TRAFFIC ANALYSIS OF ENCRYPTED MESSAGING AND NETWORK MONITORING IN VARIOUS SERVICES AND APPS

Ancy Sindhya A
PG student, Department of CSE
Ponjesly College of Engineering,
Nagercoil, Tamilnadu.

Maria Sheeba M
Assistant Professor, Department of CSE
Ponjesly College of Engineering,
Nagercoil, Tamilnadu.

Abstract-The increase in usage of messaging apps enables us to collect the encrypted internet traffic.The classification of network traffic into different types of in-app service usages can help manage bandwidth and provide quality of service. Traditional approach of classification is based on packet inspection such as parsing HTTP headers. A system named CUMMA is developed for classifying service usage of messaging Apps, modelling user behavioural pattern, network traffic characteristics and temporal dependencies. The discriminative features of traffic classification can be extracted based on packet length and time delay. The clustering Hidden Markov algorithm is used for decomposing mixed-dialogs into sub-dialog which enables analyst to identify the service usages and analyse the behaviour of end user for encrypted internet traffic. CUMMA helps the mobile analyst identify the service usage and analyse end user behaviour for encrypted internet traffic, thus improving the effectiveness and efficiency of service usage classification.

***Keywords*:** *Encrypted Internet Traffic, In-App Analytics, Service Usage Classification, Mobile Messaging App, dialog, sub-dialog.*

I. INTRODUCTION

Mobile communication plays an important role in communication that has always been focus on exchanging of information among parties at location physically apart. Initially the mobile communication was limited between one pair of users on single channel pair. The range of mobility depends on the transmitter power, type of antenna used and the frequency of operation. With the increase in the number of users, accommodating them within the limited available frequency spectrum became a major problem. Cellular telephone systems must accommodate a large number of users over a large geographic area with limited frequency spectrum. If a single transmitter/ receiver are used with only a single base station, then sufficient amount of power may not be fixed location. The mobile data communication generally refers to the infrastructure put in place in-order to ensure that seamless and reliable communication. These would include devices such as protocols, services, bandwidth, and portals necessary to facilitate and support the stated services.

II. RELATED WORK

The increased popularity of mobile messaging Apps, such as WeChat[1] and WhatsApps[2] have become the hubs for most activities of mobile users. For example, messaging Apps help people text each another, share photos, chat, and engage in commercial activities such as paying bills, booking tickets and shopping. Therefore, service usage analytics in messaging Apps becomes critical for business, because it can help understand in-App behaviour of end users, and thus enables a variety of applications. For instance, it provides in-depth insights into end users and App performances, enhances user experiences, and increases engagement, conversions and monetization. A key task of in-App usage[2] analytics is to classify Internet traffic of messaging Apps into different usage types. Traditional methods for traffic classification rely on packet inspection by analysing the TCP or UDP [8] port numbers of an IP packet or reconstructing protocol signatures in its payload. People estimate the usage types of traffic by assuming

that messaging Apps consistently transmit data using the same port numbers which are visible in the TCP and UDP[12] headers. 'However, there are emerging challenges for inspecting IP packet content. For example, messaging Apps are increasingly using unpredictable port numbers. Also, customers may encrypt the content of packets. In addition, governments have imposed privacy regulations which limit the ability of third parties to lawfully inspect packet contents. Moreover, many mobile apps use the Secure Sockets Layer (SSL) and its successor Transport Layer Security (TLS) as a building block for encrypted communications.

III. PROPOSED SYSTEM

The data mining solutions for classifying encrypted Internet traffic data can be generated by messaging Apps into different service usage types. Network Traffic Characterization can be carried out using Flow-Based Statistics. Network Monitoring System is used in collecting LAN speed and CPU speed. By reduction of unwanted memory in the cache the speed of LAN and CPU increased. It can be carried out by means of garbage collector which removes unwanted files from the cache memory.

Deployment and adaptability aspects can be achieved through mobile data traffic analysis. Security can be implemented at sender and receiver side. Authenticity is provided at the user side only authorized person can access the services. The number of user, data and time of transaction can be analyzed such that in providing security at receiver and sender side of the network.

3.1 Authentication

In this module, the user login by means of passwords the passwords may be in the form of patterns. In this authenticated user can only login such that providing security at the user side for accessing the service usage. Services provided may be Image, Text, Audio-video file.

3.2 Traffic Segmentation

Traffic segmentation is carried out by two stage segmentation, with these traffic-flows from coarse-grained level named session to fine-grained level named dialog. Session generally start when the user open the App and last until user close it. The generated internet traffic during the session is known as dialog.

3.3 Traffic Feature Extraction

The discriminative features of the network traffic data can be mined from two perspectives:

- Packet Length
- Time delay.

Packet length can be calculated based on bandwidth and size of the packet. The packet length of video stream is larger than the packet length of the text messages. This feature discreteness the distribution of packet length into several equal sized k sub-ranges.

3.4 Usage Type Prediction & Outlier Detection

Outlier detection and handling module helps to handle dialogs which are classified as an unknown mixture of usages. Dialog is classified as a mixture of usages, the objective is to identify the most probable hidden usage mixture. Exploit a clustering HMM to segment mixed dialogs into multiple

consecutive sub dialogs of single-type usage. Hidden Markov Model (HMM) is a generative model that copes with sequential data, it is used to capture temporal dependencies (which produce different result over different period of time) for enhancing classification accuracy. The CUMMA system is used such that identifying the incoming packet length and detecting number of mixed types. It helps to allocated different services with different channel such that in managing the bandwidth and provide quality of service.

3.5 Network Monitoring with Memory management

Network Monitoring System is used in collecting LAN speed and CPU speed. By reduction of unwanted memory in the cache the speed of LAN and CPU increased. It can be carried out by means of garbage collector which removes unwanted files from the cache memory.

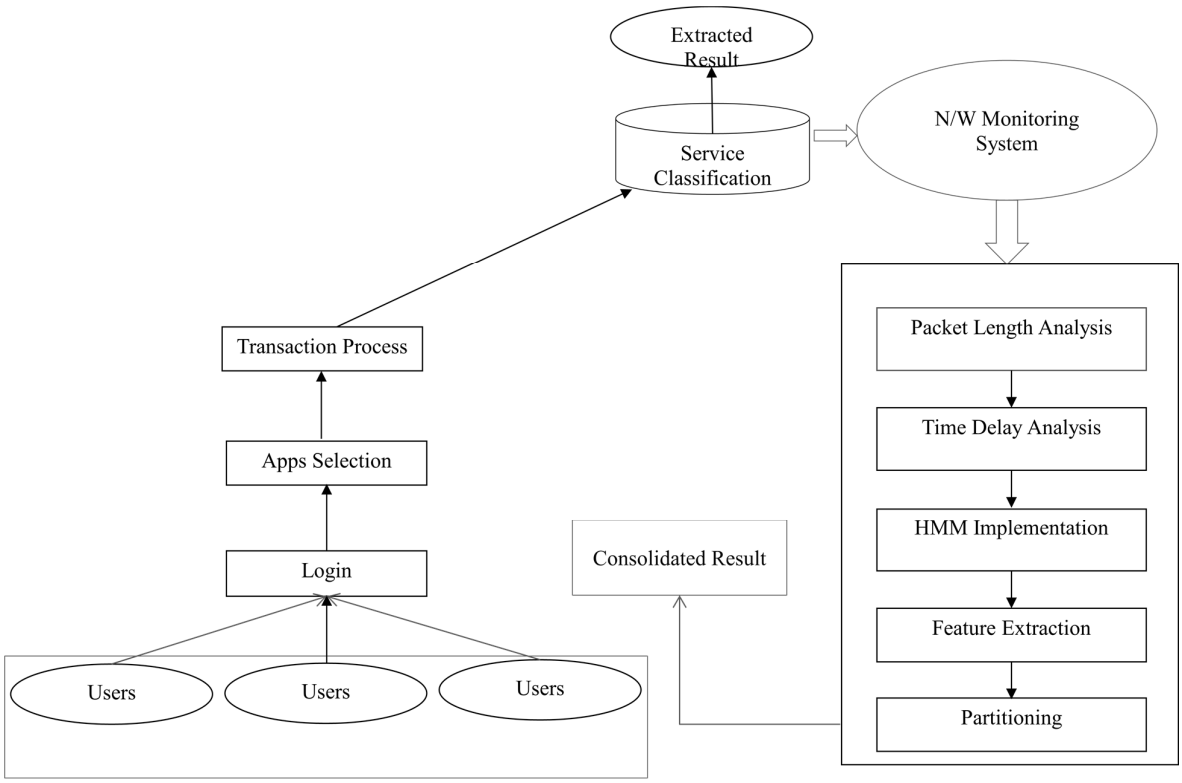

Figure 1: Architecture diagram

The figure represent the system architecture. The security is provided at the user side, security provided in the form of patterns and user login by means of app selection. The transaction process can be carried out like transfer of services like text file, image file and Audio-Video file. The services can be classified based on factors like packet length and Time delay. The Network Monitoring system is used to enhance the CPU and LAN speed. It is used such that it uses the garbage collector to remove the unwanted files from the memory. Based on the packet length and time delay analysis the HMM (Hidden Markov Model) is used which uses CUMMA system , helps in extracting traffic, identify mixed type usage, segment the session into a dialog. Session is initiated when the user open the app and end until close it and the generated traffic during the session is known as dialog. With CUMMA system the

services is split into an equal sub-ranges. Each of the services is partioned with channel 1 for the text file, channel 2 for image file and channel 3 for audio-video file. The services reach the destination with less time delay due to the removing of unwanted files from the memory.

IV. RESULT & DISCUSSION

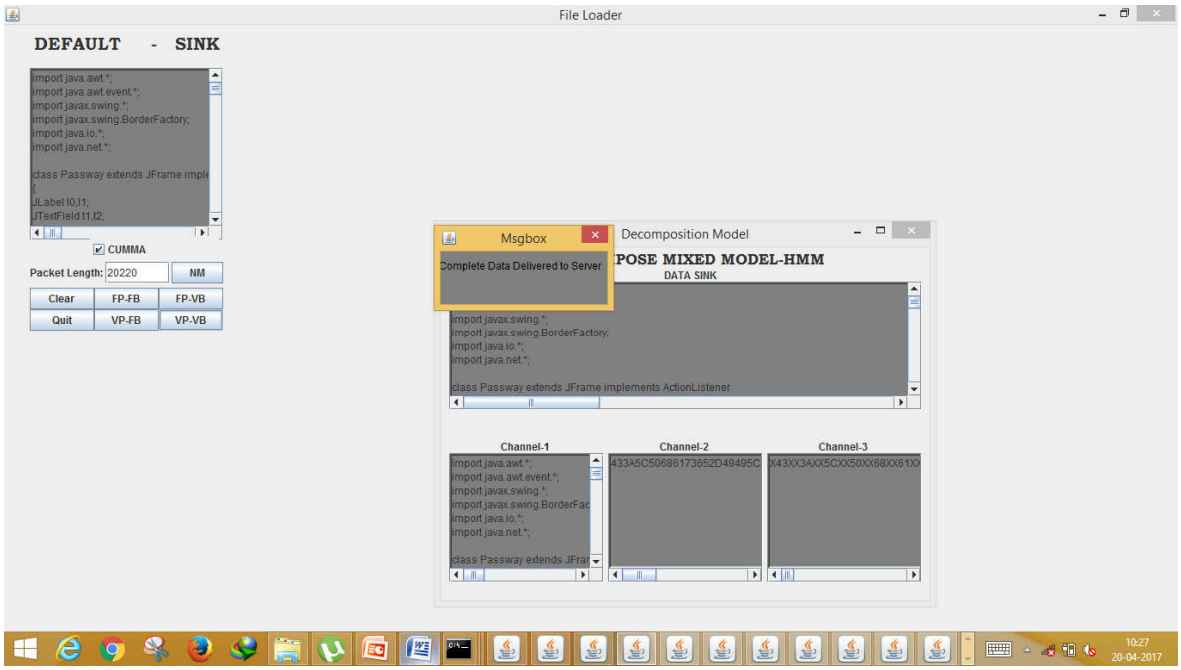

Figure 2: Snapshot of Complete Service Usage Transfer along the partioned channel.

The above snapshot shows the service usage classification along the channel using Hidden Markov Model by means of CUMMA system. Channel1 is allocated for the text file to transfer to the destination, Channel 2 is allocated for the image file to transfer to the destination, Channel 3 is allocated for the Audio- Video file to transfer to the destination.

V. CONCLUSION

Thus a system for classifying service usage using encrypted internet traffic in mobile messaging Apps is jointly modified using the temporal dependencies and traffic characteristics. Segmenting the traffic data, feature is extracted and classified as sub-dialog. Thus the in-App analyst can be able to identify service usage, effectiveness and efficiency of service usage classification by monitoring the network and providing security at the user side.

REFERENCES

[1] Johan Himberg, Kalle Korpiaho, Heikki Mannila, Johanna Tikanmaki and Hannu TT Toivonen, "Time series segmentation for context recognition in mobile devices," In ICDM, 2001.

[2] Sebastian Zander, Thuy Nguyen, and Grenville Armitage, "Automated traffic classification and application identification using machine learning," In The IEEE Conference on Local Computer Networks.

[3] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymbero poulos, Ramesh Govindan, and Deborah Estrin, "Diversity in smartphone usage," In 8th international conference on Mobile ystems, applications, and services 2010.

[4] Athula Balachandran, Vyas Sekar, Srinivasan Seshan, Ion Stoica and Hui Zhang, "A quest for an internet video quality-of-experience metric," In Proceedings of the 11th ACM workshop on Hot Topics in Networks, 2012.

[5] Anindhya Ghose and Sang Pil Han, "An emprical analysis of user content generation and usage behavior on the mobile internet," Management Science, 2011.

[6] Jun Zhang, Yang Xiang, Yu Wang, Wanlei Zhou, Yong Xiang and Yong Guan. "Network traffic classification using correlation information," Parallel and Distributed systems, IEEE Transactions on 2013.

[7] M.Isabel Sanchez "Mobility management: Deployment and adaptability aspects through mobile data traffic analysis," Simula Research Laboratory, Oslo, Norway 2016.