A Novel Approach to Network Forensic Analysis: Combining Packet Capture Data and Social Network Analysis

Irwan Sembiring[1], Suharyadi[2], Ade Iriani[3], Jenni Veronika Br Ginting[4], Jusia Amanda Ginting[5]
Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia[1, 2, 3]
Institut Teknologi dan Bisnis Indonesia, Medan, Indonesia[4]
Univeristas Bunda Mulia, Jakarta, Indonesia[5]

***Abstract*—Log data from computers used for network forensic analysis is ineffective at identifying specific security threats. Log data limitations include the difficulty in reconstructing communication patterns between nodes and the inability to identify more advanced security threats. By combining traditional log data analysis methods with a more effective combination of approaches, a more comprehensive view of communication patterns can be achieved. This combined approach can then help identify potential security threats more effectively. It's difficult to determine the specific benefits of combining Packet Capture (PCAP) and Social Network Analysis (SNA) when performing forensics. This article proposes a new approach to forensic analysis that combines PCAP and social network analysis to overcome some of the limitations of traditional methods. The purpose of this discovery is to improve the accuracy of network forensic analysis by combining PCAP and social network analysis to provide a more comprehensive view of network communication patterns. Network forensics, which combines pcap analysis and social network analysis, provides more comprehensive results. PCAP analysis is used to analyze network traffic, conversation statistics, protocol distribution, packet content and round-trip times. Social network analysis maps communication patterns between nodes and identifies the most influential key players within the network. PCAP analysis efficiently captures and analyzes network packets, and SNA provides insight into relationships and communication patterns between devices on the network.**

***Keywords*—PCAP analysis; social network analysis; network forensic; network communication pattern**

I. Introduction

Network forensic analysis is an important tool for identifying and tracking malicious activities on the network [1][2]. Traditional methods of network forensic analysis using log data are not effective in identifying specific security threats [3]. The fundamental causes of these issues are the log data's limits in recreating communication patterns between nodes and its inability to recognize more sophisticated threats. [4]. In recent years, there has been increasing focus on combining packet capture analysis and social network analysis in network forensic analysis to improve the accuracy and completeness of the analysis. PCAP analysis involves capturing and analyzing data packets transmitted over a network, while social network analysis involves visualizing and analyzing the relationships and communication patterns between devices on the network [1][5][6][7]. The research question is how social network analysis can be used in network forensics to identify potential suspects and their relationships within a network.

The combination of these two approaches has the potential to provide a more comprehensive view of network communication patterns and more effectively identify potential security threats. However, little is known about the specific benefits of combining PCAP and social network analysis in network forensic analysis. In this paper, the main contribution is to propose a new approach to network forensic analysis combining PCAP and social network analysis to address the challenges and limitations of traditional approaches.

The goal of this discovery is to combine PCAP and social network analysis to improve the accuracy of network forensic analysis by providing a more comprehensive view of network communication patterns. By combining social network analysis and PCAP analysis, it is possible to gain a deeper knowledge of network activity and communication patterns. Investigators can use this to find unusual or suspect activities on the network, such as secret or encrypted communication.

II. Related Work

A. Research Related to Packet Capture (PCAP) Analysis

Sikos [1] made a comprehensive comparison of Carnivore, Snort, Windump, Wireshark, dsniff, tcpdump, Omnipeek, Solarwind and other packet analysis tools when analyzing PCAP data. The method used is AI-based deep learning inspection combined with semi-supervised machine learning. The research aims to compare the ability to recognize patterns in different packet analyzer applications in order to find the most suitable tool for network forensic analysis activities. The results of DPI (Deep Packet Inspection), a packet analysis tool with machine learning capabilities are valid [8][9][10].

Cappers et al. [5] focused on data reduction and visualization techniques using the EventPad tool. The purpose of this study was to conduct a safety analysis in a behavioral pattern study using PCAP data. This study presents a case study of the EventPad visual analysis tool to obtain attack profiles and traffic analysis using rules and aggregations. The study did not describe any communication patterns at the application level.

Shrivastava et al. [11] focused on capturing attacks on IoT devices using Cowrie honeypots and using machine learning to classify attack types. They apply various machine learning algorithms namely Naive Bayes, J48 Decision Trees, Random Forest, and Support Vector Machines (SVM) to classify attacks such as malicious payloads, SSH attacks, XOR DDoS, espionage, suspicious and clean attacks. Perform feature selection using subset evaluation and best-first search. The training results achieved an accuracy rate ranging from 67.7% to 97.39%.

B. Network Analysis Research using Social Network Analysis (SNA)

Chakraborty et al. [12] conducted a study that advances the understanding of 5G-COVID-19 conspiracy theories. This paper conducts a social network analysis to analyze the content of Twitter data over a seven-day period (the #5GCoronavirus hashtag became trending on Twitter in the UK. The content analysis revealed that 34.8% (n=81) of a sample of 233 tweets contained references to 5G and COVID-19-related opinions, 32.2% (n=75) were critical of conspiracy theories, 33.0% (n=77) were general tweets, not disclosing views or personal opinions) tweets were from non-conspiracy theory supporters, indicating that despite interest in the topic is high, but only a small percentage of users actually believe in the conspiracy theory. Liu et al. [13] provided a large-scale group decision-making model based on the process of propagating beliefs; the process of conflict detection and resolution; and the process of selection using social network analysis methods. In the first procedure, we propose a relation strength-based belief propagation operator, which allows building a complete social network while considering the effect of relation strength on propagation efficiency. In the second procedure, we define the notion of degree of conflict and measure the degree of collective conflict in conjunction with assessments of information and belief relationships among large groups of decision makers. SNA is a modeling of users represented by nodes and interactions between users are represented by lines (edges). This analysis is needed because it brings new opportunities to understand individuals or communities regarding their social interaction patterns [20] [21]. SNA can be used to study network patterns of organizations, ideas, and people who are connected in various ways in an environment [22] [23]. Degree centrality counts the number of connections or interactions a node has. To calculate the value of centrality degree (CD), we use Eq. (1) [24].

$$CD(i) = \sum_{\substack{j=1 \\ i \neq j}}^{N} Xij \tag{1}$$

Closeness centrality (CC) calculates the average distance between a node and all other nodes on the network. This measure describes the proximity of this node to other nodes [24] as in Eq. (2).

$$cc(i) = \frac{N-1}{\sum_{J=1}^{N} dij} \tag{2}$$

Betweenness centrality (BC) calculates how often a node is passed by another node to go to a certain node in the network. This value serves to determine the role of the actor who is the bridge that connects the interaction in the network. To calculate the value of degree centrality we use Eq. (3) [24].

$$Cb(i) = \sum_{J=1}^{N} \sum_{k=1}^{j-1} \frac{gjk(i)}{gjk} \tag{3}$$

One of the most important processes at the digital forensic stage is data integrity in the preservation section. Message-digest algorithms MD5 and SHA-1 as one-way cryptographic hash functions are used in integrity validation [23] [24]. Four non-linear functions in 512bit blocks in the MD5 Algorithm as Eq. (4).

$$F(B,C,D) = (B \wedge C) \vee (\neg B \wedge D)$$

$$G(B,C,D) = (B \wedge D) \vee (C \wedge \neg D)$$

$$H(B,C,D) = B \oplus C \oplus D$$

$$I(B,C,D) = C \oplus (B \vee \neg D) \tag{4}$$

C. PCAP Data and Network Forensics Analysis

PCAP analysis in cyber forensics can be performed using a variety of methods including using software, PCAP software analysis. The software can be used to view and analyze packet content, including headers and payloads, and look for signs of malicious activity [14]. The next approach is statistical analysis, which extracts statistics from PCAP, such as packet count, network traffic and protocol statistics [15]. A rather important approach is packet and payload analysis to extract information from packet headers such as: B. Source and destination IP addresses, protocol and port used, and payload from the packet [10][11] [17][18]. Other findings indicate that computer network traffic results provide a variety of valuable information in graphical form to help identify routine banking transactions (pooled accounts, straw men, smurfing) used to hide movement of prohibited resources or obfuscation, thereby enhancing the visualization of financial analysis aspect. Packet analysis of internet network traffic is an important backtracking technique in network forensics, if the captured packet details are detailed enough, it can even show all network traffic at a specific point in time. This can be used to detect traces of malicious online behavior, data breaches, unauthorized website access, malware infections, and infiltration attempts, and to reconstruct image files, documents, email attachments, and other content sent over the network [1][15] [16][19].

We recommend combining PCAP analysis and social network analysis. This combination will demonstrate advanced network forensic analysis by revealing communication patterns between specific nodes in social media interactions. Combining PCAP and SNA to analyze network forensic activity can be a powerful method for identifying and analyzing malicious activity on the network. The integration of these two technologies can take advantage of the detailed information provided by PCAP data and the broader network-level view provided by SNA. An example of how this combination could be used is to use PCAP data to identify specific patterns or anomalies in network traffic that are consistent with known malicious activity, such as: B. Botnet command and control traffic or data exfiltration. Once these patterns are identified, SNA can be used to identify nodes and edges in the network that match these patterns, which helps identify the source and destination of malicious activity and the relationships between these entities.

III. Combining PCAP Data and Social Network Analysis for Network Forensic Activity

This research uses an experimental method in a laboratory where the independent variable is the number of nodes in social media capturing, and the dependent variable is the result of network forensic analysis, shown in Fig. 1.

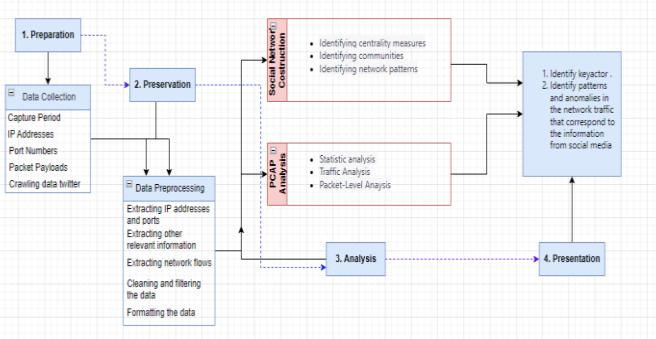


Fig. 1. Combining PCAP data and Social Network Analysis (SNA) for network forensic methodology.

A. Data Collection

The first step is to collect PCAP data from the network using an application such as Wireshark or tcpdump. This data includes information about network traffic that occurred during the collection period, including details such as source and destination IP addresses, port numbers, and packet payloads. During the collection process, the keyword "Manchester United" was used to crawl data from Twitter, and 1000 nodes were obtained. Twitter scraping was performed on January 4, 2023 using the netlytic application.

B. Data Preprocessing

This step is performed to filter out packets that are not relevant to the current investigation. For example, only save packets from suspicious IP addresses or use protocols considered important. Parsing, this step is performed to extract relevant information from the PCAP packet. Information that can be extracted includes source and destination IP addresses, protocols used, timestamps, and payload data. Anonymization, this step is performed to remove information that could be used to identify individuals participating in the communication. Information that can be removed includes IP addresses, MAC addresses, and personal information contained in payload data. Normalization, this step is performed to convert the data extracted from the PCAP packets into a format that is easier to use for analysis. For example, converting timestamps to a more readable format or converting used logs to a simpler format. Aggregation, this step is performed to combine data from multiple PCAP packets into a larger unit. For example, combining multiple packets from the same IP address into a larger unit. Enrichment, this step is performed to add additional information to the data extracted from PCAP packets. Additional information can be in the form of IP geolocation information, WHOIS information, or IP reputation information.

C. PCAP Analysis

Traffic Analysis, this step involves analyzing data to identify patterns and anomalies in network traffic. This may include identifying unusual traffic destinations, unusual traffic patterns, or specific protocols used. Packet level analysis, this step involves examining individual packets in the PCAP data such as: B. Source and destination IP addresses, source and destination ports, and packet payload. This can be used to identify specific keywords, extract files, or extract other information from the payload. Statistical Analysis, this is the process of analyzing and interpreting the data contained in a PCAP file using statistical methods. This can include identifying patterns, trends, and anomalies in network traffic, as well as estimating various metrics such as traffic volume, packet size, and packet arrival time.

D. Social Network Construction

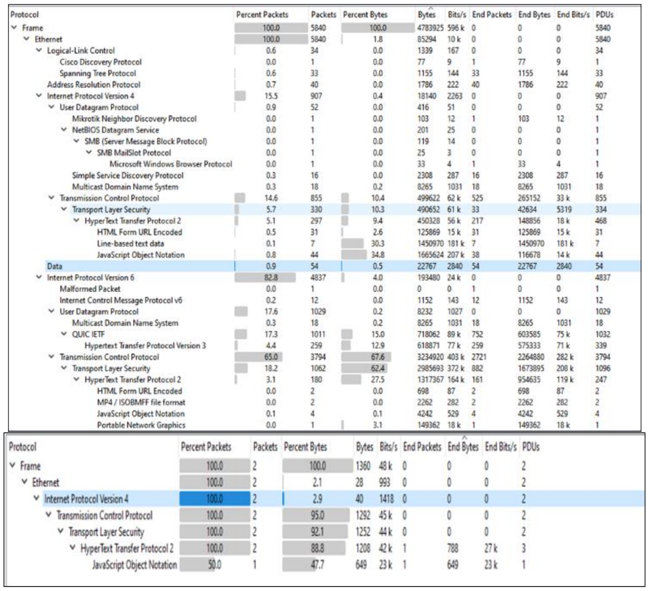
Social network construction creates representations of relationships between individuals or entities in a social network. This may involve identifying relationships between individuals, such as B. friendships, family ties, or professional relationships, and the strength of these relationships, such as B. frequency of personal interaction or communication. These centrality measures can be applied to social networks created from PCAP data and help identify key IP addresses or ports that may be important for understanding communication patterns and how information or malware spreads in the network. Degree centrality, this measure is based on the number of edges (connections) a node has. Nodes with high centrality are those that have many connections in the network. Betweenness centrality, this measure is based on the number of shortest paths through a node. Nodes with high betweenness centrality are those located on many shortest paths, considered as bridges or gatekeepers between different communities. Closeness centrality this measure is based on the average distance of a node from all other nodes in the network. Nodes with high proximity centrality are those that are close to many other nodes in the network. Community discovery is an important task in social network analysis that aims to identify groups of nodes (communities) that are more connected to each other than to the rest of the network. There are various methods for discovering communities on the web. Network pattern recognition is the process of identifying structural patterns in a network, which can provide insight into network organization and function. There are several methods for identifying patterns in networks; some of the most popular are subgraph counts and clustering coefficients.

IV. Experiment

Conversation statistics IP source 192.168.1.14 to 104.211.42.0 show the wireshark session statistics to see which devices are communicating with each other. This data also counts the traffic exchanged between these devices. It helps to understand the communication patterns in the network. Protocol Distribution Use Wireshark Hierarchical Protocol Statistics to see which protocols are used on the network and how much traffic they generate. It shows the distribution of logs and the number of packages present. Fig. 2 shows the number of packets, namely 5840, distributed among the different protocols and conversation statistics.

A. Authors and Affiliations

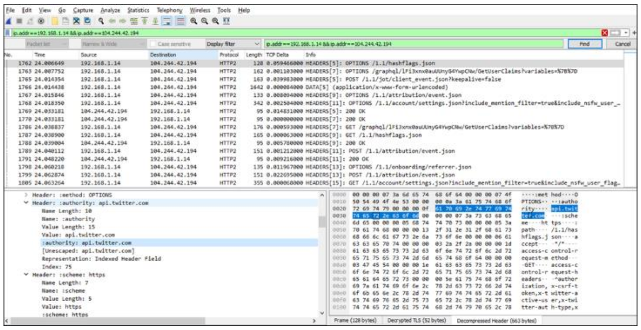
Fig. 3 shows the distribution of content packets in Wireshark's packet details panel to examine the contents of individual packets and see their structure. It can help understand the data exchanged in the network and identify potential problems. Fig. 4 shows that round trip time (RTT) is a measure of how long it takes to send a packet of data from one device to another and receive a response. RTT can use this value to measure network latency and identify potential delays. = 55 milliseconds.



Protocol	Percent Packets	Packets	Percent Bytes	Bytes	Bits/s	End Packets	End Bytes	End Bits/s	PDUs
˅ Frame	100.0	5840	100.0	4783925	596 k	0	0	0	5840
˅ Ethernet	100.0	5840	1.8	85294	10 k	0	0	0	5840
˅ Logical-Link Control	0.6	34	0.0	1339	167	0	0	0	34
Cisco Discovery Protocol	0.0	1	0.0	77	9	1	77	9	1
Spanning Tree Protocol	0.6	33	0.0	1155	144	33	1155	144	33
Address Resolution Protocol	0.7	40	0.0	1786	222	40	1786	222	40
˅ Internet Protocol Version 4	15.5	907	0.4	18140	2263	0	0	0	907
˅ User Datagram Protocol	0.9	52	0.0	416	51	0	0	0	52
Mikrotik Neighbor Discovery Protocol	0.0	1	0.0	103	12	1	103	12	1
˅ NetBIOS Datagram Service	0.0	1	0.0	201	25	0	0	0	1
˅ SMB (Server Message Block Protocol)	0.0	1	0.0	119	14	0	0	0	1
˅ SMB MailSlot Protocol	0.0	1	0.0	25	3	0	0	0	1
Microsoft Windows Browser Protocol	0.0	1	0.0	33	4	1	33	4	1
Simple Service Discovery Protocol	0.3	16	0.0	2308	287	16	2308	287	16
Multicast Domain Name System	0.3	18	0.2	8265	1031	18	8265	1031	18
˅ Transmission Control Protocol	14.6	855	10.4	499622	62 k	525	265152	33 k	855
˅ Transport Layer Security	5.7	330	10.3	490652	61 k	33	42634	5319	334
˅ HyperText Transfer Protocol 2	5.1	297	9.4	450328	56 k	217	148856	18 k	468
HTML Form URL Encoded	0.5	31	2.6	125869	15 k	31	125869	15 k	31
Line-based text data	0.1	7	30.3	1450970	181 k	7	1450970	181 k	7
JavaScript Object Notation	0.8	44	34.8	1665624	207 k	38	116678	14 k	44
Data	0.9	54	0.5	22767	2840	54	22767	2840	54
˅ Internet Protocol Version 6	82.8	4837	4.0	193480	24 k	0	0	0	4837
Malformed Packet	0.0	1	0.0	0	0	1	0	0	1
Internet Control Message Protocol v6	0.2	12	0.0	1152	143	12	1152	143	12
˅ User Datagram Protocol	17.6	1029	0.2	8232	1027	0	0	0	1029
Multicast Domain Name System	0.3	18	0.2	8265	1031	18	8265	1031	18
˅ QUIC IETF	17.3	1011	15.0	718062	89 k	752	603585	75 k	1032
Hypertext Transfer Protocol Version 3	4.4	259	12.9	618871	77 k	259	575333	71 k	339
˅ Transmission Control Protocol	65.0	3794	67.6	3234920	403 k	2721	2264880	282 k	3794
˅ Transport Layer Security	18.2	1062	62.4	2985693	372 k	882	1673895	208 k	1096
˅ HyperText Transfer Protocol 2	3.1	180	27.5	1317367	164 k	161	954635	119 k	247
HTML Form URL Encoded	0.0	2	0.0	698	87	2	698	87	2
MP4 / ISOBMFF file format	0.0	2	0.0	2262	282	2	2262	282	2
JavaScript Object Notation	0.1	4	0.1	4242	529	4	4242	529	4
Portable Network Graphics	0.0	1	3.1	149362	18 k	1	149362	18 k	1

Protocol	Percent Packets	Packets	Percent Bytes	Bytes	Bits/s	End Packets	End Bytes	End Bits/s	PDUs
˅ Frame	100.0	2	100.0	1360	48 k	0	0	0	2
˅ Ethernet	100.0	2	2.1	28	993	0	0	0	2
˅ Internet Protocol Version 4	100.0	2	2.9	40	1418	0	0	0	2
˅ Transmission Control Protocol	100.0	2	95.0	1292	45 k	0	0	0	2
˅ Transport Layer Security	100.0	2	92.1	1252	44 k	0	0	0	2
˅ HyperText Transfer Protocol 2	100.0	2	88.8	1208	42 k	1	788	27 k	3
JavaScript Object Notation	50.0	1	47.7	649	23 k	1	649	23 k	1

Fig. 2. Conversation statistic and protocol distribution.


File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
ip.addr==192.168.1.14 && ip.addr==104.244.42.194
Packet list
Narrow & Wide
Case sensitive
Display filter
ip.addr==192.168.1.14 && ip.addr==104.244.42.194
Find
Cancel

```
No.  Time       Source          Destination     Protocol Length TCP Delta   Info
1762 24.006649  192.168.1.14    104.244.42.194  HTTP2     128 0.059466000 HEADERS[5]: OPTIONS /1.1/hashflags.json
1763 24.007752  192.168.1.14    104.244.42.194  HTTP2     162 0.001103000 HEADERS[7]: OPTIONS /graphql/lFi3xnx0auUUnyG4YwpCNw/GetUserClaims?variables=%7B%7D
1765 24.014354  192.168.1.14    104.244.42.194  HTTP2     163 0.039983000 HEADERS[5]: POST /1.1/jot/client_event.json?keepalive=false
1766 24.014438  192.168.1.14    104.244.42.194  HTTP2    1642 0.000084000 DATA[5] (application/x-www-form-urlencoded)
1767 24.015846  192.168.1.14    104.244.42.194  HTTP2     133 0.000094000 HEADERS[9]: OPTIONS /1.1/attribution/event.json
1768 24.018350  192.168.1.14    104.244.42.194  HTTP2     342 0.002504000 HEADERS[11]: OPTIONS /1.1/account/settings.json?include_mention_filter=true&include_nsfw_user_…
1769 24.033181  104.244.42.194  192.168.1.14    HTTP2      95 0.014831000 HEADERS[5]: 200 OK
1770 24.033181  104.244.42.194  192.168.1.14    HTTP2      95 0.000000000 HEADERS[7]: 200 OK
1786 24.038837  192.168.1.14    104.244.42.194  HTTP2     176 0.000593000 HEADERS[7]: GET /graphql/lFi3xnx0auUUnyG4YwpCNw/GetUserClaims?variables=%7B%7D
1787 24.038900  192.168.1.14    104.244.42.194  HTTP2     165 0.000063000 HEADERS[9]: GET /1.1/hashflags.json
1788 24.039004  104.244.42.194  192.168.1.14    HTTP2      95 0.005780000 HEADERS[9]: 200 OK
1789 24.040112  192.168.1.14    104.244.42.194  HTTP2     151 0.001212000 HEADERS[11]: POST /1.1/attribution/event.json
1791 24.048220  104.244.42.194  192.168.1.14    HTTP2      95 0.009216000 HEADERS[11]: 200 OK
1798 24.060218  192.168.1.14    104.244.42.194  HTTP2     135 0.011967000 HEADERS[13]: OPTIONS /1.1/onboarding/referrer.json
1799 24.062874  192.168.1.14    104.244.42.194  HTTP2     151 0.022695000 HEADERS[13]: POST /1.1/attribution/event.json
1805 24.063264  192.168.1.14    104.244.42.194  HTTP2     355 0.000068000 HEADERS[15]: GET /1.1/account/settings.json?include_mention_filter=true&include_nsfw_user_flag…
```

```
> Header: :method: OPTIONS
v Header: :authority: api.twitter.com
    Name Length: 10
    Name: :authority
    Value Length: 15
    Value: api.twitter.com
    :authority: api.twitter.com
    [Unescaped: api.twitter.com]
    Representation: Indexed Header Field
    Index: 75
v Header: :scheme: https
    Name Length: 7
    Name: :scheme
    Value Length: 5
    Value: https
    :scheme: https
```

```
0000  00 00 00 07 3a 6d 65 74  68 6f 64 00 00 00 07 4f   ····:met hod····O
0010  50 54 49 4f 4e 53 00 00  00 0a 3a 61 75 74 68 6f   PTIONS·· ··:autho
0020  72 69 74 79 00 00 00 0f  61 70 69 2e 74 77 69 74   rity···· api.twit
0030  74 65 72 2e 63 6f 6d 00  00 00 07 3a 73 63 68 65   ter.com· ···:sche
0040  6d 65 00 00 00 05 68 74  74 70 73 00 00 00 05 3a   me····ht tps····:
0050  70 61 74 68 00 00 00 13  2f 31 2e 31 2f 68 61 73   path···· /1.1/has
0060  68 66 6c 61 67 73 2e 6a  73 6f 6e 00 00 00 06 61   hflags.j son····a
0070  63 63 65 70 74 00 00 00  03 2a 2f 2a 00 00 00 1d   ccept··· ·*/*····
0080  61 63 63 65 73 73 2d 63  6f 6e 74 72 6f 6c 2d 72   access-c ontrol-r
0090  65 71 75 65 73 74 2d 6d  65 74 68 6f 64 00 00 00   equest-m ethod···
00a0  03 47 45 54 00 00 00 1e  61 63 63 65 73 73 2d 63   ·GET···· access-c
00b0  6f 6e 74 72 6f 6c 2d 72  65 71 75 65 73 74 2d 68   ontrol-r equest-h
00c0  65 61 64 65 72 73 00 00  00 5e 61 75 74 68 6f 72   eaders·· ·^author
00d0  69 7a 61 74 69 6f 6e 2c  78 2d 63 73 72 66 2d 74   ization, x-csrf-t
00e0  6f 6b 65 6e 2c 78 2d 74  77 69 74 74 65 72 2d 61   oken,x-t witter-a
00f0  63 74 69 76 65 2d 75 73  65 72 2c 78 2d 74 77 69   ctive-us er,x-twi
0100  74 74 65 72 2d 61 75 74  68 2d 74 79 70 65 2c 78   tter-aut h-type,x
```

Fig. 3. Packet content on Wireshark app.

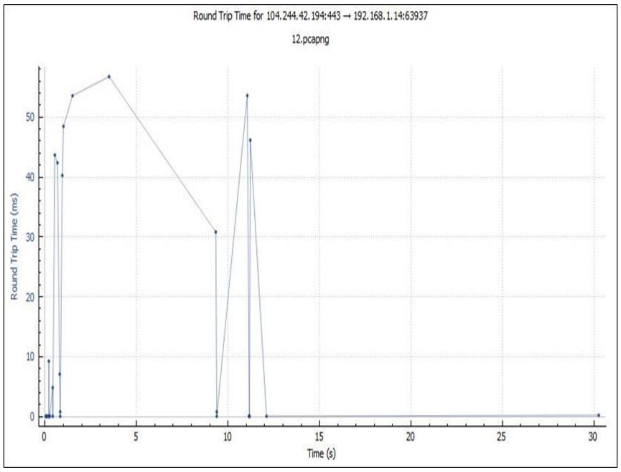
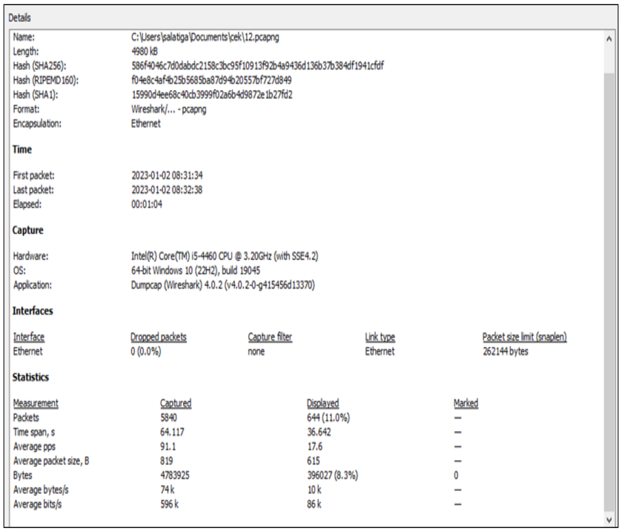


Fig. 4. Twitter server communication round trip time value.



Details

Name:	C:\Users\salatiga\Documents\cek\12.pcapng
Length:	4980 kB
Hash (SHA256):	586f4046c7d0dabdc2158c3bc95f10913f92b4a9436d136b37b384df1941cfdf
Hash (RIPEMD160):	f04e8c4af4b25b5685ba87d94b20557bf727d849
Hash (SHA1):	15990d4ee68c40cb3999f02a6b4d9872e1b27fd2
Format:	Wireshark/... - pcapng
Encapsulation:	Ethernet

Time

First packet:	2023-01-02 08:31:34
Last packet:	2023-01-02 08:32:38
Elapsed:	00:01:04

Capture

Hardware:	Intel(R) Core(TM) i5-4460 CPU @ 3.20GHz (with SSE4.2)
OS:	64-bit Windows 10 (22H2), build 19045
Application:	Dumpcap (Wireshark) 4.0.2 (v4.0.2-0-g415456d13370)

Interfaces

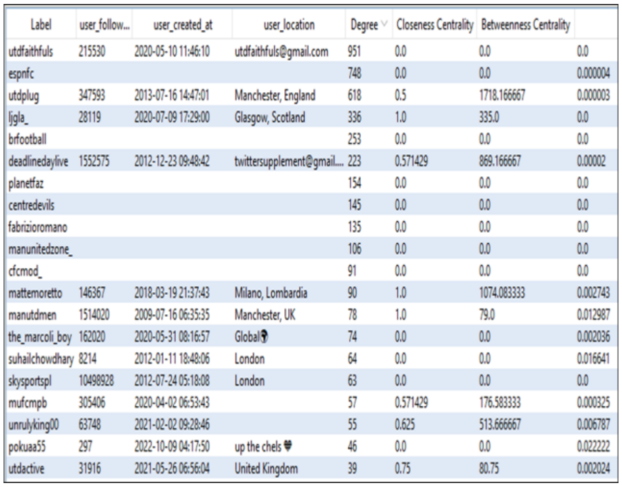
Interface	Dropped packets	Capture filter	Link type	Packet size limit (snaplen)
Ethernet	0 (0.0%)	none	Ethernet	262144 bytes

Statistics

Measurement	Captured	Displayed	Marked
Packets	5840	644 (11.0%)	—
Time span, s	64.117	36.642	—
Average pps	91.1	17.6	—
Average packet size, B	819	615	—
Bytes	4783925	396027 (8.3%)	0
Average bytes/s	74 k	10 k	—
Average bits/s	596 k	86 k	—

Fig. 5. The results of data acquisition and PCAP file integrity check.

Fig. 5 shows data collection verification and file integrity verification in network forensics refers to the process of collecting and verifying the authenticity and integrity of digital evidence from the network. This process is critical to ensuring that evidence collected is accurate and available for investigation or trial. Data collection involves capturing and copying data from the network, while file integrity checking involves checking hashes or digital signatures of captured data to ensure it has not been tampered with. These steps are important to maintain the chain of custody and maintain the authenticity of the evidence. Digital forensic analysis using SNA includes identifying primary and secondary actors. This assessment makes the investigative process more focused on specific actors.



Label	user_follow...	user_created_at	user_location	Degree	Closeness Centrality	Betweenness Centrality	
utdfaithfuls	215530	2020-05-10 11:46:10	utdfaithfuls@gmail.com	951	0.0	0.0	0.0
espnfc				748	0.0	0.0	0.000004
utdplug	347593	2013-07-16 14:47:01	Manchester, England	618	0.5	1718.166667	0.000003
ljgla_	28119	2020-07-09 17:29:00	Glasgow, Scotland	336	1.0	335.0	0.0
brfootball				253	0.0	0.0	0.0
deadlinedaylive	1552575	2012-12-23 09:48:42	twittersupplement@gmail....	223	0.571429	869.166667	0.00002
planetfaz				154	0.0	0.0	0.0
centredevils				145	0.0	0.0	0.0
fabrizioromano				135	0.0	0.0	0.0
manunitedzone_				106	0.0	0.0	0.0
cfcmod_				91	0.0	0.0	0.0
mattemoretto	146367	2018-03-19 21:37:43	Milano, Lombardia	90	1.0	1074.083333	0.002743
manutdmen	1514020	2009-07-16 06:35:35	Manchester, UK	78	1.0	79.0	0.012987
the_marcoli_boy	162020	2020-05-31 08:16:57	Global🌍	74	0.0	0.0	0.002036
suhailchowdhary	8214	2012-01-11 18:48:06	London	64	0.0	0.0	0.016641
skysportspl	10498928	2012-07-24 05:18:08	London	63	0.0	0.0	0.0
mufcmpb	305406	2020-04-02 06:53:43		57	0.571429	176.583333	0.000325
unrulyking00	63748	2021-02-02 09:28:46		55	0.625	513.666667	0.006787
pokuaa55	297	2022-10-09 04:17:50	up the chels 💙	46	0.0	0.0	0.022222
utdactive	31916	2021-05-26 06:56:04	United Kingdom	39	0.75	80.75	0.002024

Fig. 6. Degree, closeness and betweeness Centrality.

Fig. 6 generates degree centrality as the number of connections or edges a node has to other nodes in the network. Nodes with high centrality have many connections to other nodes. In this study, 10,000 nodes were obtained using the manchester united keyword. The main actor has 951 degrees on the node labeled: utdfaithfuls. Perform key player identification to determine the degree of influence a participant has in the network. It is important to identify who are the actors who play the most important role in the communication model. This betweenness centrality takes into account the number of shortest paths between other nodes passing through a given node. High centrality nodes are "bridges" between other nodes in the network. The degree, proximity, and betweenness centrality metrics tables will show the following information for each node (or vertex) in the network Degree centrality, this is a measure of how many direct connections a node has to other nodes in the network. This is usually expressed as the number of edges intersecting the node. Closeness centrality, this is a measure of how close a node is to all other nodes in the network. This is usually expressed as the sum of the shortest distances between a node and all other nodes in the network. Betweenness centrality, this is a measure of how often a node acts as a bridge between other nodes in the network. This is usually expressed as the number of shortest paths between pairs of nodes passing through a given node.

For this experiment, the tag **utdfaithfuls** Beetweenness centrality: 1718 nodes with utdplug. Graphical visualization of the top three communication modes at the highest level is shown in Fig. 7. Graph visualization is a commonly used technique in social network analysis to show the relationships between individuals or entities in the network. It visualizes the network as a graph or graph, with nodes representing individuals or entities and edges representing the relationships between them. The size and color of nodes and the thickness and direction of edges can be used to represent attributes or measures, such as relationship strength or centrality measures. This helps visualize patterns, relationships, and communities in the network and makes it easier to understand the network structure and its properties.

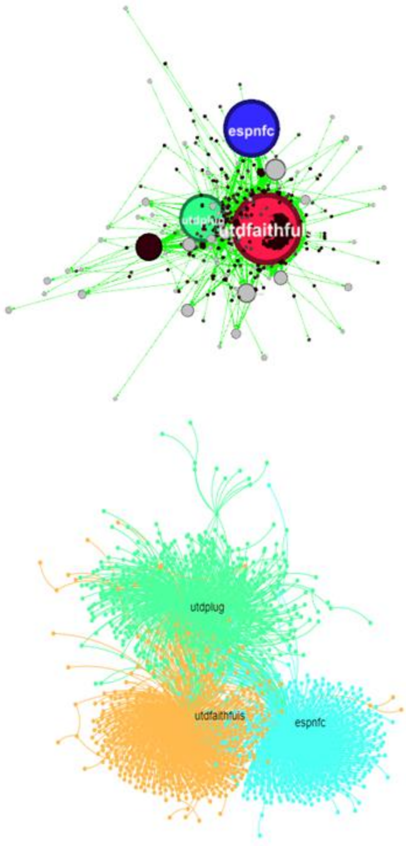


Fig. 7. Graph visualization of the top 3 pad communication patterns with the highest degree.

TABLE I. (A). RESULTS OF NETWORK FORENSIC ANALYSIS OF PCAP. AND SNA COMBINATION

Source IP	Destination IP	Packet	A >B	B>A	Date/Time	Dura Tion (s)	Name server
192.168.1.16	224.0.0.251	18	18	0	2023-01-02T08:31:34.517	4,74	
192.168.1.5	239.255.255.250	12	12	0	2023-01-02T08:31:37.513	60,00	
192.168.1.14	20.198.119.143	3	2	1	2023-01-02T08:31:40.793	0,12	
192.168.1.10	255.255.255.255	1	1	0	2023-01-02T08:31:54.895	-	
192.168.1.14	104.244.42.129	129	35	94	2023-01-02T08:31:56.481	7,37	
192.168.1.14	152.199.43.83	26	12	14	2023-01-02T08:31:57.878	0,69	
192.168.1.14	104.244.42.194	644	255	389	2023-01-02T08:31:58.043	36,60	api.twitter.com
192.168.1.14	104.244.42.5	11	5	6	2023-01-02T08:31:58.107	0,07	api.twitter.com
192.168.1.14	104.244.43.131	42	19	23	2023-01-02T08:31:59.670	11,45	api.twitter.com
192.168.1.14	192.168.1.255	1	1	0	2023-01-02T08:32:24.230	-	
192.168.1.14	239.255.255.250	4	4	0	2023-01-02T08:32:32.322	3,03	
192.168.1.13	192.168.1.14	16	16	0	2023-01-02T08:32:32.326	3,23	

(B). RESULTS OF NETWORK FORENSIC ANALYSIS OF SNA

Social Network Analysis	1.	Network Size	10000 node
	2.	Centrality / Keyactor	UtdFaithfuls
		a. User created at	10/05/20 11.46
		b. Follower	215530
		c. Pub date	04/01/23 16.54
	3	Clustering coefficient	0.005
	4	Density	0.001
	5	Network Diameter	71
	6	Cluster	3

As Table I (A), (B) in network forensics, combining PCAP analysis with social network analysis can have a big impact on research in a number of ways:

1) Better suspect identification. Based on their connections to and interactions with other network members, prospective suspects in a network can be found via social network analysis. This can give a more complete picture of the network and increase the precision with which suspects are identified.

2) An improved comprehension of network behavior can be obtained by combining social network analysis and PCAP analysis, which can give a more in-depth understanding of communication patterns and network behavior. Investigators can use this to spot unusual or suspect activities on the network, such as secret or encrypted communication.

3) Enhanced incident response. By integrating PCAP analysis with social network analysis, detectives can more precisely pinpoint the origin and consequences of a crime.

The outcomes of benchmarking the integration of PCAP analysis with social network analysis in network forensics can give important insights into the most efficient ways to carry out investigations, uncover potential security concerns, and improve network security generally.

V. Conclusion

Network forensics using PCAP analysis combined with social network analysis shows more comprehensive results. PCAP analysis is used to analyze network traffic, conversation statistics, protocol distribution, packet content and round-trip time. Social network analysis maps communication patterns between nodes and identifies the most influential key players in the network. PCAP analysis efficiently captures and analyzes network packets, while SNA provides insight into the relationships and communication patterns between devices on the network. The availability of data sources is also a factor to consider when deciding to combine PCAP and SNA. The combination of PCAP and SNA can provide a more complete view of the network when limited log data is available. When the nature of the threat includes both technical and social aspects, a combination of these two approaches may be required. PCAP and SNA can provide a more comprehensive view of the network and improve the accuracy of forensic analysis. This is because SNA can provide context about captured packets and help identify relationships and patterns that might not be apparent from packet analysis alone. Potential future suggestions for improving forensic cyber analysis include artificial intelligence and machine learning, which includes AI and machine learning algorithms that can help automate the data analysis process, making it more efficient and accurate.

References

[1] L. F. Sikos, "Packet analysis for network forensics: A comprehensive survey," Forensic Sci. Int. Digit. Investig., vol. 32, p. 200892, Mar. 2020, doi: 10.1016/J.FSIDI.2019.200892.

[2] N. Koroniotis, "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset," Futur. Gener. Comput. Syst., vol. 100, pp. 779–796, 2019, doi: 10.1016/j.future.2019.05.041.

[3] J. Lian, "Implementation of computer network user behavior forensic analysis system based on speech data system log," Int. J. Speech Technol., vol. 23, no. 3, 2020, doi: 10.1007/s10772-020-09747-2.

[4] H. Munkhondya, "Digital forensic readiness approach for potential evidence preservation in software-defined networks," 14th Int. Conf. Cyber Warf. Secur. ICCWS 2019, pp. 268–276, 2019.

[5] B. C. M. Cappers, "Eventpad: Rapid Malware Analysis and Reverse Engineering using Visual Analytics," 2018 IEEE Symposium on Visualization for Cyber Security, VizSec 2018. 2019. doi: 10.1109/VIZSEC.2018.8709230.

[6] I. Yaqoob, "Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges," Futur. Gener. Comput. Syst., vol. 92, pp. 265–275, 2019, doi: 10.1016/j.future.2018.09.058.

[7] A. Ulmer, "NetCapVis: Web-based progressive visual analytics for network packet captures," 2019 IEEE Symposium on Visualization for Cyber Security, VizSec 2019. 2019. doi: 10.1109/VizSec48167.2019.9161633.

[8] C. Yin, H. Wang, and J. Wang, "Network data stream classification by deep packet inspection and machine learning," in Lecture Notes in Electrical Engineering, 2019, vol. 518. doi: 10.1007/978-981-13-1328-8_31.

[9] B. Indira, K. Valarmathi, and D. Devaraj, "An approach to enhance packet classification performance of software-defined network using deep learning," Soft Comput., vol. 23, no. 18, 2019, doi: 10.1007/s00500-019-03975-8.

[10] J. Yoon and M. DeBiase, "Real-time analysis of big network packet streams by learning the likelihood of trusted sequences," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 10968 LNCS. doi: 10.1007/978-3-319-94301-5_4.

[11] R. K. Shrivastava, "Attack detection and forensics using honeypot in IoT environment," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, vol. 11319, pp. 402–409. doi: 10.1007/978-3-030-05366-6_33.

[12] K. Chakraborty, S. Bhattacharyya, and R. Bag, "A Survey of Sentiment Analysis from Social Media Data," IEEE Trans. Comput. Soc. Syst., vol. 7, no. 2, pp. 450–464, 2020, doi: 10.1109/TCSS.2019.2956957.

[13] B. Liu, "Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination," Eur. J. Oper. Res., vol. 275, no. 2, pp. 737–754, 2019, doi: 10.1016/j.ejor.2018.11.075.

[14] F. L. Aryeh, "Graphical analysis of captured network packets for detection of suspicious network nodes," 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment, Cyber SA 2020. 2020. doi: 10.1109/CyberSA49311.2020.9139672.

[15] S. M. Farjad, "Cluster Analysis and Statistical Modeling: A Unified Approach for Packet Inspection," 1st Annual International Conference on Cyber Warfare and Security, ICCWS 2020 - Proceedings. 2020. doi: 10.1109/ICCWS48432.2020.9292396.

[16] M. Marchetti, "READ: Reverse engineering of automotive data frames," IEEE Trans. Inf. Forensics Secur., vol. 14, no. 4, pp. 1083–1097, 2019, doi: 10.1109/TIFS.2018.2870826.

[17] S. M. Hosseini, "Digesting Network Traffic for Forensic Investigation Using Digital Signal Processing Techniques," IEEE Trans. Inf. Forensics Secur., vol. 14, no. 12, pp. 3312–3321, 2019, doi: 10.1109/TIFS.2019.2915190.

[18] P. Białczak, "Hfinger: Malware http request fingerprinting," Entropy, vol. 23, no. 5, 2021, doi: 10.3390/e23050507.

[19] S. Ali, "Effective Multitask Deep Learning for IoT Malware Detection and Identification Using Behavioral Traffic Analysis," IEEE Trans. Netw. Serv. Manag., 2022, doi: 10.1109/TNSM.2022.3200741.

[20] S. Sen Zhang, X. Liang, YD Wei, and X. Zhang, "On Structural Features, User Social Behavior, and Kinship Discrimination in Communication Social Networks," IEEE Trans. Comput. soc. syst., vol. 7, no. 2, pp. 425–436, 2020, doi:10.109/TCSS.2019.2962231.

[21] A. Matakos, C. Aslay, E. Galbrun, and A. Gionis, "Maximizing the Diversity of Exposure in a Social Network," IEEE Trans. knowl. Data

Eng., vol. 34, no. 9, pp. 4357–4370, 2022, doi:10.109/TKDE.2020.3038711.

[22] M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. Ver Steeg, and A. Galstyan, “Identifying and Analyzing Cryptocurrency Manipulations in Social Media,” IEEE Trans. Comput. soc. syst., vol. 8, no. 3, pp. 607–617, 2021, doi:101109/TCSS.2021.3059286.

[23] D. Vimalajeewa, S. Balasubramaniam, B. O'Brien, C. Kulatunga, and DP Berry, “Leveraging Social Network Analysis for Characterizing Cohesion of Human-Managed Animals,” IEEE Trans. Comput. soc. syst., vol. 6, no. 2, pp. 323–337, 2019, doi:10.109/TCSS.2019.2902456.

[24] AA Al-Shargabi and A. Selmi, “Social Network Analysis and Visualization of Arabic Tweets During the COVID-19 Pandemic,” IEEE Access, vol. 9, pp. 90616–90630, 2021, doi:101109/access.2 021.3091537.