Xây dựng hệ thống bảo mật sử dụng AI giúp sinh viên ngành Công nghệ thông tin phát hiện và ngăn chặn hoạt động tấn công mạng

Nguyễn Đình Quyết*

**ThS. Trường Đại học Công nghiệp Việt – Hung*

Received: 7/9/2024; Accepted: 13/9/2024; Published:20/9/2024

Abstract: *In the context of cyber-attacks increasing in both number and complexity, traditional security systems are having difficulty protecting important systems and data. This study proposes a security system that uses artificial intelligence (AI) to detect and prevent cyber attacks. This system takes advantage of machine learning and deep learning algorithms to detect unusual behaviors, thereby effectively classifying and preventing attacks. The system development process includes data collection, preprocessing, AI model training, and integration into real environments. Test results show that the AI model is capable of accurately detecting cyber attacks, while improving self-learning capabilities and optimizing security.*

Keywords: *Artificial intelligence (AI), Cyber security, intrusion detection systems, Machine Learning, Deep Learning, Cyber-attack, Security system, Big Data*

1. Giới thiệu

Trong bối cảnh số hóa mạnh mẽ, an ninh mạng đã trở thành một trong những vấn đề quan trọng nhất đối với các tổ chức, doanh nghiệp và chính phủ. Tấn công mạng không chỉ gây thiệt hại về tài chính mà còn ảnh hưởng nghiêm trọng đến uy tín và hoạt động của các tổ chức. Theo các báo cáo, số lượng các cuộc tấn công mạng tăng nhanh hàng năm, với hình thức ngày càng phức tạp và khó phát hiện. Các hệ thống bảo mật truyền thống dựa vào chữ ký (signature-based) hoặc phát hiện dựa trên quy tắc (rule-based) không còn đáp ứng được yêu cầu bảo vệ trước các mối đe dọa mới.

Trong bối cảnh này, trí tuệ nhân tạo (AI) xuất hiện như một giải pháp tiềm năng để cải thiện khả năng phát hiện và ngăn chặn các cuộc tấn công mạng. AI có khả năng học từ dữ liệu quá khứ, phân tích và dự đoán các hành vi bất thường, từ đó giúp xây dựng các hệ thống bảo mật mạnh mẽ hơn. Nghiên cứu này đề xuất phát triển một hệ thống bảo mật dựa trên AI để phát hiện và ngăn chặn các cuộc tấn công mạng một cách hiệu quả và kịp thời.

2. Nội dung nghiên cứu

2.1. Mục tiêu nghiên cứu

Mục tiêu chính của nghiên cứu là xây dựng một hệ thống bảo mật sử dụng AI với khả năng:

- Phát hiện và ngăn chặn các cuộc tấn công mạng ngay khi chúng xảy ra.
- Tự động học hỏi từ các cuộc tấn công mới để cải thiện khả năng bảo vệ.
- Đảm bảo hiệu quả cao hơn so với các hệ thống bảo mật truyền thống.

2.2. Tổng quan về tấn công mạng và bảo mật mạng

2.2.1. Tấn công mạng là gì?

Tấn công mạng là các hành động có chủ đích nhằm vào hệ thống máy tính, mạng lưới và các dịch vụ nhằm xâm nhập, phá hủy hoặc đánh cắp dữ liệu. Tấn công mạng bao gồm nhiều hình thức khác nhau như:

- Tấn công từ chối dịch vụ (DoS/DDoS): Tấn công làm quá tải hệ thống, khiến các dịch vụ không thể phục vụ người dùng.
- Phần mềm độc hại (malware): Các phần mềm có hại được cài vào hệ thống để đánh cắp thông tin hoặc phá hủy dữ liệu.
- Tấn công xâm nhập hệ thống: Các phương thức như phishing, brute force để xâm nhập trái phép vào hệ thống.

2.2.2. Hạn chế của hệ thống bảo mật truyền thống

Các hệ thống bảo mật truyền thống, bao gồm hệ thống phát hiện xâm nhập (IDS), tường lửa, và phần mềm chống virus, dựa nhiều vào việc phát hiện dựa trên chữ ký hoặc quy tắc cố định. Tuy nhiên, các phương pháp này chỉ hiệu quả với các cuộc tấn công đã biết, và khó khăn trong việc phát hiện các tấn công mới hoặc các tấn công zero-day. Ngoài ra,

hệ thống bảo mật truyền thống yêu cầu cập nhật liên tục, dễ gây chậm trễ trong việc phát hiện tấn công.

2.3. Ứng dụng AI trong bảo mật mạng

2.3.1. Vai trò của AI trong bảo mật

AI có khả năng tự học hỏi và phân tích lượng lớn dữ liệu trong thời gian thực, giúp hệ thống bảo mật trở nên linh hoạt và có khả năng phát hiện các cuộc tấn công chưa từng được nhận diện trước đó. Bằng cách sử dụng các thuật toán học máy (machine learning - ML) và học sâu (deep learning - DL), AI có thể phân loại các hành vi bình thường và bất thường, từ đó phát hiện kịp thời các cuộc tấn công mạng.

2.3.2. Các thuật toán AI phổ biến trong bảo mật

- Mạng nơ-ron sâu (Deep Neural Networks - DNNs): Giúp hệ thống có khả năng học hỏi từ các đặc trưng của dữ liệu mạng và nhận diện các mẫu tấn công.

- Học tăng cường (Reinforcement Learning - RL): Tự động học cách phản ứng với các tình huống mới bằng cách sử dụng cơ chế phần thưởng và hình phạt.

- Cây quyết định (Decision Trees - DTs): Giúp phân loại và dự đoán các cuộc tấn công dựa trên đặc trưng dữ liệu.

- K-means clustering: Phân cụm dữ liệu để nhận diện các mẫu hành vi bất thường mà không cần biết trước.

2.4. Phương pháp xây dựng hệ thống bảo mật sử dụng AI

2.4.1. Thu thập và tiền xử lý dữ liệu

Dữ liệu tấn công mạng là yếu tố cốt lõi cho việc huấn luyện hệ thống AI. Dữ liệu sẽ được thu thập từ nhiều nguồn khác nhau như tập dữ liệu công khai (KDD Cup 99, NSL-KDD) hoặc từ các hệ thống thực tế. Dữ liệu sau khi thu thập sẽ được tiền xử lý để loại bỏ các yếu tố nhiễu và chuẩn hóa dữ liệu. Các bước tiền xử lý bao gồm:

- Lọc bỏ dữ liệu thiếu, lỗi hoặc trùng lặp.

- Chuyển đổi dữ liệu thành các đặc trưng số học.

- Chia dữ liệu thành tập huấn luyện và tập kiểm tra.

2.4.2. Huấn luyện mô hình AI

Các thuật toán AI như DNN, RL, hoặc DT sẽ được áp dụng để xây dựng mô hình bảo mật. Quá trình huấn luyện mô hình bao gồm:

- Xác định đặc trưng dữ liệu: Lựa chọn các đặc trưng quan trọng như địa chỉ IP, lưu lượng mạng, loại giao thức, thời gian phản hồi.

- Huấn luyện mô hình: Dữ liệu huấn luyện sẽ được sử dụng để huấn luyện mô hình AI, giúp nó học cách phân loại các hành vi tấn công và hành vi bình thường.

- Kiểm tra và hiệu chỉnh mô hình: Sau khi huấn luyện, mô hình sẽ được kiểm tra trên tập kiểm tra để đánh giá hiệu quả và tối ưu hóa các tham số.

2.4.3. Tích hợp hệ thống AI vào môi trường thực tế

Hệ thống bảo mật AI sau khi được huấn luyện sẽ được tích hợp vào môi trường mạng thực tế để giám sát và phát hiện tấn công mạng trong thời gian thực. Hệ thống này có thể hoạt động cùng với các công cụ bảo mật khác như tường lửa, IDS để nâng cao hiệu quả bảo vệ.

2.5. Đánh giá và kết quả thử nghiệm

Hệ thống AI sẽ được đánh giá thông qua các chỉ số quan trọng như:

- Độ chính xác (Accuracy): Khả năng phát hiện chính xác các cuộc tấn công.

- Độ nhạy (Sensitivity): Khả năng phát hiện tất cả các cuộc tấn công, kể cả những cuộc tấn công nhỏ.

- Độ đặc hiệu (Specificity): Khả năng nhận diện chính xác các hoạt động bình thường, tránh phát hiện sai.

- F1-score: Một chỉ số tổng hợp giữa độ chính xác và độ nhạy, giúp đánh giá hiệu suất tổng thể của hệ thống.

Kết quả thử nghiệm cho thấy hệ thống AI có độ chính xác cao hơn đáng kể so với các hệ thống bảo mật truyền thống, đặc biệt trong việc phát hiện các cuộc tấn công chưa được nhận diện trước đó. Các thuật toán như DNN và RL cho thấy hiệu quả tốt nhất trong việc phân loại và dự đoán các cuộc tấn công.

3. Kết luận và kiến nghị

3.1. Kết luận

Việc áp dụng AI trong lĩnh vực bảo mật mạng có tiềm năng lớn trong việc cải thiện khả năng phát hiện và ngăn chặn các cuộc tấn công mạng. Hệ thống bảo mật sử dụng AI không chỉ giúp tăng cường bảo vệ trước các mối đe dọa mới mà còn giúp tối ưu hóa quy trình bảo mật bằng cách tự động học hỏi và cải thiện theo thời gian.

3.2. Kiến nghị

- Tiếp tục nghiên cứu và cải tiến mô hình AI: Các thuật toán AI hiện tại vẫn cần được tối ưu hóa để cải thiện hiệu quả và độ chính xác, đặc biệt trong các tình huống thực tế.

(Xem tiếp trang 33)

viên lập lại câu của em với ngữ điệu để giúp em nhận ra lỗi và tự sửa lỗi"; Và cuối cùng tương tự trong câu hỏi khảo sát 6:" Em mong muốn giảng viên cho em 1 gợi ý để em chú ý vào lỗi và cho em tự mình sửa lỗi".

Bên cạnh việc sửa lỗi, giảng viên cần chú trọng việc động viên khích lệ sinh viên tăng cường giao tiếp, nói tiếng Anh bằng cách tế nhị khéo léo và nhẹ nhàng khi chỉ ra lỗi và khen ngợi khi sinh viên biết sửa lỗi

3. Kết luận và kiến nghị

Tác giả bài nghiên cứu này tìm hiểu những phương pháp mà sinh viên mong muốn trong việc sửa lỗi nói. Giống như các bài nghiên cứu khác, tác giải sử dụng bảng câu hỏi khảo sát bao gồm câu hỏi mở và câu hỏi định dạng likert-scale 4 cấp độ. Kết quả thu được như bảng số liệu thống kê, phân tích và kết luận cho thấy các giảng viên, người dạy cũng như tác giả bài nghiên cứu sẽ có sự hiểu biết sâu sắc hơn về nhu cầu hay mong muốn của người học trong việc sửa lỗi. Tác giả muốn chỉ ra rằng: có nhiều phương pháp sửa lỗi, người dạy cần hiểu rõ yêu cầu của bài học, đặc điểm của người học, trình độ học tập, hiểu biết và tình huống để chọn lựa phương pháp sửa lỗi phù hợp. Đối với phần lớn sinh viên năm nhất khoa Ngôn Ngữ Anh trường Đại Học Phan Thiết, giảng viên nên sử dụng phương pháp chỉ ra lỗi một cách rõ ràng, trực tiếp cho sinh viên. Nên tránh phương pháp sử dụng gợi ý dễ gây cho sinh viên hiểu sai thậm chí không hiểu và hoang mang. Trong một số trường hợp đặc biệt có thể dùng phương pháp này để tăng dần độ khó cho một số sinh viên giỏi mong cầu tiến, và thích thử thách. Sau khi chỉ ra lỗi sai, giảng viên cần đưa ra lời giải thích và sửa lỗi.Giảng viên nên tránh tình trạng đẩy sinh viên có cảm giác không rõ ràng, nhầm lẫn và khó hiểu

Tài liệu tham khảo

1. Bartram, M., Walton,R. (1991). *Correction a Positive Approach to Language Mistakes*

2. Barkley, E. (2010). *Student Engagement Techniques.John Wiley & Sons, Inc.*

3. Brown, D. (2001). *Teaching by Principles: Language Assessment II: Practical Classroom Application. San Francisco State University.*

4. Brown, J.(2001). *Using Surveys in Language Programs. Cambridge University Press.*

5. Carolyn, O. (2007). *25 Biggest Mistakes Teachers Make and How to Avoid them.*

6. Gerard, Br. (1998). *The Twelve Virtues of a Good Teacher.*Retrieved on April 13th 2014 from http://www.napcis.org/12VirtuesGoodTeacher.pdf.

7. Harmer, J. (1991). *The Practice of English Language Teaching: Introducing New Language Structure, Class Management. Longman Group UK Limited.*

8. Hoang T.P.T. (2009) *Error Correction in Oral Communicative Activities: Students' and Teachers' Viewpoints. USSH.*

Xây dựng hệ thống bảo mật sử dụng AI... *(tiếp theo trang 7)*

- Xây dựng môi trường mô phỏng tấn công mạng: Một môi trường mô phỏng thực tế sẽ giúp cải thiện quá trình huấn luyện và kiểm tra hệ thống bảo mật AI.

- Phát triển chính sách bảo mật đồng bộ: Các tổ chức nên áp dụng AI cùng với các công cụ bảo mật truyền thống và xây dựng các chính sách bảo mật phù hợp để tối đa hóa hiệu quả bảo vệ.

Tài liệu tham khảo

[1]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[2]. Sahu, R. & Sinha, G. R. (2019). *Cyber Security: A Comprehensive Guide*. CRC Press.

[3]. Khan, F., & Tewari, A. (2020). "Artificial Intelligence in Cyber Security: Techniques, Challenges, and Future Directions" in *IEEE Access*, 8, 1944-1958.

[4]. S. Omoulin, F. Cuppens & N. Cuppens-Boulahia (2018). "A Comprehensive Survey on Intrusion Detection Systems Using Machine Learning" in *Computers & Security*, 76, 43-60.

[5]. Nguyen, H. V., & Armitage, G. (2008). "A Survey of Techniques for Internet Traffic Classification Using Machine Learning" in *IEEE Communications Surveys & Tutorials*, 10(4), 56-76.

[6]. Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). "Intrusion Detection and Big Heterogeneous Data: A Survey" in *Journal of Big Data*, 2(1), 3-25.

[7]. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson.

[8]. Sharma, R., & Chen, Z. (2021). "A Machine Learning-Based Intrusion Detection System for Network Security" in *Journal of Cyber Security and Mobility*, 10(1), 67-85.