Date: 17 September 2018

To: Romcholo Macatula

From: Cyber Warriors

Subject: Technical Memo 3: Brainstorming Approaches, Literature Review, Field Trip

## 1 Summary

This memo shall enumerate and describe the brainstorming ideas discussed throughout group meetings and the possible approaches for each of the project components. Methods like modeling a network of users, Markov Chains, and regression will be considered for our math modeling component. Comparably, depending on the method chosen we access different forms of collecting data through proxies or physical means. In addition, this memo addresses literature found relevant to the teams project. Topics like risk detection, situational awareness in cyberspace, as well as models other researchers have used in their approach. Unfortunately we have yet to preform our field trip, however; we have a meeting schedule with a cyber security specialist who works at the Hume Center's Intelligent Systems Laboratory in Arlington, VA Friday, Sept. 21st.

## 2 Brainstorming

**Update from Tech Memo 2:**

The team decided to condense our three components into two. The team shall move the relevant criteria from the Data Analytic component to Math Modeling. Scope shall correlate to different methods the team considers (i.e. Network of Users, Network of Networks, Markov Chains, and Regression). Ease of the model shall relate to the number of incorporated factors.

**Component 1: Math Modeling**

The team brainstormed 4 methods for development of modeling to access how an attacker shall traverse through a network.

1. Network of Servers: Each node in the method shall indicate servers with various different permission levels. Each edge shall act as a connection between each server. The placement of important nodes shall affect the model. This shall provide an assessment of how to better organize a network to prevent further privilege escalation.

2. Network of Users: Each Edge shall represent vulnerabilities such as, buffer overflow attacks, weak passwords, Denial of Service attacks, etc. The nodes that correspond shall represent access levels per user (i.e. basic to admin). The method shall access the shortest path an attack takes based on vulnerabilities available to gain access to a user with the highest level of privilege.

3. Markov Chains ("ODE"): As an attacker traverses through a network the rate at which they move from one node to another corresponds to a number of factors (i.e. weak passwords,

known vulnerabilities). The number of detections over time shall escalate the associated risk of the network. In some ways this technique shall represent an infection model that provides an answer to attacker infiltration latency.

4. Regression(Least Squares/Neural Networks): To access prediction capabilities the team considered a possible least squares and/or neural network approach. This method shall require large amounts of data to attribute to specific factors such as user access level, number of users, etc. In this method, data acts as a priority. The team shall aim to produce a system wide score of the networks risk. This technique shall determine the risk of the network.

**Component 2: Data Collection**

1. Proxies/ Synthetic: Generation of our own data through research (such as software used by a company, privilege hierarchy, etc.) to aid in the creation of one of the above models. To acquire this data the team shall use web scraping methods to pull information from important resources.

2. Online Free Sources: Data retrieved from available public sources online, used for the creation of models and data analysis. The team shall perform this form of collection through API's that collect meaningful data.

## 3 Literature Review

### 3.1 Article 1:

*A Review on Cyber Security Datasets for Machine Learning Algorithms* reviews, compares, and explains common cybersecurity datasets such as the ADFA Linux and CSIC 2010 HTTP datasets. The paper explores machine learning techniques associated with each of these datasets such as regression,support vector machines, and clustering. This paper applies to our project as it explains common cybersecurity datasets which the team shall utilize in either the creation of our model, or as a reference. The description of the machine learning techniques associated with the datasets holds a direct relation to the possible approaches for our math modeling component. The datasets apply to our data collection component[7].

### 3.2 Article 2:

*Exiting the Risk Assessment Maze: A Meta-Survey*, provided to the team by Dr. Colbert (our field trip contact), overviews cyber security risk assessment. This paper provides a high-level look at risk assessment and how analysts make assessments for a network. The team shall utilize this paper to understand risk assessment from a broader perspective without the specifics of data analysis [3].

### 3.3 Article 3:

*Biologically Inspired Risk Assessment in Cyber Security Using Neuralnetworks* explores cyber security risk assessment through neural networks. Neural networks exist as one of the possible approaches for the math modeling component of our project. The paper also includes attack graphs. Attack graphs represents all possible sequences of an attackers' actions. Our project explores the

possible actions of an attacker after premier infiltration of the network. The team shall utilize these attack graphs to understand the possible privilege escalation techniques of an attacker [4].

### 3.4 Article 4:

*A Cyber Attack Modeling and Impact Assessment Framework* explores a possible cyber attack model as well as exploitation impact assessment. This paper relates to the project as it describes a set of algorithms for security evaluation, an aspect similar to what our math modeling component involves. The paper contains attack graphs and security metric calculations which serve as possible examples for our math modeling component [6].

### 3.5 Article 5:

*System and method for risk detection and analysis in a computer network* describes an invention that provides methods for risk detection and analysis. The patent contains useful data for the data collection component. However, due to the sources' existence as a patent, the team contacted the company Skybox Security to ensure legal use of the data. The team shall wait for a response and use the data only if allowed. Otherwise, the team shall utilize the patent as a risk assessment model example for the math modeling component [1].

### 3.6 Article 6:

*Analysis of Security Data from a Large Computing Organization* provides an analysis of security incidents on a large network. Since the team aims to create a mathematical model for privilege escalation on a corporate network, the analysis in this paper shall provide a useful reference for a large network [5].

### 3.7 Article 7:

*An Evaluation Framework for Intrusion Detection Dataset* assesses and compares popular cyber security datasets such as DEFCON and KYOTO. The paper contains a table that describes the different protocols (FTP, SSH, HTTP, etc.) that each dataset contains. The team shall utilize this paper as a reference to locate relevant datasets for our project, this relates to our data collection component [2].

## 4 Field Trip

The team plans to speak with Dr. Colbert, the current director of the Hume Centers Intelligent Systems Laboratory in Arlington, VA. Dr. Colbert specializes in cyber security, cyber-physical systems, autonomy, and artificial intelligence. We shall discuss the literature we have reviewed and possible methods to improve our approach.

## References

[1] G. Cohen, M. Meiseles, and E. Reshef, *System and method for risk detection and analysis in a computer network*, (2005). US Patent 6,952,779.

[2] A. Gharib, I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, *An evaluation framework for intrusion detection dataset*, in 2016 International Conference on Information Science and Security (ICISS), Dec 2016, pp. 1–6.

[3] D. Gritzalis, G. Iseppi, A. Mylonas, and V. Stavrou, *Exiting the risk assessment maze: A meta-survey*, ACM Comput. Surv., 51 (2018), pp. 11:1–11:30.

[4] M. Ionit and V. Patriciu, *Biologically inspired risk assessment in cyber security using neural networks*, (2014), pp. 1–4.

[5] Z. Kalbarczyk, J. Barlow, A. Sharma, and R. Iyer, *Analysis of security data from a large computing organization*, in 2011 IEEE/IFIP 41st International Conference on Dependable Systems  Networks (DSN), vol. 00, 06 2011, pp. 506–517.

[6] I. Kotenko and A. Chechulin, *A cyber attack modeling and impact assessment framework*, (2013), pp. 1–24.

[7] O. Yavanoglu and M. Aydos, *A review on cyber security datasets for machine learning algorithms*, in 2017 IEEE International Conference on Big Data (Big Data), Dec 2017, pp. 2186–2193.