



DEPARTMENT OF THE ARMY  
UNITED STATES MILITARY ACADEMY  
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE  
WEST POINT, NY 10996

MADN-EC

13 November 2019

MEMORANDUM THRU

LTC Thomas Babbitt, IT Program Director  
Dr. Wenli Huang, Curriculum Coordinator

FOR COL James J. Raftery Head of Department of Electrical Engineering and  
Computer Science, West Point, NY 10996

SUBJECT: Request for Approval of Advanced Individual Study (IT493) during Term 20-2

1. Research Project Title: The Dark Web Search Engine
2. Project Summary: The Dark Web is notoriously difficult to crawl. The Hidden Services directory, which users use to find hidden services, stores hashes of domains to prevent enumeration. Hidden services, the web sites hosted on the Darknet, are not highly connected through hyperlinks like sites on the clearweb, diminishing the ability of crawlers to index the Dark Web. All users must have *a priori* knowledge of a hidden service URL. Typically, users obtain these URLs from websites on the clearweb. This project aims to create a Dark Web crawler by automating the process of finding hidden service URLs on the clearweb. Current efforts are hand-curated and do not reflect the current status of hidden services on the Dark Web.
3. Project Summary: This research proposal aims to:
  - a. Extract all hidden service URLs (i.e. URLs ending with onion) from the Common Crawl corpus.
  - b. Perform a rudimentary analysis of the top 10 most frequently occurring domains
  - c. Share results with the Department of Homeland Security's Dark Web & Crypto Currency Group
4. Project Location: TH115
5. Principle Investigator: Cadet Derek Nunn, Information Technology major.
6. Project Advisors: Kyle King will service as the primary advisor for the project and CPT Daniel Hawthorne will serve as an additional advisor.

7. Project Timeline: The following is a breakdown of projected dates and deliverables:

Lesson	Deliverable	Points
1	Abstract complete	50
3	Literature review complete	50
5	Code design complete	50
9	Common Crawl ingested	100
13	URL extractor completed	100
15	Draft paper	100
	Midpoint presentation	100
17	Analysis of top occurring domains	100
20	Customer presentation / feedback	100
30	Final Presentation	100
	Paper complete	100
Instructor Points		50
Total points available		1000

8. Academic Background: My current CQPA is 2.946 with a Core APSC of 2.746 and a major APSC of 3.516. The following is a breakdown of major courses and their grades:

Course	Grade
EE360	B-
CY300	B+
CY355	A-
EM411	A-
CY350	A-
CS384	B
IT383	A-
IT384	B+
IT394	A+

9. Request Academic Credit: 3.0 for the 120 hours in the IT493 course.

CADET DEREK NUNN  
CADET, 1ST CLASS  
COMPANY D3, USCC