TDTS21 projects vt 2022

This document includes a mix of open-ended projects and projects with specific ideas (even if only high-level descriptions here).

Data sharing statement for all projects: The datasets, tools, and analysis should not be shared publicly until we potentially publish a research article using these tools and datasets. (If needed, at such a time we hopefully have been able to sanity check, polish, and improve the tools/datasets. Postponing the release to such time also significantly improve the odds that a paper is published.) However, to enable continuous research and potential publication, already at the end of the term, all code, data, text, and results must be shared with your supervisor. Also, please discuss and update me on your ideas and progress so that we together can try to make the most of the class projects (regardless if it has potential to be published or not).

1) Descriptive title: Comparisons of the major root store programs and their roots
In our paper "Characterizing the Root Landscape of Certificate Transparency Logs" (https://www.ida.liu.se/~nikca89/papers/networking20b.pdf) we compare the root stores of the CT logs (but also include a comparison of the three most used root stores (by Apple, Microsoft, and Mozilla). Since then, Chrome has started their own root store program. In this project, you are expected to (1) collect a good longitudinal dataset of the historic root stores of the four major root stores, (2) characterize and compare how the four major root stores and their overlaps have changed over time and [if time permits] (3) take a current snapshot of currently used certificates [e.g., from CT logs or crt.sh's database certwatch] to identify (and look closer at) websites that use certificates that only chain back to a subset (or none of the root stores) but not all of the major root stores. Note that we expect the focus to be on doing a really good job on tasks (1+2) but would find some progress on (3) a nice bonus. This project requires strong programming skills and familiarity working with existing coding frameworks are highly beneficial (as you likely will work with code from various open-source tools).

2) Descriptive title: Secure BGP and its RPKI
Measure and characterize the deployment of Secure BGP (e.g., its current topology) and the keys used in this infrastructure. First step here would be to investigate to what extent we actually can develop a good methodology to collect a dataset that complements some of our past work. As a backup plan, we can create a set of graphs of the AS topology at different snapshots using public data from the RouteViews project and look at how that have changed over time.

3) Descriptive title: Data collection and analysis of IP address usage of phishing domains (and perhaps also other classes of bad domains)
In this project you will try to identify the IP infrastructure of phishing domains (and potentially other bad domains) and study to what extent such domains share underlying infrastructure. We would also like to explore to what extent this is possible to do using historic data. Ideally, we would like to do a longitudinal data analysis.

4) Descriptive title: The social networks of the gaming communities
Outside the games, users may socialize in numerous ways, including by commenting on the gamecasts (i.e., records of games) and chatting with their friends through various online resources. For example, some popular online game communities provide an interactive gamecast sharing service, wherein the creators promote their gamecasts through live streaming with on-air explanations (in audio and text format) of their game styles. In this project you will develop a measurement methodology, collect data, and present a preliminary analysis of the social networks formed in one or more such communities. Of special interest are the social interactions (which in some cases can express the strength in user relationships, for example) and the amount of additional network traffic generated around a live event (both in parallel and afterwards). Also, do you find heavy tailed relationships or other interesting characteristics?

5) Descriptive title: Longitudinal COVID-19 analysis of DNS traffic
This is a follow-up analysis to a BSc thesis from last year. In this thesis project, you will use changes in the popularity of different websites (e.g., based on DNS traffic) to characterize the impact of different classes of domains. You can find a discussion in such ranking lists here https://doi.org/10.1145/3278532.3278574. Experience with statistical analysis, signal processing, or similar may be beneficial, as the data can be noisy.

6) Descriptive title: BGP interceptions in the wild
Identify potential BGP interceptions that have taken place recently and characterize them (accidentally or intentionally) using similar methods as in our 2013 paper https://www.ida.liu.se/~nikca89/papers/pam13.pdf.

7) Descriptive title: Does being a stereotype help becoming targeted or is it enough visiting a page?
Build a Selenium-based data collection framework that generate web sessions based on different user profiles and compare what level of personalized ads (i.e., ads that match the user profile) that they see on the pages that they visit compared to a plane user (that have no history) going to the same website. Here, you will need to generate user traffic for some time and then collect how the ads presented to these users change over time. The goals here is to design a methodology and tool to (in a good way) measure the level of personalization we obtained just by ad campaigns targeting users of certain pages compared to ads that actually target individuals regardless of what website they visit.

8) Descriptive title: Recreating existing measurement work on a topic of interest ...
Pick a measurement paper of interest (e.g., perhaps from the set of papers we have seen already) and try to collect your own large-scale dataset and try to (1) validate the results (or part of the results) presented in that work and/or (2) try to extend the work in some direction. As with all projects, it is important that you discuss your research questions and data collection ideas with Niklas.

9) Descriptive title: Beyond existing measurement work on a topic of interest ...
Pick a topic of interest to your group (e.g., perhaps based on a paper or two from IMC, and identify some related issues/problems) and try to collect your own large-scale dataset that help answer some interesting/important questions. As with all projects, it is important that you discuss your research questions and data collection ideas with Niklas.