

Forensic Artifacts of Network Traffic on WeChat Calls

Da-Yu Kao^a, Ting-Chi Wang^a, Fu-Ching Tsai^b

^aDepartment of Information Management, Central Police University, Taiwan

^bDepartment of Criminal Investigation, Central Police University, Taiwan

^bCorresponding Author: fctsai@mail.cpu.edu.tw

Abstract— Voice over Internet Protocol (VoIP) applications, such as WeChat, Whatsapp, or LINE, have gained increasing popularity during the last few years. Packets analysis of VoIP is one of the critical criminal investigation strategies for law enforcement agencies (LEAs). Due to its convenience and multi-functionality, the WeChat application is accessible by a massive amount of people in Asia, especially in Taiwan and China. Therefore, this study uses Wireshark to reveal the tool marks behind the WeChat messages and voice calls. It will be an excellent help for LEAs to identify cybercriminals and bring them to justice.

Keywords— Cybercrime Investigation, Packet Analysis, VoIP, WeChat, Tool Marks

I. INTRODUCTION

An investigation is a systematic examination to identify or verify the facts from 5W1H questions in a crime or incident. 5W1H questions include who, where, what, when, why, and how. The evidence from digital devices is identified, collected, examined, analyzed, and presented in a forensically sound manner [1]. Investigators seek to gain a better understanding of digital objects and present them to a court of law [10]. The identification of an incident or a crime leads to the formation of a hypothesis about what might have happened. An investigation can focus on identifying supporting information to prove a case, identifying information that refutes an argument, or verifying the validity of any given information. The questions defined by the 5W1H model should always help us to establish a hypothesis based on the information triggering the investigation [11].

Section 2 reviews the related VOIP protocols and Deep Packet Inspection (DPI) on WeChat calls. Section 3 describes the research experiment. The discussions and analyses of this study are presented in Section 4. Section 5 is the conclusion.

II. LITERATURE REVIEW

Deri [5] introduces three leading VoIP protocol families which are based on a connectionless protocol such as UDP: standards protocols (SIP/H.323/RTP), proprietary but well-documented protocols (Cisco skinny), and proprietary protocols (Skype). The use of standard or known protocols may transport encoded proprietary data.

A. VoIP Protocols

VoIP (Voice over IP) is a technology for voice communication over Internet Protocol networks. IP network communication is composed of packets, which contain a header and a payload in transit [14]. However, circuit-switched telephony establishes a dedicated, end-to-end circuit for each call. Each packet may take a very distinct path determined by the routing tables on network traffic. VoIP. Internet telephony is a voice service that is transmitted based on an IP network after compressing a voice signal into a data packet. A telecom application service for sending voice over an open Internet is provided [7].

1) Three Steps in VOIP Communication

Every VOIP communication is made of three necessary steps [7]:

a. Step 1. Signaling Negotiation

When a caller communicates with another called party, it initiates a signaling communication for verifying credentials, negotiating the call, and agreeing on a standard codec and exchanging data ports.

b. Step 2. Communication Call

The communication call takes place on the selected ports, and the specified codec encodes the payload.

c. Step 3. Closing Call

When one of the parties decides to close the call, the signaling protocol terminates the call.

2) IP Traceback-related Protocols

a. UDP

With the User Datagram Protocol (UDP), computer applications can send messages to other computers on the Internet. Prior communications are unnecessary to set up communication channels or data paths [2].

b. RTCP

The Real-time Transport Protocol (RTP) in communication and entertainment systems delivers telephony, audio, and video applications over IP networks. RTP typically runs over UDP. While RTP carries the media streams, RTP Control Protocol (RTCP) in RFC 3550 is used to monitor the quality of service (QoS) and aids the synchronization of multiple streams. RTCP provides control information for an RTP

session in the delivery and packaging of multimedia data. It does not transport any media data itself.

c. ICMP

The Internet Control Message Protocol (ICMP) transmits control messages in the Internet Protocol (IP). It provides various problem feedbacks that may occur in the communication environment. Analysts can diagnose the problems and take appropriate measures to resolve them.

3) Other Protocols

a. TCP

Transmission Control Protocol (TCP) is a connection-oriented, reliable, byte-based stream-based transport layer protocol. It creates a virtual connection through the three-way handshake.

b. HTTP

The HyperText Transfer Protocol (HTTP) is a protocol for distributed, collaborative, and hypermedia messaging systems. HTTP is the foundation of the global information network's data communications.

c. HTTP/XML

AJAX stands for Asynchronous JavaScript And XML, which is asynchronous JavaScript and XML. AJAX uses the XMLHttpRequest object to communicate with the server. It can transmit and receive information in a variety of formats, including JSON, XML, HTML, and text files.

d. SSL

Secure Sockets Layer (SSL) is a security protocol to provide security and data integrity for Internet communications [9].

B. DPI on WeChat Calls

Deep Packet Inspection (DPI) looks at the network contents [13]. It can work with filters to find and redirect network traffic from an online service. There are multiple ways to acquire network traffic for DPI. Using port mirroring is a popular way [16].

1) Collect Data: Wireshark

Wireshark is a packet analyzer to capture the network traffic, display its data, and examine what is going on inside a network cable. It is available for Unix and Windows platforms [3]. It also helps identify different protocol packets through various devices such as Ethernet or Wi-Fi devices [12].

2) WeChat

WeChat is a popular instant messaging application on Android, iPhone, and BlackBerry smartphones [15]. With the popularization of smartphones, more forms emerge in the aspect of social network services. In Asia, WeChat has become universal communication applications in recent several years [8].

3) IP Traceback

IP traceback is the process of finding the source IP address of the user who sends messages. It is essential for identifying cybercriminals [6].

III. RESEARCH EXPERIMENT

A. Set-Up

The research has set up a controlled environment to simulate the WeChat voice calls, to collect IP address information, and to find out the personal identity of the target behind the WeChat application. In this experiment, the suspect uses the mobile phone to contact the victim's laptop. Wireshark sniffed all the communication packets. The researchers filter out the packet information of the IP address of the suspect and the victim by Excel and establish rules. We find out the patterns of the packet information between the suspect and the victim's device. By using the rules, trends, and the victim's IP address to identify the suspect's IP address. The experiment environment is shown in Table 1.

TABLE 1. EXPERIMENT ENVIRONMENT

Role (Device)	IP address	OS Version	WeChat Version
Victim (Laptop)	192.168.43.87	Windows 10 Version 1803	V 7.0.3
Suspect (Mobile Phone)	114.137.238.240	Android 8.1.0	V2.6.7.40
Analysis Toolkits	Wireshark V 2.6.5, Excel 2013, and IBM i2 Analyst's Notebook		

B. Hypothesis Testing

A hypothesis is prepared for the relationship between the two data sets or statements. The follow-up hypothesis is based on the WeChat process through a set of random communications [4].

1) Null Hypothesis (H0)

The null hypothesis, which denoted by H0, is usually assumed that sample observations are purely accidental. A simple explanation contradicts a theory that one would like to prove.

2) Alternative Hypothesis (H1)

The alternative hypothesis which denoted by H1 is the hypothesis that some non-random cause influences sample observations. A hypothesis associated with a theory one would like to prove.

3) Accept/Reject the Null Hypothesis

This comparison is considered significant if the relationship between the data sets would be an unlikely realization of the null hypothesis. Hypothesis testing can have one of two outcomes: the investigator rejects the null hypothesis or accepts the null hypothesis. Acceptance means that the null hypothesis is correct. This study follows a formal process to determine whether to reject the null hypothesis based on sample data.

C. Hypothesis Testing in WeChat Application

The process of hypothesis testing in WeChat applications consists of the following four steps.

1) Step 1: State the Hypothesis

Step one involves stating the null and alternative hypotheses. The hypothesis is stated in such a way that they are mutually exclusive.

2) Step 2: Formulate an Experiment Framework

The experiment framework describes how to use sample data to evaluate the null hypothesis. All packets are sniffed and identified in this experiment by the researchers. Based on the experiment framework in Figure 1, users can communicate with others through the WeChat server. This set of tests focuses on the victim's IP address from voice communication and examines whether the IP address of the suspect appears in the WeChat voice communication of network packets. The experiment framework was designed to find whether the suspect's IP address is shown in the packets.

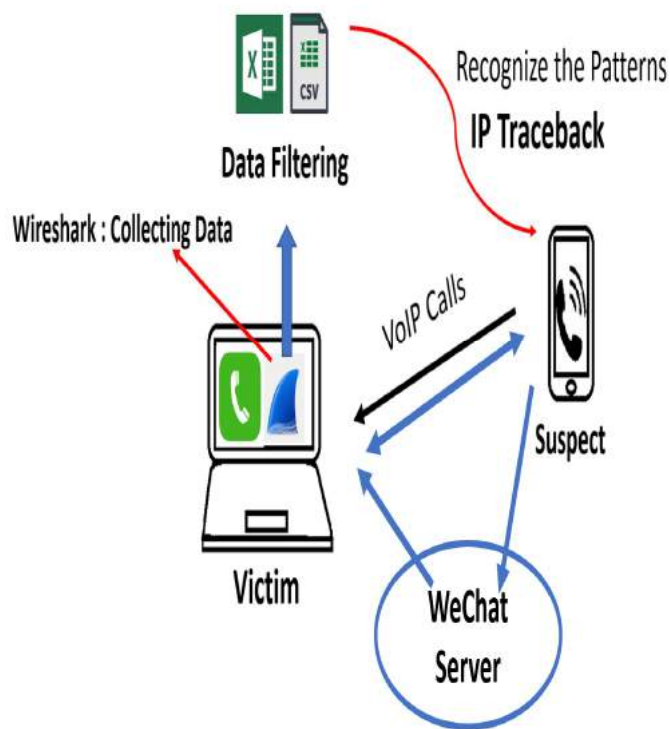


Figure 1. Experiment Framework

3) Step 3: Analyze Sample Data

The related packets on the victim's laptop are analyzed and preserved to find the WeChat tool mark. After the experiment, the researchers use pattern extraction to filter the packet and determine if the tool marks appear in the packets.

a. Data Clustering

Data clustering groups a set of objects. We use Excel to filter the raw data and visualize our data with IBM i2

Analyst's Notebook. This visual analysis tool helps to provide innovative patterns in data.

b. Pattern Extraction

Pattern extraction starts from an initial set of measured data and builds derived patterns. It reduces the number of resources, describes an extensive collection of data, and creates a better interpretation.

c. Pattern Selection

When the input data is too big to be processed, it can be transformed into a reduced set of patterns. Selected patterns contain the relevant information from the input data. The follow-up process can be performed by using this reduced data.

4) Step 4: Interpret Results

If the value of the test data is different from the null hypothesis, this alternative is accepted.

IV. DISCUSSIONS AND ANALYSES

The experiment collects information about WeChat communication. By using in-depth packet analysis, data filtering, and displaying associated data with i2 Analyst's Notebook, the operation is capable of converting the captured packets into practical statistical information.

A. Setting the Hypothesis

- 1) **H0:** Investigators cannot find the stable tool marks in the WeChat voice communication of network packets.
- 2) **H1:** Investigators find the stable tool marks in the WeChat voice communication of network packets.

B. Experiment Procedures

The experiment scenario is composed of the following steps:

1) Step 1: Understand the Background

The investigator installs Wireshark in the laptop, executes it to sniff packets without using the WeChat application to create a baseline for the current background noise of the network.

2) Step 2: Call and Sniffer the Packets

The IP address of the laptop is known to be '192.168.43.87.' The mobile phone initiated a voice call function in WeChat to call the computer. The investigator observes the change of packets filtering after the call began. The investigator stopped Wireshark and saved the packets for later analysis when the call ended.

3) Step 3: Export the Collected Data

The researchers expand all fields of packets and export packets dissections to CSV files.

4) Step 4: Data Analysis

The researchers open the CSV file in Excel and filter the data. We find out the pattern of WeChat voice calls and tool marks.

C. Interpret Results

The researchers reject the Hypothesis H0 and discover some tool marks of WeChat voice calls. Distribute the VoIP communication into three stages: (1) Signalling Negotiation, (2) Communication Call, and (3) Closing Call. The proposed rules for finding the patterns are listed below:

1) The IP Address of the Victim Laptop Appears in the Source Field

In Table 2, the packet rules of the WeChat application illustrate that the victim's laptop IP address appears in the source IP address field. The destination IP addresses belong to the suspect and the WeChat Company.

a. Emergence: WeChat Company's IP address

In stages 1 and 3, there will emerge the WeChat Company's IP address through TCP protocol or sometimes HTTP protocol. The IP address of the WeChat Company is '203.205.xxx.xxx.' By using the Whois Lookup domain search, we can discover that the IP address belongs to Tencent Building, Kejizhongyi Avenue, Hi-tech Park, Nanshan District, Shenzhen, where the headquarters of WeChat Company is located. The flags of sniffed packets are '0x4000.'

b. Emergence: Six Times of the Suspect's IP Address

In stage 2, there are many UDP and TCP protocol packets. The suspect's IP address will emerge six times through UDP protocol and sometimes RTCP protocol. The flags of UDP and RTCP protocol are '0x0000.'

c. Time to Live: 218

The value of time to live is '128' in every packet when the victim's laptop IP address show in the source field.

TABLE 2. TOOL MARK OBSERVATION ON DESTINATION IP ADDRESS

Patterns Stages	Protocol	Destination IP address	Flags	Time to live
1. Signalling Negotiation	TCP	WeChat company's IP address	0x4000	128
	HTTP*			
	SSL*			
2. Communication Call	TCP	The suspect's IP address	0x0000	
	UDP			
	RTCP*	WeChat company's IP address	0x4000	
	SSL*			
3. Closing Call	TCP	WeChat company's IP address	0x4000	
	HTTP*			
	SSL*			

Note: *The packets of the protocol will emerge randomly.

2) The IP Address of the Victim Laptop Appears in the Destination IP Address

In Table 3, the packet rules of the WeChat application illustrate that the victim's laptop IP address appears in the destination IP address field. The source IP addresses belong to the suspect and the WeChat Company.

a. Emergence: WeChat Company's IP address

In stages 1 and 3, there will emerge WeChat Company's IP address through TCP protocol or sometimes HTTP, HTTP/XML protocol. The flags of sniffed packets are '4x000.'

b. The flags in Suspect's IP Address: 0x0000

In stage 2, there are many UDP and TCP protocol packets. The suspect address will appear through the UDP protocol. The flags are '0x0000.'

c. The flags in ICMP protocol: 0x0000, 0x0000

In stage 3, there are sometimes packets through ICMP protocol, which includes the message of the suspect and the victim. The flags are '0x0000, 0x0000.'

TABLE 3. TOOL MARK OBSERVATION ON SOURCE IP ADDRESS

Patterns Stages	Protocol	Source IP address	Flags
1. Signalling Negotiation	TCP	WeChat company's IP address	0x4000
	HTTP*		
	HTTP/XML*		
	SSL*		
2. Communication Call	TCP	The suspect's IP address	0x0000
	UDP		
	SSL*	WeChat company's IP address	0x4000
3. Closing Call	TCP		
	HTTP*		
	HTTP/XML*		
	SSL*		
	ICMP*	The suspect's IP address	0x0000, 0x0000

Note: *The packets of the protocol will emerge randomly.

D. The Observation on Experiment Results

1) The Hypothesis

Since the experiment requires a user to answer the call, investigators have better to use the victim account to call/sniff the target in a controlled environment. The result of the hypothesis testing is listed in Table 4.

266

ICACT2020 February 16 ~ 19, 2020[illegible]

Figure 2. The Network Topology of WeChat VoIP

V. CONCLUSIONS

Keeping a cybercrime investigation technique up to date is a difficult task. Communication applications are constantly changing, and the follow-up methods for collecting all possible network packets are affected. If tool marks reside in the captured packets, it means that the investigator should identify its provenance and provide context as to who, when, how, and whether it has been obtained from the victim's laptop. This study has demonstrated the way to find the tool marks in the WeChat communication. The interpretation of such information will be helpful for LEAs to identify or prosecute cybercriminals. Our future researches will use AI and machine learning to find more patterns on various communication applications.

ACKNOWLEDGMENT

This research was partially supported by the Executive Yuan of the Republic of China under the Grants Forward-looking Infrastructure Development Program (Digital Infrastructure-Information Security Project-109).

REFERENCES

- [1] Årnes, A, *Digital Forensics*, John Wiley & Sons Ltd, pp. 12-50, 2018.
- [2] Androulidakis I. I., *VoIP and PBX Security and Forensics: A Practical Approach (2nd Edition)*, Springer International Publishing, pp.75-99, 2016.
- [3] Baxter, J. H., *Wireshark Essentials: Get Up and Running with Wireshark to Analyze Network Packets and Protocols Effectively*, Packet Publishing, pp. 7-142, 2014.
- [4] Bernik, I., *Cybercrime and Cyberwarfare*, John Wiley & Sons Inc., pp. 1-57, 2014.

REFERENCES

- [1] Årnes, A, *Digital Forensics*, John Wiley & Sons Ltd, pp. 12-50, 2018.
- [2] Androulidakis I. I., *VoIP and PBX Security and Forensics: A Practical Approach (2nd Edition)*, Springer International Publishing, pp.75-99, 2016.
- [3] Baxter, J. H., *Wireshark Essentials: Get Up and Running with Wireshark to Analyze Network Packets and Protocols Effectively*, Packet Publishing, pp. 7-142, 2014.
- [4] Bernik, I., *Cybercrime and Cyberwarfare*, John Wiley & Sons Inc., pp. 1-57, 2014.

- [1] Årnes, A, *Digital Forensics*, John Wiley & Sons Ltd, pp. 12-50, 2018.
- [2] Androulidakis I. I., *VoIP and PBX Security and Forensics: A Practical Approach (2nd Edition)*, Springer International Publishing, pp.75-99, 2016.
- [3] Baxter, J. H., *Wireshark Essentials: Get Up and Running with Wireshark to Analyze Network Packets and Protocols Effectively*, Packet Publishing, pp. 7-142, 2014.
- [4] Bernik, I., *Cybercrime and Cyberwarfare*, John Wiley & Sons Inc., pp. 1-57, 2014.

- [1] Årnes, A, *Digital Forensics*, John Wiley & Sons Ltd, pp. 12-50, 2018.
- [2] Androulidakis I. I., *VoIP and PBX Security and Forensics: A Practical Approach (2nd Edition)*, Springer International Publishing, pp.75-99, 2016.
- [3] Baxter, J. H., *Wireshark Essentials: Get Up and Running with Wireshark to Analyze Network Packets and Protocols Effectively*, Packet Publishing, pp. 7-142, 2014.
- [4] Bernik, I., *Cybercrime and Cyberwarfare*, John Wiley & Sons Inc., pp. 1-57, 2014.

- [1] Årnes, A, *Digital Forensics*, John Wiley & Sons Ltd, pp. 12-50, 2018.
- [2] Androulidakis I. I., *VoIP and PBX Security and Forensics: A Practical Approach (2nd Edition)*, Springer International Publishing, pp.75-99, 2016.
- [3] Baxter, J. H., *Wireshark Essentials: Get Up and Running with Wireshark to Analyze Network Packets and Protocols Effectively*, Packet Publishing, pp. 7-142, 2014.
- [4] Bernik, I., *Cybercrime and Cyberwarfare*, John Wiley & Sons Inc., pp. 1-57, 2014.

- [1] Årnes, A, *Digital Forensics*, John Wiley & Sons Ltd, pp. 12-50, 2018.
- [2] Androulidakis I. I., *VoIP and PBX Security and Forensics: A Practical Approach (2nd Edition)*, Springer International Publishing, pp.75-99, 2016.
- [3] Baxter, J. H., *Wireshark Essentials: Get Up and Running with Wireshark to Analyze Network Packets and Protocols Effectively*, Packet Publishing, pp. 7-142, 2014.
- [4] Bernik, I., *Cybercrime and Cyberwarfare*, John Wiley & Sons Inc., pp. 1-57, 2014.

- [5] Bhavani, Y., Janaki, V., and Sridevi, R., "Survey on Packet Marking Algorithms for IP Traceback," *Oriental Journal of Computer Science & Technology*, Vol. 10, No. 2, pp. 507-512, 2017.
- [6] Deepthi, A. P., "A Survey on IP Traceback Techniques," *International Research Journal of Engineering and Technology (IRJET)*, Vol. 4, No. 10, pp. 1643-1644, 2017.
- [7] Deri, L., "Open Source VoIP Traffic Monitoring," *5th System Administration and Network Engineering Conference*, The Netherlands, May 15-19, 2006.
- [8] Gao, F. and Zhang, Y., "Analysis of WeChat on iPhone," *Advances in Intelligent Systems Research*, Vol. 68, Singapore, pp. 278-281, 2013.
- [9] Heartpence, B., *Packet Guide to Voice over IP: A system administrator's guide to VoIP technologies*, O'Reilly Media, pp. 1-120, 2013.
- [10] Jonathan, J., *Machine Learning and Future Engineering for Computer Network Security*, Ph.D. thesis, Queensland University of Technology, pp. 17-64, 2017.
- [11] Kävrestad, J., *Guide to Digital Forensics: A Concise and Practical Introduction*, Springer, pp. 43-147, 2017.
- [12] Kobezak, P. D., *Frequent Inventory of Network Devices for Incident Response: A Data-driven Approach to Cybersecurity and Network Operations*, Thesis, Virginia Polytechnic Institute and State University, pp. 58-80, 2018.
- [13] Lv, Y., Duan, Y., Kang, W., Li, Z., and Wang, F. Y., "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 2, pp. 865-873, 2014.
- [14] Saxena, P. and Sharma, S. K., "Analysis of Network Traffic by using Packet Sniffing Tool: Wireshark," *International Journal of Advance Research, Ideas and Innovations in Technology*, Vol. 3, No. 6, 2017.
- [15] Silla, C., *WeChat Forensic Artifacts: Android Phone Extraction and Analysis*, Purdue University Purdue e-Pubs, pp. 52-162, 2015.
- [16] Sun, J. R., Shih M. L., and Hwang M. S., "A Survey of Digital Evidences Forensic and Cybercrime Investigation Procedure," *International Journal of Network Security*, pp. 497-509, 2015.



Da-Yu Kao is an Associate Professor at the Department of Information Management, College of Police Science and Technology, Central Police University, Taiwan. He is responsible for various recruitment efforts and training programs for Taiwan civil servants, police officers, or ICT technicians. He has an extensive background in law enforcement and a keen interest in information security, ICT governance, technology-based investigation, cyber forensics, human resource development, and public sector globalization. He was a detective and forensic police officer at Taiwan's Criminal Investigation Bureau (under the National Police Administration). With a Master's degree in Information Management and a Ph.D. degree in Crime Prevention and Correction, he had led several investigations in cooperation with police agencies from other countries for the past 20 years. He can be reached at camel@mail.cpu.edu.tw.



Ting-Chi Wang is a student at the Department of Information Management, College of Police Science and Technology, Central Police University, Taiwan.



Fu-Ching Tsai is an Assistant Professor at the Department of Criminal Investigation, Central Police University, Taiwan. He received the M.S. and Ph.D. degrees in Institute of Information management from National Cheng Kung University, Taiwan, in 2005 and 2012, respectively. His research interests include data mining, text mining, digital forensics, social network analysis, and cyber criminology.