


# THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://github.com/CaoBaKietIT/PP-NCKH>  
(ví dụ: <https://www.youtube.com/watch?v=AWq7uw-36Ng>)
- Link slides (dạng .pdf đặt trên Github):  
<https://github.com/CaoBaKietIT/PP-NCKH/blob/main/Cao%20B%C3%A1%20Ki%E1%BB%87t%20-%20xCS2205.DeCuong.FinalReport.Template.Slide%20.pdf>  
(ví dụ: <https://github.com/mynameuit/CS2205.APR2023/TenDeTai.pdf>)
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none"><li>● Họ và Tên: Cao Bá Kiệt</li><li>● MSSV: 230202008</li></ul> 	<ul style="list-style-type: none"><li>● Lớp: CS2205.CH181</li><li>● Tự đánh giá (điểm tổng kết môn): 9/10</li><li>● Số buổi vắng: 0</li><li>● Số câu hỏi QT cá nhân: 3</li><li>● Link Github: <a href="https://github.com/mynameuit/CS2205.APR2023/">https://github.com/mynameuit/CS2205.APR2023/</a></li></ul>
--	---

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

KỸ THUẬT LỪA DỐI VÀ TĂNG CƯỜNG DỮ LIỆU DỰA TRÊN CYCLEGAN CHO PHÁT HIỆN TẤN CÔNG TRONG MẠNG KHẢ LẬP TRÌNH

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CYBER DECEPTION AND CYCLEGAN-BASED DATA AUGMENTATION FOR INTRUSION DETECTION IN SDN-ENABLED NETWORK

## TÓM TẮT *(Tối đa 400 từ)*

Cùng với sự phát triển của vượt bậc của trí tuệ nhân tạo, học máy được ứng dụng để xây dựng các hệ thống phát hiện xâm nhập (IDS) có khả năng phát hiện các cuộc tấn công chưa từng được biết tới với độ chính xác khá cao. Tuy nhiên, khả năng của IDS học máy lại bị giới hạn bởi tình trạng thiếu dữ liệu huấn luyện. Hơn nữa, các tập dữ liệu huấn luyện thường bị mất cân bằng do số lượng dữ liệu bình thường lớn hơn nhiều lần so với dữ liệu độc hại.

Trong phạm vi khóa luận, để tăng cường dữ liệu độc hại huấn luyện cho IDS, chúng tôi sử dụng hai phương pháp là thu thập thêm dữ liệu và tạo ra dữ liệu giả. Cụ thể, để giúp thu thập thêm dữ liệu một cách hiệu quả, tiết kiệm, trên mạng SDN chúng tôi xây dựng hệ thống bẫy áp dụng kỹ thuật Cyber Deception nhằm thiết lập phòng thủ chủ động trước các cuộc tấn công mạng và thu thập lưu lượng độc hại của kẻ tấn công. Để tạo ra dữ liệu giả, chúng tôi thực nghiệm trên tập dữ liệu CICIDS2017 và mô hình CycleGAN. Để phục vụ cho việc đánh giá, một IDS (Imb-IDS) được huấn luyện trên tập dữ liệu mất cân bằng ban đầu làm cơ sở để so sánh với một IDS khác (Cycle-IDS) được huấn luyện trên tập dữ liệu cân bằng có bổ sung dữ liệu tạo ra bởi CycleGAN. Kiến trúc Multilayer perceptron được sử dụng cho các IDS.

Với hệ thống bẫy được xây dựng trên mạng SDN, chúng tôi chuyển hướng lưu lượng độc hại của cuộc tấn công, góp phần giảm rủi ro cho hệ thống mạng, đồng thời thu thập được lưu lượng độc hại và trích xuất các thuộc tính quan trọng lưu thành file csv để tăng cường dữ liệu cho IDS học máy. Còn việc sử dụng CycleGAN cũng đem lại kết quả khá quan trọng trong việc cải thiện khả năng phát hiện dữ liệu độc hại của IDS học máy.

## GIỚI THIỆU *(Tối đa 1 trang A4)*

Để bảo vệ cho hệ thống công nghệ thông tin của mình, các doanh nghiệp sử dụng các giải pháp phòng thủ khác nhau trong đó IDS là một phần mềm hoặc thiết bị được sử dụng để

giám sát, phân tích lưu lượng mạng trong thời gian thực, qua đó phát hiện các cuộc tấn công và cảnh báo cho quản trị viên.

Gần đây, với sự phát triển bùng nổ của trí tuệ nhân tạo, học máy đã được ứng dụng cho IDS. So với IDS truyền thống, IDS học máy có khả năng phát hiện các tấn công chưa từng được biết đến hoặc các bất thường với độ chính xác khá cao. Tuy nhiên, khả năng của chúng bị giới hạn bởi chất lượng của tập dữ liệu huấn luyện. Trong khi đó, vẫn còn tồn tại rất nhiều tập dữ liệu huấn luyện cho IDS ngày nay bị mất cân bằng do các dữ liệu tấn công ít hơn nhiều so với dữ liệu bình thường, phản ánh đúng với phân phối của lưu lượng mạng thực tế.

Để giải quyết tình trạng này, mạng sinh đối kháng (GAN) với khả năng sinh dữ liệu của mình tỏ ra triển vọng trong việc bổ sung dữ liệu cho các tập dữ liệu. Điều này đặc biệt có lợi trong các lĩnh vực mà quá trình tạo ra dữ liệu tốn kém hay cần các tập dữ liệu cân bằng.

Bên cạnh IDS, Cyber Deception cũng là một giải pháp hữu ích trong việc phòng thủ, hạn chế mức độ ảnh hưởng của các cuộc tấn công tới hệ thống công nghệ thông tin trong doanh nghiệp. Nó giúp đánh lừa kẻ tấn công, cố gắng gây ra sự nhầm lẫn để kẻ tấn công tin rằng đây là mục tiêu thật sự và ra sức tấn công vào nó. Qua đó chúng ta vừa có thể hạn chế thiệt hại, thu thập được các bằng chứng, các chiến thuật, kỹ thuật vừa thu được lưu lượng tấn công để phục vụ mục đích tái huấn luyện IDS

## MỤC TIÊU

*(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)*

Xây dựng và đánh giá IDS học máy có khả năng phân loại dữ liệu bình thường và độc hại trên tập dữ liệu bị mất cân bằng CICIDS2017 (Imbalanced IDS).

Nghiên cứu, xây dựng và đánh giá CycleGAN trong việc chuyển đổi dữ liệu bình thường sang độc hại trên tập dữ liệu CICIDS2017 nhằm bổ sung dữ liệu độc hại vào tập dữ liệu ban đầu để tái huấn luyện cho IDS (CycleGAN IDS).

So sánh hiệu năng của CycleGAN IDS với Imbalanced IDS

Triển khai Cyber Deception trong mạng khả lập trình để đánh lừa kẻ tấn công xâm nhập vào bẫy nhằm cung cấp chiến lược phòng thủ chủ động, thu thập IOCs, các chiến thuật, chiến lược tấn công và thu thập dữ liệu tấn công làm cơ sở để huấn luyện củng cố khả năng phát hiện các cuộc tấn công của IDS học máy

## NỘI DUNG VÀ PHƯƠNG PHÁP

*(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)*

**Nội dung 1:** Hệ thống phát hiện xâm nhập dựa trên phương pháp học máy

- Tìm hiểu cơ chế hoạt động của IDS dựa trên phương pháp học máy

- Tìm hiểu về ưu điểm và nhược điểm của IDS dựa trên phương pháp học máy
- Xây dựng mô hình IDS dựa trên phương pháp học máy

**Nội dung 2:** Tìm hiểu về Neural Network, Generative Adversarial Network (GAN) và Cycle Adversarial Network (CycleGAN)

- Tìm hiểu về Neural Network và các thuật toán của nó.
- Tìm hiểu về Convolutional Neural Network (CNN).
- Tìm hiểu về cấu trúc, cơ chế hoạt động của GAN.
- Tìm hiểu phương pháp tấn công đối kháng (Adversarial Attack) bằng GAN để vượt mặt IDS.
- Tìm hiểu về cấu trúc, cơ chế hoạt động của CycleGAN

**Nội dung 3:** Xây dựng hệ thống Cyber deception trong mạng khả lập trình

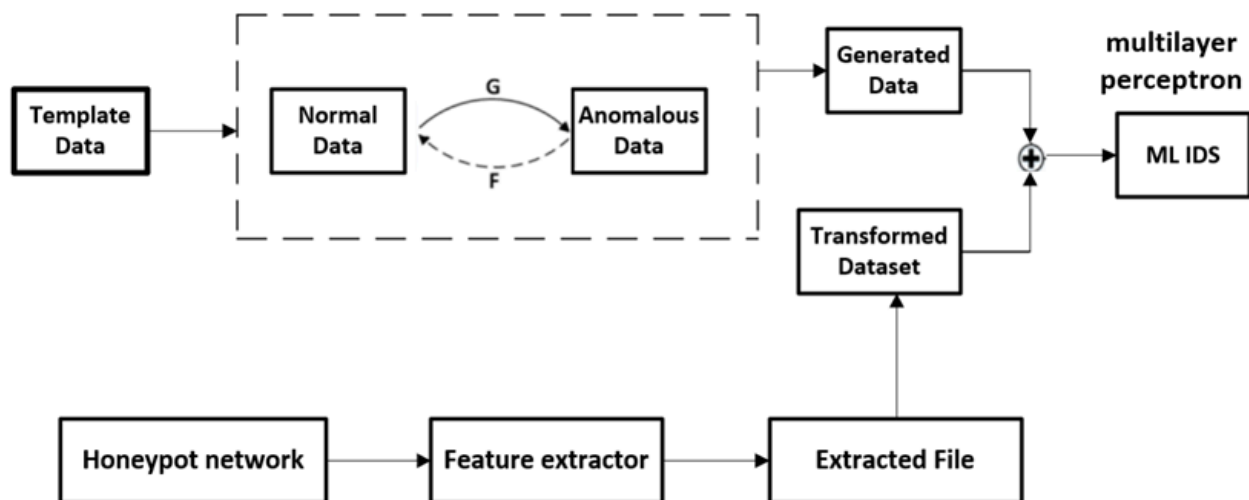
- Nắm được kiến thức tổng quan và các công nghệ chính của kỹ thuật lừa dối mạng
- Triển khai kỹ thuật lừa dối mạng trong mạng khả lập trình nhằm tạo bẫy cho kẻ tấn công.
- Xây dựng kênh giao tiếp tương tác giữa Bẫy lừa dối mạng ( Cyber Deception) và IDS cho mạng khả lập trình
- Xây dựng hệ thống Cyber Deception nhằm chủ động hơn trong việc phòng thủ giảm thiểu mức độ nguy hiểm cho hệ thống, đồng thời giám sát xuyên suốt quá trình xâm nhập, nắm bắt được mục tiêu chiến thuật, kỹ thuật của kẻ tấn công đang nhắm đến hệ thống của tổ chức, từ đó thu nhập nhật ký tấn công (log) và trích xuất các thuộc tính để phục vụ quá trình tài huấn luyện IDS cho học máy.

**Nội dung 4:** Tăng cường dữ liệu dựa trên Cyclegan cho phát hiện tấn công.

- Xây dựng ba Multi layer Perceptron phân loại dữ liệu trong tập dữ liệu CICIDS2017: Imbalanced IDS, Class weight IDS và Oversampling IDS.
- Nghiên cứu khả năng của CycleGAN trong việc chuyển đổi dữ liệu bình thường sang dữ liệu độc hại trong tập dữ liệu CICIDS2017
- So sánh hiệu năng của mô hình Multi Layer Perceptron IDS được tái huấn luyện trên tập dữ liệu được bổ sung thêm dữ liệu sinh ra bởi CycleGAN (CycleGAN IDS) với mô hình ban đầu.

**Phương pháp thực hiện:**

- Chúng tôi thiết kế mô hình kết hợp Cyber deception với CycleGAN trong việc tăng cường dữ liệu độc hại để huấn luyện cải thiện khả năng phân loại của IDS học máy.



- IDS học máy ban đầu được huấn luyện trên một tập dữ liệu dành cho IDS học máy mà ta lựa chọn (gọi là Original Dataset). Original Dataset được sao chép để tạo thành một tập dữ liệu biến động gọi là Transformed Dataset.
- Với Cyber deception, chúng tôi xây dựng hệ thống honeypot được đặt song song với hệ thống thật trong mạng SDN. Ngay khi nhận được cảnh báo lưu lượng bất thường, controller sẽ lập tức điều hướng kết nối của kẻ tấn công sang honeypots và công cụ trích xuất các thuộc tính của lưu lượng mạng sẽ tự động trích xuất các lưu lượng độc hại của kẻ tấn công và lưu thành file (gọi là extracted file).
- Sau một khoảng thời gian định kỳ nào đó (gọi là  $t_1$ ), ta lại thêm các dữ liệu trong extracted file vào Transformed Dataset.
- CycleGAN có nhiệm vụ học cách chuyển đổi dữ liệu mạng bình thường sang dữ liệu mạng độc hại trên Transformed Dataset. Do cứ sau một khoảng thời gian  $t_1$  thì Transformed Dataset sẽ bị thay đổi, nên CycleGAN sẽ được huấn luyện lại từ đầu trên tập huấn luyện mới. Trong quá trình huấn luyện, các giá trị mất mát của hai bộ sinh và hai bộ phân biệt và trọng số của CycleGAN sẽ được lưu lại sau mỗi epoch.
- Tiếp theo, ta lựa chọn một khoảng thời gian định kỳ  $t_2$  ( $t_2 < t_1$ ) cho việc sử dụng CycleGAN để sinh ra dữ liệu độc hại và tái huấn luyện IDS. Trước khi sinh dữ liệu, trọng số tốt nhất (ở epoch đã được huấn luyện có tổng mất mát nhỏ nhất) của CycleGAN sẽ được tải lại, và bộ sinh chuyển đổi dữ liệu bình thường sang độc hại sẽ chuyển một phần dữ liệu bình thường trong Transformed Dataset (Template Data) thành dữ liệu độc hại (Generated Data). IDS sẽ được tái huấn luyện trên tập dữ liệu bao gồm cả Transformed Dataset và Generated Data.
- Phân tích và đánh giá kết quả kiểm thử dựa trên các tiêu chí như số lỗi phát hiện được, độ bao phủ mã, thời gian thực hiện, v.v. và so sánh với các nghiên cứu liên quan.

## KẾT QUẢ MONG ĐỢI

*(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)*

- Triển khai IDS dựa trên phương pháp học máy
- Triển khai Cyber Deception trong mạng khả lập trình để đánh lừa kẻ tấn công, từ đó thu thập được các IOCs, các chiến thuật, chiến lược và dữ liệu tấn công nhằm xây dựng hệ thống phòng thủ chủ động, dữ liệu tấn công làm cơ sở để tái huấn luyện mô hình để củng cố khả năng phát hiện tấn công.
- Model CycleGAN có khả năng sinh dữ liệu độc hại để vượt mặt IDS và tái huấn luyện IDS.

## TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1] Trung Thanh Bui, “SOFTWARE DEFINED NETWORKING – CÔNG NGHỆ MỚI LÀM THAY ĐỔI CẤU TRÚC MẠNG”, 2006.

[2] E. Al-Shaer, J. Wei, K. W. Hamlen, C. Wang, “Autonomous Cyber Deception”, pages 102-119, 2018.

[3] W. Han, Z. Zhao, A. Doupe et al., “HoneyMix: Toward SDN-based Intelligent Honeynet”, 2016.

[4] C. Wang and Z. Lu, “Cyber Deception: Overview and the Road Ahead”, 2018.

[5] Z. Li, Z. Qin, K. Huang, X. Yang and S. Ye, "Intrusion Detection Using Convolutional Neural Networks for Representation Learning," in International Conference on Neural Information Processing ICONIP, 2017.

[6] Tang T.A., McLernon D., Mhamdi L., Zaidi S.A.R., Ghogho M., “Intrusion Detection in SDN-Based Networks: Deep Recurrent Neural Network Approach,” in Advanced Sciences and Technologies for Security Applications, volume Deep Learning Applications for Cyber Security, Springer, Cham, 2019.