ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

TIK Institut für Technische Informatik und Kommunikationsnetze
Computer Engineering and Networks Laboratory

Who turned off the Internet?

Mining Temporary Unreachability of the Internet

Master Thesis
MA-2012-11

Daniel Aschwanden , *asdaniel@ee.ethz.ch*

Advisor:
Dominik Schatzmann, schatzmann@tik.ee.ethz.ch
Dr. Bernhard Ager, bager@tik.ee.ethz.ch

Supervisor:
Prof. Dr. Bernhard Plattner, plattner@tik.ee.ethz.ch

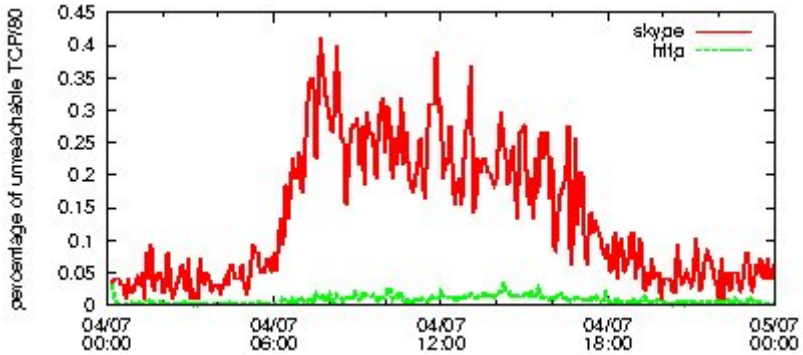
Issue Date: 17. April 2012
Submission Date: 17. Oktober 2012

1. Introduction

More than 20 years after the launch of the public Internet, operator forums are still full of reports about the temporary unreachability of complete networks. We have proposed FACT, a system that helps network operators to track connectivity problems with remote autonomous systems, networks, and hosts. In contrast to existing solutions, our approach relies solely on flow-level information about observed traffic, is capable of online data processing, and is highly efficient in alerting only about those events that actually affect the studied network or its users.

To increase the reliability of possible connectivity problem detection, the noise level is currently reduced by rating the severity of a problem by the number of internal hosts affected. However, depending on the type of service different reachability characteristics are exacerbating the reliability of this hosts aggregation. As shown in figure 1 the overall traffic to port 80 can be split up into HTTP related, Skype related and other traffic. On the one hand, HTTP traffic is predominate with a share of 80% of all port 80 traffic and implies only minor connectivity problem thus offering a reliable metric for connectivity

tracking. On the other hand, the Skype related traffic with a share of 7% has caused a lot of unbalanced connections and is thus distorting the detection.



For this reason, the goal of this thesis is to find a reliable and scalable method to extract service endpoints from the flow level data. Furthermore, the behavior of these service endpoints have to be characterized and their stability have to be rated to offer a robust set of endpoints for detecting connectivity problems with FACT.

2 Tasks

This thesis is assigned to one students. An initial task include the setup of a Subversion or Git repository for storing the code to be implemented. In addition, the student has to become acquainted with the existing code base. Furthermore, the main tasks of this thesis are:

2.1 Related Work

In a literature research existing approaches related to the topic have to be examined and potential shortcomings and possibilities of improvements have to be identified.

2.2 Import Existing Code into FlowBox

Since there is already an existing, but limited implementation of this server socket detection, it must be imported into the tool chain of FlowBox. By doing so, the student gets an in-depth insight into the FlowBox implementation.

2.3 Characterizing of Service Endpoints

The shortcomings of the existing implementation have to be detected and analyzed. Adequate improvements and enhancements have to be proposed and implemented. The found endpoint server sockets have to be rated in an appropriate way respecting the stability and significance.

2.4 Preselect Endpoint for FACT

For supporting the connectivity problem detection process of FACT the found endpoints have to preselected in an optimal set. This optimal set have to be defined combining the coverage, stability and size. The application of appropriate data structures like bloom filters, hash tables, etc. should be considered.

2.5 Adoption of FACT and its Evaluation

FACT has to be adopted to use the preselected endpoints for detecting connectivity issues. The effects of using this rated and optimized set of endpoints have to be analyzed and evaluated. Furthermore, conclusions of the server socket characterizing method and its design choices should be done.

3 Deliverables

The following results are expected:

- Source code of the implemented solution (including SCM history)
- Evaluation of the solution based on real world data
- A final report, i.e. a concise description of the work conducted in this project containing motivation, related work, approach, implementation, results and outlook.
  The abstract of the documentation has to be written in both English and German.
  The original task description is to be put in the appendix of the documentation. All deliverables need to be submitted in electronic form. The whole documentation, as well as the source code, slides of the talk, etc., needs to be archived in a printable, respectively executable version on a CD-ROM.

4 Assessment Criteria

The work will be assessed based on the following criteria:

- Knowledge and skills
- Methodology and approach
- Dedication
- Quality of results
- Presentations
- Report

5 Administrivia

5.1 Documentation

Documentation that states the steps conducted, lessons learned, 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 maintained and extended by further users. If important new research results are found, the results may be published in a research paper.

5.2 Presentation

At the end of the project, a presentation will have to be given at TIK that states the core tasks and results of this project. The presentation consists of a 20 minutes talk and reserves 5 minutes for questions.

5.3 Dates

This project starts on XX.04.2012 and is finished on XX.10.2012. At the end of the second week the student has to provide a project schedule for the thesis, which will be discussed with the supervisors. One intermediate presentation for Prof. Plattner and all supervisors will be scheduled after one month. The final presentation will be scheduled close to the completion date of the project. Informal meetings with the supervisors will be announced and organized on demand.

5.4 Infrastructure

The student is expected to complete at least part of the work on-site at the TIK ETH, where an office desk will be available. Server infrastructure, such as an SCM repository server or for use as test environment, will be provided by ETH if needed.

5.6 Copyright and Confidentiality

Copyright is regulated by the "ETH Zürich - Vereinbarung mit Projektmitarbeitern" contract. The student must respect the confidentiality of data, including but not limited to flow data, intellectual property and other ETH internal data. In particular, the student is required to adhere to all relevant Swiss law and regulations, including telecommunication secrecy and data protection. The student must make sure that the thesis documentation does not contain confidential data. If that is not possible, then the thesis documentation containing a minimal amount possible of confidential data must not be distributed outside of SWITCH and TIK.

5.7 Advisor

Dominik Schatzmann, schatzmann@tik.ee.ethz.ch, +41 44 632 54 47, ETZ G 95