HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Phan Thị Kim Hoàn

NGHIÊN CỨU CÁC PHƯƠNG PHÁP SINH DỮ LIỆU TẤN CÔNG MẠNG

Chuyên ngành : Hệ thống thông tin

Mã số: 8.48.01.04

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

HÀ NỘI – 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: PGS.TS Hoàng Xuân Dậu

(Ghi rõ học hàm, học vị)

Phản biện 1: ……

Phản biện 2: ……

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng .... năm .....

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỤC LỤC

I. MỞ ĐẦU.....1

1. Lý do chọn đề tài.....1

2. Tổng quan về vấn đề nghiên cứu 1

3. Mục đích nghiên cứu 2

4. Đối tượng và phạm vi nghiên cứu 2

5. Phương pháp nghiên cứu.....3

6. Bố cục của đề tài 3

II. NỘI DUNG.....4

CHƯƠNG 1: TỔNG QUAN VỀ SINH DỮ LIỆU TẤN CÔNG MẠNG 4

1.1. Tổng quan về dữ liệu tấn công mạng .4

1.2. Các bộ dữ liệu tấn công mạng phổ biến 4

1.3. Mục tiêu và nhu cầu của việc sinh dữ liệu tấn công mạng.....5

1.4. Thách thức trong sinh dữ liệu tấn công mạng.....5

1.5. Các nghiên cứu liên quan về sinh dữ liệu tấn công mạng.....5

1.6. Kết chương.....5

CHƯƠNG 2: CÁC KỸ THUẬT SINH DỮ LIỆU TẤN CÔNG MẠNG 6

2.1. Khái quát về các phương pháp sinh dữ liệu tấn công mạng.....6

2.2. Các phương pháp sinh dữ liệu truyền thống 6

2.3. Sinh dữu liệu bằng phương pháp học máy, học sâu. 6

2.4. Kết chương.....9

CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ.....10

3.1. Mô hình thử nghiệm.....10

3.2. Các kịch bản thử nghiệm và kết quả ...12

3.3. Nhận xét và đánh giá chung 16

3.4. Kết chương 17

III. KẾT LUẬN 18

1. Kết luận chung.....18

2. Hướng nghiên cứu tiếp theo 18

I. MỞ ĐẦU

1. Lý do chọn đề tài

Các hệ thống IDS đóng vai trò quan trọng trong phát hiện tấn công mạng, nhưng hiệu quả của mô hình học máy/học sâu bị hạn chế bởi thiếu dữ liệu tấn công chất lượng và mất cân bằng dữ liệu. Sinh dữ liệu tấn công mạng trở thành hướng tiếp cận cần thiết nhằm mở rộng dữ liệu huấn luyện, tăng tính đa dạng và cải thiện khả năng phát hiện tấn công, đặc biệt đối với các lớp hiếm.

2. Tổng quan về vấn đề nghiên cứu

2.1. Các phương pháp sinh dữ liệu tấn công mạng

Có bốn nhóm phương pháp chính: SMOTE (truyền thống) giúp cân bằng dữ liệu đơn giản; các mô hình GAN như CTGAN, WGAN sinh dữ liệu bảng phức tạp gần giống thực; phương pháp dựa trên luật cho phép kiểm soát theo kịch bản tấn công; phương pháp mô phỏng tạo dữ liệu sát thực tế nhưng tốn chi phí. Xu hướng hiện nay là kết hợp nhiều phương pháp để tạo dữ liệu đa dạng và hỗ trợ tốt hơn cho IDS.

2.2. Khảo sát một số nghiên cứu liên quan

Các nghiên cứu gần đây cho thấy dữ liệu tổng hợp ứng dụng học máy học sâu điển hình như GAN cải thiện đáng kể Recall và F1-score trong phát hiện tấn công, đặc biệt ở các lớp hiếm. Nhiều công trình chứng minh GAN và SMOTE giúp cân bằng dữ liệu và nâng cao độ chính xác của mô hình trên các bộ dữ liệu như CIC-IDS2017, UNSW-NB15 và IoTID20.

3. Mục đích nghiên cứu

Đề tài nhằm hệ thống hóa các phương pháp sinh dữ liệu tấn công, triển khai ba kỹ thuật tiêu biểu (SMOTE, CGAN, WGAN), thực nghiệm trên dữ liệu thực tế và đánh giá mức độ hữu ích của dữ liệu sinh đối với các mô hình phát hiện tấn công mạng.

4. Đối tượng và phạm vi nghiên cứu

Nghiên cứu tập trung vào các kỹ thuật sinh dữ liệu phục vụ phát hiện tấn công mạng; triển khai ba phương pháp SMOTE, CGAN và WGAN trên hai bộ dữ liệu UNSW-NB15 và CIC-IDS2017; sử dụng mô hình học máy để đánh giá chất lượng dữ liệu sinh.

5. Phương pháp nghiên cứu

Đề tài sử dụng phân tích tài liệu, xây dựng và triển khai thử nghiệm các phương pháp sinh dữ liệu, sau đó đánh giá – so sánh dựa trên phân phối dữ liệu, độ đa dạng, mức độ giống dữ liệu thật và hiệu năng mô hình IDS.

6. Bố cục của đề tài

Đề án được bố cục thành 3 chương như sau:

Chương 1: Tổng quan về sinh dữ liệu tấn công mạng

Chương 2: Các kỹ thuật sinh dữ liệu tấn công mạng

Chương 3: Thử nghiệm và đánh giá kết quả

II. NỘI DUNG

CHƯƠNG 1: TỔNG QUAN VỀ SINH DỮ LIỆU TẤN CÔNG MẠNG

1.1. Tổng quan về dữ liệu tấn công mạng

1.1.1. Khái quát về tấn công mạng

a, Khái niệm tấn công mạng

b, Mục tiêu của các cuộc tấn công mạng

c, Phân loại tấn công mạng

d, Tác động và xu hướng tấn công mạng hiện nay

1.1.2. Dữ liệu tấn công mạng

a, Khái niệm dữ liệu tấn công mạng

b, Đặc điểm của dữ liệu tấn công mạng

c, Vai trò của việc thu thập và chuẩn hóa dữ liệu tấn công mạng

1.2. Các bộ dữ liệu tấn công mạng phổ biến

1.2.1. Bộ dữ liệu UNSW-NB15

1.2.2. Bộ dữ liệu CIC-UNSW-NB15 Augmented DataSet

1.2.3. Bộ dữ liệu CIC-IDS-2017

1.2.4. Một số bộ khác (NSL-KDD, BoT-IoT, TON_IoT)

1.2.5. So sánh các bộ dữ liệu

1.3. Mục tiêu và nhu cầu của việc sinh dữ liệu tấn công mạng

1.3.1. Mục tiêu của sinh dữ liệu trong hệ thống IDS

1.3.2. Nhu cầu sinh dữ liệu tấn công mạng trong các lĩnh vực.

1.4. Thách thức trong sinh dữ liệu tấn công mạng

1.4.1. Thiếu hụt dữ liệu thực tế chất lượng cao

1.4.2. Mất cân bằng dữ liệu

1.4.3. Vấn đề riêng tư và bảo mật

1.4.4. Đặc trưng phức tạp và không đồng nhất

1.5. Các nghiên cứu liên quan về sinh dữ liệu tấn công mạng

1.5.1. Các nghiên cứu sử dụng phương pháp truyền thống

1.5.2. Các nghiên cứu sử dụng mô hình học máy, học sâu.

1.6. Kết chương

CHƯƠNG 2: CÁC KỸ THUẬT SINH DỮ LIỆU TẤN CÔNG MẠNG

2.1. Khái quát về các phương pháp sinh dữ liệu tấn công mạng

2.1.1. Phân loại các phương pháp sinh dữ liệu

2.1.2. Ưu nhược điểm, mức độ phổ biến

2.2. Các phương pháp sinh dữ liệu truyền thống

2.2.1. Phương sinh dữ liệu dựa trên lấy mẫu (sampling)

a, Oversampling (SMOTE)

b, Undersampling

2.2.2. Phương pháp sinh dữ liệu dựa trên mô phỏng

2.2.3. Phương pháp sinh dữ liệu dựa trên luật

2.2.4. So sánh các kỹ thuật

2.3. Sinh dữu liệu bằng phương pháp học máy, học sâu.

2.3.1. Tổng quan về học máy, học sâu

a, Khái niệm, ứng dụng, một số thuật toán học máy và học sâu.

b, Lợi thế của học máy, học sâu so với truyền thống

2.3.2. Mạng đối kháng tạo sinh (Generative Adversarial Networks – GANs)

a, Mạng Generative Adversarial Networks

GAN gồm hai mạng Generator và Discriminator hoạt động đối kháng: G nhận nhiễu để sinh dữ liệu giống thật, trong khi D phân biệt thật – giả; hai mạng được huấn luyện đồng thời cho đến khi D không còn phân biệt được, giúp G học được phân phối dữ liệu gốc.

b, Mạng WGAN, WGAN-GP.

WGAN cải thiện GAN bằng cách thay thế độ đo JSD bằng khoảng cách Wasserstein, cho gradient mượt hơn, giảm mode collapse và ổn định quá trình huấn luyện; Critic thay cho Discriminator và dùng weight clipping để đảm bảo tính Lipschitz. WGAN-

GP tiếp tục hoàn thiện bằng cơ chế Gradient Penalty thay vì clipping, giúp Critic học tốt hơn và đảm bảo Lipschitz mềm dẻo, từ đó tăng chất lượng dữ liệu sinh.

c, Mạng Conditional GAN (cGAN)

cGAN bổ sung biến điều kiện y vào cả Generator và Discriminator, cho phép mô hình sinh dữ liệu theo từng lớp hoặc thuộc tính cụ thể thay vì ngẫu nhiên như GAN gốc. Nhờ kết hợp điều kiện trong kiến trúc và hàm mất mát, cGAN học phân phối có điều kiện p(x|y), giúp kiểm soát chất lượng và tính định hướng của dữ liệu sinh.

d, Mạng Conditional Tabular GAN (CTGAN)

CopulaGAN chuẩn hóa dữ liệu vào không gian copula để tách phân phối biên và phụ thuộc đa biến, sau đó huấn luyện GAN trong không gian này rồi biến

đổi ngược để thu dữ liệu tổng hợp có cấu trúc phụ thuộc chính xác.

2.3.3. Bộ mã hóa tự động biến thiên (Variational Autoencoders -VAEs)

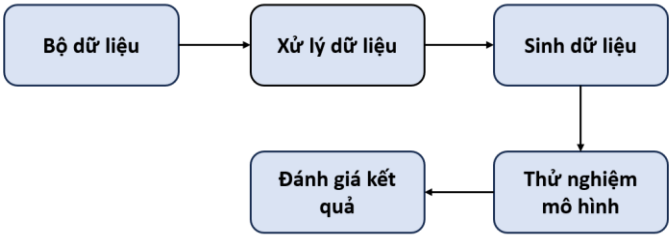
2.3.4. Mô hình ngôn ngữ lớn (Large Language Models -LLMs)

2.4. Kết chương

CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. Mô hình thử nghiệm

3.1.1. Giới thiệu mô hình thử nghiệm



Hình 3.1: Sơ đồ mô hình thực hiện đánh giá phương pháp sinh dữ liệu

Input: Bộ dữ liệu tấn công mạng.

Output: Kết quả thử nghiệm của mô hình phân loại.

Quy trình thực hiện:

Bước 1: Xử lý dữ liệu.

Bước 2: Sinh dữu liệu.

Bước 3: Thử nghiệm trên mô hình.

Bước 4: Đánh giá kết quả.

3.1.2. Thiết lập môi trường thực nghiệm

3.1.3. Các độ đo đánh giá

Các độ đo phổ biến được tính toán dựa trên các tham số của ma trận nhầm lẫn (confusion matrix) gồm: TP (True Positive – dương tính thật), TN (True Negative – âm tính thật), FP (False Positive – dương tính giả) và FN (False Negative – âm tính giả)

- Accuracy: Độ chính xác tổng thể.
- Precision: Độ chính xác dương tính.
- Recall: Độ nhạy, Tỷ lệ phát hiện dương tính.
- F1-score: Điểm hài hòa giữa Precision và Recall.
- FPR: False Positive Rate – Tỷ lệ dương tính giả.
- FNR: False Negative Rate – Tỷ lệ âm tính giả.

3.2. Các kịch bản thử nghiệm và kết quả

Kịch bản 1: Sử dụng các nhãn tấn công số lượng lớn để tiến hành thử nghiệm và đánh giá các phương pháp sinh.

- Chọn 1 nhãn tấn công có nhiều mẫu kết hợp với dữ liệu bình thường làm tập gốc. Sau đó chia 75% cho tập huấn luyện (ORIGIN_TRAIN) và 25% cho tập kiểm tra (ORIGIN_TEST).
- Sử dụng dữ liệu bình thường và tấn công trong ORIGIN_TRAIN để sinh dữ liệu mới. Tạo GEN_TRAIN với số mẫu mỗi nhãn tương đương ORIGIN_TRAIN.
- Huấn luyện mô hình gốc: ORIGIN_MODEL với (ORIGIN_TRAIN, ORIGIN_TEST). Huấn luyện mô hình sinh: GEN_MODEL với (GEN_TRAIN, ORIGIN_TEST).
- Cả ORIGIN_MODEL và GEN_MODEL đều được kiểm tra trên ORIGIN_TEST. So sánh hiệu suất theo các độ đo phân loại để đánh giá chất lượng phương pháp sinh dữ liệu.

Kịch bản 2: Xử lý mất cân bằng dữ liệu: Thực hiện dựa trên kết quả từ Kịch bản 1, sử dụng phương pháp sinh phù hợp nhất để thực hiện trên nhãn dữ liệu hiếm.

3.2.1. Kịch bản 1: Thử nghiệm và đánh giá phương pháp sinh dữ liệu.

Bảng 0.1: Hiệu suất phát hiện của mô hình ORIGIN_MODEL và các mô hình GEN_MODEL cho các nhãn tấn công trên tập dữ liệu UNSW-NB15

TT	Loại tấn công	Phương Pháp sinh	Accuracy	Precision	Recall	F1
1	Exploits	Thực	0.9969	0.9942	0.9996	0.9969
2	Exploits	SMOTE	0.9967	0.994	0.9996	0.9968
3	Exploits	CGAN	0.8028	0.9833	0.6159	0.7574
4	Exploits	WGAN	0.9905	0.9835	0.9978	0.9906
5	Reconnissance	Thực	0.9991	0.9982	1	0.9991
6	Reconnissance	SMOTE	0.9986	0.9975	0.9996	0.9986
7	Reconnissance	CGAN	0.9430	0.9836	0.9010	0.9405
8	Reconnissance	WGAN	0.9887	0.9844	0.9932	0.9888
9	Fuzzers	Thực	0.9935	0.9878	0.9994	0.9935
10	Fuzzers	SMOTE	0.9931	0.9876	0.9988	0.9931
11	Fuzzers	CGAN	0.9902	0.9831	0.9975	0.9903
12	Fuzzers	WGAN	0.9907	0.9853	0.9963	0.9908
13	Dos	Thực	0.9974	0.9954	0.9994	0.9974
14	Dos	SMOTE	0.9972	0.9954	0.9991	0.9973
15	Dos	CGAN	0.9888	0.981	0.9969	0.9889
16	Dos	WGAN	0.9879	0.9825	0.9936	0.9880

Bảng 3.2: Hiệu suất phát hiện của mô hình ORIGIN_MODEL và các mô hình GEN_MODEL cho các nhãn tấn công trên CIC-UNSW-NB15

TT	Loại tấn công	Phương Pháp sinh	Accuracy	Precision	Recall	F1
1	Exploits	Thực	0.9888	0.9808	0.9971	0.9889
2	Exploits	SMOTE	0.9877	0.9813	0.9943	0.9878
3	Exploits	CGAN	0.8899	0.9889	0.7887	0.8775
4	Exploits	WGAN	0.9805	0.9707	0.9908	0.9807
5	Reconnissance	Thực	0.9901	0.9821	0.9985	0.9902
6	Reconnissance	SMOTE	0.9879	0.9829	0.9931	0.9880
7	Reconnissance	CGAN	0.9836	0.9693	0.9988	0.9838
8	Reconnissance	WGAN	0.9849	0.9724	0.9982	0.9851
9	Fuzzers	Thực	0.9874	0.9777	0.9976	0.9875
10	Fuzzers	SMOTE	0.9873	0.9778	0.9973	0.9875
11	Fuzzers	CGAN	0.9806	0.9743	0.9872	0.9807
12	Fuzzers	WGAN	0.9849	0.9717	0.9988	0.9851

Bảng 0.3: Hiệu suất phát hiện của mô hình ORIGIN_MODEL và các mô hình GEN_MODEL cho các nhãn tấn công trên tập dữ liệu CIC-IDS-2017

TT	Loại tấn công	Phương Pháp sinh	Accuracy	Precision	Recall	F1
1	DoS GoldenEye	Thực	0.9998	1	0.9995	0.9998
2	DoS GoldenEye	SMOTE	0.9988	1	0.9976	0.9988
3	DoS GoldenEye	CGAN	0.973	0.9594	0.9879	0.9734

4	DoS GoldenEye	WGAN	0.9954	0.9985	0.9922	0.9954
5	FTP-Patator	Thực	0.9997	1	0.9994	0.9997
6	FTP-Patator	SMOTE	0.9997	1	0.9994	0.9997
7	FTP-Patator	CGAN	0.9821	0.9676	0.9975	0.9823
8	FTP-Patator	WGAN	0.9981	1	0.9962	0.9981
9	SSH-Patator	Thực	0.9996	1	0.9992	0.9996
10	SSH-Patator	SMOTE	0.9996	1	0.9992	0.9996
11	SSH-Patator	CGAN	0.9517	0.9216	0.9873	0.9533
12	SSH-Patator	WGAN	0.9771	0.9991	0.955	0.9766
13	DoS slowloris	Thực	0.9991	1	0.9983	0.9991
14	DoS slowloris	SMOTE	0.9978	1	0.9957	0.9978
15	DoS slowloris	CGAN	0.9491	0.9312	0.9698	0.9501
16	DoS slowloris	WGAN	0.9875	0.9956	0.9793	0.9874

3.2.2. Kịch bản 2: Áp dụng WGAN để khắc phục vấn đề mất cân bằng dữ liệu.

Trong kịch bản này, đề án sử dụng WGAN để khắc phục vấn đề mất cân bằng dữ liệu của tập UNSW-NB15. Đề án cho thử nghiệm UNSW-NB15-ORIGIN gồm dữ liệu gốc với số mẫu mỗi nhãn giới hạn ở 13.000, và UNSW-NB15-GEN gồm dữ liệu gốc kết hợp dữ liệu sinh thêm cho các nhãn hiếm,

trong đó tất cả các nhãn đều được cân bằng về 13.000 mẫu để khắc phục mất cân bằng dữ liệu.

Bảng 0.4: Hiệu suất phát hiện cho các nhãn trên bộ dữ liệu UNSW-NB15-GEN

TT	Nhãn dữ liệu	Các độ đo hiệu suất phát hiện		
		Precision	***Recall***	***F1-Score***
1	Benign	0.9990	0.9840	0.9914
2	Generic	0.9959	0.9730	0.9843
3	Exploits	0.6277	0.5523	0.5876
4	Fuzzers	0.8868	0.8773	0.8820
5	DoS	0.4815	0.7143	0.5752
6	Reconnaissance	0.9211	0.8287	0.8724
7	Analysis	0.9506	0.8333	0.8881
8	Backdoor	0.9858	0.8813	0.9307
9	Shellcode	0.9710	0.9813	0.9761
10	Backdoors	0.9574	0.9600	0.9587
11	Worms	0.9947	0.9960	0.9953

3.3. Nhận xét và đánh giá chung

Kết quả Kịch bản 1: SMOTE và WGAN cho dữ liệu sinh giúp mô hình đạt độ đo phát hiện ổn định

và gần tương đương mô hình dùng dữ liệu gốc trên cả ba bộ dữ liệu, trong khi CGAN cho kết quả không ổn định và thấp hơn đáng kể. SMOTE nhỉnh hơn WGAN đôi chút, nhưng WGAN phù hợp hơn với dữ liệu có phân phối phức tạp. Việc đánh giá trên nhiều nhãn tấn công và nhiều tập dữ liệu lớn giúp đảm bảo tính tin cậy của kết luận.

Kết quả Kịch bản 2: Từ kết quả của Kịch bản 1, WGAN được chọn để sinh thêm dữ liệu cho năm nhãn ít mẫu trong UNSW-NB15. Các thử nghiệm cho thấy hiệu suất phát hiện của mô hình IDS đối với các nhãn hiếm được cải thiện đáng kể, qua đó khẳng định sinh dữ liệu bằng WGAN có thể nâng cao hiệu quả phát hiện xâm nhập và giảm chi phí thu thập – xử lý dữ liệu thực.

3.4. Kết chương

III. KẾT LUẬN

1. Kết luận chung

Đề án nghiên cứu các phương pháp sinh dữ liệu tấn công mạng nhằm khắc phục tình trạng mất cân bằng dữ liệu trong hệ thống phát hiện xâm nhập. Nghiên cứu khảo sát và thực nghiệm ba kỹ thuật tiêu biểu gồm SMOTE, CGAN và WGAN; xây dựng mô hình, kịch bản thử nghiệm và đánh giá hiệu quả bằng các độ đo phổ biến. Kết quả cho thấy dữ liệu sinh giúp cải thiện chất lượng huấn luyện và nâng cao khả năng phát hiện tấn công của mô hình phân loại.

2. Hạn chế

Phạm vi thử nghiệm còn giới hạn về số lượng và quy mô tập dữ liệu; số dạng tấn công và phương pháp sinh dữ liệu được khảo sát còn ít. Việc đánh giá chủ yếu dựa trên các độ đo phân loại truyền thống, chưa áp dụng các thước đo thống kê chuyên sâu để phân tích chất lượng dữ liệu sinh.

3. Hướng nghiên cứu tiếp theo

Mở rộng thử nghiệm trên các bộ dữ liệu lớn và đa dạng hơn; nghiên cứu thêm hoặc kết hợp nhiều phương pháp sinh dữ liệu; tích hợp dữ liệu sinh vào hệ thống IDS thực tế; đồng thời xây dựng các độ đo đánh giá chuyên biệt nhằm nâng cao tính chính xác và khả năng ứng dụng thực tiễn.