Multi-functional Traffic Generation Framework Based on Accurate User Behavior Emulation

Sándor Molnár[1], Péter Megyesi
High Speed Networks Lab., Dept. of Telecomm. and Mediainformatics,
Budapest Univ. of Technology and Economics, Budapest, Hungary
Email: {molnar, megyesi}@tmit.bme.hu

Géza Szabó
TrafficLab
Ericsson Research, Budapest, Hungary
Email: geza.szabo@ericsson.com

***Abstract*—We demonstrate a multi-functional and configurable traffic generation framework which is based on accurately capturing user behavior by automatic Graphical User Interface control. The framework includes the emulation of typical user behaviors on remote controlled machines with the recording of their generated traffic. As far as we know this is the first time when a traffic generation tool is demonstrated which is able to compose arbitrary traffic mix by real generated and not simulated traffic. Our examples exhibit some of the manifold traffic analysis possibilities as use cases. The demonstrated system is open to the public in [1].**

I. Introduction

Traffic generation is a key function in several network scenarios, e.g. to test Deep Packet Inspection (DPI) techniques. DPI is important for Internet Service Providers (ISP) to properly classify the traffic on their network. DPI systems need the most up-to-date signatures to perform well in real networks. Simulators are not suitable for this task since they are not able to accurately represent real traffic characteristics. Thus a DPI signature database can only be updated by collecting real measurements. On the other hand, the network data is the property of the operator and plenty of privacy issues may arise if a DPI product vendor takes the measurements to its own site to further develop the DPI signature set. Another solution could be the manual triggering of application signatures which is a slow and unscalable process. These challenges motivated our research to create a system which is capable of generating traffic traces for testing purposes of traffic classification systems both in terms of accuracy and performance.

We demonstrate our newly developed tool the User Behavior based Emulator (UBE) which is capable of generating traffic with characteristics which is very close to the real traffic characteristics. It is due to the fact that we do not simulate or generate traffic based on some traffic model but rather we record typical user interactions with several applications on the Graphical User Interface (GUI) and construct application specific usage models which can be used later to emulate user interactions on remote controlled computers. It means that we extract typical user scenarios, e.g., used applications and their share, usage patterns, etc. from real measurements. We can generate traces and build a database when user actions e.g., mouse or keyboard events happened according to the emulated user scenarios. From these database we can construct arbitrary aggregate traffic mix having traffic characteristics very close to the real one. The generated traffic has realistic payload and traffic characteristics both in inter-packet and user level timescales and it has the advantage that it does not contain user sensitive data thus it can be distributed for wide audience.

[1]The research was supported by OTKA-KTIA grant CNK77802.

II. Demonstrated features

Figure 1 presents the architecture of the User Behavior based Emulator. This section presents the features of UBE step-by-step according to Figure 1. Table I summarizes the procedure steps in the framework which allow us to create a high speed aggregated traffic trace. We would like to encourage the research community to contribute our work by sending new network measurements or GUI automation scripts of different applications after testing these functionalities.

A. Real Measurement Processing

The first step in the operation of the framework is processing real traffic measurements. We collected several publicly available trace files which are frequently used by the research community. Due to privacy reasons these measurements contain packet header information only thus UBE uses a port base traffic classifier algorithm for converting them into per user application share in minute resolution (see Figure 1 'Timescale'). Later these reports are converted into string based descriptors introduced in [2]. In this process a character is assigned to every traffic class while minutes are separated by a time delimiter character (see Figure 1 'Transformed data'). UBE can process traffic measurements in both *tcpdump* and

TABLE I: Steps of Generation of High Speed Realistic Traffic

#	Process description
1	tcpdump / netflow log integration
2	traffic classification in minute resolution
3	changing representation format to string descriptors
4	typical pattern search in string input
5	string descriptor conversation into remote control procedure
6	running traffic measurements in remote machines
7	rebuild original measurement with typical patterns
8	merge traffic traces into high speed aggregate

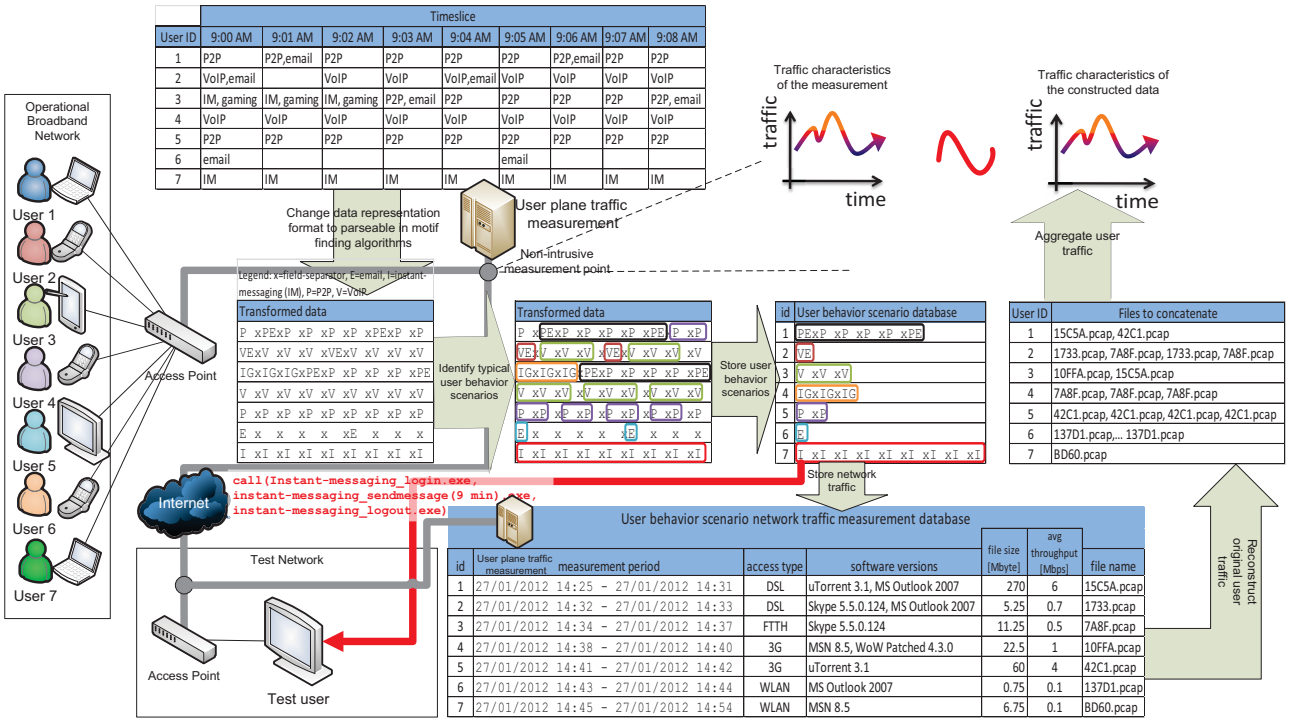


Fig. 1: The Demonstrated System.

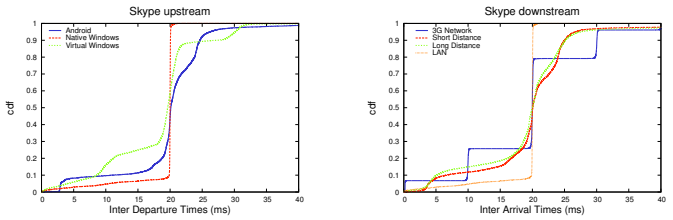
netflow formats. The next step is the extraction of typical user scenarios by searching for frequently occurring fixed time length patterns in the transformed input. Later the new results can be integrated to the 'User behavior scenario database' which is globally defined for all measurements. UBE is able to parse these string descriptors and convert them into a remote controlling procedure. The framework also contains an interface for manual definition of user scenarios.

B. Creation of Traffic Segments

In this step the creation of traffic segments are performed. When new up-to-date validation traffic is needed, the information from the 'User behavior scenario database' is grabbed and user actions are emulated by remote controlling installed test machines with the recorded user interactions. GUI testing tools drive the applications on the client machine and make them to generate real traffic on the network. The generated traffic is recorded and stored in the 'User behavior scenario network traffic measurement database'. We tested the framework on different test machines including several operating systems and geographic locations [3]. Figure 2 presents an example where the same Skype conversation was recorded on different machines. It is shown that different access technologies or operating systems can radically change the Inter Packet Timing information in both upstream and downstream directions.

C. Aggregation of Trace Files

UBE is also able to construct a high-speed aggregated traffic stream using the recorded individual trace files. This feature makes it possible to build up a realistic traffic trace



(a) Skype Upstream Inter Departure Times

(b) Skype Downstream Inter Arrival Times

Fig. 2: Skype Measurement Results

as an aggregate mix with traffic from arbitrary predefined applications. Such traces are hardly achieved in real measurements even if we neglect the privacy issues. UBE is now able to merge multiple *tcpdump* files into one. The merge tool can modify the timestamp and IP address information of the packets creating an output were the individual trace files are played simultaneously and distinct IP addresses are associated to different users. This process does not change the timing information between an individual user's packets and it will follow the same statistics that was recorded during the emulation phase of the given scenario.

References

[1] "User Behavior based Emulator DEMO portal," Dec 2012. [Online]. Available: http://ubetest1.hsnlab.tmit.bme.hu/

[2] P. Megyesi and S. Molnár, "Finding typical internet user behaviors," in *Proc. 18th EUNICE Conference on Information and Communications Technologies*, Aug 2012, pp. 321–327.

[3] S. Molnár, P. Megyesi, and G. Szabó, "Multi-functional emulator for traffic analysis," in *Proc. IEEE ICC*, June 2013.