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

Timothy R. Leschke
Cyber Defense Lab
University of Maryland, Baltimore County
Baltimore, MD 21250
tleschk1@umbc.edu

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

ABSTRACT

We present Change-Link, a customizable data exploration tool which empowers the user to see visual representations of directories that have changed over time within a computer operating system that supports the Microsoft Volume Shadow Copy Service (VSS). Change-Link displays change information in a split-screen interface comprising an overview of directory change for the entire dataset and a detail view of change for individual directories. Input to Change-Link is an evidence hard drive containing an active file system and previous versions of the directory structure that were archived by the VSS. This approach to browsing change within a directory structure helps a digital forensic examiner understand how a particular computer was used to support criminal activity. Because data that have changed are often the most important, identifying directories that have changed over time directs attention towards data of higher importance. By examining the most important data, digital forensic examiners are better able to keep pace with the data explosion that is making current digital forensic examinations unmanageable. Our contributions include the development of a *segmented box and whisker* glyph for representing change over time for individual directories, an approach for aggregating VSS data for digital forensic examinations, and a data visualization tool for exploring digital forensic data.

Categories and Subject Descriptors

H.1 [**Models and Principles**]: User/Machine Systems—*Human Information Processing*
; H.5 [**Information Interfaces and Presentation**]: User Interfaces—*Screen Design and User-Centered Design*

Keywords

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

VizSec '12 October 15 2012, Seattle, WA, USA

1. INTRODUCTION

Digital forensics involves extracting and analyzing data from digital artifacts in support of law enforcement investigations. In digital forensics, the term data explosion 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 explosion is fuelled by the increased capacity and decreased cost 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, persons engaged in criminal activity are being associated with larger amounts of digital evidence. The amount of data subject to digital forensic examinations has become unmanageable. We design, implement, and demonstrate a new data visualization tool to help address this data explosion.

One way to keep pace with the data explosion is to increase the bandwidth by which digital forensic examiners perceive forensic data. Because more information can be obtained through vision than through all other senses combined [1], obtaining information through data visualization presents the greatest bandwidth for human perception. The need for an increased perceptual bandwidth is one of the primary motivations for applying data visualization techniques to digital forensics. Other motivations include knowledge discovery, increased productivity, and better comprehension [1].

By understanding how digital evidence has changed over time, digital forensic examiners are better able to understand what happened. A notable computer process that records data that have changed over time is the Microsoft Volume Shadow Copy Service (VSS), which is found in Windows Vista and Windows 7. The repositories of data that are created by this service are known as shadow volumes.

According to Microsoft, there can be as many as 512 shadow volumes for a given volume [13]. Current digital forensic tools support the accessing of individual shadow volumes, and some provide an understanding of what changed between two selected shadow volumes. None of the known tools support an understanding of change over multiple shadow volumes, and certainly none scale well enough to convey change according to as many as 512 shadow volumes.

We have developed a prototype tool that we believe will scale well enough to support the visual representation of a directory structure that has changed according to as many as 512 shadow volumes. By visualizing change with our tool, digital forensic examiners are better able to understand "what happened." Furthermore, since data that have changed are often more important than data that have not changed, digital forensic examiners who are able to identify changed data are better able to direct their efforts toward more important data. This approach helps examiners find important evidence quickly, which reduces the amount of time needed to complete a digital forensic examination.

We have developed a tool with a split-screen user interface called Change-Link which allows the digital forensic examiner to explore shadow volume data that are archived by the VSS through an *overview+detail* view.

Our contributions include:

- a new tool, Change-Link, for exploring changes to directory structure information as found in VSS data.
- a new approach for aggregating VSS data from multiple shadow volumes to detect changes in the directory structure over time.
- a visualization technique we call a *segmented box and whisker*, which is a glyph that represents how an individual directory has changed over time.

2. RELATED WORK

There are several tools that support the forensic examination of shadow volume data including Shadow Scanner, Shadow Explorer, Shadow Analyzer, and Time Traveler. All of these tools allow the user to access data from specific shadow volumes, and most allow for the comparison of two shadow volumes to reveal the net change. None of these tools provide an overview of change throughout the entire dataset, and none support an understanding of change over more than two shadow volumes. Of these tools, only Time Traveler embraces the power of data visualization by displaying a time-line of when snapshots were taken. Our approach is an improvement over these tools because we offer an overview of the entire dataset and the ability to see change over multiple shadow volumes simultaneously.

In our previous research, we introduced the *segmented box and whisker* glyph and applied it in a *fisheye view* [11]. Our current research continues to use the segmented box and whisker glyph but applies it in a Coordinated and Multiple Views technique [17]. The Coordinated and Multiple Views approach to visualizing data uses multiple windows to show several levels of detail of a dataset simultaneously. This approach is often used by domain experts to navigate through very large datasets in search of patterns and anomalies to support a better understanding of the data.

Coordinated and Multiple Views have been subdivided into classifications based on the data, the view, and the coordination between pairs of views [16]. Change-Link is classified as a Level 0 system based on its lack of flexibility regarding the dataset, its use of a fixed view, and the fixed coordination between views. Change-Link accepts only shadow volume data, formatted according to a specific syntax, presented with a specific split-view in which the views are coordinated such that one view is the master whose manipulation affects the view provided by the other view known as the slave. This classification is in sharp contrast to other tools such as Snap [16], which is classified as a Level 3 system because it is flexible in the datasets it can work with, the views it can provide, and the coordination between the views.

Our research is concerned more broadly with the field of visual analytics, which attempts to make large datasets tractable and comprehensible to the user through visual representations of data. A portion of our research can be categorized as a linked view [4, 22], which means the representation of data in one view can be affected by the manipulation of a data representation in another view. (Change-Link derives its name from viewing *change* over time with a *linked* view). Another portion of our research can be categorized as overview+detail [3], which concerns presenting an overview of the data in one view, while presenting a simultaneous detail view of the data in another view.

The development of Change-Link is related to previous research that visualizes change or an aspect of time. The previous research includes the visualization of repetitive patterns in multivariate data [20], changes to themes over time as found in documents [6], the use of animation to show trends in changing data [18], the use of illustration techniques to convey change of position (i.e., motion) [9], and the visualization of decision paths to predict future results based on current decisions [21]. Our research is unlike these previous efforts because our data expresses change as *coming into existence* and *going out of existence* for individual Operating System directories.

Munzner et al. [15] compared the structure of large trees. Their TreeJuxtaposer tool helps Biologists understand complex relationships that exist between many phylogenetic and evolutionary trees. It compares two trees at a time, but does not scale well when visualizing multiple trees and does not deal with deleted or nonexistent nodes. By contrast, we wish to compare many (e.g., 64) directory trees simultaneously with particular attention to files that may exist or not exist during a certain time period.

Our use of color to categorize directories that either exist or do not exist at different time periods is an information layering technique that has previously been proven to be effective [1].

Previous research has been conducted to support the visualization of large hierarchies, of which our dataset is a member. The Hyperbolic Browser [10] makes strategic use of screen real estate by representing the root of the hierarchy expanding from a central point. As the hierarchy expands, the screen space available to accommodate the hierarchy increases. The Cone Tree [14] efficiently uses screen real estate by projecting node descendants onto a three-dimensional plane forming the circular base of a cone. Neither the Hyperbolic Browser nor the Cone Tree can be suitably modified to reflect additional hierarchies or time periods. Thus, neither of these visualization techniques are suitable for visualizing change in our domain.

Change-Link was specifically developed to support the domain of digital forensics. We draw inspiration from Mizbee [12], which is a coordinated and multiple view tool developed to support biologists as they explore relationships between chromosomes of different species.

3. SHADOW VOLUME DATA

Change-link works with shadow volume data. Shadow volume data are generated by the Volume Shadow Copy Service, a process found in varying degrees in several Windows operating systems, but most notably found in Windows

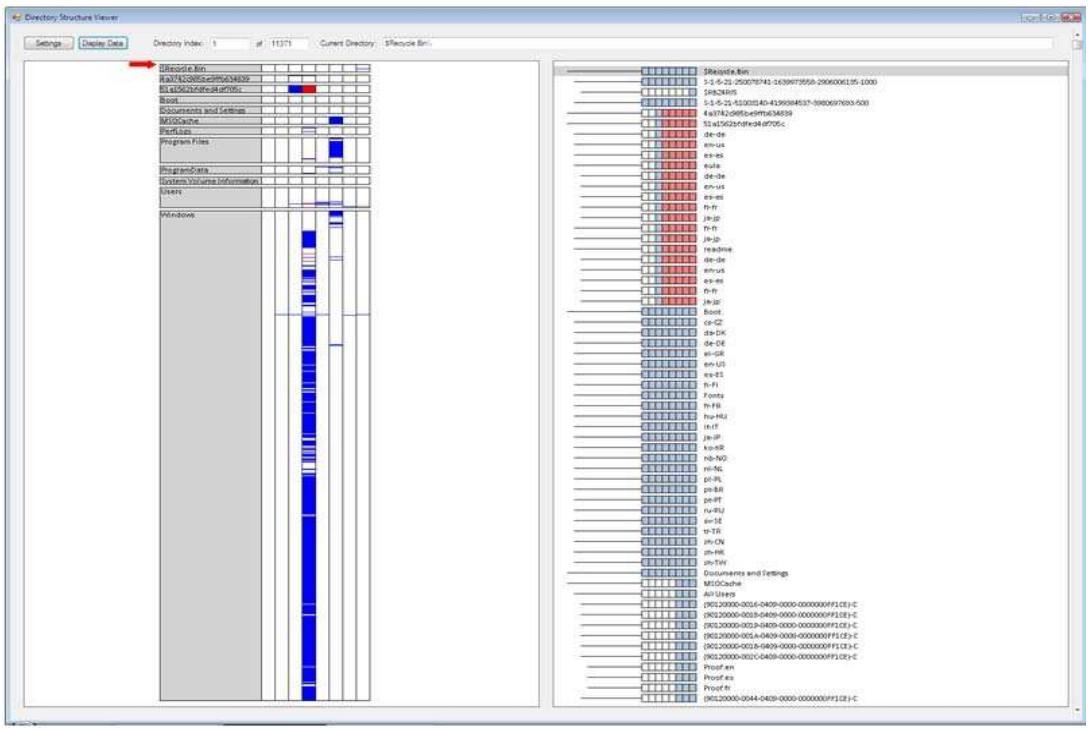

Figure 1: Change-Link user interface. The overview displayed in the left window provides the location and time period of directories that have changed, and the right window provides a detailed representation of directories that have changed.

Vista and Windows 7. This process archives the user data, application data, and operating system data, that exist at the start time of each archive process, and is used to roll back the computer system to a previous state. The major benefit of this capability is that data that have been accidentally deleted or modified can be recovered. This service also allows a computer system to be restored if a virus or the improper installation of software places the computer system into an unstable state. This service can be exploited by law enforcement to recover evidence of previous criminal activity, inlcuding the posession of illicit child images.

The repository of the archived data is commonly referred to as a *shadow volume*, and the contents are known as *shadow volume data*. Data are added to a shadow volume, when the Windows backup utility creates backups of the computer data (if configured to do so), just prior to the installation of new software, and when a user creates a restore point.

The amount of data that can be archived by the Volume Shadow Copy Service is subject to several limiting factors including the amount of physical space available to store each shadow volume, the version of the operating system, and the architecture being used (32 bit or 64 bit). A search of the literature did not identify a hard limit for the size of Shadow Volumes, but unconfirmed postings suggests the limit might be over several terabytes.

The time period between shadow volumes can be a few seconds (as when the installation of software triggers the creation of shadow volumes in quick succession) and it can be several days, weeks, or perhaps months, depending on the configuration of the system and its frequency of use. Thus, the time period spanned by each shadow volume is irregular.

4. TEST DATA AND PREPROCESSING

We created a test data set to simulate the characteristics of actual digital forensic data. We are working with law enforcement to test Change-Link with real case evidence.

Our dataset comprises eight shadow volumes that are established at the following times, respectively:

1. after installing Vista (but before activation)
2. after activating Vista
3. during the installation of Service Pack 1
4. after installing Service Pack 1
5. after installing Microsoft Office
6. after activating Microsoft Office
7. after creating user directories
8. after deleting user directories

The third shadow volume was created automatically by the operating system during the installation of Service Pack 1. The other seven shadow volumes were created manually by establishing a restore point, which triggers the archiving process of the Volume Shadow Copy Service. Each of these eight shadow volumes contains a snapshot of the directory-tree structure (and the files within each directory) at the time the shadow volume was created.

To support the creation of our test data set, we developed a data preprocessing tool that performs a breadth-first traversal of the directory tree found in each shadow volume. The tool extracts the name of each directory and assigns a nesting number. Each nesting number corresponds to the

depth of the directory in the tree structure. Directories at the root of the directory tree have a nesting number of zero; child directories have a nesting number of one; and grandchild directories have a nesting number of two. This nesting number is used as a software programming parameter to convey the nesting of child directories under parent directories with the whisker glyph, which is explained later.

When we execute our data parsing tool with user level access rights we obtain records for about 11,000 directories. Multiplying these directories by the eight time periods provides us with a dataset of about 88,000 data points. Executing the tool with system level access rights (using PsExec), results in about 700,000 directory records. For the eight time periods, this yields a large set of about 5.6 million data points. While developing our tool, we chose to work with the smaller dataset. Our future research will address how this tool scales with larger datasets.

For easier management and manipulation, data extracted by the directory parsing tool are imported into a SQLite database, along with binary values for each directory and time period indicating if the directory exists in the given time period. Change-Link accesses the SQLite database, executes queries directed by user interaction that return a subset of the data, and uses these data to update the visualizations in the browser interface.

One of our contributions to digital forensics is our combination of data from multiple shadow volumes into one conglomerate data set. This conglomeration of data, which allows for comparison of data points from different time periods, is what makes visualizing change over time possible. We are unaware of any other digital forensic tool that uses this approach to convey change in a graphical format.

5. DESIGN AND IMPLEMENTATION

Change-Link is a visual browsing tool to support forensic examiners in finding data of interest, especially changed data, within a directory hierarchy. Inputs to the system are instances of the hierarchy at different times, as recorded by VSS data. Outputs of the tool are color-coded displays of directory information in a graphical format. Work on the current prototype focused on finding effective visualization techniques for highlighting changes in the hierarchy.

Change-Link has a single interface with two display windows as shown in Figure 1. The left viewing window provides the overview, while the right viewing window provides the detail. We made this design choice to provide a "road map" (overview) which gives the user an understanding of where he is currently browsing within the data set, while also providing a detail view of the directory structure located at the focal point of the user's browsing selection.

5.1 Overview Window

The window on the left provides the overview portion of our overview+detail [3] visualization (Figure 1). This overview consists of rectangles that represent each of the directories that can be found at the root of the directory-tree structure. In Figure 1, twelve directories are represented. Figure 2 provides an example of how one of these rectangles is used to represent the Users directory.

The height of each rectangle is proportionate to the number of subdirectories that make up the ancestors of the root. For example, Figure 1 shows the Program Files directory has slightly more subdirectories than the Users directory, but

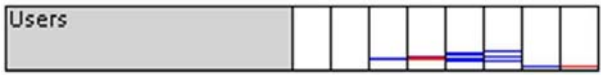

Figure 2: The Users directory. The rectangle that represents the Users directory is colored grey and has a white segment for each of the eight time periods. The segments contain blue and red horizontal lines that mark the approximate location of subdirectories that have been created (blue) or deleted (red).

each of these directories has significantly less subdirectories than the Windows directory, which appears to have about two-thirds of all of the subdirectories within the dataset. When the rectangle cannot be drawn tall enough for its name to be drawn on the rectangle, that rectangle is drawn with a minimum height of ten pixels to accommodate its name.

We considered using excentric labeling [7, 5] for associating names with directories, but we think doing so adds visual clutter. Excentric labeling is a data visualization technique that connects data labels to data points with lines. This approach allows data labels to be moved by the user so they are away from the data point being viewed to reduce congestion near small data points. Despite the intention to reduce congestion, excentric labeling still increases visual clutter.

Because the glyphs need to be large enough to support their data labels, there is a limit on the size of directories in the overview window, and on the segmented box and whiskers in the detail window. Experimentation reveals that the smallest we can make the data labels and still have them be easily readable is eight-point font.

The first half of each rectangle in the left overview window (Figure 1) is colored gray, and each has the name of the directory on it. The second half of the rectangle is partitioned into segments, one for each of the time periods. Figure 2 shows eight time periods in the dataset. Each time period segment is colored white and contains blue and red horizontal lines that mark the approximate locations within the subdirectory where a directory was either created (blue) or deleted (red). By reviewing these red and blue lines, the forensic examiner is able to understand quickly how much change has taken place, which time period contains the most change, the location within the directory tree structure where the change has taken place, and whether the change was either the creation or deletion of a directory. All of this information, conveyed in a single picture, helps the digital forensic examiner gain a better overall understanding of "what happened."

5.2 Detail Window

The detail window on the right of Figure 1 provides a scrollable column of segmented box and whiskers, one for each of the directories being displayed. The segmented box part of the segmented box and whisker is a rectangle that is divided evenly into segments; this glyph is one of our contributions to the data visualization community, which we introduced in our previous paper [11]. Each segment represents a different time period, ordered temporally with the most distant time period at the left, and the most recent time period at the right. By default, each time period segment is colored red, white, or blue. Blue denotes a directory that exists during the particular time period. White denotes

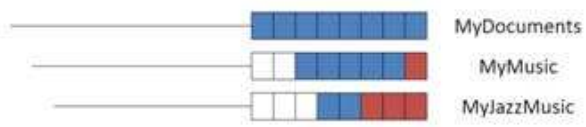

Figure 3: Segmented box and whiskers that convey MyJazzMusic directory is a child directory of MyMusic directory, which is a child directory of MyDocuments directory. Blue segments represent directories that exist, white segments represent directories that do not yet exist, and red segments represent directories that have been deleted.

a directory that does not yet exist. Red denotes a directory that once existed but has been deleted. The default values for these colors can be changed through the settings interface.

The default color values were chosen because they are visually distinct from each other and do not conflict with each other. A thorough analysis of hue, color mixing, lightness, and saturation was not considered as in previous work [2]. Previous research regarding the choice of a segmented colormap influenced our choice of colors [19].

In the right overview window (Figure 1), the whisker portion of the segmented box and whisker is a thin black line that extends to the left of each segmented box. This whisker indents the child directory under the parent directory, to express the parent-child relationship, which is referred to as an inclusion relation [8]. Our use of whiskers visually conveys nesting relationships while simultaneously permitting the segmented boxes for different files to be intuitively aligned vertically by same time periods.

Figure 3, from the bottom up, shows MyJazzMusic directory is a child directory of MyMusic directory, which itself is a child directory of MyDocuments directory. Using whiskers to express indentations allows each segmented box to be arranged such that the segments from the same time period are ordered in columns. This allows the viewer to scan one of the eight columns of segments and logically associate all of the directories that exist during a chosen time-period. For example, the far right column in Figure 3, which represents the most recent time period, shows that the directory named MyDocuments (colored blue) exists during that time period, while MyMusic and MyJazzMusic directories (colored red) do not. The benefit of using a whisker to denote nesting is that the thin black line provides just enough information to convey the nesting relationship without adding additional visual clutter that can detract from the visualization.

5.3 Navigation

By manipulating the scroll bar along the right edge of the right viewing window (Figure 1), the user can scroll through the data. Scrolling can also be achieved by changing the Directory Index value displayed in the top menu bar and selecting the Display Data button to refresh the visualization.

Manipulating the scroll bar on the right of the detail window (Figure 1) causes the Directory Index within the top menu bar to change. The Directory Index corresponds to the index value, assigned by Change-Link, of the directory that is represented by the segmented box and whisker that is located in the gray "focal bar" in the right side window. As the Directory Index value changes, so too do the representations of the segmented box and whiskers in the right window, as well as the location of the red arrow in the left overview window. This red arrow points to the relative location within the directory tree of the directory that is represented by the segmented box and whisker which is shown at the focal point of the right window. This use of linked-views [4] allows the viewer to understand their browsing position within the entire dataset (the overview) while also understanding the specific directory being viewed (the detail).

5.4 Settings

A settings button in the top menu bar allows a settings interface to be displayed. This interface permits the user to define the location of the SQLite database from which data is retrieved, and to define the heights, widths, and colors of the various aspects of the glyphs that make up the visualizations. These settings can be modified while data are being explored so the user can customize the visualization to display the representation of the data that best conveys its meaning. Being able to modify these settings proved useful when we moved Change-Link to a new operating environment and had to adjust the visualizations for different monitor resolutions and set a new path for the SQLite database.

5.5 Implementation

Change-Link is implemented with approximately 1,034 lines of C# code. We chose this language because of its availability to the author in the research lab, and because of its support of the .NET framework. We chose the .NET framework because of its support of Windows Forms, SQLite database references, and easy extraction of directory metadata from the dataset being explored.

Change-Link is a proof-of-concept prototype which ignores some functionality standards which the digital forensic community has grown to expect from a forensic tool. We have tested technology which will allow us, in our next implementation, to automate the mounting of evidence in the form of a dd-image. Although our technology requires us to operate from a Windows 7 operating system, we are optimistic that our tool will eventually be platform independent to meet the needs of both Windows and Linux users.

6. RESULTS

Applied to the test data, Change-Link reveals several notable changes to the directory structure. These notable visualizations include the installation of Service Pack 1, the renaming of a directory, and the creation and deletion of a user directory. The activation of Vista and Microsoft Office provide similar notable visualizations (not shown).

6.1 Overview Window

The overview of data presented in the left side of the interface gives the viewer a sense of which time periods experienced the most change, and therefore might be the most important and deserving of further investigation (Figure 1). The fourth and sixth time periods have the greatest concentration of change lines (whether they be blue or red), and therefore the most change. It is not surprising that the Windows directory has the most significant amount of change during these two time periods because Service Pack 1 was installed in period four and Microsoft Office was activated in time period six. It is the activation, not the installation,

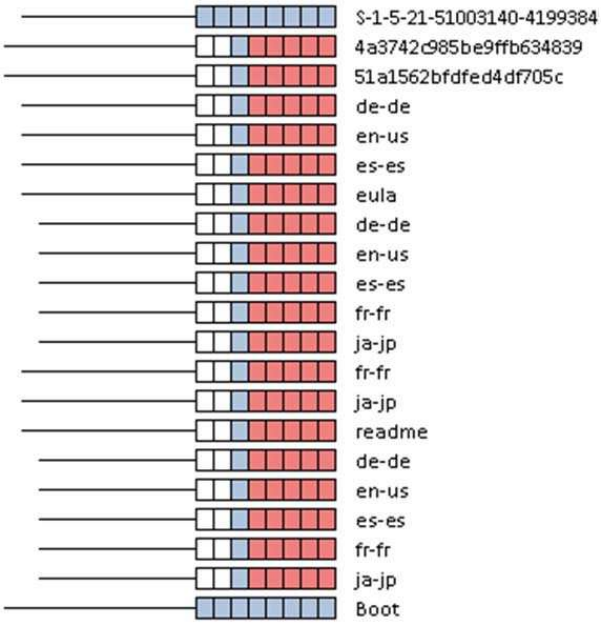

Figure 4: Installation of Service Pack 1. Segmented Box and Whiskers that contain red, white, and blue segments represent nineteen temporary directories that were created and then deleted during the installation of Service Pack 1.

of Microsoft Office that results in more directories being changed. These fourth and sixth time-periods also correspond to the most changes to the Program Files directory.

By visualizing the data through this overview method, the digital forensic examiner can devote his attention to the fourth and sixth time periods. This approach allows him to exclude about three-fourths of the extremely large data set from being considered, leading to greater productivity for the digital forensic examiner.

6.2 Service Pack 1: Temporary Directories

Figure 4 shows the effect of installing Vista Service Pack 1. The nineteen segmented box and whiskers that have red, white, and blue colored segments represent directories that were first created and then deleted during this installation.

Beginning with the directory "4a3742c985be9ffb634839" and ending with the directory "ja-jp" in Figure 4, each of the segmented box and whiskers begins with two white colored segments. This means that these directories did not exist during the first two time periods. The third segment is colored blue, which means that the corresponding directory did exist for that time period. The five red segments that follow each of the blue segments shows that the corresponding directory did not exist during those time periods. Viewed as a single event, the visualization in Figure 4 reveals that the installation of Service Pack 1 created the nineteen directories, triggered the creation of a shadow volume which archived the temporary structure of the directory tree, and then deleted the directories. Our hypothesis is that the installation of the service pack created these nineteen directories to contain data that may have been needed to roll back the system if the installation was not successful.

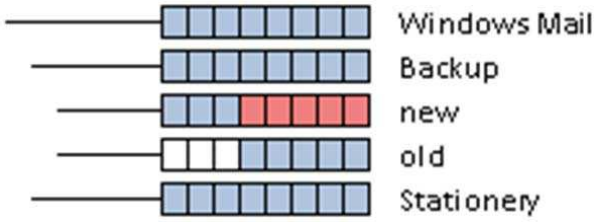

Figure 5: The installation of Service Pack 1 appears to have renamed the directory named "new" to "old."

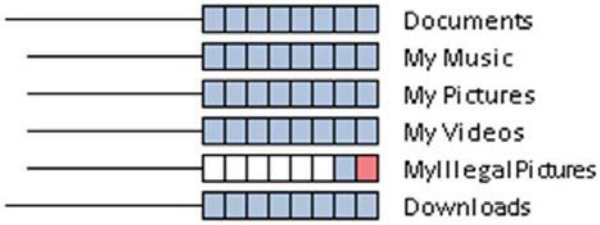

Figure 6: MyIllegalPictures. The Segmented Box and Whisker that represents the MyIllegalPictures directory highlights this directory as a directory that is worthy of furture examination.

Through color coding and aligning segments from the same time periods, Change-Link helps the examiner interpret and reason about events in Figure 4. Arranging the segments in columns clearly reveals temporal relationships among their associated directories. Change-Link empowers the examiner to interpret these changes as caused by one event. Color (red, white, and blue) distinguishes this group of directories from those nearby that have not changed.

6.3 Service Pack 1: Directory Renaming

Figure 5 depicts useful visualizations of how the directory tree structure changed. Nested under a directory named "Backup" (second from the top), there are two child directories named "new" and "old." The directory named "new" had existed for the first three time periods, until the completion of the installation of the service pack, at which time it no longer existed. The directory named "old" was created just after the installation of the service pack. Our hypothesis is that Service Pack 1 renamed the directory named "new" to the new name of "old." Considering that the parent directory is named "Backup," it makes sense that there is a child directory named "old" (perhaps for storing old versions of data) rather than a child directory named "new." Thus, the visualization created by Change-Link seems to have revealed the simple renaming of a directory by Service Pack 1. (The data we extract from the hard drive do not allow us to distinguish between a directory being renamed, and a directory being deleted and a new directory being created within the same parent directory. Both events look identical with Change-Link.) This example demonstrates how Change-Link's use of segmented box and whiskers simplifies analysis of change in specific time periods.

6.4 User Created and Deleted Directory

Figure 6 shows the effect of creating a directory in time-

period seven and then deleting it during period eight. In this case the directory is named "MyIllegalPictures" to represent a scenario in which a user has a directory containing illicit pictures of children, which is a typical discovery in some digital forensic examinations. During the creation of this data set, after Microsoft Office was installed and activated, we created this directory, established a restore point which caused a record of this directory to be archived into a shadow volume, and then deleted this directory. This scenario is meant to be similar to the situation in which a user deletes a directory in an effort to hide the illegal activity, not realizing that records of the behavior have been archived into a shadow volume. Change-Link's use of red makes this attempt to hide the event of the deleted directory easy to notice. This event is recorded by such a small anomaly in the data that current tools and digital forensic methods might encourage a forensic examiner to overlook it.

7. USER REACTIONS

We gathered preliminary informal user reactions to Change-Link from seven co-workers at the Defense Cyber Crime Center (DC3) who have work experience with forensic tool development and/or forensic tool validation. Each volunteer was seated in front of a computer running Change-Link, provided an explanation and demonstration of the tool, allowed to interact with the tool to navigate the dataset, and asked to complete a casual survey regarding their experience. The survey included free responses and seven quantitative responses on a six-point Likert scale, with 1 being "poor" and 6 being "excellent."

Based on their experience, our subjects are appropriate people to provide opinions regarding the usability of Change-Link. Further, as demonstrated by their responses to Questions 2-5, our subjects had a sufficient understanding of how to use Change-Link. In particular, each user correctly identified the time periods of the most change, the most created directories, the most deleted directories, and the root directory with the greatest number of subdirectories.

The users rated the effectiveness of the segmented box and whisker glyph for representing directories that exist over several time periods as being close to "excellent" [Four rated it a 6 and three rated it a 5]. The same scores were received in response to a question regarding the tool's ability to support navigation through the dataset.

Slightly weaker scores were received in response to a question that asked how well the user felt they understood what they are looking at [the scores were one 4, four 5s, and two 6s]. The exact same scores were received when users were asked to evaluate the use of color within the visualizations. Even weaker scores were received when users were asked to rate how effective the whisker part of the segmented box and whisker is for conveying the parent-child relationship. The scores reported are four 4s, two 5s, and one 6.

Users responded to several open-ended questions regarding their experience with Change-Link. All users stated Change-Link helps users understand more information as a result of seeing it in the visual format, helps users discover information that might be overlooked when using other tools and/or techniques, and helps digital forensic examiners be more productive in their examinations. Six of the seven users stated Change-Link left them with a better comprehension of the data set. We recognize, however, that there might be some bias given that the subjects are co-workers. Despite the casual nature of our preliminary user study, we feel that the informal user reactions provide useful feedback on the efficiency, effectiveness, and user satisfaction of Change-Link.

8. CONCLUSION

Using an overview+detail approach, we designed and implemented Change-Link, a prototype of a tool for visualizing how a directory-tree structure has changed over time. Initial informal user reactions to Change-Link are positive, with users commenting that the tool helps them to "focus on anomalies" and "find meaningful data more quickly." As a visualization-based data browsing tool to support digital forensics, our Change-Link prototype helps forensic examiners explore data more efficiently and effectively.

Change-Link offers one powerful strategy for dealing with the explosion of personal data confronted by digital forensic examiners. There are, however, limits to such tools. When data sets are sufficiently huge—which is easily possible with the virtual environments of cloud computing, for example—a single visualization that presents an overview of the entire dataset with sufficient detail is not feasible. Therefore, we offer Change-Link not as a replacement for existing tools, but rather as a companion to existing technology. Also, the criminal can attempt to limit the effectiveness of Change-Link by causing a huge number of changes to the entire directory tree concurrently with illegal activity, hiding incriminating changes among a sea of similarly modified data.

Nevertheless, Change-Link is a useful tool for exploring changes to directory tree structure. More generally, its techniques will also be useful in other contexts for exploring changes to other hierarchies. Change-Link demonstrates the power of visualization to help digital forensic examiners identify and understand significant artifacts in large data sets.

Our future goals for Change-Link include adding two additional views: a *directory view*, which displays the contents of a selected directory and how those contents may have changed over time, and a *file view*, which displays a visual representation of file attributes (metadata) and how these values may have changed over time. By adding these views to the overview and tree view already found in Change-Link, we believe we will produce a coordinated and linked visual representation of the data that will more strongly support navigation through a large data set while also supporting analysis of how the data has changed over time.

Acknowledgements

We are grateful to the many participants in our usability study and also to Josiah Dykstra, Eoghan Casey, Dhananjay Phatak, Charles Nicholas, and Clay Shields for their helpful comments. Leschke is supported as the Research and Development Lead at the Defense Cyber Crime Center. Sherman is supported in part by the Department of Defense under IASP Grants H98230-09-1-0404, H98230-10-1-0359, and H98230-11-1-0473.

9. REFERENCES

[1] Thomas Ball and Stephen Eick. Software visualization in the large. *Computer*, 29(4):33–43, 1996.

[2] Cynthia Brewer. Color use guidelines for data representation. In *Proceedings of the Section on*

Statistical Graphics, pages 55–60. American Statistical Association, 1999.

[3] Andy Cockburn, Amy Karlson, and Benjamin Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys*, 41(1):1–31, 2008.

[4] Robert Erbacher and Deborah Frincke. Hierarchical linked views. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, pages 1–12, 2007.

[5] Jean-Daniel Fekete and Catherine Plaisant. Excentric labeling: Dynamic neighborhood labeling for data visualization. Technical report, University of Maryland at College Park Human-Computer Interaction Laboratory, 1999.

[6] Susan Havre, Elizabeth Hetzler, Paul Whitney, and Lucy Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.

[7] Natalie Henry and Jean-Daniel Fekete. Matrixexplorer: A dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684, 2006.

[8] Danny Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.

[9] Alark Joshi and Penny Rheingans. Illustration-inspired techniques for visualizing time-varying data. *Visualization*, pages 679–686, 2005.

[10] John Lamping, Ramana Rao, and Peter Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*, pages 401–408, 1995.

[11] Timothy R. Leschke, Alan T. Sherman, and Penny Rheingans. Change-link: A tool for exploring how a directory-tree structure has changed over time in support of digital forensic examinations. In *Government Forum of Incident Response and Security Teams (GFIRST)*, December 2011.

[12] Mariah Meyer, Tamera Munzner, and Hannspeter Pfister. Mizbee: A multiscale synteny browser. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):897–904, 2009.

[13] Microsoft. Volume shadow copy service. http://technet.microsoft.com/en-us/library/ee923636%28v=ws.10%29.aspx" Accessed 14 September, 2011.

[14] Tamara Munzner and Paul Burchard. Visualizing the structure of the world wide web in 3d hyperbolic space. In *First Symposium on Virtual Reality Modeling Language (VRML '95)*, pages 33–38. Association of Computing Machinery, 1995.

[15] Tamara Munzner, Fracois Guimbretiere, Sardar Tasiran, Li Zhang, and Yunhong Zhou. Treejuxtaposer: Scalable tree comparison using focus+context with guaranteed visibility. *ACM Transactions on Graphics*, 22(3):453–462, 2003.

[16] Chris North and Ben Shneiderman. Snap-together visualization: A user interface for coordinating visualizations via relational schema. In *Advanced Visual Interfaces*, pages 1–9, 2000.

[17] Jonathan C. Roberts. State of the art: Coordinated and multiple views in exploratory visualization. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, pages 1–11, 2007.

[18] George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6): pages 1,325–1,332, 2008.

[19] Bernice Rogowitz and Lloyd Treinish. How not to lie with visualization. `http:\\www.research.ibm.com` Accessed 14 September, 2011.

[20] Jarke van Wijk and Edward van Selow. Cluster and calendar based visualization of time series data. In *Information Visualization*, pages 4–9, 1999.

[21] Jurgen Waser, Rahael Fuchs, Hrvoje Ribici, Benjamin Schindler, Gunther Bloschl, and M. Edward Groller. World lines. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1458–1467, 2010.

[22] Graham Willis. Linked data views. *Statistical and Computing Graphics Newsletter*, 10(1):20–24, Summer 1999.