Ecole Doctorale STIC

Sciences & Technologies de l'Information et de la Communication

ED STIC - Proposition de Sujets de Thèse

pour la campagne d'Allocation de thèses 2011

Titre du sujet :	Internet/Intranet Traffic classification
Mention de thèse :	Informatique
HDR Directeur de thèse inscrit à l'ED STIC :	Urvoy-keller Guillaume

Co-encadrant de thèse éventuel :

Nom :	
Prénom :	
Email :	
Téléphone :	

Email de contact pour ce sujet :	urvoy@unice.fr
Laboratoire d'accueil :	I3S

Description du sujet :

The TCP/IP protocol suite does not enable by itself an accurate identification of the application that generated a particular IP packet. The port number is a valuable piece of information but there is no way to enforce HTTP to use port 80 or 8080 for instance. Still, the knowledge of which application is behind a specific flow is of interest in a number of situations, e.g., an ISP that wants to monitor the trends in traffic to anticipate its link provisioning process. Also, an enterprise that consists in many branches interconnected via the public network (the Internet) and wants to manage the traffic flowing in its tunnels needs a precise traffic classification function, to grant each flow with the correct grade of service in the extended VPN of the company.

In the field of traffic classification, two main approaches have been investigated. First, deep packet inspection, where the payload of packets is inspected for specific strings or data structures that are considered as signatures of applications. Second, statistical approaches where

a pre-labeled trace, i.e. a trace for which one knows the mapping between flows and applications, is used to train a statistical model to recognize an application based on non-payload related information, e.g., packet size and directions [4]. Recent studies [1] have suggested that statistical approaches, though less intrusive, suffer from a number of shortcomings. Hybrid approaches [2], where one makes use of both signatures and classical statistical indicators, appear as a promising avenue but are not mature yet.

In the context of this thesis, we aim at investigating new problems related to traffic identification:
- As a continuation of [1,2], we want to further explore the process of adding features to the classification tool. In addition to adding features, we would like to investigate new applications that might represent a minority of bytes when observing the overall traffic on the long run but might be considered as crucial by the ISP, e.g. streaming or social network traffic. Also, encrypted traffic is of high interest [5].

- So far, all works that rely on statistical approaches use deep packet inspection tools for annotating the pre-labelled trace used to train the classifier. No study has considered the reverse problem of how results of statistical tools can help improving deep packet inspection tools.

- Intranet type of traffic is IP-based but consists of different applications than Internet traffic, e.g., NFS or LDAP [6]. One would like to assess which technique can classify Intranet traffic on the fly.

- Traffic classification might be used to profile groups of users. It can also be used to inform anomaly detection, e.g., abnormal trends in a specific application. We started to investigate users profiling in [3] but this area needs to be further explored.

This work will be partly carried out in cooperation with Orange Lab, Sophia-Antipolis. The candidate should have a solid background in networking and programming.

References :

[1] Pietrzyk, Marcin;Costeux, Jean-Laurent;Urvoy-Keller, Guillaume;En-Najjary, Taoufik
Challenging statistical classification for operational usage : the adsl case
IMC 2009, 9th ACM SIGCOMM Internet Measurement Conference, November 4-6, 2009,
Chicago, IL, USA , pp 122-135

[2] Pietrzyk, Marcin; En-Najjary, Taoufik; Urvoy-Keller, Guillaume; Costeux, Jean-Laurent
Hybrid traffic identification
Report RR-10-238

[3] Pietrzyk, Marcin; Plissonneau, Louis; Urvoy-Keller, Guillaume; En-Najjary, Taoufik
On profiling residential customers
TMA 2011, 3rd International Workshop on Traffic Monitoring and Analysis, 27 April 2011, Vienna, Austria

[4] Thuy T. T. Nguyen, Grenville J. Armitage:
A survey of techniques for internet traffic classification using machine learning.
IEEE Communications Surveys and Tutorials 10(1-4): 56-76 (2008)

[5] Laurent Bernaille, Renata Teixeira:
Early Recognition of Encrypted Applications.
PAM 2007: 165-175

[6] Ruoming Pang, Mark Allman, Mike Bennett, Jason Lee, Vern Paxson, Brian Tierney:
A First Look at Modern Enterprise Traffic.
Internet Measurment Conference 2005: 15-28

English version: