THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):
  (https://youtu.be/9KKbkR9mXYE)
- Link slides (dạng .pdf đặt trên Github):
  (https://github.com/7zomke/CS2205.MAR2024)
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

• Họ và Tên: Nguyễn Tấn Kiệt • MSSV: 230202009 	• Lớp: CS2205.CH181 • Tự đánh giá (điểm tổng kết môn): 7.5/10 • Số buổi vắng: 0 • Số câu hỏi QT cá nhân: 0 • Link Github: https://github.com/7zomke/CS2205.MAR2024

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

PHÂN TÍCH HÀNH VI NGƯỜI DÙNG VÀ THỰC THỂ VỚI MẠNG NƠ-RON TỰ TỔ CHỨC TĂNG CƯỜNG TRÊN NHẬT KÝ HỆ THỐNG

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

USER AND ENTITY BEHAVIOR ANALYTICS WITH SELF-ORGANIZING INCREMENTAL NEURAL NETWORK ON SYSTEM LOG

TÓM TẮT *(Tối đa 400 từ)*

Phân tích hành vi người dùng và thực thể là một phần quan trọng trong việc bảo vệ an ninh mạng nhằm phát hiện sớm các hành vi bất thường và nhận biết các nguy cơ tiềm ẩn từ người dùng và thực thể trong mạng doanh nghiệp. Các phương pháp tiếp cận truyền thống thường xuyên dựa vào các phương pháp thống kê hoặc theo quy tắc cố định và phải nhóm vận hành an ninh mạng phải luôn tự điều chỉnh liên tục để phù hợp cho môi trường doanh nghiệp, điều này hạn chế trong việc phát hiện các rủi ro bảo mật mới và tinh vi, những phương pháp này cũng có thể tạo ra nhiều cảnh báo sai, gây quá tải cho nhóm vận hành an ninh mạng và bỏ sót các mối đe dọa thực sự.

Vì những vấn đề trên, chúng tôi đề xuất một giải pháp bảo mật mới, sử dụng mô hình mạng nơ-ron tự tổ chức tăng cường . Giải pháp này cung cấp khả năng tự động thích ứng với sự biến đổi trong hành vi của người dùng và thực thể theo thời gian và có khả năng học liên tục, cho phép phân tích dữ liệu hoạt động của người dùng trong thời gian thực. Ngoài ra, còn có thể xử lý các tập dữ liệu lớn về hành vi của người dùng và thực thể, giúp tổ chức phát hiện và xử lý các mối đe dọa một cách toàn diện và hiệu quả.

Với giải pháp này, doanh nghiệp có thể nâng cao khả năng phát hiện các mối đe dọa mới và tinh vi, tự động hóa quy trình phân tích hành vi , tiết kiệm thời gian và nguồn lực, phản ứng nhanh chóng và hiệu quả trước các mối đe dọa tiềm ẩn trong môi trường doanh nghiệp.

GIỚI THIỆU *(Tối đa 1 trang A4)*

An ninh mạng đang trở thành mối quan tâm hàng đầu cho các tổ chức trong mọi quy mô. Các cuộc tấn công mạng ngày càng gia tăng về số lượng và mức độ tinh vi, đe dọa nghiêm trọng đến dữ liệu và hệ thống của doanh nghiệp. Trong đó Người dùng nội bộ trở thành mối đe dọa đáng kể, chiếm 22% tổng số vụ vi phạm dữ liệu (theo Cybersecurity Ventures), trong bảng báo cáo DBIR của Verizon, yếu tố con người có

liên quan chiếm 68% trong các vụ rò rỉ an ninh mạng được báo cáo trong năm 2023. Điều này cho thấy yếu tố con người đóng vai trò quan trọng trong các vi phạm an ninh mạng và cần phải được chú ý đặc biệt . Các hành động của người dùng này thường có xu hướng khác biệt so với người dùng nội bộ thông thường có thể là tải hàng trăm tài liệu nội bộ trong một ngày trong khi với người thường chỉ khoản 5 -10 lần. Do có quyền truy cập hợp pháp, họ có thể dễ dàng tương tác với các thực thể trong hệ thống doanh nghiệp, thực hiện các hành vi dễ gây hại hơn so với những kẻ tấn công bên ngoài. Các phương pháp bảo mật truyền thống như SIEM và DLP dựa trên quy tắc và chữ ký, thường không hiệu quả trong việc phát hiện các mối đe dọa mới và tinh vi.
User and Entity Behavior Analytics (UEBA) là một khái niệm phát triển từ cụm từ User Behavior Analytics (UBA) được Gartner đề cập lần đầu vào năm 2017, UEBA tập trung vào việc phân tích hành vi của người dùng và các thực thể (như máy chủ, ứng dụng, thiết bị di động) để phát hiện các hoạt động bất thường và nguy cơ tiềm ẩn. Bằng việc sử dụng các thuật toán học máy và trí tuệ nhân tạo để phát hiện các hành vi bất thường mà không cần dựa vào các quy tắc cố định. AI giúp UEBA tự động điều chỉnh và cải thiện khả năng phát hiện mối đe dọa theo thời gian. Các mô hình học máy có thể xử lý một lượng lớn dữ liệu và tìm ra các mẫu phức tạp mà con người hoặc các hệ thống dựa trên quy tắc khó có thể nhận ra. Điểm khác biệt chính của hệ thống so với các phương pháp bảo mật truyền thống dựa trên quy tắc là khả năng học hỏi và thích nghi liên tục. Thay vì dựa vào các quy tắc cố định có thể bị lừa bởi những kẻ tấn công tinh vi, UEBA sử dụng các thuật toán ML để tự động xây dựng mô hình hành vi bình thường cho mỗi người dùng và thực thể bằng cách theo dõi hành vi theo thời gian và so sánh nó với mô hình đã học.

Với mô hình mạng tự tổ chức tăng cường (SOINN) cung cấp khả năng phân tích hành vi người dùng và thực thể theo thời gian thực giúp UEBA nhanh chóng nhận biết các mẫu hành vi bất thường và đưa ra cảnh báo kịp thời về các mối đe dọa tiềm ẩn. Khả năng học tập liên tục của mô hình cho phép không ngừng cập nhật và hiệu quả trong việc phát hiện các mối đe dọa mới và phức tạp, đặc biệt trong môi trường mạng đang thay đổi liên tục. Mô hình cũng đóng vai trò trong việc xây dựng các hồ sơ hành vi chi tiết cho người dùng và thực thể, từ đó phát hiện các thay đổi nhỏ nhất trong hành vi một cách chính xác. Bằng cách giảm thiểu báo động giả và tập trung vào các mối đe dọa thực sự, SOINN giúp cải thiện hiệu quả công việc của nhóm an ninh mạng và giảm bớt sự mệt mỏi do xử lý quá nhiều cảnh báo không cần thiết.

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

- Có góc nhìn sâu hơn về hành vi người dùng trong hệ thống doanh nghiệp
- Tăng cường khả năng giám sát và phát hiện các hành động bất thường thời gian

thực

- Cảnh báo bảo mật với độ chính xác cao, giảm thiểu số lượng cảnh báo giả

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Nội dung :

- Thu thập dữ liệu người dùng từ nhiều hệ thống khác nhau và trích xuất đưa dữ các dữ liệu người dùng về một dạng chung
- Nghiên cứu các thuật toán SOINN, và biến thể của SOINN như : T-SOINN[1], RF-SOINN [2] để tìm ra thuật toán phù hợp nhất cho việc phân tích hành vi người dùng
- Huấn luyện các mô hình trên bằng việc sử dụng bộ dữ liệu CLUE-LDS [2] gồm thông tin nhật ký của hơn 5000 user được thu thập trong vòng 5 năm từ ngày 07-07-2017 đến 29-09-2022 (1910 ngày) để mô hình có thể học từ tập dữ liệu này
- Xây dựng thành một ứng dụng mô đun có thể tích hợp vào hệ thống SIEM

Phương pháp :

- Tìm hiểu và ứng kỹ thuật phân tích dữ liệu phân cụm để nhóm các nhóm người dùng và thực thể theo hành vi vào mô hình
- Tìm hiểu và ứng dụng phương pháp học tăng tiến cho phép mô hình học tập và thích nghi với dữ liệu mới vào mô hình
- Tìm hiểu các cấu trúc và mô hình thuật toán SOINN và các biến thể của SOINN : T-SOINN[1], RF-SOINN [2]
- Tìm hiểu đánh giá mô hình bằng độ đo Silhouette
- Huấn luyện các mô hình chạy trên bộ dữ liệu sử dụng so sánh và đánh giá kết quả dựa trên độ đo Silhouette
- Xây dựng, triển khai tích hợp vào hệ thống quản lý sự kiện và thông tin mã nguồn mở

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

- Có góc nhìn thực tế về hành vi của người dùng và thực thể trong hệ thống mạng doanh nghiệp
- Báo cáo các phương pháp và kỹ thuật của các thuật toán SOINN được sử dụng trong bài toán phân tích hành vi người dùng và thực thể.
- Mô hình đánh giá, phân tích người dùng và thực thể thời gian thực đạt hiệu suất cao
- Tích hợp vào hệ thống quản lý sự kiện và thông tin mã nguồn mở ELK

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1]. Ren, Jing, et al. "SOINN Intrusion Detection Model Based on Three-Way Attribute Reduction." *Electronics* 12.24 (2023): 5023.

[2]Voltan, Gabriele, Gian Luca Foresti, and Marino Miculan. "Pairing an Autoencoder and a SF-SOINN for Implementing an Intrusion Detection System." *CEUR WORKSHOP PROCEEDINGS*. Vol. 3486. CEUR-WS, 2023.

[3]. Landauer, Max, et al. "A User and Entity Behavior Analytics Log Data Set for Anomaly Detection in Cloud Computing." *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022.