**SEG3904 Project Proposal:** Network Traffic Classification for Detection of Malicious TCP SYN Flood as a Distributed Denial of Service Attack

Student Name: Iymen Abdella, [iabde058@uottawa.ca](mailto:iabde058@uottawa.ca), 7724153

Supervisor Name: Miguel Garzon, mgarz042@uottawa.ca

Overview:

The purpose of this project is to analyze and classify models for the detection of malicious attacks, more specifically: Distributed Denial of Service attacks using TCP SYN Flood. A literature review of the material will have to preface the project in order to gain more information on the topic of the research project. Collection of the necessary network traffic data will be orchestrated using the appropriate industry standard data mining techniques. The collection of data will then be prepared, processed, and discretized for ease of classification using industry standard data processing methods. Once the data has been gathered, different classification models will be analyzed to categorize malicious attacks to a sufficient level of accuracy. To conclude the project, a report will be authored that will provide a detailed discussion of the various classification models and which of them performed most accurately.

Learning Outcomes:

At the end of the project the student (Iymen Abdella) will have learned:

* Distributed Denial of Service attack mitigation and prevention for increased network security.
* How to use distributed file systems to store, retrieve, edit and analyze large data sets.
* To build classification models to analyze large data sets.
* To evaluate the accuracy of classification models.
* How to analyze data packets and netflows for increased network security.

Technologies:

* Python is a general use high level programming language.
* Pyspark is a Python API that gives access to SPARK.
* Scikit-learn, numpy, pandas are machine learning library for Python.
* Docker is a containerization software to perform operating system level virtualization

Resources:

● Various literature works that will be discovered during the literature review.

● Technology websites mentioned above.

* William Stallings, Data and Computer Communications, 9th edition, Prentice-Hall, 2011.

**Deliverables**

|  |  |
| --- | --- |
| **Deliverable** | **Weight** |
| Literature review report | 15% |
| Environment setup | 10% |
| Implementation code: Part a) Data Collection Module | 10% |
| Implementation code: Part b) Data Cleansing, Preprocessing, and Discretization | 20% |
| Implementation code: Part c) Classification of data | 20% |
| Implementation code: Part d) Complete code | 5% |
| Final Report with final results, analysis of results, challenges, code structure overview, self­assessment of learning, and possible future work items. | 20% |

**Work Plan (135 hours)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Meet?** | **Action** | **Hours** |
| 1 | Y | Project plan and expectations | 10 |
| 2/3 | Y | Performing Literature review and reporting results | 20 |
| 4 | Y | Environment Setup and Initial preprocessing of data set | 10 |
| 5 | Y | Data Cleansing and preprocessing | 10 |
| 6 | Y | Feature generation | 15 |
| 7/8 | Y | Classifying network Traffic- Building a binary classifier  - | 35 |
| 9/10/11 | Y | Compiling data and Creating report | 35 |