

Extracting Suspicious IP Addresses from WhatsApp Network Traffic in Cybercrime Investigations

Da-Yu KAO*, En-Cih CHANG*, Fu-Ching TSAI**

*Department of Information Management, Central Police University, Taoyuan 333, Taiwan

** Department of Criminal Investigation, Central Police University, Taoyuan 333, Taiwan

dayukao@gmail.com, dorislovesnoopy@gmail.com, fctsai@mail.cpu.edu.tw

Abstract—Sniffers are among the commonest approaches for capturing network traffic activities and collecting digital evidences in cybercrime investigations. The ubiquity of instant messaging (IM) apps on smartphones has provided criminals with communication channels that are difficult to decode. Moreover, investigators and analysts of cybercrimes are encountering increasingly large datasets. To combat criminal activity, law enforcement agencies (LEAs) often rely on call-record analysis. In this paper, cybercriminals are investigated by network forensics and sniffing techniques. Retrieving valuable information from specific IM apps is difficult because the criminal's IP address records are not easily recognisable on the Internet. Here, a criminal's identity is located more effectively by a packet filter framework that isolates the WhatsApp communication features from huge collections of network packets. A rule extraction method for sniffing packets is proposed that retrieves the relevant attributes from high-dimensional analysis based on geolocation and a pivot table. The utility of this methodology is illustrated on real-time network forensics and a lawful interception system in Taiwan. The methodology also meets the ISO/IEC 27043:2015 standards of fear, uncertainty, and doubt avoidance. Besides supporting LEAs in discovering criminal communication payloads, prosecuting cybercriminals and bringing them to justice, it improves the effectiveness of modern call-record analysis.

Keyword—Cybercrime Investigation, Network Forensics, Packet Analysis, VoIP, WhatsApp, Lawful Interception, ISO/IEC 27043: 2015

I. INTRODUCTION

WhatsApp is a cross-platform application enabling instant communications on electronic devices such as smartphones, tablet computers and personal computers. More than 1.5 billion active WhatsApp users were estimated in December 2017 [11]. The worldwide popularity of WhatsApp is attributable to a range of attractive features at low subscription cost. New features allow people to group chat and send texts, pictures and other multimedia elements along with their messages. Since WhatsApp was acquired by Facebook in 2014, more users have communicated through this platform by the snowball effect [11]. Unfortunately, the convenience and high functionality of WhatsApp has facilitated effective and secret communications among criminals. The present study attempts to recognise WhatsApp communication features among huge collections of network logs and packets, and thereby locate criminal activities more effectively. Discovering criminal communication contents among vague connections helps law enforcement agencies (LEAs) to better filter criminal activities.

Call-record analysis ranks among the critical criminal investigation strategies of LEAs. Call records provide important information for crime-scene investigations, such as the dates, times, and lengths of outgoing and incoming calls [1]. However, the ubiquity of instant messaging (IM) apps on smartphones has provided criminals with communication channels that are difficult to track by traditional investigation technologies. Nowadays, most criminals communicate through IM apps rather than voice phones to prevent detection by LEAs. Identifying a cybercriminal without the help of foreign authorities is difficult on the Internet, which provides complete anonymity and privacy and consequently hinders an investigation [2]. New techniques for analysing modern call records are urgently required.

The main difficulty of retrieving valuable information from specific IM apps is filtering the massive volume of network connection records on the Internet. Raw data captured from the Internet are full of packets produced by different apps from various devices, each with differing protocols, ports, and connection frequencies. Moreover, smartphones can establish connections through different network interfaces. Despite the challenges of retrieving call records or network connection logs from smartphones, Internet data provide more advanced

Manuscript received Dec. 19, 2017. This work was a follow-up of the invited journal to the accepted & presented paper of the 20th Conference on Advanced Communication Technology (ICACT2018), and this research was partially sponsored by the Executive Yuan of the Republic of China under the Grants Forward-looking Infrastructure Development Program (Digital Infrastructure-Information Security Project-107) and the Ministry of Science and Technology of the Republic of China under the Grants MOST 107-2221-E-015-002.

Da-Yu Kao is with the Department of Information Management, Central Police University, Taoyuan 333, Taiwan (Corresponding Author phone: +886-3-328-2321; fax: +886-3-328-5189; e-mail: dayukao@gmail.com).

En-Cih CHANG is with the Department of Information Management, Central Police University, Taoyuan 333, Taiwan (phone: +886-3-328-2321; fax: +886-3-328-5189; e-mail: dorislovesnoopy@gmail.com).

Fu-Ching TSAI is with the Department of Criminal Investigation, Central Police University, Taoyuan 333, Taiwan (phone: +886-3-328-2321; fax: +886-3-328-5189; e-mail: fctsai@mail.cpu.edu.tw).

and detailed information than traditional phone records. For example, the geographic information system or Internet protocol (IP) address reveals the call locations, while the captured network packets provide the multimedia content of the communications.

The remainder of this paper is organised as follows. Section 2 reviews packet analysers, the Voice-over Internet Protocol (VoIP), and WhatsApp. Section 3 describes the research design. Section 4 proposes a cybercrime investigation framework of network traffic compliant with ISO/IEC 27043:2015, and experimentally demonstrates its effectiveness. The last section concludes the paper and suggests ideas for future work.

II. LITERATURE REVIEW

A. Packet Analysers

Packet analysers are widely applied to raw-traffic analysis, attack detection, sniffing and network troubleshooting in the network security field [6]. As shown in Fig. 1, a packet analyser performs several functions [3]: reverse engineering, storing and accessing packets, detecting improper data transfer, monitoring network statistics, assisting intrusion detection systems, and handling network problems. Packet analysers can play different roles in various applications. From a moral perspective, packet analysers assist with security audits of data packets; for network administrators, they provide diagnostic tools for network problems. White-hat hackers study the reports of packet analysers to find vulnerabilities in software applications, and thereby issue an early warning before cyber-attackers can launch serious attacks. Protocol developers use packet analysers to diagnose protocol-related issues. Packet analysers can also be used in immoral ways, for example, inspecting packet payloads to decrypt passwords or sniffing traffic to deploy a man-in-the-middle attack. Packet analysis is the process of capturing and interpreting live data flowing across a network, and hence understanding the network dynamics. Most packet analyses are performed by a packet sniffer, which captures the raw network data traversing wires or wireless interfaces. Packet analysis can help with understanding the network characteristics, determining who or what is utilising the available bandwidth, finding unsecured and bloated applications, identifying summit network usage times, and detecting malicious activities.

Packet-sniffing programs are varied in type, and can be free or commercial. Each program is designed for different goals. A few popular packet-analysis programs are Tcpdump, OmniPeek, and Wireshark. Tcpdump is a command-line program, while OmniPeek and Wireshark have graphical user interfaces (GUIs) [10]. Wireshark, one of the most well-known open-source packet analysers, provides both an easy-to-use GUI and a command-line utility with very active community support [7]. It also supports offline and online modes for flexible capturing operations. The features of Wireshark are live-packet capture, a user-friendly GUI and command-line interface, data filtering, GNU open-source software, generation of various statistics, and decoding of sets of protocols [7] (see Fig. 2).

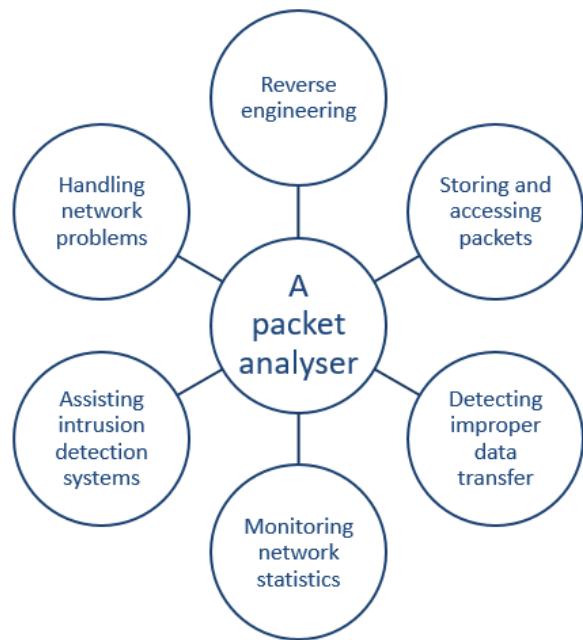


Fig. 1. Functions of a packet analyser

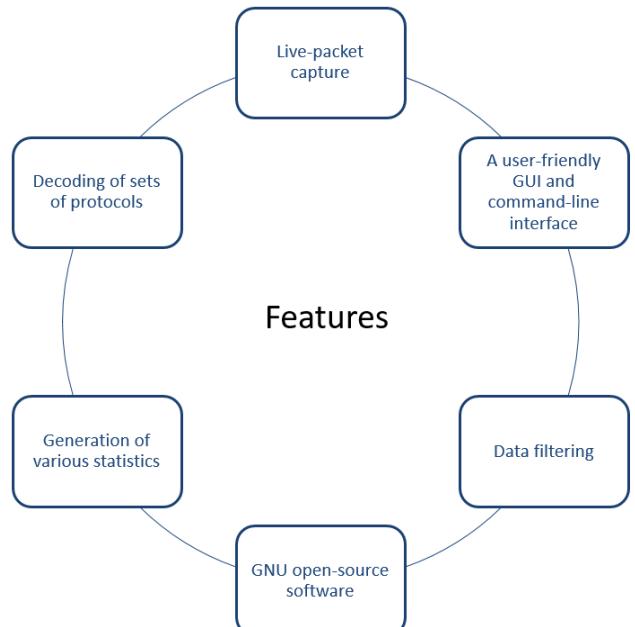


Fig. 2. Features of the Wireshark packet sniffer

B. VoIP and WhatsApp

1) VoIP

VoIP, which sends voices over an IP-based network, totally differs from circuit-switched public telephone network [8]. Whereas circuit switching allocates resources to each individual call, IP networks are packet switched. Each packet is semi-autonomous, with its own IP header and forwarded separately by the routers. VoIP manages the signalling, set-up, and tear-down of calls by session control and signalling protocols. It cooperates with several protocols such as Session Initiation Protocol, H.323, Session Description Protocol, Real-time Transport Protocol, and Inter-Asterisk eXchange. A traditional system requires much control signalling to accomplish the various tasks, but VoIP collects these signalling messages and places them inside IP packets.

It is worth mentioning that because an IP can and does run over almost all types of low-layer communication architectures, VoIP can as well. Researchers can compare the topologies of different VoIP architectures, and can short-list the basic skills required to work on VoIP and traditional telephony. Both VoIP and telephony serve the same functions with the same equipment, but using different techniques with completely different sets of protocols [3].

2) WhatsApp

Network sniffing is a vital strategy in modern crime investigation [1]. With the rapid evolution of the Internet, communication has transformed from traditional phone calling to network-based VoIP interactions. The low cost and interactive features (with delivery of multimedia elements) of IM applications have encouraged a large number of users to almost abandon traditional phones. The most commonly used feature of WhatsApp is voice calling. When a user starts a call to a private IP address behind a network address translation (NAT) firewall, the packet routing should be assisted by a STUN (Session Traversal Utilities for NAT) protocol, which allows the end computer to discover the public IP address, and permits NAT traversal of real-time voices, messages, and other interactive communications [9]. The anonymous nature of the Internet limits the abilities of LEAs to monitor the communications of criminal activities. Therefore, efficient network sniffing technologies are demanded for cybercrime investigation.

3) ISO/IEC 27043:2015

The purpose of network sniffing is to discover criminal activities. To bring criminals to justice, the integrity of digital evidence should be maintained by procedures that collect and analyse network packets. The 2015 ISO/IEC 27043 standard provides readiness, initialisation, acquisitive, and investigative guidance for criminal investigations [4]. However, the ISO standards have been rarely applied in practical solutions. This study simulates a network sniffing scene that collects packets between the victim and suspect following the recommendation processes in ISO/IEC 27043: 2015. The present paper demonstrates the framework of the network sniffing strategy for LEAs operating under lawful interception warrant procedures.

III. RESEARCH DESIGN

This paper simulates the communications between the victim and suspect, and extracts the likely incriminating features in the communication. Using these features, it proposes filtering rules by which LEAs can effectively target suspects in WhatsApp packets. Our research design comprises four phases: data collection, data preparation, feature recognition, and result evaluation (see Fig. 3).

A. Research Experiment

The sniffing of WhatsApp voice calls, collection of IP address information, and personal identification of the WhatsApp application target, were conducted in a controlled environment.

1) Software Environment

Within the experimental environment, the transmission time and packet size were controlled by varying the bandwidth and traffic congestion. All devices were initially configured as follows [5]:

a) Victim: Cellphone at Domain A

- Android Operating System v5.0
- WhatsApp Ver. 2.17.146

b) Investigator: Computer at Domain A

- Wireshark v2.2.5
- Windows 10.0.14393
- Excel 2013
- I2 Analyst's Notebook 8 v8.5.5
- NodeXL Basic Excel Template 2014

c) Suspect: Cellphone at Domain B

- iOS 10.3.2
- WhatsApp Ver. 2.17.146

2) Participants

The experimental participants included a victim, an investigator and a suspect (Fig. 3). Domain A was configured by the investigator or the victim. Domain B was used by the suspect, criminal, or target. All communication packets were sniffed by Wireshark.

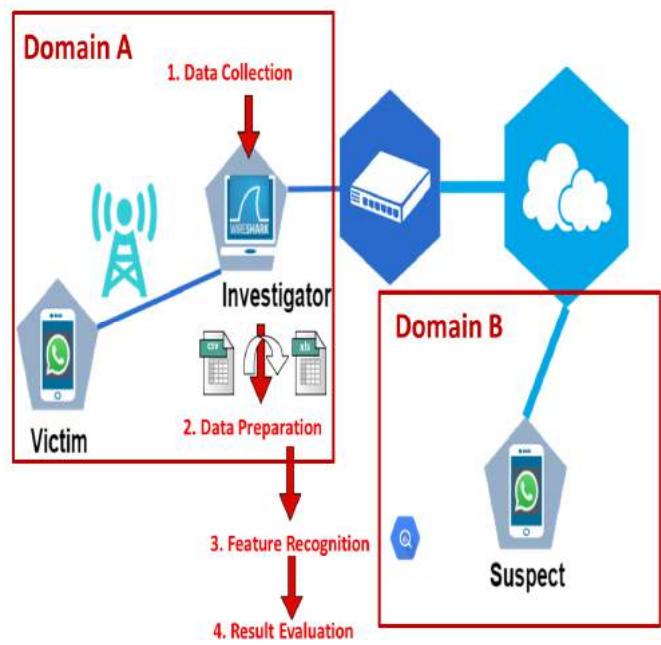


Fig. 3. Research design

B. Experimental Phases in the Research Design

The four experimental phases in our research design are discussed below.

1) Data Collection

The hotspot of the investigator computer in Fig. 3 shared its network connections with the victim's cellphone. The researchers monitored and eavesdropped (by copying) the traffic to and from the investigator's computer. The eavesdropping included the packets from the victim's phone to the Internet. By setting a midpoint in the investigator host,

the researchers were able to use Wireshark, capture all network traffic, and investigate the criminal behaviour along the victim–suspect route. Routine data transmission was prioritised over the copying process. This priority might have caused dropped Ethernet frames when collecting the incriminating evidence.

2) Data Preparation

To capture general traffic, the researchers installed the packet-sniffing software, configured the network interface controller (NIC) in promiscuous mode, and collected all network traffic addressed to the MAC address of the NIC. From the collected data, the researchers could overview the WhatsApp performance and tentatively identify the suspect. For this purpose, the data passing through the investigator computer were captured and analysed, then presented in an easy-to-read format.

3) Feature Recognition

Common tools for collecting network traffic, such as pcap (for Unix-like systems) and libcap (for Windows systems), collect thousands of small data packets that are sent across the Internet. Such numerous small packets can be difficult to navigate. The main purposes of the present study are listed below:

- Assess the overall traffic flow through the network
- Exactly copy the network traffic for predictive analysis
- Identify how WhatsApp applications generate the VoIP traffic
- Identify the IP address of the suspect WhatsApp user
- Highlight the features of the WhatsApp packets in the suspect's IP address

4) Evaluation Results

We monitored only the traffic to and from the investigator computer. While two users conducted voice calls through

WhatsApp, the researchers assumed that the Wireshark deployment node was lawfully intercepted by the warrant procedures of the victim's agreement. To start a sniffing procedure, the investigator computer must be on the same network as the snuffed cell phone. Packets can be very useful for tracking suspects or offenders in cybercrime investigations.

IV. PROPOSED CYBERCRIME INVESTIGATION FRAMEWORK OF NETWORK TRAFFIC

The storage and handling of network traffic requires the processing of massive numbers of packets, maintaining the integrity of the digital evidence, and preserving the digital evidence during the investigative period. These requirements present significant challenges for LEAs. The ISO/IEC 27043: 2015 international standards provide instructional guidance for the readiness, initialisation, acquisitive, and investigative processes. Our network-based sniffer framework helps to address the above challenges and formalises what should be logged for an appropriate cybercrime investigation.

A. Materials and Methods

The collected digital evidence should increase the conviction rate and restore the truth. The following standardised procedures are vital to the validity and reliability of the collected digital evidence. The network-based sniffer experiments in this study were based on the ISO/IEC 27043: 2015 international standards of incident investigation processes. The various process classes are shown in Fig. 4 and discussed below.

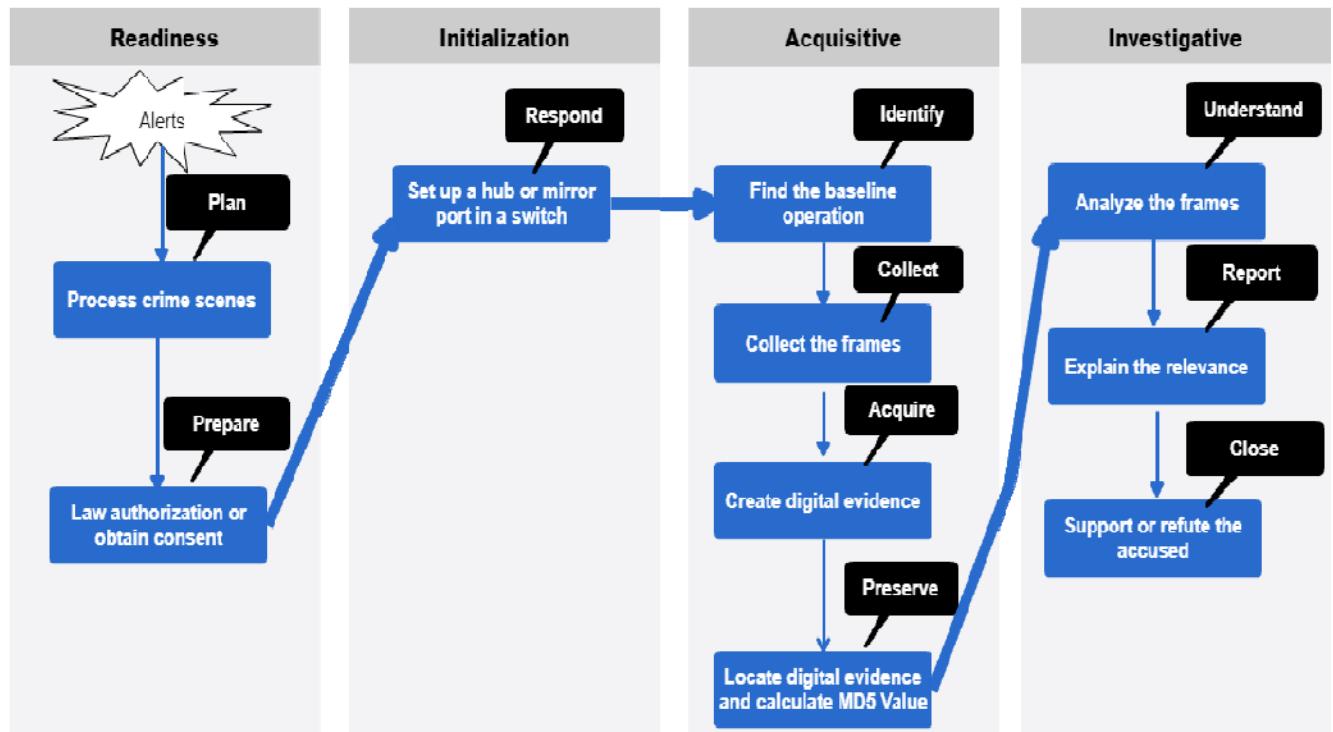


Fig. 4. Network-based sniffer framework for cybercrime investigation

B. ISO/IEC 27043:2015 Process Class

In recent years, network sniffing in criminal investigations has been conducted under lawful interception warrant procedures. Network sniffing poses great challenges to LEAs because unlike traditional call interception, it lacks any systematic procedure. The network sniffing framework proposed in this study is guided by the ISO/IEC 27043: 2015 standards. In particular, it follows the ten steps in ISO/IEC 27043: 2015 to preserve the integrity and prevent damage of the digital evidence. LEAs can adopt the framework as a standard operation procedure to facilitate an efficient network traffic analysis.

1) Readiness Process Class

a) Plan

The plan phase consolidates the scope and purpose of the investigation. Using Wireshark, the researchers captured all network traffic along the victim–suspect route. The WhatsApp network traffic was sniffed to identify the calling and receiving phones. The personal computer was configured as a hotspot for sharing network connections to the cell phone, and as the node for capturing the network packets utilised by Wireshark. The study included the detailed routing information, such as the IP address, protocol, time, and packet length. The investigative tasks were assisted by careful planning.

b) Prepare

Network sniffing is an interdisciplinary process. The team members responsible for this task should possess knowledge of packet analysis, technology devices construction and criminal investigation. Therefore, LEAs should provide multi-domain training courses for their team members. Good preparation ensures that criminal investigators can cope with various crime scenes. Once the collection process is complete, the data integrity can be documented by the MD5 value of the pcap file, and the data can be preserved on a write-only medium [1].

2) Initialisation Process Class

a) Respond

One law enforcement strategy in criminal investigations is a dedicated middle node for packet sniffing. To this end, we set up a network sniffing framework that efficiently responds to a crime case. To handle the massive volume of network packets on the Internet, we erected a hub or mirror port in a switch that probed the routing nodes containing the targeted WhatsApp connections. The collected information provides LEAs with quick responses to various crime scenes.

3) Acquisitive Process Class

a) Identification

LEAs should convert the huge number of Internet packets to readable information. To identify the criminal activities, we imported the connecting information to the pivot table. Having identified the facts of the crimes, LEAs can process the investigation by various systematic approaches.

b) Collection

Internet packets were collected by Wireshark software. The collector should be placed en-route between the caller and receiver. Under the lawful interception warrant procedures, the network sniffing node should be the Internet data centre owned by either caller, or the telecommunication service provider of the receiver.

c) Acquisition

Having confirmed the sniffing nodes, the LEA should deploy the packet analyser that collects the network packets. Most of the packet analysers store the packets in their own formats. To analyse the payload information more effectively, we imported the files produced by various packet analysers into a normalised database table using extract–transform–load tools.

d) Preservation

Digital evidence is commonly acquired by live investigation or dead forensics. Live investigation is conducted on a system running at the scene, and dead forensics is usually performed in a trusted laboratory environment. In both investigation modes, the data should be preserved to maintain its integrity. In this study, the data integrity was verified by the MD5 hash value.

4) Investigative Process Class

a) Understanding

This stage analyses the collected digital evidence. The modus operandi of criminal activities in a huge database is probed by forensic tools. Open-source toolkits, data mining, and machine learning approaches that reflect the features or contextual information in a crime case, are also available.

b) Reporting

A criminal investigator must document the processes and results of the case. The report should not only detail the crime case, but should also provide testimony in court. Moreover, it should be precise, easily read, and clearly understandable. The report can also contain multimedia elements such as video, audio and pictures.

c) Close

After checking that all evidence is well-protected and safely deposited, the criminal investigation is closed. The storage should be regulated by strict rules, preventing the evidence from being changed, lost, stolen or destroyed.

C. Research Findings

1) Feature Recognition in Frequency Distribution Analysis

a) Frequency Distribution Analysis

The data packets were captured in the pcap file format, and imported to Wireshark for demonstrating their header and payload information. The pcap files were then exported to excel, where the high-dimensional data were viewed from different angles in a pivot table. To investigate the features of the WhatsApp communications, we imported the headers and payloads of the captured packets into the pivot table. In a frequency distribution analysis, most of the packet fields

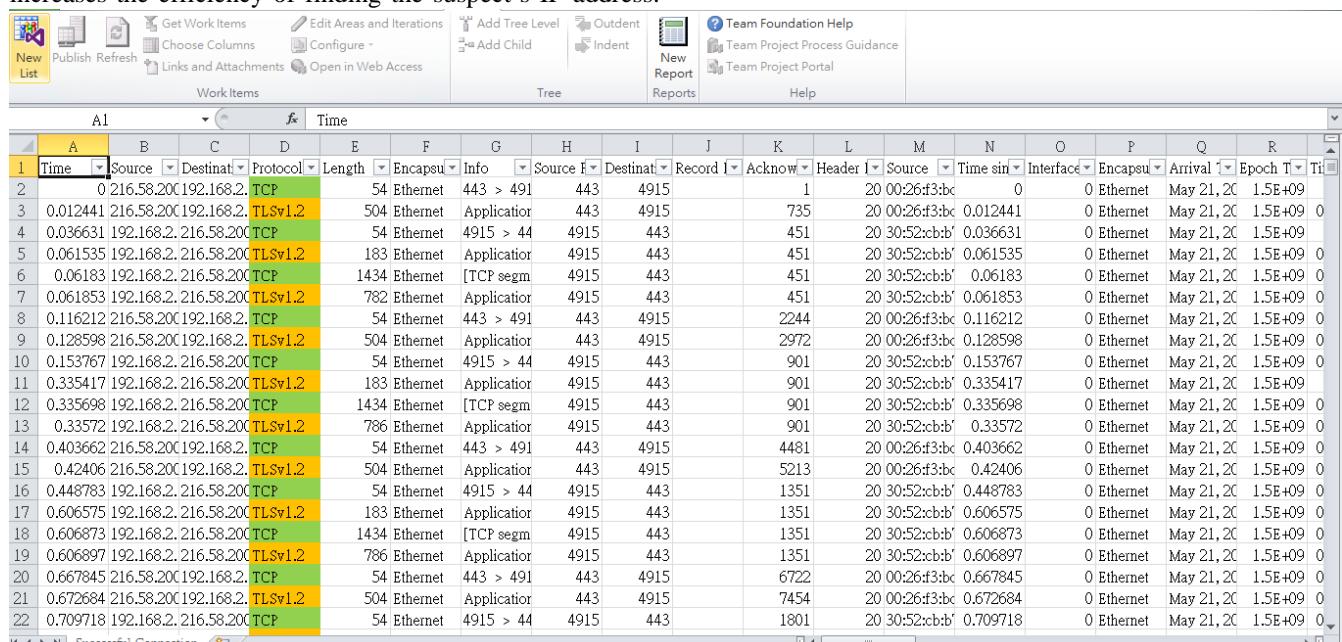
consisted of random values with no relevance to communication features. However, the values of several attributes, such as Differentiated Service Field, Flags, and Differentiated Services Codepoint, were fixed. These fixed-value attributes were selected as the criteria of feature recognition in the WhatsApp communications. The derived packet attributes and their contents are shown in Table 1.

TABLE I.
CRITERIA AND CONTENTS OF FEATURE RECOGNITION IN WHATSAPP
COMMUNICATIONS

Packet Attribute	Content
Differentiated Services Field	0x38
Flags	0x00
Differentiated Services Codepoint	Assured Forwarding 13

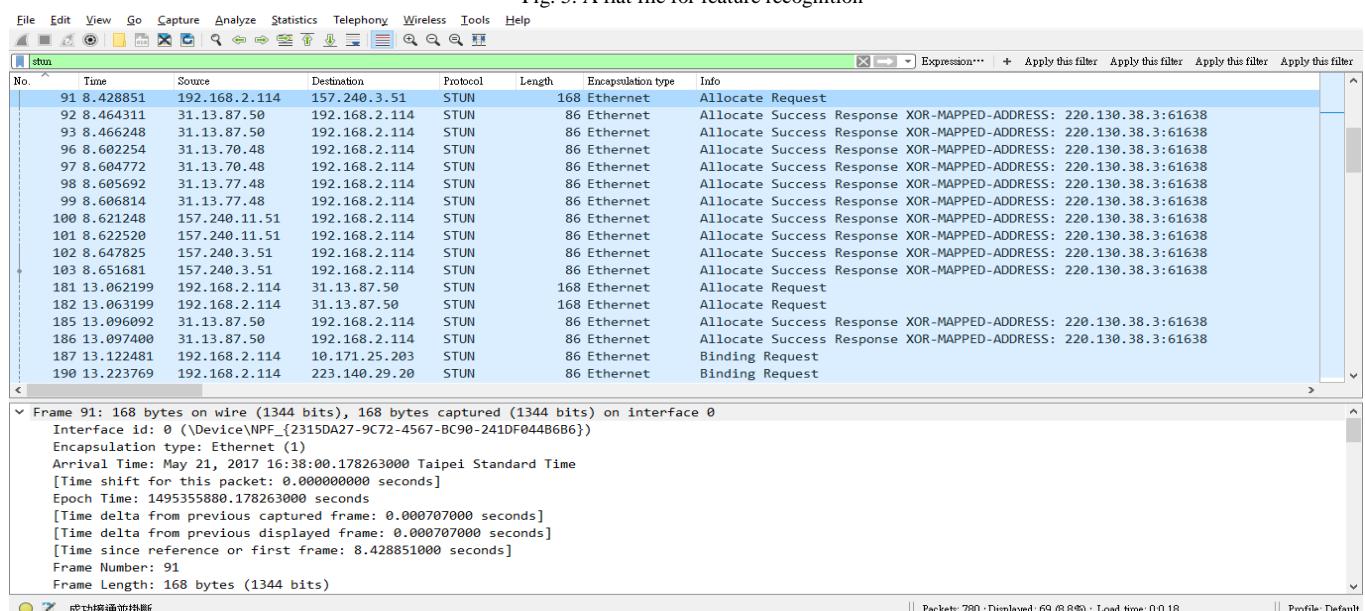
b) Feature Recognition

The feature-recognition phase identifies the features generated by WhatsApp in the packet records. This phase increases the efficiency of finding the suspect's IP address.



Time	Source	Destination	Protocol	Length	Encapsulation type	Info	Record ID	Acknowledge	Header	Source	Time stamp	Interface	Encapsulation	Arrival time	Epoch time	Time
0 216.58.200.192.168.2. TCP	54	Ethernet	443 > 491	443	4915		1	20 00:26:f3:bc	0	0 Ethernet	May 21, 2017 15:49:09	0				
0.012441 216.58.200.192.168.2. TLSv1.2	504	Ethernet	Applicator	443	4915		735	20 00:26:f3:bc	0.012441	0 Ethernet	May 21, 2017 15:49:09	0				
0.036631 192.168.2.216.58.200.TCP	54	Ethernet	4915 > 44	4915	443		451	20 30:52:cbb'	0.036631	0 Ethernet	May 21, 2017 15:49:09	0				
0.061535 192.168.2.216.58.200.TLSv1.2	183	Ethernet	Applicator	4915	443		451	20 30:52:cbb'	0.061535	0 Ethernet	May 21, 2017 15:49:09	0				
0.06183 192.168.2.216.58.200.TCP	1434	Ethernet	[TCP segm]	4915	443		451	20 30:52:cbb'	0.06183	0 Ethernet	May 21, 2017 15:49:09	0				
0.06183 192.168.2.216.58.200.TLSv1.2	782	Ethernet	Applicator	4915	443		451	20 30:52:cbb'	0.06183	0 Ethernet	May 21, 2017 15:49:09	0				
0.116212 216.58.200.192.168.2. TCP	54	Ethernet	443 > 491	443	4915		2244	20 00:26:f3:bc	0.116212	0 Ethernet	May 21, 2017 15:49:09	0				
0.128598 216.58.200.192.168.2. TLSv1.2	504	Ethernet	Applicator	443	4915		2972	20 00:26:f3:bc	0.128598	0 Ethernet	May 21, 2017 15:49:09	0				
0.153767 192.168.2.216.58.200.TCP	54	Ethernet	4915 > 44	4915	443		901	20 30:52:cbb'	0.153767	0 Ethernet	May 21, 2017 15:49:09	0				
0.335417 192.168.2.216.58.200.TLSv1.2	183	Ethernet	Application	4915	443		901	20 30:52:cbb'	0.335417	0 Ethernet	May 21, 2017 15:49:09	0				
0.335698 192.168.2.216.58.200.TCP	1434	Ethernet	[TCP segm]	4915	443		901	20 30:52:cbb'	0.335698	0 Ethernet	May 21, 2017 15:49:09	0				
0.33572 192.168.2.216.58.200.TLSv1.2	786	Ethernet	Application	4915	443		901	20 30:52:cbb'	0.33572	0 Ethernet	May 21, 2017 15:49:09	0				
0.403662 216.58.200.192.168.2. TCP	54	Ethernet	443 > 491	443	4915		4481	20 00:26:f3:bc	0.403662	0 Ethernet	May 21, 2017 15:49:09	0				
0.42406 216.58.200.192.168.2. TLSv1.2	504	Ethernet	Application	443	4915		5213	20 00:26:f3:bc	0.42406	0 Ethernet	May 21, 2017 15:49:09	0				
0.448783 192.168.2.216.58.200.TCP	54	Ethernet	4915 > 44	4915	443		1351	20 30:52:cbb'	0.448783	0 Ethernet	May 21, 2017 15:49:09	0				
0.606575 192.168.2.216.58.200.TLSv1.2	183	Ethernet	Application	4915	443		1351	20 30:52:cbb'	0.606575	0 Ethernet	May 21, 2017 15:49:09	0				
0.606873 192.168.2.216.58.200.TCP	1434	Ethernet	[TCP segm]	4915	443		1351	20 30:52:cbb'	0.606873	0 Ethernet	May 21, 2017 15:49:09	0				
0.606897 192.168.2.216.58.200.TLSv1.2	786	Ethernet	Application	4915	443		1351	20 30:52:cbb'	0.606897	0 Ethernet	May 21, 2017 15:49:09	0				
0.667845 216.58.200.192.168.2. TCP	54	Ethernet	443 > 491	443	4915		6722	20 00:26:f3:bc	0.667845	0 Ethernet	May 21, 2017 15:49:09	0				
0.672684 216.58.200.192.168.2. TLSv1.2	504	Ethernet	Applicator	443	4915		7454	20 00:26:f3:bc	0.672684	0 Ethernet	May 21, 2017 15:49:09	0				
0.709718 192.168.2.216.58.200.TCP	54	Ethernet	4915 > 44	4915	443		1801	20 30:52:cbb'	0.709718	0 Ethernet	May 21, 2017 15:49:09	0				

Fig. 5. A flat file for feature recognition



No.	Time	Source	Destination	Protocol	Length	Encapsulation type	Info	Expression...	+ Apply this filter	Apply this filter	Apply this filter	Apply this filter
91	8.428851	192.168.2.114	192.168.3.51	STUN	168	Ethernet	Allocate Request					
92	8.464311	31.13.87.50	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
93	8.466248	31.13.87.50	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
96	8.602254	31.13.70.48	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
97	8.604772	31.13.70.48	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
98	8.605692	31.13.77.48	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
99	8.606814	31.13.77.48	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
100	8.621248	157.240.11.51	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
101	8.622520	157.240.11.51	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
102	8.647825	157.240.3.51	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
103	8.651681	157.240.3.51	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
181	13.062199	192.168.2.114	31.13.87.50	STUN	168	Ethernet	Allocate Request					
182	13.063199	192.168.2.114	31.13.87.50	STUN	168	Ethernet	Allocate Request					
185	13.096092	31.13.87.50	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
186	13.097400	31.13.87.50	192.168.2.114	STUN	86	Ethernet	Allocate Success Response XOR-MAPPED-ADDRESS: 220.130.38.3:61638					
187	13.122481	192.168.2.114	10.171.25.203	STUN	86	Ethernet	Binding Request					
190	13.223769	192.168.2.114	223.140.29.20	STUN	86	Ethernet	Binding Request					

Fig. 6. Packets of STUN Protocol

b) IP Geolocation

The geolocation of an IP address is important for transforming the location from the network space to physical space. The Whois database (originally designed for Unix) has become the commonest mechanism for locating the registration information of IP resources registered in Internet-number resource organisations. After querying a registry, an open-source Whois, Lookup, or IP location tool returns rich geolocation information, including the domain ownerships, addresses, locations, and phone numbers of the queries. To evaluate the effectiveness of our proposed framework, we collected network packets during the 34.505698-second period containing the WhatsApp communications. Although the traffic was only captured from the local area network, the IP list was rendered complicated by additional connections with its software companies and Internet service providers. The geolocations of the IP addresses in this experiment were transformed by whois lookup tools and are listed in Table 2.

TABLE II.
GEOLOCATIONS OF THE IP ADDRESSES IN THE PRESENT EXPERIMENT

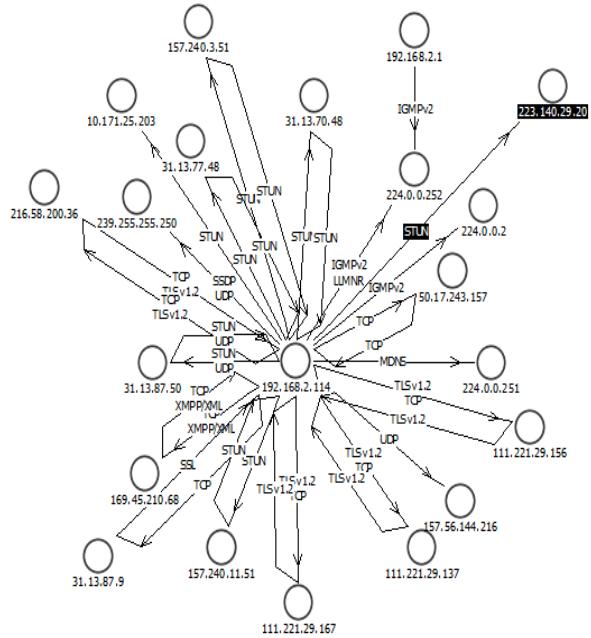
IP Address	Whois Lookup
192.168.2.114	Victim
224.140.29.20	Suspect(EMOME-IP.hinet.net, Taiwan)
157.240.4.51	United States Menlo Park Facebook Inc.
31.14.87.50	Taiwan, Province Of China Taiwan, Province Of China Taipei Facebook Ireland Ltd
31.14.70.48	United States United States Los Angeles Facebook Ireland Ltd
157.240.11.51	United States United States Menlo Park Facebook Inc.

3) Observation Rules on Suspect IP Addresses

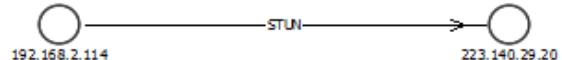
Several phenomena were observed in the WhatsApp network traffic (see Table 3). The rules were established to improve the efficiency of selecting WhatsApp communications from the Internet data. The rules should filter out the noise connections produced by the software companies, manufacturers of network devices and ISPs. To demonstrate the effectiveness of our approach, we applied the generated rules to the collected packets containing the WhatsApp communications. Fig. 7 shows the network topology before and after implementing the rules. The proposed rules successfully pruned the complicated topology and revealed the victim–suspect connections. The above findings will assist LEAs in their cybercrime investigations of criminals contacting through WhatsApp.

TABLE III.
THE RULES

Rule 1	IF Differentiated Service Field = 0x38, Flags = 0x00, and Differentiated Services Codepoint = Assured Forwarding 13 THEN Source IP address = suspect
Rule 2	IF Protocol = STUN, Length = 86, and Info = Binding Request THEN Destination IP address = suspect



(a) Before the rule implementation



(b) After the rule implementation

Fig. 7. Network topology of WhatsApp communications after pruning by our proposed rules

V. CONCLUSION AND FUTURE WORKS

Modern call-record analysis is an expected future trend of criminal investigation strategies. Recognising the communication features of IM software on smartphones is essential for revealing the locations of suspects, providing clues that improve the efficiency of investigative work by LEAs. This study has developed a rule extraction framework that reveals the WhatsApp communications and eliminates the impact of disordered Internet connections. In an experimental test, the generated rules successively simplified the complexed network topology to simple connections between the suspect and victim. The criteria discovered in the sniffed packets provided instructive information for identifying the characteristics of WhatsApp communications. Furthermore, the proposed framework can explore the features produced by other IM software. Extracting the connections made by criminals will improve the prosecution and conviction rates by LEAs. To keep pace with the rapid developments of IM software, future research should consider the software upgrade problem, and minimise the impact of updating the IM software. The application of the proposed framework to encrypted communication should be examined from various perspectives. As the proposed methodology is intended to help LEAs, it should also provide a more complete coverage of IM feature-recognition applications. For this purpose, additional IM software should be considered in future work.

ACKNOWLEDGMENT

The authors would like to thank Enago for the English language review.

REFERENCES

- [1] Casey, E., *Digital Evidence and Computer Crime*. Waltham, MA: Elsevier Inc., pp. 727-735, 2011.
- [2] EC-Council, *Computer Forensics: Investigating Network Intrusions and Cyber Crime*. Boston, MA: EC-Council Press, pp. 27-60, 2010.
- [3] Hartpence, B., *Packet Guide to Voice over IP: A System Administrator's Guide to VoIP Technologies*. Sebastopol, CA: O'Reilly Media Inc., pp. 2-5, 2014.
- [4] International Organization for Standardization (ISO), "ISO/IEC 27043: 2015 Information Technology – Security Techniques - Incident Investigation Principles and Processes," Switzerland: ISO Office, pp. 5-20, 2015.
- [5] Kao, D. Y. and Wu, W. Y., "Practical Packet Analysis: Exploring the Cybercriminal behind the LINE Voice Calls," 2017 19th IEEE International Conference on Advanced Communications Technology (ICACT), Pyeong Chaung, South Korea, Feb. 19-22, 2017, 2011.
- [6] Kizza, J. M., *A Guide to Computer Network Security (3rd Edition)*. Swindon, UK: Springer-Verlag London Ltd., pp. 299-324, 2015.
- [7] Nath, A., *Packet Analysis with Wireshark*. Birmingham, UK: Packet Publishing Ltd., pp. 56-146, 2015.
- [8] Rahbar, A. G., *Quality of Service in Optical Packet Switched Networks*. Danvers, MA: IEEE Press, pp. 19-43, 2015.
- [9] Roy, R. R., *Handbook on Session Initiation Protocol: Networked Multimedia Communications for IP Telephony*. Boca Raton, FL: CRC Press, pp. 1-350, 2016.
- [10] Sanders, C., *Practical Packet Analysis: Using Wireshark to Solve Real-World Network Problems (3rd Edition)*. San Fransciso, CA: No Starch Press, pp. 53-102, 2017.
- [11] Statista—the Statistics Portal, "the Number of monthly active WhatsApp users worldwide from April 2013 to December 2017 (in millions)." [Online]. Available: <https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/>



Da-Yu Kao received the B.S. and M.S. degree in information management from Central Police University, Taiwan, in 1993 and 2001, the Ph.D degrees in Crime Prevention and Correction from Central Police University, Taiwan, in 2009, respectively. From 1993 to 1996, he was with Taipei City Police Department, Taiwan, where he was an information technology police officer involved in the development of policing information systems. From 1996 to 2007, he was with Criminal Investigation Bureau, National Police Administration, Taiwan, where he was a detective and forensic police officer in cybercrime investigation and digital forensics. From 2007 to 2013, he was with Maritime Patrol Directorate General, Coast Guard Administration, Taiwan, where he was an information technology section chief in the department of information and communication. Since 2013, he has been with Central Police University, Taiwan, where he is currently an associate professor in the Department of Information Management. His research interests include cybercrime investigation, digital forensics, digital evidence, information management, criminal profiling, and cyber criminology.



En-Cih CHANG received the B.S. degree in information management from Central Police University, Taiwan, in 2018. In 2018, she is studying in the College of Communication and Information, Florida State University, Tallahassee, FL, USA. Her current research interests include information security, incident response, cybercrime investigation, digital forensics, information systems management, criminal profiling, cyber criminology, and machine learning.



Fu-Ching TSAI received the B.S. degree in Information Management from Central Police University, Taiwan, in 2001, the M.S. and Ph.D degrees in Institute of Information Management from National Cheng Kung University, Taiwan, in 2005

and in 2012, respectively. From 2001 to 2010, he was with Pingtung County Police Bureau, Taiwan, where he was a lieutenant involved in the development of policing information systems. From 2010 to 2014, he was with National Police Agency, Taiwan, where he was a division assistant in network & security incident investigation. From 2014 to 2017, he was with Changhua County Police Department, Taiwan, where he was an Information Management Division chief. Since 2017, he has been with Central Police University, Taiwan, where he is currently an assistant professor in the Department of Criminal Investigation. His research interests include data mining, text mining, digital forensics, social network analysis, and cyber criminology.