UVisP: User-centric Visualization of Data Provenance with Gestalt Principles

Jeffery Garae
Cyber Security Lab
The University of Waikato
Hamilton, New Zealand
Email: jg147@students.waikato.ac.nz

Ryan K L Ko
Cyber Security Lab
The University of Waikato
Hamilton, New Zealand
Email: ryan@waikato.ac.nz

Sivadon Chaisiri
Cyber Security Lab
The University of Waikato
Hamilton, New Zealand
Email: chaisiri@waikato.ac.nz

Abstract—The need to understand and track files (and inherently, data) in cloud computing systems is in high demand. Over the past years, the use of logs and data representation using graphs have become the main method for tracking and relating information to the cloud users. While being used, tracking related information with 'data provenance' (i.e. series of chronicles and the derivation history of data on metadata) is the new trend for cloud users. However, there is still much room for improving data activity representation in cloud systems for end-users. We propose "User-centric Visualization of data provenance with Gestalt (UVisP)", a novel user-centric visualization technique for data provenance. This technique aims to facilitate the missing link between data movements in cloud computing environments and the end-users uncertain queries over their files security and life cycle within cloud systems. The proof of concept for the UVisP technique integrates an opensource visualization API with Gestalts theory of perception to provide a range of user-centric provenance visualizations. UVisP allows users to transform and visualize provenance (logs) with implicit prior knowledge of 'Gestalts theory of perception.' We presented the initial development of the UVisP technique and our results show that the integration of Gestalt and 'perceptual key(s)' in provenance visualization allows end-users to enhance their visualizing capabilities, to extract useful knowledge and understand the visualizations better.

Index Terms - Visualization; Data Provenance; Security; User Centric.

I. INTRODUCTION

The complex changes on networks raise the question of how to secure them. Before securing these networks and systems, one needs to "know" or "see" what is going on in these systems. A proposed solution to this is visualization. With the state-of-the-art 'data visualization' techniques are the means of providing a visible solution to knowing what is going on in these systems. It also helps users to actively identify useful patterns and gens within the data when picturing different visualizations [1].

In the area of security, data provenance is a burgeoning field for data security particularly for knowing what goes on with the data. Data provenance enables the ability of keeping track of file movements over the network and on the Internet. However, in the area of provenance, visualization is not a well-studied topic. Most visualizations in this area of study are for exploratory, discovery and reporting purposes. Therefore, the

need to bring visualization into provenance at the end-users perspective is a high demand.

The main objective of this research is to explore potential methods and techniques of visualizing provenance to benefit the end-users of the cloud network. This research presents a user-centric visualization technique (UVisP) centred on visualizing provenance. It incorporates the concepts of Gestalt theory of perception [2] to provide a user-centric approach which would relate provenance to the users benefits. We define 'end-users' as the users of the product (cloud services) with minimal savvy on technologies, i.e., not technology specialists, data analysts, but someone who relies on products and cloud services such as Dropbox.

1) Research Goals: The goal of UVisP is to transform data into visualization for end-users to identify useful patterns, gens, trends, and relationships by observing activities of interested files over the network in a given time. Such observations leads to new insights which will be discussed in the later section of this paper.

2) Key Contributions: The main contributions of this paper are as follows:

- We propose a novel user-centric visualization technique (UVisP) for visualizing provenance. Our technique equips end-users with prior-knowledge, methods, and ways of visualizing data provenance as a means of security for their benefits.
- We implemented our novel technique by integrating D3, an existing open source visualization API with the Gestalt theory of perception. To the best of our knowledge, this is the first time, such a user-centric technique is proposed for Data Provenance.
- We introduced and defined *"perceptual key"*, a core aspect to identifying useful patterns and information indicators in a user-centric provenance visualization.

II. BACKGROUND AND RELATED WORK

We begin this section by analyzing the past and present work carried out in this research field.

A. The Visualization Discovery Process

Visualization is a common method for analysing, relating and understanding scientific datasets. Previous research shows that visualization exists for several reasons: technology, art and for an empirical science [3]. Different visualizations target specific user groups for specific goals, needs, environment and time [4]. Most visualization gives the users the flexibility of customizing their displays and visual view in an understandable way [3] [5] [6]. In order to acquire these user views, one must understand the visualization discovery process as presented in Fig. 1.
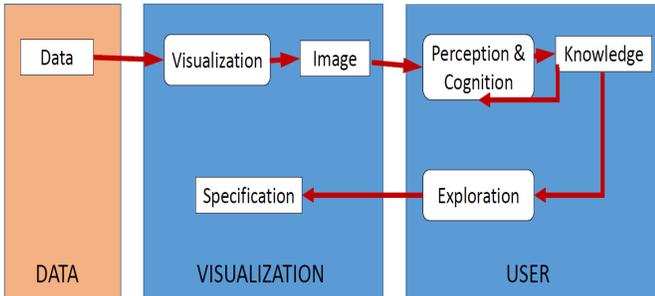

Fig. 1. Views on Visualization

B. Evaluation of Related Work

As shown previously in Fig. 1, users are the active component of the visualization discovery process. They process and transform raw data into knowledge from the visualizations. A vast amount of work concentrating on visualization and security specifically targeting IT expects and analysts who wishes to analyse data and produce reports out from the findings [28] [29]. However, less work has been done in the field of visualizing data provenance for security purposes. This is to address the end-users needs. Visualizing data provenance can help users to understand the life cycle of his/her data and observe potential security breaches.

Several types of past and existing related work are discussed as follows:

1) Visualization in General: We begin with setting the foundation of visualization by addressing the general use of line/bar graphs, pie charts, donut charts and heat maps to explore and analyse data through visualization [7] [8] [9]. These types of visualizations have paved a way forward to new and intractable visualizations which will be discussed later in this chapter. For example, in [10], bar graphs are used to show traffic resources based on an online survey and are centered towards research purposes. Their existence came about as a better way to display data in a visual representation than using tables [11]. However, to date, the presence of visualization in the research fields has attracted researchers to adopt graphs, pie-charts, and other common graphs into the visualizations to address user interactions and a means of providing exploratory information.

2) Circular Patterns Visualization: The circular pattern representation visualization is another form that addresses large datasets. The so-called "AnalyticaR, a visualization technique for data exploratory)" helps users to view data with a reasonable amount of space without much effort [12]. Similarly in [13] [14], visualization of large datasets are represented using circle segments and recursive patterns for 'data exploratory' purposes and multidimensional data representations. However, the emphases of existing examples of these visualizations are exploring and analysing data for better understanding. There is less focus on who the targeted audiences are. With different levels of visual evaluations by targeted audiences, end-users may struggle to understand such visualizations.

3) Visualization of Provenance: As visualization in general is a popular means of data representation, visualizing provenance is still in its beginning years, whereby researchers are gradually changing from the current graphing (use of graphs) techniques, circular patterns and moving onto applying other existing visualization techniques [15] [16]. This is due to the rapid study and use of data mining and analysis of large datasets [17].

Another example of visualization of provenance is discussed in the paper "Visualization of network data provenance" [18] [19]. Chen et al. aided their study on large provenance datasets for exploratory and explanatory purposes [19] showing the chain of history of datasets. This visualization uses the 'layout' technique to improve the researchers understanding of large datasets. The important components here in the visualization are; 'the nature of provenance data' and the 'user requirements.' For example, we see that the nature of the provenance dictates the outcome of the visualization [19].

4) Gestalt Theory of Perception: The Gestalt principles established by Max Wertheimer, Wolfgang Klher and Kurt koffka [20] [21] states the key ideas behind Gestalt Theory and the principles of the Gestalt perception model [22]. While Gestalt theory has its advantages, it does possess its challenges and limitations. These challenges are closely related to the perceptual grouping and figure-ground organization. For example, there is a need for further analysis on the factors that are common in both the grouping and figure organization. There is also a need to fully distinguish between perceptual grouping and figure-ground organization [22]. Such challenges and limitations can affect visualizations when trying to apply both the perceptual grouping and figure-ground organization into one visualization.

III. D3 AND GESTALT INTEGRATION DESIGN

In this section, we outline the Visualization API - Data-Driven Documents (D3.js) [25] and the Gestalt integration design. Firstly, D3.js is a JavaScript library for manipulating documents based on data. The adoption of D3 into visualizations has become the new trend for new visualizations. With high and multiple visualization dimensions and the interactions between D3 visualizations, it became the choice our research.

A. UVisP's User-Centric Technique

With the foundation knowledge of D3.js and how it works, we now take a look at an important key phases of the UVisP design. This research refers to the key phases as the UVisPs U-C methodology. The methodology consists of (in an orderly manner) Fig. 2:

1) **Prior-knowledge of the Gestalt Theory of perception:** In this step, a user of the UVisP has to understand what the Gestalt theory of perception is. A fair understanding of the Gestalt Laws / Principles is the minimum requirement for the 'Step 1' of this U-C methodology.

2) **Processing of the raw data (Progger log) Type, User, Filename, FD, [26]:** In this step, a user with the preferred visualization in mind, should have some knowledge of what type of data is being process for the visualization. For example, tracking or visualizing the data movement requires certain data types. This knowledge will provide the user with some clues to how the visualization will look like. In any visualizations, like any other art portraits, clues and prior knowledge are the key sources of building that image and final knowledge/story from that particular visualization.

3) **Building data-input format (.json & .csv):** This step allows the UVisP to produce better user-centric visualizations. In other words, the visualization tells a story and are viewed over and over again by the users of the UVisP technique.

4) **Visualization type & animation (D3 Library):** This step creates the connection to 'Step 1' again in order for whole methodology to focus only on the output of the visualization.

5) **Colour Theory:**

   - Colour selection for each visualization: The right selection of colour choices is very important for different types of visualizations.
   - The required colours needed to use for each visualization: The need to use right colours to represent each visualization. This method helps end-users to quickly identify perceptual keys and form visual images, therefore produce and extract useful information from visualizations.
   - Applying the colour wheel/circle (a tool for combining colours): This is a tool/method provided to users of the visualizations with common colour (Red, Yellow and Blue) usages to enhance their ability when confronting different types of visualizations. For example, Yellow and Green can be used to represent "Create" and "Read" system calls since both actions represent an action done to the file.

6) **Gestalt-&-D3 Association Visualization Approach (as shown in Fig. 3):** This step integrates the whole process of the U-C methodology back to the human perception where the

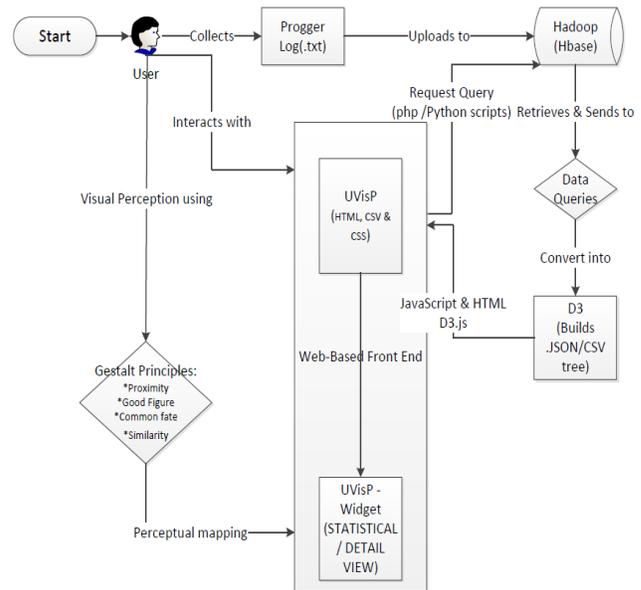user has the power of perceiving useful information and knowledge from the visualizations.

Fig. 2. UVisP design flowchart

B. Gestalt Integration Design with D3 Visualization

A model design of the Gestalt application and the user's visualization experience when interacting with UVisP is created by incorporating both human experiences with the Gestalt applications . The emphasis is that humans are natural at perceptual mapping (a technique) and often their perception are permanent [24]. Therefore, applying Gestalt Theory into the users experiences provides a new method of visualization.

Fig. 3 shows the type (Vis_GP1 - Vis_GP6) of visualizations scenarios (1), Gestalt laws of perception (2) and the common grounds between the two areas which is the human perception (3). The visualization types in (1) are added to act as a form of ID or tagging which contains different Progger scenarios. These visualization types produces provenance visualizations. The visualizations are then added to the different Gestalt laws/principles in (2) to compare and analyse for common grounds (union) as shown in (3).

The idea of using Fig. 3 is to show the integration of the visualization scenarios and the Gestalts theory of perception. This integration produces the third section as shown, whereby the 'union (3)' integrates 'visualization type (1)' and 'Gestalt theory (2)' together. This union is the 'human perception.' Therefore human perception in visualizations has a big impact on the presented knowledge to the users. The knowledge gained when visualizing using the UVisP technique are directed towards the provenance of data.

With that knowledge in hand, a model design of how a user (human) would interact with the UVisP technique with the
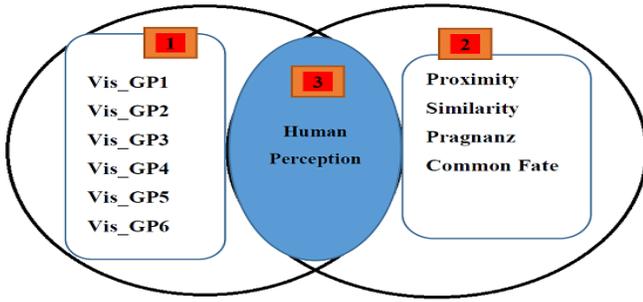
Fig. 3. The UVisP D3 - Gestalt Integration Technique

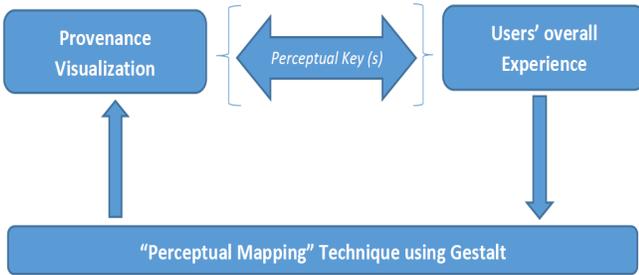aim of extracting useful information from the visualizations is shown in Fig. 4:

Fig. 4. The UVisP Visualization user approach model

1) Perceptual Key Function in Visualizations: The model design in Fig. 4 shows that a 'perceptual key' in humans plays an important role in the UVisP technique, i.e. with the perceptual key present in humans, the user uses the perceptual mapping technique to build/form mental images, patterns from the visualizations given using UVisP. 'Perceptual key(s)' in this paper is defined as; the act of recognizing the very first key-identifier, eye-catching, common and recurring pattern(s) in a given visualization that are sourced out by the users of UVisP technique. For example, in Fig. 5, there are several perceptual key indicators such as the dogs 'head', 'body' and or 'legs.' Without such perceptual key indicators in a visualization that visualization in Fig. 5 is just a group of dotted spots. In addition, the order at which the perceptual key(s) appear depends on the users of the UVisP technique. The "perceptual mapping" technique is applied by the user when confronting different visualizations of UVisP technique [3] [24]. This technique helps users to build the knowledge faster, retrieve and extract useful knowledge from the visual representation.

2) Associating D3 and Gestalt Theory of Perception: Another important aspect to the "D3 - Gestalt visualization Model" involves manually identifying user queries when using the cloud services (See Table I). This has been done by carrying out a user survey (questionnaire) to categorize queries from most important to least important. We begin by providing a step-by-step method of how the user survey was carried out.

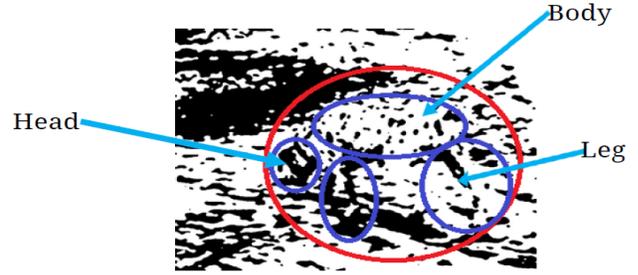- Step 1: Develop/brain-storm up to 100 user queries when using a cloud service (See Table I).

Fig. 5. Identifying Perceptual Key(s) in a Gestalt Visualization

- Step 2: Categorize all user queries into common groups. For example, 'file creation' and 'file movement.'
- Step 3: Narrow down the user queries to the top 20 user queries by selecting the top 20 queries that goes in line with what the users of the cloud consider most important when tracking and monitoring data activities.
- Step 4: Using the top 20 user queries, carryout a survey/questionnaire by asking random users of the cloud to select their top 3 user queries based on the importance of data stored in the cloud. The top 3 selected queries are for the survey.
- Step 5: Based on the survey/ questionnaire, the top 3 user queries are associated with the Gestalt theory for visualization.

TABLE I
UVISP PROGGER-D3-GESTALT INTEGRATION METHOD

Vis-Type	Goal-Vis	D3.Vis-Rep	GA-Sim	GA-close	GA-Prox
F-Open	stat	tr-mp	✓		
F-Creat	report	heat-mp		✓	✓
F-copy	track	buble	✓	✓	
F-Mv	track	FD-gr	✓		
F-ren	stat	tr-mp	✓		
Dir-ren	track	tr-mp	✓	✓	✓

Table I shows the user survey on the first few user queries. The aim of this user survey is to identify:

- types of visualization (Based on Progger system call actions) (Vis-Type).
- goal of visualization (Goal-Vis).
- types of D3 visual representation (D3.Vis-Rep).
- which Gestalt law approach will match the type and D3 visualization (GA-Sim, GA-close, GA-Prox).

A justifiable method is used to design and obtain the 'types of visualizations', 'goals of visualizations' and link them up with the Gestalt approach. This method not only has to work for the end visualization, but also to link up all components such as the:

1) Progger scenarios: For example, A progger log collected from executing a "Create file, change content, update file, and finally close the file."

2) Reasons for the scenarios: For example, why is it important to visualize?
3) What type of visualization model is suitable for this scenarios: For example, which D3 visualization environment and type is best for each scenarios; i.e. D3 Force direction-layout visualization; Bubble visualization; Treemaps.
4) Gestalt approach proposed for each scenarios: Which Gestalt Laws will apply to this scenarios based on the D3 visualization model chosen.

Table I shows that to link all components together requires two processes:

1) Process 1: Identify and analysing the existing D3 visualization templates and manually linking them up with how the Progger scenarios are executed. For example, a 'D3 Treemap' visualization can be linked to the Progger scenario that allows the user to "create, read, write and finally save/close" a file.
2) Process 2: Once Process 1 is complete, the type of Gestalt approach will be added to the visualization based on the behaviour of the data when executing the scenario from Progger.

The findings should provide a justifiable indication to which D3 visualization representation will associate with the Progger scenarios. This also provides a complete visualization environment to the users by applying the missing UVisP components (1 and 2) to the "Visualization Discovery Process (Fig. 1)" as they apply the UVisP technique to obtain the final visualization. This is also illustrated in Fig. 6. The missing components between users and existing visualizations is the UVisP technique. This consists of Progger-D3-Gestalt Integration (Applied UVisP 1); Perceptual Key(s) (Applied UVisP 2); and finally the UVisP U-C methodology (Applied UVisP 2).
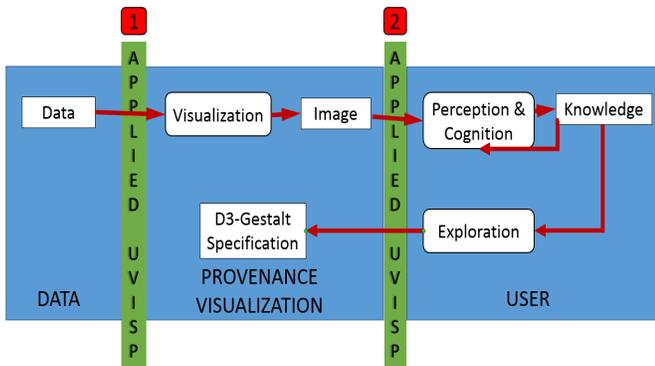

Fig. 6. UVisP components added to the 'Visualization Discovery Process'

Finally this section of the research covers the most sensitive aspect of the visualization technique proposed. The overall structure, flow-functions and output is tailored toward user-centricity therefore it targets simplicity for both the users and also the UVisP operations.

IV. EVALUATION

In this section, we describe the novel technique used in this research, its findings and challenges encountered when carrying out the research. We begin with analyzing the "D3-Gestalt Model of Association" method used in Fig. 6. We will then discuss the challenges encountered while carrying out this research. Finally, we analyze the visualization findings using the sample visualizations.

A. Visualization Attributes

The purpose of creating and consolidating all information regarding the Gestalt laws and the different types of Progger scenario is to create a link between the two entities and see what is common in them (Shown in Table I). The common factors between the two entities provides the user-centric portion of each visualizations in the UVisP technique. Based on the idea stated above, it indicates that most types of visualization which are highly applicable with applying Gestalt laws /principles are the process calls that combines several system calls together. For example, a Progger operation to create, open, write, and close a file is linked to the 'Gestalt approach' of Similarity, Closure, Continuation, and Proximity. Another example is the 'scp operation' of a file from one VM to another VM uses the Similarity and Proximity approach of Gestalt. However, multiple 'scp operations' from VM to VMs are justifiable with the Similarity, Closure, Continuation, and Pragnanz approach of Gestalt.

Throughout the course of this research, we had encountered challenges. These challenges are as follows:

B. Supporting the full data life cycle (Provenance) with visualization

1) Data acquisition The ability to constantly carry out data logging and visualizing it was a challenge. In other words, being able to evaluate a full data life cycle and visualize it was a challenge for us.
2) Integration The ability for us to incorporate all data life cycle (Provenance) into one visualization was a challenge. Most visualization portrays parts of the data life cycle and not all.
3) Analysis Analysing the full data life cycle (Provenance) with visualization was a challenge due to several reasons. The first being, the amount of space needed for a user-centric visualization was not possible. Secondly, analysing multiple datasets and its visualization was time consuming. However, the visualization output provided by UVisP had to be static and covers only parts of the data life cycle and not all.
4) Dissemination The ability to spread information using visualization was a challenge as well for us. All visualizations are developed with the idea of targeting the end-users, and most information extracted from the given visualizations were tailored towards end-users.

C. Factors affecting end-users ability to visualize

An important aspect for this novel technique to fully function, is to "understand the targeted audiences" and how these audiences (audiences are the end-users) perceive patterns and information. Different range of visualization, images are put out to test, and feedbacks are collected. Based on that, we provide the end-user guidelines when developing this UVisP visualization technique. These are:

- Visualization view must be of reasonable size.
- Visualization must be user-intractable.
- Visualization must contain less clicks if required inputs.
- Visualization must be simple and attractive, i.e. captures the viewer's eyes and mind.
- Visualization must contain reasonable colours (common colours - Primary, Secondary and Tertiary Colours).

Based on the end-users requirements, another challenge which we encountered is the ability to decide and allocate which colours, colour categories, and how many sets of colours would fully represent each visualizations. One main factor affecting this colour allocation, is the ratio or visualization nodes and actions to the different colours available. Another factor is the types of colours that are by research or theories identified as user-centric colours. Up till today, there has still been the debate of which colours, colour categories are considered user-centric. Therefore using a colour in any visualizations does have an effect on viewers, especially how they perceive the visualization and draw out useful information.

D. UVisP User Test Results

For this section we have carried out some testing with the UVisP technique to see how it performs. Our results are summarized and presented in Table I. But before we get to the Table II, we elaborate more on the Headers of Table II:

1) 'Vis-Type' in Table II means Visualization Type derived from the various Progger Scenarios.
2) This column of Table II shows how users provide a rating to different UvisP visualizations in terms of whether the UVisP Visualization provides useful information to users, and if knowledge has been extracted from the visualizations. 'Informative' means Informative (Scale of 1 - 5) 5 is most informative.
3) 'Easy/hard' in Table II means Easy/Hard understand (scale of 1 - 5) 5 is hardest. This refers to complexity level of understanding the various UVisP visualizations.
4) 'Gestalt Pr' in Table II means Gestalt Laws Visualizations (G.S-V, G.P-V, G.Pr-V, G.CF-V). This column refers to the different Gestalt Laws put into Visualizations; Gestalt Simplicity Visualization; Gestalt Proximity Visualization; Gestalt Pragnanz Visualization; and Gestalt Common Fate Visualization.
5) 'AvTime' in Table II means Average Time of visualization (sec). These values are measures by taking the time a random user begins observes the UVisP

visualization sample to the moment, he/she identifies useful information, knowledge from the visualization presented.

TABLE II
UVisP User Test Results

Vis-Type	Informative	Easy/Hard	Gestalt Pr	AvTime
Vis_Type_1	4	2	G.S-V, G.CF-V	60
Vis_Type_2	4	1	G.P-V	120
Vis_Type_3	5	1	G.Pr-V	60
Vis_Type_4	2	3	G.S-V	180
Vis_Type_5	4	3	G.S-V, G.Pr-V	180
Vis_Type_6	3	3	G.S-V, G.CF-V	120

Table II shows the relation between the different UVisP visualizations and the related time spent by random users to visualize and obtain useful knowledge, patterns and information from them. There are six different visualization samples (Vis_Type_1 to 6) used for the user test and based on the users observations, they have provided us with such information such as how informative the visualizations are; what time of Gestalt's Principle was used to identify the visualizations and relevant information needed. Finally, the time spent on observing the images for each random users was recorded as well, which indicated that the users took between zero seconds up to 180 seconds to process each UVisP visualizations and provide useful feedback on their observations. For example, Vis_Type_1 (in Table II) took up tp 60 seconds to view the visualization by the users of the UVisP technique.

Having to spend as less time on visualizing the different UVisP visualization samples, we can say that the UVisP's User-centric technique has proved to be useful and it provided the relevant missing links between a user and the different visualizations that are available to view. These missing links are the channel/bridge between the data-segment, visualization-segment and the user-segment of the whole visualization discovery process as shown in Fig. 1.

Another factor is the appearance, design and methods used to create various UVisP visualization. For example, Vis_Type_1 is created in a circular visualization form to show completeness, proximity, and Pragnanz. This relates to the scenarios of the visualizations. this also shows simplicity in visualizations. Vis_Type_2 (Fig. 6) UVisP visualization shows a detailed viewed visualization. it also relates to the time, relations of specific data. The idea of having it in a form of Tree visualizations is because it links the relations between the movements of the data and related files required for certain data process.

E. Analysing User Perception

We further carry out other methods to measure how effectiveness our UVisP technique. This is done by measuring user perception using several key indicators such as; user response to the clarity of visualizations, colour factors used to show certain key patterns/ideas in visualizations, and the time taken from when the user confronts the visualization.

Application	Vis Clarity	Com-Colour Factor	Av-Min Time Required
Patterns	High	Brigther	3 minutes
Vis Layout	Medium	No Effect	5 minutes

Table III shows the relationship and performance between applying visualization patterns and layouts (Fig. 9) into UVisP against visualization clarity and the use of common colours in visualizations (Fig. 7). An average time is taken across different users with different age groups and different level of knowledge on visualization. The results shows that having prior knowledge of possible patterns (Fig. 8) and applying it to UVisP technique with common colours enhances higher visualization clarity. This results a mean time taken of three minutes between several UVisP visualization samples.
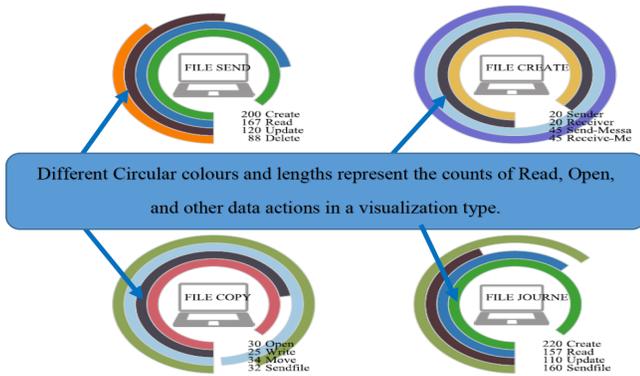

Fig. 7. UVisP Sample 2 - UVisP Visualization showing system call statistics.
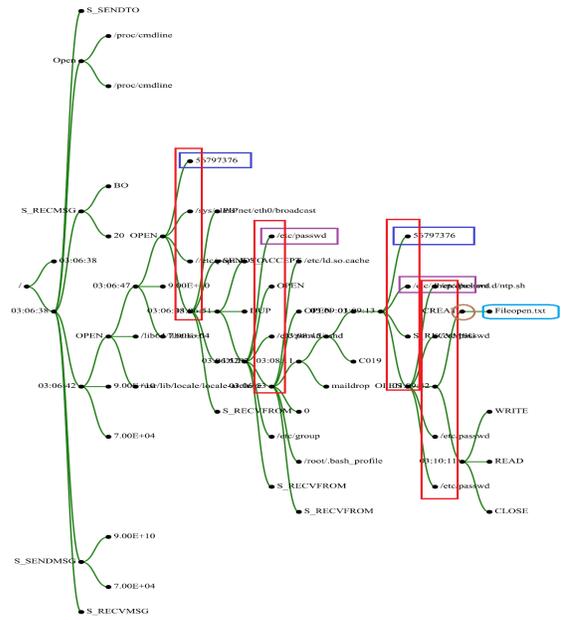

Fig. 8. UVisP Sample 1 - UVisP Visualization showing patterns and the provenance of Fileopen.txt.
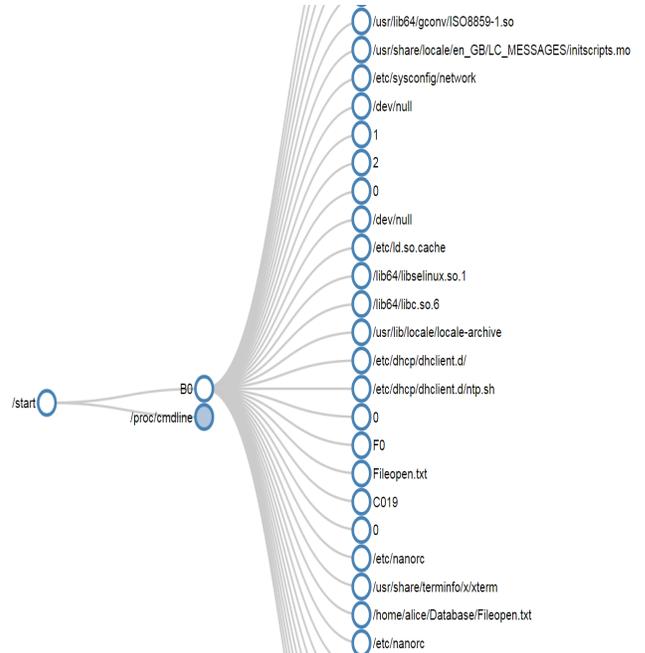
Fig. 9. UVisP Sample 3 - Detailed view Visualization.

V. CONCLUSIONS AND FUTURE WORK

We aimed to find and propose techniques that will transform provenance in the form of raw data (logs) into a final output that end-users could easily understand. Visualization is proposed solution. This research presents "User-centric Visualization of Data Provenance with Gestalt (UVisP)", a novel user-centric visualization technique to analyze existing visualization which empowers users to interpret their data provenance elements and security throughout their data life cycle. The inclusion of the Gestalt theory of perception to the technique, provides a range of visualization that allows the end-users to visualize and interact with an aim to obtain useful information such as tracking data provenance. This UVisP technique proves to work well for users who have the prior Gestalts theory of perception knowledge and have identified the perceptual key (s) within the visualization. However, for the Gestalt theory of perception and the perceptual key (s) to fully function, there has to be ways of integrating the Gestalt

Laws / principles with the types of visualization. The UVisP technique with its methodology provides the procedure and steps on how to obtain the final user-centric visualizations. With UVisP technique in place, end-users also have the opportunity to interact with different log files in the form of visualization. These logs represents the type of visualizations, which are then transformed and integrated with the Gestalt theory of perception to produce a range of user-centric visualizations.

Finally we provide a user-centric visualization technique that offers less clicks front-end to visualize provenance. Results from the UVisP technique indicated that our visualization samples are considered user-centric. Therefore we conclude that, the novel technique is user-centric in a sense that the visualizations presented are easy to understand, informative and does indicate the Gestalts Laws of Proximity, Similarity, Pragnanz and Common Fate. This work will form the basis of several new research directions which will be discuss in the following section.

While conducting this research, the following potential future work areas were identified:

1) Provenance Database System: The need to store data into a well-managed Database system is required. This will allow fast queries requests from the front-end visualization technique. For example, such database systems as 'CouchDB' and 'Hadoop' should be further studied for better implementation.

2) Processing Time: A Data processor to speed up the time between processing the data and loading it into D3 for visualization. For example caching and distributed algorithms can be implemented.

3) A Middle-man Processor / Filter: The purpose to have a middle-man processor/ filter between the front-end interface and the storage database, is to sort out and process the datasets (Provenance) without tampering with the dataset. The process data is then channelled into the front end for visualization.

4) The Need to Merge UVisP with Web Security Tools: Further investigation into merging UVisP with existing web security tools such as "Escrow" [27] in order to visualize the findings/logs as means of reporting and to provide a proof of concept to the Escrow end-users.

ACKNOWLEDGEMENT

REFERENCES

[1] R. Marty, Applied Security Visualization. Addison-Wesley Professional, 2008.

[2] Gestalt Psychology in German Culture, 1890-1967 Holism and the Quest for Objectivity — History of science and technology, Cambridge University Press.

[3] C. D. Green, Classics in the History of Psychology – Wertheimer (1923). [Online]. Available: http://psychclassics.asu.edu/Wertheimer/Forms/forms.htm. [Accessed: 04-Jan-2015].

[4] I. Wassink, O. Kulyk, B. van Dijk, G. van der Veer, and P. van der Vet, Applying a User-centered Approach to Interactive Visualisation Design, in Trends in Interactive Visualization, R. Liere, T. Adriaansen, and E. Zudilova-Seinstra, Eds. Springer London, 2009, pp. 175-199.

[5] A. Komlodi, P. Rheingans, U. Ayachit, J. R. Goodall, and A. Joshi, A user-centered look at glyph-based security visualization, in IEEE Workshop on Visualization for Computer Security, 2005. (VizSEC 05), 2005, pp. 21-28.

[6] Apache Pig, Hortonworks. [Online]. Available: http://hortonworks.com/hadoop/pig/. [Accessed: 01-Feb-2015].

[7] J. Clark, Neoformix - Blog, 05-Nov-2014. [Online]. Available: http://www.neoformix.com/archive.html. [Accessed: 11-Jan-2015].

[8] M. Bruls, K. Huizing, and J. J. van Wijk, Squarified Treemaps, in Data Visualization 2000, D. ir W. C. de Leeuw and ir R. van Liere, Eds. Springer Vienna, 2000, pp. 33-42.

[9] B. B. Bederson and B. Shneiderman, The Craft of Information Visualization: Readings and Reflections. Elsevier, 2003.

[10] C. Vermehren, The Art of Web Analytics, 12-Oct-2009. . [Online]. Available: http://www.theartofwebanalytics.com/. [Accessed: 19-Jan-2015].

[11] S. Vieira, Easily create stunning animated charts with Chart.js, Webdesigner Depot, 04-Nov-2013. .

[12] Ward, Matthew O., Georges Grinstein, and Daniel Keim. Interactive data visualization: foundations, techniques, and applications. CRC Press, 2010.

[13] M. Ankerst, D. A. Keim, and H.-P. Kriegel, Circle Segments : A Technique for Visually Exploring Large Multidimensional Data Sets, 1996.

[14] D. A. Keim, M. Ankerst, and H.-P. Kriegel, Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data, in Proceedings of the

[15] J. Cheney, S. Chong, N. Foster, M. Seltzer, and S. Vansummeren, Provenance: A Future History, in Proceedings of the 24th ACM SIGPLAN Conference Companion on Object Oriented Programming Systems

[16] Y. L. Simmhan, B. Plale, and D. Gannon, A Survey of Data Provenance in e-Science, SIGMOD Rec., vol. 34, no. 3, pp. 31-36, Sep. 2005.

[17] Y. Simmhan, B. Plale, and D. Gannon, A Survey of Data Provenance Techniques. Technical Report IUB-CS-TR618.

[18] P. Buneman, S. Khanna, and W. C. Tan, Data Provenance: Some Basic Issues, in Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science, London, UK, UK, 2000, pp. 87-93.

[19] P. Chen, B. Plale, Y. Cheah, D. Ghoshal, S. Jensen, and Y. Luo, Visualization of network data provenance, in 2012 19th International Conference on High Performance Computing (HiPC), 2012, pp. 1-9.

[20] Gestalt Psychology in German Culture, 18901967 Holism and the Quest for Objectivity — History of science and technology, Cambridge University Press. [Online]. Available: http://www.cambridge.org/nz/academic/subjects/history/history-science-and-technology/gestalt-psychology-german-culture-18901967-holism-and-quest-objectivity. [Accessed: 04-Jan-2015].

[21] K. Koffka, Perception: an introduction to the Gestalt-Theorie., Psychological Bulletin, vol. 19, no. 10, pp. 531-585, 1922.

[22] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt, A century of Gestalt psychology in visual perception: I. Perceptual grouping and figureground organization., Psychological Bulletin, vol. 138, no. 6, pp. 1172-1217, 2012.

[23] V. Tan, P. Groth, S. Miles, S. Jiang, S. Munroe, S. Tsasakou, and L. Moreau, Security Issues in a SOA-Based Provenance System, in Proceedings of the 2006 International Conference on Provenance and Annotation of Data, Berlin, Heidelberg, 2006, pp. 203-211.

[24] Boundless, Perceptual Mapping, Boundless, Nov. 2014.

[25] M. Bostock, D3.js - Data-Driven Documents, Data-Driven Documents, 2013. [Online]. Available: http://d3js.org/. [Accessed: 20-Apr-2015].

[26] R. K. L. Ko and M. A. Will, Progger: An Efficient, Tamper-Evident Kernel-Space Logger for Cloud Data Provenance Tracking, in 2014 IEEE 7th International Conference on Cloud Computing (CLOUD), 2014, pp. 881-889.

[27] B. Delamore and R. K. L. Ko, Escrow: A Large-Scale Web Vulnerability Assessment Tool, in 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2014, pp. 983-988.

[28] Harrison, R. Spahn, M. Iannacone, E. Downing, J. R. Goodall, and J. R. Goodall, NV: Nessus Vulnerability Visualization for the Web, 2012. .

[29] P. Trinius, T. Holz, J. Gobel, and F. C. Freiling, Visual analysis of malware behavior using treemaps and thread graphs, in 6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009, 2009, pp. 3338.