CERIAS

the center for education and research in information assurance and security

Strengthening Distributed Digital Forensics

Jeremiah Nielsen

Some Problems

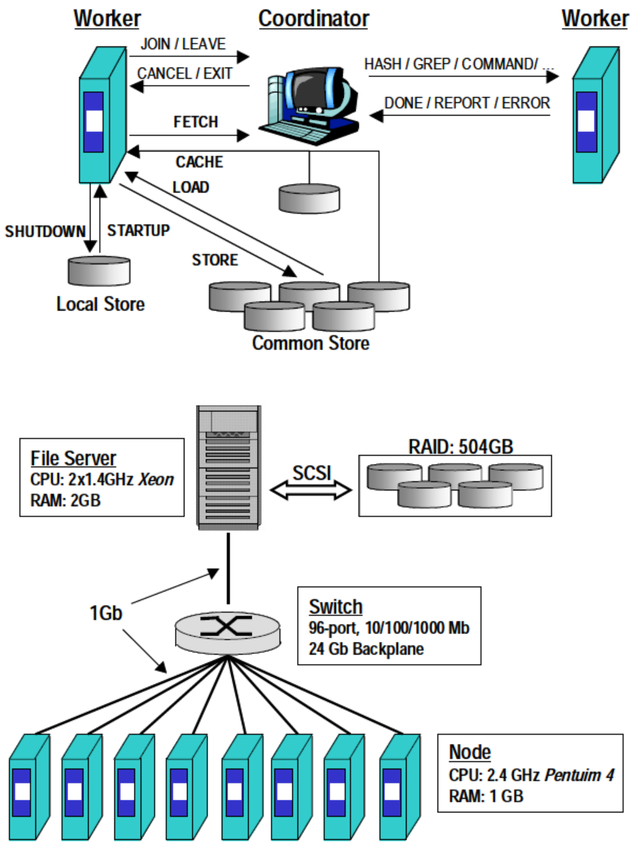
- File sizes continue to increase
  - 1080p Blu-Ray images 4GB – 11GB+ per
- Its OK, hard drives are massive and cheap! (3TB for $180)
  - Cheap easy to use NAS devices to (4TB for $340)
- Current tools were not created with these sizes in mind
  - Still utilize single work station processing
- Analysis processes are inefficient
  - Must capture everything and analyze everything

Potential Solutions

- Wait for SSDs to replace all magnetic disks
  - Magnetic disk of the future?
- Selective digital forensics using known goods/bads
  - Fresh Windows 7 x64 install takes 22GB
- Combine static and live analysis methods
  - Can help to pin point items of interest
- Develop more intelligent image capture and analysis
  - Apparently not there yet, still using FTK / Encase

Distributed Digital Forensics Prototype

(Richard III & Roussev, 2004)



The Temporary Band-Aid Solution

- Decrease dataset sizes using intelligent imaging and analysis
  - Nothing yet but still need to do investigations
- Why not spread the analysis load across several machines?
  - Distributed Digital Forensics!
  - Analogous to a criminal investigation in that resources are added to speed completion (diminishing returns?)
- Could also parallelize apps such as FTK
  - Create split image on SAN device and have workstations index specific pieces of large image

Prototype Results

(Richard III & Roussev, 2004)

	Search time: String Expression (mm:ss)	Search time: Regular Expression (mm:ss)
FTK	08:27	41:50
8-node System	00:27	00:28

Initial Operation	Time (hh:mm:ss)
FTK "Open"	1:38:00
CACHE	0:09:36
8-node LOAD	0:03:58
1-node LOAD	0:05:19

Strengthening Distributed Digital Forensics

- Research community is dealing with PB datasets (Hadron Collider) (NASA)
  - Why is digital forensics finding it difficult to deal with TB datasets?
    - Inefficient imaging and analysis approaches
- What constitutes an effective digital forensics network?
  - Scalable, reliable, high speed, secure, and needs little administration
- Apply data intensive computing research to digital forensics
  - Use a reliable file transfer protocol such as GridFTP
  - Use high speed RAM pools for storage
  - Use peer to peer VPNs for security
  - Use super peers for increased reliability
  - Use data management frameworks for security / reliability
- Use resources from existing underutilized machines
  - No need to invest in dedicated distributed digital forensics infrastructure

Prototype Issues

- Only a small image was utilized for testing due to age of the article
  - 6GB image loaded completely in RAM of nodes
- Resource saturation means machines can only be used for DDF
  - Expensive network / resources required if a grid does not exist
- No methods to address node reliability and scalability
  - Needs to be dealt with at the software level but also network level
- A homemade clear text message protocol was used as an MPI
  - Insecure, as messages can be captured and injected
- Insecurities justified through the use of a private network
  - Potentially financially infeasible and could conflict with exiting infrastructure

References available upon request

PURDUE UNIVERSITY

CERIAS

Discovery Park e-Enterprise Center