**BỘ THÔNG TIN VÀ TRUYỀN THÔNG**



**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**A red and white logo with a star

Description automatically generated with low confidence**

Báo cáo cuối kỳ môn:

AN TOÀN MẠNG NÂNG CAO

**Đề tài 9:**

**Phát hiện tấn công bằng phương pháp học máy**

Sinh viên thực hiện: Nguyễn Công Lực

MSSV: N19DCAT048

Lớp: D19CQAT01-N

Giáo viên hướng dẫn**:** Đàm Minh Lịnh

A. Tên đề tài: Phát hiện tấn công bằng phương pháp học máy

Nội dung thực hiện:

1 Đánh giá hiệu quả các mô hình: CNN, LSTM, CNN + LSTM, Markov; sau khi thực hiện Epoch = 200 cho mỗi mô hình, cho mỗi loại tấn công.

2 Nghiên cứu các dạng chiết suất đặc trưng: IFCM (Interval Fuzzy c-Means) và FCAI (Fuzzy Clustering Algorithm for Intervals)

3 Phương pháp đánh giá Confusion Matrix

4 Dùng mô hình ML/DL để huấn luyện dataset: NSL-KDD.

5 Thu thập và phát hiện dữ liệu các tấn công: nhận dạng dữ liệu bình thường; dữ liệu tấn công và đưa ra cảnh báo.

B.

Báo cáo đồ án cuối kỳ - Môn học An Toàn Mạng Nâng Cao

A. Tổng quan đề tài: Phát hiện tấn công bằng phương pháp học máy

I. Giới thiệu:

An Toàn Mạng là một lĩnh vực quan trọng trong hệ thống thông tin hiện đại. Trong đó, việc phát hiện tấn công là một nhiệm vụ cốt lõi để đảm bảo tính bảo mật của mạng. Trong đồ án này, chúng ta tập trung vào việc phát hiện tấn công bằng phương pháp học máy. Chúng ta sẽ đánh giá hiệu quả của các mô hình máy học như CNN, LSTM, CNN + LSTM, và Markov thông qua việc thực hiện huấn luyện và đánh giá trên dataset NSL-KDD. Đồng thời, chúng ta sẽ nghiên cứu các phương pháp chiết suất đặc trưng như IFCM (Interval Fuzzy c-Means) và FCAI (Fuzzy Clustering Algorithm for Intervals), cùng với phương pháp đánh giá Confusion Matrix để đánh giá hiệu suất của các mô hình.

II. Nội dung thực hiện:

1. Đánh giá hiệu quả các mô hình: CNN, LSTM, CNN + LSTM, Markov

a. Thực hiện Epoch = 200 cho mỗi mô hình và mỗi loại tấn công trên dataset NSL-KDD.

b. Đánh giá hiệu suất của từng mô hình thông qua các chỉ số đánh giá như độ chính xác, độ nhạy, độ đặc hiệu và F1-score.

2. Nghiên cứu các dạng chiết suất đặc trưng: IFCM và FCAI

a. IFCM (Interval Fuzzy c-Means): Dựa trên lý thuyết tập mờ và giá trị khoảng.

- Cho phép mỗi điểm dữ liệu có một giá trị khoảng, không chỉ một giá trị duy nhất.

- Phù hợp khi dữ liệu không chính xác hoặc không chắc chắn và có tính chất mờ.

b. FCAI (Fuzzy Clustering Algorithm for Intervals): Dựa trên lý thuyết tập mờ và giá trị khoảng.

- Tạo ra các cụm dữ liệu dựa trên tính mờ của các điểm dữ liệu.

- Cho phép sử dụng các giá trị khoảng để đo đạc sự tương đồng giữa các điểm dữ liệu.

c. So sánh và đánh giá hiệu suất của IFCM và FCAI trong việc phân loại tấn công mạng trên dataset NSL-KDD