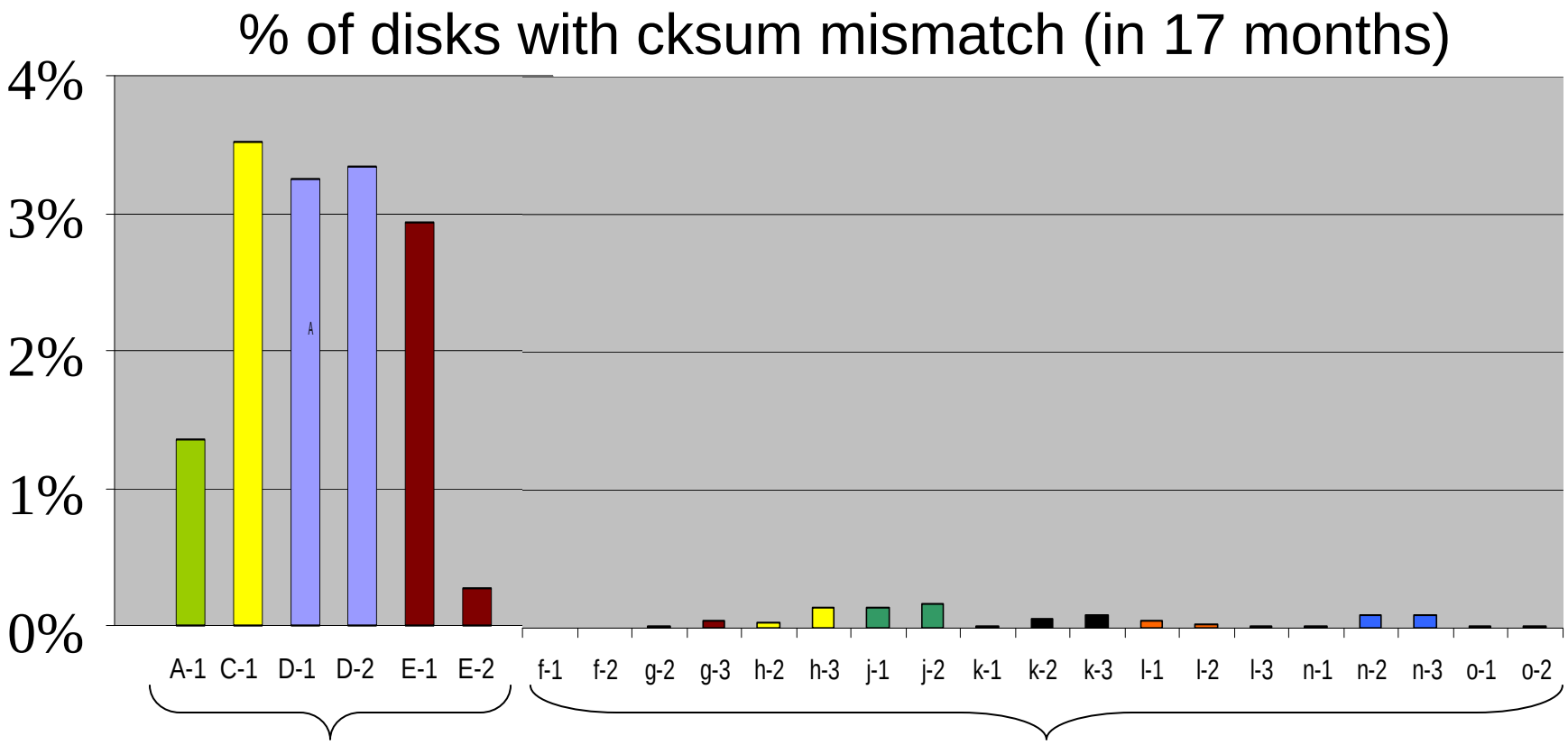
- Scrub vs. FS op vs. reconstruction

The data

- Total 1.53 million disks
  - Nearline & enterprise class drives
  - 15 different drive families
  - 26 different drive models
- Time period: Jan 2004 to Jun 2007
- Detecting corruption:
  - Netapp metadata for every 4KB of data with checksum
  - Verification during all operations
- We study <u>checksum mismatch</u> events
  - (Also looked at other ways, not part of this talk).

<u>Important note:</u> Checksum allows us to identify corruption, but not the source of the corruption!

How common is data corruption?

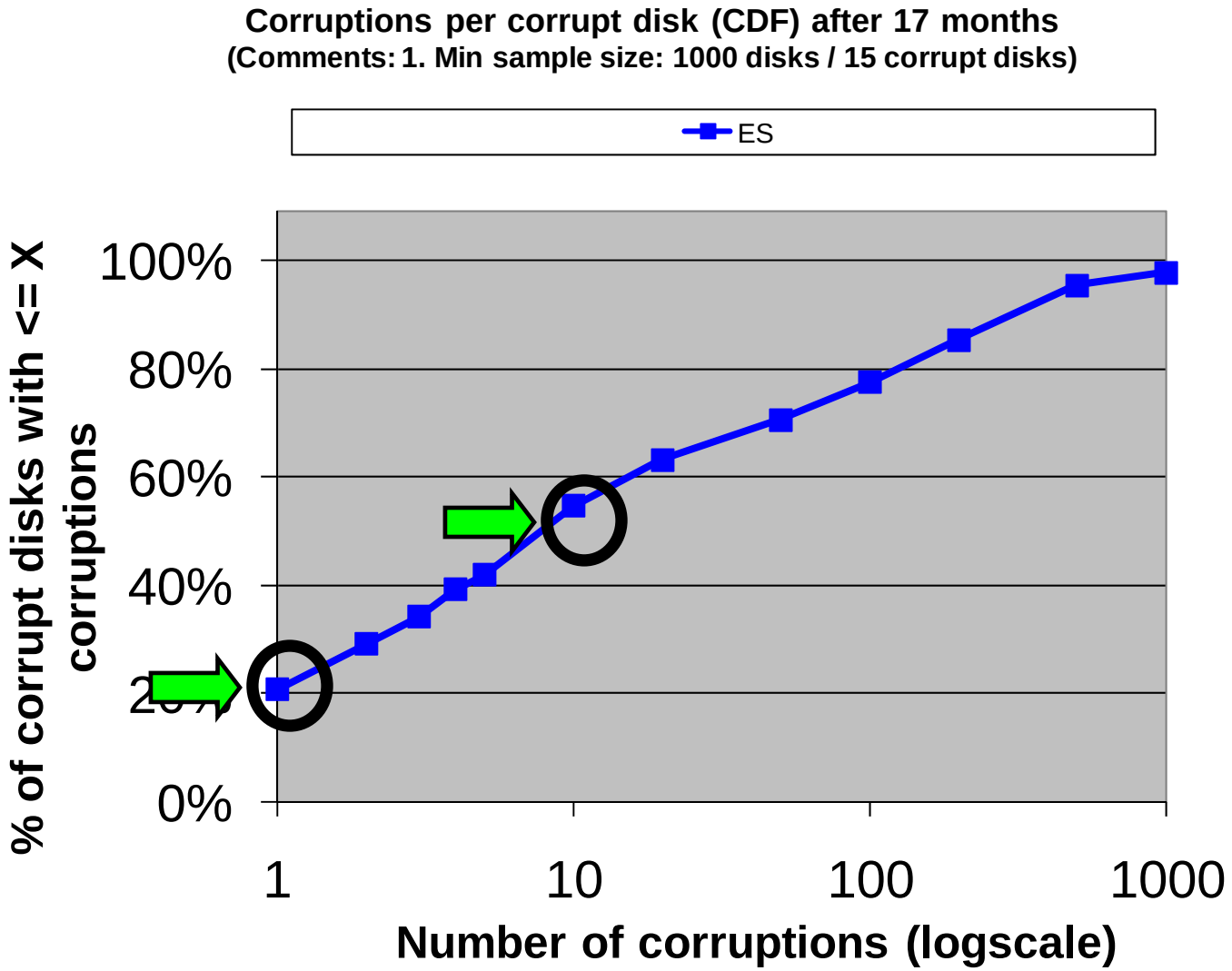


Nearline: 0.9 %

Enterprise: 0.07 %

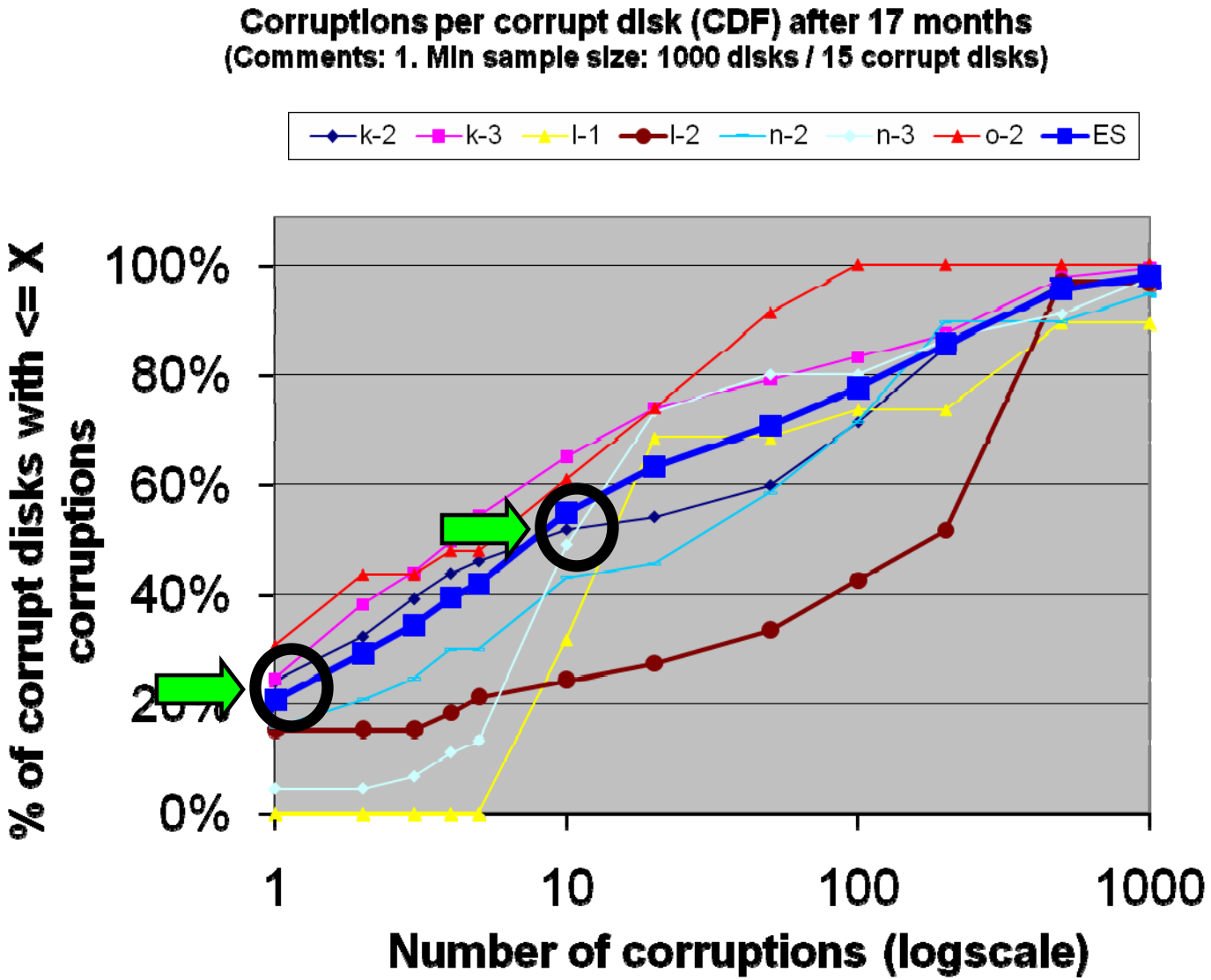
- More than 400,000 checksum mismatch events
- Frequency depends greatly on class and model!

Corruptions per corrupt disk (Enterprise)



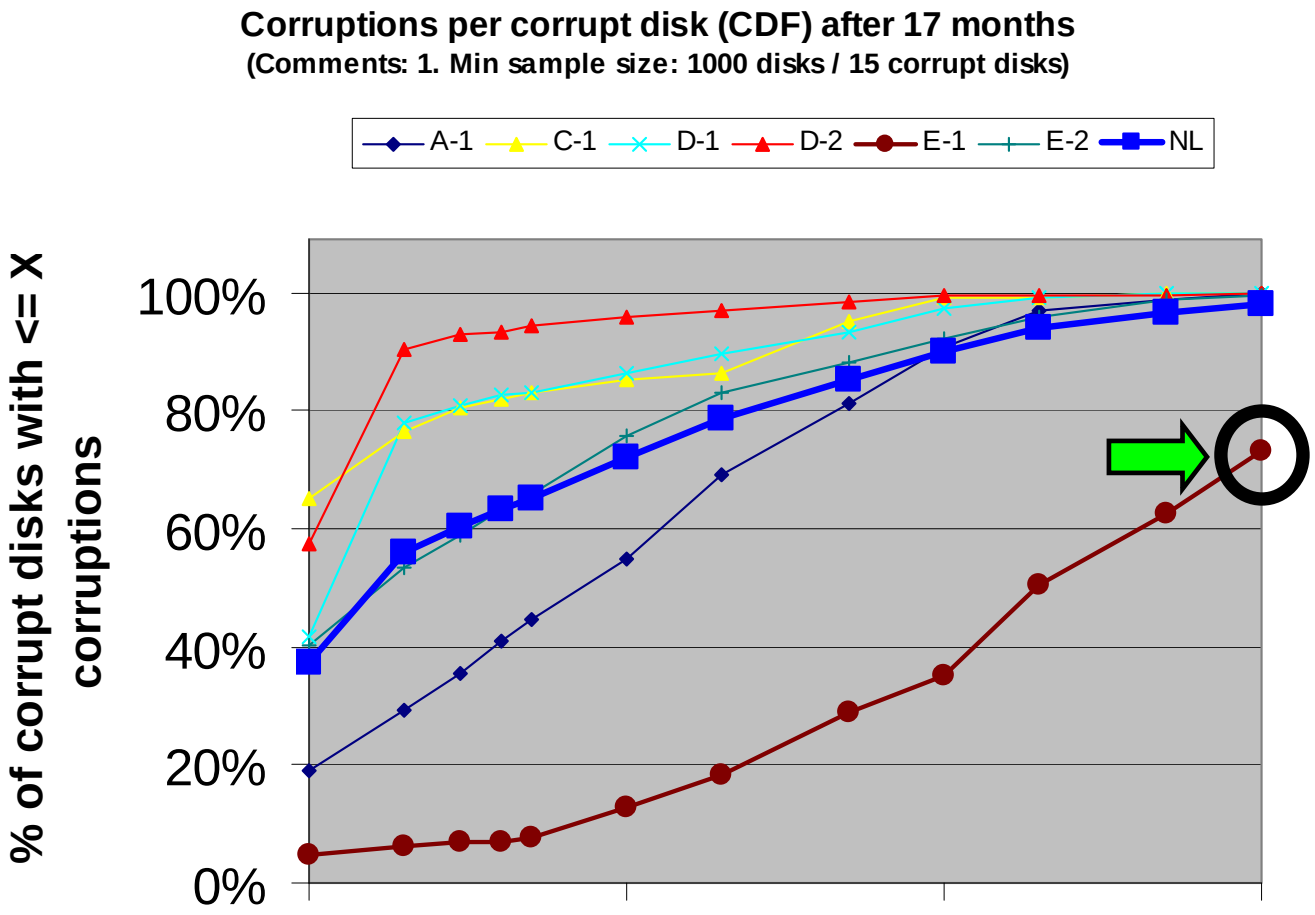
<Bianca Schroeder> © August 08

Corruptions per corrupt disk (Enterprise)



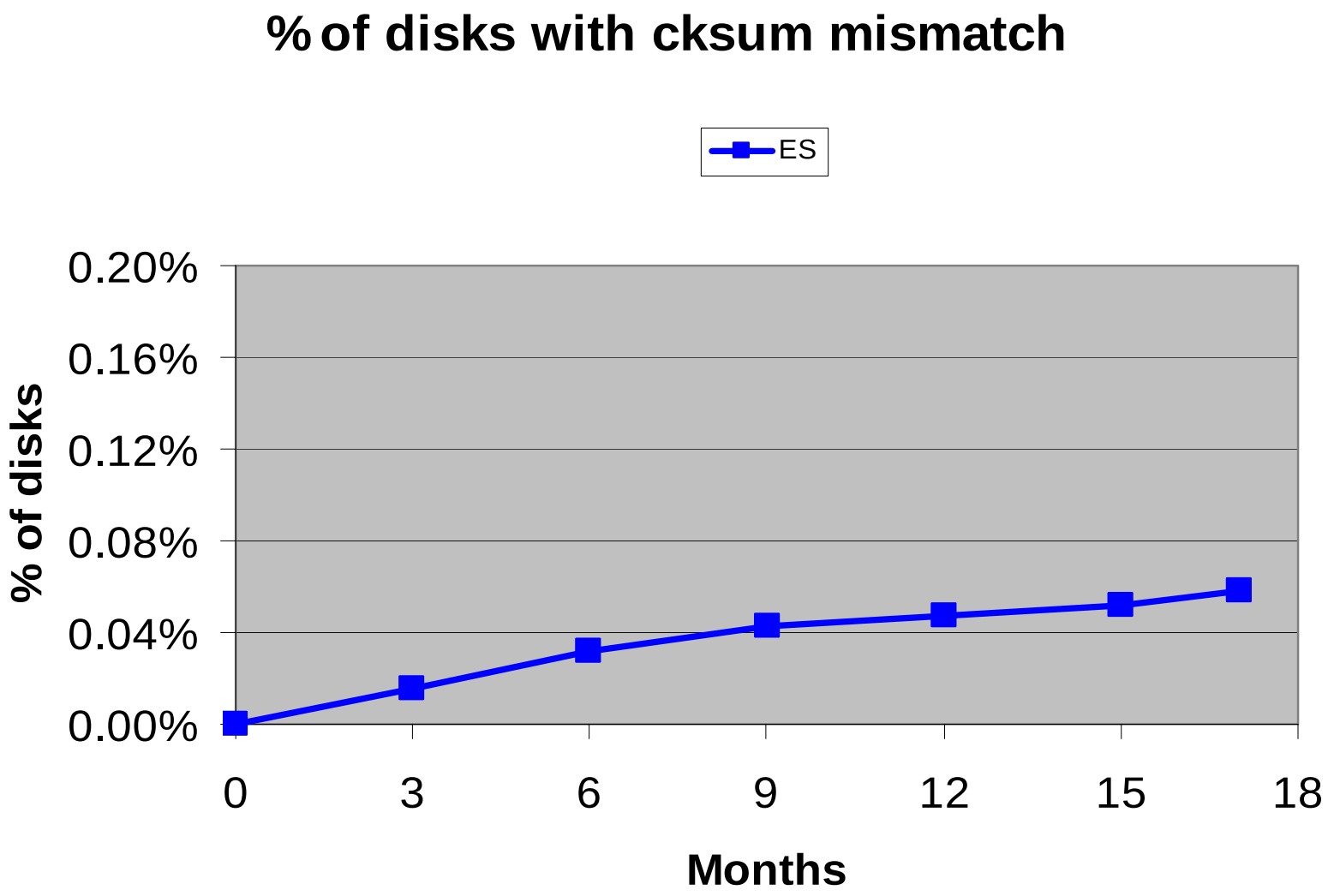
<Bianca Schroeder> © August 08

Corruptions per corrupt disk (Nearline)



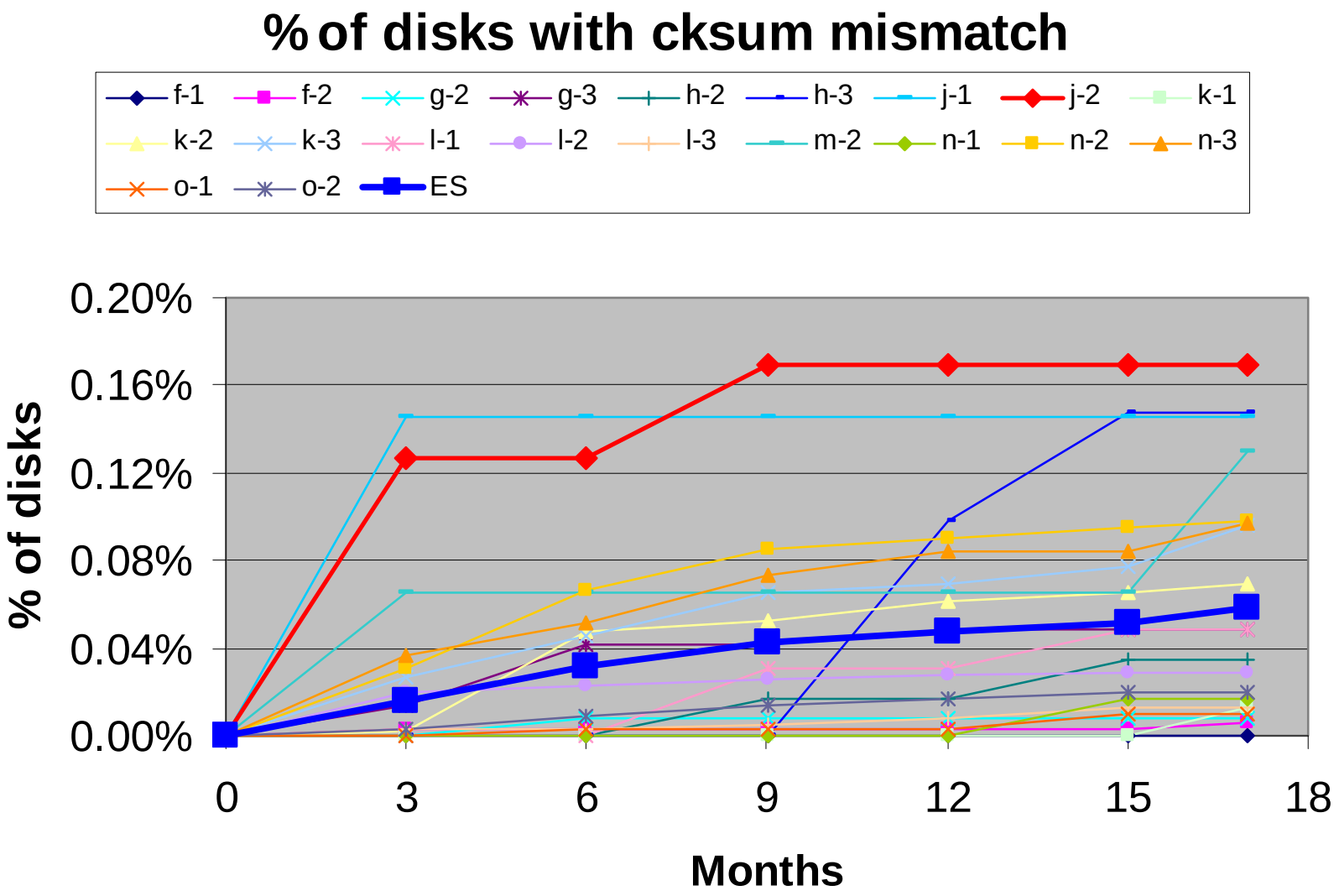
Huge differences between models
Enterprise drives more likely to develop more corruptions

Effect of Disk Age – Enterprise



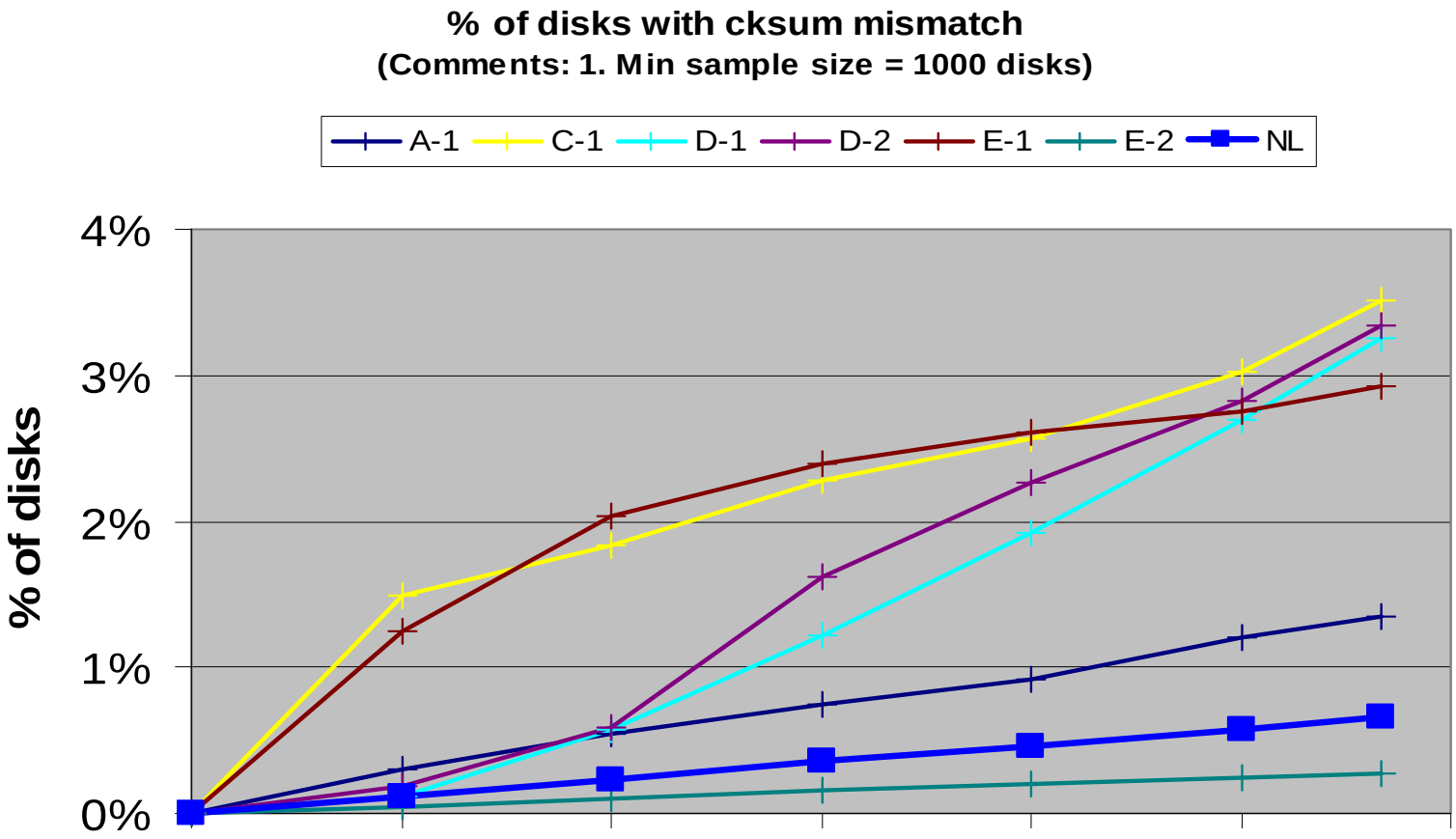
<Bianca Schroeder> © August 08

Effect of Disk Age – Enterprise



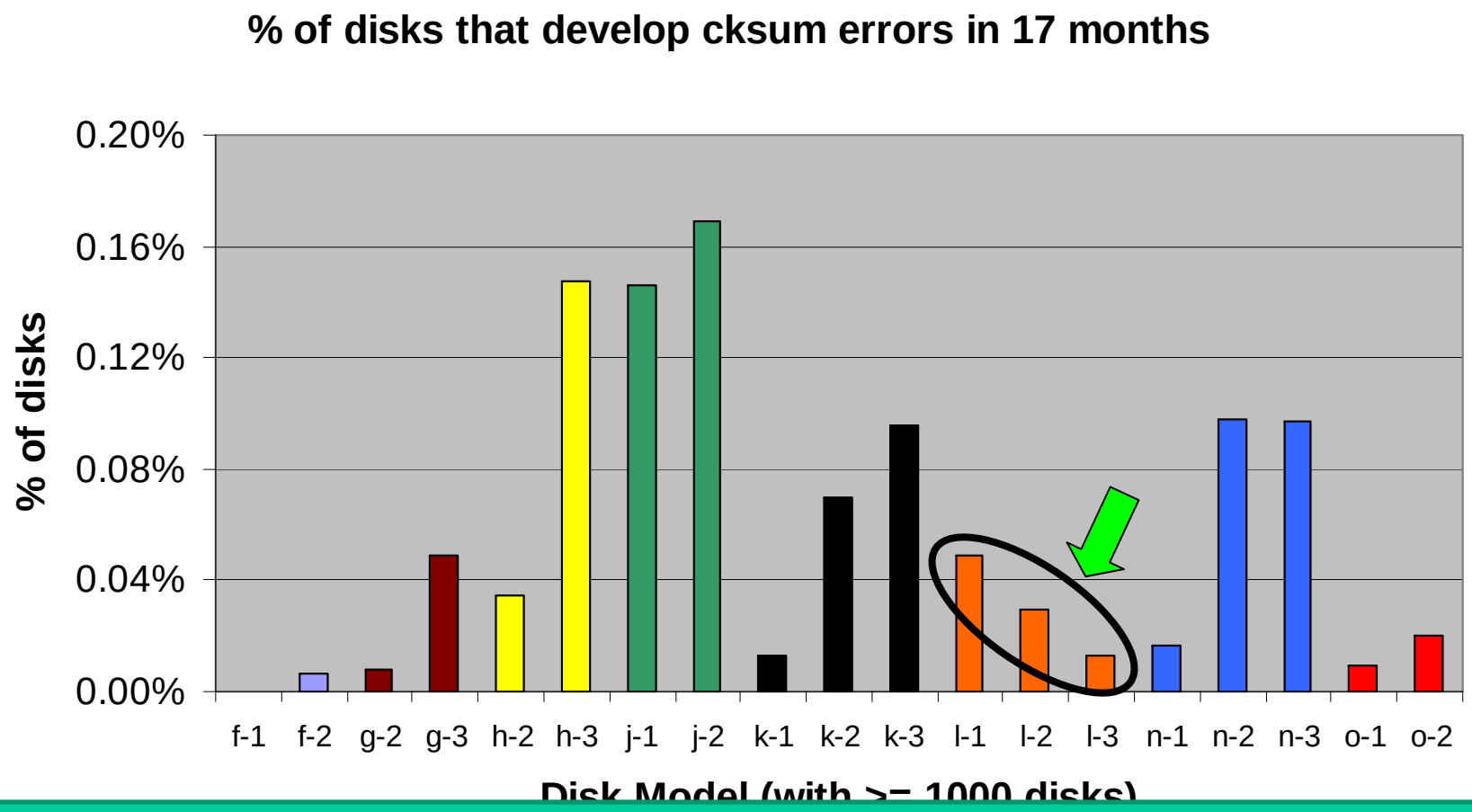
<Bianca Schroeder> © August 08

Effect of Disk Age – Nearline



Nearline drives: fairly independent of age
Enterprise drives: rate slows down with age

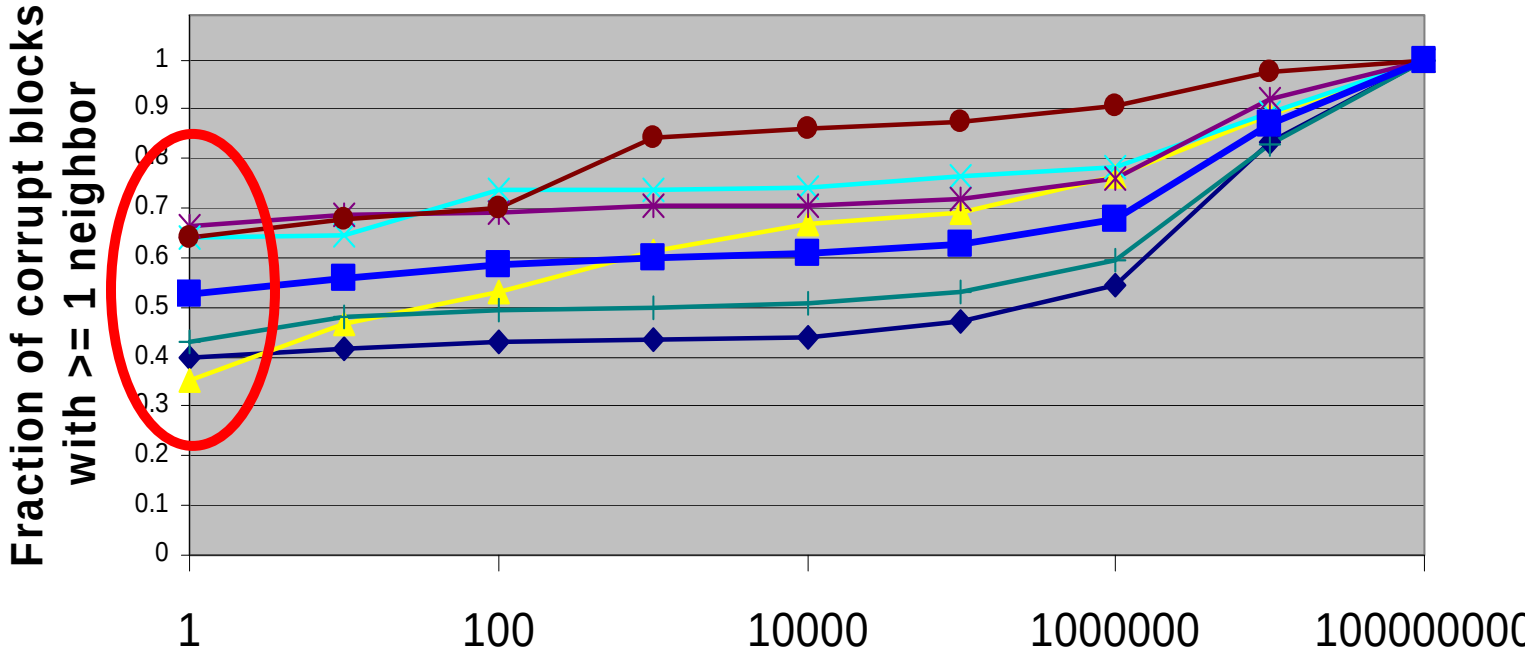
Effect of Disk Size – Enterprise



- Enterprise drives: %affected disks increases with size
- Nearline drives: effect of size not clear

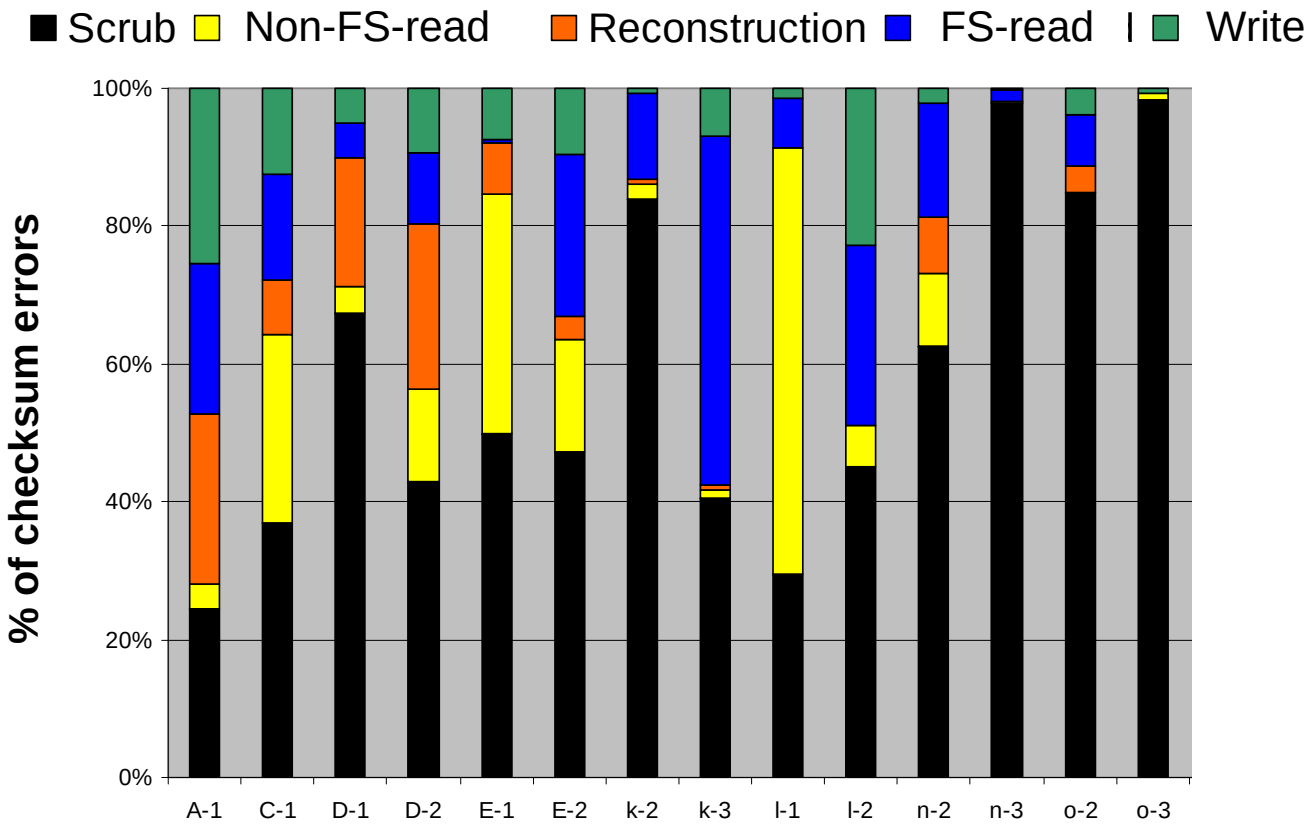
Spatial Locality – Nearline

What fraction of corrupt blocks have a corrupt neighbor within a radius of X blocks?



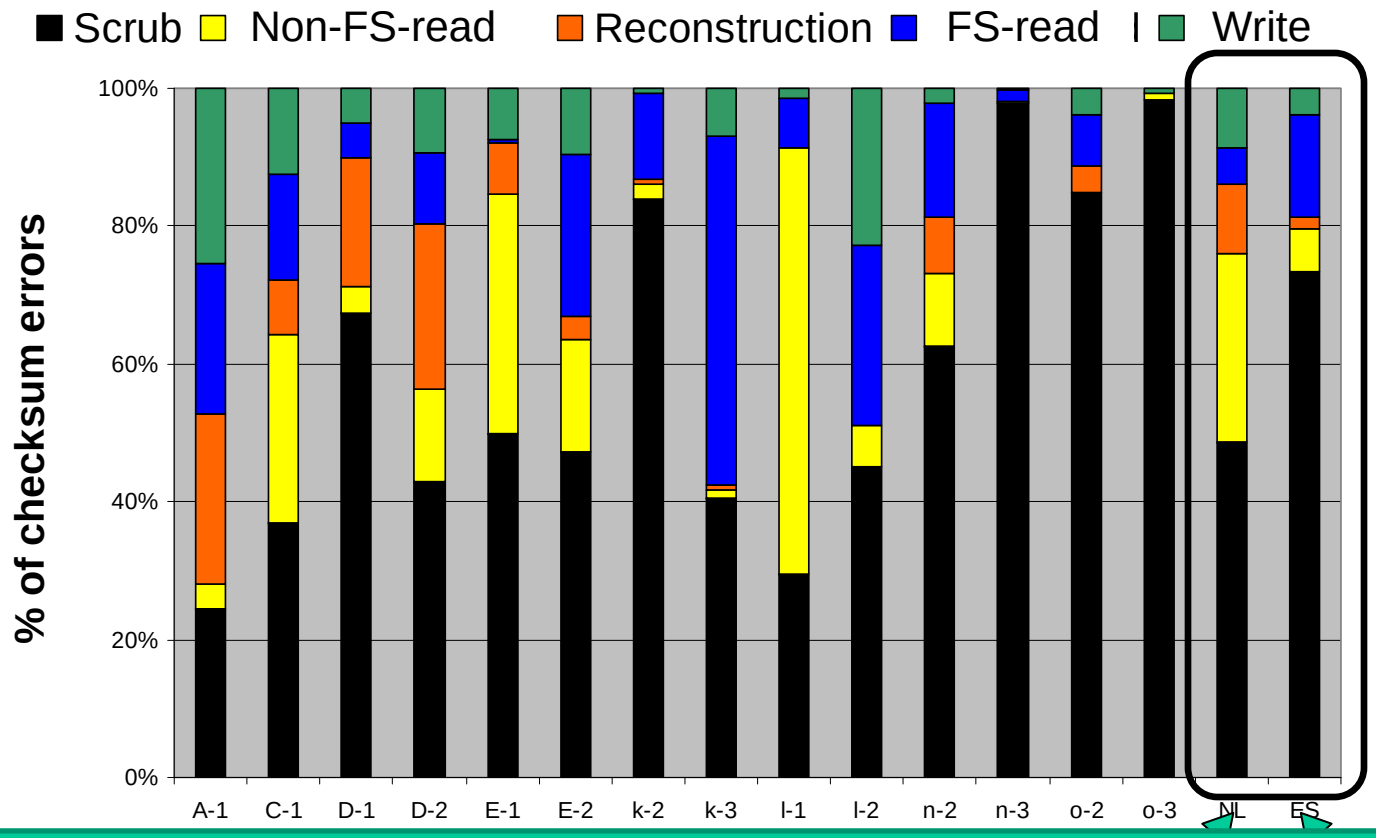
- High spatial locality for very small radius
- Low spatial locality for higher radius
- Very similar behavior for nearline & enterprise drives

How are corruption events detected?



<Bianca Schroeder> © August 08

How are corruption events detected?



- Majority detected during scrubs
- Significant number detected during reconstruction!
  - (8% for nearline drives)

Summary

- Silent data corruption happens!
  - More than 400,000 instances in our study
  - For nearline drives, 8% discovered during RAID reconstruction
  - Nearlines drives are affected an order of magnitude more often than enterprise
  - Affected enterprise drives develop more corruptions than nearline drives
- Strong spatial locality
- Strong dependence in time
- Next: design lessons?

<Bianca Schroeder> © August 08

Thank you!

Design lessons?

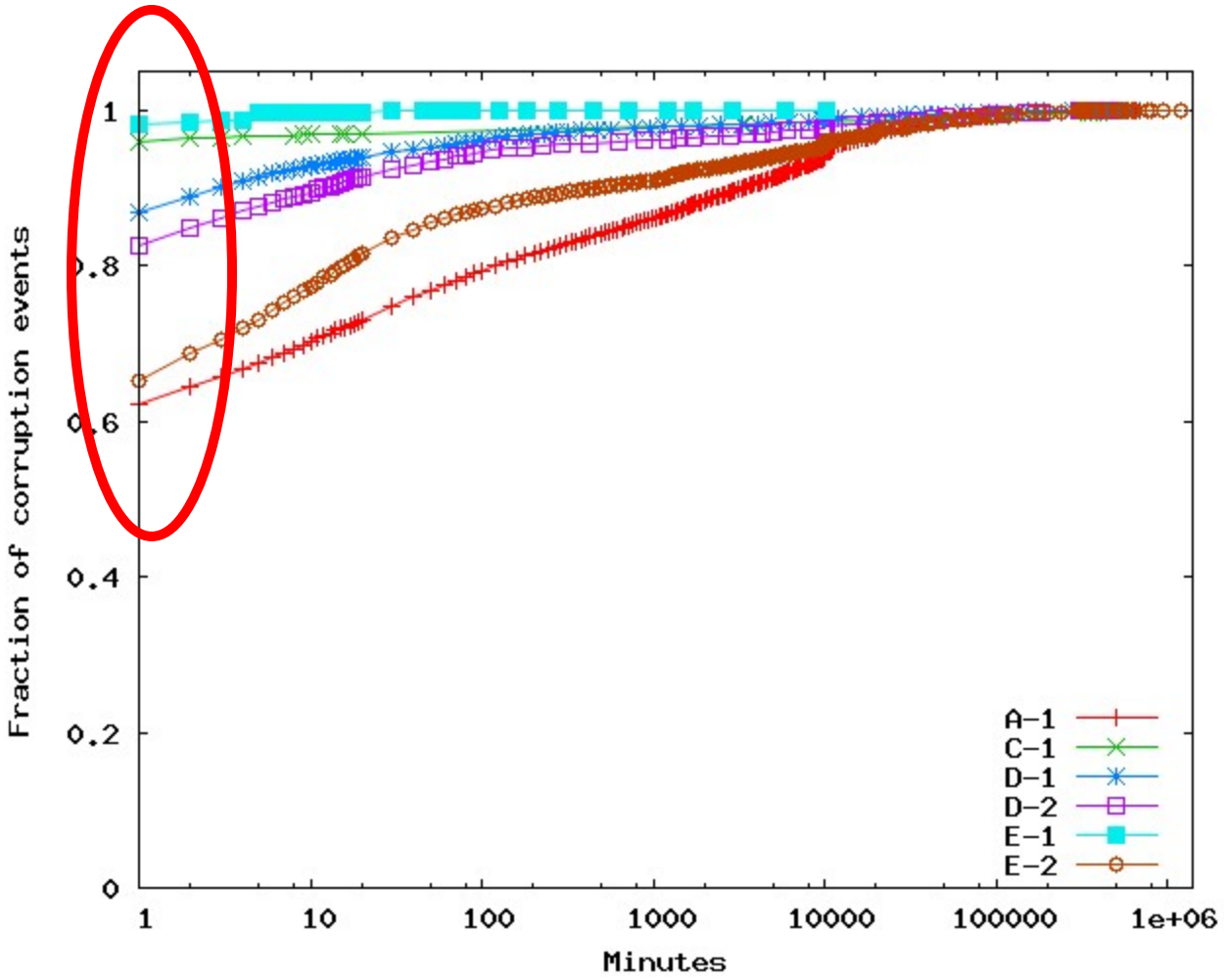
- Silent corruption does occur
  - Checksum protection is well-worth the space and performance overhead
- Very few enterprise disks develop corruption
  - “Fail-out” the disk when first corruption is detected
- High temporal & spatial locality
  - Write redundant data at different times
  - Smarter scrubbing?
- Corruption detected during reconstruction
  - More aggressive scrubbing?
  - Smarter scrubbing?

<Bianca Schroeder> © August 08

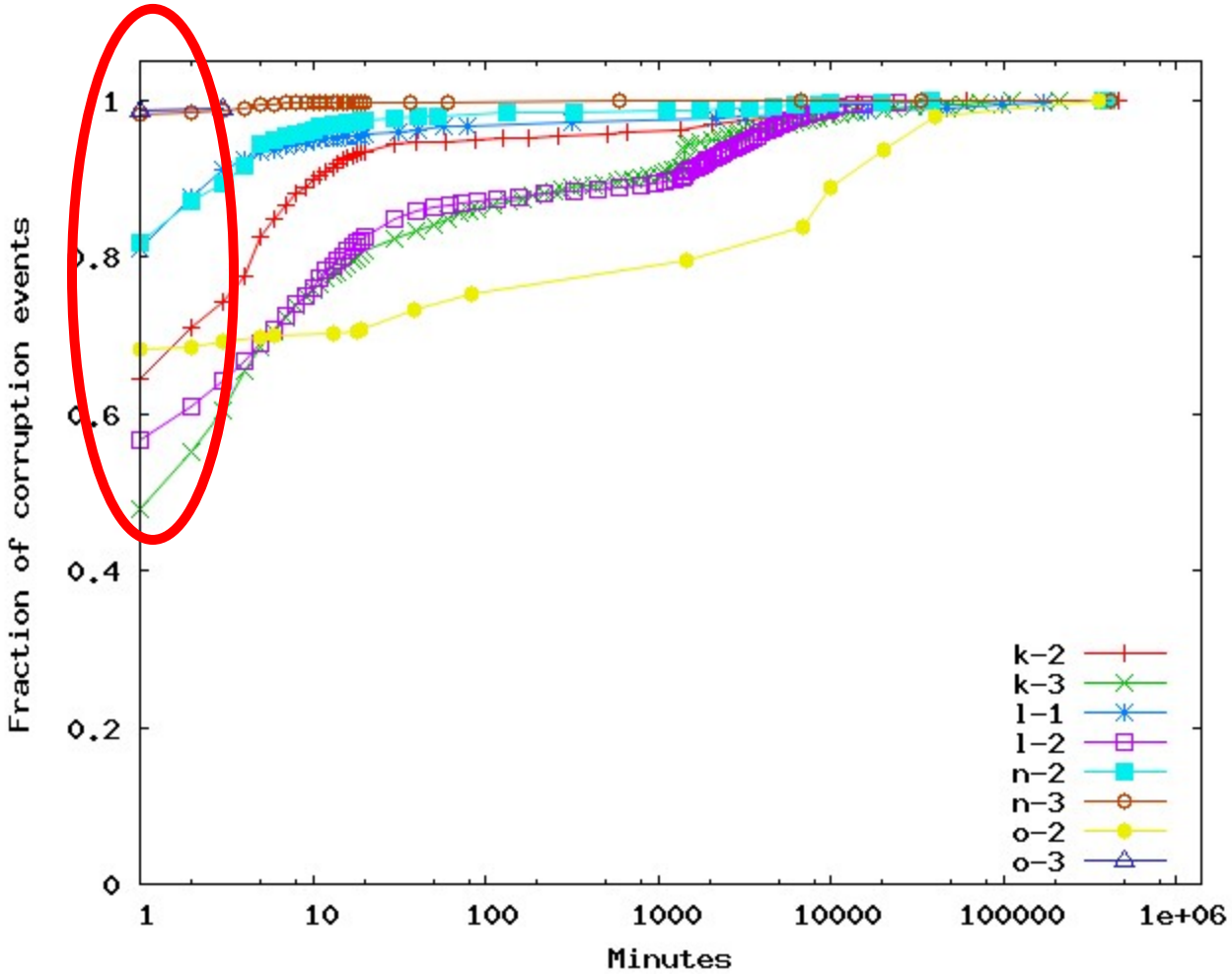
Corruptions detected in other ways

- Identity Mismatch (Lost writes)
  - Order of magnitude less often than random corruption
- Parity Inconsistencies
  - About 5 times less often than random corruption

Temporal Locality (Inter-arrival Time) - Nearline

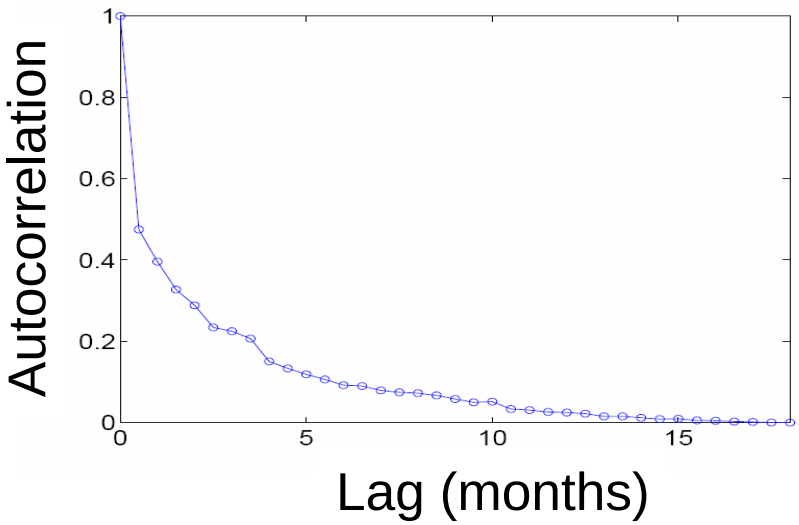


Temporal Locality (Inter-arrival Time) - Enterprise



Temporal Locality

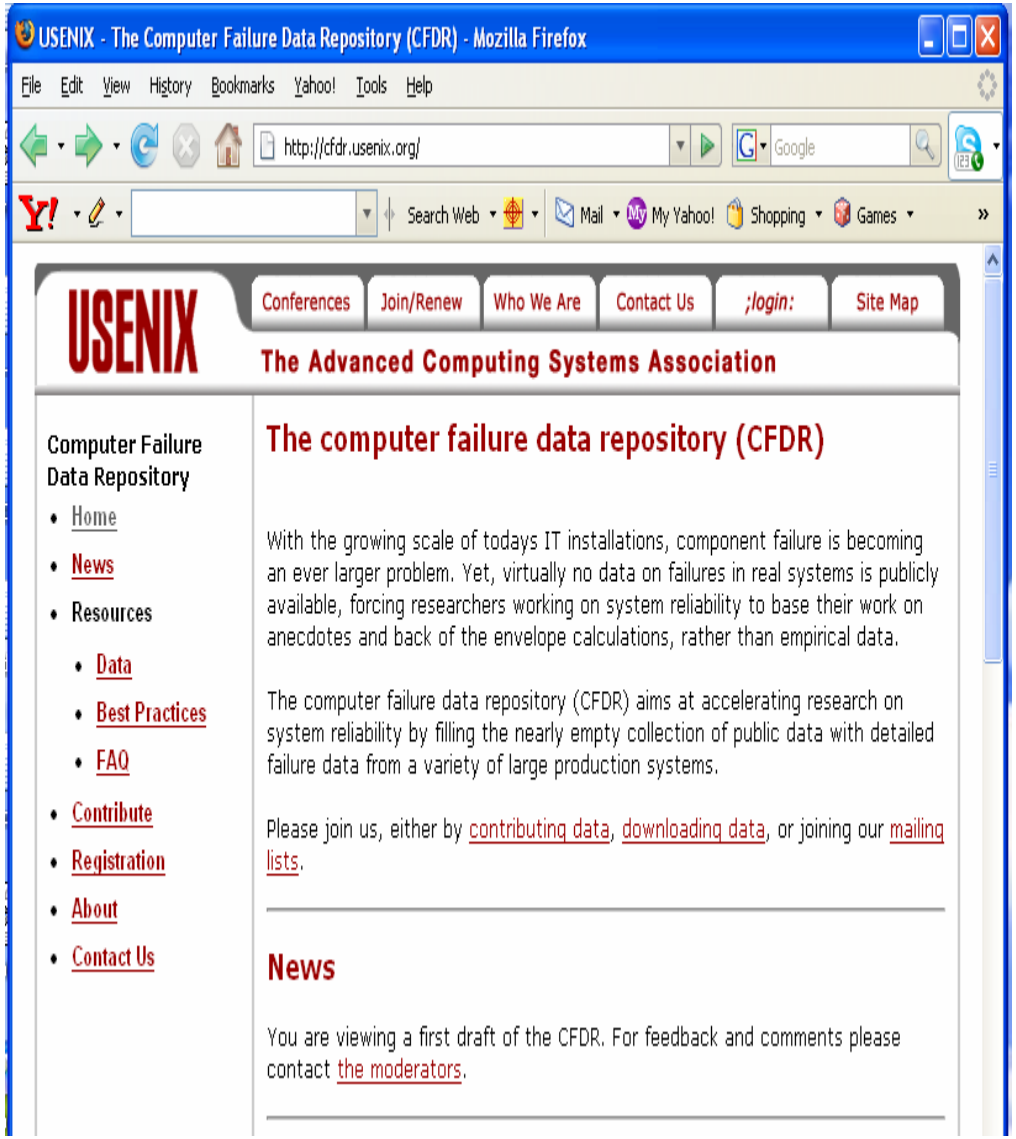
- High temporal locality
  - **<u>But:</u>** Reflects the fact that the errors were discovered around the same time
- Study temporal locality over longer time periods to remove effect of detection time
  - Test auto-correlation over 2-week bins



Temporal locality exists beyond the effect of detection time.

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



Available data

- Downloaded 900 times in 6 months
- Used in at least 3 SC'07 papers

9 years of node outages
[DSN'06,TDSC]
[SciDAC'07]

Error logs
[DSN'07]

I/O specific failures

Hardware / disk drive failures
[FAST'07, TOS]

Na			
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. Some job 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.

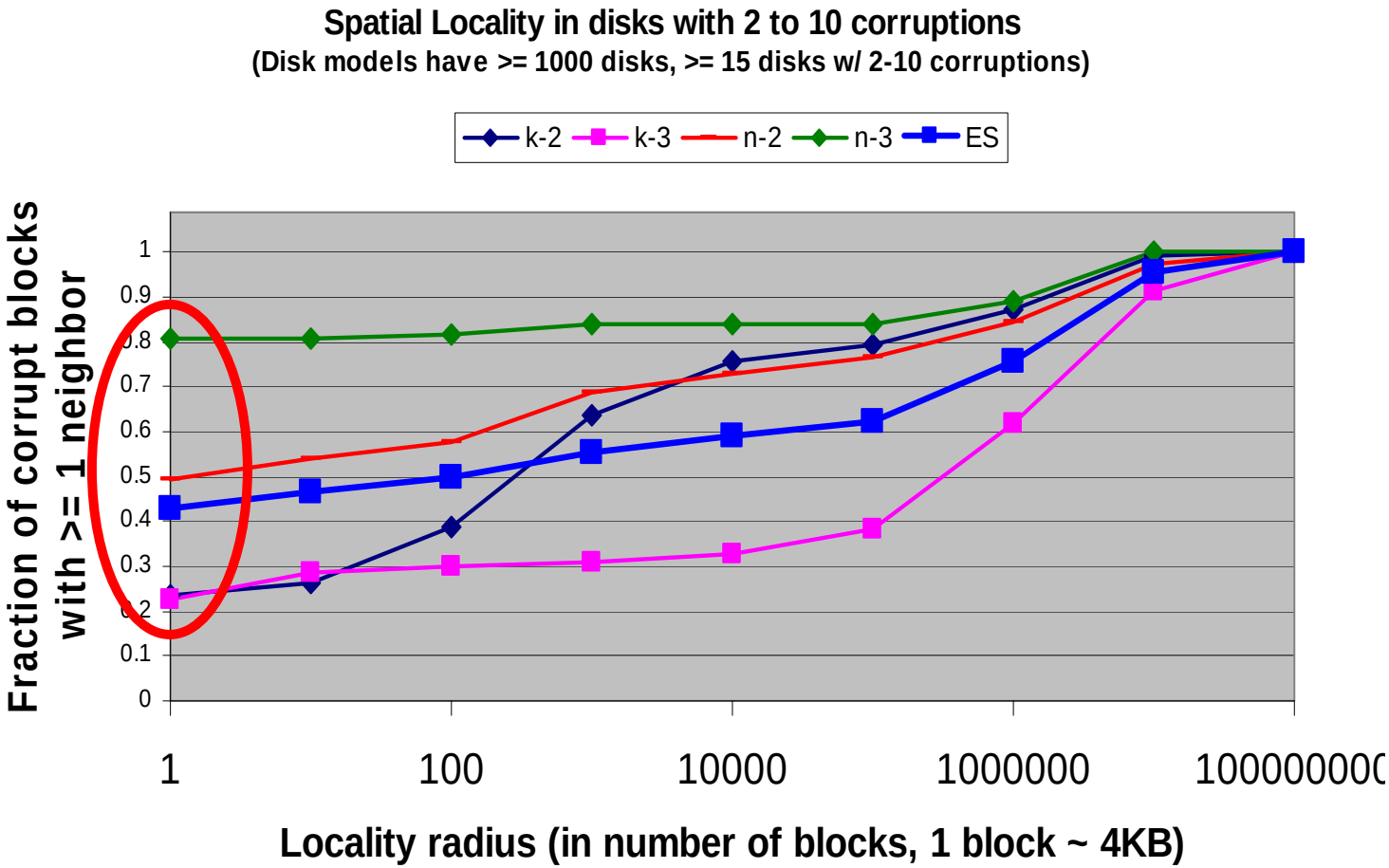
Data not available (yet?):

- [FAST'07 Google] study of hard drive replacements
- [Sigmetrics'07 NetApp] study of media errors

Corruptions per Corrupt Disk

- Big differences between disk models
  - 2 orders of magnitude difference in median
- Nearline somewhat better than enterprise drives
  - Median of 2 versus 10 corruptions
  - 80th percentile of 20 versus 100 corruptions
- Some disk models can be really bad
  - Model E-1: 3% of disks have corruption and 25% of those have > 1000 errors (all within 17 months)

Spatial Locality – Enterprise



Spatial Locality

- Bi-modal behavior
  - High spatial locality for very small radius
    - 50% of corrupt blocks have adjacent block corrupt
  - Low spatial locality for higher radius
- Very similar behavior for nearline & enterprise drives

How are corruption events detected?

- Majority of corruptions detected by scrubs
  - 50% of corruptions in nearline drives
  - 73% of corruptions in enterprise drives
- A significant number detected during reconstruction
  - In particular for nearline drives (8% on average)
  - 20% for some drive models