**CHƯƠNG 1. TỔNG QUAN ĐỀ TÀI**

**CHƯƠNG 2. KIẾN THỨC TỔNG QUAN**

1. Machine Learning IDS

2. Generative Adversarial Network

**CHƯƠNG 2. PHƯƠNG ÁN ĐỀ XUẤT**

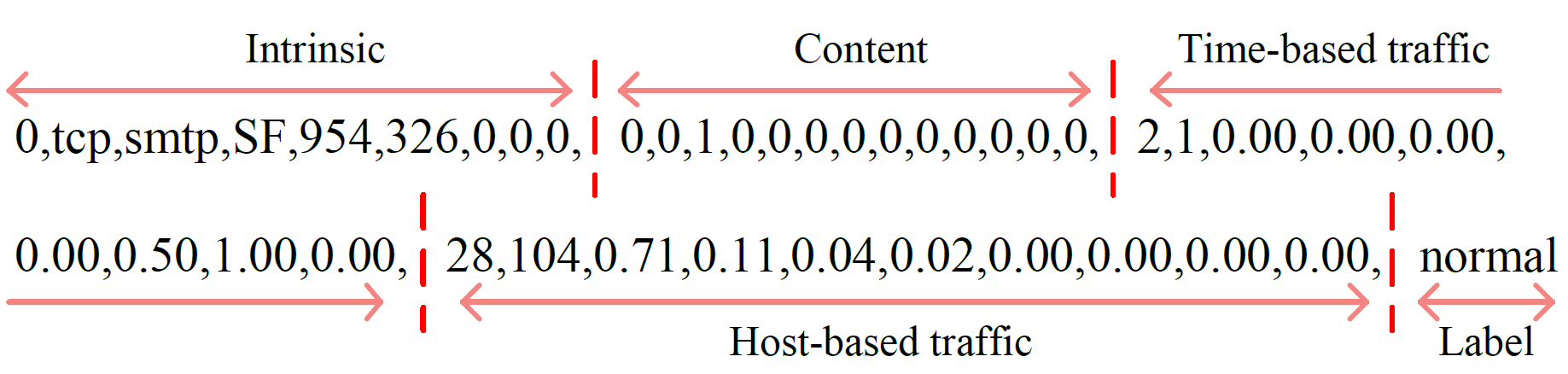
1. Dataset: Mô tả bộ dữ liệu NSL-KDD

Được cải tiến từ KDD’99, NSL-KDD được sử dụng làm bộ dữ liệu điểm chuẩn để đánh giá IDS ngày nay [1]. Trong NSL-KDD, bộ dữ liệu bao gồm tập huấn luyện KDDTrain+ và tập kiểm thử KDDTest+.

Để mô phỏng môi trường mạng thực, dữ liệu lưu lượng chứa lưu lượng truy cập bình thường và bốn loại lưu lượng tấn công chính: Probing (Probe), Từ chối dịch vụ (DoS), User to Root (U2R) và Root to Local (R2L).

Các bản ghi lưu lượng trong NSL-KDD được trích xuất thành các chuỗi kết hợp các đặc tính, như mô tả trừu tượng về lưu lượng truy cập mạng bình thường và tấn công. Có 9 đặc tính rời rạc và 32 đặc tính trị liên tục, tổng cộng có 41 đặc tính. Theo ý nghĩa của các đặc tính,chúng bao gồm bốn bộ bao gồm: “intrinsic”, “content”, “time-based traffic” và “host-based traffic” [2, 3], như được hiển thị trong **Hình 1**. Mô tả chi tiết về một tập hợp các đặc tính tập trung vào được liệt kê bên dưới:

* Đặc tính “intrinsic”: phản ánh các đặc điểm vốn có của một kết nối đơn cho phân tích mạng chung.
* Đặc tính “content”: đánh dấu nội dung của các kết nối cho biết liệu một số hành vi liên quan đến cuộc tấn công có tồn tại trong lưu lượng hay không.
* Đặc tính “time-based traffic”: kiểm tra các kết nối trong 2 giây vừa qua, có cùng máy chủ đích hoặc cùng dịch vụ với kết nối hiện tại, bao gồm các đặc tính “same host” và các đặc tính “same service”.
* Đặc tính “host-based traffic”: giám sát các kết nối trong 100 kết nối vừa qua, có cùng máy chủ đích hoặc cùng dịch vụ với kết nối hiện tại, tương tự “time-based traffic”.



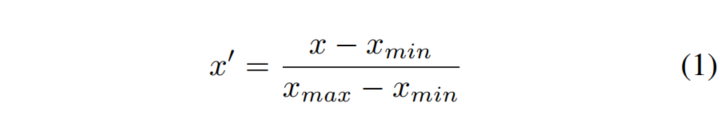
**Hình 1**: Phân phối các bộ đặc tính trong một dòng bản ghi lưu lượng trong NSK-KDD, bao gồm các tính năng và nhãn.

2. Tiền xử lý dữ liệu

Trong quá trình tiền xử lý dữ liệu, dữ liệu trong NSL-KDD cần xử lý chuyển đổi các đặc tính sang dạng số và chuẩn hóa để trở thành các vectơ đầu vào của các mẫu lưu lượng cho IDSGAN

Trong 9 đặc tính rời rạc, có 3 đặc tính không phải là số và 6 đặc tính là các giá trị nhị phân 0 hoặc 1. Để chuyển đổi các chuỗi đặc tính thành các vectơ số đầu vào, trước tiên cần thực hiện chuyển đổi các đặc tính không phải số thành dạng số, bao gồm “protocol type”, “service” và “flag”. Chẳng hạn, “protocol type” có 3 loại thuộc tính: TCP, UDP và ICMP, sẽ được mã hóa thành các giá trị số rời rạc là 1, 2 và 3.

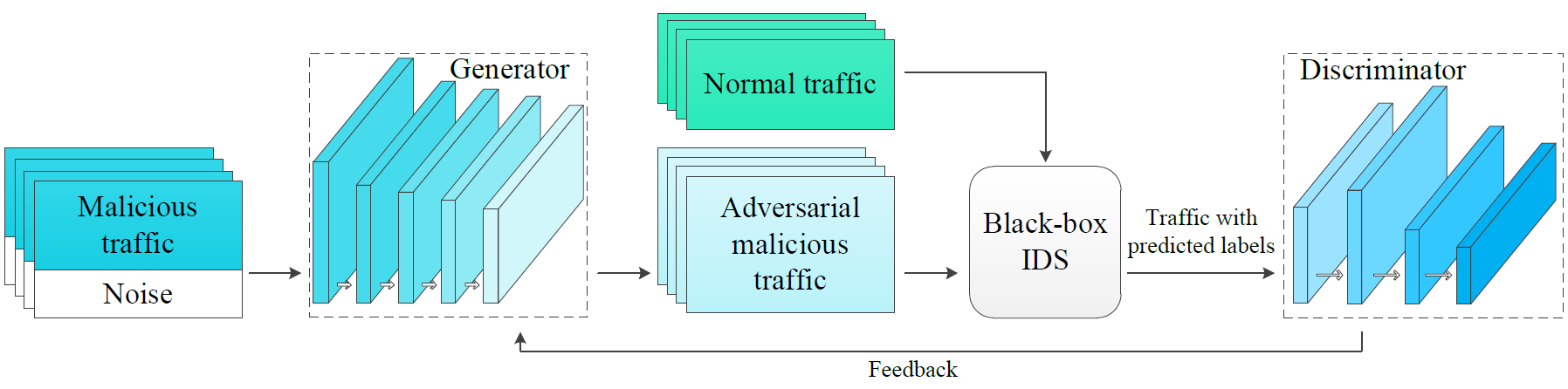
Sau đó, để loại bỏ tác động chiều giữa các giá trị đặc tính trong các vectơ đầu vào, vô hướng chuẩn được sử dụng để chuẩn hóa các đặc tính nguyên bản và đặc tính được chuyển đổi sang số vào một phạm vi cụ thể. Để đào tạo và kiểm tra tốt hơn, chúng tôi thực hiện phương pháp chuẩn hóa Min-Max để chuyển đổi dữ liệu thành khoảng [0; 1], phù hợp cho tất cả các đặc tính rời rạc và liên tục. Chuẩn hóa tối thiểu được tính như sau:



Trong đó *x* là giá trị đặc tính trước khi chuẩn hóa và *x'* là giá trị đặc tính sau khi chuẩn hóa. Ngoài ra, *xmax* và *xmin* tương ứng là giá trị tối đa và giá trị tối thiểu của các đặc tính này trong bộ dữ liệu.

3. Cấu trúc của IDSGAN

Với sự phát triển nhanh chóng của GAN, nhiều phiên bản GAN được thiết kế cho các yêu cầu khác nhau. Để tránh sự không hội tụ và không ổn định của GAN, IDSGAN được thiết kế dựa trên cấu trúc của Wasserstein GAN. Trong IDSGAN, trình tạo (generator) sửa đổi một số đặc tính cụ thể để tạo các mẫu lưu lượng tấn công đối nghịch. Trình phân biệt (disciminator) được đào tạo để bắt chước IDS hộp đen và hỗ trợ đào tạo generator. IDS hộp đen được triển khai bằng các thuật toán học máy để phát hiện các cuộc tấn công. Bằng cách làm cho các trọng số của generator khác với IDS trong đào tạo, các mẫu đối nghịch có thể được tạo để trốn tránh việc phát hiện IDS. Khung của IDSGAN được mô tả trong **Hình 2**.

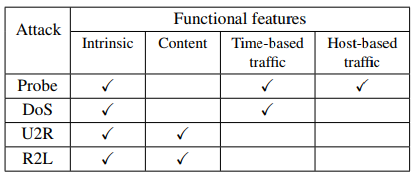


**Hình 2**: Phần huấn luyện của IDSGAN.

3.1. Hạn chế trong việc tạo các mẫu đối nghịch

Mặc dù mục đích chính của việc tạo các mẫu tấn công đối nghịch là để trốn tránh IDS, tiền đề là mẫu đối nghịch được tạo ra nên giữ lại chức năng tấn công của lưu lượng.

Theo các nguyên tắc và mục đích tấn công, hiển nhiên là mỗi loại tấn công đều có các đặc tính chức năng (functional features), đại diện cho chức năng cơ bản của cuộc tấn công này. Nói cách khác, thuộc tính tấn công vẫn không bị xáo trộn nếu chúng ta chỉ thay đổi các đặc tính không chức năng (nonfunctional features), không phải các đặc tính chức năng. Do đó, để tránh làm mất hiệu lực lưu lượng phải giữ nguyên functional features của mỗi cuộc tấn công. Đối với nonfunctional features của một cuộc tấn công, không thể hiện chức năng liên quan đến cuộc tấn công đó, chúng ta có thể thay đổi hoặc giữ lại chúng. Các đặc tính được giữ lại trong thế hệ này, bao gồm functional features, hay được gọi là các đặc tính chưa sửa đổi (unmodified features) trong bài viết này. Như thể hiện trong [3], functional features của từng loại tấn công trong NSL-KDD được trình bày trong **Bảng 1**.



**Bảng 1**: Các đặc tính chức năng (functional features) của mỗi thể loại tấn công.

3.2. Trình tạo (Generator)

Là một phần quan trọng của mô hình, trình tạo (generator) đóng vai trò tạo ra mẫu lưu lượng tấn công đối với cuộc tấn công trốn tránh IDS.

Để chuyển đổi một mẫu gốc thành một mẫu đối nghịch, mỗi vectơ đầu vào của các mẫu lưu lượng phải bao gồm một vectơ mẫu gốc M m-chiều chiều và vectơ nhiễu N n-chiều. Là phần mẫu gốc ban đầu, M đã được xử lý trước. Để phù hợp với vectơ được xử lý trước, các phần tử của N phần nhiễu bao gồm các số ngẫu nhiên trong một phân bố đồng đều trong phạm vi [0,1].

Cấu trúc đề xuất của generator có một mạng thần kinh (neural network) với 5 lớp tuyến tính. Phi tuyến tính ReLU F = max (0; x) được sử dụng để kích hoạt đầu ra của 4 lớp tuyến tính trước. Để làm cho các mẫu đối nghịch đáp ứng công thức của vectơ mẫu gốc M, lớp đầu ra phải có m đơn vị. Việc cập nhật các tham số trong mạng này dựa trên phản hồi từ disciminator.

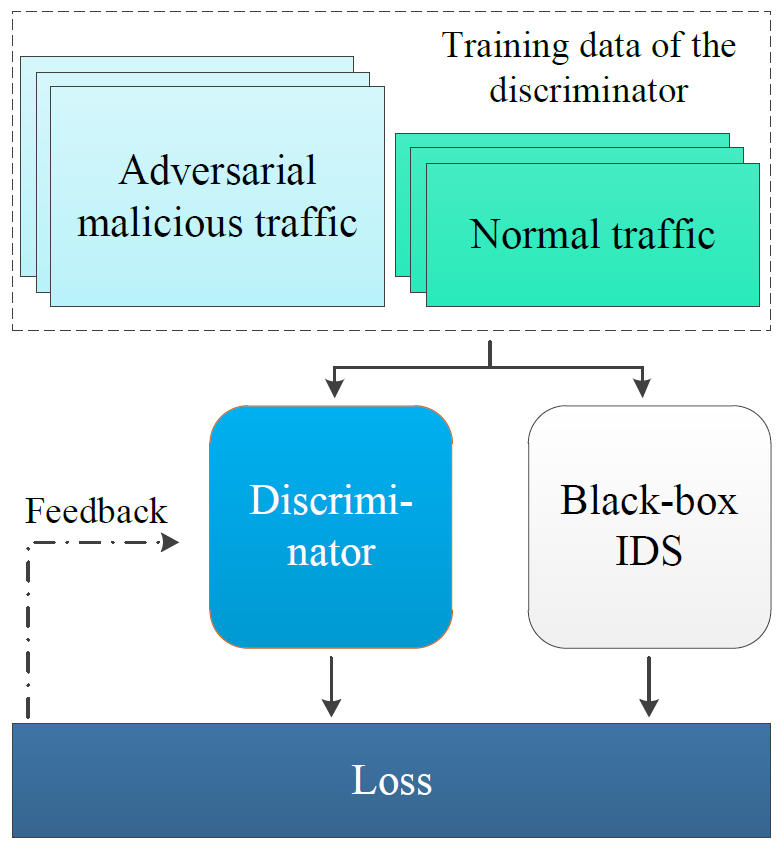
Ngoài ra, có một số thủ thuật trong việc xử lý các đặc tính được sửa đổi. Để giới hạn các phần tử đầu ra trong phạm vi [0,1], phần tử trên 1 được đặt là 1 và phần tử dưới 0 được đặt là 0. Ngoài ra, xem xét các đặc tính "intrinsic" là các đặc tính chức năng trong tất cả các cuộc tấn công trong NSL-KDD, không sửa đổi các đặc tính riêng biệt có nhiều hơn hai giá trị, tất cả thu thập trong đặc tính "intrinsic". Vì vậy, các đặc tính riêng biệt được sửa đổi là các đặc tính rời rạc nhị phân. Các giá trị của các đặc tính riêng biệt được sửa đổi này sẽ được chuyển thành giá trị nhị phân. Ngưỡng trong chuyển đổi nhị phân là 0,5. Các giá trị trên hoặc dưới ngưỡng sẽ được chuyển thành 1 hoặc 0.

3.3. Trình phân biệt (Disciminator)

Không có kiến thức về cấu trúc của mô hình IDS hộp đen, bộ phân biệt được sử dụng để bắt chước IDS hộp đen dựa trên các mẫu lưu lượng tấn công đối nghịch và các mẫu lưu lượng truy cập thông thường. Việc bắt chước giúp đào tạo generator vì các cuộc tấn công đối nghịch có thể cố gắng bỏ qua IDS trong quá trình đào tạo generator. Trong khi đó, là một công cụ phân loại, bộ phân biệt đối xử phân loại đầu ra của trình tạo và cung cấp phản hồi cho trình tạo.

Trình phân biệt (discriminator) là một mạng lưới thần kinh nhiều lớp để phân loại lưu lượng tấn công và lưu lượng bình thường. Dữ liệu huấn luyện của nó bao gồm lưu lượng tấn công đối nghịch và lưu lượng bình thường.

Là một trong những mục đích chính của discriminator, công việc bắt chước khả năng của IDS hộp đen cần kết quả phát hiện của IDS hộp đen đối với các mẫu lưu lượng tấn công bình thường và đối nghịch. Đầu tiên, các mẫu lưu lượng thường và các mẫu lưu lượng tấn công đối nghịch được phân loại theo IDS hộp đen. Sau đó, kết quả phát hiện của IDS hộp đen được sử dụng làm nhãn mục tiêu của dữ liệu huấn luyện của discriminator để làm cho phân loại của discriminator, tương tự như IDS hộp đen. Quy trình của discriminator bắt chước mô hình IDS hộp đen được mô tả trong **Hình 3**.



**Hình** **3**: Quá trình mô phỏng IDS hộp đen của trình phân biệt (discriminator).

3.4. Thuật toán huấn luyện

Trong quá trình đào tạo generator, các kết quả thu được từ phân loại của discriminator cho các mẫu đối nghịch cung cấp thông tin độ dốc cho quá trình huấn luyện của generator. Hàm mất mát của generator được định nghĩa như sau:



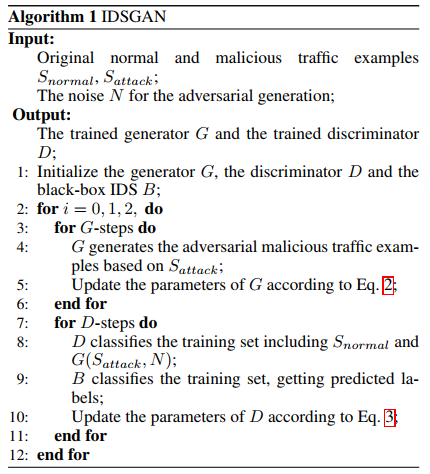
trong đó Sattack là mẫu lưu lượng tấn công ban đầu; G đại diện cho generator và D đại diện cho discriminator . Để đào tạo và tối ưu hóa generator để đánh lừa IDS hộp đen, chúng ta cần giảm thiểu LG.

Đối với đào tạo discriminator, các mẫu lưu lượng tấn công đối nghịch là một phần của tập huấn luyện. Theo phần giới thiệu ở trên về mối quan hệ giữa tập huấn luyện và nhãn dự đoán của IDS hộp đen, mất mát của discriminator được tính bằng nhãn đầu ra của discriminator và nhãn dự đoán từ IDS hộp đen. Do đó, định nghĩa về hàm mất mát của discriminator cho việc tối ưu như sau:



trong đó s là các mẫu lưu lượng cho việc đào tạo discriminator; Bnormal và Battack tương ứng là các mẫu bình thường và các mẫu tấn công đối nghịch được dự đoán bởi IDS hộp đen.

Theo Wasserstein GAN, RMSProp là trình tối ưu hóa IDSGAN để tối ưu hóa các tham số trong mạng. Thuật toán 1 cho thấy một phác thảo về đào tạo IDSGAN chung.



**CHƯƠNG 3. THỰC NGHIỆM VÀ KẾT QUẢ**

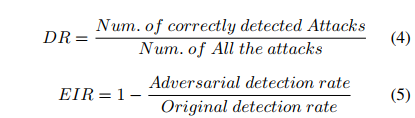
Trong thử nghiệm, PyTorch được thông qua làm deep learning framework để triển khai IDSGAN [4]. Mô hình đề xuất được chạy và đánh giá trên PC Linux với Intel Core i7-2600.

IDSGAN được đào tạo với batch size 64 cho 100 epochs. Learning rate của discriminator và generator là 0,0001. Kích thước của vectơ nhiễu (noise vector) là 9. Ngưỡng weight clipping cho đào tạo discriminator được đặt là 0,01.

Để đánh giá năng lực của mô hình một cách toàn diện và sâu sắc, các loại thuật toán học máy khác nhau được sử dụng làm IDS hộp đen trong các thực nghiệm. Dựa trên các nghiên cứu có liên quan trong phát hiện xâm nhập, các thuật toán được sử dụng của IDS hộp đen trong các thí nghiệm bao gồm: Support Vector Machine (SVM), Naive Bayes (NB), Multilayer Perceptron (MLP), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and K-Nearest Neighbor (KNN). Các mô hình IDS hộp đen đã được đào tạo với tập huấn luyện của chúng trước khi đánh giá IDSGAN.

Tập huấn luyện và tập kiểm thử được thiết kế dựa trên bộ dữ liệu NSL-KDD bao gồm KDDTrain+ và KDDTest+. Tập huấn luyện cho IDS hộp đen gồm một nửa số bản ghi trong KDDTrain+, bao gồm các bản ghi lưu lượng bình thường và tấn công. Tập huấn luyện của discriminator bao gồm các bản ghi lưu lượng thường trong nửa còn lại của KDDTrain+ và các mẫu về lưu lượng tấn công đối nghịch từ generator. Vì GAN được áp dụng trong phân loại nhị phân, IDSGAN có thể tạo các mẫu đối nghịch cho chỉ một loại tấn công mỗi lần. Vì vậy, trong mọi thực nghiệm, tập huấn luyện của generator là bản ghi của một loại tấn công trong nửa còn lại của KDDTrain+. Các bản ghi của một tấn công trong KDDTest+ tạo nên tập kiểm thử của một loại tấn công đối nghịch cho generator.

Đối với các số liệu thực nghiệm, tỷ lệ phát hiện (Detection Rate - DR) và tỷ lệ tăng trốn tránh (Evasion Increase Rate - EIR) được đo lường, cho thấy hiệu suất của IDSGAN trực tiếp và tương đối. Tỷ lệ phát hiện (DR) phản ánh tỷ lệ giữa các số lượng lưu lượng tấn công được phát hiện chính xác bởi IDS hộp đen và số lượng tất cả các lưu lượng tấn công, trực tiếp cho thấy khả năng trốn tránh của mô hình và độ mạnh mẽ của IDS hộp đen. Tỷ lệ phát hiện ban đầu và tỷ lệ phát hiện đối nghịch tương ứng với tỷ lệ phát hiện đối với các lưu lượng tấn công ban đầu và tỷ lệ phát hiện lưu lượng tấn công đối kháng tương ứng. Ngoài ra, tỷ lệ tăng trốn tránh (EIR) là tỷ lệ gia tăng của các mẫu lưu lượng truy cập tấn công không bị phát hiện bởi IDS so với các mẫu lưu lượng tấn công ban đầu, phản ánh tương đối khả năng của IDSGAN, đặc biệt là trong các nền tảng khác nhau. Các số liệu này được tính như sau:



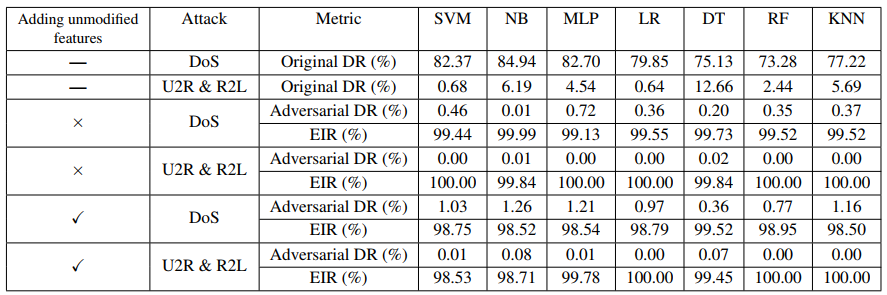
Tỷ lệ phát hiện thấp hơn có nghĩa là có nhiều những lưu lượng tấn công hơn có thể trốn tránh IDS hộp đen, phản ánh trực tiếp sự mạnh mẽ hơn của IDSGAN. Ngược lại, tỷ lệ trốn tránh thấp hơn phản ánh nhiều những mẫu đối nghịch hơn bị phát hiện bởi IDS hộp đen, nghĩa là có sự suy giảm về khoảng cách tương đối khả năng tấn công trốn tránh giữa lưu lượng tấn công ban đầu và lưu lượng tấn công đối nghịch. Vì vậy, động lực cho IDSGAN là đạt được tỷ lệ phát hiện (DR) thấp hơn và tỷ lệ tăng trốn tránh (EIR) cao hơn.

1. Hiệu suất của IDSGAN trong các cuộc tấn công khác nhau

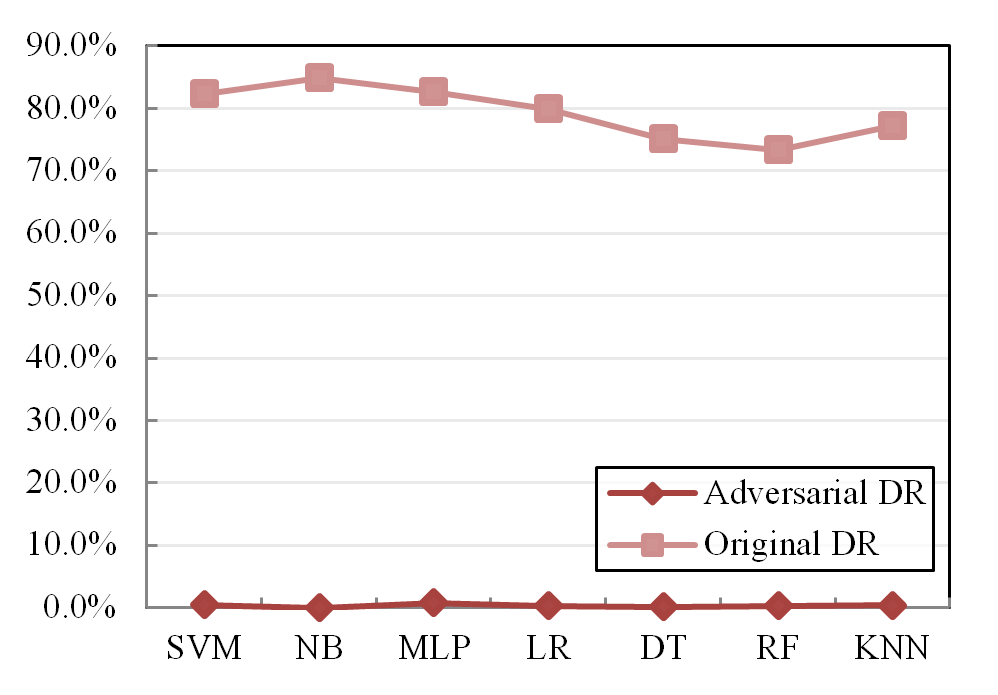
Để đánh giá mô hình một cách toàn diện, IDSGAN được đào tạo và sau đó, tạo ra lưu lượng tấn công đối nghịch trong các thử nghiệm dựa trên KDDTest+. Xem xét rằng DoS và Probe đều là các cuộc tấn công dựa trên mạng, chúng tôi chỉ thử nghiệm và phân tích trên DoS để phản ánh hiệu suất của IDSGAN đối với các loại tấn công này. Ngoài ra, các cuộc tấn công dựa trên nội dung lưu lượng truy cập như U2R và R2L cũng được thử nghiệm. Do các dữ liệu U2R và R2L trong NSL-KDD có số lượng nhỏ và có đặc tính tương tự, U2R và R2L được tập hợp thành một nhóm tấn công trong các thực nghiệm.

Trước khi áp dụng IDSGAN, tỷ lệ phát hiện ban đầu của DoS, U2R và R2L được đo trên IDS hộp đen đã đào tạo, được hiển thị trong **Bảng 2**. Do số lượng nhỏ hồ sơ của U2R và R2L trong tập huấn luyện, việc học không đủ làm cho tỷ lệ phát hiện gốc của U2R và R2L thấp.

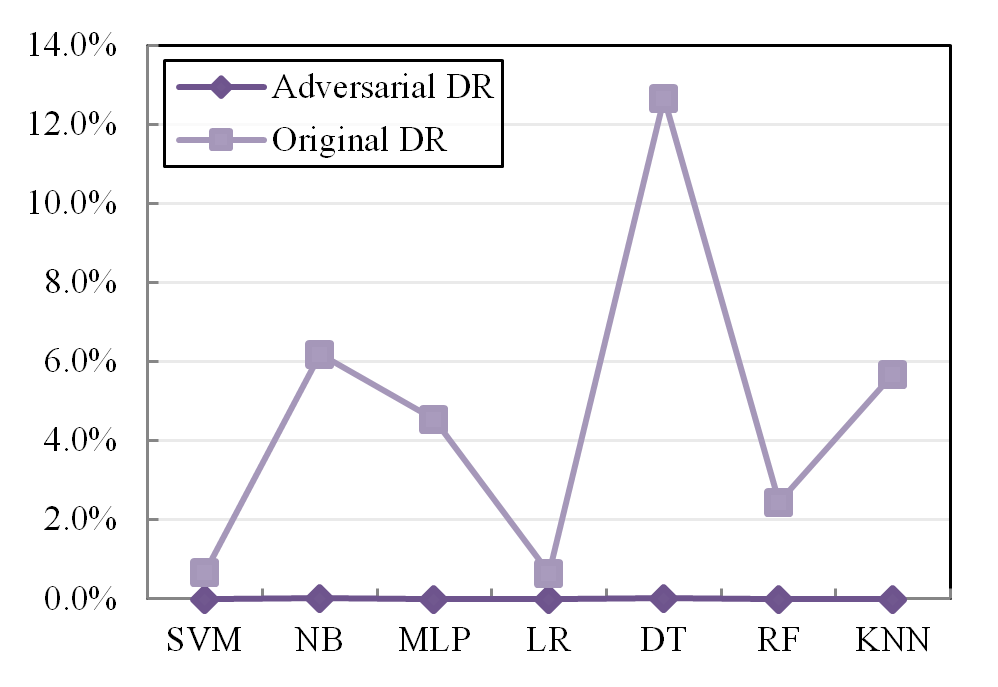
Đầu tiên, chúng tôi kiểm tra khả năng của IDSGAN trong các cuộc tấn công khác nhau chỉ với các đặc tính chức năng chưa được sửa đổi (functional features unmodified). Kết quả của các thí nghiệm được trình bày trong **Bảng 2** và **Hình 4**. Theo kết quả, tất cả các tỷ lệ phát hiện đối nghịch của DoS, U2R và R2L trong các mô hình IDS hộp đen khác nhau đều giảm và gần tới 0 sau xử lý của IDSGAN, có nghĩa là các mô hình IDS gần như không thể phân loại bất kỳ ví dụ lưu lượng truy cập độc hại nào. Khoảng cách giữa tỷ lệ phát hiện ban đầu và tỷ lệ phát hiện đối nghịch của các cuộc tấn công này cũng rất lớn và rõ ràng.



**Bảng 2**: Hiệu suất của IDSGAN đối với DoS, U2R và R2L



**(a)**



**(b)**

**Hình 4**: So sánh tỷ lệ phát hiện đối nghịch và tỷ lệ phát hiện ban đầu trong các mô hình IDS hộp đen khác nhau chỉ có các tính năng chức năng chưa được sửa đổi. (a) là kết quả của DoS và (b) là kết quả của U2R và R2L.

Như được hiển thị trong **Hình 4 (a)**, tỷ lệ phát hiện đối nghịch nghịch của DoS trong tất cả các thuật toán phát hiện giảm đáng kể từ khoảng 80% xuống dưới 1%. Mặc dù MLP thể hiện sự mạnh mẽ nhất trong danh sách của tất cả các mô hình IDS hộp đen, tỷ lệ phát hiện đối nghịch của DoS chỉ là 0,72%. Tỷ lệ tăng trốn tránh đều trên 98,5%. Kết quả cho thấy hiệu suất tuyệt vời của IDSGAN trong DoS. Hơn 99% các ví dụ lưu lượng truy cập DoS đối nghịch có thể tránh được sự phát hiện của mô hình IDS thực nghiệm trong mỗi thử nghiệm.

Đối với U2R và R2L được hiển thị trong Hình 4 (b), mặc dù sự khác biệt về tỷ lệ phát hiện ban đầu giữa các thuật toán là đáng chú ý, tất cả các tỷ lệ phát hiện đối nghịch đều bằng hoặc gần bằng 0 sau sự tạo mẫu đối nghịch, nghĩa là hầu hết các mẫu lưu lượng tấn công ban đầu của U2R và R2L mà IDS phát hiện được có thể đánh lừa và trốn tránh IDS sau xử lý của IDSGAN. Số lượng lớn các mẫu không bị phát hiện được tăng lên sau khi qua generator dẫn đến tỷ lệ tăng trốn tránh cao, tất cả đều trên 99,5%.

Tỷ lệ phát hiện đối nghịch thấp và tỷ lệ tăng trốn tránh cao trong thử nghiệm dưới các cuộc tấn công khác nhau phản ánh rằng IDSGAN cho thấy khả năng tuyệt vời của nó đối với cuộc tấn công đối nghịch trong các thực nghiệm. Đối mặt với các cuộc tấn công đối nghịch, có một số khác biệt nhỏ về khả năng của IDSGAN và sự mạnh mẽ của các mô hình IDS hộp đen trong các loại tấn công khác nhau và các mô hình IDS khác nhau.

2. Hiệu suất của IDSGAN đối với các loại GAN khác nhau

Trong nghiên cứu về khả năng và sự mạnh mẽ của IDSGAN, số lượng các đặc tính chưa sửa đổi(unmodified features) là một yếu tố quan trọng quyết định sự thành công của các cuộc tấn công đối kháng bằng mô hình đề xuất của chúng tôi. Để đánh giá mối quan hệ giữa IDSGAN và số lượng unmodified features, các thực nghiệm ngược lại được thực hiện trên DoS, U2R và R2L, thay đổi số lượng unmodified features. Với việc xem xét rằng các đặc tính chức năng của mỗi cuộc tấn công là các đặc tính ít nhất đại diện cho chức năng tấn công, cách để thay đổi số lượng đặc tính không được sửa đổi là thêm các đặc tính không hoạt động (nonfunctional features) trên cơ sở các đặc tính chức năng. Để kiểm tra ảnh hưởng không thể thiếu của việc thêm các đặc tính tiềm năng trên IDSGAN, unmodified features được thêm này được chọn ngẫu nhiên từ mỗi bộ đặc tính khác. Các tỷ lệ của các đặc tính được thêm vào trong phần còn lại của các bộ là như nhau. Các unmodified features được thêm vào trong các thực nghiệm được liệt kê dưới đây.

* DoS:

“Content”:

hot, num\_failed\_logins, logged\_in, num\_compromised, num\_root, num\_file\_creations, is\_guest\_login;

“Host-based traffic”:

dst\_host\_count, dst\_host\_rerror\_rate, dst\_host\_serror\_rate, dst\_host\_same\_srv\_rate, dst\_host\_same\_src\_port\_rate;

* U2R&R2L:

“Time-based traffic”:

count, srv\_count, serror\_rate, srv\_serror\_rate;

“Host-based traffic”:

dst\_host\_srv\_diff\_host\_rate, dst\_host\_srv\_serror\_rate,

dst\_host\_srv\_count, dst\_host\_diff\_srv\_rate, dst\_host\_srv\_rerror\_rate.

Kết quả của các thực nghiệm được trình bày trong **Bảng 2 và Hình 5**. Khả năng và sự mạnh mẽ của IDSGAN theo số lượng khác nhau các đặc tính chưa được sửa đổi (unmodified features) được trình bày rõ ràng và có thể so sánh bằng tốc độ tăng trốn tránh trong **Hình 5**. So với các thực nghiệm chỉ có unmodified features, tốc độ tăng trốn tránh trong các thực nghiệm tương phản giảm nhẹ hoặc duy trì.

Những thay đổi nhỏ của tất cả các tỷ lệ tăng trốn tránh trong **Hình 5** phản ánh sự mạnh mẽ mạnh mẽ của IDSGAN với số lượng lớn hơn unmodified features. Trong khi đó, với sự gia tăng của unmodified features, nhiều thông tin gốc trong một bản ghi lưu lượng được giữ lại sau khi tạo lưu lượng đối nghịch. Với sự giúp đỡ của kiến thức thêm này, IDS hộp đen có thể phát hiện tốt hơn dựa trên thông tin rõ ràng hơn, đưa ra những đánh giá chính xác hơn trong các bài kiểm tra. Nhưng do sự nhiễu loạn mạnh mẽ từ các tính năng đã được sửa đổi, việc bổ sung unmodified features sẽ không giúp cho việc phát hiện của một số mô hình IDS như Logistic Regression khi phát hiện U2R và R2L. Do đó, mức độ tăng độ chính xác phát hiện của IDS hộp đen là khác nhau trong các mô hình tấn công và IDS khác nhau.

Mức độ giảm của tỷ lệ tăng trốn tránh có liên quan đến các loại tấn công. Ví dụ, trong Hình 5(a), việc giảm tỷ lệ tăng trốn tránh của DoS trong K-Nearest Neighbor là khoảng 1,00% trước và sau khi thêm các tính năng chưa được sửa đổi. Tuy nhiên, tỷ lệ tăng trốn tránh của U2R và R2L theo thuật toán phát hiện này không thay đổi sau khi thêm vào được thể hiện ở Hình 5 (b).

Hơn nữa, sự sụt giảm của tỷ lệ tăng trốn tránh trong cùng loại tấn công giữa các mô hình IDS khác nhau cũng khác nhau . Như được thể hiện trong Hình 5 (a), những thay đổi về tỷ lệ tăng trốn tránh trong Naive Bayes trong DoS lớn hơn nhiều so với những cái khác, có nghĩa là việc bổ sung unmodified features có tác động lớn hơn đến hiệu suất của tấn công DoS đối nghịch dưới sự phát hiện của Naive Bayes.

Do đó, theo kết quả của các thực nghiệm, tỷ lệ tăng trốn tránh vẫn cao chứng tỏ sự mạnh mẽ của IDSGAN với hầu hết lưu lượng tấn công né tránh sự phát hiện của IDS hộp đen, mặc dù việc thêm unmodified features ảnh hưởng một chút tới việc đánh lừa IDS của các mẫu tấn công đối nghịch. Nhưng mức độ ảnh hưởng đến hiệu suất của các mẫu tấn công đối nghịch là khác nhau tùy theo các loại tấn công và các mô hình IDS hộp đen khác nhau.

**CHƯƠNG 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

Với mục đích tạo ra các cuộc tấn công đối kháng để trốn tránh hệ thống phát hiện xâm nhập, IDSGAN là một khuôn khổ mới của các mạng đối kháng tạo sinh (GAN) dựa trên Wasserstein GAN, bao gồm trình tạo (Generator), trình phân biệt (Discriminator), IDS hộp đen. IDSGAN cho thấy khả năng tốt của mình trong việc tạo ra các mẫu lưu lượng tấn công của các tấn công khác nhau, khiến tỷ lệ phát hiện của các mô hình IDS hộp đen khác nhau giảm xuống xấp xỉ 0. Hơn nữa, để đánh giá mức độ ảnh hưởng của việc thay đổi số lượng các đặc tính chưa được đổi, nó cho thấy rằng khả năng trốn tránh của các mẫu lưu lượng tấn công đối nghịch sẽ duy trì hoặc giảm nhẹ khi thêm các đặc tính không đổi (unmodified features), phản ánh sự mạnh mẽ mạnh mẽ của IDSGAN. Mức độ thay đổi hiệu suất của IDSGAN sau khi bổ sung các đặc tính không đổi dựa trên loại tấn công và loại thuật toán phát hiện của IDS hộp đen. Hiệu năng tuyệt vời được trình bày ở trên bởi mô hình cho thấy tính khả thi và tính linh hoạt rộng rãi của IDSGAN.

Trong tương lai, chúng tôi sẽ tập trung hơn vào việc cải thiện IDSGAN. Cải tiến này sẽ tập trung vào hai khía cạnh: thứ nhất, chúng tôi sẽ áp dụng IDSGAN trong nhiều loại tấn công xâm nhập hơn; Thứ hai, với mục đích dứt khoát là phát triển IDS, nâng cao sức mạnh của IDS là công việc quan trọng của chúng tôi.

**TÀI LIỆU THAM KHẢO**