MSDS 7349
Project List

I. On Research Problems

The fields encompassing security have a long and illustrious history marked with periods of high innovation followed by periods of incremental improvements. The driving force for all of the innovation and improvements come from the broad range of important but open problems that are analyzed, addressed, and then change as either the underlying technologies change (such as microprocessor performance improvements that make "computationally infeasible" problems suddenly feasible in real-time), the applications and data types and quantity change (such as the Internet of Things and Big Data revolutions), or both. For this reason, some old open problems are new open problems again. Traffic characterization is the most obvious of these types of "solved" problems that are constantly becoming unsolved. As applications change, so do traffic usage patterns. This makes it useful to perform again traffic studies to identify the characteristics of the current traffic usage.

The following list of security problems provides a small subset of the open problem topics that currently exist. For this class, choose one of the problems given in this document. Note that these are problem *topics* and not predefined specific problems.

Be sure to choose a project topic that interests you.

Once you have chosen a problem topic, you need to further define the problem into a problem that is solvable within the confines of this term project. You have roughly two months, limited time, limited resources and will be performing "research," so your problem needs to be scoped accordingly. You should begin scoping your problem by looking for technical publications in the related topic area within the main computer security publication venues. For journals, the primary publication venues are:

- ACM TISSEC
- Journal of Computer Security
- IEEE Transactions on Dependable and Secure Computing
- Journal of Information Security
- Computers and Security

For conferences, the primary publication venues include:

- ACM Conference on Computer and Communications Security
- IEEE Symposium on Security and Privacy
- IFIP SEC
- USENIX
- ACM SACMAT

Start with a Scholar Google search for papers published since 2014 on your chosen topic. This will give you an idea of what you are expected to produce as well as the types of problems that are being addressed. Note that just because a paper has been published on a specific problem does not mean that the problem is solved or that you cannot attempt to solve that problem in a different way or come up with a better algorithm than what has already been published.

The goal of your research is to *explore the solution space.* That is, don't just come up with *a solution* (although you will do this). You also will evaluate what are the other possible solutions including what would be the optimal solution (even if it's not a practical solution) and possibly including what impacts small changes to the problem definition would have on the goodness/badness of possible solutions including yours (in effect, evaluating how fragile your solution is to minor changes in the problem definition).

Your term paper will present your results and only those critical portions of your technical solution required to understand your solution and your results. Note that this means that the bulk of your work, particularly if software is required to be written to simulate or otherwise evaluate your solution, will likely be presented in two to three paragraphs (less than one half page) of a six plus page paper. For a better idea of what I mean, read the papers in the journals and conferences identified above.

The term paper will be evaluated on its technical merit and depth, description and comparison with related work, originality and presentation, as is done in all of the journals and conferences cited above.

II. Open Problems

1) *Statistical Analysis of AES-128* AES is the standard block cipher used for secret communications (confidentiality) worldwide. AES-128 utilizes a 128-bit key and has a 128-bit block. In this project, use the SMU ManeFrame to generate as many AES encryption blocks as possible and perform an extensive statistical analysis on the resulting data. The standard NIST randomness tests should be used, plus additional tests and analysis should be performed to identify any bias in either the entirety of the output or in sequences, such as the sequence of encryption blocks generated from input values created using a counter or a linear feedback shift register (LFSR). Data should be generated using the encryption function of AES-128 using 128-bit input values.
2) *Blockchain Analysis* Blockchain technology is at the heart of bitcoin and is increasingly being developed for use in a broad range of financial applications. The goal of this project is to provide a survey and tutorial on blockchain technology. The project and writeup will include an evaluation of an experimental blockchain database.

3) *Privacy in a Data Centric World* Data capture, transmission and distribution is regulated, often highly so, yet we continue to live in a data rich cyber world that seems to capture and share our every point and click across cyberspace. The goal of this project is to perform a comprehensive evaluation of the legal and social constraints on data collection, transmission and dissemination. While the scope is global, particular attention should be upon the laws governing the EU and the USA as these two regions have diverged on their legal requirements. Your analysis should focus on how these laws impact data throughout that data's lifespan.
4) *The Economics of Privacy* Privacy has a cost, particularly in our data centric world. The goal of this project is to identify and explore the economics of privacy. (A key resource to get your ideas going is Ross Anderson's page on Security Economics http://www.cl.cam.ac.uk/ rja14/econsec.html)
5) *Anonymizing Data* Large "anonymized" data sets are often made available to researchers, but just how anonymous are these data sets? Researchers have found that with very little information from a data set, especially when combined with other publicly available data sets, easily compromises the anonymity and privacy of individuals. In this project, the goal is to identify methods to further anonymize data sets by changing the actual data values without changing the statistics of the overall data set. That is, how much does the data need to be modified to provide for true anonymity while still allowing for the analysis to yield results that are within the 95% confidence interval of the original data?
6) *Security Threats of Large Data* Large data sets, such as machine activity logs, can contain a significant amount of information about the activities of the users and the software on a system. This information may be used to identify potential avenues of system attack or even clearly identify a security bug or simply learn more about a particular vulnerability or attack approach. In this project, the goal is to analyze one or more data sets from the `azsecure-data.org` site in order to learn something security related from the data.