Signature Identification and User Activity Analysis on Whatsapp Web through Network Data

Ramraj S (ramrajs@srmist.edu.in)
SRM University https://orcid.org/0000-0002-8322-9921
Usha G
SRM Institute of Science and Technology

Research Article

Keywords: WhatsApp Web, Network traffic, User activities, Encryption, Pattern identification, Read Receipts, SSL encryption, Network Security, Traffic Analysis, Signature Identification

Posted Date: May 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-498274/v1

Signature Identification and User Activity Analysis on Whatsapp Web through Network Data

Ramraj S[1*]
Research Scholar, Department of Computer Science and Engineering
SRM IST, Kattankullathur

Dr Usha G[2]
Associate Professor, Department of Software Engineering
SRM IST, Kattakullathur

Abstract

WhatsApp messenger is a popular instant messaging application that employs end-to-end encryption for communication. WhatsApp Web is the browser-based implementation of WhatsApp messenger. Users of WhatsApp communicate securely using SSL protocol. Encryption and use of common port for communication by multiple applications poses challenge in traffic classification for application identification. It is highly needed to analyze the network traffic for the purpose of QoS, Intrusion Detection and application specific traffic classification. In this paper, we have done traffic analysis on the network packets captured through data transfer in whatsapp web. In the result, we have explored the user activities such as message texting, contact sharing, voice message, location sharing, media transfer and status viewing. Packet level traffic analysis of user activities reveal patterns in the encrypted SSL communication. This pattern is identified across SSL packet lengths for WhatsApp media transfer and voice message communication. Other important features WhatsApp is the ability to view the status of the message being sent. We have identified the read and unread message status in these data packets by exposing signatures in the network layer. These signatures are identified with the help of the SSL lengths in the TLS header information of WhatsApp Web network traffic traces. Various other information on WhatsApp traffic presented in our study is relevant to the version of WhatsApp Web v0.3.2386.

Keywords - WhatsApp Web; Network traffic; User activities; Encryption; Pattern identification, Read Receipts, SSL encryption, Network Security, Traffic Analysis, Signature Identification

1.Introduction:

With social media being one of the most important services of the Internet, WhatsApp outshines its competitors such as WeChat, Viber, Hangout, Telegram and Skype in the world of instant messaging applications with over 300 million users everyday [1]. It is known for its availability in multiple platforms and responsive user interface with various features as mentioned in [2]. It supports instant text messaging, voice calls, video calls, and media sharing like images, videos, documents. WhatsApp also provides a web interface known as WhatsApp Web [3] that can be accessed using any web browser on a desktop. WhatsApp Web gives a seamless connection between the browser and the messenger application in the smartphone by providing secure and reliable advantages with amazing keyboard shortcuts. WhatsApp Web allows users to access WhatsApp on various platforms like desktop or laptop with the help of a web browser. This reduces the burden of handling multiple devices for those who predominantly work on computers. Being the most widely used instant messaging application all over the world, WhatsApp definitely presents an opportunity to conduct research on its network traffic. Communication between users in the Internet is made possible through transfer of network packet data, which contains packet header information and the message payload. It can be brought to light that WhatsApp provides end-to-end encryption [2] with SSL (TLS) protocol. Encryption ensures that only the sender and recipient has access to what they have exchanged using WhatsApp. The communication flow of the WhatsApp Web in the network layer can be examined with the help of the network traffic. After the addition of end-to-end encryption, WhatsApp has revamped their network security protocols making it resistant against intrusion. However, the paper brings out the possible information that can be obtained by examining the network traffic of the WhatsApp Web application between a sender and a receiver. One of the characteristics that can be retrieved from the network traffic is the payload content. This can be utilized to derive a signature using payload-based traffic classification. Since the WhatsApp traffic is encrypted, analysis of the network layer payload content does not reveal any pattern. To overcome this situation presented by SSL encryption, an approach based on inspection of packet length is considered in this paper, to have insight into the working of WhatsApp Web.

Also, WhatsApp gives the user the ability to see the status of a chat message being exchanged by providing three statuses [3], namely, (i) Message successfully sent (single check mark in grey color in grey color) (ii) Message successfully delivered to the recipient's phone (two check marks in grey color) (iii) Recipient has read the message (two check marks in blue color). This proves to be an advantage as the user who sends a message (text/media) to a person or a group has the means to see whether the message is delivered or whether it has been read by the recipient(s). This functionality is provided to both the WhatsApp mobile application and WhatsApp Web. However, WhatsApp also offers the user the option of turning off the read receipt in which case the sender cannot see if the message has been read by the recipient. With WhatsApp being the most popular messaging application and known for its renowned end-to-end encryption of messages, this provides an opportunity to examine the strength of its security on its read receipts. By identifying a signature to differentiate the read receipts, it is possible to

demonstrate an inadequacy in the encryption presented by WhatsApp in terms of leakage of information. Enterprises and Institutes can reproduce the methodology formulated in this paper to help impose policies for efficient network monitoring of WhatsApp Web. Along with network traffic pattern analysis, this paper exhibits an extensive study of WhatsApp Web's network traffic by considering traffic trace information, primarily the SSL length, to arrive at a signature that could possibly classify WhatsApp's read receipts.

2.Key Contributions:

2.1) Considering the need for research in whatsapp web data, this paper proposes a study on various user activities of WhatsApp Web and signature analysis of read receipts.

2.2) For our study, user activities such as text-messaging, media sharing, contact sharing, location sharing, voice messaging and viewing statuses, which are cited as "user activities" throughout the paper is included.

2.3) The whatsapp web traffic network dataset is generated in real-time for various scenarios which contains only SSL data packets.

2.4) The generated data is captured through wireshark v2.4.4 snipping tool and is preprocessed.

2.5) Pattern of SSL packet length is analysed to identify significant information about WhatsApp Web user activities.

2.6) From the TLS record layer length field, dimensions such as read and unread messages from network packet level is explored.

3. Paper organization:

The paper first highlights the related work on network traffic analysis and signature identification in Section 4, then proceeds with the procedure of traffic generation, capture and filtering in Section 8. Section 8 and Section 9 discusses the analysis of the captured network traffic. The algorithmic process of the traffic pattern identification and signature verification is explained in Section 6 followed by the conclusion and references.

4. RELATED WORK

One of the most important aspects of network management has been network traffic classification. With the skyrocketing growth of the Internet, the composition of network traffic is becoming complex and varied. Accordingly, network traffic classification has come a long way from the traditional methods of port-based classification to the Deep Packet Inspection method to flow-based statistical method to host-based behavior classification. Introduction of encryption and obfuscation techniques has resulted in a significant rise in encrypted network applications and traffic in network [15]. According to the Dornger in [16], encrypted network traffic classification can be studied by considering classification based on packet information or based on host/social (popularity of the host which is the number of hosts communicating with it) information.

With the popularization of the Internet and high speed networks, network traffic is being generated at an enormous rate. The introduction of encryption has helped in securing the user's

personal information and maintaining their privacy.Consequently, a lot of research is being carried out to test the credibility of encrypted traffic. One way is to identify a signature in the network traffic that corresponds to a particular application or an user activity of an application. Network traffic consists of header information like ports, IPs, payload length, protocol, etc which are currently used to classify applications [1]. Identification of signature in the payload length is a technique commonly used to classify an application with high accuracy [2][3][4]. Some researchers found a way to automate the process of signature identification by finding a string or hex subsequence in the payload [5][6]. [7] proposed a behaviour signature for internet traffic identification that assessed the first few packets of multiple traffic flows. This signature was proved to be successful by obtaining a precision of 100% in categorizing ten different applications. [8] was able to discern SSL/TLS traffic using signature matching methods.

Instant Messaging applications are a primary medium for communication on the Internet. Thus, significant research is being done to identify signatures in these applications to validate their encryption and privacy features. [9] displayed a dependable framework to identify Viber traffic over the network layer and also went on to classify voice/video calls and chat messages (voice/text/media) using signatures found in TCP payload sizes. [10] has provided a comprehensive
study that investigates information leakages by analyzing packet sizes. Operating System (iOS/OSX) of devices used has been identified based on packet lengths and direction pairs. Language used and plaintext length of the message has been characterized by examining the count of packet lengths and direction pairs. Finally, the study also distinguishes user actions with the help of signatures that have been found in the packet lengths. [11] implements a system to meticulously
identify Skype traffic over the network by revealing signatures with regard to the UDP flows of Skype. Skype VoIP calls are detected by finding a signature in the port usage and payload size of the packets in a session [12]. [13] shows that the behaviour of WhatsApp voice calls can be studied by analyzing the WhatsApp traffic between users. An algorithm was also developed to characterize voice calls from other applications such as text messaging or media sharing which considered UDP flow information and traffic time series to perform the classification.

This study is centralized on WhatsApp Web network traffic to identify signatures of read receipts based on the SSL packet length field in TLS header. It also exposes WhatsApp's end-to-end encryption by revealing some user information in the network layer.

There are many researches on the encryption aspect of instant messaging applications to validate their secured communication methodologies. TCP payload sizes have been used to distinguish Viber traffic through a framework, which classifies voice/video calls, and chat messages over the network layer by Sudozai et al [17]. Scott et al. [18] used packet size, direction pairs and count of packet lengths of iMessage (Apple's messenger application) network traffic to perform various classifications like OS of device, language used and plain text length. Jayeeta et al. [19] presented a case study on Google Hangouts by studying the domain name servers (DNS) and further using it to classify the Google application traffic as Gmail, Hangout or Google plus. This is done by following a packet-based identification (ports and packet lengths). Alshammari et al. [21] used a flow-based feature set consisting of 22 features such as protocol, duration of flow, number of packets in forward direction, mean forward packet length to identify VoIP traffic of

Gtalk and Skype. The authors performed this classification with the help of three machine-learning algorithms, namely, C4.5, AdaBoost and Genetic Programming and concluded that C4.5 had the best classification performance. Wongyai et al. [22] conducted traffic analysis of packets that are generated when the Facebook homepage is being loaded. After investigating these packets, the authors summarized the number of components loaded and the order they are loaded in, number of TCP streams used and the number of servers accessed while the Facebook homepage is being loaded.

Recently, new studies are being conducted on the widely known WhatsApp application since there is not a lot of research work present. Yanjie Fu et al. [23] tried to classify user activities of WeChat and WhatsApp based on the packet lengths in the traffic and the time delay between the sessions. Antonio et al. [13] proposed a blind traffic classification for VoIP calls in WhatsApp messenger in android. They have obtained results based on the data rate and protocols that were used by the WhatsApp messenger application.

In this paper, the aim is to study various user activities of WhatsApp Web. These activities include, text-messaging, media sharing, contact sharing, location sharing, voice messaging and viewing statuses, which are cited as "user activities" throughout the paper. It is done through the inspection of the network traffic packets. It also concentrates on the SSL packet lengths to identify any significant information. To the best of authors' knowledge, not much work has been reported in this area.

PROPOSED METHOD

5. Whatsapp User Activity Analysis:

The raw network data is monitored through Wi-Fi port and captured. Then the captured packets are saved as Pcap files through wireshark. The Pcap data contains data of all applications in the network from which the whatsapp data is alone filtered through the SSL header file. Then the network traffic pattern is analyzed.

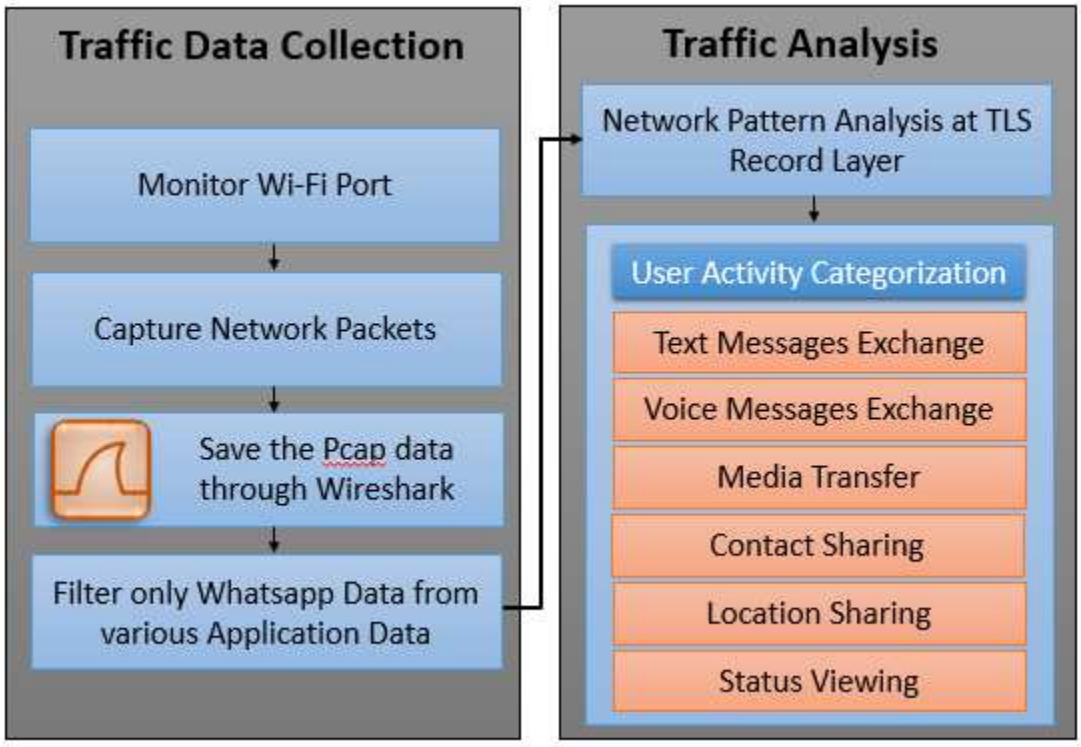


Figure 1: Methodology: Whatsapp User Activity Analysis

6. Signature Identification:

After the signatures that have been identified by traffic analysis, an algorithm is created to help classify the different read receipts from the traffic traces. In order to apply this algorithm, a methodology is followed which is represented by the flow chart given in Figure 2.

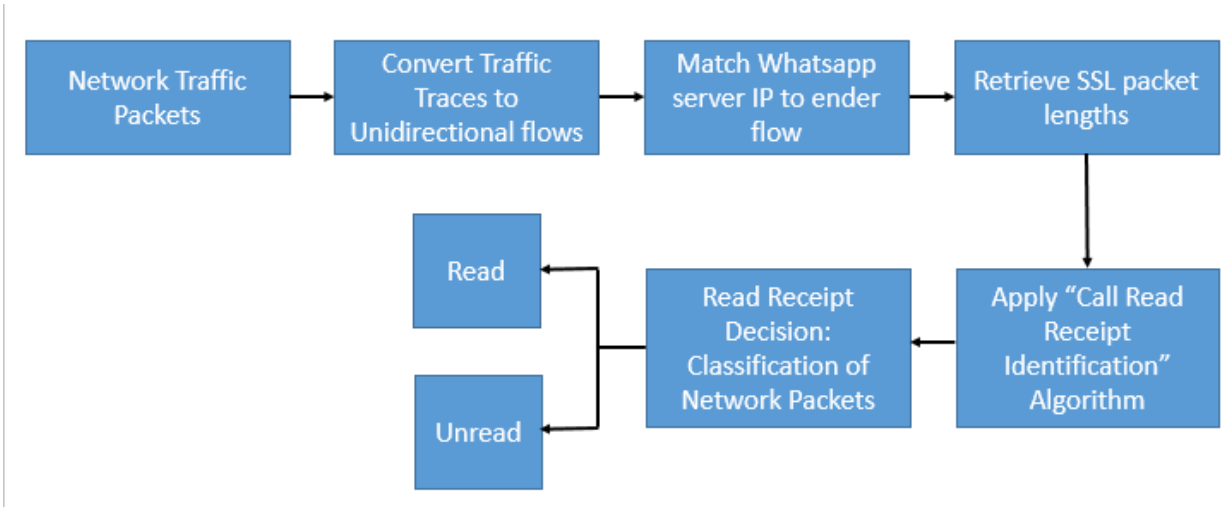


Figure 2: Methodology: Signature Identification

6.1) Unidirectional flow creation: Since the signature is obtained from one direction of the communication (WhatsApp server to sender), traffic traces are converted to unidirectional flows. A flow is a 5-tuple that consists of source IP, source port, destination IP, destination port, and transport protocol.

6.2) Extraction of WhatsApp Server to Sender Flow: Many unidirectional flows are created from each source IP to each destination IP. From these set of flows, only the WhatsApp server (source IP) to sender flow (destination IP) is carefully extracted as the signature is found only in this flow.

6.3) SSL Length retrieval: The unidirectional flows consists of various header information. However, only the SSL lengths are retrieved from the flow because the signature identified based on it.

6.4) Algorithm Implementation: The purpose of this algorithm is to verify whether the SSL lengths from the WhatsApp server to sender flow contain the signature. The retrieved SSL lengths are grouped in an array which is fed as the input to the algorithm. The pseudocode of the algorithm used is shown below:

Algorithm 1. Read Receipt Identification

```
procedure: signature_check(p_lengths)
        for Pi in p_lengths
            if Pi = 40 and Pi+1 = 69, then
                if Pi+2 = 149, then
                    return two check marks
                else if Pi+2 = 150, then
                    return two blue check marks
                else
                    return single check mark
            end if
            if Pi = 47 and Pi+1 = 76, then
                if Pi+2 = 155, then
                    return two check marks
                else if Pi+2 = 156, then
                    return two blue check marks
                else
                    return single check mark
            end if
            i++
        end for
```

The algorithm parses through the array of SSL lengths and compares each value of the array to the identified signatures. When a value of the array matches the first value of either signatures, the subsequent values of the array are checked sequentially against the read receipts signatures. If a match is found for one of the signatures, that particular read receipt is returned by the algorithm.

7. EXPERIMENTAL SETUP

The whatsapp web traffic data is analyzed in the real time data captured as below:

8. TRAFFIC COLLECTION

Availability of data is the basis of any traffic classification research. The focus of this paper is on identification of traffic pattern and signature in WhatsApp Web user activities. Hence, data or traffic traces pertaining to these activities of WhatsApp Web communication is the primary requirement.

8.1)Traffic Generation

The initial requirement is WhatsApp Web network traffic which should be generated without any interference from other applications. The network traffic is introduced by exchanging any message or media with a recipient using WhatsApp Web as the medium. It is made sure that only WhatsApp Web is running in the host machine. No other applications should be active in the background to avoid noise. The version of WhatsApp Web used in this study is v0.3.2390.

The common user activities in whatsapp web are described in Table 1.

TABLE 1: USER ACTIVITIES IN WHATSAPP WEB

Activity	**Description**
Text Messages exchange	Users exchange messages in plain text or emojis
Voice Messages exchange	Users exchange voice message by recording an audio clip.
Media transfer	Users exchange media files like images, documents, videos and audios
Contact sharing	Users share the contacts available in their phonebook.
Location Sharing	Users share any location or a live location using Google

	Maps.
Status viewing	Users view the media statuses uploaded by their contacts.

It is to be noted that calling features like voice call and video call are not available in WhatsApp Web. There are no traffic trace data of WhatsApp user activities available in literature. Therefore, it is required to devise a method for WhatsApp Web traffic generation. The traffic, thus, generated is considered as the ground truth in this paper. In this traffic generation activity, it is ensured that network traffic pertaining to only WhatsApp Web is generated by the machine. This is done by ensuring that no other application responsible for generating network traffic is running on the machine while WhatsApp communication is happening. In addition, apart from WhatsApp Web traffic, there is no browsing traffic is generated. It is also ensured that WhatsApp communication is only between two communicating parties. The version of WhatsApp application in use is mentioned in Table 2.

TABLE 2: WHATSAPP APPLICATION VERSION

Application	**Version**	**Dated**
WhatsApp Messenger (Android)	v2.19.69	January 2021
WhatsApp Web	V0.3.2386	January 2021

The various scenarios based on WhatsApp application type and network connectivity for the traffic generation of the different WhatsApp Web user activities are listed below:

8.1.1) Sender and Receiver using WhatsApp Web on Wi-Fi:

8.1.1.1) Media Transfer: In this scenario, both the sender and receiver exchange media files like documents, images, audios, videos using WhatsApp Web. Both the sender and receiver use Wi-Fi connectivity. It is to be noted that, while exchanging media files, the sender needs to transfer a new file every time because WhatsApp maintains a local cache of files that are already exchanged and does not retransmit it. In case a file that has been sent earlier needs to be resent, it has to be deleted from the conversation chat history in WhatsApp. It is also ensured that only one media file is transferred at a time and text messages are not sent.

8.1.1.2) Text Messages: In this scenario, the sender and receiver exchange only text messages of any length.

8.1.1.3) Contact Sharing: In this scenario, the sender and receiver exchange only contacts.

8.1.1.4) Voice Message: In this scenario, the sender and receiver exchange

only voice messages.

8.1.1.5) Location Sharing: In this scenario, only the sender can send the location using the mobile application as WhatsApp Web does not provide the ability to share locations.

8.1.1.6) Status: In this scenario, the traffic is generated by viewing the status uploaded by any of the user's contacts. It is to be noted that WhatsApp Web does not provide the user the ability to upload media status.

8.1.2) Sender using WhatsApp Web and Receiver using WhatsApp Messenger on Wi-Fi or vice versa:

In this scenario, the traffic is generated in the same way as the first case for all activities with the only difference being that the receiver uses the WhatsApp messenger on a mobile device.

8.1.3) Sender using WhatsApp Web on Wi-Fi, Receiver using WhatsApp Messenger on mobile data or vice versa:

In this scenario, the traffic is generated in the similar way as the first case for all activities with the only difference being that the receiver uses WhatsApp messenger on a mobile device and is connected to the Internet using mobile data instead of Wi-Fi.

For signature identification of these whatsapp data, three different experiments are carried out to analyze the three read receipts as below:

8.1.3.1) Receiver is offline: The sender sends a message or media in WhatsApp Web while the receiver is offline. This gives a single check mark for the message that has been sent.

8.1.3.2) Receiver is online but not chatting with the sender: This experiment is same as the above but the receiver is online and hasn't read the message yet. This gives two check marks for the message that has been sent.

8.1.3.3) Receiver is online and chatting with the sender: In this experiment, the receiver is in the sender's chat when the message is being sent. This makes the message being instantly read by the receiver and gives two blue check marks for the message in the sender end.

8.2) Traffic Capture

The WhatsApp traffic, thus, generated needs to be captured for further analysis. The traffic is always captured at the communicating endpoint, which uses WhatsApp Web irrespective of sender or receiver. Wireshark is a GUI based packet sniffing tool, used to capture the traffic of various user activities of WhatsApp Web, store in traffic

trace files, and analyze the network protocols and various other header information packets. It allows interactive viewing of packet data from live network or from a previously saved capture file. Wireshark's native capture file format for the traffic trace files is pcap format, which is also the format used by tcpdump.

8.3) Automation of Traffic Generation and Capture

In order to generate and capture the WhatsApp communication traffic with minimum manual user intervention, automation of the generation and capture process is devised. This allows huge collection of traffic traces at a shorter period and makes the traffic generation and capture process uniform and less cumbersome. Macro Recorder [15] tool is used to automate the traffic generation process. Macro Recorder is a record and playback automation tool for Windows that records the mouse and keyboard actions and replay them for a specified number of times. The tool is mainly used for software testing and system maintenance but, in this study the tool serves the purpose of automating the traffic generation and collection process. Macro Recorder version 1.0.58f is used in this paper.

8.4) Traffic Filtering

After the generation and capturing of network packets, there is still a possibility for the existence of other packets in the captured traffic trace file that belong to non-WhatsApp applications. This is due to several undetected operating system background applications. This noisy traffic needs to be removed from the captured traffic trace file. Therefore, filtering of traffic is done to ensure that all packets in the captured traffic trace file are relevant to the WhatsApp Web communication only. It is to be noted that WhatsApp uses SSL/TLS protocol for the packet transfer. The filtering of traffic is achieved through Wireshark such that all SSL traffic is retained. In addition, it is confirmed that either of the source IP or destination IP of the packets in the traffic trace file belongs to the pool of WhatsApp server IPs [16]. The communication between a sender and a receiver always occur with a server as a medium. All the traffic traces, generated due to the activities, from the sender is first passed to the server and then it reaches the receiver from the server. This server IP can be used to excerpt the WhatsApp traffic traces from the non-WhatsApp SSL traffic traces for further analysis.

9.RESULT ANALYSIS

The filtered WhatsApp Web traffic is analyzed to identify the presence of patterns in the packet stream and the user read signature.

Traffic pattern analysis:

The WhatsApp Web traffic that has been captured contains only SSL packets as WhatsApp uses SSL for communication. As SSL data is encrypted, inspecting the payload of

SSL packets does not reveal any pattern. The information available in the SSL header for multiple packets of an SSL captured traffic trace and across multiple traffic traces are analyzed. An important parameter in the SSL header is the length field of the TLS Record Layer [17]. This length field for consecutive packets of a WhatsApp capture traffic trace is analyzed. The research findings post traffic analysis is elaborated below:

9.1) Media Transfer and Voice Messages:

Analysis for multiple media transfer and voice message traffic traces has revealed that WhatsApp Web uses TLSv1.2 and TLSv1.3 for these communications. Further inspection of traffic traces using TLSv1.3 protocol revealed a specific signature or pattern across the SSL length field. This SSL length field is cited as "SSL packet length" for the rest of the paper. It is observed that certain sequences of SSL packet lengths always occur in these captured WhatsApp traffic traces. Few of the length sequences have been presented in Table 3 and they form the pattern for the identification of media transfer and voice messages.

The pattern identified always starts with the SSL packet length 69, which occurs between 5th to 13th packets from the WhatsApp SSL client hello packet. This SSL packet length may be followed by the SSL packet lengths 26 and/or 30. x in the pattern represents any single SSL packet length. The regular expression for all possible patterns can be written as:
/69(\,\d{0-4})(\,26){0-2}(\,?\d{0-4})(\,30)?(\,?\d{0-4})/

The regex is written with respect to PCRE format [18] that is a commonly used standard for regular expressions. It is also found that, certain media transfer and voice messages traffic traces use TLSv1.2. However, these traffic traces do not uncover any relevant pattern or other useful information.

TABLE 3: SSL PACKET LENGTH PATTERNS IN MEDIA TRANSFER, VOICE MESSAGE PATTERNS AND STATUS VIEWING

69,x,..
69,26,x,..
69,26,30,x,..
69,26,26,x,..
69,26,26,30,x,..
69,26,26,x,30,x,..
69,x,26,30,x,..
69,x,26,26,30,x,..
69,x,26,26,x,30,x,..
69,x,26,x,30,..
69,30,x,..
69,x,30,x,..
69,x,26,x,..

9.2) Text Messages and Contact Sharing:

Following the analysis of traffic traces of text messages and contact sharing, it is observed that both these activities use TLSv1.2 protocol. There is no presence of significant signatures or patterns in the SSL packet lengths.

9.3) Status:

Upon studying the traffic traces collected for media statuses, it is found that WhatsApp Web uses TLSv1.3 while viewing media statuses and follows the same pattern as mentioned in Table 3. It is to be noted that WhatsApp Web uses TLSv1.2 for statuses that have already been viewed by the user and no pattern is identified in the SSL packet lengths.

9.4) Location Sharing:

It is observed that unlike other user activities that use TLS, WhatsApp Web facilitates location sharing using GQUIC protocol. As the location is shared using Google Maps, the Google servers are accessed instead of WhatsApp. Therefore, the location is transmitted through Google server IPs.

Signature Identification:

WhatsApp is SSL encrypted, hence, the application payload data does not reveal any signature. But, after inspecting the TLS record layer length field in all the capture files, signatures are seen. It is to be noted that the signatures are identified only in the traffic traces using TLSv1.2 protocol. The scenario is illustrated in Figure 3. The packets of the traffic traces are marked in Figure 3 [A, B, C]. They are found to be common in all the WhatsApp capture files with the source being a WhatsApp server and the destination being the sender.

No.	Time	Source	Destination	Protocol	Length	Info
17	4.351354	31.13.79.53	192.168.0.101	TLSv1.2	88	Application Data
19	4.866173	31.13.79.53	192.168.0.101	TLSv1.2	102	Application Data
24	7.068507	31.13.79.53	192.168.0.101	TLSv1.2	88	Application Data
25	7.068837	31.13.79.53	192.168.0.101	TLSv1.2	225	Application Data
53	7.128046	31.13.79.53	192.168.0.101	TLSv1.2	106	Application Data
271	7.914484	31.13.79.53	192.168.0.101	TLSv1.2	349	Application Data
Ⓐ 278	8.164034	31.13.79.53	192.168.0.101	TLSv1.2	99	Application Data
Ⓑ 280	8.721906	31.13.79.53	192.168.0.101	TLSv1.2	128	Application Data
Ⓒ 282	9.326913	31.13.79.53	192.168.0.101	TLSv1.2	209	Application Data
303	23.575007	31.13.79.53	192.168.0.101	TLSv1.2	158	Application Data

Figure 3. Common Traffic Traces

Figure 4 depicts the SSL length for consecutive three packets for the scenario when receiver has seen the image sent by a sender and the sender has received the read receipts in the form of two check marks in blue color. These sequence of packet length form a signature for the read receipt

indicated by two blue check mark This signature always occur towards the end of WhatsApp server to sender traffic trace. In each experiment, two significant signatures are found for each read receipt.

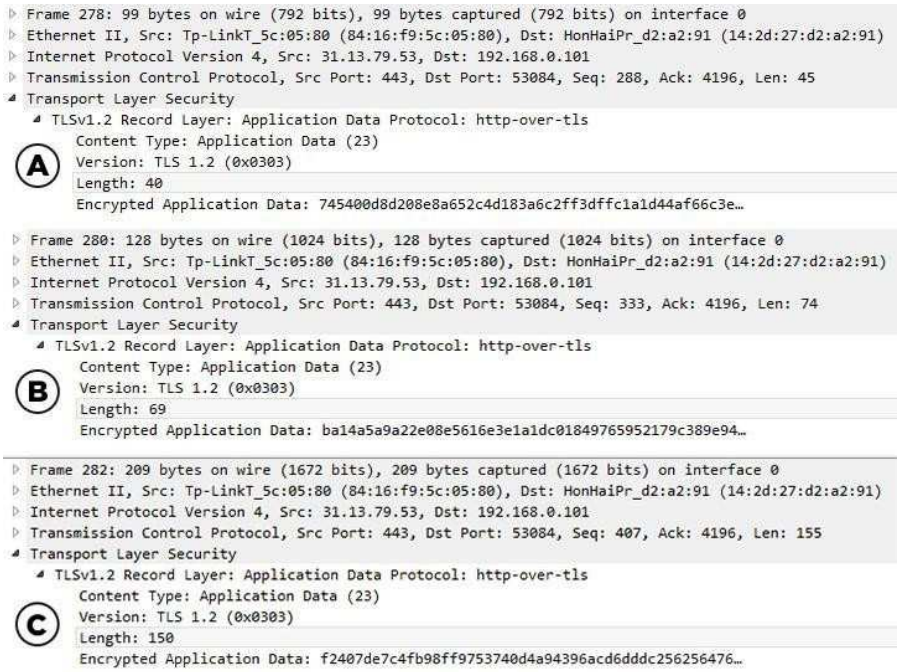


Figure 4. Signature revealed as SSL packet length for two blue check marks in SSL length

TABLE 4: SIGNATURES OF READ RECEIPTS

	Signature 1	Signature 2
Single check mark in grey color (Sent)	40,69	47,76
Two check marks in grey color (Send and Received)	40,69,149	47,76,155
Two Check Marks in blue color (Send and Read)	40,60,150	47,76,156

Both the signatures found during traffic analysis for all three read receipts are tabulated in Table4.

10. CONCLUSION

The research conducted in this paper highlights the presence of pattern in encrypted whatsapp web communication. Network traffic analysis at packet level reveal distinct pattern for various whatsapp web user activities. This pattern is revealed on analysis of the SSL packet lengths for the user activities of media transfer, statuses and voice message in WhatsApp Web. Also, the paper proposes a methodology to identify signatures in SSL lengths to classify the various read receipts provided by WhatsApp. This work can be further extended for whatsapp communication over mobile application as well.

11. DECLARATIONS:

FUNDING

The authors express their sincere gratitude to the SRM Institute of Science and Technology, Tamil Nadu, India for aiding in facilitating the research. Our project is funded by the university in the way of grants such as resources and guidance. For the project, the dataset was derived by the Port Mirroring Technique, which involved the use of a Port Mirroring Switch obtained from the college lab. Resources like stable network, labs, systems etc. were also granted by the college. Throughout the software development cycle and research, the funding was offered by the university. No external funding was involved in this project.

CONFLICTS OF INTEREST/ COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AVAILABILITY OF DATA AND MATERIALS

The datasets generated during and/or analyzed during the current study are available in the https://github.com/nishantuzir/just_a_naive_flowmeter/tree/master/sample

CODE AVAILABILITY -

All code for data cleaning and analysis associated with the current submission is available at https://github.com/nishantuzir/just_a_naive_flowmeter repository. and can be accessed freely by anyone.

References

[1] Yoon, S.H., Park, J.S. and Kim, M.S. (2012). Signature maintenance for Internet application traffic identification using header signatures. In *2012 IEEE Network Operations and Management Symposium,* 1151-1158. http://dx.doi.org/10.1109/NOMS.2012.6212042

[2] Goo, Y.H., Shim, K.S., Lee, S.K. and Kim, M.S. (2016). Payload signature structure for accurate application traffic classification. In *2016 18th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 1-4. https://doi.org/10.1109/APNOMS.2016.7737287

[3] Shim, K.S., Ham, J.H., Sija, B.D. and Kim, M.S. (2017). Application traffic classification using payload size sequence signature. *International Journal of Network Management*, *27*(5). https://doi.org/10.1002/nem.1981.

[4] Lee, S.H., Park, J.S., Yoon, S.H. and Kim, M.S. (2015). High performance payload signature-based Internet traffic classification system. In *2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS),* 491-494.
https://doi.org/10.1109/APNOMS.2015.7275374

[5] Feng, X., Huang, X., Tian, X. and Ma, Y. (2010). Automatic traffic signature extraction based on Smith-waterman algorithm for traffic classification. In *2010 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, 154-158.
https://doi.org/10.1109/ICBNMT.2010.5704886

[6] Park, B.C., Won, Y.J., Kim, M.S. and Hong, J.W. (2008). Towards automated application signature generation for traffic identification. In *NOMS 2008-2008 IEEE Network Operations and Management Symposium*, 160-167. https://doi.org/10.1109/NOMS.2008.4575130

[7] Yoon, S.H., Park, J.S. and Kim, M.S. (2015). Behavior signature for fine-grained traffic identification. *Appl. Math*, *9*(2L), 523-534.

[8] Sun, G.L., Xue, Y., Dong, Y., Wang, D. and Li, C. (2010). An novel hybrid method for effectively classifying encrypted traffic. In *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, 1-5.
https://doi.org/10.1109/GLOCOM.2010.5683649

[9] Sudozai, M.A.K., Habib, N., Saleem, S. and Khan, A.A. (2017). Signatures of viber security traffic. *Journal of Digital Forensics, Security and Law*, *12*(2), 11.
https://doi.org/10.15394/jdfsl.2017.1477

[10] Coull, S.E. and Dyer, K.P. (2014). Traffic analysis of encrypted messaging services: Apple imessage and beyond. *ACM SIGCOMM Computer Communication Review*, *44*(5), 5-11. https://doi.org/10.1145/2677046.2677048

[11] Yuan, Z., Du, C., Chen, X., Wang, D. and Xue, Y. (2014). Skytracer: Towards fine-grained identification for skype traffic via sequence signatures. In *2014 International Conference on Computing, Networking and Communications (ICNC),* 1-5. https://doi.org/10.1109/ICCNC.2014.6785294

[12] Sven Ehlert, Sandrine Petgang. "Analysis and Signature of Skype VoIP Session Traffic". Fraunhofer FOKUS Technical Report NGNI-SKYPE-06b

[13] Cuadra-Sanchez, A. and Aracil, J. (2017). A novel blind traffic analysis technique for detection of WhatsApp VoIP calls. *International Journal of Network Management*, *27*(2). https://doi.org/10.1002/nem.1968

[14] Kumar, N. and Sharma, S. (2016). Survey Analysis on the usage and Impact of Whatsapp Messenger. *Global Journal of Enterprise Information System*, *8*(3), 52-57.

[15] Cao, Z., Cao, S., Xiong, G. and Guo, L. (2012). Progress in study of encrypted traffic classification. In *International Conference on Trustworthy Computing and Services*, 78-86. https://doi.org/10.1007/978-3-642-35795-4_10

[16] Dorfinger, P. (2010). *Real-time detection of encrypted traffic based on entropy estimation*. na.

[17] Nguyen, T.T. and Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE communications surveys & tutorials*, *10*(4), 56-76. https://doi.org/10.1109/SURV.2008.080406

[18] Shen, G. and Fan, L.(2008). Network traffic classification based on message statistics. In *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 1-4. https://doi.org/10.1109/WiCom.2008.1046

[19] Datta, J., Kataria, N. and Hubballi, N. (2015). Network traffic classification in encrypted environment: a case study of google hangout. In *2015 twenty first national conference on communications (NCC)*, 1-6. https://doi.org/10.1109/NCC.2015.7084879

[20] Alshammari, R. and Zincir-Heywood, A.N. (2009). Machine learning based encrypted traffic classification: Identifying ssh and skype. In *2009 IEEE symposium on computational intelligence for security and defense applications*, 1-8. https://doi.org/10.1109/CISDA.2009.5356534

[21] Alshammari, R. and Zincir-Heywood, A.N. (2010). An investigation on the identification of VoIP traffic: Case study on Gtalk and Skype. In *2010 International Conference on Network and Service Management,* 310-313. http://dx.doi.org/10.1109/CNSM.2010.5691210

[22] Wongyai, W. and Charoenwatana, L. (2012). Examining the network traffic of facebook homepage retrieval: An end user perspective. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, 77-81. http://dx.doi.org/10.1109/JCSSE.2012.6261929

[23] Fu, Y., Xiong, H., Lu, X., Yang, J. and Chen, C. (2016). Service usage classification with encrypted internet traffic in mobile messaging apps. *IEEE Transactions on Mobile Computing*, *15*(11), 2851-2864. https://doi.org/10.1109/TMC.2016.2516020

Figures

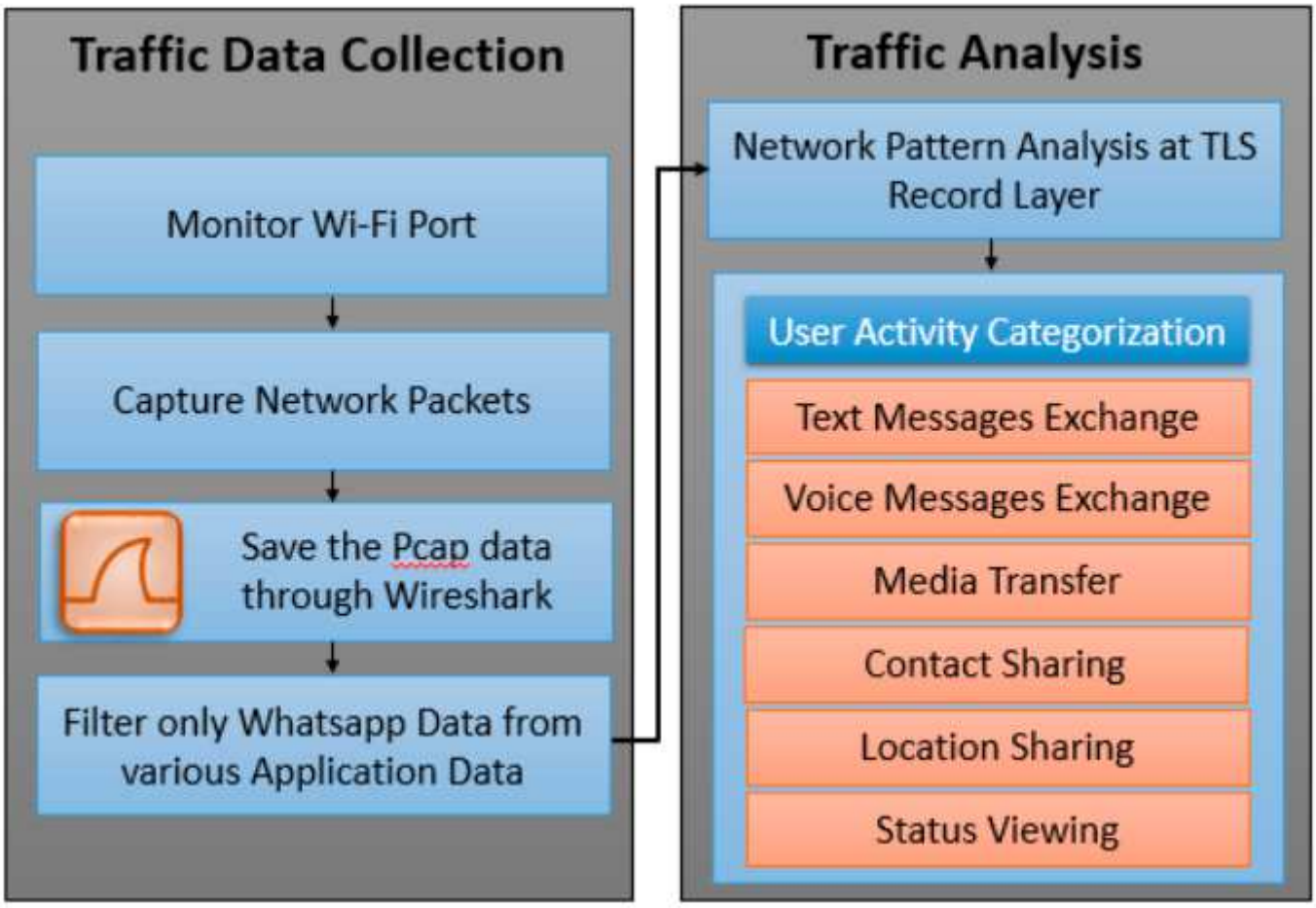


Figure 1

Methodology: Whatsapp User Activity Analysis

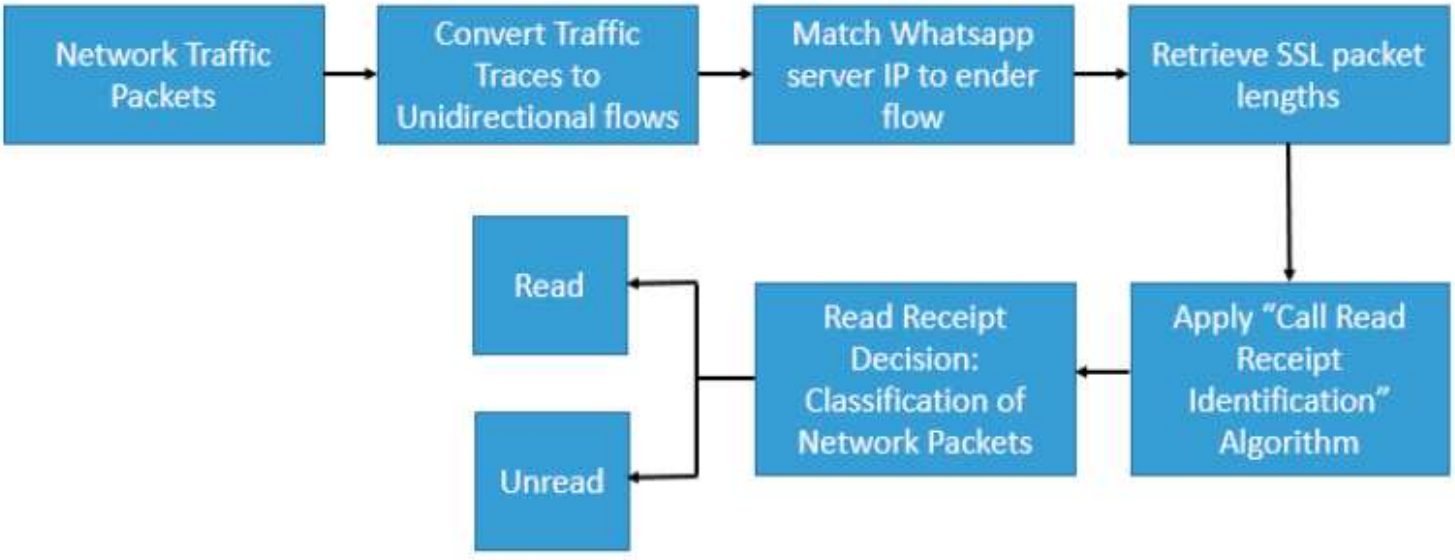
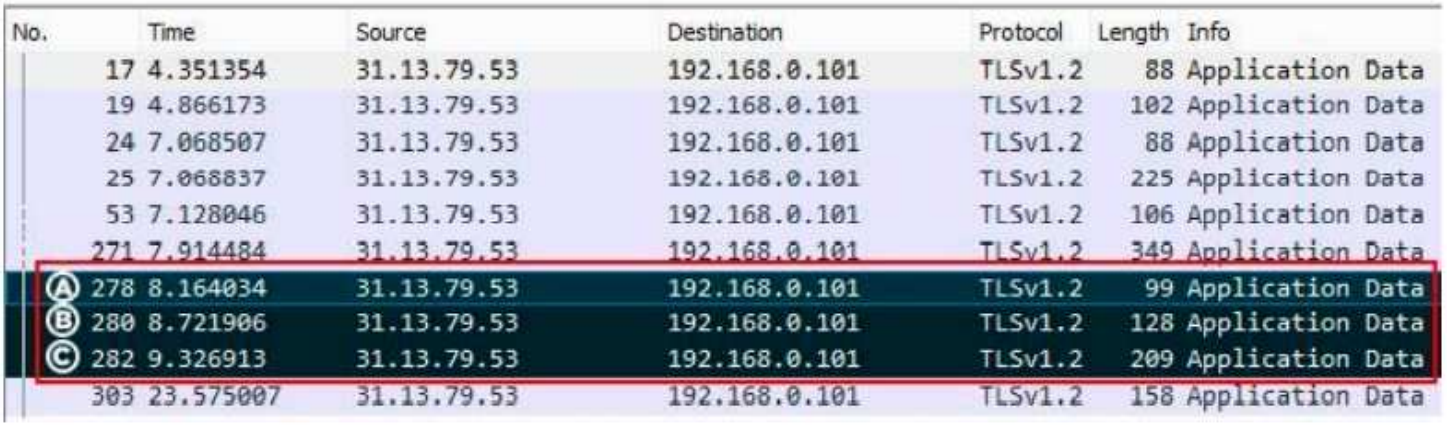


Figure 2

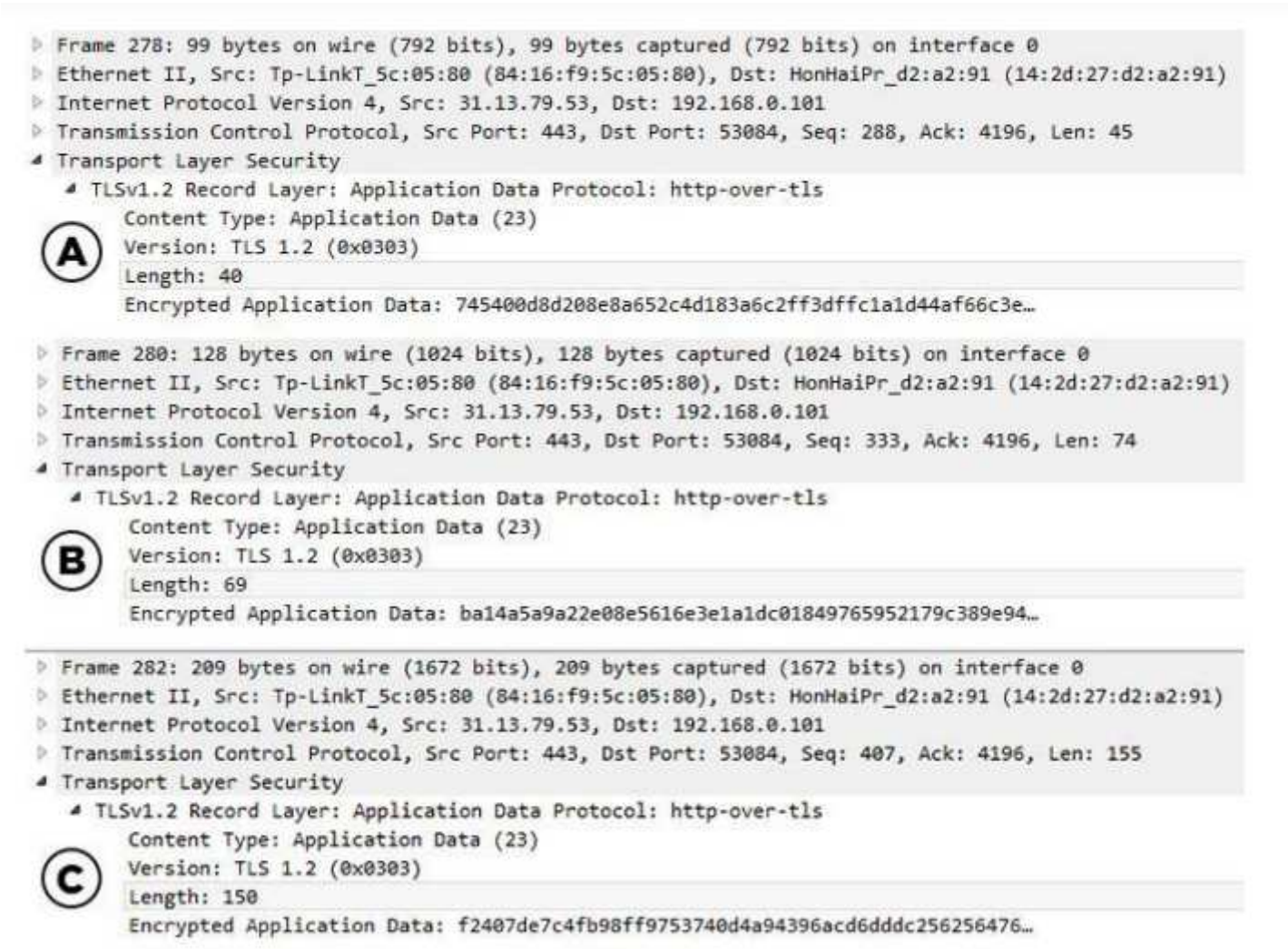
Methodology: Signature Identification



No.	Time	Source	Destination	Protocol	Length	Info
17	4.351354	31.13.79.53	192.168.0.101	TLSv1.2	88	Application Data
19	4.866173	31.13.79.53	192.168.0.101	TLSv1.2	102	Application Data
24	7.068507	31.13.79.53	192.168.0.101	TLSv1.2	88	Application Data
25	7.068837	31.13.79.53	192.168.0.101	TLSv1.2	225	Application Data
53	7.128046	31.13.79.53	192.168.0.101	TLSv1.2	106	Application Data
271	7.914484	31.13.79.53	192.168.0.101	TLSv1.2	349	Application Data
Ⓐ 278	8.164034	31.13.79.53	192.168.0.101	TLSv1.2	99	Application Data
Ⓑ 280	8.721906	31.13.79.53	192.168.0.101	TLSv1.2	128	Application Data
Ⓒ 282	9.326913	31.13.79.53	192.168.0.101	TLSv1.2	209	Application Data
303	23.575007	31.13.79.53	192.168.0.101	TLSv1.2	158	Application Data

Figure 3

Common Traffic Traces



```
▷ Frame 278: 99 bytes on wire (792 bits), 99 bytes captured (792 bits) on interface 0
▷ Ethernet II, Src: Tp-LinkT_5c:05:80 (84:16:f9:5c:05:80), Dst: HonHaiPr_d2:a2:91 (14:2d:27:d2:a2:91)
▷ Internet Protocol Version 4, Src: 31.13.79.53, Dst: 192.168.0.101
▷ Transmission Control Protocol, Src Port: 443, Dst Port: 53084, Seq: 288, Ack: 4196, Len: 45
◢ Transport Layer Security
  ◢ TLSv1.2 Record Layer: Application Data Protocol: http-over-tls
Ⓐ     Content Type: Application Data (23)
      Version: TLS 1.2 (0x0303)
      Length: 40
      Encrypted Application Data: 745400d8d208e8a652c4d183a6c2ff3dffc1a1d44af66c3e…

▷ Frame 280: 128 bytes on wire (1024 bits), 128 bytes captured (1024 bits) on interface 0
▷ Ethernet II, Src: Tp-LinkT_5c:05:80 (84:16:f9:5c:05:80), Dst: HonHaiPr_d2:a2:91 (14:2d:27:d2:a2:91)
▷ Internet Protocol Version 4, Src: 31.13.79.53, Dst: 192.168.0.101
▷ Transmission Control Protocol, Src Port: 443, Dst Port: 53084, Seq: 333, Ack: 4196, Len: 74
◢ Transport Layer Security
  ◢ TLSv1.2 Record Layer: Application Data Protocol: http-over-tls
Ⓑ     Content Type: Application Data (23)
      Version: TLS 1.2 (0x0303)
      Length: 69
      Encrypted Application Data: ba14a5a9a22e08e5616e3e1a1dc01849765952179c389e94…

▷ Frame 282: 209 bytes on wire (1672 bits), 209 bytes captured (1672 bits) on interface 0
▷ Ethernet II, Src: Tp-LinkT_5c:05:80 (84:16:f9:5c:05:80), Dst: HonHaiPr_d2:a2:91 (14:2d:27:d2:a2:91)
▷ Internet Protocol Version 4, Src: 31.13.79.53, Dst: 192.168.0.101
▷ Transmission Control Protocol, Src Port: 443, Dst Port: 53084, Seq: 407, Ack: 4196, Len: 155
◢ Transport Layer Security
  ◢ TLSv1.2 Record Layer: Application Data Protocol: http-over-tls
Ⓒ     Content Type: Application Data (23)
      Version: TLS 1.2 (0x0303)
      Length: 150
      Encrypted Application Data: f2407de7c4fb98ff9753740d4a94396acd6dddc256256476…
```

Figure 4

Signature revealed as SSL packet length for two blue check marks in SSL length