# Proposal: A Tool to Classify P2P Traffic Based on Application Signnatures and SVM

Li Jiaying, Dong Fei

October 16, 2013

## 1 Introduction

The identification and classification of P2P network traffic is drawing more and more attention because it is crucial to a wide range of network operations including capacity planning, service differentiation and rate-limiting. Traditional traffic, mapping to TCP or UDP network port, is easy to detect, while P2P traffic may take too much effort to identify because of incorporating various strategies to avoid detection. In this project, we intend to employ an efficient approach for identifying P2P network traffic through application level signatures combining with a machine learning methods, support vector machine (SVM) algorithm. As detecting application level signature is an intuitive way, it will be fast and won't take much memory resource for identification. Besides, SVM is a very effective approach of classification in machine learning, thus it is a very efficient way of combining two methods to improve accuracy of P2P traffic classification.

## 2 Problem Statements

The problem we focus on is how to ccurately identify and classify P2P network traffic from the whole network traffic. How to implement the tool using application signatures[1] and SVM method[2-3] to a real platform is also a big challenge. Another thing to mention is that, we should train the program to increase its total accuracy.

## 3 Objectives

The objective we develop a tool is to find out the p2p traffic among the whole network flow. The tool just read its input from captured network packages from other program, like WireShark. It will figure out a set of packages, which belongs to P2P application, as the output, out of the whole set of packages.
Finally, we hope our tool for detecting P2P traffic can achive more than 90% accuracy on average.

## 4 Methodology and Plan

In our project, to realize classification of P2P traffic properly, we employ application-level signature identification approach combining with SVM. Application layer signatures for P2P protocols can achieve a high accuracy and robustness classification even if applied to Gigabit Ethernet speeds[1].
We focus on the download phase of P2P protocols to track some signatures. SVM is a kind of classification and regression method which has been described in Vapniks Statistical Learning Theory[4]. To classify network flow, it is necessary to find a hyperplane in a high dimensional feature space, used to split two type of traffic, P2P traffic and Non-P2P traffic. As P2P traffic can be divided into two parts: data and signaling traffic. We try to consider two kinds of traffic with separated weights in SVM methods and use some training examples to construct a splitting hyperplane. After employing application level signature classification, we use SVM methods to classify those who is difficult to classify, aiming at improve the accuracy of classification.

## 5 References

[1]Sen S, Spatscheck O, Wang D. Accurate, scalable in-network identification of p2p traffic using application signatures. Proceedings of the 13th international conference on World Wide Web. ACM, 2004: 512-521.
[2]Liu C, Yang Y, Tang C. A Classification Method of Unstructured P2P Multicast Video Streaming Based on SVM. Multimedia Information Networking and Security, 2009. MINES'09. International Conference on. IEEE, 2009, 1: 68-72.
[3]Wang R, Liu Y, Yang Y, et al. Solving the app-level classification problem of P2P traffic via optimized support vector machines. Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on. IEEE, 2006, 2: 534-539.
[4]Vapnik V. The nature of statistical learning theory. springer, 2000.