Semester Thesis:

Visualisation of Internet Host
Client/Server Behaviour
for Bernhard Tellenbach <betellen@ee.ethz.ch>

1 Introduction

Host- vs. Server-like Behaviour

This thesis in the context of DDoSVax\(^1\) will develop new visualization techniques that show the current state and behaviour of Internet hosts based on real NetFlow data.

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\(^2\). 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

After characterizing server vs. client behaviour of Internet hosts, a new type of graphical plot will be developed and implemented to show the current distribution of client and server behaviour in the Internet. A second type of graphical plot will be used for visualizing how the behaviour of Internet hosts changes over time. For both plots, measurement parameters, efficient algorithms and visualization plots will be conceived.

The task of the student is split in three major subtasks: Specification, Implementation, and Validation.

\(^1\)See http://www.tik.ee.ethz.ch/~ddosvax/
\(^2\)See http://www.switch.ch/
**Specification**

Efficient aggregation algorithms and meaningful visualization plots for the current state of observed hosts and for showing the change of host states over time have to be specified. A rough estimation on the needed processing power will have to be made. If performance will be a problem, a simple online and a more detailed but slower offline algorithm might be considered.

**Implementation**

The goal is to generate the new visualization plots for client/server behaviour by a plug-in for UPFrame and to make the plots accessible on the web. In addition, characteristic parameter values should be made available for further processing by anomaly detection tools.

**Validation**

As a validation of the usefulness of the plots and the efficiency of the algorithms, the archived NetFlow data prior and during the Blaster worm of August 2003 will be used. The plots should show significant visible changes during the outbreak of the worm.

### 3 Deliverables

The following results are expected:

1. **Specification** Before implementation can start, an exact specification of the aggregation algorithms and visualization plots must be written down and discussed with the supervisors.

2. **UPFrame plug-ins** The algorithms will be implemented as plug-ins for UPFrame. The code must be tested, run stable and feature short and precise comments.

3. **Visualization plots** The scripts and tools needed for creating the visualization plots and providing them on the web. A short “installation”- readme must be provided along with the code.

4. **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).

Further optional components are:

- Further graphical plots of aggregated statistics extracted from NetFlow data
- Anomaly detection algorithms and tools that use the parameters calculated by the algorithms of this thesis as input.
• Proposition of further aggregation algorithms that could yield characteristic parameters or meaningful plots.

• Generalizations of the algorithms and application to other problem domains.

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 March 25th, 2004 and is finished by mid-July, 2004. 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 organized on demand.

Supervisors

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