# 4 April 16

MEMORANDUM FOR: AFIT/ENG

ATTENTION: MAJ BRIAN WOOLLEY

FROM: 1ST LT PAUL JORDAN, (GCS-17M)

SUBJECT: Thesis Prospectus: Online Failure Prediction in a Microsoft Enterprise Domain Controller

1. The purpose of this research is to explore the use of machine learning techniques to predict failure in our Microsoft Enterprise Domain Controllers (DC) to avoid network service outages. An expected result is to develop a method of automatically predicting failure based on existing log messages and system health information in order to inform operators of pending failures so that they may take action to avoid those failures. The method developed will be an extension of a recently published framework that automatically trains and employs a predictor after perceived underlying system changes.

2. Previous work has been done to identify and classify nearly one hundred methods of predicting failure in computer systems (Salfner et al., 2010). Many of these methods are not in use today due to the rapid changes in underlying systems that often render prediction algorithms useless until retrained. Furthermore, since failure is a relatively rare event, obtaining data with real failure occurrences is difficult. A framework was published in 2015 that attempts to solve this problem through automated retraining and realistic failure simulation (Irrera et al., 2015). Unfortunately, the conditions under which this framework was tested no longer exist and the framework is not easily applied to modern Microsoft Windows Server operating systems. This research seeks to extend the new framework by implementing the components of the framework in a way that is compatible with the modern Microsoft Windows Server operating systems. Further, a case study will be conducted using the new implementation in order to demonstrate its efficacy and generalize the existing framework.

3. The primary objective of this research will be to train a prediction algorithm using simulated failure data obtained by injecting faults into a loaded DC until failure occurs as outlined in the Adaptive Failure Prediction (AFP) framework (Irrera et al., 2015). The simulation and training will occur in a virtual environment that has recently been designed and built, running on hardware that has already been allocated for this research. The load generator has also recently been built, tested, and validated in a paper that has been submitted for publication. The fault injection will be done by a tool that has been developed for this research that implements the General Software Fault Injection Technique (G-SWFIT; Duraes et al., 2001). Tools written before this research were not capable of injecting faults into the primary executable responsible for authentication on a modern Microsoft Windows DC. Finally, extensions to the AFP framework will be explored. These extensions include obtaining labelled failure data after forced failure due to a corrupted active directory database, and load induced failure. This research will be considered successful if the AFP framework is accurately implemented and tested for a modern Microsoft Windows DC.

4. Results and common measures of performance such as precision, accuracy, and recall of predicted failures recorded both before and after software updates will be reported. The expected result of this research is that the AFP framework will be generalizable and work with modern Microsoft Windows Server operating systems. Since a prediction method is not presently deployed on the Air Force network any level of dependable prediction will be better than what is currently available. However, this research will attempt to show that after an underlying system change, the extended AFP framework will be capable of automatically training a more effective prediction algorithm.

5. Results from this research will be implemented and used by the Cyber Security and Control System (CSCS) weapon system employed at the 561st and 83d Network Operation Squadrons (NOS) and their associated detachments to reduce the number of network service outages increasing uptime leading to improved mission effectiveness in both the support and operational domains. This research is funded by the National Security Agency and as a result could also be employed outside of the Air Force to increase mission effectiveness across the Department of Defense (DOD). External to the DOD, this research further generalizes an approach that could be used to help increase availability of nearly any computer system.

6. Proposed thesis committee:

a. Dr. Gilbert Peterson, Chair / Thesis advisor \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

b. Dr. Michael Mendenhall, Committee member \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

c. Maj. Alan Lin, Committee member \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

d. Maj. Andrew Sellers, Committee member \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

7. Sponsor: National Security Agency

a. Names: Glenn M­­­. Ellison, and Alice E. Shaffer

b. Organization: National Security Agency

National Information Assurance Education and Training Program (NIETP)

9800 Savage Road

Fort Meade, Maryland 20755-6744

c. Telephone number: 410-854-6206

d. E-mail address: gmellis@nsa.gov, and aeshaff@nsa.gov

8. The coursework completed at AFIT has been planned to provide the foundation of knowledge required to successfully complete this thesis. Courses of particular applicability to this research topic include:

* CSCE 523 Artificial Intelligence
* CSCE 623 Statistical Machine Learning
* CSCE 699 Network Device Failure Prediction
* CSCE 823 Introduction to Artificial Neural Networks

Additionally, four credits of special study focusing on network device failure prediction have been taken to conduct literature review and gain familiarity with state of the art machine learning failure prediction techniques. As previously mentioned, a load generation tool capable of sufficiently loading a domain controller has been developed through a course project. Additionally, a tool that implements a general and realistic software fault injection technique for x86-64 executables has already been developed through self study. Finally, familiarity with machine learning techniques and the R programming is required to complete this thesis and has been or will be gained through self study, course projects, and assignments.

Paul L. Jordan, 1st Lt, USAF

GCS-17M

1st Ind, AFIT/ENG

MEMORANDUM FOR AFIT/ENG

I approve/disapprove the above thesis prospectus and thesis committee. This prospectus will be maintained in the student’s file. The thesis should be prepared in accordance with the AFIT Thesis Guide. Good luck!

BRIAN WOOLEY, Maj, USAF

Chief, Computer Science Division

Department of Electrical and Computer Engineering