HAL open science

Designing a Decision-Support Visualization for Live Digital Forensic Investigations

Fabian Böhm, Ludwig Englbrecht, Günther Pernul

To cite this version:

Fabian Böhm, Ludwig Englbrecht, Günther Pernul. Designing a Decision-Support Visualization for Live Digital Forensic Investigations. 34th IFIP Annual Conference on Data and Applications Security and Privacy (DBSec), Jun 2020, Regensburg, Germany. pp.223-240, 10.1007/978-3-030-49669-2_13 . hal-03243642

HAL Id: hal-03243642

https://inria.hal.science/hal-03243642v1

Submitted on 31 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Designing a Decision-Support Visualization for Live Digital Forensic Investigations

Fabian Böhm[1 [0000−0002−0023−6051]] (corresponding author), Ludwig Englbrecht[1 [0000−0002−8546−3017]], and Günther Pernul[1]

Universität Regensburg, 93053 Regensburg, Germany
{fabian.boehm | ludwig.englbrecht | guenther.pernul}@ur.de

Abstract. Fileless Malware poses challenges for forensic analysts since the infected system often can't be shut down for a forensic analysis. Turning off the device would destroy forensic artifacts or evidence of the fileless malware. Therefore, a technique called Live Digital Forensics is applied to perform investigations on a running system. During these investigations, domain experts need to carefully decide what tools they want to deploy for their forensic analysis. In this paper we propose a visualization designed to support forensic experts in this decision-making process. Therefore, we follow a design methodology from the visualization domain to come up with a comprehensible design. Following this methodology, we start with identifying and defining the domain problem which the visualization should help to solve. We then translate this domain problem into an abstract description of the available data and user's tasks for the visualization. Finally, we transform these specifications into a visualization design for a Live Digital Forensics decision-support. A use case illustrates the benefits of the proposed method.

Keywords: Digital Forensics · Visual Analytics · Live Forensics · Visualization Design.

1 Introduction

Malware has been around since the early days of computers. While traditional malware relies on malicious executable files, there is one particularly evil type of malware: Fileless Malware (FM). This type is hard to detect as it hides itself in locations that are difficult to analyze [31]. It exists exclusively in memory-based areas like the RAM instead of being written directly on the target's hard drive. This complicates forensic investigations of FM as most traditional Digital Forensic analysis techniques are designed to work on computers after they got turned off [16]. However, as FM solely exists in memory, turning off the target would lead to significant loss of evidence. Although some evidence of FM can be acquired through traditional DF analysis techniques, keeping the potentially infected system running allows the investigator to gather additional evidence occurring during an incident. Moreover, there are mission-critical systems that simply cannot be shut down in order to not disrupt business operations. Therefore, Live Digital Forensics (LDF) is necessary.

LDF allows domain experts to investigate a running system, identify artifacts and collect evidence. This helps to understand FM-based attacks but at the same time requires fast and careful decisions about the LDF tools used to carry out the analysis. A poor choice of the analysis tool could destroy or compromise important artifacts.

In order to support forensic analysts to make faster and better decisions upon which tools should be used during an LDF investigation or upon which indicators might need additional attention, we propose to apply Visual Security Analytics (VSA). VSA allows domain experts to interactively explore the data of the system under investigation. It supports the decision-making process by allowing the forensic investigators to assess the current situation with a tailor-made visualization approach for a specific situation [29]. Therefore, they can lead the attention towards possible indicators for FM and deploy the respective LDF analysis tools like *volatility*[1] or *SysAnalyzer*[2].

This paper shows our process of developing a visual representation aimed to help Digital Forensic experts with directing their attention throughout their analyses. We follow a methodological design approach to bridge the gap between domain (digital forensic) and visualization experts [17, 30]. Our main contribution is the methodological design of a visual decision-support system aiding forensic experts to direct their further investigations during a live forensic analysis. We introduce the methodology, derive a design from the requirements and problems within the LDF domain, and evaluate our design by showcasing the identification of a fileless malware's artifacts within a live forensic analysis.

The remainder of this work is structured as follows: Section 2 identifies and summarizes related work within the digital forensic analysis domain and existing visualization approaches. We describe the applied methodology to design the visualization in Section 3. The first step of our methodology is a characterization of the domain problem in Section 4. Section 5 follows the remaining steps of the methodology to design a comprehensible visualization for the characterized domain problem. This design is afterwards evaluated in Section 6 by showcasing how artifacts of the fileless malware *Poweliks* can be identified and how this helps to guide further investigations. We conclude our work and point to further possible research in Section 7.

2 Related Work

A Live Digital Forensic analysis is performed on a running system during an ongoing incident. The data is collected and analyzed simultaneously. The focus is on the preservation and processing of semi-persistent or volatile traces. This could be the content of the RAM, active network connections or running processes and programs [1]. Since these traces are no longer available after a system restart, they cannot be extracted from a disk image by post-mortem analysis [13]. Live analysis is therefore useful if volatile data is essential for reconstructing an

[1] `https://www.volatilityfoundation.org/`
[2] `http://sandsprite.com/iDef/SysAnalyzer/`

incident. This is the case if the system cannot be shut down for reasons of availability or dependency, or if encrypted data systems can no longer be accessed after a restart, for example when analyzing a fileless malware [11].

A disadvantage of live analysis is that the process can often not be repeated after leaving the location of the seizure [11]. In addition, the analysis takes place in a potentially compromised environment, so that relevant traces can be hidden, for example by using rootkits [1]. Furthermore, in the context of live analysis, a modification of the system by the investigation activities is almost unavoidable [1]. These modifications should be as limited as possible and all activities in the system must be precisely documented [13].

It is challenging to prove in court that the data integrity of the digital evidence has been preserved throughout the entire digital investigation. This may lead to a reduction in the admissibility of the evidence or even to a prohibition of its use. However, there are methods for comprehensible documentation and differentiation between the actions of an Incident Response team and the activities of an active attacker [8]. Providing a profound and tamper-proof documentation of analysis steps reduces the possible impact on the admissibility of volatile evidence and/or its modification. Nevertheless, these methods usually have to be implemented in advance as Digital Forensics Readiness measures.

Additionally, post-mortem and live analysis are not competing approaches, but rather complement each other. Live analysis enables the extraction and processing of additional traces, which can considerably support post-mortem analysis and the reconstruction of the course of events [1].

We identify several related visualization approaches originating from both the Visual Analytics (VA) and the Digital Forensics research domains. Within the VA domain, the designs are often based on user-centered approaches to provide a solution for a specific, relevant task of forensic experts. These visualizations feature a broad variety of use-cases ranging from the forensic investigation of shadow volumes and directories [15, 14] to live monitoring of network traffic [4, 3]. Tools like EventPad [6] allow the interactive and explorative analysis of large, dynamic data sets to identify malware and its behavior. The KAMAS solution is a tool providing not only innovative automated malware analysis features but also the functionality for malware analysis experts to exchange domain knowledge with the automated analysis methods [28]. Although a variety of related visualization designs exists in the VA domain, none of these visual representations is specifically designed to support the decisions forensic investigators need to make during an ongoing live forensic investigation. The same applies to the VA approaches introduced in the DF research domain. Tools like Timelab [23], LogAnalysis [7], MalViz [22], Vera [26], or Devise [27] allow a visual representation of different types of data for static forensic investigations but are by no means capable to support fast, dynamic decisions for live forensics.

None of the above-described visualization approaches pays special attention to the decision-support required throughout an LDF investigation. Additionally, to the best of our knowledge there is no existing work on bridging the gap between the domains of Live Digital Forensics and Visual Analytics by applying

methodologies to develop comprehensive and reproducible visualization designs. Therefore, the knowledge from the Visual Analytics domain is beneficial as it pays attention to design aspects that are being neglected up to now in the LDF domain. We aim to contribute to a transfer of knowledge from the VSA towards the LDF domain in this research as it has been done in other security-related domains within the last years [25].

3 Methodology

This section summarizes the methodology which we follow throughout this work. A methodological approach allows our design decisions to be reproducible and comprehensible. Especially in visualization design this is of utmost importance because even methodologically based decisions remain subjective [18]. Therefore, we follow the Nested Blocks and Guidelines Model (NBGM) which is a well-established methodology for designing visualizations [19]. Another important aspect of the NBGM is that it is aimed to support the collaboration between domain and visualization experts and, therefore helps to close the aforementioned gap in LDF visualization designs [30]. The high-level layers of the NBGM are depicted in Figure 1 and described in the subsequent sections [19, 21].

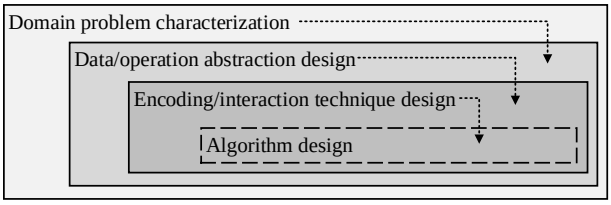

Fig. 1. Nested layers of the NBGM [19, 21].

Domain problem characterization: The main task in this first layer is the identification of the specific situation and problem for which the visualization should be designed. The tasks and data of the target group are identified including their workflows and processes. Each target domain has its own descriptive vocabulary and it is important within this phase to work with the target users using their familiar vocabulary. This layer of the nested model bridges the gap between visualization experts and domain experts as it allows designers to understand the world the domain experts work in and which problems they face [32].

Data/Operation Abstraction Design: The second level abstracts the domain problem characterization using the vocabulary of the visualization domain. Therefore, it describes visualization tasks and the data relevant to

the design. Domain-specific tasks and data descriptions are translated into a visualization-specific vocabulary. This way, visualization designers identify what tasks (e.g. finding outliers, identifying trends) the domain experts have to solve from a visualization point of view. The tasks of the users in visualizations can be derived from a variety of existing task taxonomies [5, 28]. Additionally, the data abstraction allows designers to describe data transformations of available data identified within the domain problem characterization into a different format if necessary, for subsequent encoding technique decisions.

Encoding/Interaction Technique Design: This layer describes the visualization (encoding) techniques and the necessary interactions for users. Both, encoding and interactions must be aligned together and are derived from the visualization tasks in combination with the data at hand from the data / operation abstraction design-layer. Encoding and interaction techniques combine the first two nested layers with a design that instantiates the abstract visualization for the domain problem.

Algorithm Design: The innermost layer of the NBGM requires to create appropriate algorithms carrying out the beforehand designed encodings and interactions. We do not consider this step in our current work and focus on the first three layers of the model. The final implementation of the design is part of our further research.

4 Decision-support for Live Forensics

In the case of an LDF investigation, decisions can be directly linked to the risks involved. Therefore, it is important to make well-considered decisions when choosing the right techniques, tools, and artifacts. In this section, we characterize the domain problem to enable a suitable visualization design helping domain experts facing this domain problem. We emphasize the supporting effect of the visualization for a digital forensic examiner. In particular, the tasks during a live digital forensic investigation are discussed. The goal is to apply the design methods of visualization experts to support better decision-making for a domain expert.

4.1 Live Digital Forensics Process

The collection and analysis of digital evidence should be based on a defined comprehensive process model. A common description of a forensic investigation process is represented by the model of Kent et al. [13]. The investigation process is divided into four phases as depicted in Figure 2. We have extended the original approach to include an overarching decision-support by an interactive visualization at every stage. The following paragraphs describe the different original stages, which need a decision-support:

Collection: Data related to the criminal activity are identified, labeled, recorded, and secured from all potential sources of relevant data [13]. Possibly relevant

additional data sources might be identified, and respective data needs to be collected during an LDF analysis.

Examination: The data collected in the previous phase is evaluated. The aim is to identify and extract relevant data [13]. Since our approach is applied to a live investigation, a visual analysis of the data allows the decision to include additional data sources for the analysis. Consequently, the visual decision-support creates a return to a previous phase.

Analysis: The results of the previous phases are analyzed in depth and interpreted to establish connections between persons, places, objects, and events and to obtain useful information regarding specific questions [13]. Findings from the visual analysis are directly incorporated into the analysis process. Malicious activities can be better understood through a visual representation of the data.

Reporting: The results of the analysis are prepared and presented, including important information. The format and content of the report depend on the type of recipient [13]. Especially, visual representations can contribute considerably to the understanding of the incident. Particularly, if the attack is complex, spikes in network traffic or system performance can provide a good insight on the activities during the incident.

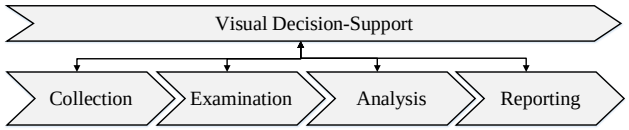

Fig. 2. A high-level process for Live Digital Forensics and Visual Decision-Support.

4.2 Tasks of Domain Experts in Live Digital Forensics

Mistry and Dahiya [20] discuss the volatile memory forensics approach in detail. Using live forensics, real-time data is analyzed and stored based on the system activities. The analysis of the memory (RAM) is very important while considering live computer forensics. The approach of live forensics plays an important role in identifying Indicators of Compromise (IoCs) and recording volatile data, which would be lost after shutting down the system. The authors use *memory forensics* to run through various challenging scenarios and prove their approach based on previously extracted and identified data in real-time. Since their approach provides a good description of the domain experts' workflow and it has been used by the authors in several scenarios, this is further considered. We abstract and extend the original approach as a baseline (see Table 1) to identify the main tasks during an LDF analysis:

Table 1. Summarized expert tasks in LDF.

Task	Details
Data Acquisition	– Identify suspected devices and media – Dump RAM, cache, and network traffic – Acquire an image of system (if possible)
Establish Intelligence	– Parse memory structure – Identify relevant memory segments – Identify loaded modules – Identify running processes and file accesses – Identify established network connections
Memory & Data Analysis	– Search outliers and irrelevant information – Extract additional relevant data – Verify findings for further decisions – Decide the next analysis steps
Documentation	– Document interesting findings – Document artifacts and evidence

- **Data Acquisition:** Within this task, investigators need to decide which data they export from the device under investigation. During an LDF analysis, only a limited amount of data can be extracted. An additional limitation for this task is often, that data only can be extracted with a-priori implemented functionalities.
- **Establish Intelligence:** This step is very much based on the present situation and requires that the investigator has a good sense of the specific case. Usually this is due to the prior knowledge of the investigator. It is important that in this step no analysis in the actual sense is carried out, but rather the region for possible purposeful evidence is identified. A graphical processing by means of VSA can contribute significantly to this. Especially decisions about the inclusion of further areas are very time-critical and a visual representation can contribute to a fast identification.
- **Memory & Data Analysis:** In this step, the previously identified data is examined for suspicious features. In addition, the findings are put into context to reconstruct the course of events. By a supporting effect of VSA, outliers and correlations can be better found.
- **Documentation:** The aforementioned tasks are documented during the whole digital forensic investigation to be used in the final report. This is an essential component to make the investigation comprehensible. During an analysis using VSA, findings based on a graphical preparation can be documented in the figures using markers (e.g. at peak values).

4.3 Available Data in Live Digital Forensics

Harichandran et al. [10] formed the term *curated forensic artifact (CuFA)* to specify the scope of forensic artifacts and their supervised attributes. The Arti-

fact Genome Project (AGP), based on CuFA's principles, was launched in 2014 and has received 1099 forensic artifacts within the last few years [9]. It reached an acceptable level of maturity, as registered participants can contribute to this project by uploading artifacts along 19 categories.

Crimes are committed in several ways, and the expedient evidence is accumulated by different forensic artifacts. Depending on the peculiarity of a case, digital evidence either adds more value to an investigation or is completely inappropriate. The ontology of crimes by Kahvedzic et al. [12] provides a specification of past criminal cases and offers the possibility to specify almost every cyber case. We summarize the sub crime cases and focus only on cyber-crime cases.

The violation of the quality of forensic artifacts influences their admissibility at courts. Because of the fast-moving nature of digital evidence, we adopt the legal requirements by Antwi-Boasiako et al. [2] due to their overall completeness and applicability. This framework is appropriate for forensic investigations and reduces the overall scope of common data quality dimensions. These legal requirements cannot be circumvented, as admissibility in court is indispensable. AGP represents an open-source platform based on the CuFA principles. The following forensic artifacts categories have been extracted: Windows registry, memory, file, network packet, process, email message, address, code, disc partition, account, network socket, disk, user account, X509 certificate, user session, windows event log, volume, and Linux packages. These categories are further reduced since our concept focuses on LDF and not all categories are available in this type of investigation. To illustrate the possibilities of our approach, we will focus on the following categories of data sources that can be accessed during a live forensic analysis (without major interference due to the installation and execution of additional applications): *file access, network packets, process-lists, event logs (including PowerShell) and system statistics.*

VSA can support understanding and interpreting the context data in combination with stored data. Therefore, VSA allows experts to make better, context-based decisions for further investigations.

5 A Design for Visual Decision-Support in Live Forensics

Based on the domain problem which arises for forensic experts during a live forensic investigation (see Section 4) we derive appropriate visualization tasks, visual encoding and necessary interaction functionalities for a visual decision-support system within this section. Figure 3 depicts an overview over the respective, fully defined NBGM model for this problem domain.

The central contribution of this part of our work is the innovative application of different encoding techniques combined as an interactive, coordinated view where interactions in one view influence the representation in others. This allows a lateral, visual movement in the data enabling forensic analysts to browse through the data and identify artifacts that need further investigation with specific forensic analysis methods. However, without the visual decision-support

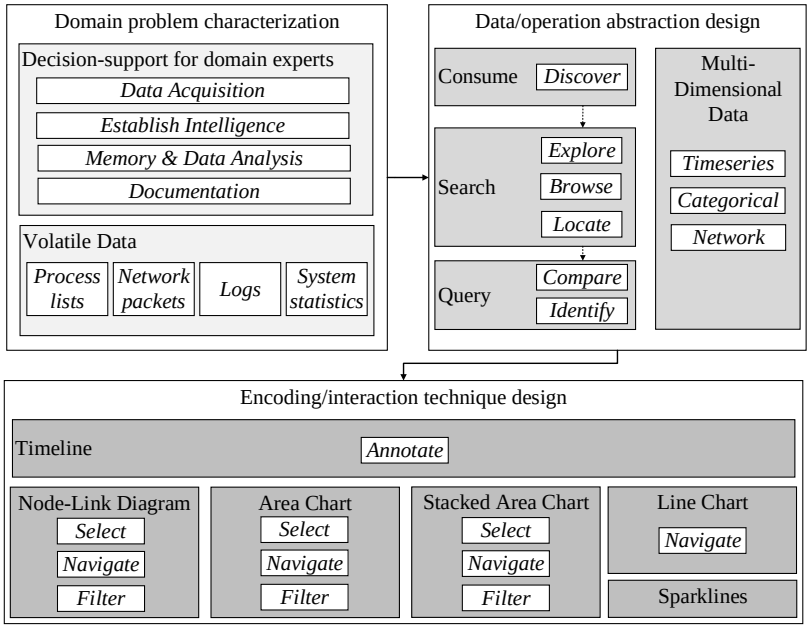

Fig. 3. NBGM applied for LDF covering the results Sections 4 and 5.

they might not even have spotted the artifact. Therefore, we strengthen the necessity of a visual design as proposed by us in the following sections.

5.1 Data/operation abstraction design

This section covers the abstraction of the domain problem characterization above. This step in the NBGM is carried out via designing and identifying data blocks and task blocks that describe the needs, requirements, and problems of the forensic experts. Defining these blocks, subsequently allows a comprehensible decision for specific visual encoding and interaction techniques.

Task blocks In Section 4.2 we identify several tasks that are important for forensic experts when they are performing an LDF investigation. A visual design for decision-support during these investigations needs to support the experts in these tasks. To be able to transfer the domain problems into a suitable visual design, we identify abstract visualization tasks from the *Why?* part of Brehmer and Munzner's task typology [5]. This typology describes why a specific task is performed in terms of which goal a user is pursuing.

The abstract, process-like overall task of forensic experts is to get an idea or indication where to further direct the ongoing LDF investigation and therefore, which tools they should deploy (see Section 4.1). We summarize this high-level

task as "Decision-support for domain experts" in Figure 3. However, the process-based task of decision-support can be split into several abstract visualization tasks. The main goal for forensic experts from a visualization point of view is to *Discover* which is a task formed around the generation and verification of hypotheses. An exemplary hypothesis in this context might describe which malware is acting on a device or which forensic tools need to be deployed to continue the investigation.

The *Discover* task is further specified depending on whether the analyst has a hypothesis in mind when using the visualization or not. When the LDF investigation is carried out to find evidence for a specific malware with known indicators acting on the device, this corresponds with the *Locate* task. *Browsing* outlines actions to search through suspicious indications within the visualization to find out which investigative tool might help to continue the analysis. The remaining task at this level, *Explore*, represents an analyst exploring the displayed data to identify possible suspicious patterns in the data and therefore, to make decisions on possible malware types on the device or at least additional LDF tools to deploy on the device. Once a hypothesis about malware or LDF tools is made, the forensic expert continues to *Identify* additional characteristics of the malware or tries to further strengthen the need for the LDF tool. This is possible by using the visual representation of our design.

Data blocks The available volatile data described in Section 4.3 is mostly defined as multi-dimensional data from a visualization point of view. However, there are some relevant subcategories of data types and formats which significantly influence the decision for corresponding visual encoding.

All available data that needs to be visualized within the decision-support for LDF is time-based data as it relates to some characteristics or actions of the system at a specific point on time, e.g. the network connections at a specific time or the respective CPU and RAM workload. Additional data categories can be categorical data like the different severity types of collected logs and network data.

5.2 Encoding/interaction technique design

This section connects the different aspects between the problems of the forensic experts and the abstract blocks derived in Section 5.1 by introducing the visual encoding and corresponding interaction techniques that are necessary to enable a visual decision-support for LDF investigations. The visual encodings described in the following section are mainly well-known and established visualization techniques. We decided to only use these to ensure easy and fast perception of the design for forensic investigators. The encoding techniques are derived from the available data blocks while the interactions are necessary for the forensic experts to follow their tasks using the visual encoding of the data. The design sketch for the decision-support visualization is shown in Figure 4. It comprises five interactive main components that are further detailed within this section. In terms

of more abstract visualization tasks, the design allows *Navigation*, *Selection*, *Filtering*, and *Annotation* [5].

Fig. 4. Resulting design sketch of a decision-support visualization for Live Digital Forensics investigations.

Investigation Timeline The first component is the overarching *Investigation Timeline*. This component is necessary since most of the data represented within the design is time series data (see Section 5.1). Therefore, a timeline allows navigating through different points in time of the collected data and analysts can select a specific time window for their analysis. The selected time window is indicated by the small white box-shaped overlay on the timeline and the time-range on the right side. In the design sketch of Figure 4 a window of two minutes between 12:00:00 and 12:02:00 is selected. However, the box overlay can be moved across the timeline and can also be resized to allow the selection of different time ranges. The other four components of the visualization design display only data from the selected time window. An additional functionality of the *Investigation Timeline* is the event annotation. Forensic analysts can mark and label specific points in time when they identified possible evidence or interesting artifacts. This allows to come back to these events in a later investigation or even the collaboration of multiple analysts where one can pick up the investigation on a mark added to the timeline by another analyst.

Network Activity The next component displayed in the upper left corner of Figure 4 is the *Network Activity*. This view aims to give an overview over the

device's external activities regarding the endpoints and IP addresses it communicated with. In the center of this view the device under investigation and process IDs (PIDs) is shown. To keep this representation clearly laid out only PIDs with active connections during the selected time frame are displayed. The connection partners are illustrated with ellipses labeled by IP addresses. The connection targets are clustered by IP address range allowing to distinguish different networks. In the exemplary design sketch, for example, the local network of the device is clustered on the left of the *Network Activity* clearly separated from external connection targets on the right.

We include the connections between a process and its communication partner by adding directed links for both incoming and outgoing communication. The color coding of the links is dependent on the cumulative number of bytes sent through the connection with a scale from blue (i.e. few bytes or "cold connection") to red (i.e. many bytes or "hot connection"). The distinguishable and color-coded links for up-link and down-link connections allow to quickly detect large data flows and to identify the process responsible for this data flow. Examples for possible artifacts needing additional analysis with sophisticated LDF tools are the imaginary process with the PIDs 2345 and 3456. The first one is sending a lot of data to an external IP address while the other one is downloading numerous bytes from another address.

Clicking on a PID highlights the connections of the selected process in this view but also in the *System Performance* and the *Read/Write Entropy* views where the activities of the process are highlighted respectively. This allows a quick indication about a process's overall statistics including its network activity as well as its CPU and RAM activity. Hovering a connection opens a thumbnail with additional information on this specific network communication between a process and an external IP address. This additional information contains the exact number of bytes sent over the connection as well as the port and protocol used to open it. A similar hovering interaction is also provided for the nodes depicting the communication partners of the device under investigation. The thumbnail for these nodes contains the total amount of bytes sent from and to this node as well as ports that were used to connect to. Hovering a node simultaneously also highlights the processes that established a connection with the corresponding IP address.

Read/Write Entropy In the upper right corner the decision-support visualization design features a *Read/Write Entropy* display. The area charts of this view show the entropy of both read and write operations on mounted drives of the investigated device. The x-axis of the charts encodes the selected time frame from the *Investigation Timeline* while the positive y-axis displays the entropy of the data read from the specific drive at a point in time on a range from 0 to 1 (0% to 100%). Analogously, the negative y-axis represents the same indicators but for data written onto the drive. Therefore, the entropy for write operations is indicated as a negative value in our design. This only serves to clearly distinguish positive and negative y-axis values in the area charts. In addition to

indicating the difference of read and write operations by indicating them with different vertical directions, they also are encoded with different colors. In the top right corner of this view, the drives for which the entropy values should be displayed can be selected via check-boxes.

This view allows a zooming interaction, preferably by mouse-wheel, where zooming in narrows down the displayed time window and zooming out analogously widens the time span. If experts zoom into this view, the time window is adjusted respectively for all other views. Possible artifacts that catching an expert's eye in this view are unusually high entropy scores for read or write operations. As an example, serves the increasing entropy scores in the design sketch towards the end of the selected time frame for both drives. This might indicate the writing of a lot of encrypted data on the two drives.

System Activity The left view in the bottom row of our design is a visual representation is also an area chart, but a stacked version. It provides a visual encoding of *System Activity* by displaying a count of system events (e.g. Windows Event Logs, Powershell Events, Syslogs). The events are colored depending on their type or severity allows experts to detect a rising number of errors or similar indications of artifacts. The check-boxes on the top right of the view allow enabling different event types to be displayed. The x-axes of the area chart are like a timeline while the y-axes indicate the cumulative count of currently displayed event types at a specific point in time. The stacked area chart allows identifying trends and changes in the logged activities of the system.

This chart is also allowing a zoom interaction like the previous *Read/Write Entropy* views. Within this view, an unusually high number of error logs in the Windows Event Log that is constantly appearing throughout the whole two minutes currently under investigation could be an artifact for further analysis.

System Performance The last view that is part of our visualization design located on the bottom right of Figure 4. It is split into two smaller views which in combination give an indication of the *System Performance*. The upper part of this view is occupied by a line chart with two different lines. The blue line depicts CPU performance while the second, orange line indicates memory or RAM activity. Both lines are on a relative scale, meaning that the y-axis ranges from 0 to 1. Both lines on the chart are again displayed for the selected window of time. The line chart allows a zoom interaction similar to the interaction described within the *Read/Write Entropy* and the *System Activity* views.

The lower part of the display contains a table with active processes during the time which is currently defined for analysis. The table has four columns for the process name, its PID, and a spark line visualization allowing a fast perception of this process's CPU and RAM activities. A spark line is a special, word-sized type of line chart. They are not displayed with any axes and serve a single purpose: to give an indication about the trend of a single indicator. The table might be longer than the five exemplary rows from our design sketch and therefore, needs to be scrollable. Rows can also be selected leading to a highlighting of the

corresponding process in the table and in the *Network Activity* view. Selecting a process also changes the *Read/Write Entropy* view by now only showing the entropy of the read or write operations performed on behalf of this specific process. This enables forensic analysts to conclude on the influence of a process on the systems performance and possible correlations with network activities.

6 Use Case

To show how our visualization design can support forensic experts in their LDF investigations, we go through a short use case featuring a well-known and documented fileless malware attack. We describe how indicators of this malware become apparent within our visual decision-support and how we support the tasks of forensic experts during an LDF investigation identified in Section 4.2.

The use case features the fileless-malware *Poweliks* which attacks Windows-based systems. This malware became known as a file-based piece of malicious code but in 2014 it moved to a file-less variant. After computers are infected they are part of a click-fraud botnet where bots request advertisement data from a central Command-and-Control (C&C) server, load the ads and click them to generate revenue [24]. As a side effect, *Poweliks* often acts a door-opener for other malware as it clicks up to 3000 ads per day on a single computer and does not care about whether the ads are malicious or not. Although this malware attracted attention back in 2014, its design is special in two aspects. *Poweliks* acts without leaving a file on the computer's file system. It stores all the data it needs in the registry and memory by injecting code into legitimate processes currently running. Therefore, it is hard to detect once it gained a foothold on the system. The second interesting aspect of *Poweliks* is, that, despite being a fileless malware, restarting the infected device does not remove it as it reboots itself from altered registry keys. This makes *Poweliks* a very special and dangerous type of fileless malware [33]. Because of those characteristics, we choose to describe how indicators for the *Poweliks* malware are visible within our visualization design.

Based on publicly available details and threat hunting details about the ad-fraud variant of *Poweliks*, we describe indicators that can be detected within our design, helping forensic analysts to make decisions where to guide their attention for further analyses. We structure the indicators and their identification according to the different views of our visualization design (see Section 5.2).

Network Activity: Regarding the network activity of an infected system, there are several indicators becoming apparent within a visual display. First, *Poweliks* is known to download the Powershell as well as the .NET framework from official Microsoft download pages if not available on the computer. The respective connections might appear in the view as connections of processes to official Microsoft IP addresses and a high payload on the down-link transfer, i.e. the link between the Microsoft IP and the process turns red.
Additionally, as the malware acts as a botnet, it regularly connects to its C&C server. These are only short connections with a very limited payload.

However, as they appear on a very regular basis, they can be identified as an indicator for further analysis why the system is connecting to the respective IPs.

Another suspicious activity to be spotted via the proposed design is the ad-clicking component of *Poweliks*. The behavior of requesting ad data from the C&C server, contacting a search page for the URL of the ad, and clicking the loaded advertisements becomes recognizable as the network activity would show many small-scale connections to a lot of different, external IP addresses. This is all more suspicious when the respective network connections are originating from a single process.

Read/Write Entropy: Overall, activities on the file system is less apparent as *Poweliks* is a file-less malware. However, as the malware can request up to 3000 ads per day on a single computer it is very likely that the malware also "clicks" other malicious ads. The entropy of read and write operations on different drives of a computer shall light on possible ransomware being active due to *Poweliks*' activities. Increasing entropy values in this view indicate the transfer of encrypted data. This highlights the necessity for domain experts to further investigate this malware since it could have features of a ransomware.

System Activity: *Poweliks*' special fileless persistence method uses a watchdog and PowerShell scripts when it is establishing its foothold. It also modifies many key registry entries trying to lower or disable browser security settings to be able to perform the ad-clicking behavior. Both of these actions produce log events (Windows Event Logs, PowerShell Events, etc.) with different severity. However, as the performed behavior is rather uncommon, the *System Activity* view shows several warnings and errors to the domain experts. Therefore, they might for example decide to analyze the changes made to the key registry in depth.

System Performance: Also, the system performance is not too bad during the execution of the *Poweliks* malware. This is because the malware does not want to significantly affect the performance of the infected computer. However, with our concept it can be seen that the CPU and RAM are used when a web page is accessed in the background for a few moments. This is due to the fact that the browser has to interpret and render the website. If the observed computer is not running other programs, this is also a possible indication of the malware. These findings by using a visual display during a live investigation help the forensic examiner to better assess the current situation and to make a well-considered decision for the use of certain tools. Also, the display of the processes and their RAM as well as CPU indicates possible further investigation needs. In the case of *Poweliks* which is hiding in different common processes (e.g. cmmon32.exe, dllhost.exe, logagent.exe), unusual activity of those processes indicates further investigation potential. Especially, when these processes are involved in anomalous network activity as well.

However, the malware *Poweliks* will be detected by current virus scanners but a coming back by a modification of the malicious code or behavior is very likely since file-less techniques evolved in the last few years. Nonetheless, they are relevant artifacts helping forensic experts to guide their further analyses.

7 Conclusion and future work

Within this work we made a contribution utilizing Visual Security Analytics as a decision-support approach for Live Digital Forensic investigations. We describe and abstract the problem of forensic investigators which have a wide variety of tools at hand for their analyses but need to decide quickly which of them need to be deployed in the current situation. To support them in this decision-making process we applied a methodology derived from the visualization research domain. Contributing to this domain problem with a tailor-made visualization approach enables forensic investigators to make faster and well informed decisions. We described the proposed visualization design and evaluated the visual representation with a simple use case. Summarizing, we showcased how Visual Security Analytics could help to solve an existing problem on the domain of LDF.

For future work we mainly see two different directions to follow. First of all, we want to apply our visual security analytic approach to a more sophisticated malware using *Process Doppelgänging*[3] where current anti-virus software and forensic tools reach their limits. Process Doppelgänging refers to a file-less code injection that uses a Windows native function and an undocumented implementation of the Windows Process Loader. This technique leaves no traces and is very difficult to detect. Our approach can highlight malicious activities and assist the digital forensics examiner during a live forensics investigation. Furthermore, another path to pursue in future work is the generalization of our approach. This requires to identify inherent characteristics of FMs and their classification based on a subset of those characteristics. A more holistic and modular version of our design approach would allow to have a specific encoding for each malware characteristic. This would support the work of forensic investigators even further as they can define individual dashboards as subsets of the available designs fitting their need to identify known and unknown FM.

References

1. Adelstein, F.: Live forensics: diagnosing your system without killing it first. Communications of the ACM **49**(2), 63–66 (2006)
2. Antwi-Boasiako, A., Venter, H.S.: Implementing the harmonized model for digital evidence admissibility assessment. In: Advances in Digital Forensics XV - 15th IFIP WG 11.9 International Conference. IFIP Advances in Information and Communication Technology, vol. 569, pp. 19–36. Springer (2019)

[3] `https://www.blackhat.com/docs/eu-17/materials/eu-17-Liberman-Lost-In-Transaction-Process-Doppelganging.pdf`

3. Arendt, D., Best, D., Burtner, R., Lyn Paul, C.: Cyberpetri at cdx 2016: Real-time network situation awareness. In: 2016 IEEE Symposium on Visualization for Cyber Security (VizSec). pp. 1–4. IEEE (2016)
4. Boschetti, A., Salgarelli, L., Muelder, C., Ma, K.L.: Tvi: A visual querying system for network monitoring and anomaly detection. In: Proceedings of the 8th International Symposium on Visualization for Cyber Security - VizSec '11. pp. 1–10. ACM Press, New York, New York, USA (2011)
5. Brehmer, M., Munzner, T.: A multi-level typology of abstract visualization tasks. IEEE Transactions on Visualization and Computer Graphics **19**(12), 2376–2385 (2013)
6. Cappers, B.C., Meessen, P.N., Etalle, S., van Wijk, J.J.: Eventpad: Rapid malware analysis and reverse engineering using visual analytics. In: 2018 IEEE Symposium on Visualization for Cyber Security (VizSec). pp. 1–8. IEEE (2018)
7. Catanese, S.A., Fiumara, G.: A visual tool for forensic analysis of mobile phone traffic. In: Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence - MiFor '10. p. 71. ACM Press, New York, New York, USA (2010)
8. Englbrecht, L., Langner, G., Pernul, G., Quirchmayr, G.: Enhancing credibility of digital evidence through provenance-based incident response handling. In: Proceedings of the 14th International Conference on Availability, Reliability and Security, ARES 2019. pp. 26:1–26:6. ACM (2019)
9. Grajeda, C., Sanchez, L., Baggili, I., Clark, D., Breitinger, F.: Experience constructing the artifact genome project (agp): Managing the domain's knowledge one artifact at a time. Digital Investigation **26**, S47 – S58 (2018)
10. Harichandran, V.S., Walnycky, D., Baggili, I., Breitinger, F.: Cufa: A more formal definition for digital forensic artifacts. Digital Investigation **18**, S125 – S137 (2016)
11. Hoelz, B.W.P., Ralha, C.G., Mesquita, F.: Case-based reasoning in live forensics. In: Advances in Digital Forensics VII - 7th IFIP WG 11.9 International Conference on Digital Forensics. IFIP Advances in Information and Communication Technology, vol. 361, pp. 77–88. Springer (2011)
12. Kahvedzic, D., Kechadi, M.T.: Dialog: A framework for modeling, analysis and reuse of digital forensic knowledge. Digital Investigation **6**, 23–33 (2009)
13. Kent, K., Chevalier, S., Grance, T., Dang, H.: Guide to integrating forensic techniques into incident response. NIST Special Publication **10**(14), 800–886 (2006)
14. Leschke, T.R., Nicholas, C.: Change-link 2.0: A digital forensic tool for visualizing changes to shadow volume data. In: Proceedings of the Tenth Workshop on Visualization for Cyber Security - VizSec '13. pp. 17–24. ACM Press, New York, New York, USA (2013)
15. Leschke, T.R., Sherman, A.T.: Change-link: A digital forensic tool for visualizing changes to directory trees. In: Proceedings of the Ninth International Symposium on Visualization for Cyber Security - VizSec '12. pp. 48–55. ACM Press, New York, New York, USA (2012)
16. Mansfield-Devine, S.: Fileless attacks: compromising targets without malware. Network Security **2017**(4), 7–11 (2017)
17. Marty, R.: Applied Security Visualization. Safari Tech Books Online, Addison-Wesley (2009)
18. McCurdy, N., Dykes, J., Meyer, M.: Action design research and visualization design. In: Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16. pp. 10–18. ACM Press, New York, New York, USA (2016)

19. Meyer, M., Sedlmair, M., Quinan, P.S., Munzner, T.: The nested blocks and guidelines model. Information Visualization **14**(3), 234–249 (2015)
20. Mistry, N.R., Dahiya, M.S.: Signature based volatile memory forensics: a detection based approach for analyzing sophisticated cyber attacks. International Journal of Information Technology **11**(3), 583–589 (2019)
21. Munzner, T.: A nested model for visualization design and validation. IEEE Transactions on Visualization and Computer Graphics **15**(6), 921–928 (2009)
22. Nguyen, V.T., Namin, A.S., Dang, T.: Malviz: an interactive visualization tool for tracing malware. In: Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis - ISSTA 2018. pp. 376–379. ACM Press, New York, New York, USA (2018)
23. Olsson, J., Boldt, M.: Computer forensic timeline visualization tool. Digital Investigation **6**, 78–87 (2009)
24. O'Murchu, L., Gutierrez, F.P.: The evolution of the fileless click-fraud malware poweliks (2015), `https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/evolution-of-fileless-click-fraud-15-en.pdf`, last accessed: 24.02.2020
25. Puchta, A., Böhm, F., Pernul, G.: Contributing to current challenges in identity and access management with visual analytics. In: Data and Applications Security and Privacy XXXIII, Lecture Notes in Computer Science, vol. 11559, pp. 221–239. Springer International Publishing, Cham (2019)
26. Quist, D.A., Liebrock, L.M.: Visualizing compiled executables for malware analysis. In: 2009 6th International Workshop on Visualization for Cyber Security. pp. 27–32. IEEE (2009)
27. Read, H., Xynos, K., Blyth, A.: Presenting devise: data exchange for visualizing security events. IEEE Computer Graphics and Applications **29**(3), 6–11 (2009)
28. Rind, A., Aigner, W., Wagner, M., Miksch, S., Lammarsch, T.: Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. Information Visualization **15**(4), 288–300 (2016)
29. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.A.: Knowledge generation model for visual analytics. IEEE Transactions on Visualization and Computer Graphics **20**(12), 1604–1613 (2014)
30. Simon, S., Mittelstädt, S., Keim, D.A., Sedlmair, M.: Bridging the gap of domain and visualization experts with a liaison. In: Eurographics Conference on Visualization (EuroVis) - Short Papers. The Eurographics Association (2015)
31. Sudhakar, Kumar, S.: An emerging threat fileless malware: a survey and research challenges. Cybersecurity **3**(1), 1–12 (2020)
32. van Wijk, J.J.: Bridging the gaps. IEEE Computer Graphics and Applications **26**(6), 6–9 (2006)
33. Wueest, C., Anand, H.: Internet security threat report: Living off the land and fileless attack techniques (2017), `https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/istr-living-off-the-land-and-fileless-attack-techniques-en.pdf`, last accessed: 24.02.2020