Digital Forensics as a Service - A Game Changer

By

Ruud van Baar, Harm van Beek and Erwin van Eijk

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2014 EU

Amsterdam, NL (May 7th - 9th)

Digital Investigation 11 (2014) S54–S62

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

Digital Forensics as a Service: A game changer

R.B. van Baar*, H.M.A. van Beek, E.J. van Eijk

Netherlands Forensics Institute, Laan van Ypenburg 6, 2497 GB The Hague, The Netherlands

Keywords:
Digital forensics
DFaaS
Digital forensic process
Process model
Xiraf

A B S T R A C T

How is it that digital investigators are always busy and still never have enough time to actually dig deep into digital evidence? In this paper we will explore the current implementation of the digital forensic process and analyze factors that impact the efficiency of this process. Next we explain how in the Netherlands a Digital Forensics as a Service implementation reduced case backlogs and freed up digital investigators to help detectives better understand the digital material.

Introduction

It is impossible to imagine life today without digital material. Who does not use a computer, smartphone, tablet or other digital device nowadays? As a result of the explosive growth in the number of devices and their use, the traces produced by the use of these devices have become more and more important in combating crime. This growth requires a new understanding of forensic data analysis: of the manner in which the data on these devices is processed and of the manner in which the traces collected by this processing is analyzed.

Since December 2010, in the Netherlands a new approach is used for processing and investigating the high volume of seized digital material, viz. Digital Forensics as a Service (DFaaS). Now, three years later, this approach has become a standard for hundreds of criminal cases and over a thousand detectives. This paper describes our approach and the impact on both the digital and tactical investigative process.

This paper starts with describing related work in the next section. In Section Traditional digital investigation process we describe the traditional digital investigation process, that we analyze in Section Analysis of the traditional process. The service model helps to solve a number of bottlenecks. The DFaaS model is described in Section Digital Forensics as a Service and analyzed in Section Analysis of the Digital Forensics as a Service Process. Despite the big changes this model causes, there is still room for improvement. In Section Experience and future work these improvements are discussed. Section Conclusions will complete this paper with final conclusions.

Related work

In this paper we apply a digital forensic process model to the previous and current digital forensic process in the Netherlands. In the related work we discuss process models, techniques that can help optimize the current process and expected developments that have an impact on the forensic process.

Process model

Even though the digital forensic process model is not standardized, consensus on the abstract level about the digital forensics process exists. The latest effort by Kohn et al. (2013) to propose a model contains an overview of the most significant models described over the years. On a high level, Kohn described six processes: *documentation*, *preparation*, *incident*, *incident response*, *digital forensic*

* Corresponding author.
E-mail addresses: ruud@holmes.nl (R.B. van Baar), harm@holmes.nl (H.M.A. van Beek), eijk@holmes.nl (E.J. van Eijk).

http://dx.doi.org/10.1016/j.diin.2014.03.007

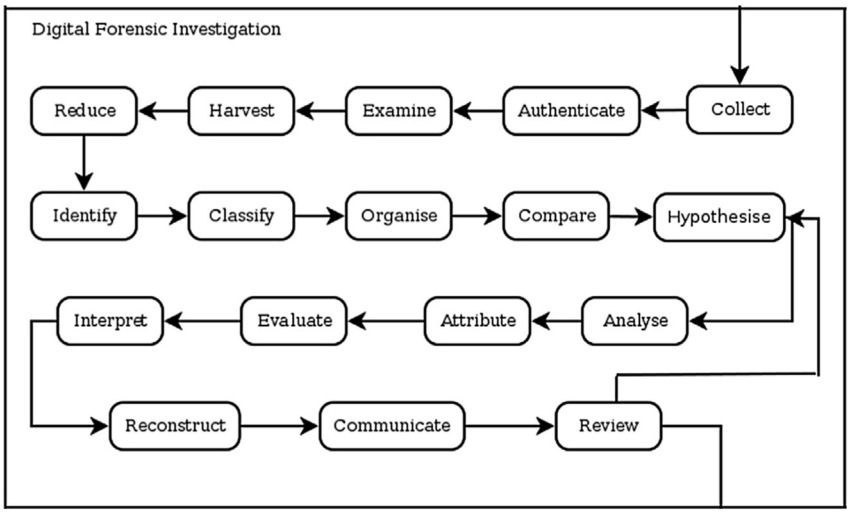

Fig. 1. IDFPM: Digital forensic investigation (Kohn et al., 2013).

investigation and *presentation*. In 2003, Carrier and Spafford already described the digital investigation process, where they defined five groups: *readiness*, *deployment*, *physical crime scene investigation*, *digital crime scene investigation* and *review*. Casey (2011) summarized the different steps as *preparation*, *survey/identification*, *preservation* and *examination/analysis*.

In addition to forensic investigations, eDiscovery exists. Their processes are very similar. Chisholm (2010) explains that the primary difference between the two is the scope of work. For eDiscovery, the Electronic Discovery Reference Model (EDRM)[1] is leading.

This paper focuses on the examination of the digital traces, defined by different authors as the *digital forensic investigation process* (Kohn), the *digital crime scene investigation* (Carrier) or *examination/analysis* (Casey). We explain how the digital forensics process is implemented in the Netherlands. We do this by using the integrated digital forensic process model (IDFPM) and terminology as described by Kohn et al. (2013), but other models can be applied just as easily. The *digital forensic investigation* step is presented in Fig. 1. Other parts of the IDFPM are described where applicable.

Technical implementations

Multiple next generation forensic analysis systems are under development or already implemented. These systems are generally built to automate and speed-up the indexing of the images, which is a good starting point to set up Digital Forensics as a Service (DFaaS).

In 2004, Roussev and Richard described a distributed processing system many times faster than AccessData FTK.[2] This was a lab setup. Since then FTK 3 and higher support a total of four so-called workers to automatically process data in parallel. Research on the automated processing of seized material was coined in 2006 by Alink et al. (2006). Ayers (2009) put down the need for such a system and described the requirements that such a system must, should or may meet. In 2012, Bhoedjang et al. explained how the Xiraf system is engineered and in use in the Netherlands. One of the efforts to build a DFaaS system is proposed by Lee and Un (2012). They focused on speed and provided the end-user with a web interface to search through the data.

A lot of tools exist that support eDiscovery in both law enforcement and businesses, like *ZyLAB eDiscovery OnDemand*[3] and *Symantec eDiscovery Platform*,[4] powered by Clearwell.

Expected developments

Some interest has already gone out to speculate on what will happen in the near future. Much of this speculation stays within the current boundaries and describes improvements in tooling and standardization. Richard and Roussev (2006) and Garfinkel (2010) described several developments they expected to happen. Some of these developments, like distributed processing, are already in production, as described before. Garfinkel expected a crisis in digital forensics if no efficient method is found to analyze all data. Some reasons for this are the increase in data, encryption and proliferation of operating systems and file formats.

Traditional digital investigation process

We discuss how the digital forensics process as described in Section Process model is implemented in the Netherlands. We do this by using the model and

[1] http://www.edrm.net/, visited Feb, 2014.
[2] http://www.accessdata.com/products/digital-forensics/ftk, visited Feb, 2014.
[3] http://www.zylab.com/services/ediscovery-ondemand- saas.aspx, visited Feb, 2014.
[4] http://www.symantec.com/ediscovery-platform/, visited Feb, 2014.

terminology as described by Kohn et al. (2013), because it is the most recent and contains an overview of the other relevant models.

Fig. 2 shows the traditional digital investigation process as practiced in the Netherlands. On the left, the digital devices are shown, on the right the detectives and analysts that are interested in information stored on these devices.

Depending on the type of case, a digital investigator may be involved sooner or later. Or perhaps not at all. A case can start in a number of ways: a victim reports a crime, law enforcement officers are called in to a crime scene, an intrusion detection system reports an intrusion or any of many other ways. The IDFPM describes this as the *incident* process, with a subsequent *incident response* process. Both traditional crimes like arson, child abuse and murder, and cybercrimes like hacking, phishing or denial of service attacks generally have a digital component.

Detectives generally have limited or no training in how to handle digital devices. They involve a digital investigator (see Fig. 2) to answer a specific question related to the digital devices. This digital investigator has no detailed knowledge about the case since he was not involved from the start, did not read any statements and did not attend any briefings. He only focuses on the question at hand and starts with a series of standard tasks. The first task is generally the same: create forensic copies of the digital devices (*collection* and *authentication*). After that, numerous tools or scripts may be run to recover deleted files, carve unallocated space, unpack archives, etcetera. These standard tasks may be strung together using scripts (Carrier, 2003–2013). In this process (*examination*), digital evidence is made visible and allows for a detective to look at the information.

A possible next step is to index the data, *harvest*, to structure the data in a logical format. Indexing can mean a number of things, ranging from creating a keyword index, store recognized timestamps in a database, extract all metadata to be queried later or a combination of these methods. To *reduce* the amount of data to be analyzed, it is possible to remove known files, for example by using hash sets. Hash sets can also be used to *identify* an incident, but identification may not be necessary or even possible in a lot of the digital evidence. Identification of the incident may not occur until the very end of the investigation, since it is not always clear if a digital device even contains relevant trace information.

The IDFPM relies on the identification part of the digital investigation to determine if and what incident occurred. It may not immediately be clear if a digital device is relevant. Without identification, it is difficult to perform *classification*, *organization* and *comparison*. A more common practice in criminal cases is that a detective formulates questions based on non-digital case information. These questions are answered by digital investigators. Generally, an (implicit) *hypothesis* forms the basis of these questions. If a question is asked to retrieve all emails sent on a certain date, the underlying hypothesis may be that somebody contacted someone else and email was the preferred means of communication. A digital investigator can help to formulate these hypotheses and aid in coming up with more ways to test a hypothesis.

To answer the question or test the hypothesis, the digital evidence is *analyzed*. This may be done in a number of ways: a keyword search can be performed using an indexed or live search, the detective may manually search for Internet history or use a script to search for all chat messages, a timeline may be created using a forensic tool or spreadsheet application or other types of investigations may be performed.

In the current process in the Netherlands, the steps *attribution*, *evaluation*, *interpretation* and *reconstruction* are optional in a digital investigation. If the original question was “Give me all email communication”, there is not much interpretation or reconstruction needed. If these steps are to be taken, results are generally *communicated* with relevant parties for discussion before continuing with these steps. Communication may be done in a number of ways. Common methods are printing out relevant information, putting results on a portable media or central storage or upload the results to another system that can correlate information. Depending on the type of question, this may lead to a lot of results. The detective has to sieve through these results before either finding relevant information or asking a follow-up question to explain results or reduce the number of results. If relevant information is found, any previously skipped steps may be taken to reconstruct the

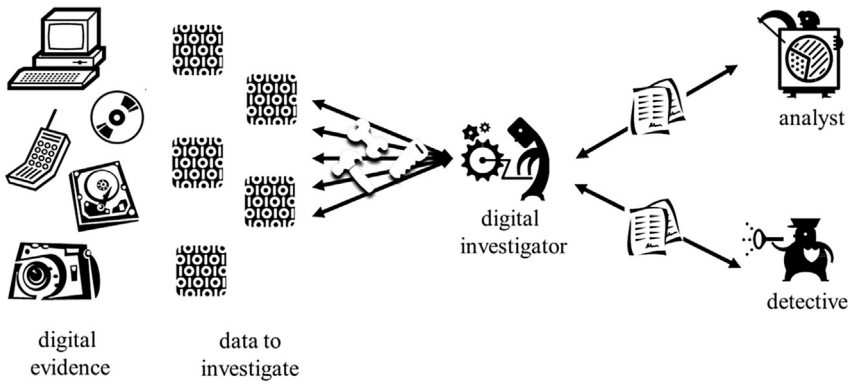

Fig. 2. Traditional digital forensics process.

events that lead to the resulting traces. This means that the cycle as described in the IDFPM is more of a tangle than a straightforward line with an optional cycle.

Eventually, an official report needs to be made that can for example be *presented* in court or to management. Sometimes even at this stage new questions arise, making it necessary to redo parts of the investigation or perform additional research. This can be done by either the original detectives, other detectives, digital investigators or even counter experts.

Analysis of the traditional process

In general, the process described in the previous section varies depending on a number of factors. During our years of experience running a DFaaS with hundreds of cases and over a thousand users, we identified a number of factors with a high impact on efficiency. We discuss resource management, types of questions, required time frame, collaboration and research, development and sharing knowledge.

Resource management

Digital investigators are generally responsible for their investigative environment (*infrastructure readiness*). This goes for storage, backups, network, software, security and many other tasks that come with system administration. Although it is not their primary task, they may feel (and possibly be held) responsible if something goes wrong. In general, they are not equipped to administer the investigative environment. This may lead to security breaches, failing backup systems, downtime due to incorrect installations, use of obsolete software and many other things associated with system administration. This obviously leads to more time performing system administrative tasks and less time available to perform digital investigations.

Digital investigators are also expected to be a Jack of all trades. Besides system administration, they have a lot of other tasks. They are needed at crime scenes with non-trivial digital devices, perform incident response jobs, create disk images, phone images, memory dumps, network captures, perform analysis, research unknown file formats, and so on. The saying "Jack of all trades, master of none" is applicable here. While there are definitely digital investigators that succeed at combining all the different roles that are expected of them, they are generally burdened the most and cannot be everywhere at the same time. This would not be a bad thing if they succeeded to at least make the most efficient use of their skills. But is it really efficient to have someone create disk images while he could also reverse engineer file formats?

Questions

In general, three types of questions related to digital material can be defined. The first type is not really a question, but more of a request. Examples are: "Give me all communication information", "Search for the name Pete" and "I want all information related to drugs". These requests generally result in a lot of work on the part of the digital investigator, mostly because the requests are ambiguous. What is defined as communication? Chat, phone calls, text messages come to mind, but what about an email message saved in the drafts folder of Gmail (Leger and Alcindor, 2012)? Does information related to drugs include contact information, bank account statements and pictures of expensive cars?

The next type of question is somewhere in the middle. It is still vague, but more specific than the first type. Examples are: "What did the suspect search for?", "What are the origins of this document?" and "Was the clock running on time?". These seem to be questions that may give a more narrow answer, but still leaves a lot of work to be done for the digital investigator. How are searches defined? Is this online, on the local machine, in the email client? And how does a digital investigator link these results to a specific person? There are methods to check if a clock was changed and if it synchronized, but usually these are vague indications for a longer period of time.

The last type of question is very specific. The detective has a hypothesis he wants tested and asks questions like "With whom did the user chat on March 20th?", "Were these pictures taken with this digital camera?" or "Did a user search online for any of the following medicines?". These questions are very specific and give a good direction to the digital investigator on what to look for. The problem is that this pulls the digital investigator into a tunnel. If a relevant chat message was sent out by the user on March 19th close to midnight or a picture was taken with a cell phone that also belongs to the suspect, the digital investigator may not trigger on these results. As said before, he generally has no more knowledge about a case other than what is part of the question. As a result, he answers the question, thereby potentially missing crucial information.

Time frame

Another issue with the traditional model is the time frame. In the first couple of days of an investigation, hypotheses are formed and leads based on those hypotheses are followed-up (U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, 1998; Joyce, 2012). Since the results of a digital investigation usually are not available in the first couple of days, they are not taken into account. The growth in number of digital investigators is very limited compared to the growth in number of digital devices, size of digital data carriers and number of cases with a digital component.

If a digital investigator starts to work on a case, the first couple of steps are generally the same. Create a forensic copy, recover deleted files, retrieve Internet history, etc. This is one of the reasons why digital investigators are almost always busy. Since the next case is already screaming for attention, there is practically no time left for innovation or sharing of knowledge. A digital investigator is also always busy due to other obligations, which leads to cases where digital technology is a very important component, but the case is put on the shelf or not handled at all.

Collaboration

Current procedures make it hard to collaborate. If a digital investigator reports an entire result set, this set is scattered across multiple detectives. This can be done in a number of ways, e.g. by number of pages, by person of interest or by task. The detectives can then individually work through their part. If the set is split so that one detective looks at contact information, another looks at communication and a third one looks at the Internet history, there may be a lot of overlap. If a suspect communicates through Facebook with one of his contacts there is an overlap, so who is responsible for looking at these traces? Are they all three responsible? Or is no-one responsible and is the trace ignored or overlooked? Another way of splitting up the work is that each investigator looks at a set of devices, but there is probably some overlap between devices that is overlooked. Just like with a printed dossier, splitting up the evidence based on the size or complexity of the material makes it harder to keep a bird's-eye view.

Research, development and sharing knowledge

Digital investigators do case driven research, for example finding out the forensic meaning of certain timestamps. This information is reported to the investigation team, which uses the information in the case. Of course, the research results themselves can be applied to other cases. In general, he can only apply this knowledge if the digital investigator is associated with a case where the same question arises. If a similar case is investigated without the digital investigator knowing about it, either because he is working another case or because the case is run in a different part of the country, his knowledge is not reused: the research is redone or the traces are ignored. Every now and then, small tools are built or research results are blogged to distribute the knowledge to other digital investigators.

All digital investigators have to keep track of new developments in their field. As said before, the digital investigators are generally overburdened and do not always have time to keep their knowledge up to date.

Digital Forensics as a Service

Since December 2010, we use a new service-based approach for processing and investigating the high volume of seized digital material: Digital Forensics as a Service (DFaaS). This service is based on Xiraf, a closed-source, non-commercial product, developed at the Netherlands Forensics Institute,[5] funded by the Dutch Government.

Fig. 3 shows the procedure how forensic cases are handled using this approach. On the right, there are still detectives and analysts that have questions related to the digital material shown on the left. To guarantee forensic integrity, images are still needed. So as in the traditional process, the first task is to create forensic copies of the digital devices (*collection* and *authentication*). The big difference is that images are copied to a central storage, processed (*examined*) using a standard set of tools, ranging from tools that extract file systems, files and carve unallocated space, to tools that parse chat logs, Internet history and mail databases. The results of these tools are stored (*harvested*) in a centralized database. After storing these traces, they can be queried (*reduced* and *analyzed*) using multiple methods: detectives can log on using a web browser, digital investigators can use the programming interface to run automated tools and analysts may want to retrieve all information and analyze the results using data visualization tools, integrate data sources or build a network of contacts. This makes it possible to *identify*, *classify*, *organize* and *compare* the traces within seconds, based on *hypotheses* and questions the investigators have. This can be done at any time during the investigation.

It is generally difficult for detectives without a digital background to *attribute*, *evaluate*, *interpret* and *reconstruct* digital traces. For these tasks they need to understand how and why certain traces exist and what events can lead to a given collection of digital traces. If a relevant trace is found, it is generally a good idea for a detective to find assistance with a digital investigator that can help the detective perform these tasks. To do this, good methods of *communication* are needed.

Analysis of the Digital Forensics as a Service Process

Some factors that have a high impact on efficiency on the digital forensics process are discussed in Section Analysis of the traditional process. These same factors are discussed as they are observed in the DFaaS process.

Resource management

Administration of the DFaaS is performed by a team of service operators. These operators are system administrators without a digital forensic background. They can upload images to central storage, index the images, give users access to cases and perform other tasks related to system administration. During indexing, all metadata including timestamps is extracted from digital devices and a keyword index is created.

Apart from these operators there are also application administrators, database administrators, storage administrators, infrastructure administrators and other system administrators that are not specific to the field of digital forensics and are generally needed for other tasks as well. They have very specific knowledge about a part of system administration and can use this knowledge to optimize systems, prevent data loss, make sure data is protected and generally make sure the service is available.

By centralizing the software, we make sure that any spare capacity, both storage and processing power, can be used by different users. Traditionally, if an investigation in one part of the country required a lot of storage or processing power, these resources were purchased. If another department had this capacity available and was not actively using it, there is no mechanism to share these resources.

[5] The authors are employed by the Netherlands Forensics Institute and work on Xiraf and its successor Hansken.

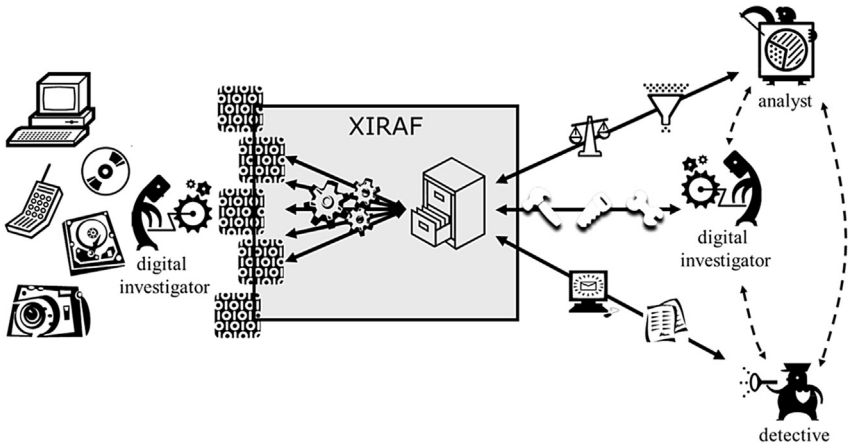

Fig. 3. Digital Forensics as a Service.

With a service model, there is a centralized capacity that can be shared among all investigations.

By centralizing the data, a backup mechanism has to be implemented only once and security can be arranged centrally. Moreover, if more than one department had to perform investigation on the same data, multiple copies of the data had to be made and distributed and the copies had to be analyzed multiple times. With a service model, the only effort required is to grant a detective access to a case, which can be done in a matter of minutes rather than days or weeks.

Questions

By enabling detectives to directly query the digital material, they can use their detailed knowledge of a case or field of expertise. They can come up with more hypotheses. They can also identify if there are any leads that may support these or other hypotheses. Because they have specific knowledge of a case, like the interests of a suspect or lingo used in a certain environment, they may trigger more quickly on results than digital investigators. A detective that wants to know if a person unjustly received social benefits may look at pictures to see if his suspect went abroad on vacation. A detective performing a drug investigation knows dozens of words that can be used to describe drugs. A detective looking at an alibi of somebody claiming he was sitting behind his computer at a certain time may take a couple of hours extra into account to see if a suspect could have been at a given place at a given time.

Previously, if detectives asked questions regarding digital material, they could be either too broad or too narrow. Now, they are confronted with the results immediately and adjust their searches accordingly. A keyword search resulting in hundreds of thousands of results needs additional filtering before it is useful. If the investigator looks at a specific form of communication on a given date about a specific subject, the detective may tune any of the filters to find additional information.

Time frame

As described in Section Technical implementations, with a service model, it makes sense to use a distributed implementation for harvesting, reducing and analyzing traces from seized material. For a single department with a couple dozen of detectives it may not be feasible to implement a multi-rack system that is idling most of the time. At a large scale it makes sense to implement a central system that can by used by multiple departments. With this model, it is expected that the system spends less time idling and more time actually processing data. If digital material becomes available sooner in the investigation, it can be used to form hypotheses instead of only using it to test hypotheses.

Providing digital forensics as a service frees up time from digital investigators. On the one hand, it reduces the administrative tasks that are not part of their actual job and on the other hand the investigative tasks are executed by detectives, who have much more case knowledge. This gives the digital investigator the possibility to do more in-depth (possibly case related) research. Eventually this leads to a positive spiral upwards that gives a more solid base to investigations. If this new knowledge is embedded in the service, the system doubles as a knowledge center. Any new research results stored this way helps other investigation teams that struggle with the same problems. The digital material becomes an essential part of any investigation.

Using the service model allows detectives to work from their own work station at their own desk. This reduces overhead time and lowers the threshold to fill in empty hours. If they run into leads during an investigation, the digital material can be looked at instantaneously. This immediately speeds up the investigation.

Collaboration

Any forensic analysis system should make it possible to share progress on a case. This means that when a detective

finds an interesting trace, it should be possible to annotate the trace so other detectives see the annotation. As shown in Fig. 3 there are links between detectives, digital investigators and analysts. If a detective runs into a trace that he does not understand, it is easy to ask a digital investigator for help. The digital investigator can help by explaining the meaning of traces, conduct additional research or assist with writing a statement.

Splitting the data is much easier with a service model that allows collaboration. Investigative topics are generally already distributed amongst detectives and if it is possible to generically search through the data of all devices it is not needed to split the data according to size or type of device. It should not matter if an email was read on a mobile phone, tablet, desktop computer running Windows or a laptop running OS X. At first, a detective wants to know if an email was sent, later on he wants to know how the email was sent. These types of queries are supported by Xiraf (Bhoedjang et al., 2012).

Research, development and sharing knowledge

We expect that digital investigators specialize on one or more forensic tasks. Currently, three tasks are identified. The first task is creating forensic images. In order to guarantee the integrity of the images, this is a procedure that needs to be performed. The second task is to interpret results found by detectives, to explain the meaning of traces that they found and how they can use the traces in the actual case. The third task is digging down into the bits and bytes of the images in order to find out specific information relevant to the case. This final task is often driven by traces found by detectives that require in-depth digital investigation.

With DFaaS we have seen a lot of digital forensic departments with backlogs getting back in line with the investigations. Digital investigations are performed more efficiently and digital investigators are freed to perform more in-depth research. As described, research performed — whether it is case driven, technology driven, trend driven or curiosity driven — can be incorporated in the DFaaS. The service now serves as a knowledge center. With over a million apps in just the Apple App Store,[6] it is crucial to share knowledge about how applications can be forensically examined (Garfinkel, 2010). It must be easy to add tools that parse forensic traces left behind by an application to a centralized environment. New tools become available to all running investigations, leveraging the results of the research to more than just the single case, potentially to a significant amount of cases. Even if the results are used in only one other case, just the fact that the application is analyzed may lead to results that would otherwise not have been obtained.

Experience and future work

Even though the service model fits nicely on digital forensic investigations and the Xiraf implementation already has proven itself in numerous cases, there is still a number of possible improvements that enhance the experience and reduce the lead time for results to be available. In this section a number of improvements are discussed.

Redundant storage

As mentioned in Section Resource management, in traditional investigations, if a case needs to be handled by multiple departments, multiple copies of the data need to be made and shared. A national investigation where all departments need to have access to parts of the digital material results in numerous copies. Not only does this require a lot of storage, but the logistics in handling this type of data transport is cumbersome and error prone. In the current Xiraf procedures, images created of digital devices are copied to a central storage that can be made available to all departments. Although this optimizes the amount of storage required and reduces the number of steps to distribute data, it is still possible to optimize this step by uploading the images directly to the storage from the digital devices. This can be done using so-called upload stations.

Another optimization is that harvesting traces from the forensic image can start while the device is being imaged. Even better, the analysis process can influence the imaging process by asking for certain blocks of data to become available with priority. If a hard drive containing an NTFS file system is imaged, the master file table (MFT) is the first thing to analyze. The MFT is generally not stored at the beginning of the disk. Serving the blocks containing the MFT first, makes it possible to start analyzing it. This may lengthen the time it takes to image a disk due to random access, but the total processing time including the analysis step might shorten.

The result of this optimization is that the *authenticate* task as shown in Fig. 1 is temporarily set on hold. This means that any result obtained is not yet authenticated until this task is complete.

Indexing performance

As described in Section Related work, a lot of thought has already gone into parallelizing the indexing process. The only way so far of parallelizing this process with Xiraf is by using multiple indexing machines and separately index images. After the indexing, the results are merged in a single database, the so called publishing. This process is cumbersome and requires the index to be fully available before the database can be queried. Parallelizing the Xiraf backend requires a full redesign to be able to make it scalable and query the results while the index is still running.

Additional query capabilities

Xiraf only allows for indexed queries, which reduces flexibility. Sometimes a user wants to ask a question that is not possible by just querying the metadata database. More complex queries, like matching full keywords, regular expressions or running additional tools, are difficult to

[6] http://148apps.biz/app-store-metrics/, visited Feb, 2014.

implement in Xiraf and far from efficient. In the current implementation, if a new tool is added to the system and new results are available, these results have to be published to the database before they can be queried. Adding new tools is a laborous task and these new tools are not run automatically for every running case. To apply these tools on all current cases is a full time job for operators. This should be optimized so that new tools are automatically run on all cases and users can create their own data intensive queries, e.g. regular expression searches or image processing tools.

Disadvantages of Software as a Service

The same disadvantages for the Software as a Service model apply to the DFaaS model. Some disadvantages are latency, dependency on a working Internet connection and using the stored data in other applications. Any DFaaS implementation has to make sure these problems are either acknowledged or tackled.

Security and privacy by design

Xiraf started as a master thesis' project and grew to a funded research project (Alink et al., 2006). It was never meant to be used in real life cases at its current scale. Any measures taken to implement security and privacy are put on top of the current implementation. Newer Xiraf releases incorporate user management, user roles, management tooling and other measures that are expected in a service solution. However, if security and privacy are part of the design process, any code, process or infrastructure takes these measures into account. In technical design, documentation, code review and infrastructure design, security and privacy should be a second nature to anybody involved. Adding security and privacy in a later stage leaves unexpected holes in place. Security and privacy measures should make sure that no-one has access to any information they are not authorized to access, including developers, system administrators, malicious hackers and cleaners. Especially since, as described in Section Resource management, system administration is no longer done by digital investigators but by system administrators. This is security and privacy by design.

Conclusions

In this paper we analyzed the traditional digital forensics investigation process and the digital forensics as a service (DFaaS) model. We compared the implementations using the integrated digital forensic process model (IDFPM) by Kohn et al. (2013).

Digital investigators should not be tasked with system administrative tasks that are better performed by dedicated system administrators. In the traditional process, digital investigators are held responsible for their entire investigation environment (storage, network, software, security, etc.). Combined with their central role in securing and analyzing digital evidence this leads to a lot of administrative overhead and possible security breaches, failing backup systems and the use of obsolete software, amongst others. Digital investigators are either underqualified or overqualified for a lot of the tasks they perform on a daily basis. In the DFaaS setup in the Netherlands, digital investigators focus on the forensic tasks, i.e. seizing material and extracting data from it. The data is sent to a centralized system that automatically extracts traces from the data and gives digital investigators, detectives and analysts access to the traces. Several administrators execute domain-specific tasks related to this service, like application administrators, database administrators, storage administrators and infrastructure administrators.

It should be possible to use any capacity currently available within the same organization or even across organizations to process cases. A lot of departments have spare storage and processing capacity while other departments may temporarily require this capacity for a large case that is handled. Traditionally, there is no method to share this capacity. By centralizing this capacity, departments can use all processing power available to speed up the processing of digital material. Cross department investigations have access to one copy of the pre-processed data, granting access to this data to detectives in another department can be done in minutes. A disadvantage of the centralization is that the data has to be made available to this centralized system. In the Netherlands, we currently create an extra copy of the data, making it possible to locally analyze the material as well.

Detectives should be the ones looking at the digital material. They have knowledge about a case that digital investigators do not have. In the traditional process, detectives ask digital investigator to provide information. The lack of critical case-related information makes it hard if not impossible to determine if traces relate to a specific crime or not. As a result, in the traditional process, the digital investigator is more a harvester of traces than an analyst. The turnaround time for this can be days or even weeks. The digital investigator provides the harvested traces (which may or may not relate to the requested information) to detectives who subsequently reduce these traces to a small set of relevant ones. In the service model, all traces extracted by the service are delivered to the detectives. They can query the traces, filtering out irrelevant ones within seconds. If relevant traces are found, they have direct access to the source material, like pictures, document and emails. While it is true they run into problems understanding certain concepts, like carved files or the interpretation of a timestamp, they can be educated to learn this. As described by Bhoedjang et al., 2012, the benefits of nonexperts investigating digital material outweigh the risks. Digital investigators can help detectives by explaining the query results and do more in-depth investigation on technical details of relevant traces.

Digital material should be made available in the first couple of days in the investigation so it can be used to form hypotheses. In the traditional process, results from digital investigations are not available in these first couple of days. As a result, traces found in digital material are often used for validating instead of forming hypotheses. Providing digital forensics as a service speeds up the trace harvesting process, making traces available to

detectives sooner in the investigative process at their own desk. This reduces overhead time and lowers the threshold to fill in empty hours. If they run into leads during an investigation, the digital material can be looked at instantaneously.

Collaboration between detectives and digital investigators is crucial in understanding digital evidence. Traditional procedures make it hard to collaborate. Digital investigators report large result sets that are scattered across multiple detectives. Just like with a printed dossier, splitting up the evidence makes it harder to keep a bird's-eye view. In general, instead of the material, investigative topics are distributed among detectives. A service model makes it possible to analyze all data in the context of a topic. A trace that is irrelevant to one topic, might be decisive to another topic in the same case. If a detective found an interesting trace or a trace that he does not understand, the service model makes it possible to annotate the trace. Other detectives or digital investigators have access to the annotation and can act on it.

Freeing up a digital investigator to perform more in-depth research will result in a positive spiral upwards. If more investigative methods are developed, hopefully turnaround time and time spent for digital investigations will be reduced. By embedding newly gathered knowledge in the service, the system is used as a knowledge center. Any new research results stored this way helps other investigative teams that struggle with the same problems.

Since December 2010, we implement digital forensics as a service (DFaaS) in the Netherlands using Xiraf (Alink et al., 2006; Bhoedjang et al., 2012). Now, three years later, this approach has become a standard for hundreds of criminal cases and over a thousand detectives, with great success. At the moment, we are working on the successor of Xiraf, called Hansken, adding the things we learned from three years of servicing Dutch law enforcement agencies.

References

Alink W, Bhoedjang R, Boncz PA, de Vries AP. XIRAF – XML-based indexing and querying for digital forensics. Digit Investig 2006;3:50–8.

Ayers D. A second generation computer forensic analysis system. Digit Investig 2009;6:S34–42. http://dx.doi.org/10.1016/j.diin.2009.06.013.

Bhoedjang RAF, van Ballegooij AR, van Beek HMA, van Schie JC, Dillema FW, van Baar RB, et al. Engineering an online computer forensic service. Digit Investig 2012;9(2):96–108. http://dx.doi.org/10.1016/j.diin.2012.10.001.

Carrier B. Autopsy http://www.sleuthkit.org/autopsy/; 2003–2013.

Carrier B, Spafford EH, et al. Getting physical with the digital investigation process. Int J Digital Evid 2003;2(2):1–20.

Casey E. Digital evidence and computer crime: Forensic Science, Computers, and the Internet. 4th ed. Elsevier Science; 2011.

Chisholm C. Integrating forensic investigation methodology into eDiscovery; Jan 2010. SANS Forensics Whitepapers.

Garfinkel SL. Digital forensics research: the next 10 years. Digit Investig 2010;7(Suppl.):S64–73. http://dx.doi.org/10.1016/j.diin.2010.05.009.

Joyce T. Closing the case: solving violent crimes quickly and efficiently with public records. The Police Chief 2012;79:50–6.

Kohn M, Eloff M, Eloff J. Integrated digital forensic process model. Comput Secur 2013;38(0):103–15. http://dx.doi.org/10.1016/j.cose.2013.05.001. Cybercrime in the Digital Economy.

Lee J, Un S. Digital forensics as a service: a case study of forensic indexed search. In: ICT Convergence (ICTC), 2012 International Conference on; 2012. pp. 499–503. http://dx.doi.org/10.1109/ICTC.2012.6387185.

Leger DL, Alcindor Y. Petraeus and Broadwell used common e-mail trick. USA Today; 2012. URL, http://www.usatoday.com/story/tech/2012/11/13/petraeus-broadwell-email/1702057/.

Richard III GG, Roussev V. Next-generation digital forensics. Commun ACM 2006;49(2):76–80. http://dx.doi.org/10.1145/1113034.1113074.

Roussev V, Richard III GG. Breaking the performance wall: the case for distributed digital forensics. In: Proceedings of the 2004 Digital Forensics Research Workshop, vol. 94; 2004.

U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. When your child is missing: A family survival guide. URL, http://www.ojjdp.gov/pubs/childismissing/; May 1998.