2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS) | 978-1-6654-4399-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/NTMS49979.2021.9432677

WhatsApp Network Forensics: Discovering the IP Addresses of Suspects

Waqas Ahmed
National Center for Cyber Security
Air University, Islamabad, Pakistan
waqaskhattak99@gmail.com

Faisal Shahzad
National Center for Cyber Security
Air University, Islamabad, Pakistan
faisal.rwp@gmail.com

Abdul Rehman Javed
Department of Cyber Security
Air University, Islamabad, Pakistan
abdulrehman.cs@au.edu.pk

Farkhund Iqbal
College of Technological Innovation
Zayed University, United Arab Emirates
farkhund.iqbal@zu.ac.ae

Liaqat Ali
Department of Cyber Security
Air University, Islamabad, Pakistan
liaqat.ali@mail.au.edu.pk

***Abstract*—Call record analysis is the most critical task for the Law Enforcement Agencies (LEAs) in a cyber-investigation process. It provides valuable information in the investigation, such as time and date and the duration of incoming and outgoing calls. The technological advancement of smartphones and the versatility of Instant Messaging (IM) applications provide multiple communication channels to cybercriminals for communication, making it difficult for the LEAs to monitor/investigate using traditional forensics tools and techniques. The most challenging part is to retrieve specific information from the network traffic of a particular IM Application such as WhatsApp. This research article's primary purpose is to find the IP address of the cybercriminal using WhatsApp through existing sniffing techniques and tools. A method called rule-based extraction for sniffing packets is proposed for extracting the most relevant data from the network traffic. The results support LEAs to identify the cybercriminals' specific traffic and help in analyzing and comparing the mobile phone data with the network traffic.**

***Index Terms*—WhatsApp, Forensics, Suspects, Cybercriminals, Wireshark**

I. Introduction

The different applications of Information Technology (IT) have now become an essential part of our daily life routine [1], [2]. This involvement of IT can be easily found in both work and leisure times of our life. Checking email accounts [2], visit favorite social media applications and websites [3], or an IT-based technology for professional duties purposes, it is impossible to imagine a day without these and so many other activities like this. The determining needs for smartphones and devices and their applications have encouraged the fast development of portable technologies that accurately fit modern life demands. Such portable devices like smartphones help us to accomplish our daily computing tasks. These smart devices are considered more suitable to hold and carry, keeping in view their compact size, and ever-increasing functions [4]. The concept of smart devices communication applications like WhatsApp has become more widely adopted due to the point that they can easily be operated on smartphones as compared to other Personal Computers (PCs) or notebooks. The different types of service provision backends, from mobile telecommunication operators to the Internet, also made it attractive to accomplish routine communication tasks, which were difficult to do otherwise using smartphone-based communication applications. These include sending pictures, making video calls, sharing documents, sharing location, and not but least, making audio calls and messaging.

The ubiquity of the IM applications on different android phones helps different types of criminals communicate through available applications, making them difficult to track using traditional investigation techniques. These days, criminals use IM applications in traditional voice phones, which protect them from LEAs. It is not easy to find a particular smartphone's identity without the support of different foreign authorities on the Internet, which required complete anonymity and privacy of the investigation process [5]. WhatsApp data and call analysis are critical criminal investigations approach for different LEAs. WhatsApp provides different types of important data, such as date, time, and length of incoming and outgoing calls received and sends messages, images, and videos to individuals and groups, highly relevant for examining [6].

Filter valuable information from the large dataset of a particular IM application like WhatsApp is the main difficult task. We required network traffic for call analysis. The captured network packets from the Internet contain different types of information regarding connected devices. Every application has its protocols, port numbers, and connection frequencies. Smartphones can establish connections through different network interfaces, like mobile data or WiFi. Retrieving the call details, logs of the connected smartphones to the network, and finding detailed user data from the dataset is a great challenge. If we find the IP addresses of cybercriminals from network traffic, we can quickly reveal the IP addresses' particular IP locations and other communication content.

WhatsApp is the only IM application that worked on almost all available Operating systems and devices, including smartphones, tablets, and PC. WhatsApp is a popular worldwide application not because of the low-cost subscription, but because of its different features, which help people send and receive messages, multimedia files, group chats, calls, etc. The number of users increased since Facebook purchased the application in 2014. Unfortunately, its additional comfort level, well-designed, and functional features pay the criminals' rough usage to connect more secretly and effectively. Fig. 1 describes the history of different WhatsApp features.

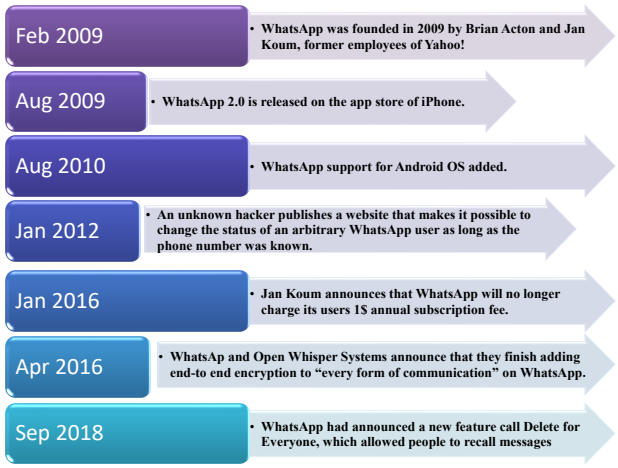


Fig. 1. History of WhatsApp Features

This research article extracted the WhatsApp artifacts from network communication using patterns matching to find criminal events more successfully. To find criminals' network communication details like IP addresses and other contents of the connection from different connected devices will help LEAs find criminals' misbehavior more efficiently. The paper's structure is organized according to Section 2 presents a review of the current research work. Section 3 presents the research methodology where the last section of the research paper presents the conclusion, discussion, and suggestions for future work.

II. Literature Review

Nowadays, many researchers and organizations are working on different mobile forensics [7]. Mobile forensic is a multidisciplinary research area that includes memory forensic and network forensic. Data acquisition, data presentation, and other operational processes are included in mobile forensic different digital data preservation. The main idea is based on what type of digital evidence is required, which preserves its significance and integrity, and from where and how it can be acquired. To conduct any professional mobile forensic, good knowledge of both hardware and software is required. Fig. 2 presents different WhatsApp Forensic Areas.

Fig. 2. WhatsApp Forensic Areas

Mobile devices are considered minicomputers; digital investigation and forensic are applied concepts from several mobile concepts. Mobile forensic investigators required knowledge of the mobile structure and different methods of acquiring, analyzing, and interpreting the residing data on mobile [8]. [9] compared different available WhatsApp tools included WhatsApp DB/Extractor and Belkasoft, to retrieve WhatsApp data from mobile memory. The comparison shows that WhatsApp DB/Extractor is faster than Belkasoft, but in terms of in-depth analysis and evidence accuracy, Belkasoft is more efficient than WhatsApp DB/Extractor. WhatsApp communication between two devices is end-to-end encrypted. The encryption techniques are in use for privacy-sensitive applications for data de-identification [10]. Authors in [6] described how we could decrypt the network traffic of WhatsApp users. A forensic investigator can obtain WhatsApp artifacts from network packets, including user's call records, IP addresses, WhatsApp numbers, and many more. In the research paper, the authors explain the tools and methods used for decrypting network traffic. The authors also analyzed and examined the handshaking between client and WhatsApp server, authentication process of clients, and analyzed the clients from relay servers. The paper also provides exciting findings such as metadata of call duration, time date stamps, and server IP address used during calls [11].

Through network traffic analysis of WhatsApp calls using different packet capturing tools, we can collect, analysis and examine WhatsApp packets. [12] presented a complete code that filters out WhatsApp network packets from non-WhatsApp traffic. The author tests the code on the network dataset; WhatsApp packets are successfully identified from the dataset. The information in the packets about users is identified. The filter WhatsApp packets are stored in separate files for further analysis and examination [12]. [13] presented a research work in which they understand audio messages in WhatsApp groups. They collected a large dataset of different

audios messages from different openly available groups and analyzed their content. Automatically find and analyze the audio messages in WhatsApp group chat, they analyzed different psychological topographies between the most shared audios, finding out that audio messages with a more sad connotation and closely related to time, work, family, and money are the most shared.

A. NIST Forensic Methodology

National Institute of Standard and Technology (NIST) defines a digital forensic guideline called SP 800-86. Digital evidence on mobile devices included messages, audio, videos, call records, images, and many more. Digital evidence is an important part of any digital investigation process since data pertinent to the crime may be store in some digital form. Critical parts of the investigation processes include extracting evidence from devices, storing the evidence, and analyzing the evidence efficiently and quickly. NIST standard SP 800-86 provides a four-phase digital forensic process [14]. Fig. 3 present the NIST digital forensic process phases.

Fig. 3. NIST Forensic Process

B. Packet Sniffing

A packet sniffer is a software tool used to intercept, log, and analyze additional network traffic and data types. These packet sniffing tools help in the packet identification, classification, and troubleshooting of network traffic by source, destination, and application type. Several tools are available; most of them are on Application Program Interfaces (API) known as pcap (for Unix OS) or libcap (for Windows OS) to sniff network traffic. Packet sniffers applications help us analyze the capture data, find the source of data, and prevent it from happening again in the future. Fig. 4 presents the different functionality of packet sniffer applications.

Fig. 4. Packet Analyzer Functions

Wireshark is the most well-known, usable and open-source network packet sniffer from the list. Wireshark is available in both command-line and GUI usefulness [15]. Also, supports both online and offline modes for network packet capturing. Fig. 5 presents features of the wireshark.

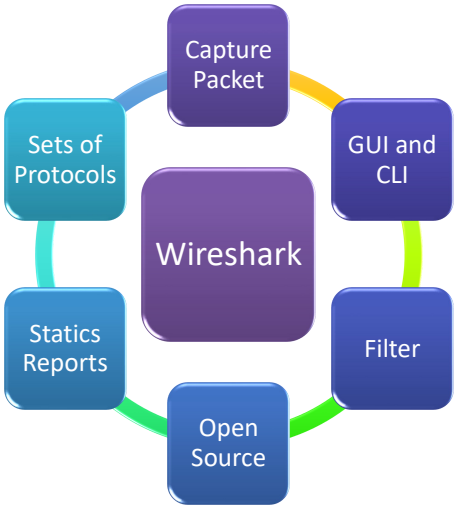


Fig. 5. Wireshark Features

III. WhatsApp Communication

Technically, people always think about how messages are transferred from their phones to the receiver phones without giving attention to what types of processes are performed on their phones' messages. When the sender clicks on the send button, the message is saved in the sender's phone, forwarded to the server, and received by the receiver, and save in the receiver's phone. Fig. 6 presented the communication process of WhatsApp.

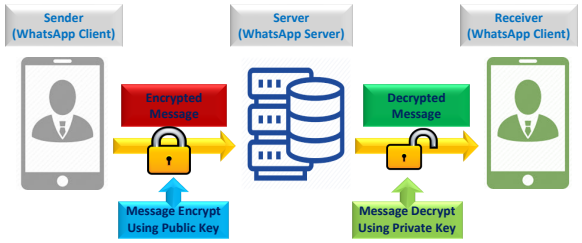


Fig. 6. WhatsApp Communication Process

Forensics investigators should know about the data storage place of a particular application in phone memory to help in data acquiring and analyzing. According to WhatsApp Terms and conditions [16], the sent messages are stored in WhatsApp servers for a limited period; the messages will be deleted after receiving the receiver. The send messages are stored on the WhatsApp server for 30 days, if not received by the receiver, not online [16]. From the forensics point of view, WhatsApp stored different artifacts with high evidentiary values and provided help to investigators in the evidence collection process, like log files and databases. Therefore, the investigator must focus on the criminal's phones to extract and analyze the WhatsApp data. The forensics tools should extract the WhatsApp data and help in analyzing all of these extracted artifacts. WhatsApp data are saved in different files and databases at different memory locations. Some of them are saved in phone internal memory, and others are saved in a Secure Digital Card (SDC) [17]. Some of the WhatsApp databases' essential components are shown in Fig. 7.

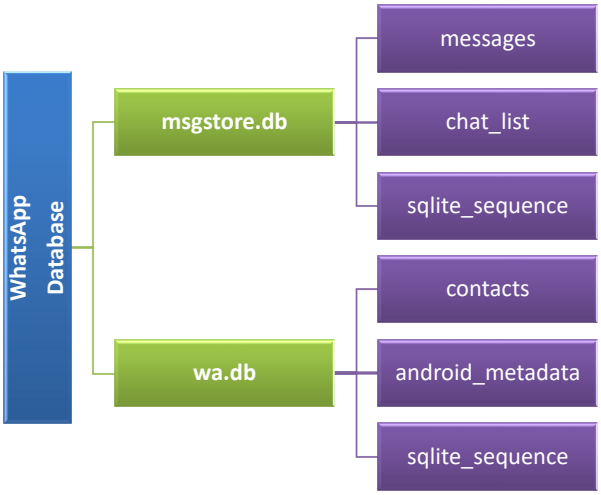


Fig. 7. WhatsApp Database Structure

IV. Research Framework

This research work-simulated and recognized the communications between two WhatsApp users (victim and suspect) and retrieved IP addresses of particular WhatsApp users from an extensive network data set of Local Area Network (LAN) LEAs to find suspects and their location effectively. Fig. 8 of the research paper presents four different parts of our proposed framework, which include capturing of WhatsApp packets from the network, data preparation, data pattern recognition, and evaluation results.

Fig. 8. Proposed Framework

1) **Data Collection**
The proposed framework's first step is data collection from the network; we used Wireshark (network packet sniffer) to capture the LAN packets between the victim machine and the suspect machine. One of law enforcement agencies' strategies is to set a central location for sniffing network packets for investigating criminal behavior. To capture the network packets between WhatsApp users, Wireshark is deployed according to the law enforcement procedures. WhatsApp does not allow users to make voice calls from computers, and we cannot set up our experiment for this. To captured network traffic through Wireshark, we installed Wireshark on our PC for the experiment. We convert our PC to a hotspot device for sharing the Internet to the near smartphones and used our PC to capture network packets using Wireshark.

2) **Data Preparation**
The second part of the proposed framework is to prepare the PCAP data captured through Wireshark and demonstrate the packets' header and payload. WhatsApp uses STUN protocol for voice calls; therefore, we used Wireshark filtering functionality to filter only STUN protocol packets from the captured network packets. Fig. 9 of the research paper shows the STUN protocol packets listed by time-stamp from smuggled PCAP network files.

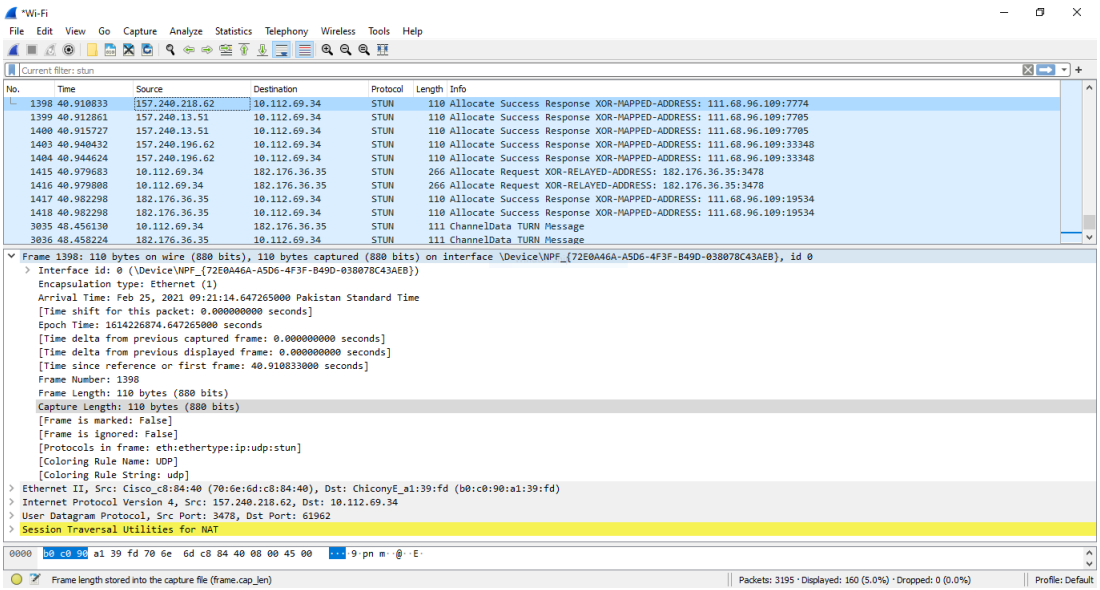


Fig. 9. STUN Protocol Packets

Besides the STUN protocol using the filtering technique, the IP addresses' Geo-Location plays an essential part in finding the physical space network locations. First available in Unix Operating System, the most common and useful method for finding the location and other important information related to a registered IP address is the Whois tool. By inserting the IP address into the open-source Whois tool, it will return different types of information, including the physical location of IP address, phone numbers, domain ownership, and much other important information. To view data from different sides and angles, we export the PCAP files to excel for the Pivot table.

3) **Pattern Recognition** The third step of the proposed framework is pattern recognition. In the pattern recognition step, we analyzed the captured packets generated by WhatsApp to find patterns for efficiently identifying the suspect's IP from the packets. There are 90 different attributes in the flat files, as shown in Fig. 10. After a more in-depth analysis of the captured packets and applying the Pivot table to view the captured network packets from all the angles, we have found some key packet attributes which helped us to create different network patterns for the LEAs, including flags, differentiated services code-point, and differentiated services field. Differentiated services fields provide a low latency rate during streaming and voice to critical network traffic and provide best-effort services to noncritical network services, such as file transferring and web-traffic. The value of the field enclosed in the flag is well-defined in different sections associated with the data structure, and the bit field value is normally contained with privileges or with properties.
4) **Evaluation Results**

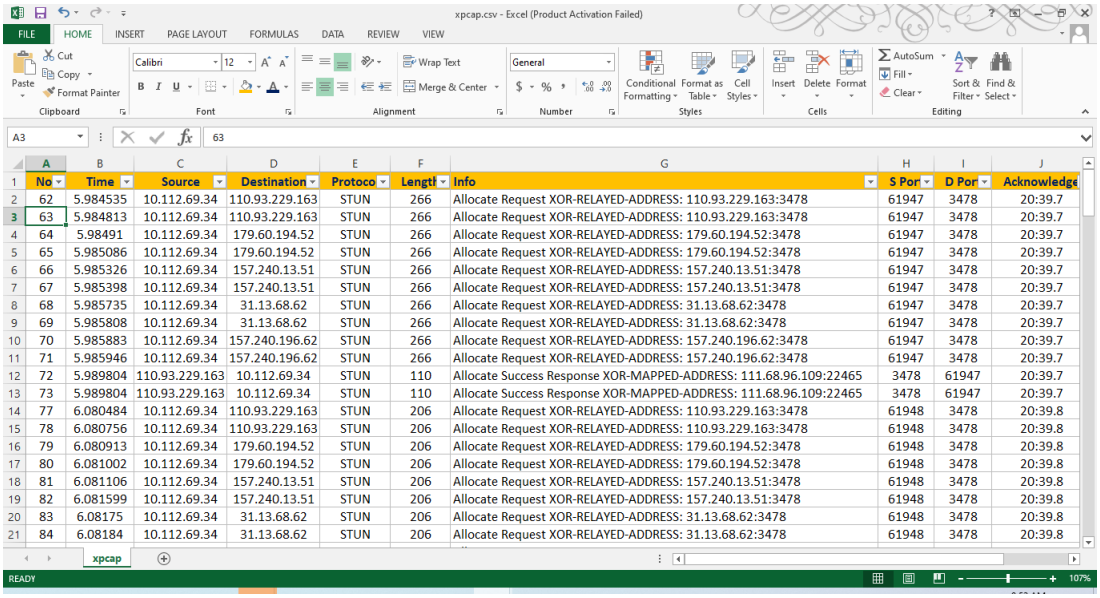
Different experiments are conducted to find the helpfulness of our proposed methodology for WhatsApp forensics. We captured network traffic for 44 seconds, and the captured packets contain WhatsApp communication between different users. Captured packets belong to the Local Area Network (LAN) only; to find a particular packet or IP address in the list is very difficult because of additional connections to different software and Internet Service Providers (ISP). Whois tool helped us in finding the Geo-Location of a particular IP address in the experiments.

For additional information, we import packet headers and detail payloads of captured network packets into the Pivot table to find specific WhatsApp communications patterns. With the help of frequency dissemination analysis methods, we found that the network packet's maximum fields contained some random numbers that cannot recognize the communication packets. Moreover, values of other types of fields such as Flags (F), Differentiated Service Field (DSF), and Differentiated Services Code-point (DSC) are fixed. For the recognition of the WhatsApp communications pattern, we choose the attributes which have fixed values. Table 1 presents the attribute of derived packets and their content.

TABLE I
THE ATTRIBUTE OF THE CONTENT

Packet Attributes	Content
Flags	0x00
DSF	0x38
DSC	Assured Forwarding 13

Table 2 presents the packet filtering rules, which help us smear different resultant principles for network packet examination. The main idea to generate the rules is to find WhatsApp communication more effectively for the



xpcap.csv - Excel (Product Activation Failed)

No	Time	Source	Destination	Protoco	Length	Info	S Port	D Port	Acknowledge
62	5.984535	10.112.69.34	110.93.229.163	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 110.93.229.163:3478	61947	3478	20:39.7
63	5.984813	10.112.69.34	110.93.229.163	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 110.93.229.163:3478	61947	3478	20:39.7
64	5.98491	10.112.69.34	179.60.194.52	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 179.60.194.52:3478	61947	3478	20:39.7
65	5.985086	10.112.69.34	179.60.194.52	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 179.60.194.52:3478	61947	3478	20:39.7
66	5.985326	10.112.69.34	157.240.13.51	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 157.240.13.51:3478	61947	3478	20:39.7
67	5.985398	10.112.69.34	157.240.13.51	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 157.240.13.51:3478	61947	3478	20:39.7
68	5.985735	10.112.69.34	31.13.68.62	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 31.13.68.62:3478	61947	3478	20:39.7
69	5.985808	10.112.69.34	31.13.68.62	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 31.13.68.62:3478	61947	3478	20:39.7
70	5.985883	10.112.69.34	157.240.196.62	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 157.240.196.62:3478	61947	3478	20:39.7
71	5.985946	10.112.69.34	157.240.196.62	STUN	266	Allocate Request XOR-RELAYED-ADDRESS: 157.240.196.62:3478	61947	3478	20:39.7
72	5.989804	110.93.229.163	10.112.69.34	STUN	110	Allocate Success Response XOR-MAPPED-ADDRESS: 111.68.96.109:22465	3478	61947	20:39.7
73	5.989804	110.93.229.163	10.112.69.34	STUN	110	Allocate Success Response XOR-MAPPED-ADDRESS: 111.68.96.109:22465	3478	61947	20:39.7
77	6.080484	10.112.69.34	110.93.229.163	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 110.93.229.163:3478	61948	3478	20:39.8
78	6.080756	10.112.69.34	110.93.229.163	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 110.93.229.163:3478	61948	3478	20:39.8
79	6.080913	10.112.69.34	179.60.194.52	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 179.60.194.52:3478	61948	3478	20:39.8
80	6.081002	10.112.69.34	179.60.194.52	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 179.60.194.52:3478	61948	3478	20:39.8
81	6.081106	10.112.69.34	157.240.13.51	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 157.240.13.51:3478	61948	3478	20:39.8
82	6.081599	10.112.69.34	157.240.13.51	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 157.240.13.51:3478	61948	3478	20:39.8
83	6.08175	10.112.69.34	31.13.68.62	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 31.13.68.62:3478	61948	3478	20:39.8
84	6.08184	10.112.69.34	31.13.68.62	STUN	206	Allocate Request XOR-RELAYED-ADDRESS: 31.13.68.62:3478	61948	3478	20:39.8

Fig. 10. A File for Pattern Recognition

Internet. The extra connections included manufacturers of network devices, software companies, and ISPs filtered after implementing the rules. We implemented the rules on the network captured packets that contained the WhatsApp communications to demonstrate the effect. The proposed rules' results successfully reveal the connection between two malicious user devices. The research paper conclusions help the LEAs in different cybercrime, where cybercriminals use WhatsApp application for communicating with each other.

TABLE II
PACKET FILTERING RULES

Rule 1	IF DSF = 0x38 Flag = 0x00 DSC=Assured Forwarding 13 THEN Sender IP = suspect
Rule 2	IF IM Protocol = STUN Packet Length = 266 Information = Binding Request THEN Receiver IP = suspect

V. CONCLUSION

Call record analysis provides valuable information in the investigation, such as time and date, the duration of incoming and outgoing call records. The most challenging part was how to retrieve specific information from the dataset of a particular IM Application. For the LEAs, the most critical part of the investigation strategy is analyzing such IM application data. This research work proposed a rule-based data extraction methodology to disclose the WhatsApp network communication between two users and remove disordered connections' influence. The experiments' results prove that the generated rules simplify a complex network to a simple network connection between two user's devices. Our proposed methodology can discover different network patterns generated by other IM applications, which will help the LEAs find connections between criminals. As future work, we intend to extract WhatsApp data from phone memory and compare it with the network data.

REFERENCES

[1] A. R. Javed and Z. Jalil, "Byte-level object identification for forensic investigation of digital images," in *2020 International Conference on Cyber Warfare and Security (ICCWS)*. IEEE, 2020, pp. 1–4.

[2] A. Basit, M. Zafar, A. R. Javed, and Z. Jalil, "A novel ensemble machine learning method to detect phishing attack," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*. IEEE, 2020, pp. 1–5.

[3] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of ai-enabled phishing attacks detection techniques," *Telecommunication Systems*, pp. 1–16, 2020.

[4] A. R. Javed, M. O. Beg, M. Asim, T. Baker, and A. H. Al-Bayatti, "Alphalogger: Detecting motion-based side-channel attack using smartphone keystrokes," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[5] EC-Council, *Computer Forensics: Investigating Network Intrusions and Cyber Crime*. Nelson Education, 2009.

[6] E. Casey, *Digital evidence and computer crime: Forensic science, computers, and the internet*. Academic press, 2017.

[7] R. Ayers, S. Brothers, and W. Jansen, "Guidelines on mobile device forensics (draft)," *NIST Special Publication*, vol. 800, p. 101, 2013.

[8] H. F. Tayeb and C. Varol, "Android mobile device forensics: A review," in *2019 7th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE, 2019, pp. 1–7.

[9] S. Yadav, S. Prakash, N. Dayal, and V. Singh, "Forensics analysis of whatsapp in android mobile phone," *Available at SSRN 3576379*, 2020.

[10] F. Shahzad, W. Iqbal, and F. S. Bokhari, "On the use of cryptdb for securing electronic health data in the cloud: A performance study," in *2015 17th International Conference on E-health Networking, Application Services (HealthCom)*, 2015, pp. 120–125.

[11] F. Karpisek, I. Baggili, and F. Breitinger, "Whatsapp network forensics: Decrypting and understanding the whatsapp call signaling messages," *Digital Investigation*, vol. 15, pp. 110–118, 2018.

[12] C. Shubha, S. Sushma, and K. Asha, "Traffic analysis of whatsapp calls," in *2019 1st International Conference on Advances in Information Technology (ICAIT)*. IEEE, 2019, pp. 256–260.
[13] A. Maros, J. Almeida, F. Benevenuto, and M. Vasconcelos, "Analyzing the use of audio messages in whatsapp groups," in *Proceedings of The Web Conference 2020*, 2020, pp. 3005–3011.
[14] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Sp 800-86. guide to integrating forensic techniques into incident response," 2006.
[15] A. Nath, *Packet Analysis with Wireshark*. Packt Publishing Ltd, 2019.
[16] J. Levin, *Android Internals: a Confectioner's Cookbook: Volume 1: the Power Users's View*. Technologeeks. com, 2017.
[17] B. Nelson, A. Phillips, and C. Steuart, *Guide to computer forensics and investigations*. Cengage Learning, 2018.