

Diploma Thesis:

Near Real-Time Detection of Traffic Usage Rhythm Anomalies in the Backbone

for René Gallati <rgallati@student.ethz.ch>

1 Introduction

The problem

Internet attacks such as massive worm spreading events or denial of service attacks have increased in frequency and impact over the last few years. Attacks that involve several thousand hosts are a reality today.

The setting

In the context of the project DDoSVaX¹ a long-term (>1 year) archive of flow-level Internet backbone border traffic data in the form of Cisco Netflow v5 records was established. In several offline analyses of Internet attacks, characteristic patterns that possibly could be used for online detection were obtained. An online traffic data analysis framework named UPFrame and several near-realtime plugins were already developed and successfully used to detect worm outbreaks.

2 The Task

The usage patterns of certain protocols (such as the amount of e-mail traffic in bytes per hour) show a clearly visible daily rhythm. In case of an attack, this rhythm is disturbed.

The task of the student is split in four major subtasks.

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

Related Work

A search for publications about algorithms that can be used to track and model seasonal or rhythmic changes in somewhat noisy measurements against a time-line is conducted. Inputs from research about technical measurements in various fields (medicine, weather, network traffic, electricity etc.) and also established economic algorithms (stock market etc.) should be considered.

Specification of anomaly detection algorithms

A small set of promising algorithms for tracking seasonal changes in noisy measurements will be selected, adjusted, extended and/or newly developed. An already existing robust algorithm developed by T. Dübendorfer that detects relative minima and maxima in a noisy signal will also be studied and extended for seasonal anomaly detection.

For each algorithm a specification is written and the algorithm's strengths and limitations are analysed and described. The algorithms should be flexible enough to allow for tracking measurement values and fluctuations at different orders of magnitudes. Anomalies that must be detected are spikes (momentary high values), surges (prolonged high values), faults (momentary blackout), blackouts, sags (momentary low values) and brownouts (prolonged low values). In addition, the seasonal behaviour and deviations from "normal" must be characterized with parameters that are easy to interpret and understand.

The algorithms will be evaluated and tested on network traffic of a restricted set of specific services (e.g. SMTP, HTTP). They must be flexible enough to recognize a "normal" seasonal traffic rhythm and adapt to it. After an adaptation or learning phase, they must be able to detect anomalies in the network traffic of such a service.

UPFrame plugin for Netflow data

After familiarizing with the DDoSVax cluster "Scylla", the archived CISCO Netflow traffic data and "UPFrame", a near-realtime plugin will be developed that incorporates the specified algorithms.

For efficient measurements and tests, the plugin must support reading input data through the UPFrame framework as well as directly from the file system. The output produced by the plugin in the form of log files will be further processed by independent tools that interact with the user of the plugin. For graphical plots, tools such as gnuplot or RRD can be used. The developed plugin will be released under GPL.

As the plugin will run under tight resource constraints (CPU, RAM) in near real-time, special attention will have to be paid to efficient resource use. It must be possible to restrict the plugin to a fixed amount of RAM usage.

Validation and Parametrization

It is important to reduce false positives in anomaly detection. Therefore a large set of measurements will be carried out with archived DDoSVax Netflow data. Special attention will have to be paid to outbreak phases of e-mail worms such as Sobig.F or Mydoom.A. A cross-validation on “normal” traffic without outbreaks of larger worms or large network incidents is also vital. This will help to validate, parametrize and optimize the algorithms.

3 Deliverables

The following results are expected:

1. *Related work survey* A short but precise survey that explains current approaches in the field of tracking and modelling measurements exhibiting seasonal behaviour.
2. *Specification of various seasonal algorithms* A specification and analysis of at least three different algorithms that proved useful for tracking seasonal measurements and detecting anomalies in the network traffic of a restricted set of specific services (e.g. SMTP, HTTP).
3. *UPFrame plugin* A near-realtime implementation of the seasonal anomaly detection algorithms. The code should be documented well enough such that it can be extended by another developer within reasonable time.
4. *Diploma thesis documentation* A concise description of the work conducted in this thesis (task, related work, environment, algorithms, measurements, results and outlook). The survey and the specification of the algorithms are also part of this main documentation.

Further optional components are:

- Alarming mechanisms for important anomalies detected
- Further analysis of the real cause of detected anomalies
- Generalizations of the algorithms used

Presentations and Dates

There will be one informal intermediate presentation before the end of the first half of the total thesis duration. 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. This diploma thesis starts on October 25th, 2004 and is finished on February 24th, 2005. It lasts 4 months in total.

Contacts

Tutor: Thomas Dübendorfer, duebendorfer@tik.ee.ethz.ch, +41 1 632 71 96, ETZ G95

Co-Tutor: Arno Wagner, wagner@tik.ee.ethz.ch, +41 1 632 70 04, ETZ G95

Supervisor: Bernhard Plattner, plattner@tik.ee.ethz.ch, +41 1 632 70 00, ETZ G89