PHÁT HIỆN XÂM NHẬP MẠNG SỬ DỤNG MẠNG NƠ-RON ĐỒ THỊ VỚI ĐẶC TRƯNG NÚT VÀ CẠNH

Nguyễn Thị Thanh Thủy, Nguyễn Ngọc Điệp
Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Bài báo này trình bày một phương pháp mới sử dụng mạng nơ-ron đồ thị với cơ chế tập trung và nhúng kết hợp cả nút và cạnh để phát hiện xâm nhập mạng. Nhờ khả năng tận dụng thông tin toàn diện về hoạt động mạng bằng cách kết hợp nhúng nút và cạnh trong đồ thị, mô hình đề xuất có thể học được các mối quan hệ phức tạp giữa các luồng dữ liệu mạng. Kết hợp với cơ chế học tập trung mạnh mẽ của mô hình mạng nơ-ron đồ thị tập trung (GAT), mô hình đồ thị đề xuất có thể học được nhiều thông tin hữu ích trong đồ thị biểu diễn mạng với các luồng dữ liệu giàu thông tin. Mỗi nút trong đồ thị đại diện cho một luồng dữ liệu mạng, được khởi tạo bằng các đặc trưng của luồng dữ liệu. Các cạnh của đồ thị được xây dựng dựa trên mối quan hệ giữa các luồng dữ liệu có chung địa chỉ IP nguồn. Kết quả thực nghiệm trên tập dữ liệu tiêu chuẩn CIC IDS 2017 cho thấy, mô hình đề xuất có hiệu năng vượt trội hơn các phương pháp học máy truyền thống và học sâu, cũng như mô hình mạng nơ-ron đồ thị đơn giản.

Từ khóa: phát hiện xâm nhập, mạng nơ-ron đồ thị, đặc trưng nút và cạnh, học sâu

I. GIỚI THIỆU

Trong bối cảnh xã hội hiện đại, mạng máy tính giữ vai trò thiết yếu trong đời sống và sản xuất, đồng thời cũng trở thành mục tiêu của nhiều mối đe nhằm thực hiện các hoạt động tấn công độc hại. Phát hiện xâm nhập mạng là một yếu tố vô cùng quan trọng trong lĩnh vực an ninh mạng. Hoạt động này đóng vai trò là hàng rào bảo vệ, giúp giám sát và kịp thời phát hiện các hoạt động xâm nhập trái phép vào hệ thống mạng. Các hệ thống phát hiện xâm nhập mạng (IDS) hiện đang được triển khai rộng rãi trong các mạng của doanh nghiệp và tổ chức.

Theo báo cáo của công ty IBM, chi phí trung bình cho một vụ vi phạm dữ liệu vào năm 2024 đã đạt 4,88 triệu đô la Mỹ, là mức trung bình cao nhất từng được ghi nhận [1]. Chi phí liên quan đến tội phạm mạng dự kiến sẽ đạt 10,5 nghìn tỷ đô la Mỹ vào năm 2025 [2]. Tại Việt Nam, trong 8 tháng đầu năm 2024, các cơ quan cũng đã ghi nhận hơn 4.000 sự cố tấn công mạng [3]. Những con số này nhấn mạnh sự cần thiết cần phải phát triển các hệ thống phát hiện xâm nhập mạng hiệu quả.

Ngày nay, các cuộc tấn công mạng ngày càng trở nên tinh vi, tạo ra những thách thức lớn cho các hệ thống phát hiện xâm nhập mạng. Những kẻ tấn công không ngừng tìm kiếm các lỗ hổng để khai thác, với sự đa dạng và phức tạp của các kỹ thuật tấn công, làm cho việc xác định và phân biệt các mối đe dọa tiềm ẩn càng trở nên khó khăn hơn. Để giải quyết các vấn đề phát hiện xâm nhập mạng, có nhiều phương pháp đã được thực hiện có thể kể đến như phương pháp phát hiện dựa trên chữ ký, phương pháp phát hiện dựa trên học máy cơ bản và phương pháp phát hiện dựa trên học sâu. Các phương pháp phát hiện dựa trên chữ ký [4] là các phương pháp truyền thống còn gặp nhiều hạn chế. Kẻ tấn công có thể dễ dàng vượt qua các phương pháp này bằng cách tạo ra biến thể của các cuộc tấn công đã biết hoặc nhắm mục tiêu vào lỗ hổng hoàn toàn mới. Phương pháp phát hiện dựa trên học máy cơ bản thường đạt được độ chính xác cao khi được huấn luyện và đánh giá với lưu lượng truy cập của cùng một mạng, nhưng các phương pháp này đặc biệt kém hiệu quả với các cuộc tấn công đối kháng, hoặc có những thay đổi các đặc trưng luồng theo thời gian để tránh bị phát hiện [5]. Phương pháp phát hiện dựa trên học sâu thông thường như CNN hay LSTM có khả năng tự học đặc trưng và giảm thiểu vấn đề quá khớp dữ liệu trong các mô hình học máy truyền thống, do đó đã đạt được nhiều kết quả khả quan trong lĩnh vực an ninh mạng [6,7]. Tuy nhiên, các phương pháp này vẫn có một số thách thức khi thực hiện với bài toán phát hiện xâm nhập mạng do dữ liệu mạng thường ở dạng không có cấu trúc và có nhiều tính chất phức tạp về mô hình mạng, các thực thể mạng và các mối quan hệ đan xen trong hệ thống mạng. Khi không thể nắm bắt đầy đủ các thông tin này sẽ dẫn đến giảm hiệu suất của các phương pháp phát hiện xâm nhập.

Gần đây, mạng nơ-ron đồ thị (Graph Neural Network-GNN) đã trở thành một phương pháp tiềm năng trong lĩnh vực phát hiện xâm nhập mạng [8]. GNN là một nhóm mạng nơ-ron mới, được thiết kế đặc biệt để xử lý dữ liệu dưới dạng đồ thị. Ưu điểm vượt trội của GNN so với các phương pháp trước đây là khả năng học hỏi và khái quát hóa từ dữ liệu có cấu trúc đồ thị, đồng thời nắm bắt và mô hình hóa các mẫu ẩn trong đồ thị. Điều này mang lại sức mạnh dự đoán cao hơn trong nhiều bài toán có dữ liệu được tổ chức dưới dạng đồ thị.

Tuy nhiên, hệ thống phát hiện xâm nhập dựa trên mạng nơ-ron đồ thị vẫn chưa có được mức độ tin cậy và ổn định mong muốn, vì dữ liệu đầu vào của mô hình chưa được tối ưu hóa trước khi huấn luyện. Các nghiên cứu trước đây chủ yếu tiếp cận dựa trên khai thác thông tin chỉ trên nút hoặc của cạnh. Trong khi đó, các kịch bản truyền thông trên mạng máy tính, dữ liệu mạng thường nằm trên cả 2 thành phần là nút và cạnh. Cả nút và cạnh đều là những yếu tố quan trọng của đồ thị cần được khai

Tác giả liên hệ: Nguyễn Ngọc Điệp
Email: diepnguyenngoc@ptit.edu.vn
Đến tòa soạn: 11/2024, chỉnh sửa: 12/2024, chấp nhận đăng: 12/2024.

thác đồng thời để mô hình học được thông tin có liên quan về ngữ cảnh. Ví dụ, một luồng dữ liệu có dung lượng lớn bất thường hoặc được gửi đến một địa chỉ IP đáng ngờ có thể là dấu hiệu của một cuộc tấn công DDoS. GNN có thể tận dụng các thông tin này để có được cái nhìn toàn diện hơn về hoạt động mạng, giúp phân biệt hiệu quả giữa hoạt động bình thường và tấn công mạng. Ngoài ra, khi xem xét về các hoạt động xâm nhập mạng trên đồ thị, một số nút hoặc cạnh có thể đóng vai trò quan trọng hơn trong việc phát tán mã độc hoặc thực hiện tấn công. Nếu ứng dụng cơ chế học tập trung vào các nút và cạnh này trên đồ thì có thể cải thiện được độ chính xác của việc phát hiện xâm nhập.

Trong nghiên cứu này, chúng tôi đề xuất mô hình mạng nơ-ron đồ thị để xử lý dữ liệu đồ thị từ luồng mạng. Dữ liệu đồ thị được xây dựng dựa trên tập các đặc trưng quan trọng nhất của luồng dữ liệu được biểu diễn dưới dạng nút. Đồng thời, các cạnh của đồ thị được hình thành bằng cách sử dụng mối tương quan giữa các luồng chia sẻ cùng một địa chỉ IP nguồn. Cách tiếp cận được này gợi ý từ nghiên cứu [5], giúp tạo ra dữ liệu đồ thị đã chứa thông tin về các mối quan hệ bên trong chúng. Phương pháp này bảo toàn lượng thông tin mạng tối đa vì nó coi toàn bộ tập dữ liệu là một tổng thể chứ không xử lý từng điểm dữ liệu độc lập. Ngoài ra, để nâng cao hiệu quả của bài toán phát hiện xâm nhập mạng, chúng tôi đề xuất sử dụng mô hình mạng nơ-ron đồ thị nhúng kết hợp cả nút và cạnh, cùng với cơ chế tập trung (Graph Attention Networks - GAT). Các đóng góp chính của bài báo như sau:

1) Giới thiệu một phương pháp mới sử dụng đồ thị mạng nơ-ron với cơ chế tập trung và nhúng kết hợp cả nút và cạnh để phát hiện xâm nhập mạng. Phương pháp này sử dụng được thông tin toàn diện về hoạt động mạng thông qua kết hợp nhúng nút và cạnh trong đồ thị. Đồng thời, kết hợp với cơ chế học tập trung mạnh mẽ của mô hình GAT, mô hình đồ thị này có thể học được nhiều thông tin hữu ích trong đồ thị biểu diễn mạng máy tính với các luồng dữ liệu giàu thông tin.

2) Đánh giá hiệu năng trên bộ dữ liệu tiêu chuẩn CIC IDS 2017 [9]. Đây là một bộ dữ liệu phổ biến được sử dụng để đánh giá các hệ thống phát hiện xâm nhập mạng. Việc sử dụng bộ dữ liệu này cho phép so sánh trực tiếp với các phương pháp hiện có và chứng minh hiệu quả của mô hình đề xuất trong việc phát hiện tấn công.

Phần còn lại của bài báo được tổ chức như sau. Phần II mô tả các nghiên cứu liên quan. Phần III trình bày cơ sở lý thuyết về GNN. Phần IV đề xuất phương pháp phát hiện phát hiện xâm nhập mạng sử dụng mạng nơ-ron đồ thị tập trung với nhúng kết hợp. Kết quả và những phân tích thực nghiệm được trình bày trong Phần V. Cuối cùng, Phần VI là kết luận của bài báo.

II. CÁC NGHIÊN CỨU LIÊN QUAN

Trong những năm đầu, nhiều nhà nghiên cứu [4,10,11] đã áp dụng các phương pháp học máy như SVM (Support Vector Machine), Random Forest vào phát hiện xâm nhập để nhằm phát hiện các tấn công mới mà các hệ thống dựa trên chữ ký không thể phát hiện. Mặc dù đạt được hiệu suất cao trong việc phát hiện xâm nhập, nhưng các phương pháp học máy truyền thống thường không có tính tổng quát cao, do đó không hiệu quả trong các môi trường mạng thực tế, khi lưu lượng thay đổi liên tục theo thời gian.

Để khắc phục những hạn chế này, một số nghiên cứu sử dụng các mô hình học sâu để cải thiện độ chính xác trong phát hiện xâm nhập. Các kỹ thuật học sâu để phát hiện xâm nhập mạng được sử dụng bao gồm MLP, CNN, RNN, và CNN-BiLSTM [4,6,7].

Mặc dù có hiệu suất tốt trong nhiều lĩnh vực như xử lý ảnh, xử lý văn bản, nhưng các mô hình này vẫn gặp thách thức trong phát hiện xâm nhập mạng, do các dữ liệu cần xử lý trong mạng thường là các dữ liệu phi cấu trúc, có nhiều tính chất phức tạp về mô hình mạng. Để xử lý các dữ liệu dạng phi cấu trúc, các nghiên cứu gần đây sử dụng cấu trúc mạng nơ-ron sâu mới là mạng nơ-ron đồ thị (GNN). Trong [12], Zhou và đồng sự đã sử dụng mạng GCN để phát hiện tấn công botnet thông qua các nhiệm vụ phân loại nút, bằng cách mô phỏng lưu lượng botnet song song với lưu lượng mạng bình thường. Tuy nhiên, nhiệm vụ phân loại nút thường không phù hợp cho các tấn công mạng phổ biến khác như DoS hay quét cổng do dữ liệu tấn mạng thường không nằm trong nút mạng. Để phát hiện các cuộc tấn công trên lưu lượng truy cập và luồng, nghiên cứu [10] đã áp dụng nhiệm vụ phân loại cạnh trên các bộ dữ liệu luồng và đạt được độ chính xác khá cao. Các nghiên cứu [13,14] sử dụng phương pháp học biểu diễn đồ thị hiện có và kết hợp các đặc trưng cạnh để nâng cao biểu diễn nút nhằm cải thiện hiệu suất.

Mô hình đề xuất của chúng tôi chuyển đổi dữ liệu mạng sang đồ thị trong đó luồng dữ liệu mạng được dùng để tạo thành nút, còn cạnh là các địa chỉ IP nguồn có chung trong luồng dữ liệu, tương tự như [5]. Tuy nhiên, mô hình đề xuất dựa trên kiến trúc mạng nơ-ron đồ thị tập trung có thể học được nhiều và chính xác hơn các thông tin hơn từ luồng mạng nhờ sử dụng GNN với nhúng kết hợp nút và cạnh, cùng với cơ chế tập trung trong việc phát hiện xâm nhập mạng.

III. CƠ SỞ LÝ THUYẾT

Phần này trình bày ngắn gọn cơ chế hoạt động của mạng nơ-ron đồ thị và mô hình GAT được sử dụng trong mô hình đề xuất.

A. Mạng nơ-ron đồ thị

Trong những năm gần đây, mạng nơ-ron đồ thị [15,16] đã thu hút sự chú ý lớn nhờ khả năng xử lý dữ liệu có cấu trúc đồ thị, cho phép học các đặc trưng của mạng và các mối quan hệ giữa các nút. GNN có thể được áp dụng cho các tác vụ ở các cấp độ khác nhau, bao gồm dự đoán cấp độ nút, cấp độ cạnh và cấp độ đồ thị. Trong nghiên cứu này, vấn đề phát hiện xâm nhập được chuyển thành bài toán phân loại nhị phân, trong đó mục tiêu là dự đoán xem một nút v có phải là độc hại (lớp 1) hay không (lớp 0).

Đối với một nút v trong đồ thị G, GNN học một biểu diễn ẩn, h_v , bằng cách sử dụng cả đặc trưng của nút x_v và cấu trúc đồ thị (các cạnh) E làm đầu vào. Đa số các mô hình GNN tiên tiến theo chiến lược truyền thông điệp và tổng hợp, gọi là kiến trúc MPNN (Message-Passing Neural Network), trong đó thông tin của một nút được truyền lần lượt đến các nút khác dọc theo các cạnh, và sau đó, biểu diễn của nút được cập nhật bằng cách tổng hợp và học thông tin của chính nó cùng với các nút láng giềng. Khi thực hiện truyền thông điệp với GNN có l lớp, mỗi nút có thể thu nhận thông tin của các nút láng giềng trong phạm vi l bước nhảy. Không làm mất tính tổng quát, biểu diễn ẩn ở lớp thứ l của GNN có thể được mô tả như sau:

$$h_v^l = f^l(h_v^{l-1}, \mathrm{AGG}^l(\{h_u^{l-1}, \forall u \in N_v\})) \quad (1)$$

Trong đó, N_v đại diện cho các nút láng giềng của nút *v*. AGG^l đại diện cho các phép tổng hợp. f^l là một hàm với các tham số có thể học. Sự kết hợp giữa AGG^l và f^l xác định loại toán tử đồ thị, ví dụ như toán tử tích chập đồ thị và toán tử đồ thị tập trung.

Đối với dự đoán cấp độ nút, đầu ra của lớp ẩn cuối cùng có thể được truyền trực tiếp vào lớp đầu ra Φ, ví dụ như lớp kết nối đầy đủ. Đối với dự đoán cấp độ đồ thị, đầu ra sẽ được truyền vào một hàm đọc *R* (để kết hợp thông tin từ các lớp ẩn của mô hình thành một đầu ra có ý nghĩa) với các tham số có thể học. Một cách chính thức, biểu diễn của lớp đầu ra có thể được biểu diễn như sau:

$$h_v = \Phi(h_v^l) \quad (2)$$

$$h_G = R(h_v^l \mid v \in V) \quad (3)$$

B. Mạng nơ-ron đồ thị tập trung (GAT)

Cơ chế tập trung được giới thiệu trong [17] để học trọng số cạnh giữa hai nút kết nối, trọng số tập trung α_{ij} được tính toán rõ ràng theo cơ chế tự tập trung cho các nút láng giềng (*i* và *j*), được mô tả theo công thức:

$$\alpha_{ij} = \frac{exp\ (LeakyReLU(a^T[Wh_i \| Wh_j]))}{\sum_{k \in N_i} exp\ (LeakyReLU(a^T[Wh_i \| Wh_k]))} \quad (4)$$

Trong đó, *a* và *W* là các tham số có thể học được, và ký hiệu ‖ biểu thị phép nối, hàm kích hoạt phi tuyến LeakyRELU giúp tránh hiện tượng gradient bằng không [17]. N_i đại diện cho các nút láng giềng của nút *i*. Tiếp đó, h_i' , biểu diễn ẩn cho nút i có thể được tính bằng cách học các biểu diễn ẩn có trọng số của các nút láng giềng h_j, với σ là một hàm kích hoạt phi tuyến:

$$h_i' = \sigma(\sum_{j \in N_i} \alpha_{ij} W h_j) \quad (5)$$

Cơ chế tập trung có thể được mở rộng với cơ chế tập trung nhiều đầu:

$$h_i' = \|_{k=1}^{K}\ \sigma(\sum_{j \in N_i} \alpha_{ij}^k W^k h_j) \quad (6)$$

Nhúng cạnh trong GAT: Mô hình GAT chủ yếu tập trung vào nhúng nút, nhưng có thể mở rộng hoặc điều chỉnh để học nhúng cho các cạnh trong đồ thị. Tuy nhiên, GAT không trực tiếp thiết kế để học nhúng cho các cạnh, mà thông tin về các cạnh có thể được gián tiếp học qua các nhúng của các nút đầu và cuối của chúng. Cách tiếp cận để tạo nhúng cho cạnh (v_i, v_j) là:

$$h_{(i,j)} = f(h_i, h_j) \quad (7)$$

Trong đó *f* có thể là một hàm kết hợp (chẳng hạn như phép ghép nối) hoặc một hàm mạng nơ-ron đơn giản. Một lựa chọn khác là sử dụng các hệ số tập trung α_{ij} để tạo ra nhúng cho cạnh, coi chúng như là trọng số của các cạnh trong đồ thị [17].

Đồ thị đầu vào cho GNN: Đầu vào cho các mô hình GNN cần là dữ liệu có cấu trúc đồ thị, do đó cần xây dựng các đồ thị từ mạng máy tính, có thể được đưa vào các mô hình GNN để học.

IV. PHƯƠNG PHÁP ĐỀ XUẤT

Bài toán phát hiện xâm nhập là một nhiệm vụ phân loại, có thể là phân loại các mẫu đầu vào thành hai loại là lành tính hoặc độc hại (phân loại nhị phân), hoặc theo cách chi tiết hơn là dự đoán các kịch bản tấn công cụ thể (phân loại đa lớp). Trong bài báo này, đầu ra của mô hình GNN là xác suất của các nút trong đồ thị đầu vào được coi là lành tính hoặc độc hại, tức là xác định luồng nào trong mạng là độc hại.

Phần dưới đây sẽ mô tả bài toán và đề xuất phương pháp mới để phát hiện xâm nhập trong mạng dựa trên mạng nơ ron đồ thị. Mô hình đề xuất gồm 2 phần chính: (1) xây dựng đồ thị để biểu diễn luồng dữ liệu mạng và (2) kiến trúc mạng nơ-ron đồ thị dựa trên GAT và nhúng kết hợp để đưa ra dự đoán về một nút có phải là độc hại/luồng tấn công hay không.

A. Biểu diễn dữ liệu mạng dưới dạng đồ thị

Để biểu diễn dữ liệu hoạt động mạng dưới dạng đồ thị, ta cần xác định rõ ràng cách định nghĩa nút và cạnh của đồ thị.

Nút: Mỗi nút trong đồ thị đại diện cho một luồng dữ liệu mạng. Các nút này được khởi tạo bằng dữ liệu được lấy từ các đặc trưng của luồng. Trong bài toán này, tập dữ liệu luồng mạng được sử dụng là CIC IDS 2017 [9]. Do đó nút được định nghĩa bằng các đặc trưng về thông tin luồng mạng, ví dụ như:

- Lượng dữ liệu: Tổng số byte hoặc gói tin được truyền trong luồng mạng.
- Tốc độ truyền dữ liệu: Tốc độ truyền dữ liệu trung bình của luồng mạng.
- Số lượng gói tin: Số lượng gói tin được truyền trong luồng mạng.
- Thời gian trễ: Thời gian trễ trung bình của các gói tin trong luồng mạng.

Và các đặc trưng khác còn lại. Trong bài báo này, chúng tôi tận dụng tất cả các thông tin đặc trưng trong luồng mạng, kể cả địa chỉ IP nguồn và IP đích, nhằm tránh mất thông tin trong quá trình huấn luyện.

Cạnh: Các cạnh của đồ thị được xây dựng dựa trên mối quan hệ giữa các luồng dữ liệu có chung địa chỉ IP nguồn. Nghĩa là, nếu hai luồng dữ liệu có cùng địa chỉ IP nguồn, chúng sẽ được kết nối với nhau bằng một cạnh.

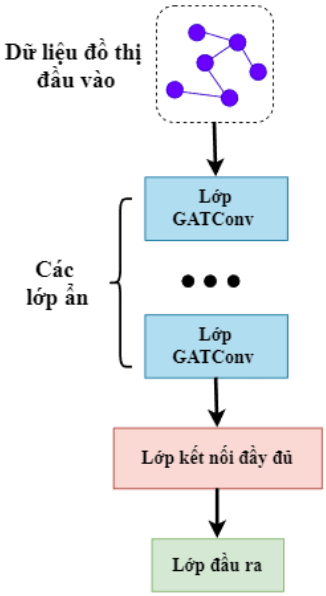
Phương pháp này giúp tạo ra một đồ thị chứa thông tin về mối quan hệ giữa các luồng dữ liệu mạng. Việc sử dụng cả nút và cạnh để biểu diễn đặc trưng của luồng dữ liệu giúp mô hình GNN đề xuất khai thác hiệu quả hơn các mẫu đặc trưng trong dữ liệu, từ đó đạt được độ chính xác cao hơn so với các mô hình chỉ sử dụng một số đặc trưng hạn chế.

Sau khi tạo được đồ thị với dữ liệu của các nút và các cạnh, đồ thị này sẽ được sử dụng làm dữ liệu đầu vào cho mô hình phân loại nút để tìm ra nút độc hại trong đồ thị, đại diện cho luồng dữ liệu độc hại.

B. Kiến trúc mạng nơ-ron đồ thị đề xuất

Nghiên cứu này đề xuất sử dụng phiên bản cải tiến của mô hình mạng nơ-ron đồ thị tập trung GAT với khả năng kết hợp cả nhúng nút và nhúng cạnh, để thực hiện nhiệm vụ phân loại nhị phân. Mô hình đồ thị này có thể học được

nhiều thông tin hữu ích trong đồ thị biểu diễn mạng máy tính với các luồng dữ liệu giàu thông tin.



Hình 1. Kiến trúc mô hình phát hiện xâm nhập đề xuất với các lớp GAT Conv

Mô hình được trình bày trong Hình 1 với các lớp GATConv kết hợp thông tin từ cả nhúng nút và cạnh để cập nhật nhúng nút (được cập nhật dựa trên nhúng của các nút lân cận và nhúng của các cạnh kết nối với nút đó), một lớp kết nối đầy đủ và lớp đầu ra. Các hàm kích hoạt ReLU và dropout cũng được sử dụng ngay sau mỗi lớp GATConv. Hàm softmax ở cuối mô hình giúp tạo ra các dự đoán hiệu quả nhất dựa trên lớp có xác suất đầu ra cao nhất. Kết quả của mô hình là phân loại các nút thành các loại bị tấn công hoặc hoạt động bình thường.

GAT có hai phiên bản là GATv1 và GATv2 [18]. GATv1 sử dụng một cơ chế tập trung tĩnh (static attention), nghĩa là các hệ số tập trung giữa các nút được tính toán một lần và cố định trong suốt quá trình học. Điều này có thể hạn chế khả năng của mô hình trong việc thích nghi với các thay đổi trong cấu trúc mạng hoặc các loại tấn công mới. Trong khi đó, GATv2 đưa vào cơ chế tập trung động, cho phép các hệ số tập trung thay đổi tùy thuộc vào từng truy vấn (query) khác nhau. Nói cách khác, mỗi khi một nút đóng vai trò là nút truy vấn, nó sẽ có một cách đánh giá khác nhau về tầm quan trọng của các nút láng giềng, dẫn đến các hệ số tập trung khác nhau. Nhờ cơ chế chú ý động, GATv2 cho phép mô hình linh hoạt hơn trong việc học các mối quan hệ phức tạp giữa các nút, đặc biệt trong các đồ thị không đồng nhất, đồng thời thích ứng tốt hơn với các dữ liệu đầu vào khác nhau. Điều này giúp mô hình học được những mối quan hệ phức tạp và động trong mạng, đặc biệt hữu ích trong việc phát hiện các loại tấn công mới nổi hoặc các biến thể của các cuộc tấn công đã biết. Mô hình đề xuất sẽ thử nghiệm cả 2 phiên bản này trong phần thực nghiệm.

V. THỰC NGHIỆM VÀ KẾT QUẢ

A. Tập dữ liệu

Tập dữ liệu CIC IDS 2017 [9], được công bố vào năm 2017 bởi Đại học New Brunswick, là một nguồn dữ liệu mạng đa dạng và hiện đại, được đánh giá cao trong cộng đồng nghiên cứu về phát hiện xâm nhập. Tập dữ liệu này bao gồm các gói tin mạng thu thập được trong một môi trường mạng mô phỏng, cùng với các luồng (flows) đã được phân loại chi tiết, bao gồm các loại tấn công như Brute Force, DDoS, và các hoạt động mạng bình thường. Đặc biệt, CIC IDS 2017 nổi bật với việc mô phỏng gần với môi trường mạng thực tế, cung cấp một nền tảng thử nghiệm hiệu quả cho các thuật toán phát hiện xâm nhập mới.

Cụ thể, bộ dữ liệu CIC IDS 2017 bao gồm lưu lượng truy cập bình thường và bất thường, với 13 loại tấn công được gán nhãn. Dữ liệu thô được lưu dưới dạng file .pcap và được chuyển đổi sang dữ liệu luồng mạng (flow) để giảm kích thước và thông tin dư thừa, sử dụng công cụ CIC-FlowMeter. Quá trình chuyển đổi đảm bảo các đặc trưng dữ liệu được trích xuất một cách nhất quán từ cùng các đặc trưng có trong dữ liệu thô ban đầu. Kết quả có được là bộ dữ liệu được gán nhãn chứa nhiều luồng dữ liệu được lựa chọn ngẫu nhiên, mỗi luồng có hơn 80 đặc trưng được trích xuất từ lưu lượng mạng thu thập được, gồm thời gian, địa chỉ IP nguồn và đích, cổng nguồn và đích, giao thức và loại tấn công. Tương tự với một số nghiên cứu trước đó [5,19,20], chúng tôi cũng thử nghiệm bằng cách sử dụng ngẫu nhiên 424.155 luồng dữ liệu từ bộ dữ liệu, bao gồm 80% luồng dữ liệu bình thường và 20% luồng dữ liệu độc hại (số lượng luồng bình thường nhiều hơn nhiều luồng dữ liệu độc hại). Mỗi luồng dữ liệu tương ứng với 1 nút trong đồ thị, do đó đồ thị được tạo sẽ có số lượng nút bằng số lượng luồng dữ liệu đã lựa chọn.

B. Thiết lập thực nghiệm

Hiệu năng của mô hình đề xuất được đo bằng độ đo F_1, được tính từ độ chính xác (precision), độ bao phủ (recall) theo các công thức như sau:

$$Precision = \frac{|A \cap B|}{|A|} \quad (8)$$

$$Recall = \frac{|A \cap B|}{|B|} \quad (9)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

Các tham số A và B ở công thức trên tương ứng là tập các nhãn được phát hiện và tập các nhãn đúng (được gán nhãn bởi người gán nhãn).

Để đánh giá hiệu năng của mô hình đề xuất, chúng tôi tiến hành thử nghiệm trên tập dữ liệu được chia ngẫu nhiên thành tập huấn luyện (train) và tập kiểm thử (test) với tỷ lệ 80% và 20%. Sau đó, 20% dữ liệu huấn luyện được trích ra làm dữ liệu đánh giá (valid) trong quá trình huấn luyện mô hình. Phương pháp đánh giá trong thực nghiệm là kiểm tra chéo 5 lần để đảm bảo độ tin cậy trong việc phân loại các cuộc tấn công mạng.

Các thực nghiệm đều sử dụng bộ tối ưu Adam optimizer với tham số *weight decay* là $1e^{-5}$. Hàm mất mát được sử dụng là BinaryEntropyLoss để phân loại một nút. Số lượng epoch là 100. Để chọn các giá trị phù hợp cho kích thước vector đặc trưng của nút và tốc độ học (learning rate), chúng tôi đã thực hiện nhiều thử nghiệm trong khi giữ các tham số khác không đổi và điều chỉnh một tham số cụ thể. Kết quả cho thấy mô hình đạt hiệu suất tối ưu khi

kích thước vector đặc trưng của nút 64, tốc độ học là 0,001. Chúng tôi cũng áp dụng cơ chế dừng sớm, quá trình huấn luyện sẽ dừng khi hiệu suất trên tập dữ liệu kiểm thử không được cải thiện nào trong ít nhất 5 epoch liên tiếp.

C. Kết quả thực nghiệm

Phần dưới đây sẽ mô tả các thực nghiệm để đánh giá hiệu năng của mô hình phát hiện tấn công mạng khi so sánh với các mô hình cơ sở khác.

1) Hiệu năng của phương pháp nhúng kết hợp so với chỉ nhúng nút

Bảng I. Hiệu năng của nhúng kết hợp và chỉ nhúng nút đối với mô hình đồ thị đề xuất

STT	Mô hình	Precision (%)	Recall (%)	F_1 (%)
1	Mô hình GATv2 nhúng nút	88,62	96,84	92,55
2	Mô hình GATv2 nhúng kết hợp	91,74	98,72	95,1

Kết quả thử nghiệm trong Bảng I cho thấy, hiệu suất của mô hình đồ thị nhúng kết hợp vượt trội lên tới 2,55% trên độ đo F_1 khi so với mô hình đồ thị chỉ có nhúng nút. Kết quả này thể hiện mô hình GATv2 nhúng kết hợp (cả nhúng nút và nhúng cạnh) đã tận dụng học được nhiều thông tin hơn từ cả nút và cạnh trong đồ thị, và do đó phân loại tốt hơn so với mô hình chỉ có nhúng nút. Trong những thử nghiệm dưới đây, chúng tôi sẽ chỉ so sánh mô hình đồ thị nhúng kết hợp với các mô hình cơ sở khác.

2) So sánh hiệu năng của mô hình đề xuất với các mô hình khác

Để đánh giá hiệu năng của kiến trúc mạng nơ-ron đồ thị đề xuất, chúng tôi sẽ so sánh kết quả với các mô hình học sâu của các nghiên cứu khác gần đây về phát hiện xâm nhập mạng trên cùng tập dữ liệu CIC IDS 2017, bao gồm MLP[7], CNN-BiLSTM[6]. Ngoài ra, chúng tôi cũng so sánh với các mô hình học máy truyền thống khác như Random Forest, SVM [11], và một mô hình mạng nơ-ron đồ thị GCN [22] đơn giản. Hiệu năng của các mô hình được so sánh được trình bày trong Bảng II.

Bảng II. Hiệu năng của các mô hình phát hiện xâm nhập

STT	Mô hình	Precision (%)	Recall (%)	F1 (%)
1	Random Forest [11]	88,47	90,25	89,35
2	SVM [11]	90,26	91,67	90,96
3	MLP [7]	81,7	99,5	87,3
4	CNN-BiLSTM [6]	95,9	86,2	90,8
5	GCN	89,43	93,68	91,51
6	GATv1 nhúng kết hợp	91,4	96,27	93,77
7	GATv2 nhúng kết hợp	91,74	98,72	95,10

Như thể hiện trong Bảng II, so với các mô hình học máy truyền thống khác thường được sử dụng phổ biến là Random Forest và SVM, các mô hình học sâu hầu hết đều có hiệu suất vượt trội do khả năng xử lý tốt các loại dữ liệu phi cấu trúc phức tạp, nhất là các mô hình GNN. Trong đó, hiệu năng của mô hình đề xuất vượt trội hơn với độ đo F_1 là 95,10 %, tốt hơn 4,14% so với mô hình SVM, hơn 5,75% so với mô hình Random Forest, và lên tới 7,8% so với mô hình MLP [7]. Khi so sánh với mô hình học sâu đề xuất trong các nghiên cứu khác như CNN-BiLSTM [6], mô hình mạng nơ-ron đồ thị đề xuất cũng có hiệu năng cao hơn tới 4,3% (tính theo độ đo F_1).

Khi so sánh với mô hình đồ thị GCN đơn giản, hiệu năng của mô hình đề xuất theo độ đo F_1 cũng cao hơn 3,59%. Chú ý rằng mô hình GCN cài đặt trong thử nghiệm cũng là mô hình nhúng nút nên không thể tận dụng tất cả các thông tin của mạng như trong mô hình đồ thị nhúng kết hợp đề xuất. GCN [22] cũng có hiệu năng thấp hơn GATv2 nhúng nút như kết quả trong Bảng I. Điều này cũng thể hiện sự hiệu quả của cơ chế tập trung trong đồ thị.

Chúng tôi cũng so sánh hiệu năng của mô hình GATv2 nhúng kết hợp đề xuất với phiên bản GATv1 như trong Bảng II, và kết quả cũng cho thấy rằng, khi sử dụng GATv2 cải tiến thì hiệu năng của mô hình đề xuất được tăng lên tới 1,33% (tính theo độ đo F_1). Điều này cho thấy, GATv2 có khả năng thích ứng tốt với mô hình đồ thị biểu diễn cho dữ liệu luồng mạng máy tính có tính phức tạp, hay thay đổi, nhờ khả năng tự động đánh giá tầm quan trọng của từng mối quan hệ giữa các nút trong đồ thị.

VI. KẾT LUẬN

Nghiên cứu này đã đề xuất một mô hình mạng nơ-ron đồ thị mới với nhúng kết hợp cả nút và cạnh, cùng với cơ chế tập trung (GAT) để phát hiện xâm nhập mạng. Mô hình tận dụng thông tin toàn diện về hoạt động mạng bằng cách kết hợp nhúng nút và cạnh trong đồ thị. Kết hợp với cơ chế học tập trung mạnh mẽ của GAT, mô hình đồ thị này có thể học được nhiều thông tin hữu ích từ đồ thị biểu diễn mạng máy tính với các luồng dữ liệu giàu thông tin.

Kết quả của nghiên cứu cho thấy mô hình đề xuất vượt trội hơn các phương pháp hiện có về độ chính xác và khả năng chống lại các tấn công mạng. So sánh với các mô hình học máy truyền thống (Random Forest, SVM [11]) và học sâu đề xuất trước đó (MLP [7], CNN-BiLSTM [6]), mô hình đề xuất đạt được kết quả tốt hơn đáng kể. Kết quả này là do việc sử dụng nhúng kết hợp cả nút và cạnh mang lại hiệu suất cao hơn so với chỉ sử dụng nhúng nút. Thêm vào đó, mô hình GATv2 với cơ chế tập trung động hiệu quả hơn GATv1 trong việc thích ứng với các thay đổi trong cấu trúc mạng và các loại tấn công mới. Nghiên cứu này cũng chứng minh sự hiệu quả của việc sử dụng GNN với nhúng kết hợp cả nút và cạnh, cùng cơ chế tập trung trong việc phát hiện xâm nhập mạng.

LỜI CẢM ƠN

Nghiên cứu này được hỗ trợ bởi Đề tài mã số 10-2024-HV-CNTT1, Học viện Công nghệ Bưu chính Viễn thông.

TÀI LIỆU THAM KHẢO

[1] IBM, "Cost of a Data Breach Report 2024," https://www.ibm.com/reports/data-breach.

[2] Steve Morgan, "Cybercrime To Cost The World $10.5 Trillion Annually By 2025," https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025/.

[3] "Việt Nam ghi nhận hơn 4.000 sự cố tấn công mạng trong 8 tháng đầu năm 2024," https://antoanthongtin.gov.vn/chinh-sach---chien-luoc/viet-

nam-ghi-nhan-hon-4000-su-co-tan-cong-mang-trong-8-thang-dau-nam-2024-110642.

[4] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, 2021, doi: 10.1002/ett.4150.

[5] D. H. Tran and M. Park, "FN-GNN: A Novel Graph Embedding Approach for Enhancing Graph Neural Networks in Network Intrusion Detection Systems," Applied Sciences (Switzerland), vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14166932.

[6] A. I. Getman, M. N. Goryunov, A. G. Matskevich, D. A. Rybolovlev, and A. G. Nikolskaya, "Deep Learning Applications for Intrusion Detection in Network Traffic," Proceedings of the Institute for System Programming of the RAS, vol. 35, no. 4, 2023, doi: 10.15514/ispras-2023-35(4)-3.

[7] O. Belarbi, A. Khan, P. Carnelli, and T. Spyridopoulos, "An Intrusion Detection System Based on Deep Belief Networks," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2022. doi: 10.1007/978-3-031-17551-0_25.

[8] T. Bilot, N. El Madhoun, K. Al Agha, and A. Zouaoui, "Graph Neural Networks for Intrusion Detection: A Survey," IEEE Access, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3275789.

[9] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy, 2018. doi: 10.5220/0006639801080116.

[10] M. Aamir, S. S. H. Rizvi, M. A. Hashmani, M. Zubair, and J. A. . Usman, "Machine Learning Classification of Port Scanning and DDoS Attacks: A Comparative Analysis," Mehran University Research Journal of Engineering and Technology, vol. 40, no. 1, 2021, doi: 10.22581/muet1982.2101.19.

[11] S. S. Panwar, Y. P. Raiwani, and L. S. Panwar, "An Intrusion Detection Model for CICIDS-2017 Dataset Using Machine Learning Algorithms," in 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM), 2022, pp. 1–10. doi: 10.1109/ICACCM56405.2022.10009400.

[12] J. Zhou, Z. Xu, A. M. Rush, and M. Yu, "Automating botnet detection with graph neural networks," arXiv preprint arXiv:2003.06344, 2020.

[13] L. Gong and Q. Cheng, "Exploiting edge features for graph neural networks," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019. doi: 10.1109/CVPR.2019.00943.

[14] M. Mirlashari and S. A. M. Rizvi, "Enhancing IoT intrusion detection system with modified E-GraphSAGE: a graph neural network approach," International Journal of Information Technology (Singapore), vol. 16, no. 4, 2024, doi: 10.1007/s41870-024-01746-9.

[15] K. Xu, S. Jegelka, W. Hu, and J. Leskovec, "How powerful are graph neural networks?," in 7th International Conference on Learning Representations, ICLR 2019, 2019.

[16] J. Zhou et al., "Graph neural networks: A review of methods and applications," 2020. doi: 10.1016/j.aiopen.2021.01.001.

[17] P. Veličković, A. Casanova, P. Liò, G. Cucurull, A. Romero, and Y. Bengio, "Graph attention networks," in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018. doi: 10.1007/978-3-031-01587-8_7.

[18] S. Brody, U. Alon, and E. Yahav, "HOW ATTENTIVE ARE GRAPH ATTENTION NETWORKS?," in ICLR 2022 - 10th International Conference on Learning Representations, 2022.

[19] Z. Sun, A. M. H. Teixeira, and S. Toor, "GNN-IDS: Graph Neural Network based Intrusion Detection System," in ACM International Conference Proceeding Series, Association for Computing Machinery, Jul. 2024. doi: 10.1145/3664476.3664515.

[20] H. D. Le and M. Park, "Enhancing Multi-Class Attack Detection in Graph Neural Network through Feature Rearrangement," Electronics (Switzerland), vol. 13, no. 12, Jun. 2024, doi: 10.3390/electronics13122404.

[21] M. Fey and J. E. Lenssen, "Fast Graph Representation Learning with PyTorch Geometric," in ICLR Workshop on Representation Learning on Graphs and Manifolds, 2019.

[22] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," Comput Soc Netw, vol. 6, no. 1, 2019, doi: 10.1186/s40649-019-0069-y.

NETWORK INTRUSION DETECTION USING GRAPH NEURAL NETWORKS WITH NODE AND EDGE FEATURES

Abstract: This paper proposes a novel approach using a Graph Neural Network with an attention mechanism and combined node and edge embedding for network intrusion detection. Thanks to the ability to leverage comprehensive information about network activity by combining node and edge embedding in the graph, the proposed model can learn complex relationships between network data flows. Combined with the robust focused learning mechanism of the Graph Attention Networks (GAT) model, the proposed graph model can learn much useful information in the graph representing computer networks with information-rich data flows. In the proposed model, each node in the graph represents a network data flow, initialized with all the features of the data flow. The edges of the graph are constructed based on the relationship between data flows sharing the same source IP address. The performance of the model is evaluated on the standard CIC IDS 2017 dataset. Experimental results show that the proposed model outperforms traditional machine learning methods and deep learning methods, as well as the simple GCN graph neural network model.

Keywords: *Intrusion Detection System, graph neural network, node and edge features.*

Nguyễn Thị Thanh Thủy, nhận học vị Tiến sĩ năm 2023. Hiện đang công tác tại Khoa Công nghệ Thông tin 1, và Lab Học máy và ứng dụng, Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực nghiên cứu: trí tuệ nhân tạo, học máy, xử lý ngôn ngữ tự nhiên.

Email: thuyntt@ptit.edu.vn

Nguyễn Ngọc Điệp, nhận học vị Tiến sĩ năm 2017. Hiện đang công tác tại Khoa An toàn thông tin, Lab Học máy và ứng dụng, Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực nghiên cứu: học máy, an toàn thông tin, xử lý ngôn ngữ tự nhiên.

Email: diepnguyenngoc@ptit.edu.vn