Analyzing the Security Implications of Popular Video Conferencing Applications in Classified Environments: A Comprehensive Network Traffic Review

Lai Pin Cheng
Forensics and Cybersecurity Research Center, FSeC
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
tp060190@mail.apu.edu.my

Ng Li Sheng
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
tp060612@mail.apu.edu.my

Julia Juremi
Forensics and Cybersecurity Research Center, FSeC
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
julia.juremi@apu.edu.my

***Abstract*—The security concern of video conference applications has been a rising issue, especially when artificial intelligence (AI) technology is advancing drastically. AI technologies, such as deepfakes and AI voice-over, and eavesdropping can be threatening when being performed by hackers or third parties in top-level government video conferences. This paper aims to review the network traffic of several commonly used video conferences applications and to determine the security of them when being used in classified environments.**

Keywords—network traffic analysis, video conference, cyber security

I. Introduction

The popularity of video conferencing has been growing rapidly in this modern era for both businesses and individuals and is expected to have an ever-growing market size for the ease of use and convenience it brings to the table (Castillo, 2023). However, the security of video conferencing has been questionable as technologies such as deepfake and voice-over advance along with the fast-growing capability of artificial intelligence (Miller, 2023). This report documents the tracking of the network traffic that happened during the video conference using different applications such as Vidcall, Google Meet, Zoom, Microsoft Teams, Skype, and Webex. All video conferencing sessions done in this report are conducted between two devices within the same local area network in Malaysia.

II. Aim and Objectives

The aim of this paper is to determine the security of multiple video conferencing applications by reviewing the location of network packets source and destination IP addresses. To harvest the desired outcome from this paper, a number of objectives have been meticulously outlined:

1. To capture and examine the IP address of the network packets generated during video conference sessions.
2. To review the location of the IP addresses extracted from the network packets.
3. To compare and contrast each of the video conferencing applications based on the locations involved in the generated network packets.

III. Literature Review

A. Video Formats (YUV, H264)

A camera is an important component in video conferencing for image capturing. The images captured by cameras are in raw data format taken straight from the sensors of the camera. The raw data format will then be processed into the more commonly known formats such as PNG and JPEG. YUV is one of the raw and uncompressed image formats that is outputted by the cameras. The color model of YUV is made up of the luma (Y) and two chrominance (UV) components, which was used by televisions back in the days to support both color and monochrome televisions (Sarbolandi, 2013). The raw image or video in YUV format is large in file size and is not ideal for real-time video conferencing. Therefore, a compressed video format like H.264 is used for the video data transmission during video conferencing. H.264 standard features both an encoder to perform video frame prediction, transformation and encoding process to produce a compressed H.264 bitstream while a decoder to perform decoding, inverse transformation, and video frame reconstruction to output sequential video frames. H.264 format is widely used for video transmission and storage in the telecommunication field such as video conferencing, video streaming, and television (Richardson, 2007).

B. Secured Real-time Transport Protocol

The Secured Real-time Transport Protocol (SRTP) is extremely significant and solely used for real-time data transmission, which uses Voice over IP technology (VOIP) since real-time video conferencing applications are growing significantly. SRTP is closely working over Real-time Transport Protocol (RTP) which is responsible for undergoing encryption among the process of video or audio transmission in the real-time media stream. SRTP commonly using AES encryption which is a symmetric key encryption technique to encrypt the video or audio payload with 128 or 256-bit keys in order to prevent unauthorized listening or sneaking, particularly crucial for private

conversations during medical consultation or business (Sen et al., 2020). A type of keyed hash function which is known as HMAC-SHA1 will be generated for each packet in SRTP to verify the identity of the sender to prevent potential disruptions and impersonation.

By leveraging AES encryption and HMAC-SHA1 features, SRTP enhances the security of a video conferencing system by safeguarding against replay attacks. This is achieved by detecting and either blocking or identifying outdated packets in order to thwart potential denial of service attempts. As remote work increases significantly, the people will be conducting meetings through video conferencing applications more frequently. Thus, the real-time encryption of SRTP plays a vital role to ensure the sensitive discussion for the business or health sector to be protected. In addition, it will act as a shield against common threats such as information tampering, eavesdropping and impersonation by leveraging the power of encryption, integrity, and authentication checks in SRTP (Digital Samba, 2023).

C. General Video Conferencing Procedure

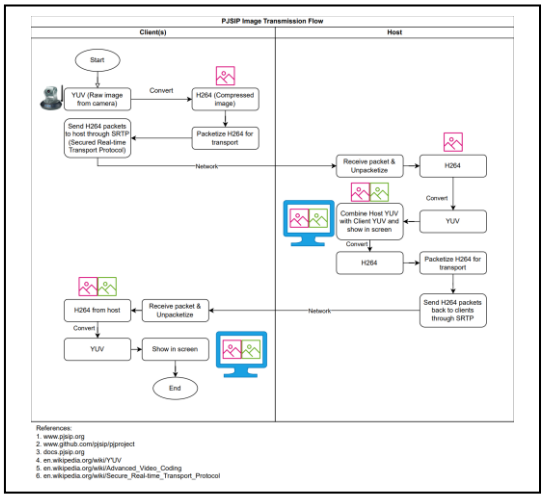

Figure II.a: PJSIP Image Transmission Flow

As a reference for the general video transmission procedure, the working mechanism of PJSIP library is studied and made into a simplified flowchart. PJSIP library is a free and open-source library developed for multimedia communication with the implementation of protocols such as SIP, SDP, RTP, STUN, TURN, and ICE. It features API for high level multimedia communication that is convenient and compatible for most type of systems (PJSIP, 2012).

The flow of the image during a peer-to-peer call using PJSIP library is as shown in the image above (Figure II.a). In this case, the host will act as a media server that handles the video media processing. The camera source of client PC will be captured in the format of YUV and is converted to H264 format before sending out to host through SRTP as encrypted data. The host will receive the H264 packet from the client, decrypt it, convert it back to YUV and combine it with host PC camera source before showing on the screen. Then, the host will convert the YUV back to H264 for the STRP encrypted transmission back to the client. Lastly, the client will receive the H264 packet from the host and convert it to YUV to show on the screen.

On the other hand, a media server is usually required for the hosting of meeting calls when the number of participants is large as the workload for video processing is also larger. In this case, the media server will act as the host while the original meeting host will be considered as a client as shown in the image transmission flow.

IV. Problem Statement

A. Deepfake Technology

Deepfake technology involves the usage of generative artificial intelligence (GAI) model, which is also known as Generative Adversarial Networks (GANs) to manipulate an audio or video footage, usually featuring an influential individual, and replace with the face and voice of another person or modifying the original statement from the individual. Simply put, deepfakes are artificially modified media files that often serve the purpose of manipulating the intention of the original media files. As the encrypted data of video conferences are to be decrypted on the hosting side or media server for processing, there is a possibility for the decrypted data to be leaked. With that, modification such as deepfakes can be done to the leaked data or footages. Deepfakes pose a cyber threat to not only the social media and entertainment industries but also to the boardroom of higher management and government. As a way of social engineering attacks, deepfakes can be used to create false public announcements or statements that are misleading and can cause havoc to the society (Miller, 2023). For example, fake news from a deepfake video featuring footages of boardroom members of a public listed company taken from a high management video conference could result in fluctuated stock prices.

B. Personal Data Protection Law of Malaysia

According to Personal Data Protection Act (PDPA) 2010 Malaysia, any personal data processed outside of Malaysia is not applicable to be protected under the act (Personal Data Protection Act 2010, 2010). With that being said, video conferencing applications may require media data to be sent to a designated media server to be processed and redirected to the participants in the conference. In most cases, users are unable to select the location of the media server to be used and is up to the application itself and the availability of the media server prepared by the application provider. Therefore, it is not guaranteed that the data is to be processed within Malaysia and stay protected under the act in the case of any data breaching.

V. Methodology

A. Data Collection

In this review, Wireshark will be used primarily as a tool for network traffic tracking. Wireshark is a free and open source network packet analyzer that can capture live packet data from network interfaces and provide in-depth analysis features such as protocol information, packet filtering and packet searching (Wireshark, 2008).

To be standardized, all network traffic and packets are generated by conducting video conferences using two laptops that are connected to the same local area network

TABLE I. SUMMARY OF COMPARISON

	Vidcall	Vidcall	Google Meet	Zoom	Webex	Skype	Microsoft Teams
Call type	Peer-to-peer Call	Video Conference (Meeting Call)					
Scenario	Two laptops in the same local area network						
Media Server Used	No	Yes	Yes	Yes	Yes	Yes	Yes
Server Location	-	Johor, Malaysia	United States	United States	United States	Singapore, Japan	Japan
Server IP Address	-	124.13.45.205	2001.4860:4864:6:4000::13	2a06:98c1:52::3	170.72.88.241	2603:1046:c01:2488::2 & 20.63.155.173	52.112.182.7

*The information is based on the data and network packets collected in this paper and may not necessarily be identical in all scenarios.

with access to the public Internet, where one of the laptops will be the meeting host while the other laptop serves as the meeting participant. The video conferencing applications involved in this study are Vidcall, Google Meet, Zoom, Microsoft Teams, Skype, and Webex.

B. Data and Packet Analysis

From the captured packets, the IP addresses involved will be examined to locate the source and destination of them. Websites used for the IP address lookup function are **WhatIsMyIPAddress.com** for IPv4 addresses and **DNSChecker.org** for IPv6 addresses to check the location. Finally, the findings will be summarized and compared based on the location of the network traffic source and destination.

VI. FINDINGS AND DISCUSSIONS

Based on the IP addresses of network packets captured from Wireshark, the summary of results is recorded in Table I. It shows that Google Meet, Zoom, and Webex use media servers located in the United States. For one video conference session, Skype uses multiple media servers located in both Singapore and Japan. Microsoft Teams uses media servers located in Japan. Vidcall is the only video conferencing application that uses media server located in Malaysia during a video conference.

With that said, users of Google Meet, Zoom, Webex, Skype, and Microsoft Teams will not have their data protected under the PDPA due to the foreign location of their media servers. On the other hand, using Vidcall will ensure that user's data is processed in Malaysia and remain protected under the PDPA.

It is important to highlight that Vidcall also offers a direct call feature that can connect two clients in a private peer-to-peer call without routing the network traffic to any third-party server but directly between the two clients. The media processing would all be done on the devices. This method of calling should be preferred for highly confidential communications for maximum network privacy and security.

VII. CONCLUSION AND RECOMMENDATION

In the setting of this paper, two laptops under the same local network with access to the public Internet are used to gather network traffic and packets generated by the video conference sessions conducted with different video conferencing applications such as Vidcall, Google Meet, Zoom, Microsoft Teams, Skype, and Webex. In regard to the paper findings, Vidcall is the only application that hosts its media servers in Malaysia while the other mentioned applications are having media servers outside of Malaysia, which are either in the United States, Singapore, or Japan.

A limitation of this paper is that the number of devices involved is limited and confined to one local area network, which may not necessarily reflect the scenario of a real video conference session. Besides, the paper only points out the potential risk of deepfake technology to video conferencing applications without implementing or testing deepfake as a third party in the video conference session with these applications.

Future studies or review papers should consider allocating additional devices in multiple locations or even countries to simulate a real international video conference scenario so that the results can be more generalizable, convincible, and applicable for larger companies and authorities that span across the globe.

ACKNOWLEDGEMENTS

The authors would like to thank all school of computing and FSeC members who were involved in this study. The opportunity to the early access of Vidcall provided by Advanced Product Design Sdn. Bhd. was greatly appreciated as the paper was made possible and valuable with the involvement and data of Vidcall.

REFERENCES

Castillo, L. (2023, December 16). *Must-Know Video Conferencing Statistics [Current Data]*. Retrieved from gitnux.org: https://gitnux.org/video-conferencing-statistics/

Digital Samba. (2023, November 30). *digitalsamba*. Retrieved from Securing Real-Time Communication: An In-Depth Exploration of the SRTP Protocol: https://www.digitalsamba.com/blog/srtp-protocol

Miller, M. (2023, September 27). *Deepfakes: Real Threat.* Retrieved from kpmg.com: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/deepfakes-real-threat.pdf

Personal Data Protection Act 2010. (2010, June 2). Retrieved from www.kkd.gov.my: https://www.kkd.gov.my/pdf/Personal%20Data%20Protection%20Act%202010.pdf

PJSIP. (2012, May 22). *About PJSIP*. Retrieved from www.pjsip.org: https://www.pjsip.org/about.htm

Richardson, I. (2007). White Paper: An Overview of H.264 Advanced Video Coding.

Sarbolandi, H. (2013). Simultaneous 2D and 3D Video Rendering.

Wireshark. (2008, June 24). *Chapter 1. Introduction.* Retrieved from www.wireshark.org: https://www.wireshark.org/docs/wsug_html_chunked/ChapterIntroduction.html

APPENDIX

A. *Vidcall (Video Conference)*

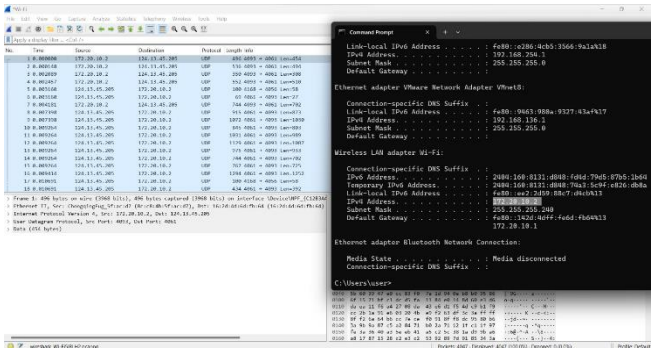

Figure Appendix.A.a: Vidcall Network Packets (Video Conference)

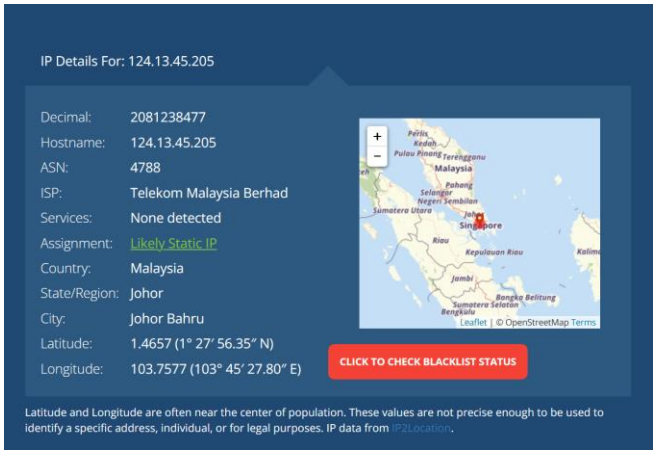

Figure Appendix.A.b: Location of Vidcall Media Server IPv4 Address

The UDP traffic is direct communication between the meeting host and the media server located in Johor, Malaysia

B. *Vidcall (Direct / Peer-to-peer Call)*

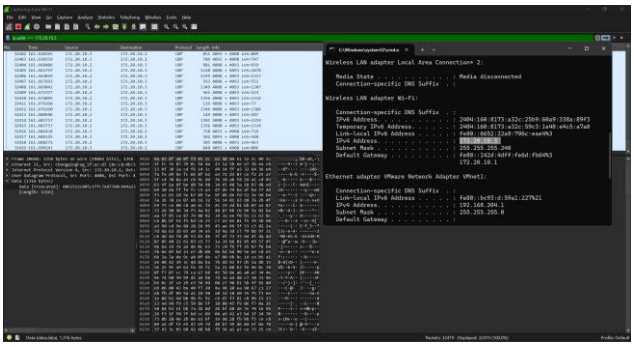

Figure Appendix.B.a: Vidcall Network Packets (Direct / Peer-to-peer Call)

The UDP traffic is direct communication between the call host and the client as both IP addresses are private.

C. *Google Meet*

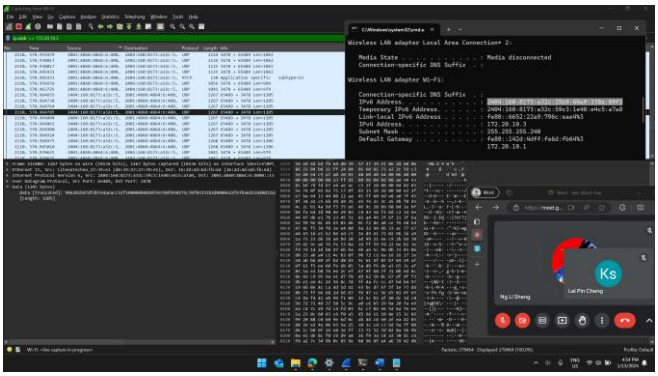

Figure Appendix.C.a: Google Meet Network Packets

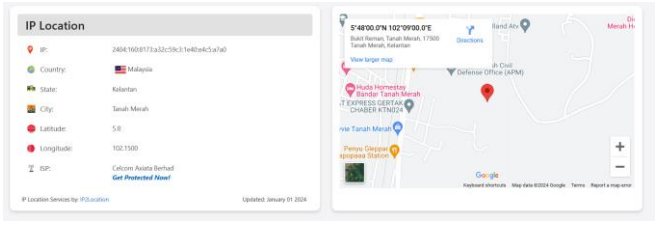

Figure Appendix.C.b: Location of Meeting Host Public IPv6 Address

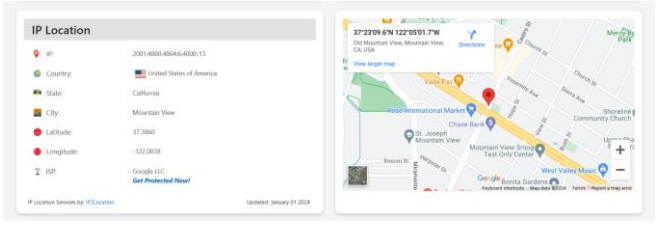

Figure Appendix.C.c: Location of Google Meet Media Server Public IPv6 Address

The UDP traffic is between clients in Malaysia and the media server located in United States.

D. Zoom

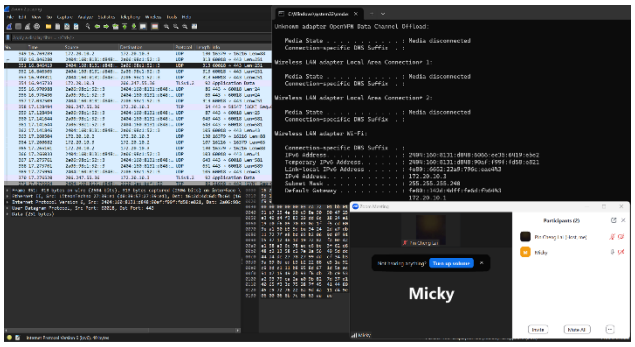

Figure Appendix.D.a: Zoom Network Packets

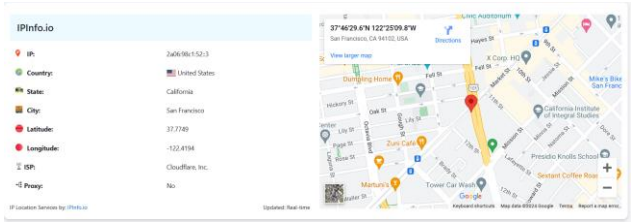

Figure Appendix.D.b: Location of Zoom Media Server IPv6 Address

Based on the captured UDP packets, it seems that most of the communications are direct communication between the two devices. However, it is noted that TCP handshake and packets containing application data are being sent occasionally to Zoom located in the United States.

E. Webex

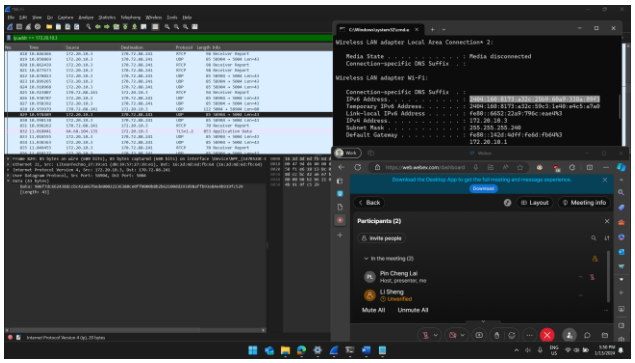

Figure Appendix.E.a: Webex Network Packets

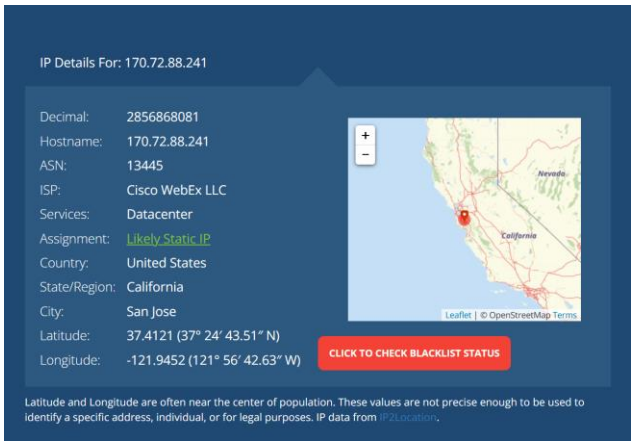

Figure Appendix.E.b: Location of Webex Media Server Public IPv4 Address

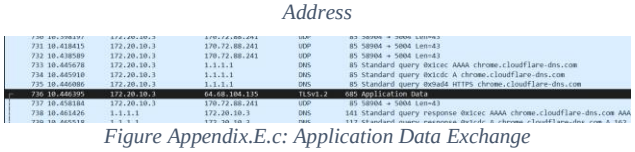

Figure Appendix.E.c: Application Data Exchange

The captured UDP packets shows that the communication is between the devices and Cisco data centre located in United States. It is noted that application data are exchanged occasionally with the server.

F. Skype

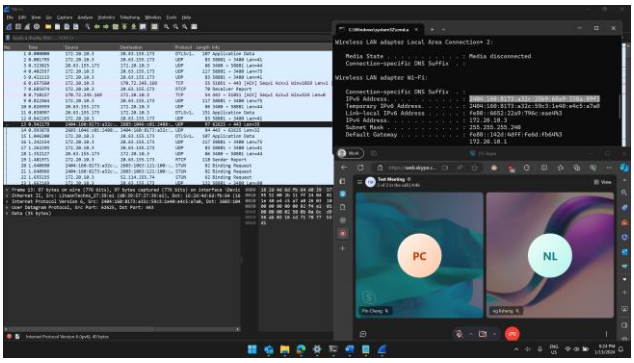

Figure Appendix.F.a: Skype Network Packets

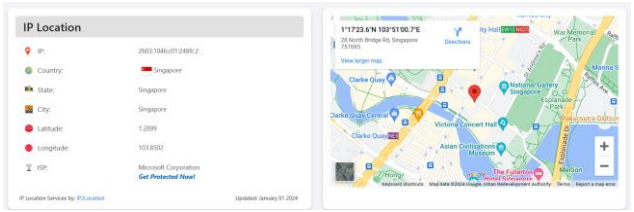

Figure Appendix.F.b: Location of Skype Media Server Public IPv4 Address 1

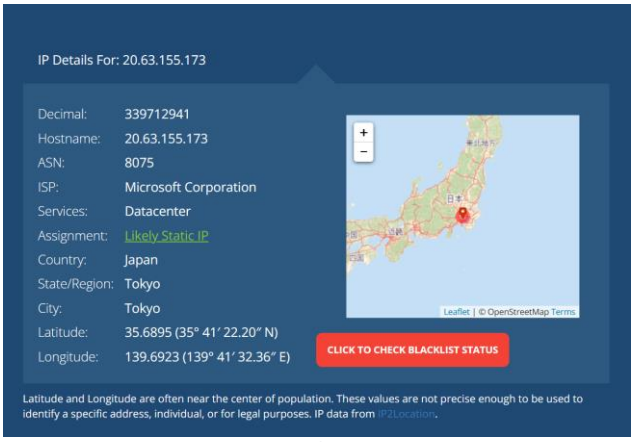

Figure Appendix.F.c: Location of Skype Media Server Public IPv4 Address 2

The captured UDP packets shows that the traffic during the video conference has reached two different countries, which are Singapore and Japan.

G. Microsoft Teams

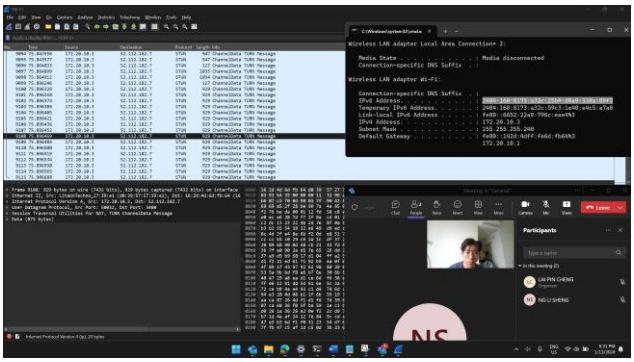

Figure Appendix.G.a: Microsoft Teams Network Packets

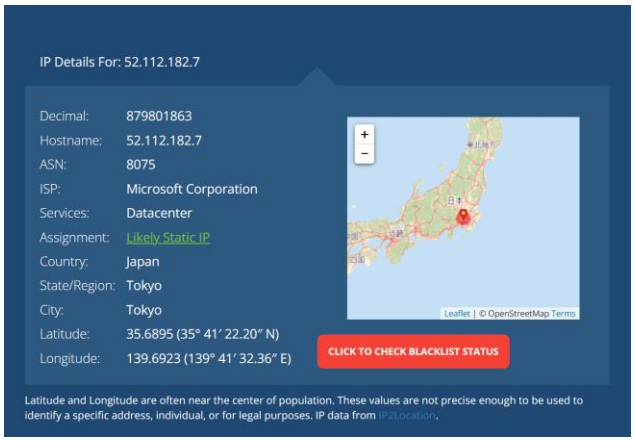

Figure Appendix.G.b: Location of Microsoft Teams

When using Microsoft Teams for video conferencing, the STUN UDP traffic seems to travel to Japan based on the IP address that the device is communicating with.