

## Semester Thesis:

# 2P File-Sharing Traffic Identification Method Validation and Verification for Roger Kaspar <kasparr@ee.ethz.ch>

## 1 Introduction

### P2P Filesharing: Traffic Identification Method Validation and Traffic Characterisation

This thesis in the context of DDoSVax<sup>1</sup> will continue the work done by Lukas Hämmerle (Master Thesis 2004) and Philipp Jardas (Bachelor Thesis 2004) that is targeted at identification of traffic generated by P2P filesharing systems for the long-term purpose of intrusion detection. The primary goal is passive detection in NetFlow data collected by the DDoSVax project. Identification of P2P users or shared content is explicitly not in the focus of this thesis and no efforts will be made in that direction.

### The DDoSVax Project

In the joint ETH/SWITCH research project “DDoSVax” aggregated network traffic data (Cisco NetFlow) is collected at all border gateway routers of the Internet backbone operated by SWITCH<sup>2</sup>. This data contains information about which Internet hosts were connected to which others and how much data was exchanged over which protocols.

The DDoSVax project provides archived NetFlow data as well as a near real-time framework (named UPFrame) with plug-in support for online processing of NetFlow data received from routers.

## 2 The Task

The overall task is split into two phases. The first phase consists in adding real-time validation to the detection algorithms defined by the thesis of Hämmerle. Real-time validation tries to

---

<sup>1</sup>See <http://www.tik.ee.ethz.ch/~ddosvax/>

<sup>2</sup>See <http://www.switch.ch/>

validate that a host is actually running a specific P2P system before adding it to the pool of identified P2P participants. The second phase is then devoted to measurements of P2P traffic to determine traffic characteristics that can be used to do anomaly detection. Some thought should be spent on realistic worm propagation scenarios and hosts they would change P2P traffic characteristics.

## **Literature Study**

The student will read and understand the thesis by Hämmerle. The Software of Hämmerle should be understood and operated to get a working knowledge of its characteristics.

## **Implementation of Validation Mechanism**

The validation mechanisms outlines in the thesis by Hämmerle should be implemented, tested and evaluated. If they do not work to a satisfactory degree, the student might need to design and implement refinements and additional mechanisms.

## **Measurements**

The student will use the implemented software to measure P2P population sizes, traffic generated and specifically traffic characteristics of the different P2P filesharing systems under study. The measurements should be based on NetFlow data records identified to have been exchanged between participants in a P2P system identified by the detection algorithms. The primary focus is to determine "normal" P2P traffic characteristics. Traffic characteristics do not only include amount of traffic, but also average time a pair of hosts communicates with each other, average download sizes, typical number of overlay network maintenance connections between hosts, etc..

## **P2P Worm Scenarios**

The Student will explore the possibilities for P2P worms and what kind of traffic they would generate and how this traffic would change the overall traffic characteristics of the P2P systems under study. This exploration serves to validate that the identified traffic characteristics are actually useful for anomaly detection based worm detection in P2P systems.

## **3 Deliverables**

The following results are expected:

1. *Algorithm documentation* for the used polling algorithms, their effectiveness and limitations.

2. *Measurement documentation* of validation measurements and traffic figures for the validated hosts.
3. *Thesis documentation* A concise description of the work conducted in this thesis (task, related work, environment, design decisions and functionality of delivered implementations, results and outlook).

## **Documentation and Presentation**

A documentation that states the steps conducted, lessons learnt, algorithm design and implementation, major results and an outlook on future work and unsolved problems has to be written. The code should be documented well enough such that it can be extended by another developer within reasonable time. At the end of the thesis, a presentation will have to be given at TIK that states the core tasks and results of this thesis. If important new research results are found, a paper might be written as an extract of the thesis and submitted to a computer network and security conference.

## **Dates**

This thesis starts on October 20th, 2004 and is finished by March 9th, 2005. It lasts approximately one semester. The student is expected to spend 250 hours on the thesis.

An intermediate informal presentation for Prof. Plattner and the supervisors will be scheduled for a date about 5-8 weeks after the thesis has started.

A final presentation at TIK will be scheduled close to the completion date of the thesis.

Informal meetings with the supervisors will be announced and organised on demand.

## **Supervisors**

Arno Wagner, wagner@tik.ee.ethz.ch, +41 1 632 70 04, ETZ G64.1 Co-Tutor: Thomas Dübendorfer, duebendorfer@tik.ee.ethz.ch, +41 1 632 71 96, ETZ G64.1