**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THỰC PHẨM TP. HCM**

**KHOA: CÔNG NGHỆ THÔNG TIN**

-----o0o----

****

**ĐỀ CƯƠNG KHOÁ LUẬN**

**NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG DDOS**

|  |  |
| --- | --- |
| **Nhóm:**  1: Trần Minh Trí – 2033190135  2: Đặng Hoàng Duy – 2033190158  3: Đinh Phạm Quốc Quyền – 2033190112 | **Giảng viên hướng dẫn: Ths. Trần Đắc Tốt** |

**Thành phố Hồ Chí Minh, tháng 6 năm 2023**

**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THỰC PHẨM TP. HCM**

**KHOA: CÔNG NGHỆ THÔNG TIN**

-----o0o----

****

**ĐỀ CƯƠNG KHOÁ LUẬN**

**NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG DDOS**

|  |  |
| --- | --- |
| **Nhóm:**  1: Trần Minh Trí - 2033190135  2: Đặng Hoàng Duy – 2033190158  3: Đinh Phạm Quốc Quyền - 2033190112 | **Giảng viên hướng dẫn:**  **Ths. Trần Đắc Tốt** |

**Thành phố Hồ Chí Minh, tháng 6 năm 2023**

**NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG DDOS**

**Lời cảm ơn**

Lời đầu tiên chúng em xin gửi lời cảm ơn chân thành đến quý thầy, cô giáo trong khoa Công nghệ Thông Tin trường Đại học Công nghiệp thực phẩm Tp Hồ Chí Minh đã truyền đạt kiến thức và kinh nghiệm quý báu cho chúng em trong suốt quá trình học tập và rèn luyện .

Đặc biệt chúng em xin gởi lời cảm ơn đến thầy Trần Đắc Tốt, người đã trực tiếp hướng dẫn và tận tình giúp đỡ chúng em trong suốt thời gian học môn “Khóa Luận Tốt Nghiệp” cũng như trong thời gian chúng em thực hiện thực bài tập kết thúc môn học.

Mặc dù đã cố gắng trong quá trình làm bài nhưng không thể tránh khỏi những sai sót. Chúng em mong nhận được sự góp ý, nhận xét của quý thầy cô và các bạn về nội dung cũng như hình thức trình bày để bài báo cáo được hoàn thiện.

Chúng em xin chân thành cảm ơn thầy cô và các bạn đã xem.

Tp. Hồ Chí Minh, ngày 13 tháng 6 năm 2023

SV thực hiện SV thực hiện SV thực hiện

Trần Minh Trí Đặng Hoàng Duy Đinh Phạm Quốc Quyền

**MỤC LỤC**

[**PHẦN MỞ ĐẦU 1**](#_Toc137554925)

[**1. Lý do chọn đề tài 1**](#_Toc137554926)

[**2. Mục đích và nhiệm vụ nghiên cứu 2**](#_Toc137554927)

[**2.1. Mục đích nghiên cứu 2**](#_Toc137554928)

[**2.2. Nhiệm vụ nghiên cứu 2**](#_Toc137554929)

[**3. Phương pháp nghiên cứu 2**](#_Toc137554930)

[**4. Đối tượng và phạm vi nghiên cứu 3**](#_Toc137554931)

[**PHẦN NỘI DUNG 4**](#_Toc137554932)

[**CHƯƠNG 1: TỔNG QUAN VỀ DDOS VÀ BOTNET 4**](#_Toc137554933)

[**CHƯƠNG 2: NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC TRONG PHÁT HIỆN TẤN CÔNG DDOS 4**](#_Toc137554934)

[**CHƯƠNG 3: XÂY DỰNG MÔ HÌNH MÁY HỌC VÀ THỰC NGHIỆM 5**](#_Toc137554935)

[**CHƯƠNG 4: THỰC NGHIỆM MÔ HÌNH MÁY HỌC 5**](#_Toc137554936)

# PHẦN MỞ ĐẦU

# 1. Lý do chọn đề tài

Trong thời đại công nghệ 4.0 hiện nay thì hệ thống mạng lưới Internet đang phát triển rất nhanh ở nước ta. Bằng chứng là nhiều doanh nghiệp, trường học và các hộ gia đình đều đang sử dụng mạng Internet để phục vụ cho nhu cầu, lợi ích của riêng họ. Việc sử dụng mạng lưới Internet góp phần không nhỏ giúp kinh tế, chính trị các nước phát triển gắn bó với nhau do có thể kết nối thông qua mạng lưới. Không chỉ vậy, cuộc sống của mỗi người trong các công việc hằng ngày đang ngày một tối ưu, nhanh chóng hơn thông qua lợi ích của mạng lưới Internet.

Song song với sự phát triển và lợi ích của mạng lưới Internet thì cũng tồn tại không ít các mối nguy hại tiềm ẩn cho mạng lưới. Một trong các mối nguy hại ví dụ như là những cuộc tấn công vào hệ thống ngân hàng, trường học, tổ chức, cơ quan truyền thông nhằm đánh cắp các tài liệu mật, thông tin riêng tư, nhạy cảm nhằm phá hoại hoặc trục lợi cho bản thân của kẻ xấu.

Các cách tấn công phổ biến thường thấy hiện nay là trang web lừa đảo, email độc hại, virus phá hoại, vv. DDoS - tấn công từ chối dịch vụ phân tán là 1 kiểu tấn công phổ biến làm tắc nghẽn máy chủ đang hoạt động bằng cách yêu cầu quyền truy cập vào máy chủ với số lượng lớn, để làm được như vậy thì cần 1 đội máy tính lớn truy cập. Botnet ra đời để giải quyết điều này, tức 1 đội máy tính zombie mạng lưới được chỉ huy để gửi các gói tin truy cập vào máy chủ.

Hậu quả của các cuộc tấn công DDoS có thể gây ra tổn thất lớn về kinh tế, về tính tiện lợi của mạng Internet, tầm ảnh hưởng không chỉ có các doanh nghiệp mà còn ảnh hưởng tới người dùng Internet. Và hiện nay, nó đang được phát triển với nhiều loại, hình thức tấn công mới, biến thể mới, gây ra không ít thách thức lớn cho các nhà quản trị an ninh, bảo mật mạng.

Vì vậy việc phát hiện để ngăn chặn hay giảm thiểu các cuộc tấn công DDoS là một điều hết sức cần thiết trong giai đoạn Internet phát triển như hiện nay ở nước ta, do đó, chúng em quyết định lựa chọn đề tài cho cuối kì là “Nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDos”.

# 2. Mục đích và nhiệm vụ nghiên cứu

## **2.1. Mục đích nghiên cứu**

Mục tiêu nghiên cứu của nhóm là nghiên cứu về các thuật toán máy học có giám sát và các kỹ thuật xử lý, đồng thời cũng tìm ra cách để áp dụng mô hình máy học cho ra kết quả là một ứng dụng/ công cụ có thể phát hiện được cuộc tấn công DDoS.

## **2.2. Nhiệm vụ nghiên cứu**

Để thực hiện đề tài “Nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS” nhóm sẽ giải quyết các nhiệm vụ cơ bản sau:

* Tìm hiểu về nguồn gốc Botnet, cách chúng được tạo ra như thế nào, nhằm mục đích gì, các loại mô hình Botnet, có bao nhiêu loại và cách phân biệt loại Botnet, cách chúng lây nhiễm và cơ chế hoạt động của chúng. Làm sao để điều khiển Botnet và tìm ra giải pháp phòng chống.
* Tìm hiểu về học có giám sát và các thuật toán máy học phân loại có giám sát, các kỹ thuật sử dụng trong tiền xử lý dữ liệu.
* Sau khi có hướng giải pháp thì xây dựng mô hình máy học, sử dụng bộ dữ liệu gì, giải thích thực nghiệm triển khai mô hình máy học và tiến hành đánh giá kết quả.

# 3. Phương pháp nghiên cứu

Đầu tiên nhóm sẽ chủ động nghiên cứu tham khảo tài liệu (trong hoặc ngoài nước), đọc hiểu và chọn lọc các dữ liệu nội dung có nguồn gốc xuất sứ rõ ràng để chuẩn bị cho bài làm.

Trong quá trình làm sẽ có không ít các khó khăn và thử thách đối với nhóm, với bài toán phát hiện tấn công từ chối dịch vụ phân tán DDoS thì khó khăn đầu tiên ở đây đó là nhóm phải có tập dữ liệu đầu vào cho các thuật toán máy học, vì trong thực tế các tập dữ liệu này thường chỉ là tập dữ liệu mô phỏng, không theo một chuẩn nhất định và có thể bị nhiễu. Thứ hai là việc mất cân bằng trong các tập dữ liệu khiến cho các thuật toán máy học có thể dự đoán sai kết quả khi đưa vào dữ liệu trong thực tế. Cho nên để tối ưu các thuật toán máy học sẽ tốn rất nhiều thời gian và kết quả có thể không chính xác tuyệt đối khi đưa vào dữ liệu.

Phương pháp mà nhóm đề xuất tiếp cận để phát hiện tấn công từ chối dịch vụ phân tán DDoS sẽ thông qua phát hiện lưu lượng mạng Botnet trên mô hình máy học bởi vì tấn công DDoS được thực hiện chủ yếu thông qua mạng Botnet. Thêm vào đó nhóm sẽ áp dụng thêm các kỹ thuật xử lý dữ liệu cho bộ dữ liệu để tăng độ chính xác cho các dữ liệu máy học. Giới hạn của nhóm trong đề tài sẽ là phát hiện tấn công DDoS từ mạng Botnet.

# 4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu chính ở đây chính là những thuật toán máy học đã có sẵn, ta sẽ nghiên cứu chúng và áp dụng để phát hiện tấn công DDoS.

Phạm vi về không gian:

+ Trong nước, với các doanh nghiệp cty, các hộ gia đình có nhu cầu, những cá nhân hoặc nhóm người sử dụng mạng lưới Internet.

Phạm vi về thời gian: việc nghiên cứu tài liệu, tìm ra được giải pháp và tiến hành thực nghiệm trong khoảng thời gian 1 năm (2023-2024).

# PHẦN NỘI DUNG

## [**CHƯƠNG 1: TỔNG QUAN VỀ DDOS VÀ BOTNET**](#_heading=h.2et92p0)

[1.1 Giới thiệu về Botnet](#_heading=h.tyjcwt)

[1.2 Giới thiệu mô hình của Botnet](#_heading=h.3dy6vkm)

[1.2.1](#_heading=h.1t3h5sf) Mô hình Star

[1.2.2](#_heading=h.17dp8vu) Mô hình Random

[1.2.3](#_heading=h.26in1rg) Mô hình Multi - Server

[1.2.4](#_heading=h.35nkun2) Mô hình Hierachical

[1.3 Phân loại botnet](#_heading=h.44sinio)

[1.3.1](#_heading=h.2jxsxqh) Phân loại theo chức năng hoặc mục đích

[1.3.2](#_heading=h.z337ya) Phân loại botnet dựa trên mô hình của chúng

[1.3.3](#_heading=h.3j2qqm3) Phân loại theo nền tảng hệ điều hành để tấn công mục tiêu

[1.3.4](#_heading=h.1y810tw) Phân loại dựa theo phần mềm độc hại được cài đặt trên botnet

[1.4 Cách Botnet lây nhiễm](#_heading=h.2xcytpi)

[1.4.1](#_heading=h.1ci93xb) Đính kèm mã độc lừa người dùng chạy chúng

[1.4.2](#_heading=h.3whwml4) Tấn công vào các lỗ hổng chưa được vá (Zero-day)

[1.4.3](#_heading=h.2bn6wsx) Backdoor nạn nhân bằng Trojan

[1.4.4](#_heading=h.qsh70q) Khả năng lây nhiễm của botnet

[1.5 Cơ chế hoạt động của botnet](#_heading=h.1pxezwc)

[1.6](#_heading=h.2p2csry) Điều khiển botnet

[1.7 Giải pháp phòng chống DDoS](#_heading=h.41mghml)

## [**CHƯƠNG 2: NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC TRONG PHÁT HIỆN TẤN CÔNG DDOS**](#_heading=h.2grqrue)

[2.1 Học có giám sát (Supervised Learining)](#_heading=h.vx1227)

[2.2 Các thuật toán máy học phân loại trong học có giám sát](#_heading=h.3fwokq0)

[2.2.1](#_heading=h.1v1yuxt) Decision Tree

[2.2.2](#_heading=h.2u6wntf) Random Forest

[2.2.3](#_heading=h.3tbugp1) Navie Bayes

[2.2.4](#_heading=h.37m2jsg) K-nearest neighbors (KNN)

[2.2.5](#_heading=h.2lwamvv) Logistic Regression

[2.2.6](#_heading=h.3l18frh) XGBoost

[2.3 Các kỹ thuật sử dụng trong tiền xử lý dữ liệu](#_heading=h.206ipza)

[2.3.1](#_heading=h.4k668n3) Kỹ thuật xử lý các giá trị rỗng trong bộ dữ liệu

[2.3.2](#_heading=h.2zbgiuw) Xử lý các biến giá trị phân loại

## [**CHƯƠNG 3: XÂY DỰNG MÔ HÌNH MÁY HỌC VÀ THỰC NGHIỆM**](#_heading=h.2dlolyb)

[3.1 Xây dựng mô hình máy học](#_heading=h.sqyw64)

[3.2 Bộ dữ liệu CTU-13](#_heading=h.1rvwp1q)

[3.3 Giải thích thực nghiệm triển khai mô hình máy học](#_heading=h.25b2l0r)

[3.3.1](#_heading=h.kgcv8k) Tiền xử lý dữ liệu

[3.3.2](#_heading=h.43ky6rz) Mô hình kịch bản sử dụng để thực nghiệm

[3.4 Tiến hành đánh giá kết quả thực nghiệm](#_heading=h.xvir7l)

## **CHƯƠNG 4: THỰC NGHIỆM MÔ HÌNH MÁY HỌC**

4.1 Tổng quan về mô hình thực nghiệm

4.2 Xây dựng máy Bot header để điều khiển botnet

4.3 Tạo botnet để phân tán cho các Bot Client

4.4 Giả lập tấn công DDoS và cách phát hiện