TDTS21 projects vt 2021

This document includes a mix of open-ended projects and projects with specific ideas to be discussed in person.

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: Post-quantum crypto (PQC) analysis of one or more internet protocols
Muller et al. [https://doi.org/10.1145/3431832.3431838] recently presented an interesting study in which they present a case study in which they look at how DNSSEC can be retrofitted to work in the post-quantum cryptography (PQC) context. Here, they consider the requirements imposed by using larger signatures, keys, and more computational power, for example, and the implication this may have on DNSSEC. In this project, you are expected to identify one or more additional internet protocols were PQC may need to be considered and try to apply the same (or similar) methodology on that/those protocol(s).

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 phishing domains
In this project you will try to identify as many complementary data sources (e.g., public records, CT logs, traffic traces, routing tables based on prefixes, geo-data, etc.) as possible that provide information about known phishing domains (e.g., https://www.phishtank.com), and then perform a preliminary data collection and analysis. As part of this project, you will help implement a data collection framework, perform large-scale data collection using this framework, perform careful pre-processing of the dataset so to create a cleaned-up dataset, and ideally also perform a preliminary analysis in which we try to characterize how phishing domain across several dimensions. As part of this project, it would be good to see if there are other interesting datasets that can be used (e.g., as shared in prior measurement conferences and via other related/non-related papers).

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: Performance evaluation of the impact of individual third-party domains and having them blocked/unblocked
Build an evaluation framework in Selenium that evaluates the performance tradeoffs (e.g., load times etc.) associated with having individual third-party ad services blocked/unblocked. Careful selection of the sites visited is expected here (and may need to be coordinated with other parallel projects).

7) Descriptive title: Which users see most targeted adds and which ad-service are most successful in the ads they present?
Build a Selenium-based framework that generate web sessions based on different user profiles and try to determine which user profile sees the largest fraction of targeted ads (i.e., adds that matches the user profiles). Here, you will need to generate user traffic for some time and then collect how the ads presented to these users change over time.

8) Descriptive title: Comparing tails: Identification and collection of heavy-tailed datasets
Create a repository of uniformly formatted datasets with potential heavy-tailed properties. This would include a combination of identification of public distribution-related datasets (with different forms of network related data), but could also include collection/creation of additional datasets (e.g., by extracting the distribution data from other sources). All files should follow a uniform row-column format (e.g., csv), and a high-level analysis that compares the dataset properties should be performed. A quick read on heavy-tailed distributions: A. Mahanti, N. Carlsson, A. Mahanti, M. Arlitt, and C. Williamson, "A Tale of the Tails: Power-laws in Internet Measurements", IEEE Network, Vol. 27, No. 1, Jan/Feb. 2013, pp. 59--64.

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 project, it is important that you discuss your research questions and data collection ideas with Niklas.