Asia Pacific Fraud Journal
E-ISSN: 2502-695X, ISSN: 2502-8731
Volume 6, Issue 2 (July-December) 2021
*Asia Pacific Fraud Journal now has been accredited "**SINTA 3**" by Ministry of Research and Technology of The Republic of Indonesia (No SK. 85/M/KPT/2020).*
Availaeble online at: http://apfjournal.or.id/index.php/apf

Email Analysis in Fraud Investigation: Digital Forensic and Network Analysis Approach

✉Wishnu Agung Baroto
Direktorat Jenderal Pajak, Indonesia

ARTICLE INFORMATION
Article History:
Received September 06. 2020
Revised November 08, 2020
Accepted December 3, 2021

DOI:
10.21532/apfjournal.v6i2.212

ABSTRACT
Email is an imperative method of communication that is changing the way people share their data and information. It provides effective and efficient communication, especially in business, convenience, and easy access and replication. Those electronic data should be considered by a fraud investigator to comprehend the investigation. Email can be divided into two parts: the head of the email and the email body. The head of the email is metadata that consists of unstructured data, and the body and its attachment consist of semi-structured data. The email data usually comes in large volumes and ranges of types. Therefore, a manual investigation of an email should be avoided. This paper uses the Design Science Research Methodology to discover the most profound framework in an email fraud investigation. Using email metadata and email body, this research performs a digital forensic framework: preparation, gathering, processing, and presentation,combines with social network analysis to be applicable in the investigation. The result shows that digital forensics process, network analysis, data visualization provides a more valuable and comprehensive insight into email analysis.

Keyword: Digital Forensic, Network Analysis, Email, Fraud

1. INTRODUCTION

The existence of the internet and the development of information technology have significant changes in people's behavior. E-commerce, social media, communication, and digital currency are examples of ICT use that ease human life. Electronic data produced due to the massive use of ICT is an increase in intensity and types. It creates a new term called Big Data. The term describes not only the amount of digital data produced, but also describes the variety of data, the velocity of data, and the veracity of data, as typical abbreviated as 4V's (Volume, Variety, Velocity, and Veracity). It has recently moved to 5V's by considering Value as the last V. The fast growth of data is mostly in the form of unstructured data. It is a type of data that has no structured pre-defined data model or schema. Examples of this data is an email message, audio files, video files, pdf reports, and other digital information (Baroto and Prasetyo, 2020).

Email is one of the primary communication tools in this internet era, which produces a lot of electronic data. It is effective and efficient communication, simple and convenient, and easy access, and replication. However, it is quite common that fraudsters use email to commit fraud or to collaborate with their co-inspirators. Email scam in terms of spoofing, phishing, and bogus offer are some types of the use of email in cybercrime. There is also another term to explain the use of email as a means of fraud, the cyber-related crime. This is related to the use of email

✉Corresponding author :
Email: wishnu.ab@gmail.com

to communicate between fraudsters, store data, or collaborate in planning fraud. Financial fraud, asset misappropriation, tax fraud, and data fraud are the most frequent types of fraud in cyber-related fraud.

Fraud examiner or fraud investigator must gather all available and related evidence, including digital or electronic data. Investigators should understand and capable of obtaining and analyzing digital data as evidence besides physical evidence. The problem may arise because one email may comprise thousands of mail data and contacts. Thus, the investigator should utilize digital forensic tools to process the email's contents. On the other hand, the header of the email also comprises metadata that beneficial to comprehend the investigation, especially to find intention or *mens rea* of the fraudster.

This paper elaborates on the investigation of email using a general framework of digital evidence. All the procedures performed to ensure the process is forensically sound manner. However, the general process of digital forensics has not captured the procedure of link analysis. This paper searches for an appropriate framework and process to conduct email analysis, both the email body and email header. Thus, the objectives of this research are:

a. As proof of the concept that digital forensic and network analysis beneficial on fraud investigation.
b. Test the tools for conducting digital forensic and network analysis.
c. All the processesare forensically sound manner as a requirement of digital evidence in a trial.

2. LITERATURE REVIEW AND HYPOTHESIS

Fraud Investigation

In an organization, fraud examination is carried out for various objectives as follows (ACFE, 2019): identifying improper conduct, identifying the persons responsible, stopping fraud, sending a message that fraud will not be tolerated, determine the extent of potential losses, facilitate the recovery, prevent future losses, mitigate other consequences, and strengthen internal control.

Fraud Examiners' role in an investigation is mostly divided into four activities: obtaining evidence, reporting, testifying, and assisting in fraud detection and prevention.

a. Obtaining evidence

The value of a fraud examination rests on the credibility of the evidence obtained. Evidence of fraud usually takes the form of documents or statements by witnesses; therefore, fraud examiners must know how to obtain documentary evidence and witness statements legally and adequately.

b. Reporting

Once the evidence has been obtained and analyzed, and findings have been drawn from it, the fraud examiner must report the results to the designated individuals (e.g., management, the board, or the audit committee). A fraud examination report is a narration of the fraud examiner's specific activities, findings, and, if appropriate, recommendations.

c. Testifying

Often, fraud examiners are called upon to provide testimony and report their findings at a deposition, trial, or other legal proceedings. When providing testimony, fraud examiners must be truthful. They should also communicate clearly and succinctly.

d. Assisting in fraud detection and prevention

Fraud examiners are not responsible for preventing fraud; such responsibilities belong to management or other appropriate authority. Nevertheless, fraud examiners are expected to actively pursue and recommend appropriate policies and procedures to prevent fraud.

An investigator may utilize digital forensics tools to recover and investigate material found in a digital device to support the investigation.

Digital Forensic

Digital forensics is a branch of forensic science discipline where scientific principles, methodologies, and techniques are used in the investigation (Sachowski, 2016). Data in digital forensic is divided into two categories: the volatile and nonvolatile data. Each of the categories has a different method in managing the digital evidence. Digital forensic usually initialize after an incident occurs. (Baroto and Darajat, 2020). First, the Digital forensic examiner assesses the information system and other preparation, and then acquire or collect and preserve digital evidence.

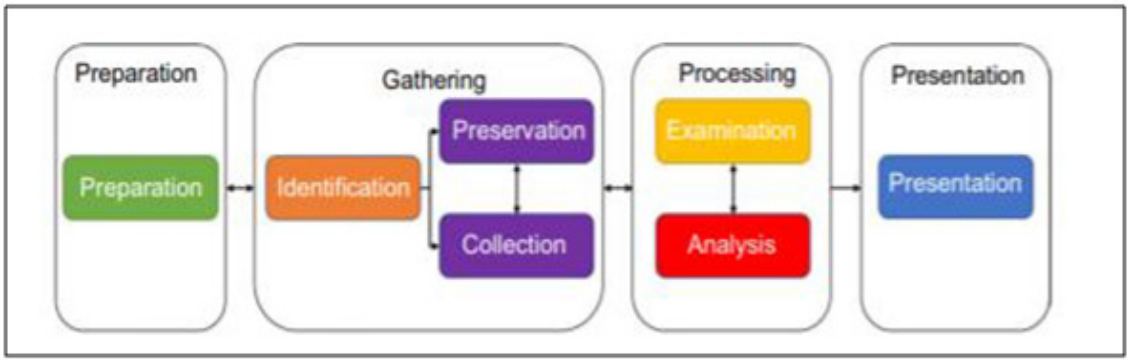
Moreover, a digital forensic examiner then conducts an examination, analysis, and presentation of the investigation's results and findings. From 1995 until 2011, at least 21 proposed frameworks in digital forensic procedures have been issued by many scholars (Oettinger, 2020). However, the most general steps are preparation, identification, collection, preservation, examination, analysis, and presentation (Sachowski, 2016).

a. Preparation includes actions to guarantee equipment and personnel are organized;
b. Identification contains detection of an incident;
c. The collection covers any evidence acquisition using standardized techniques;
d. Preservation creates proper evidence collection and the chain of custody;
e. The examination evaluates digital evidence volumes, protected files, registry analysis;
f. The analysis examines the content and context of digital evidence, determine relevancy, link, and analysis of the root cause of the incident;
g. The presentation comprehends the reports for documentation of all processes.

According to the literature, there are four general principles as a useful practice guide for digital evidence (ACPO,2012).

a. No action taken by law enforcement agencies, persons employed within those agencies or their agents should change data which may subsequently be relied upon in court.
b. In circumstances where a person finds it necessary to access original data, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
c. An audit trail or other record of all processes applied to digital evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
d. The person in charge of the investigation has overall responsibility for ensuring that the law and these principles are adhered to.

Figure 1. **Digital Forensic Framework**



Source: Sachowski, 2016

Network Analysis

Data has three types: ideational data, attribute data, and relational data. Ideational data gives motivations, meanings, and definitions into conceptions and typological analysis. Attribute data is about indexes like attitude, characteristics, choices, and preferences, which can be examined using correlation analysis and regression analysis (Knoke et al., 1982). The last, relational data describe relations among units, either static or dynamic (Knoke and Yang, 2008). The essential concept of network analysis is centrality, related to the data on the nodes' structural dimension (Li, 2013).

a. Centrality

Centrality indicates which node is essential in a network. The more node has a relationship with other nodes; it goes to the center of the network. Therefore, the node has a higher power or influence in the network (Sparrowe, 2001). The most common centrality measures are degree centrality, betweenness centrality, and closeness centrality.

b. Degree Centrality

Degree centrality is calculated by the total amount of links of a node with other nodes.

c. Betweenness Centrality

Betweenness centrality measures the role of a node in a network.

d. Closeness Centrality

Closeness centrality measures the distance of one node to other nodes in a network.

Network analysis, or sometimes called Social Network Analysis, is a non-financial fraud detection tool (Omar et al., 2014). Furthermore, network analysis is also useful to support financial fraud (Alamsyah et al., 2013).

Previous Study

Many researchers over the last decades have discussed the use of e-mail forensic in an investigation. Devendran et al. (2015), examines a set of standard features of open-source e-mail forensic tools and suggests a combination of analysis tools to comprehend the investigation. The author explores Mailxaminer, Add4Mail, eMailTrackerPro, Digital Forensic Framework, and Paraben E-Mail Examiner and examines those tools in nine criteria: input file in disk, search option, information provided, recovery capability, format supported, visualization format supported, the operating system, export format, and extended device support. The result is almost like the work of Banday (2011). Moreover, Banday (2011) also portrays e-mail actors, roles, and their responsibilities using meta-data contained in e-mail message.

Paper in e-mail forensic commonly explores forensic tools to recover, search, and visualize the result, as mentioned by Devendran (2015). Banday (2011) explains the role of actors in e-mail forensic to describe a more comprehensive analysis. This paper combines the work of Banday (2011) and Devendran (2015) using a real case of investigation to comprehend the investigation, as shown in table 1.

3. METHODS

This research utilizes Design Science Research Methodology as a method suggested in Information Science and Computer Science research (Hevner, 2004). The process of DSR are as follows:

a. Identify problem and motivation

Define the research problem and justify the value of a solution. This is the stage to show importance of the research.

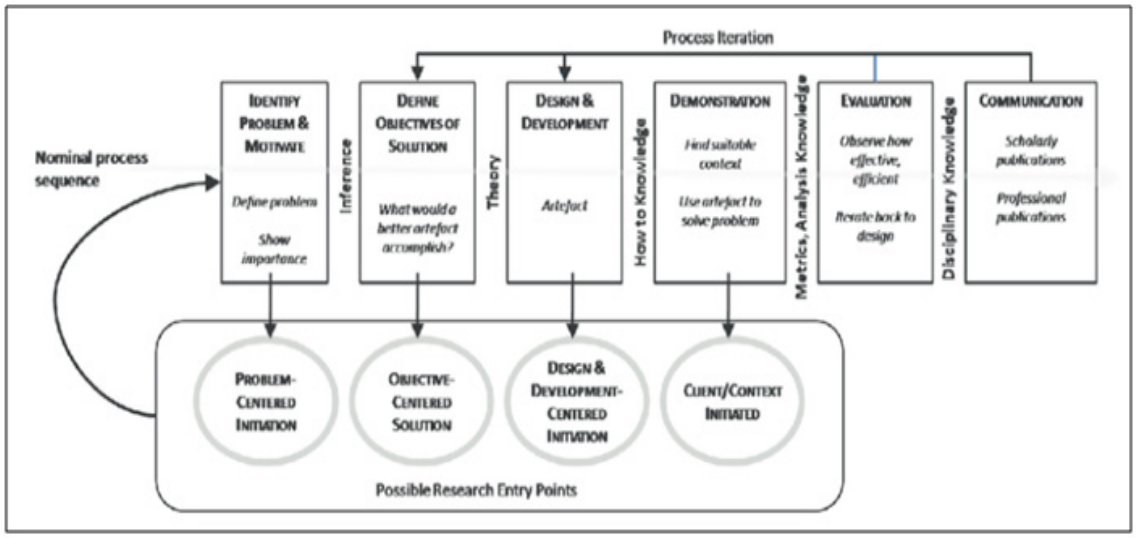
b. Define objectives

In addition to general objectives such as feasibility and performance. This includes a specific criterion of a solution.

Table 1. **Previous Study and This Paper Novelty**

	Forensic Framework	Network Theory	Descriptive	Case
Devendran	X			X
Banday	X	X	X	
This Research	X	X	X	X

Figure 2. **Design Science Research Methodology**



Source: Peffers et al., 2008

c. Design & development

Create constructs, models, or methods in which a research contribution is embedded.

d. Demonstration

Prove that the artifact works by solving one or more instances of the problem.

e. Evaluation

Observe and measure the artifact supports a solution to the problem.

f. Communication

Communicate the problem, its solution, and the utility, novelty, and effectiveness of the solution to researchers and other relevant audiences.

4. RESULTS AND DISCUSSION

The process of this research is divided into six stages: identify problem, define objectives, design & development, demonstration, evaluation, and communication, as required by the research methodology. In this section, the procedures and result on digital forensics process is the identical as Wishnu and Prasetyo (2020).

a. Identify problem and motivation.

The problem of this research is mentioned in the introduction of this paper, which is to establish an appropriate framework of email analysis using digital forensics process and social network analysis.

b. The objectives of the process are:

- As proof of the concept that digital forensic and network analysis beneficial on fraud investigation.
- Test the tools for conducting digital forensic and network analysis.
- All the processes are forensically sound manner as a requirement of digital evidence in a trial.

c. Design & development

This stage is one of the most essential in the whole research process. This paper first analyze the current digital forensics framework, demonstrate the process, analyzed the results, and analyze email using social network analysis.

a. Preparation

First, the preparation process consists of an analysis of whether the fraudster uses email to communicate. Manual internet history analysis or tool-based analysis can find digital footprints of email in the internet history. In this paper, we utilize the internet history browser as part of DART (Digital Advanced Response Toolkit) (Table 2).

b. Gathering

- Importation

  This is an additional process in the digital forensic framework to accommodate the network analysis process. This process covers an email

client's installation consisting of an email login (username and password) in Outlook, setting the IMAP, and then synchronizing the web-based email and email client in the Send/Receive menu.

- Collection

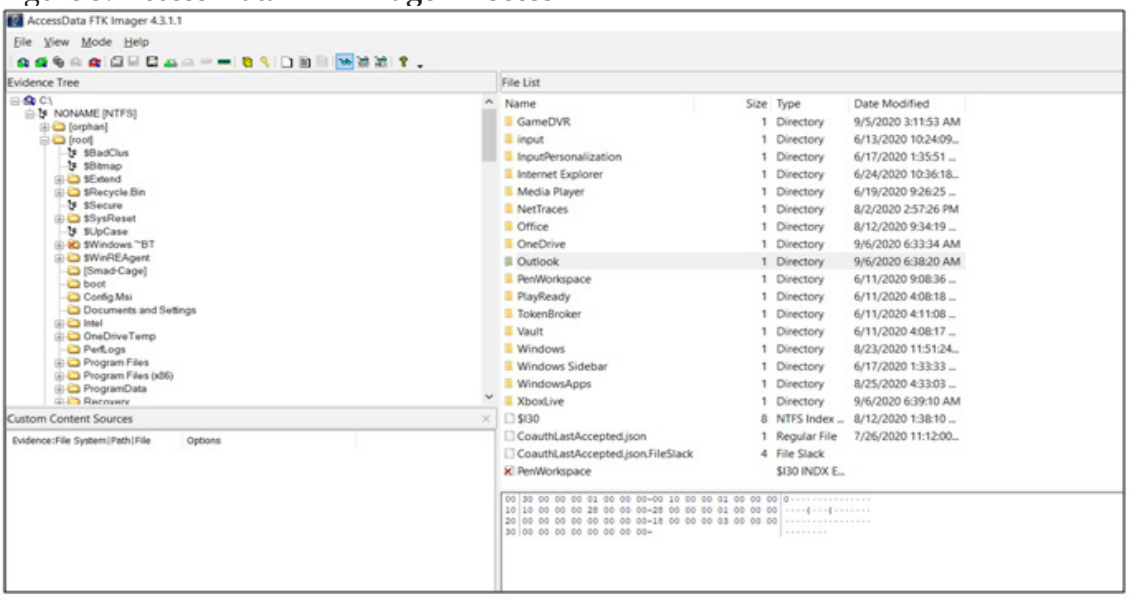
After all the content of the email synchronize, then find the backup file (.ost or .pst) in the root of the Outlook folder. Generally, it is stored in: C:\Users\$USER$\AppData\Local\Microsoft\Outlook.

Table 2. **Tools Used in This Research**

No.	Tools	Use	Information
1	Outlook 365	To process email from web-based email to client-based email and export the metadata of email.	Outlook is an anchor app in Microsoft that mainly used for sending and receiving emails. This tool suggested by Devemdran (2011)
2	FTK Imager 4.3.1	To create an image file of the backup email file to maintain its integrity.	FTK Imager is part of AccessData tools on forensic investigation. This tool suggested by Baroto and Prasetyo (2020)
3	Autopsy 4.15.0	To analyze metadata and content of an email, indexing, and keyword searching.	Autopsy is a digital forensics platform and graphical interface of The Sleuth Kit form Basis Technology. This tool is suggested by Baroto and Prasetyo (2020)
4	Maltego Case File 4.2.11	To perform a social network analysis from email metadata.	CaseFile is visual intelligence application that can be used to determine the relationships and real world links between hundreds of different types of information.

Source: Data Processed, 2020

Figure 3. **Access Data FTK Imager Process**



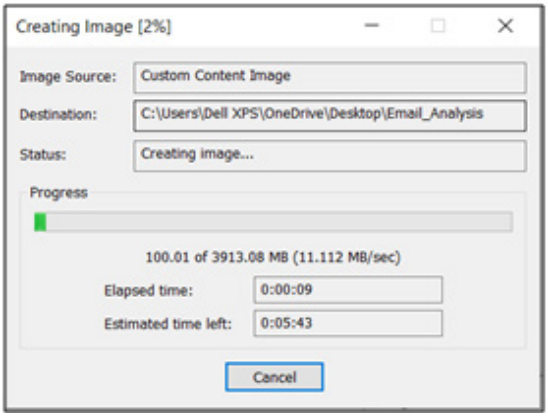
Source: Data Processed, 2020

- Preservation

  Using FTK Imager, the process of imaging the evidence and documenting the result of the image.

Furthermore, using Content of Folder feature to image only designated data then image the evidence.

Figure 4. **AccessData FTK Imager Process**



Source: Data Processed, 2020

The result of this process is at least two files: one is the image of the evidence, and the other one is the manifest of the image, which includes the MD5 and SHA1 as a digital fingerprint of the evidence. This digital fingerprint should be documented and maintained as legal evidence in the court or other disputes.

c. Processing

The next phase of digital forensic analysis is examining and analyzing the evidence. First, the evidence should be copied, and the investigator works on a copy of the evidence. To support the process, Autopsy Digital Forensic can be used as a tool for processing.

- Examination

  To process the evidence, first, we must create a new case. In this paper, the Suspect Case is the name of the case. Then, the investigator must conduct the following steps:

  - Add the evidence
  - Run ingest module

  This is the step to configure several evidence processing. The most common process for email investigation is:

  - Email parser

    This module detects and parsers file .mbox and .pst or.ost files and populates email artifacts.
  - Keyword search

    Performs file indexing and periodic search using keywords and regular expressions in lists.
  - Embedded file extractor

    Extracts embedded files, schedules for ingestion, and populates directory tree.
  - File type identification

    Matches file types based on binary signatures.
  - Extension Mismatch detector

Figure 5. **Digital Fingerprint in FTK Imager**

Source: Data Processed, 2020

Flags files that have a non-standard extension based on their file type.

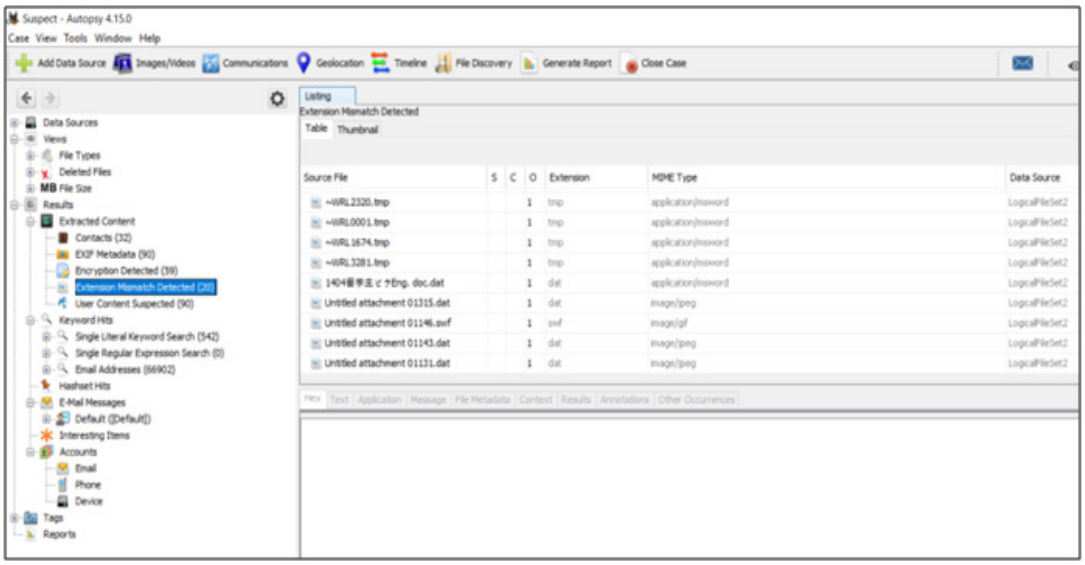
- Encryption detection
  Detect encrypted files with specified minimum entropy.

d. Analysis
Analyzing the evidence depends on the purpose of the investigation. For example, the most common tax fraud investigation cases are fraudulent tax invoices or usually called as *fakturpajak yang tidak berdasarkan transaksi yang sebenarnya* (tax invoices that are not based on actual transactions). Therefore, a keyword search of "fakturpajak" will be necessary to be performed. The result shows that several hits in the "fakturpajak" word search, which support investigators for further examination.

e. Centralities
This part also an addition in the digital forensic framework to expand the information in email metadata. The procedures are as follows:

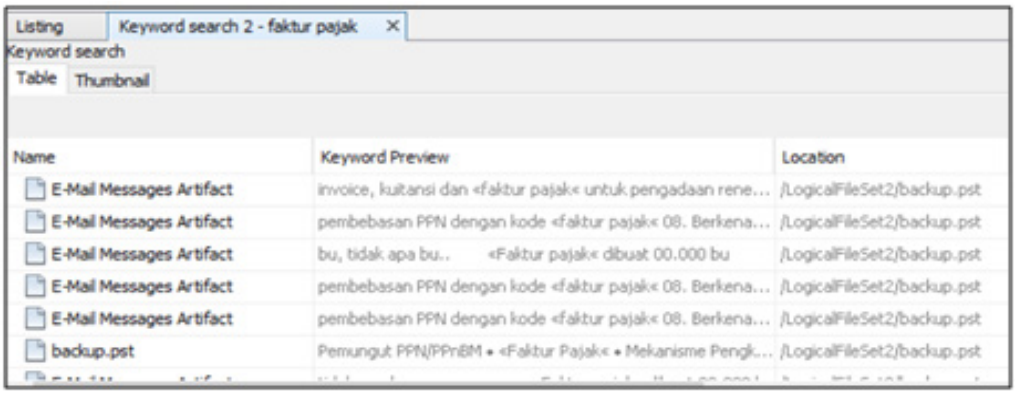
- Exporting metadata
  Using Outlook, the metadata of email can be extracted easily. Using the export file menu, then all contacts can be exported to a tabular spreadsheet.

Figure 6. **Autopsy Process Analysis**



Source: Data Processed, 2020

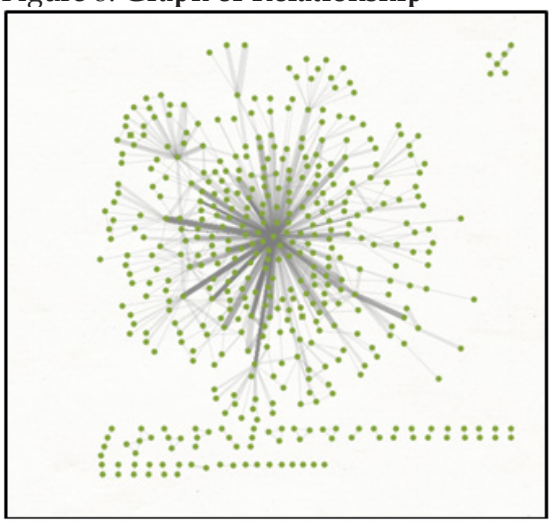
Figure 7. **Keyword Search in Autopsy**



Source: Data Processed, 2020

- Analyze
  The file then analyzes whether the results are in a readable format.
- Graph
  Using Maltego, then all the contacts graphed to depicts the relationship among parties in the email.

Figure 8. **Graph of Relationship**



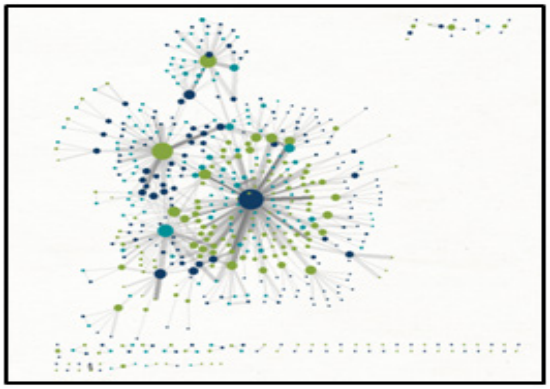
Source: Data Processed, 2020

- Network Analysis
  The graph without considering centralities measurement will not provide information onthe role of each node. Therefore, we need to add a degree, betweenness, and closeness to provides more information.

f. Presentation
The last phase is to present the result, which consists of visualizing the network analysis and creates a standard report of a comprehensive examination.

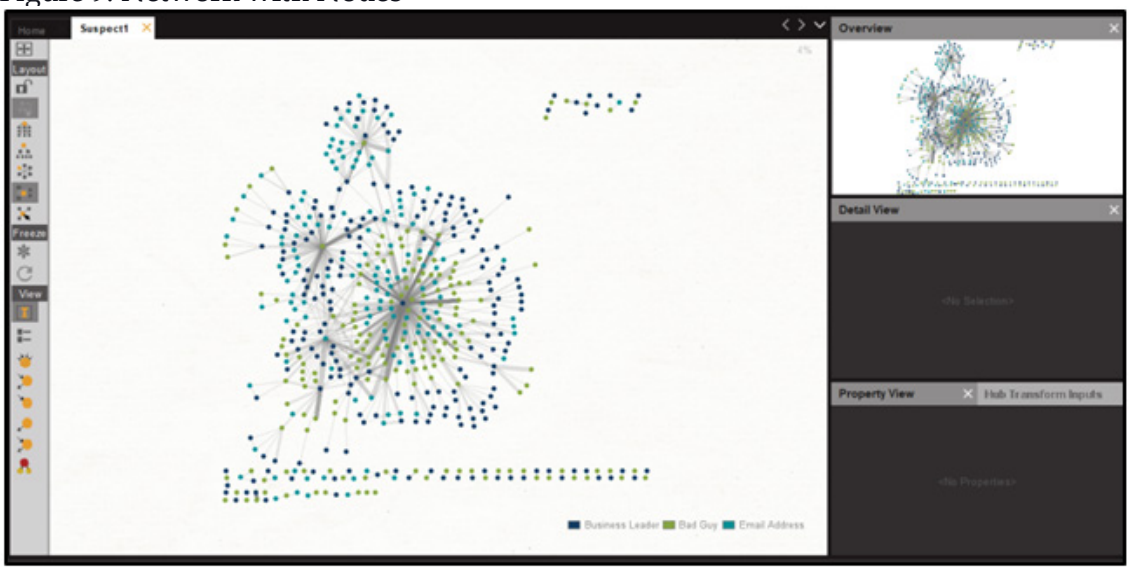
- Visualization
  Using Maltego, we add all the centralities measurement: degree centralities, betweenness, and closeness, and the result shows the relationship of nodes more informative, especially communication intensity and closeness in the network.

Figure 10. Network Analysis



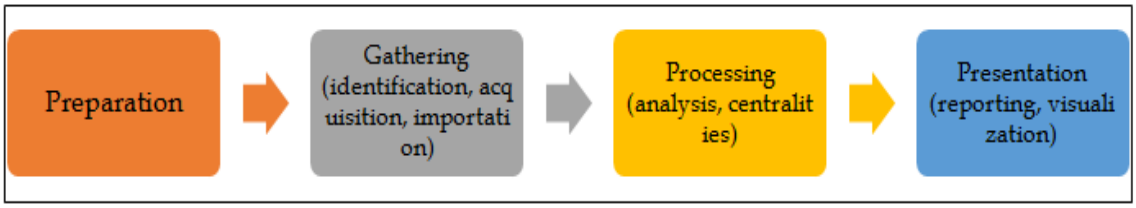
Source: Data Processed, 2020

Figure 9. **Network with Nodes**



Source: Data Processed, 2020

Figure 11: The Process



Source: Data Processed, 2020

- Reporting
  Both Maltego and Autopsy provide a standardized report that can be performed in the reports tab to assist the investigator in creating a report, communicating, or disseminating the investigation results.

d. Demonstration
   All the process above includes a demonstration of the process.
e. Evaluation and Communication
   These parts are reflected in the conclusion and discussion of this paper.

A combination of digital forensic process and network analysis creates a new integrated approach on email analysis. The summary of process can be described as follows (Figure 11).

5. CONCLUSION

This research uses the Design Science Research Methodology to determine the process of digital forensic and network analysis in the email fraud investigation. Based on the research process and demonstration, we can conclude that: 1) The integrity of data in emails (metadata and the contents of the email) can be maintained by a forensically sound manner process. 2) The body of email can be extracted for further analysis (keyword search) and more advance analysis such as sensitivity analysis. 3) Header of email consists several useful data which need to be extracted for further analysis. 4) Network theory able to support investigator to find suspects, eliminate unnecessary data, and visualize relationship. 5) The need of email forensic framework which combines digital forensic and social network analysis.

REFERENCES

Alamyash, Andry, Rahardjo Budi.(2013). Financial Fraud Detection using Social Network Analysis, e-Indonesia Initiatives (eII-Forum).

Association of Certified Examiners. (2019). Fraud Examiners Manual.

Banday, M. Tariq. (2011). Techniques and Tools for Forensic Investigation of E-Mail, International Journal of Network Security & Its Application (IJNSA), Vol 3, No. 6.

Baroto, Wishnu Agung and Darajat, Firman, Digital Forensic Readiness for Micro, Small, and Medium Enterprise in Indonesia. (2020). International Journal of Management and Applied Science Vol 6, Issue 1, 25-30.

Baroto, Wishnu Agung and Prasetyo, Ardhianto H. (2020). Digital Forensic Process in Fraud Investigation: A Case Study on Email Analysis. International Journal of Scientific Engineering and Science Volume 2, Issue 9, 36-40.

Devendran, V., Shahriar, H., and Clincy, V. (2015). A Comparative Study of Email Forensic Tools, Journal of Information Security, 6, 111-117.

Hevner R., A., Salvator T., Jinsoo Park, & Sudha Ram. (2004). Design Science in Information Science.

Knoke David, H., James, Kuklinski, Network Analysis. (1982). Beverly Hills:Sage Publication.

Knoke David, H., Yang S. (2008). Social network analysis. Sage Publication.

Li, Mingxiang. (2013). Social Network and Social Capital in Leadership andManagement Research: A Review of Causal Methods." LeadershipQuarterly 24.5(2013):638–665.

Oettinger, William. (2020). Learn Computer Forensics, Packt.

Omar, Normah, Mohamed, Ismail Sanusi, Zuraidah, and Prabowo, Hendi Yogi. (2014). Understanding Social Network Analysis in Fraud Detection, Recent Trends in Social and Behavior Scieance, Taylor & Francis Group.

Peffers, Ken, Tuunanen, Tuure, Rothenberger, Marcus A., and Chatterjee, Samir (2007). A Design Science Research Methodology for Information Systems Research, Journal of Management Information Systems, Volume 24 Issue 3.

Sachowski, Jason (2016). Implementing Digital Forensic Readiness from Reactive to Proactive Process,Elsevier.

Sparrowe, Raymond T., Liden, Robert C., Wayne, Sandy J., Kraimer, Maria L. (2011). Social Networks and the Performance of Individuals and Groups. The Academy of Management Journal.

Zhang, Junlong and Luo, Yu. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network, Advances in Intelligent Systems Research volume 132.