submit date:2022-02-01

DFRWS Challenge 2021

Summary Report

(Team: DFRC)

KOREA UNIVERSITY

Digital Forensic Research Center

Summary Report_DFRC

🕵️Contact Information

Team Name : DFRC

Team Member :

Jung Heum Park (jungheumpark@korea.ac.kr)

Sang Hyuk An (starson1@korea.ac.kr)

Seung Ah Kang (kyn0503121@korea.ac.kr)

Byeongchan Jeong (naaya@korea.ac.kr)

Sueun Jung (whrgk10@korea.ac.kr)

Jaeseong Kwon (kjs000804@korea.ac.kr)

Team Nationality : Republic of Korea 🇰🇷

Contact : starson1@korea.ac.kr (+82-10-4170-2391)

🔧 Tool

Tool Table

Name	Type	Part	Version	URL
Autopsy	Freeware	QNAP NAS Raspberry PI Samsung Smartphone Skimming device	4.19.2	https://sleuthkit.org/autopsy/
Sonic Visualizer	Freeware	Skimming device	4.4	https://www.sonicvisualiser.org/download.html
CC Checker	web-open	Skimming device	NaN	https://www.mobilefish.com/services/credit_card_number_checke
Magstripper	Freeware	Skimming device	0.3a	https://sourceforge.net/projects/magstripper/
AccessData FTK Imager	Freeware	QNAP NAS Raspberry PI Samsung Smartphone Skimming device	4.3.0	https://accessdata.com/product-download/ftk-imager-version-4-3-
Notepad++	Freeware	Raspberry PI	7	https://notepad-plus-plus.org/downloads/v7.0/
Ultimaker Cura	Freeware	Raspberry PI	4.13	https://ultimaker.com/software/ultimaker-cura
DB browser for SQLite	Freeware	QNAP NAS	3.12.2	https://sqlitebrowser.org/
ChromeCacheView	Freeware	QNAP NAS	2.30	https://www.nirsoft.net/utils/chrome_cache_view.html
jadx	Freeware	Samsung Smartphone	1.3.1	https://varaneckas.com/jad/
CyberChef	web-open	Samsung Smartphone	NaN	https://gchq.github.io/CyberChef
Carpe	Developed Freeware	Multisource Analysis	20210620	https://github.com/dfrc-korea/carpe

📖 Solution

Part 0. Answer

Preconditions

[Integrity]

Every Image file hash was checked at the introduction of each part.

Exporting files and reading files were done by **AccessData FTK Imager** in read-only mode.

[Time Zone]

Every time given without a specific timezone is considered as **UTC +01:00, UTC +02:00 (DST)**

Since it was broadly found in the given image files.

Daylight Saving Time (DST) period in Europe runs from 01:00 UTC (Coordinated Universal Time) on the last Sunday of March to 01:00 UTC on the last Sunday of October every year.

[The Challenge Answer]

Table captured below out team’s answers to the challenge questions.

Challenge Questions Response

Question	Answer
1. Skimming Device	- Card number found in recording.mp3. (4334 2250 2436 4939) - Train ticket found from unallocated area (Lausanne → Aosta)
2. Raspberry Pi	- Raspberry Pi has been used to control the 3D printer. - 3D printer printed illicit “Liberator” at “2021-04-13 14:27:57.389504 ~ 2021-04-14 17:29:46.340420”.
3. Samsung Smartphone	- Encoded SMS messages was decoded. - Conversations done with E-Mail.
4. QNAP NAS	- Mattermost server hosted for communication system. - traces of Vault was found. file including CC number was found.
5. Multisource analysis and correlation	- Proposed Process includes 6-steps. - Process will cut off time to get forensics by selection and relation analyzing.

Multisource Analysis

[Introduction]

With the incredible speed of developments in electronic and internet industry, criminal acts became more advanced and meticulous furthermore not only a single digital source is used nowadays. In this situation, investigators must have a unified methodology to analyze and correlate potential digital evidence from multiple sources.

With the DFRWS Challenge 2021, it was truly challenging to examine each source of interest and to correlate them with a single story. There has been practical guidelines for a single source such as mobile, desktop, etc. However, research efforts connecting in between different sources has not been done enough.

So, in this Part 5, a unified guideline is proposed that is specialized with multi-sources. Believing that it will help investigators to investigate in efficiency with increasing amounts of data from multitude of sources.

Briefly, the process being introduced is designed “Top-Down” in 6-steps, with analysis done in each step, gradually helps building logical timeline that explains users’ activities at last.

[Approach]

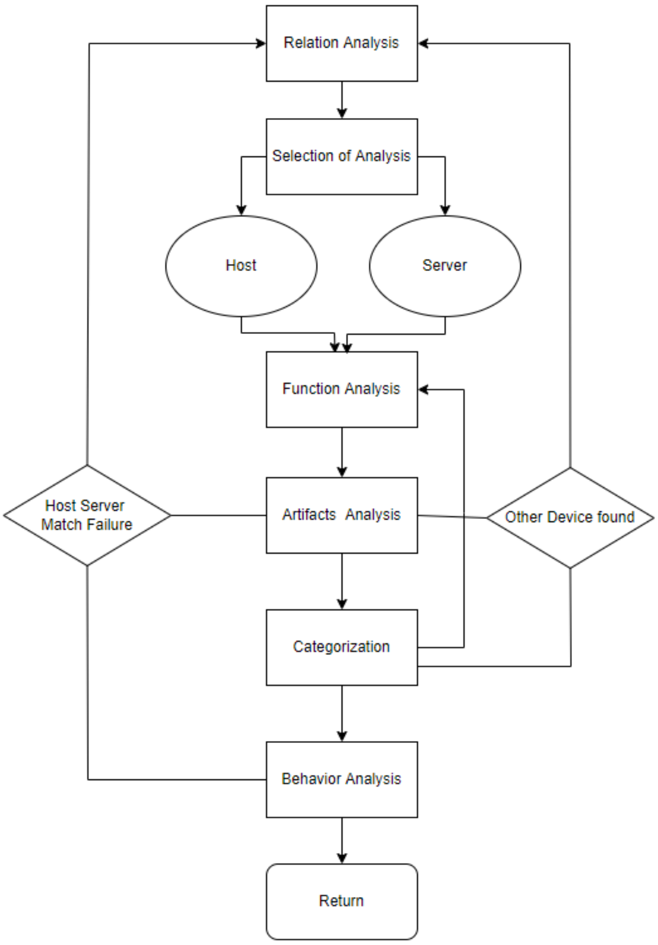
The goal of this part is to advance the state-of-the-art in multisource analysis and correlation. When there is a case with multisource, the most important thing is to save time for investigators and seeing a new view through correlation analysis on multi-source data. Because any kind of analysis method which saves lots of time needs a new approach because of characteristic of multisource analysis. This idea comes from the importance of recognizing the relationship between multiple sources not focusing on the individual one.

By solving the challenge given, the main theme was to design a new process that will cut off time no matter what kind of source device is given. The process presented in here includes 6-steps. Relation analysis, Selection of analysis, Function analysis, Artifact analysis, Categorization, and Behavior analysis. Time saving was achieved in Selection of analysis and Categorization parts, which gives convenience inspecting the entire case. In the next section, detailed description of the proposed approach will be provided.

[Preperation]

Preparing forensic knowledge for various H/W and S/W. For multisource analysis, there needs to understand basic knowledge of individual potential source of evidence, and to prepare forensic tools and techniques for handling it properly. This preparation is necessary not only for multisource analysis but also for normal forensic analysis. Therefore, we explain the proposed process on the premise that basic knowledge for handling target H/W and S/W are already shared between stakeholders. Also, it is necessary to understand about the case in detail. That is to say, digital forensic examiners need to check what case is about, and think about which of events to focus on. For example, in the case of a confidential information leakage, we should focus on file copying, moving traces, or messenger artifact.

[Proposed Process]



[Figure 5-1] A workflow of the proposed multisource analysis process

1. **Relation Analysis**

   The first step of the whole process is to define relationship between sources. Between sources, any type of relation can occur thorough linkages such as Bluetooth, LTE, Wi-Fi, USB interface, and so on. Based on the characteristic of the source device, every linkage should be listed up before analysis based on possibilities.

2. **Selection of Analysis**

   After the previous step, privilege has to be set based on the relations of source devices. Every relation is defined based on information, and the part that is significant is where these data intersects. It basically means that giving roles are needed, which device will be client or server (host). With both types, analysis works must be done in parallel, in order to validate their relationship. For this reason, top-down and recursive method was chosen.

   After selection of sources to analyze, individual source has to be analyzed as normal single-source case is. Analysis of individual source needs understanding the source of interest, preparing appropriate forensic tools and techniques to start analysis. Then, acquisition and preservation of the evidence is needed.

3. **Function Analysis**

   Analysis of function is necessary for logical reasons. Logical means that there must be a justification when moving on to the next step. Although this process does not practically find forensically meaningful data, it helps to discover what situation could happen. It distinguishes specific functions that could have been used. Experiments and validation are required to apprehend existing features of interest.

4. **Artifact Analysis**

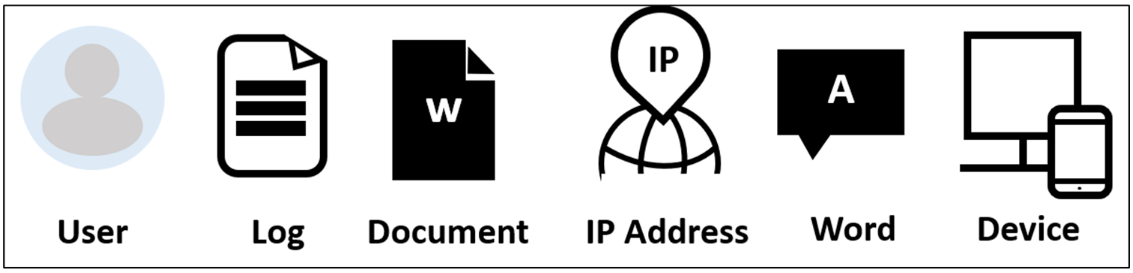
   Artifacts in this part mean as follows. It means directly or indirectly obtaining data remaining on the acquired device. This information is essential for cross checking data and putting strength on the evidence that is previously found or even helps to find a new data that is not found. Academical skills are required if necessary, such as reverse engineering knowledge and filesystem concepts. Utilizing existing tools and developing new tools are also important for doing that.

5. **Categorization**

   Results of analysis can be generated by using automated tools. However, investigators had to make their own decision based on the results and then find the path or evidence to make a conclusion. So, it would be more efficient if the results can be categorized based on various results. Categorization consists of two parts: (1) integrating data identified from multiple sources, and (2) correlating multisource information.

   "Integration" can be done by various factors. Most commonly used factor which is recommended is listed as follows:

   - Based on timestamps
   - Based on identities (name , ID, nickname, email, IP address, etc.)
   - Based on communication activities

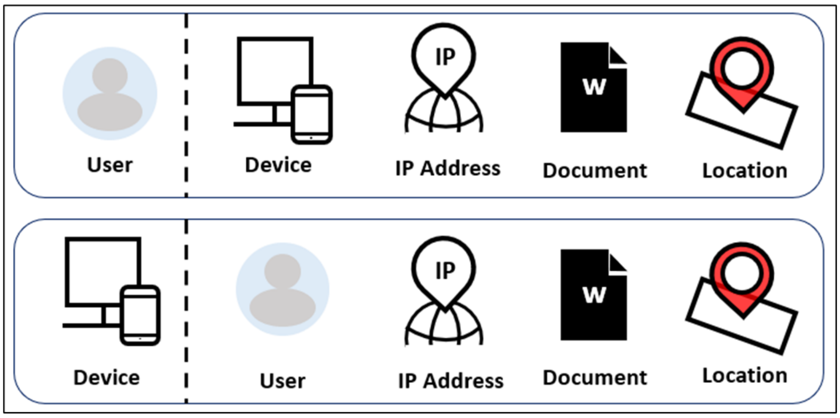


[5-2] Categorization Example

"Correlation" of multisource information also has many factors. Some examples of those factors are provided as follows:

- Analysis of communication activities (SMS, messenger, SNS, etc.)
- Analysis of suspicious activities related to the case
- Analysis of techniques used for criminal activities (destruction, deletion, encryption, obfuscation, etc.)

Also, it is necessary to match each factor with the other factors. Factors in the "User" category can match with device, IP address, location, or etc. This step can be of great help in analyzing user behaviors.

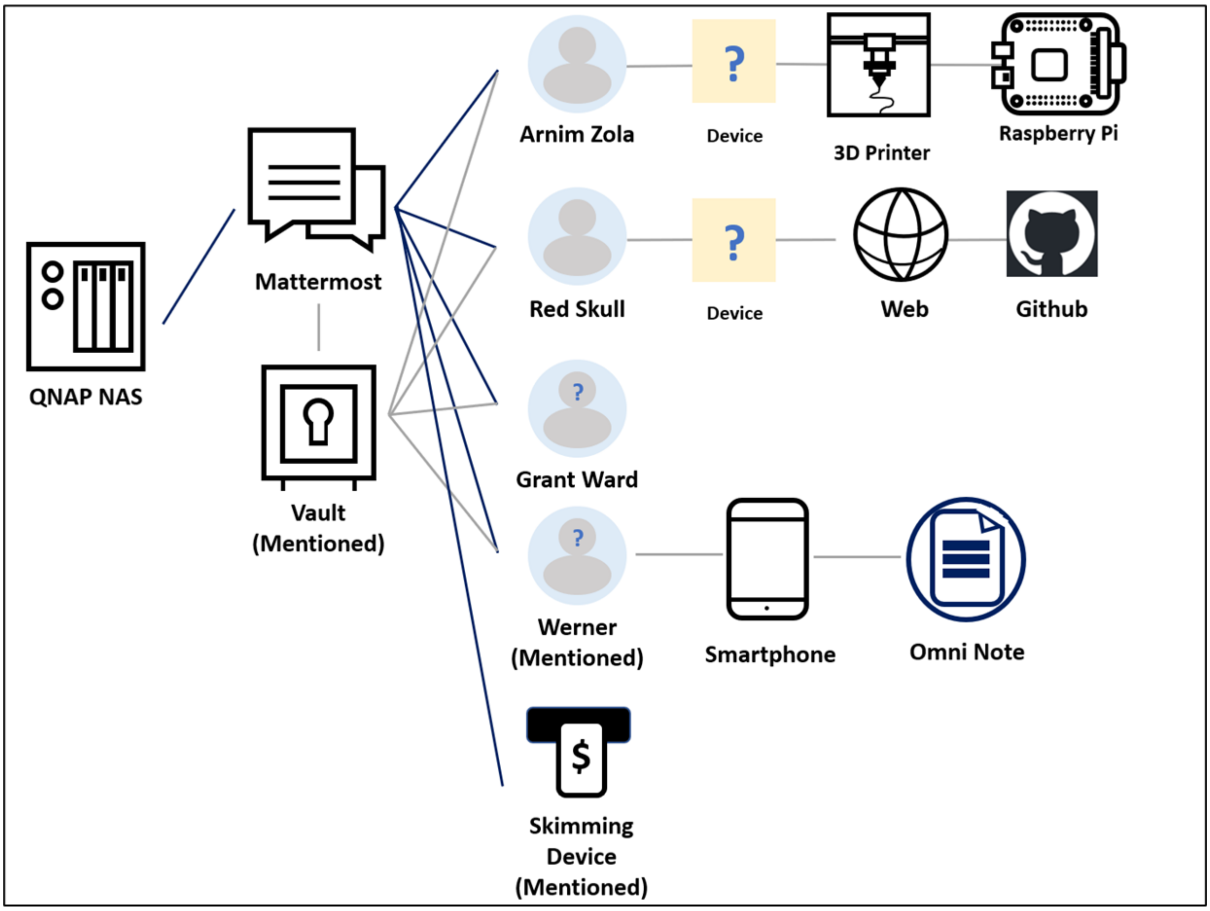


[5-3] Matching Example

6. **Behavior Analysis**

Using the categorized data, logical results are derived. Including the relationships between devices a specific result of analysis has to be mentioned. With this process, advanced timeline is constructed. Through 1~5 step, We established the relation of each image and the behavior of user more easily.

[Interpretation & Examination of Challenge Source]



[5-4] Integrative Process of Challenge Case

1. **Relation Analysis**

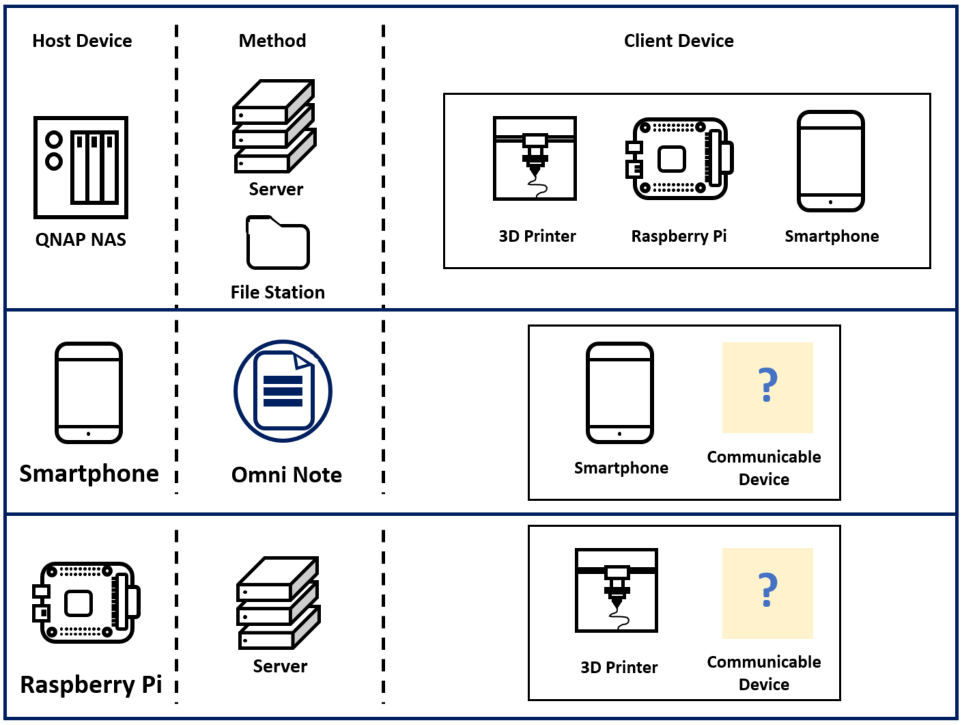
   The QNAP NAS uses “qpkg” to host several services including Mattermost, which can be used as a communication system. And these services can be used with other devices such as Raspberry Pi. Even though describing these relationships are important, listing these possibilities in tables are more readable as Table [5-1].

[Table 5-1] Relation Analysis of Challenge Case

Aa Method	☰ source1 (Host)	☰ source2 (Client(
Messenger	QNAP NAS Samsung Smartphone	Raspberry Pi Samsung Smartphone Unkown Device
File Station	QNAP NAS	Raspberry Pi Samsung Smartphone Unkown Device
Encrypted Folder/File Station	QNAP NAS	Raspberry Pi Samsung Smartphone Unkown Device
User Distribution	QNAP NAS Raspberry Pi	Unkown Device
Customized Application	Samsung Smartphone	Raspberry Pi Samsung Smartphone Unkown Device
Physical Connection	QNAP NAS	Raspberry Pi Samsung Smartphone Skimming Device Unkown Device
Physical Connection	Raspberry Pi	3D Printer Samsung Smartphone Unkown Device
Phone Call	Samsung Smartphone	Unkown Device

2. **Selection of Analysis**

   Thinking about the image we have, there are six interaction cross-sections that QNAP NAS have. Therefore, the first forensic investigation can be performed for the QNAP NAS with the most interaction cross-sections.

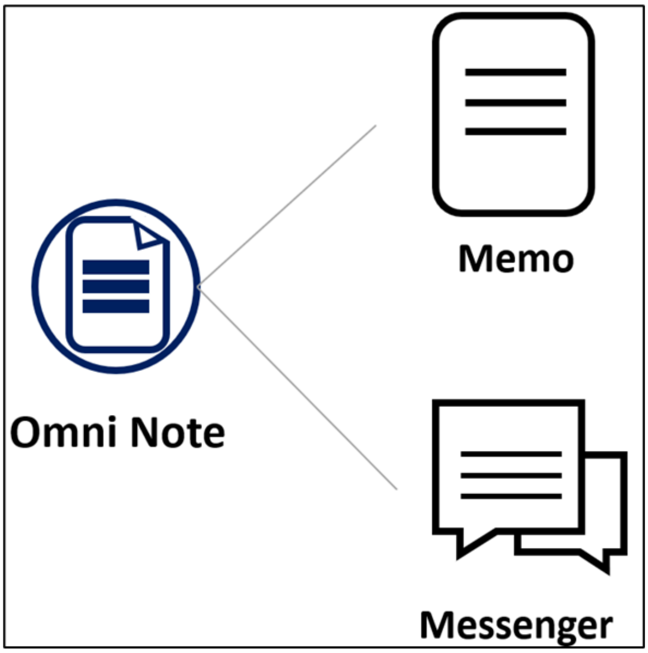


[5-5] : Real-relation anlysis of Challenge Source

The real-relation analysis picture of the device and the rest of the devices identified in this case is shown in Figure [5-5]. Among them, it can be seen that the QNAP NAS is most likely to be related with three client devices using two methods. In Table [5-1], it was confirmed that the NAS had the most connection points with other devices. Therefore, if examiners investigate and analyze the NAS first, they will be able to find connections with other deices, and it will be easy to connect to things common to analysis of other devices. There is a high probability that a device, which has the most intersections in the previous 'Relation Analysis', will eventually have the most intersections when it comes to real-relation analysis. Due to this reason QNAP NAS was selected initially.

3. **Function Analysis**

   A complete understanding of applications in each devices given. For a Samsung smartphone, full functions of the suspicious Omni Note application was needed to identify the message sending page. In case of the QNAP NAS, Mattermost is the application that requires knowledge of database structure of PostgreSQL. In this case, Omni Note has a hidden function on sending messages. Examiners can determine the range of checking artifacts if it is possible to know the fact that these functions exist and are executable.



[5-6] Function Analysis on Challenge Sources: Omni Note as an Example

4. **Artifact Analysis**

   In order to analyze the given images, examiners should know and understand some forensic knowledge, including but not limited to NAS, smartphone, Linux, Android, APK, web browser application, messaging service, RAID, magnetic stripe, and S/W reverse engineering skills. In addition, for every image, unallocated areas were analyzed for deleted files that could have meaningful information. Although it was not useful for every image, meaningful files were carved from the skimmer device related image. Beside the data which was manually found, existing tools (e.g., Autopsy) were used to get artifacts such as e-mails and registered users, web browser history, messages, OS information, and SNS related data. Every data found in the current step will be used in the next step.

5. **Categorization**

   Using existing tools like Autopsy, it was able to get forensically meaningful artifacts. But the tools usually do not provide functions to correlate between those artifacts. It is up to investigators relating the given artifacts. So, in order to do that, artifacts found in the previous step are needed to be categorized properly in this step. Table [5-2] lists the categorized results associated with multiple artifacts processed in the 'Artifact Analysis' step. Through categorization, relationships are made simply connected by each categorization factor. The path of evidence in the source is also given if it is a parsed information. For reference, any source is not given when it is carved from unallocated areas.

[Table 5-2] Categorized information

Categorization Factor	sources
Skimmer Device(Keyword)	Samsung smartphone Skimmer device
Leaving to Geneve(Event)	Samsung smartphone Skimmer device
Symbol of Hydra(img)	QNAP NAS Samsung smartphone Skimmer device
3D printed Liberator(Keyword)	QNAP NAS Raspberry Pi
Skimmer installation(Event)	Samsung smartphone Skimmer device
Mattermost(Service)	QNAP NAS Samsung smartphone
Vault(Service)	QNAP NAS Raspberry Pi Skimmer device
(Person)	QNAP NAS Raspberry Pi Samsung smartphone Skimmer device
CC data shared	Samsung smartphone Skimmer device

6. **Behavior Analysis**

In this step, details on the case will be analyzed by every categorized factors. Information such as “why it is related” and “what is learned from the relation” can be obtained as results. For the challenge dataset, “Description” in Table [5-3] explains the detailed behaviors revealed through this analysis.

After analysis, integrating and rearranging every events from multiple sources individually by timestamp, an integrated timeline on the case is built as listed in Table [5-4].

[Table 5-3] Behavior Analysis Result

Categorization Factor	sources	Connected	Description
Skimmer Device(Keyword)	Samsung smartphone Skimmer device	O	ATM skimmer machine was filmed by smartphone.
Leaving to Geneve(Event)	Samsung smartphone Skimmer device	X	SNS article that victim is leaving matches with the calendar information. Train ticket from Lausanne to Geneve was found. Date mismatches.
Symbol of Hydra(img)	QNAP NAS Samsung smartphone Skimmer device	O	Symbol image was found in every images.
3D printed Liberator(Keyword)	QNAP NAS Raspberry Pi	O	Same blueprint was found, which turns out to be Liberator Also found by DEFCAD.com File stored in QNAP NAS Vault
Skimmer installation(Event)	Samsung smartphone Skimmer device	O	Skimmer installed
Mattermost(Service)	QNAP NAS Samsung smartphone	O	Email on the smartphone implies another channel. Mattermost found in QNAP NAS
Vault(Service)	QNAP NAS Raspberry Pi Skimmer device	O	Decoded CC numbers stored in Vault Blueprints(STL) stored in Vault Vault is serviced in QNAP NAS

Categorization Factor	sources	Connected	Description
(Person)	QNAP NAS Raspberry Pi Samsung smartphone Skimmer device	O	**Arnim Zola** → Skimmer device, Raspberry Pi, QNAP NAS **Redskull** → Samsung Smartphone, QNAP NAS **Werener Von Strucker** → Samsung Smartphone, QNAP NAS **Grant Ward** → QNAP NAS
CC data shared	Samsung smartphone Skimmer device	O	identical mp3 file was found in each device since the file time is faster in smartphone, and data was not extracted in the file, this file seems to be a sample file.
Omninote installed	Samsung smartphone		

[Table 5-4] Timeline of Entire Case

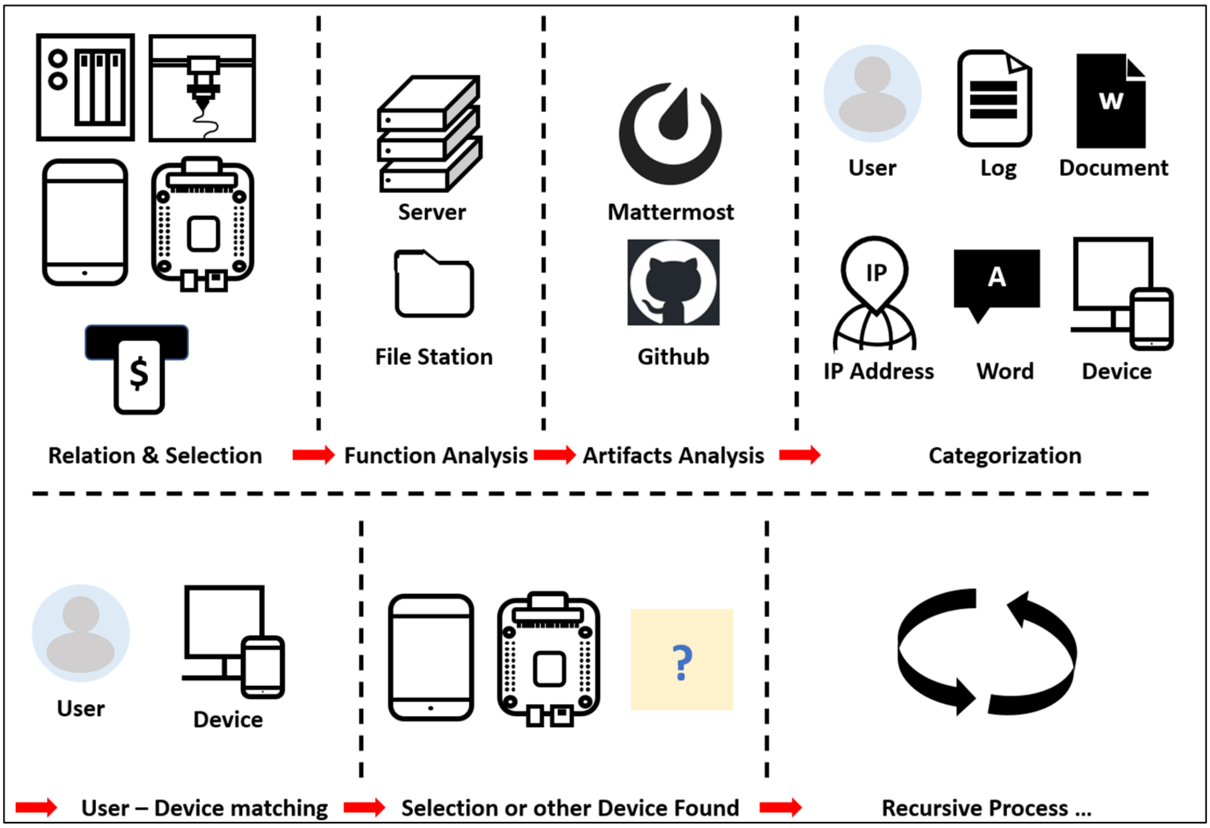
Event	Source	Date (UTC +02:00)	Description
Laussane → Geneve	Skimmer Device	2021/03/28 09:21	carved from unallocated space
Omninote installed	Smartphone	2021/04/08 11:59	
Omninote conversation	Smartphone	2021/04/09 16:10 ~ 2021/04/09 16:15	
Email	Smartphone	2021/04/08 17:04 ~ 2021/04/08 17:06	redskull & werner
Omninote conversation	Smartphone	2021/04/08 18:09 ~ 2021/04/08 18:16	
Sample mp3 file	Smartphone	2021/04/09 17:24	identical mp3 file was found in each device since the file time is faster in smartphone, and data was not extracted in the file, this file seems to be a sample file.
Sample mp3 file	Skimmer Device	2021/04/09 16:22	identical mp3 file was found in each device since the file time is faster in smartphone, and data was not extracted in the file, this file seems to be a sample file.
Victim Data skimmed	Skimmer Device	2021/04/09 16:20 ~ 2021/04/09 16:25	Modified time of Recording.mp3 which has the CC number.
Skimmer Discovered	Skimmer Device	2021/04/09 16:25	information given in the introduction.
Vault Created	QNAP NAS	2021/04/11 20:49	found in event.log file
DEFCAD visited	QNAP NAS	2021/04/11 21:38	by Arnim Zola
Mattermost conversation	QNAP NAS	2021/04/11 23:55	by Arnim zola & Red Skull
3D printer Browsing started	QNAP NAS	2021/04/12 14:06	
3D printer information Download start	QNAP NAS	2021/04/12 14:08	
Images of Liberator shared	QNAP NAS	2021/04/12 14:48	via Mattermost

Event	Source	Date (UTC +02:00)	Description
Mattermost conversation	QNAP NAS	2021/04/12 14:48 ~ 2021/04/12 14:57	by Arnim Zola & Red Skull
Octoprint Server Turned On	Raspberry Pi	2021/04/13 12:38	
Liberator parts printed	Raspberry Pi	2021-04-13 14:27 ~ 2021-04-14 17:29	
Mattermost visited	QNAP NAS	2021/04/17 20:34	via Chrome Cache by redskull
Mattermost conversation	QNAP NAS	2021/04/17 20:07 ~ 2021/04/17 20:19	by Arnim Zola & Red Skull
Raspberri Pi Lab found	Raspberry Pi	2021/04/18	Informaiton given in the introduction
Instagram	Smartphone	2021/04/20 13:01	“ready to leave”
Lausanne→Geneve	Smartphone	2021/04/20 16:45	Found in google calendar reminder
Samsung Smartphone extracted	Smartphone	2021/04/21	Information given in the introduction
Arrested in Geneve	Smartphone	2021/04/21 18:30	information given in the introduction
Vault		2021-04-28 14:49	found in Mattermost messages
Mattermost conversation	QNAP NAS	2021/04/28 14:49 ~ 2021/04/28 14:51	by Arnim Zola & Red Skull
Mattermost conversation	QNAP NAS	2021/04/29 07:51 ~ 2021/04/29 07:52	by Grant Ward
NAS discovered	QNAP NAS	2021/04/29	Information given in the introduction

Overall Process

Figure [5-7] shows an integrated process on handling the challenge dataset. As we explained above, the QNAP NAS with the most interactions are selected as the first target source. Then, examiners analyze functions that can be using the QNAP NAS and checking artifacts corresponding to it. The artifacts analyzed by using existing or self-developed tools can be categorized. Various factors in each category can be matched with factors in another category, and while proceeding with this, other devices can be detected as new sources of interest. The next step is recursively operated again. The device with many connections is selected, and then functions and artifacts are analyzed.

After this repetitive step, it is easy to judge the behavior of the user by using the categorization and matching factors. Therefore, the proposed process can discover correlation between multisource information as well as reduce the time for forensic analysis.



[5-7] Apply Process - Challenge Source

[Conclusion]

The most challenging part of multisource analysis is difficult to determine which source to analyze first and which artifacts to investigate in each source. With the proposed methodology, investigators will not get lost in the flood of information, by selection a source to analyze first and then relating sources logically. In multi source analysis, the most important thing is the correlation between different sources, and it can be found in places where data intersects.