MODIFIED Charter:

1. Proposed Title of Capstone Project: **A Quest Called TRIBE: Clustering Malware Families for Enhanced Triage and Analysis**
2.
a. Type of Project: indicate which of the following 3 groups your project best fits:

TECHNICAL

3. Name of Faculty Advisor:

SCY: Dr. Dane Brown	**Cyber Science**	**Primary Advisor**
CS: CDR Edgar Jatho	**Computer Science**	**Secondary Advisor**

a. Subject Matter Experts (SMEs):

LtCol Solomon Sonya, USAF, Purdue University
Dr. Dongyan Xu, Director of CERIAS at Purdue University

b. Battle Rhythm: proposed meeting schedule:

MIDN will meet once weekly on Wednesday during the lunch period. MIDN also have classes together and will use that time to communicate. Faculty meetings will go for 30 minutes every Thursday. A github repo and google calendar will be kept up to date.

(1) Paper format:

c. Link to shared google drive folder for your project materials/files you create in this shared drive folder:
https://drive.google.com/drive/folders/0AJH3ew5_kNFXUk9PVA

d. Name of all midshipmen on this Project and indicate any who are dual majors or not SCY majors:

Michael Chen - Computer Science
Chris Kim - Data Science
John Jenness - Computer Science/Cyber Science
Justin Liaw - Computer Science/Cyber Science

e. Detailed description of research project:

The project aims to enhance machine learning-based malware classification by developing a novel model, named Tribal Relation Inferential Binary Encoder (TRIBE), which groups malware into 'tribes' based on binary data from a malware file. Utilizing a transformer-based sequence-to-sequence variational autoencoder, the project seeks to improve the speed and accuracy of malware classification, overcoming the limitations of traditional

methods for malware triage. This approach looks to enhance malware incident response by providing an effective and efficient immediate classification system.

This project is originally based on the proposed Trident by Michael Chen. We have access to a multi-thousand dollar GPU hosted outside of the USNA to do analysis of malware binaries. We are looking to publish two papers. The first of which will cover the data collection and a review of our custom autoencode. The second will cover the malware analysis and ML grouping we do with the data. Our larger goal is to improve the classification time of malware so customers like SOCs can quickly and efficiently respond to threats. This is a unique methodology approach to malware classification.

Note: Micheal Chen's Proposal Paper: ChenTridentApp25.docx (1).pdf

4. Individual Committment

a. 1st Week: Group formation
b. 2nd week:
- **Literary Review & Catchup - Liaw & Jenness**
- **Auto Encoder Github Searches - Chen**
- **Non-Malicious Data Hunting - Liaw & Jenness & Kim**

c. 3rd week:
- Weekly Meeting Establishment
- **Literary Review- Liaw & Jenness**
- **Auto Encoder - Chen & Liaw**
- **Non-Malicious Data Formatting - Jenness & Kim**

d. 4th week:
- **Auto Encoder - Chen & Liaw**
- **Non-Malicious Data Classification - Jenness & Kim**

e. 5th Week.
- **Writing Sprint & Reevaluate - All**

f. 6th Week (Note Lit Review is due)
- **Literary Review- Liaw & Jenness**
- **Writing Sprint - Kim and Chen**

g. 7th Week.
- **Autoencoder Revision- Chen & Jenness**
- **Malicious Data Corpus Creation - Liaw and Kim**

h. 8th Week.
- **Autoencoder Revision- Chen & Jenness**
- **Malicious Data Corpus Creation - Liaw and Kim**

i. 9th Week.
- **Autoencoder Revision- Chen & Jenness**
- **Malicious Data Corpus Creation - Liaw and Kim**

j. 10th Week

- **ML Intro Dev- KIM**
- **Malicious Data Corpus Creation - Chen and Kim**

k. 11th Week **Writing Paper**
.

l. 12th Week: **Writing Paper - All**

m. 13th Week: **Writing Paper - All**

n. 14th Week: **Editing - All**

o. 15th Week: **Reevaluation**

p. 16th Week (DRAFT Paper is due).

Missed Deadlines:

Following the failure to meet a deadline, all group members will schedule a meeting with both faculty members the week of the failure. A report no shorter than one page will be written documenting the reasons for the failure. This report will be signed by all members of the group and submitted to faculty members.

Unacceptable Work/Performance:

1st Occurance: On the first assignment submission in which the performance of a team member was unacceptable, a MIDN only meeting will occur to address the performance.

2nd Occurance: On the second occurrence, a full meeting will occur with the faculty and the MIDN. The group will decide an outcome for this infraction.

3rd Occurance: The MIDN will have a reduction in grade.

Other Concerns:

The scope of this project is vast. A more detailed timeline must be constructed.

Agreement Statement

WE

ALL MIDSHIPMEN ON CAPSTONE PROJECT NAMES

recognize that the Capstone Project is a labor-intensive enterprise that demands a high level of personal and team commitment, time, and effort. This is particularly true when the Capstone Project counts towards the SY401/SY402 final grade and the project must be completed within the temporal limitations of two semester-long courses. By signing this document, we promise to dedicate the necessary time and effort to complete this project in accordance to the schedule drawn above. We have also reviewed our institution's academic integrity policies (3920.3A) and we are fully aware of the seriousness of these issues and of the consequences of violating such policies. If this research project involves the recruitment and testing of human subjects, we agree to take a tutorial on the protection of human subjects before commencing work on the project. We will reach out to the UNSA HRPP Office and complete the necessary training and approval process. We shall also abide by the stipulation that all research data (e.g., questionnaires, data files, records, observations) from this project become the property of the institution and will be retained by the faculty member who will determine who and under what circumstances others may have access to such data. We also understand that authorship of any resulting conference presentation or journal article will depend on the extent of our contributions to this project. We agree that this project is a team effort and the workload shall be shared equally among the team. We are subject to periodic evaluations of our level of effort during the Fall and Spring Semester — individually and by our project peers — that may result in penalties for poor effort and teamwork.

We understand that Capstone Day is the last day of classes 30 APR 2025!

MIDN Signatures

Justin Liaw
John Jenness
Christopher Kim
Michael Chen

Faculty member

Dane Brown

Edgar Jatho