Information Assurance in a Distributed Forensic Cluster

• Nick Pringle\textsuperscript{a*}, Mikhaila Burgess\textsuperscript{a}

\begin{itemize}
\item \textsuperscript{a} University of South Wales (formerly University of Glamorgan), Treforest, CF37 1DL, UK
\end{itemize}
• This is a short presentation of the work presented at DFRWS Europe 2014
• [www.fcluster.org.uk](http://www.fcluster.org.uk)
• PhD published at the end of the year
Operation Big Wing, 24th April 2014
3,300 Metropolitan Police Officers
Targeting Co-ordinated arrests of 630 persons across London as burglary and theft crackdown

National Crime Agency
Operation Notorise resulted in 660 persons arrested in regards child abuse.
Specifically 9,172 devices seized.
Source media
- Write Protected Copy

Our Media
- Copy

8GB USB 2 Stick
- Image on Local harddisk
- 2 minutes

3TB hard Disk 75 Mbyte/s
- Copy
- 11 hours

Our Media
- Copy

SATA III but HD 85MB/s
- 2 MBytes/sec per core
- 20 minutes

12 cores

Results Database

Analysis Reporting

3TB hard Disk 75 Mbyte/s
- Direct Write Protected Copy
- 11 hours

SATA III but HD 85MB/s
- 2 MBytes/sec per core
- 70 hours

12 cores

All Cores at 100%

Results Database

Analysis Reporting

3TB hard Disk 75 Mbyte/s
- Direct Write Protected Copy
- 11 hours

SATA III SSD at 450 MB/s
- Still 2 MBytes/sec per core
- 70 hours

12 cores

Results Database

Analysis Reporting

No Control over Source device speed

$\$\$\$ Expensive

Processor Bound
3TB hard Disk 75 Mbyte/s

11 hours

Direct Write Protected Copy

SATA III SSD at 450 MB/s

2 MBytes/sec per core

13 hours

Xeon 64 cores

Results Database

Analysis Reporting

$\$\$\$ Expensive! $150,000? Unacceptable

SGI Altix 4700?

3TB hard Disk 75 Mbyte/s

11 hours

Direct Write Protected Copy

SATA III SSD at 450 MB/s

Gigabit Network

100 MB/s = 10 hours

10 Gigabit Network

1000 MB/s

$\$\$\$ Expensive

For 25 W/S $25,000

2 MBytes/sec per core

25 x i7 = 300 cores

2 hours

Results Database

Analysis Reporting
This problem is not going away. It’s going to get worse!

<table>
<thead>
<tr>
<th>RCFL Figures</th>
<th>2004</th>
<th>2012</th>
<th>%</th>
<th>2020?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examinations</td>
<td>1304</td>
<td>8566</td>
<td>657</td>
<td>56270</td>
</tr>
<tr>
<td>Total Volume Examined</td>
<td>229TB</td>
<td>5886 TB</td>
<td>2570</td>
<td>151 EB</td>
</tr>
<tr>
<td>Average Case</td>
<td>175GB</td>
<td>680 GB</td>
<td>300</td>
<td>2.6 TB</td>
</tr>
</tbody>
</table>
Latency, Multi-threading and Parallel Processing

Task setup  
Processing  
Task closure

Linear Processing

Multi-Threaded/Parallel/Distributed Processing
We lose “Chain of Evidence”

• We’re not longer using a simple system with one file store and a few PCs

• In this world of distributed storage and processing we need to revisit and re-establish “Chain of Evidence” within the computer system

• We’re back a decade and can’t move on until we do
Jigsaw Imaging

• Reads the Source media and follows the file system (not sector by sector)

• Creates Digital Evidence Bags for each file while simultaneously creating the conventional image
Data Acquisition Triage

- Uses a Bayesian approach that directs the Jigsaw Imaging process to prioritise files considered to be of higher likelihood of yielding evidence.
Fclusterfs
A file system for Digital Evidence Bags

A FUSE file system that:

• Stores the original file meta-data in the file system
• Gives access to files stored as whole, encrypted DEBs
• Has access control by user and file system and file
• Is Read Only
• Logs movement of, and access to, data
• Allows (most) unaltered legacy software
• Allows non-parallel-aware software to run across multiple nodes
Why is this the right approach?

• This could be achieved within an application program but each package would have to implement it and gain approval.
• Working at file system level the efficacy is global
• Interaction with FClusterfs is unavoidable
• FClusterfs controls data access and maintains Assurance
Funded by...
Information Assurance in a Distributed Forensic Cluster Questions?