Research Project

Tổng Quan

I - Phát hiện xâm nhập (IDS – Intrusion Detection System)

II - AI trong IDS – Ứng dụng học máy, học sâu

I - Phát hiện xâm nhập (IDS – Intrusion Detection System)

Khái Niệm:

IDS (Intrusion detection system) hoặc là hệ thống phát hiện xâm nhập, rà sát mạng hoặc những hệ thống (phần mềm) để tìm những hành vi độc hại hoặc những vi phạm điều lệ (policy violation) sau đó thông báo cho các administrator khi hành vi khả nghi được phát hiện.

Phân loại IDS theo nguồn dữ liệu:

NIDS (Network-based IDS)

HIDS (Host-based IDS)

PIDS (Provenance-based IDS)

NIDS (Network-based IDS)

Nguồn dữ liệu là lưu lượng mạng giữa các máy chủ. Hiệu quả trong việc đối phó với các cuộc tấn công đa máy chủ quy mô lớn như Distributed Denial-of-Service (DDoS).

Ví dụ: *Trong mạng* ***IoT chăm sóc sức khỏe****,* ***NIDS giám sát*** *luồng dữ liệu giữa các thiết bị đeo (wearable devices) của bệnh nhân và máy chủ bệnh viện để phát hiện hoạt động truy cập hoặc truyền dữ liệu bất thường.*

HIDS (Host-based IDS)

Nguồn dữ liệu là các sự kiện hệ thống bên trong máy chủ, như thay đổi hệ thống tệp và hoạt động tiến trình.

Ví dụ: *Một* ***thiết bị IoT y tế*** *có* ***HIDS*** *cài trong hệ điều hành của nó. Khi có một tiến trình lạ cố gắng ghi đè lên file hệ thống hoặc mở cổng mạng ngầm,* ***HIDS*** *sẽ* ***cảnh báo*** *có thể* ***mã độc đang chạy trong thiết bị.***

PIDS (Provenance-based IDS)

Một loại HIDS sử dụng dữ liệu nguồn gốc (data provenance).

Ví dụ:*Trong hệ thống* ***IoT chuỗi cung ứng, PIDS theo dõi nguồn gốc của mỗi gói dữ liệu từ cảm biến hàng hóa****.*
Nếu phát hiện một gói dữ liệu về nhiệt độ hàng ***bị thay đổi bởi*** *một thiết bị không có quyền,* ***PIDS*** *sẽ* ***phát hiện và ngăn chặn hành vi giả mạo dữ liệu*** *(****data tampering****)*

Tình trạng những hệ thống IDS trong IoT hiện nay.

- Tài nguyên trên thiết bị IoT có hạn tạo lên khó khăn trong việc áp dụng những hệ thống IDS phức tạp.

- Không có sự tiêu chuẩn hóa cho những thiết bị IoT: Mỗi thiết bị IoT không được tiêu chuẩn hóa cho hệ điều hành, phụ kiện nên mỗi thiết bị IoT đòi hỏi một hệ thống IDS riêng biệt.

Tình trạng những hệ thống IDS trong IoT hiện nay.

- Khả năng mở rộng khó khăn: Số lượng những thiết bị IoT cũ và mới rất lớn tạo lên khó khăn cho những hệ thống trung tâm IDS xử và quản lý.

- Những cuộc tấn công không-ngày và đang phát triển: Những hệ thống IDS truyền thống rất yếu trong việc phát hiện những cuộc tấn công thể loại mới, không biết hoặc những cuộc tấn công không-ngày. Trong đó số lượng những cuộc tấn công này ngày càng tăng.

II - AI trong hệ thống IDS (Intrusion detection system).

Sử dụng trí tuệ nhân tạo (AI) và mô hình học sâu (deep learning) trong những hệ thống IDS đã thay đổi cốt lõi của những hệ thống IDS từ việc tìm kiếm những nguy cơ đã được biết đến tới việc xác nhận những cuộc tấn công phức tạp chưa được biết đến.

Phân loại theo Phương pháp Phát hiện

A. Phát hiện Dựa trên Chữ ký (Signature-based Detection)

Nguyên tắc: So sánh lưu lượng/hoạt động hiện tại với một cơ sở dữ liệu khổng lồ chứa các mẫu (chữ ký) của các cuộc tấn công đã biết (ví dụ: chuỗi byte cụ thể của một virus, hoặc chuỗi lệnh tấn công).

Ưu điểm: Độ chính xác cao và rất nhanh trong việc phát hiện các tấn công phổ biến.

Nhược điểm: Hoàn toàn không thể phát hiện các cuộc tấn công mới (zero-day) hoặc các biến thể tấn công đã được sửa đổi.

Phân loại theo Phương pháp Phát hiện

B. Phát hiện Dựa trên Bất thường (Anomaly-based Detection)

Nguyên tắc: Sử dụng AI/Học máy để xây dựng một mô hình hành vi cơ sở (baseline) của hệ thống hoặc mạng khi nó hoạt động bình thường.
Hoạt động: Bất kỳ hoạt động nào lệch quá xa so với mô hình cơ sở này đều được gắn cờ là bất thường (anomalous).
Ưu điểm: Có khả năng phát hiện các cuộc tấn công zero-day và các hành vi độc hại mới.
Nhược điểm: Tỷ lệ cảnh báo sai (False Positive) cao hơn, vì đôi khi hành vi người dùng hợp lệ nhưng mới cũng có thể bị xem là bất thường.

Phân loại theo Phương pháp Phát hiện

C. Phát hiện Dựa trên Trạng thái Giao thức (Stateful Protocol Analysis)

Nguyên tắc: Tạo ra một mô hình được chấp nhận của các giao thức mạng (ví dụ: TCP, HTTP).
Hoạt động: Giám sát các gói tin để đảm bảo chúng tuân thủ trạng thái và luật lệ của giao thức.
Mục đích: Phát hiện các tấn công cố gắng vi phạm cấu trúc hoặc logic của giao thức, chẳng hạn như tấn công tràn bộ đệm (buffer overflow) hoặc thiết lập phiên TCP bất thường.

Ưu điểm của áp dụng AI so với phương pháp truyền thống.

- Phát hiện không-ngày: Khả năng tìm lạ thường của AI cho phép sự xác nhận của những cuộc tấn công chưa từng được thấy bao giờ.

- Ít thông báo dương tính sai lệch hơn so với những hệ thống IDS truyền thống.

- Phân tích thời gian thực: AI có thể phân tích và xử lý thông tin nhanh hơn nhiều lần so với con người, cho phép sự phát hiện và xử lý gần như tức thời.

Ưu điểm của áp dụng AI so với phương pháp truyền thống.

- Khả năng tiến hóa: Những mô hình AI có thể liên tục học hỏi và phát triển khi mối đe dọa phát triển, giữ vững độ hiệu quả mà không cần đến sự duy trì liên tục bởi con người.

- Khả năng tự động thực thi: Những hệ thống IDS sử dụng AI có thể tự động thi hành những quá trình được lập sẵn để xử lý cuộc tấn công đã được phát hiện bằng cách chặn địa chỉ IP độc hại hoặc cách ly một host đã bị xâm lược. Tối thiểu hóa thiệt hại được gây ra.

Nhược điểm của áp dụng AI so với phương pháp truyền thống.

- Sự cần thiết của một lượng lớn dữ liệu chất lượng tốt cho việc huấn luyện mô hình AI: Dữ liệu sai lệch hoặc chất lượng kém có thể dẫn đến mô hình AI có độ hiệu quả, độ chính xác thấp hoặc không thể dự đoán những cuộc tấn công nhất định.

- Những người tấn công có thể đưa vào AI dữ liệu độc hại khi trong quá trình huấn luyện. Hoặc đưa ra những hướng dẫn được tạo ra để tránh sự phát hiện của AI.

Nhược điểm của áp dụng AI so với phương pháp truyền thống.

- Phức tạp và khó giải thích: AI có thể được coi là hộp đen chứa thông tin trong dạng con người không thể hiểu được khiến cho độ phức tạp trong việc xác nhận và sửa lỗi do AI gây ra khó khăn và phức tạp.

- Sự đòi hỏi tài nguyên máy tính: Sự huấn luyện của những mô hình AI đòi hỏi một lượng lớn tài nguyên máy tính (GPU hoặc graphics processing unit) làm tăng chi phí một cách đáng kể.

Thank You