

Change-Link: A Digital Forensic Tool for Visualizing Changes to Directory Trees

Change-Link: A Digital Forensic Tool for Visualizing Changes to Directory Trees

Timothy F. Leschke
Cyber Defense Lab
University of Maryland, Baltimore County
Baltimore, MD 21250
leschke1@umbc.edu

Alan T. Sherman
Cyber Defense Lab
University of Maryland, Baltimore C
Baltimore, MD 21250
sherman@umhc.edu

ABSTRACT

[illegible]

Keywords

coincident and multiple views, linked view, change over time, overview+detail, digital forensics, data visualization.

INTRODUCTION

Digital fernandes involve entrusting and analyzing data from digital artifacts in support of law enforcement investigations. In digital forensics, the words data exposure is used to describe the rapid growth in recent years of the amount of data that is subject to a digital forensic examination. This data exposure is the increased number and depth of the stored data of computer hard drives, and the proliferation of smart phones, GPS receivers, and other portable devices that have growing digital storage capacities. As a result, someone engaged in criminal activity can bring sophisticated digital forensic tools to the investigation and use them to a variety of digital forensic examinations has become as manageable. We design, implement, and demonstrate a new data visualization tool that helps address this data explosion.

One way to keep pace with the data explosion is to increase the bandwidth by which digital evidence is examined. This is the focus of the research presented in this paper. We describe through video the data visualization tool that we have

Categories and Subject Descriptors

© 2012 Association for Computing Machinery. ACM acknowledges that this contribution is authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the United States Government retains a nonexclusive, irrevocable and exclusive right to publish or reproduce this article or to allow others to do so, for Government purposes only.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the identification on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ocls support the accounting of individual shadow volumes and some provide an understanding of what changed between two selected shadow volumes. None of the known ocls support an understanding of change over multiple shadow volumes, and currently none seem well enough to capture change according to as many as 512 shadow volumes.



Timothy R. Leschke, M.S.
Doctoral Student
tleschk1@umbc.edu



Alan T. Sherman, Ph.D.
Associate Professor
sherman@umbc.edu

- Tim Leschke
 - Doctoral Student in Computer Science at the University of Maryland Baltimore County.
 - My research involves using data visualization techniques to support digital forensics. Digital forensics involves extracting and analyzing data from digital artifacts (such as computers and cell phones) in support of law enforcement investigations.
 - I am conducting my research under the supervision of my advisor, Dr. Alan Sherman.
 - Presenting a paper titled “Change-Link: A Digital Forensic Tool for Visualizing Changes to Directory Trees”
 - This paper is about a prototype tool which I developed called Change-Link, which supports a user’s ability to see visual representations of an operating system directory structure that has changed over time.
 - This paper represents a starting point for my research in which I am investigating technology that supports the browsing of digital forensic data, including large dataset and data that has changed over time.

Motivation

- Visualizing Change
- Digital Forensics
 - Data Explosion
 - Shadow Volume Data



<http://www.smeadvisor.com/wp-content/uploads/2011/07/Data-explosion.jpg>, <http://datastoragealongbeach.com/wp-content/uploads/2012/02/data-backup-Long-Beach.jpg>

2

• I am interested in visualizing change because it is an interesting academic exercise that is ripe with opportunities to conduct interesting research.

• I am also interested in visualizing change because I believe the technology can be used to address two issues within digital forensics. These issues are:

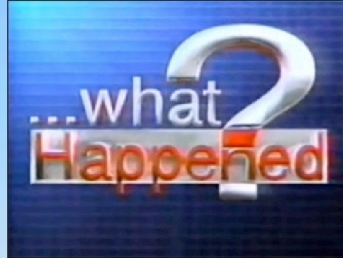
- Dealing with the “data explosion” that we are currently seeing in digital forensics and
- Dealing with Shadow Volume Data, which can be a significant issue for digital forensic examinations.

• The issue of the “data explosion” is the result of personal cell phones/smart phones, GPS navigation aides, portable computers, and other digital devices have become ubiquitous within modern society. Furthermore, with digital storage devices surpassing the terrabyte level, the amount of personal data that is associated with each of these devices has grown significantly. Because there is more data, there is more work to be done by digital forensic examiners.

• The issue of “shadow volume data” is an issue of both quantity and complexity of the data. As I will explain in greater detail later, certain computer systems contain “shadow volume data” which record multiple versions of file, application, and operating system data. The multiple versions of data makes analyzing this data more complex. Plus, because there can be many very large sets of shadow volume data, this extra quantity of data adds to the data explosion that forensic examiners are trying to keep pace with.

Benefits of Visualizing Change

- “...direct their efforts toward more important data.”



- “...digital forensic examiners are better able to understand what happened.”

<http://techmehigh.com/wp-content/uploads/2011/06/how-to-restore-3.jpg>, http://sharetv.org/images/what_happened-show.jpg

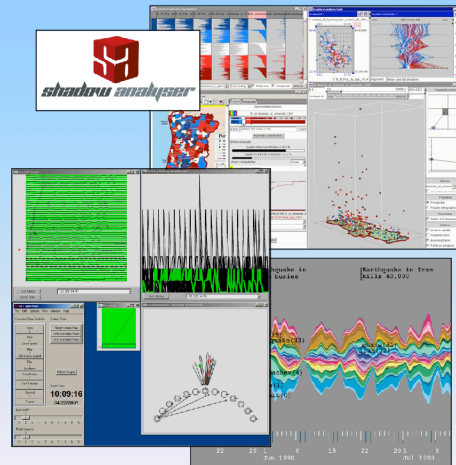
3

There are several benefits to visualizing change:

1. Because there is too much data that can be subject to a digital forensic investigation, examiners cannot look at all of the data. A better approach might be to look at the most important data first. I believe data that has changed within a computer system is the most important data. For example, the fresh installation of an operating system is generally not as interesting as how that operating system data may have been changed by alleged criminal activity. Thus, by using data visualization techniques to identify data that has changed, digital forensic examiners might be able to complete their examinations with a smaller sub-set of the evidence and thereby keep pace with the recent “data explosion.”
2. A second benefit of visualizing change is it will help the digital forensic examiner understand how the data was used and changed to support criminal activity. This supports a better understanding of “what happened.”

Related Work

- Shadow Volume Tools
- Coordinated and Multiple Views
- Visual Analytics
- Linked Views
- Overview+Detail
- TreeJuxtaposer & Mizbee
- Visualization of Change
 - Few address *coming into* and *going out of* existence.



<http://www.prlog.org/10766636-shadow-analyser-computer-forensics-tool-for-digital-forensic-investigators.jpg>, State of the Art: Coordinated & Multiple Views in Exploratory Visualization by Jonathan C. Roberts, Hierarchical Linked Views by Erbacher and Frincke, Themriver: Visualizing thematic changes in large document collections, by Haime, Hetzler, Whitney and Nowell.

4

There is significant related work to this research.

First, within digital forensics, there are many tools that support the accessing of shadow volume data. These tools include

- Shadow Scanner
- Shadow Explorer
- Shadow Analyzer
- Time Traveler, and others

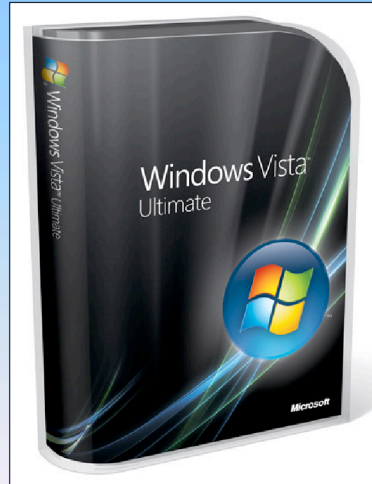
None of these tools provide an overview of change, and none support an adequate understanding of change over more than two shadow volumes at a time.

Within the data visualization community, the previous research that has influenced this research includes:

- Coordinated and multiple views – using several windows to show multiple levels of detail
- Visual analytics – attempts to make large dataset tractable through visualization
- Linked views – which allows the manipulation of one view to be reflected in another view.
- Overview+Detail – which helps the user retain the context and browsing position within a large dataset.
- TreeJuxtaposer – which provides a nice visualization of how a tree structure (similar to our directory tree) has changed from one time period to the next.
- MizBee – which inspires us to explore relationship among data points.

Shadow Volume Data

- Volume Shadow Copy Service
 - Windows Vista, Windows 7, and others
- When data is archived
 - Backup utility
 - Prior to installation
 - Restore point
- Why data is archived
 - Rollback data to restore stability and recover lost data.



http://home.comcast.net/~SupportCD/Images/Windows_Vista_Ultimate_Box_Hires.jpg

5

The data set that Change-Link works with is derived from shadow volume data, which is a product of the Volume Shadow Copy Service. This is a service that is found in several Microsoft Operating Systems, but perhaps most notably in Windows Vista and Windows 7.

- Intended to prevent accidental data loss (due to user error)

The Volume Shadow Copy Service archives data

- When configured to do so (nightly, or weekly)
- Prior to new software installation
- At the creation of a restore point

This archiving of data allows a system to be rolled back to a previous state

- To recover lost data
- To put the system back into a stable state if a virus or the incorrect installation of software has made the system unstable.

Test Data

- Eight shadow volumes
 - After installing Vista
 - After activating Vista
 - During installation of SP1
 - After installing SP1
 - After installing Office
 - After activation of Office
 - After creating directories
 - After deleting directories
- 11,000 directories x 8 shadow volumes = 88,000 data points.
- 700,000 directories x 8 \approx 5 million data points!



http://laptoping.com/wp-content/vista_and_office2007.jpg

6

I am working with law enforcement to test future tools with real evidence/case data, but for the Change-Link prototype tool, I created a simple test dataset to work with.

On a Windows Vista machine, I created a shadow volume (my manually creating a restore point)

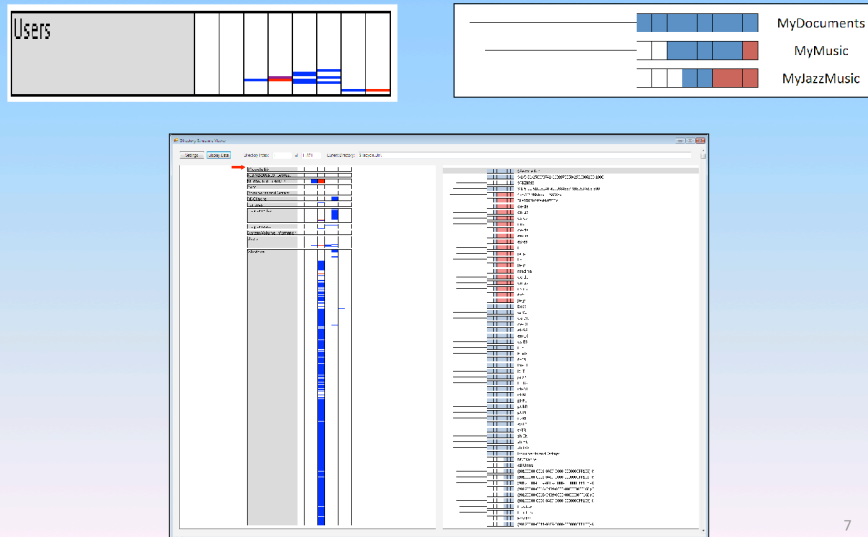
1. After installing Vista
2. After activating Vista
3. During the installation of Service Pack 1
4. After the installation of Service Pack 1
5. After the installation of Microsoft Office
6. After the activation of Microsoft Office
7. After creating some user directories
8. After deleting some user directories

By including just user-level directories in my dataset, my dataset had 11,000 directories, in each of the 8 shadow volumes, for a total of about 88,000 data points.

When I included all directories (including system level directories), my dataset had 700,000 directories in each of the 8 shadow volumes, or about 5.6 million data points.

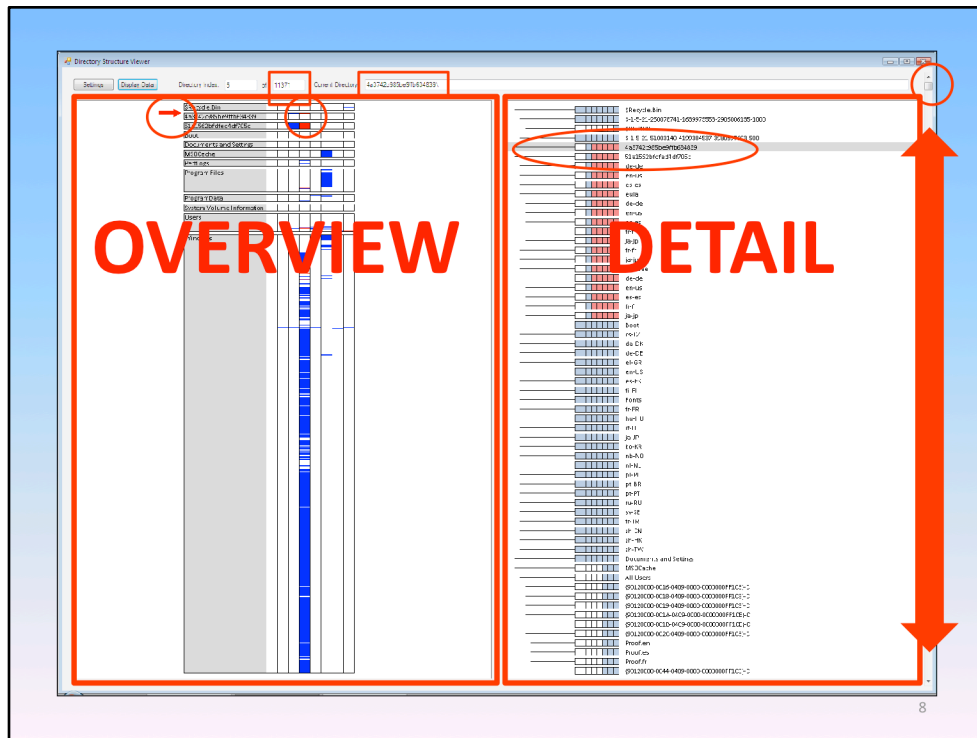
For simplicity, I conducted my research with the smaller (88,000 data point) data set.

Design and Implementation



As I explain the Design and Implementation of Change-Link, I will be showing the user interface (which is seen at the bottom center of this slide), and also zooming into each of the left-side and right-side windows to show the details of those visualizations as shown by the pictures at the top left and top right of this slide.

Lets first look at the user interface as a whole.



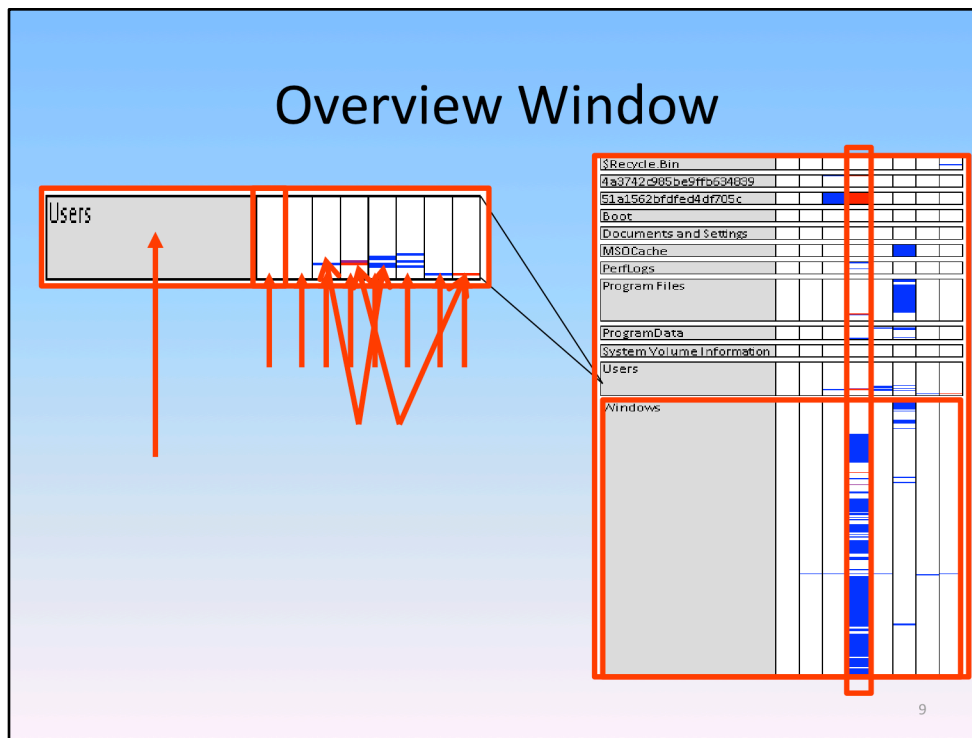
This is the Change-Link user interface.

There are two main windows:

The left window provides an overview and the right window provides the detail.

As the user clicks with a pointer in the left overview window, a red arrow moves to mark that point in the overview visualization, and the right side windows scrolls the data to reflect the details of the part of the dataset that has been chosen.

Lets zoom into the left side window first to get a better understanding of the overview visualization.



The picture on the right of this screen is the overview visualization. What this shows is a visual representation of the operating system directories that are located at the root of the directory. Thus, we see directories with names like \$Recycle.Bin, Boot, program Files, User, and Windows (among others).

To the left is an enlarged view of the Users directory.

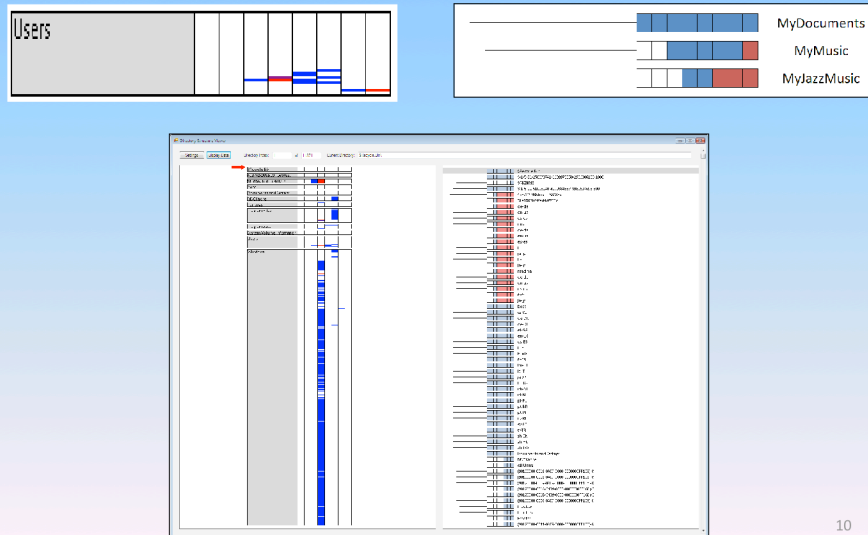
Each directory is colored grey, has its name on it, and contains eight (8) white partitions – one for each of the eight time periods that correspond to the eight (8) shadow volumes.

The blue and red lines that appear in these segments represent the approximate location within the sub-directory of that directory where a directory was either created (blue) or deleted (red).

The height of each directory reflects the number of sub-directories within that directory.

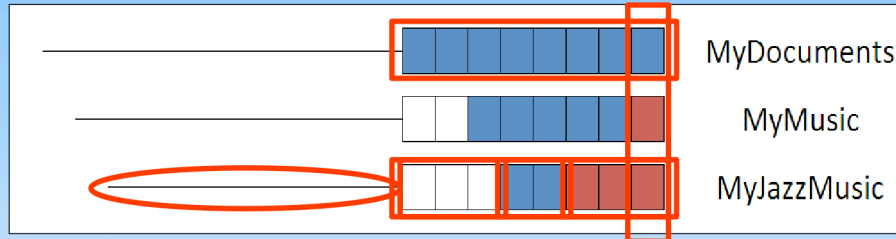
Thus, the overview shows the Windows directory has the most sub-directories, and also had the most created (blue lines) directories in time period 4. Time period 4 is also when the most directories were deleted as seen in the third directory from the top. Areas of concentrated creation and deletion of directories are easily identifies, as are single creation and deletions of directories.

Design and Implementation



If we zoom into the right visualization window, we see a glyph we developed called a segmented box and whisker. This glyph is used to show the change history of an individual directory.

Segmented Box and Whisker



11

In this example, there are three segmented box and whisker glyphs.

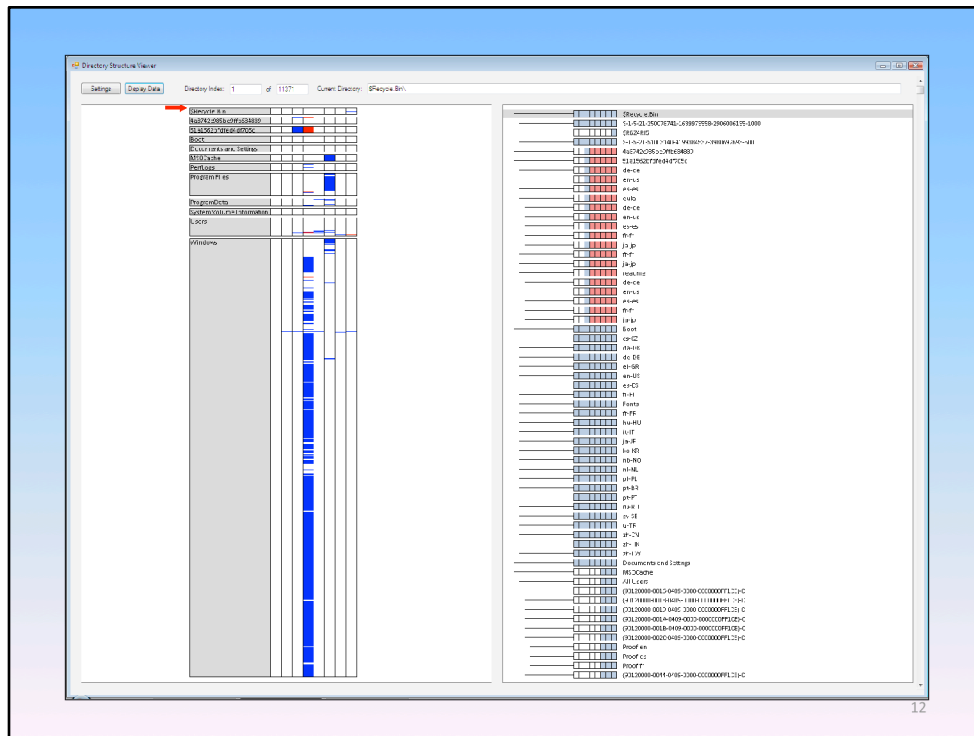
The thin black line that extends to the left of each glyph is the “whisker.” and it is used to represent the nesting of a child directory under a parent directory. In this example, the MyJazzMusic directory is a child directory of the MyMusic directory, which is a child directory of the MyDocuments directory.

The segmented box and whisker glyph contains eight (8) boxes or segments – one for each of the eight time periods of the eight shadow volumes. They are by decreasing age from left to right. Thus, the oldest time period is represented by the left most box and the most recent time period is represented by the rightmost box.

The boxes are colored blue for when the directory does exist, red for when the directory does no longer exist, and white if the directory does not yet exist. For example, the MyJazzMusic directory did not exist for the first three time periods, existed for the next two time periods, and then did not exist for the last three time periods.

By ordering the glyphs so that the time periods are aligned with each other, it is easy to compare the existence of the directories for any time period. For example, in the last time period, only the MyDocuments directory existed.

The name of each directory is provided to its right.

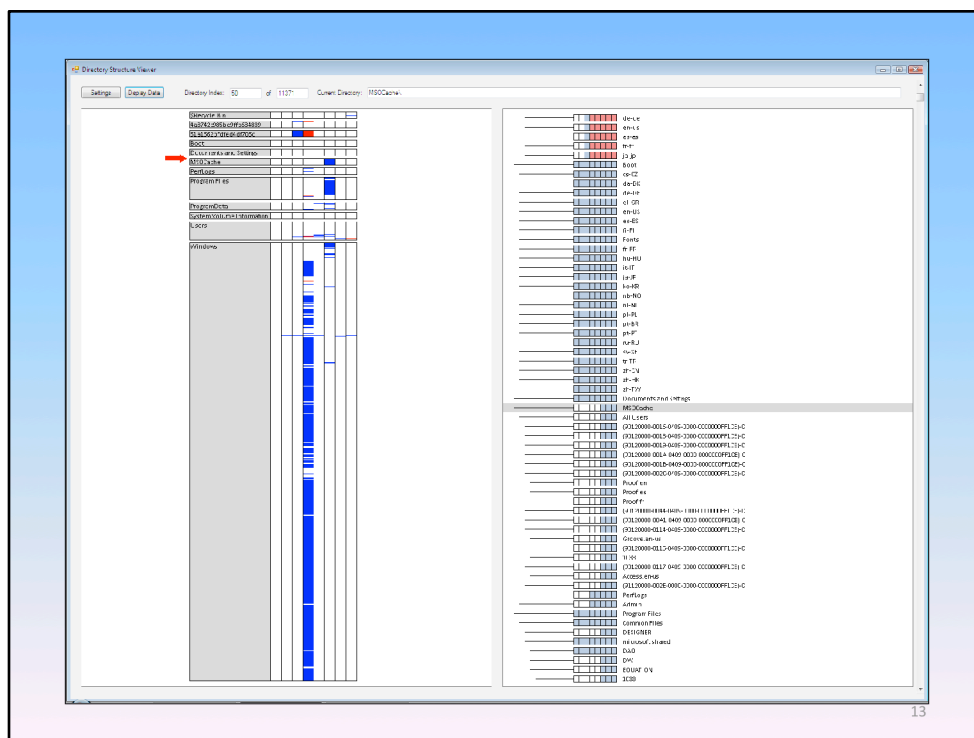


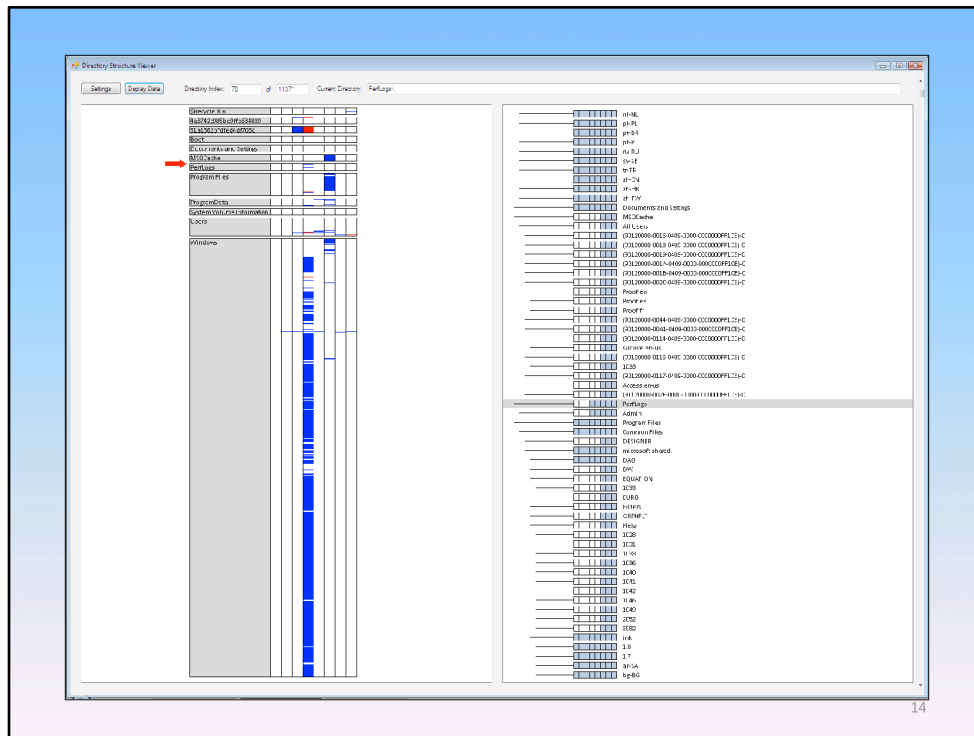
Next, I am going to flip forward and back with the next six (6) slides to simulate how both the right and left side pictures change with user interaction. You can assume the user is clicking on specific change lines found in the overview window and the right side is scrolling to that data points location.

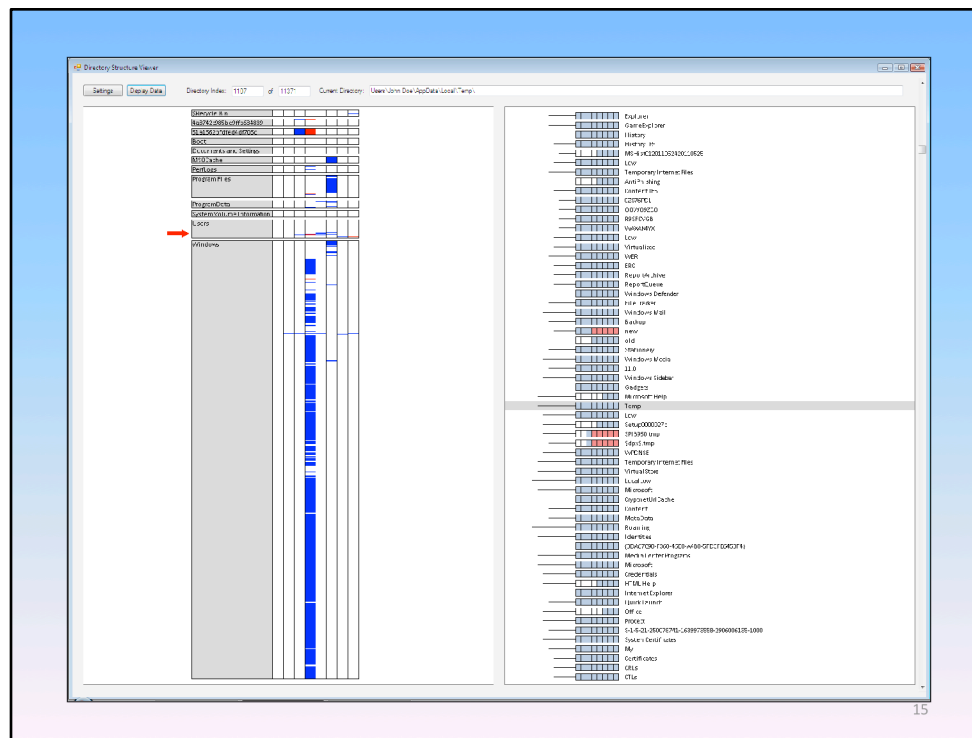
As the user clicks on a change line, the red arrow in the left overview window moves to point to the general location of the change line that was selected.

The right side will appear to change by scrolling to the selected data point.

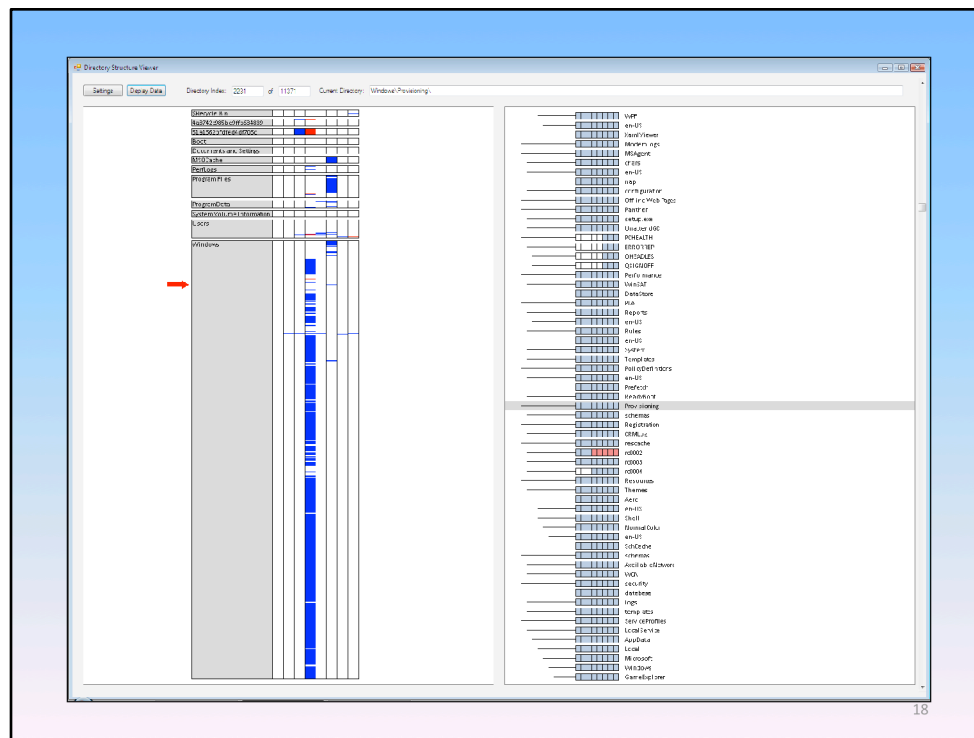
A grey focus bar highlights the exact directory that has been selected. The visualization is configured so that this grey focus bar stays in the center of the visualization, whenever possible.



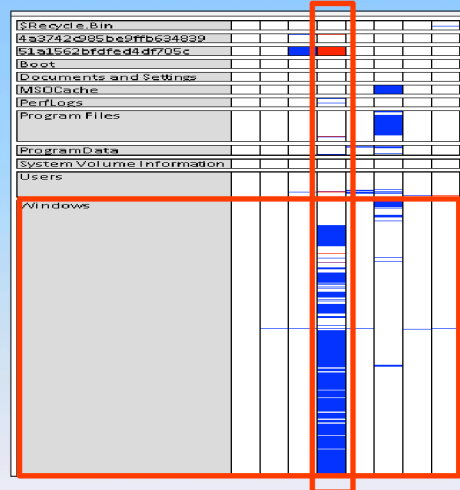








Results



Overview of all Directory Change

19

This picture shows an overview of our dataset.

Time period 4 has the most change (both blue and red lines).

Time period one has no change (and never will, since there is no time period before it that it can be compared to)

The Windows directory appears to have about half or two thirds of all of the sub-directories.

The most deleted directories are in time period 4, in the third directory from the top.

Etc.

Results (2)



Installation of Service Pack 1

20

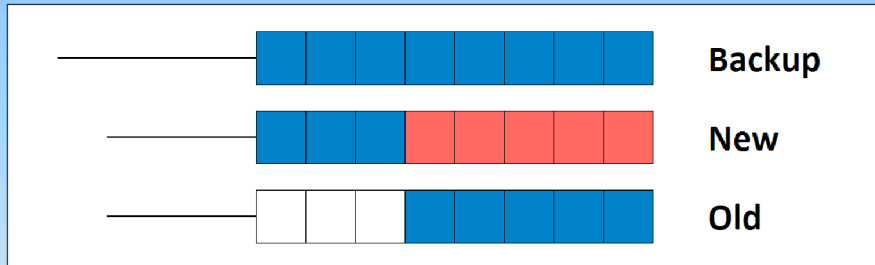
These pictures show changes that occurred during the Installation of Service Pack 1. This event occurred in time period 3 and time period 4.

Looking at the picture on the left, we see in time period three, about 19 directories were created in this part of the directory structure. Then in time period 4, the boxes are red which means the directories were deleted. View as a single event, the left picture shows Service Pack 1 creating backup folders to store data temporarily, to be used in case the installation is not successful and the installation has to be rolled-back.

The picture on the right shows an interesting event. Near the top, the fourth directory from the top, there is a directory named "Backup". Below it is a directory named "new" which existed for the first three time periods and is then deleted. Below that directory is a directory named "old", which is created in time period 4 and continues to exist for the duration of the next five time periods.

Viewing this part of the picture on the right side as a single event, we hypothesize that Service Pack 1 is simply renaming the Backup sub-directory from "new" to "old." After all, it seems to make more sense that the backup directory should have a sub directory for storing "old" data rather than "new" data.

Renaming of a Sub-Directory



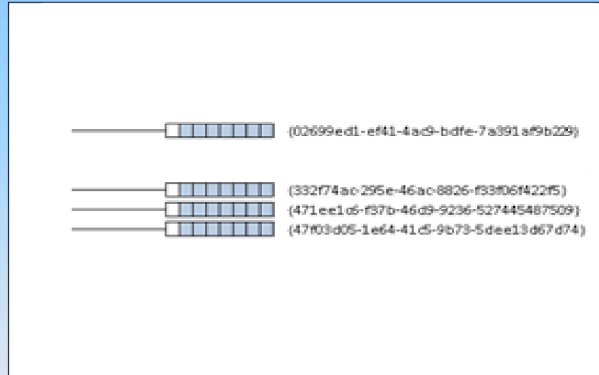
21

I redrew the visualization to make it easier to see.

This shows a directory named “Backup” that has a sub-directory named “New” for the first three (3) time periods and a sub-directory named “Old” for the last five (5) time periods. This seems to show that the “New” directory was renamed to “Old” at the transition from time period three (3) to time period four (4). This seems to mean that Service Pack 1 simply renamed the directory.

It makes more sense to have a sub-directory within the Backup directory named “Old” for perhaps old data rather than a sub-directory named “New.”

Results (3)



Activation of Vista

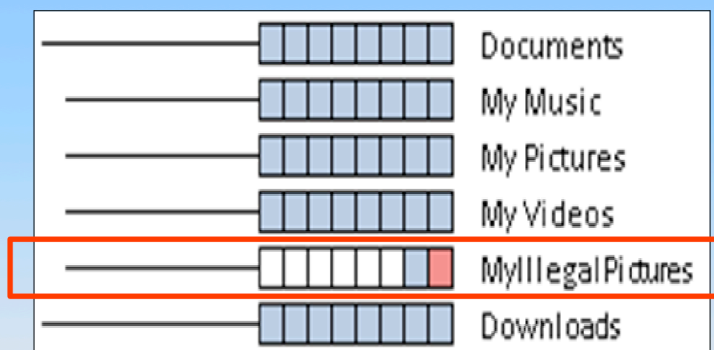
22

This shows a surprising event. The change reflected in this picture occurred during the period in which the Vista Operating System was activated. (the surrounding directories were whited-out so you can focus just on the four directories in the picture.)

This picture implies that the four directories shown above were created when Vista was activated. We expected the activation to maybe update a file. We are surprised to see that it may have created several directories.

Although this visualization cannot confirm a cause and effect relationship between the activation of Vista and the creation of the directories, like every good data visualization, this picture makes the user ask “why is the data like this.” this picture helps direct one’s attention toward an interesting change in the data that might require further examination.

Results (4)



User Created and Deleted Directory

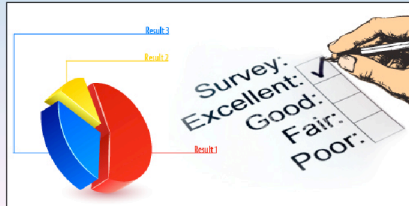
23

This is a simple example of a user creating a directory named “MyIllegalPictures” and then deleting it (after a record of it was archived into a shadow volume).

This example simulates what is commonly seen in a child pornography investigation. The person might create a directory to store their illicit images, and then delete that directory as a way to hide their activity. However, if the information is archived by the volume shadow copy service, a record of that illegal activity is retained.

User Reactions

- Users identified periods of...
 - Most change
 - Most created directories
 - Most deleted directories
 - Root directory with the most sub-directories
- Effectiveness of segmented box and whisker glyph
- Navigation support
- Comprehension
- Color
- Whisker
- “helps understand...”
- “discover information...”



<http://www.experiment-resources.com/images/survey-research-design.jpg>

24

I conducted an informal usability study with about seven (7) co-workers. I sat them in front of a computer that was running Change-Link, asked them to interact with it and become familiar with how it works, and then asked them a few test questions to verify if they understood the visualizations that they were seeing. Once they understood the visualization and how to use the tool, I asked them to identify certain patterns in the data. They were able to;

1. Identify the time period of most created directories
2. Identify time period of most deleted directories.
3. Identify time period of most change (which is most deleted and most created directories)
4. Identify directory located at the root that has the most sub-directories (done by comparing the height of the glyphs)
5. The users stated the segmented box and whisker glyph was effective for showing the child-parent relationship among directories, and also showing how the existence of the directory changes over time.
6. Change-Link was found to provide strong navigation support and support good comprehension of the data.
7. The colors that were used were good, with the exception that some users thought blue and red in one window mean the same thing in the next window.

Conclusion

- Change-Link helps
 - “focus on anomalies”
 - “find meaningful data more quickly”
 - Explore data more efficiently and effectively
- Companion to existing technology
- Future Goals
 - Add two additional views:
 - Overview
 - Tree View
 - Directory View
 - File View



<http://datacentricjob.com/wp-content/uploads/2011/10/data-science.jpg>

25

Change-Link does help the user understand how the directory structure has changed over time.

Anomalies were easy to spot and understand. Could better focus on anomalies.

Users could explore data more efficiently and effectively. Could find meaningful data more quickly.

Our initial results suggest

1. Visualizing data in this way might provide a better understanding.
2. Being able to see the most important data first (that being changed data) might help the digital forensic community deal with the data explosion.

Change-Link is a prototype of a more robust tool which is currently under development. The next generation tool will provide a linked-view approach to browsing an evidence hard drive. The next tool will include an Overview and Treeview (like Change-Link), but also include a Directory View and a File View. This will allow the user to see an overview of change over the entire data set, see how the directory structure has changed in a specific area, see how the contents of a specific directory has changed over time, and see the change history of a specific file. This all supports a better understanding of “what happened.”

Thank You



<http://www.turnbacktagod.com/wp-content/uploads/2008/12/questions.jpg>