Nghiên cứu giải pháp phát hiện xâm nhập dựa trên học sâu cho hạ tầng điện toán đám mây

Nguyễn Văn Demo[1], Trần Thị Kịch Bản[2]

[1]Học viện Kỹ thuật mật mã · [2]Viện Khoa học Công nghệ mật mã

Tóm tắt

Bài báo trình bày một giải pháp phát hiện xâm nhập (IDS) dựa trên mô hình học sâu áp dụng cho hạ tầng điện toán đám mây nhiều thuê bao. Nhóm tác giả đề xuất kiến trúc lai giữa mạng nơ-ron tích chập một chiều và mạng bộ nhớ dài ngắn hạn nhằm khai thác đồng thời đặc trưng không gian của gói tin và đặc trưng thời gian của luồng kết nối. Giải pháp được đánh giá trên hai tập dữ liệu chuẩn là CIC-IDS2017 và UNSW-NB15 với các chỉ số độ chính xác, độ nhạy, độ đặc hiệu và điểm F1. Kết quả thực nghiệm cho thấy mô hình đề xuất đạt độ chính xác trung bình 98,4 phần trăm, cao hơn 2,7 điểm phần trăm so với mô hình cơ sở sử dụng rừng ngẫu nhiên, đồng thời giảm 31 phần trăm tỷ lệ cảnh báo sai. Bài báo cũng phân tích chi phí tính toán khi triển khai mô hình ở biên của hạ tầng đám mây và đề xuất cơ chế cắt tỉa trọng số để giảm độ trễ suy luận.

Từ khóa

Từ khóa: phát hiện xâm nhập, học sâu, điện toán đám mây, an toàn thông tin, mạng nơ-ron tích chập.

Abstract

This paper presents a deep learning based intrusion detection solution for multi tenant cloud computing infrastructure. The authors propose a hybrid architecture combining a one dimensional convolutional neural network and a long short term memory network in order to jointly exploit the spatial features of packets and the temporal features of connection flows. The solution is evaluated on two benchmark datasets, CIC-IDS2017 and UNSW-NB15, using accuracy, recall, specificity and F1 score. Experimental results show that the proposed model reaches an average accuracy of 98.4 percent, which is 2.7 percentage points higher than the random forest baseline, while reducing the false alarm rate by 31 percent. The paper also analyses the computational cost of deploying the model at the edge of the cloud infrastructure and proposes a weight pruning mechanism to reduce inference latency.

Keywords

Keywords: intrusion detection, deep learning, cloud computing, information security, convolutional neural network.

1. GIỚI THIỆU

Điện toán đám mây đã trở thành nền tảng hạ tầng chủ đạo cho các hệ thống thông tin của cơ quan nhà nước và doanh nghiệp tại Việt Nam. Việc tập trung dữ liệu và dịch vụ trên một hạ tầng dùng chung mang lại hiệu quả về chi phí và khả năng mở rộng, nhưng đồng thời làm gia tăng bề mặt tấn công. Một sự cố an toàn thông tin trên hạ tầng dùng chung có thể ảnh hưởng dây chuyền tới nhiều thuê bao khác nhau, gây thiệt hại lớn hơn nhiều so với mô hình hạ tầng truyền thống.

Hệ thống phát hiện xâm nhập là một trong những lớp phòng thủ quan trọng nhất trong kiến trúc bảo vệ nhiều lớp. Các hệ thống dựa trên dấu hiệu tuy có độ chính xác cao với các tấn công đã biết nhưng gần như bất lực trước các biến thể mới và các tấn công khai thác lỗ hổng chưa được công bố. Ngược lại, các hệ thống dựa trên bất thường có khả năng phát hiện tấn công mới nhưng thường có tỷ lệ cảnh báo sai cao, làm giảm hiệu quả vận hành của trung tâm điều hành an ninh mạng.

Trong những năm gần đây, các kỹ thuật học sâu đã được ứng dụng rộng rãi trong lĩnh vực phát hiện xâm nhập nhờ khả năng tự động trích rút đặc trưng từ dữ liệu thô mà không cần quá trình thiết kế đặc trưng thủ công. Tuy nhiên, phần lớn các nghiên cứu hiện có tập trung vào việc nâng cao độ chính xác trên tập dữ liệu chuẩn mà chưa xem xét đầy đủ tới chi phí tính toán và độ trễ khi triển khai thực tế trên hạ tầng đám mây.

Bài báo này đề xuất một kiến trúc lai nhằm cân bằng giữa độ chính xác phát hiện và chi phí triển khai. Đóng góp chính của bài báo gồm ba điểm. Thứ nhất, nhóm tác giả đề xuất kiến trúc kết hợp mạng tích chập một chiều với mạng bộ nhớ dài ngắn hạn. Thứ hai, nhóm tác giả xây dựng quy trình tiền xử lý dữ liệu luồng mạng phù hợp với môi trường nhiều thuê bao. Thứ ba, nhóm tác giả đề xuất cơ chế cắt tỉa trọng số giúp giảm độ trễ suy luận mà chỉ làm suy giảm không đáng kể độ chính xác.

2. CÁC NGHIÊN CỨU LIÊN QUAN

Các nghiên cứu về phát hiện xâm nhập có thể chia thành ba nhóm chính. Nhóm thứ nhất sử dụng các thuật toán học máy cổ điển như máy véc tơ hỗ trợ, cây quyết định và rừng ngẫu nhiên. Nhóm này có ưu điểm là dễ huấn luyện, dễ giải thích và chi phí tính toán thấp, nhưng phụ thuộc nhiều vào chất lượng của quá trình thiết kế đặc trưng thủ công.

Nhóm thứ hai sử dụng các mô hình học sâu thuần túy như mạng nơ-ron nhiều lớp, mạng tích chập hoặc mạng hồi quy. Các nghiên cứu theo hướng này thường cho độ chính xác cao hơn nhóm thứ nhất trên cùng tập dữ liệu, nhưng đòi hỏi lượng dữ liệu huấn luyện lớn và tài nguyên tính toán đáng kể.

Nhóm thứ ba, cũng là hướng tiếp cận của bài báo này, sử dụng các kiến trúc lai nhằm khai thác điểm mạnh của nhiều loại mạng khác nhau. Các kết quả đã công bố cho thấy

kiến trúc lai thường đạt hiệu năng tốt hơn khi dữ liệu vừa có cấu trúc không gian vừa có tính chất chuỗi thời gian, đúng như đặc điểm của dữ liệu luồng mạng.

3. PHƯƠNG PHÁP ĐỀ XUẤT

Kiến trúc đề xuất gồm bốn khối chức năng. Khối tiền xử lý thực hiện chuẩn hóa dữ liệu luồng, mã hóa các trường phân loại và cân bằng lại phân bố nhãn bằng kỹ thuật lấy mẫu tổng hợp. Khối trích đặc trưng không gian sử dụng ba lớp tích chập một chiều với kích thước bộ lọc lần lượt là bảy, năm và ba, mỗi lớp đi kèm một lớp chuẩn hóa theo lô và hàm kích hoạt tuyến tính chỉnh lưu.

Khối trích đặc trưng thời gian gồm hai lớp bộ nhớ dài ngắn hạn hai chiều với số đơn vị ẩn lần lượt là một trăm hai mươi tám và sáu mươi tư. Đầu ra của khối này được đưa qua cơ chế chú ý nhằm gán trọng số cao hơn cho các bước thời gian mang nhiều thông tin phân biệt. Khối phân lớp gồm hai lớp kết nối đầy đủ và một lớp đầu ra sử dụng hàm kích hoạt phi tuyến chuẩn hóa xác suất.

Để giảm chi phí triển khai, nhóm tác giả áp dụng cơ chế cắt tỉa trọng số theo độ lớn kết hợp với lượng tử hóa tám bit. Quá trình cắt tỉa được thực hiện theo nhiều vòng, mỗi vòng loại bỏ mười phần trăm trọng số có giá trị tuyệt đối nhỏ nhất và huấn luyện lại mô hình trong năm chu kỳ để phục hồi độ chính xác.

4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Thực nghiệm được tiến hành trên máy chủ trang bị bộ xử lý đồ họa chuyên dụng với bộ nhớ hai mươi tư gigabyte. Hai tập dữ liệu được sử dụng là CIC-IDS2017 và UNSW-NB15, chia theo tỷ lệ bảy mươi phần trăm cho huấn luyện, mười lăm phần trăm cho kiểm định và mười lăm phần trăm cho kiểm thử. Các chỉ số đánh giá gồm độ chính xác, độ nhạy, độ đặc hiệu và điểm F1.

Kết quả cho thấy mô hình đề xuất đạt độ chính xác chín mươi tám phẩy bốn phần trăm trên tập CIC-IDS2017 và chín mươi sáu phẩy chín phần trăm trên tập UNSW-NB15. So với mô hình cơ sở rừng ngẫu nhiên, mức cải thiện lần lượt là hai phẩy bảy và ba phẩy một điểm phần trăm. Tỷ lệ cảnh báo sai giảm ba mươi mốt phần trăm, đây là chỉ số có ý nghĩa vận hành quan trọng đối với trung tâm điều hành an ninh mạng.

Sau khi áp dụng cắt tỉa và lượng tử hóa, kích thước mô hình giảm từ bốn mươi hai megabyte xuống mười một megabyte, độ trễ suy luận trung bình giảm từ hai mươi ba mili giây xuống chín mili giây trên mỗi lô dữ liệu, trong khi độ chính xác chỉ suy giảm không quá không phẩy năm điểm phần trăm. Kết quả này cho thấy mô hình hoàn toàn khả thi khi triển khai ở biên hạ tầng.

Nhóm tác giả cũng tiến hành thực nghiệm loại bỏ từng thành phần nhằm đánh giá đóng góp riêng của mỗi khối. Khi loại bỏ cơ chế chú ý, điểm F1 giảm một phẩy hai điểm. Khi

thay khối bộ nhớ dài ngắn hạn bằng một lớp kết nối đầy đủ, điểm F1 giảm ba phẩy sáu điểm, cho thấy đặc trưng thời gian đóng vai trò quan trọng trong bài toán này.

5. KẾT LUẬN

Bài báo đã đề xuất và đánh giá một giải pháp phát hiện xâm nhập dựa trên kiến trúc học sâu lai cho hạ tầng điện toán đám mây. Kết quả thực nghiệm trên hai tập dữ liệu chuẩn cho thấy giải pháp đạt độ chính xác cao đồng thời giảm đáng kể tỷ lệ cảnh báo sai so với mô hình cơ sở. Cơ chế cắt tỉa và lượng tử hóa giúp mô hình đủ nhẹ để triển khai ở biên hạ tầng.

Hướng nghiên cứu tiếp theo của nhóm tác giả là mở rộng mô hình cho dữ liệu mã hóa đầu cuối, nơi nội dung gói tin không còn khả dụng, và nghiên cứu cơ chế học liên kết nhằm huấn luyện mô hình trên dữ liệu của nhiều thuê bao mà không cần tập trung dữ liệu về một nơi.

Tài liệu tham khảo

[1] Nguyễn Văn A, Trần Văn B (2024), Tổng quan các kỹ thuật phát hiện xâm nhập dựa trên học máy, Tạp chí An toàn thông tin, số 3, trang 12-25.

[2] Sharafaldin I., Lashkari A. H., Ghorbani A. A. (2018), Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization, ICISSP.

[3] Moustafa N., Slay J. (2015), UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems, MilCIS.

[4] Hochreiter S., Schmidhuber J. (1997), Long Short-Term Memory, Neural Computation, vol. 9, no. 8, pp. 1735-1780.

[5] Han S., Pool J., Tran J., Dally W. (2015), Learning both Weights and Connections for Efficient Neural Networks, NeurIPS.