TRIỂN KHAI HONEYPOT KẾT HỢP VỚI DEEP LEARNING ĐỂ MÔ PHỎNG HÀNH VI MÁY CHỦ

Nguyễn Hoàng Thảo Quyên - 230202014

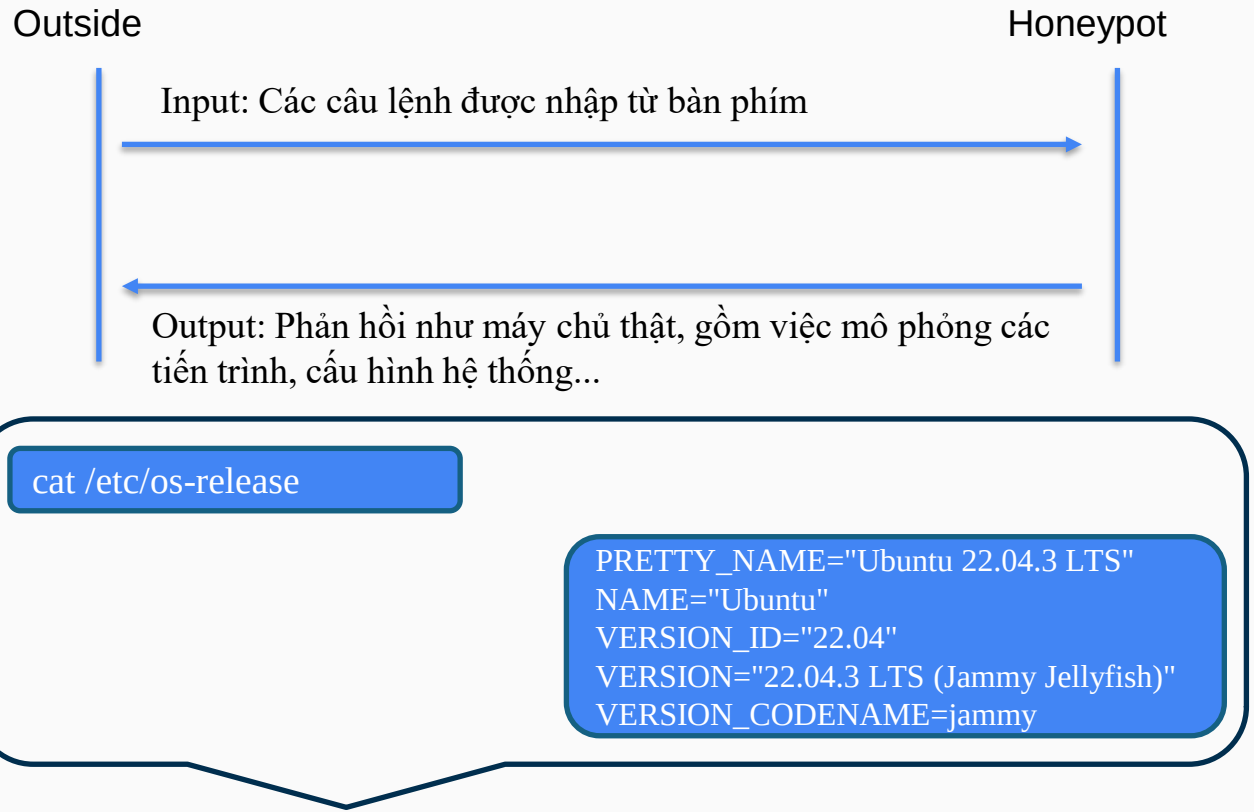
Tóm tắt

- Lớp: CS2205.CH181
- Link Github: https://github.com/quyen66/CS2205.CH181_PPLNCKH
- Link YouTube video: https://youtu.be/W6MpErFaMc4
- Ảnh + Họ và Tên: Nguyễn Hoàng Thảo Quyên

Giới thiệu

- Honeypot là hệ thống mồi nhử cố ý cài đặt không an toàn, được sử dụng để phát hiện và cảnh báo về hoạt động độc hại của kẻ tấn công.
- HoneyPots tương tác cấp thấp là loại đơn giản nhất trong các loại honeypot và không mang lại nhiều nguy cơ tiềm ẩn như các loại tương tác cấp cao nhưng chúng lại dễ bị phát hiện bởi nhiều đặc trưng.
- Nghiên cứu này sẽ đề xuất phương pháp để xây dựng honeypot thông minh có khả năng mô phỏng hành vi máy chủ chân thực từ cách giao tiếp đến phản ứng với các câu lệnh nhận được.

Giới thiệu


Outside
Honeypot
Input: Các câu lệnh được nhập từ bàn phím
Output: Phản hồi như máy chủ thật, gồm việc mô phỏng các tiến trình, cấu hình hệ thống...
cat /etc/os-release
PRETTY_NAME="Ubuntu 22.04.3 LTS"
NAME="Ubuntu"
VERSION_ID="22.04"
VERSION="22.04.3 LTS (Jammy Jellyfish)"
VERSION_CODENAME=jammy

Mục tiêu

- Xây dựng một mô hình mô phỏng hành vi cho honeypot.
- Sử dụng dữ liệu từ nhiều nguồn khác nhau để huấn luyện mô hình máy học này, giúp tăng tính chân thực và hiệu quả của honeypot.

Phạm vi và đối tượng

- Dữ liệu được thu thập từ máy chủ sẽ bao gồm cấu hình hệ thống, các tiến trình của các máy chủ thật, các dữ liệu từ syslog để biết được các phản hồi của máy chủ thực đến các câu lệnh nhận vào.
- Máy chủ honeypot sẽ được giới hạn là một máy chủ web nên sẽ nhận thêm các dữ liệu về gói tin mạng như HTTP, HTTPS, DNS.
- Sử dụng thuật toán GAN và RNN để tạo mô hình và sử dụng trong đề tài này.

Nội dung và Phương pháp

- Tổng hợp tất cả các dữ liệu thành ba bộ chính để huấn luyện, kiểm tra và xác thực.
- Nghiên cứu các thuật toán học sâu GAN VÀ RNN, xây dựng mô hình học máy, huấn luyện mô hình để xác định độ đo của từng mô hình trên tập dữ liệu.
- Chọn dùng các bộ dữ liệu như Syslog Dataset, HTTP Archive,… để huấn luyện cấu hình hệ thống và các bộ dữ liệu như Apache HTTP Server Logs, Nginx HTTP Server Logs, … để huấn luyện về các gói tin mạng.
- Huấn luyện mô hình theo các bộ dữ liệu trên và đánh giá hệ thống.
- Xây dựng honeypot và chạy thử trên môi trường thực tế.

Nội dung và Phương pháp

- Nghiên cứu các thuật toán tự sinh câu lệnh và phản hồi của máy Linux như BART-GAN, CodeGAN, LLM-GAN, GPT-3, Linux Command Response Prediction, GAN-based Linux Command Generation …
- Huấn luyện mô hình học sâu đã nghiên cứu, đánh giá kết quả.
- Dựng honeypot trong môi trường thử nghiệm và tìm cách áp dụng mô hình đã huấn luyện vào.

Kết quả dự kiến

- Một mô hình học sâu có khả năng phản hồi lại các câu lệnh với mức độ chân thực đạt yêu cầu.
- Một honeypot áp dụng mô hình máy học đã huấn luyện với các thông số đạt yêu cầu có thể phân tích các loại tấn công theo mẫu tấn công và phản hồi dựa theo kết quả phân tích.
- Báo cáo các phương pháp và kỹ thuật của mô hình máy học đã phát triển được. Kết quả thực nghiệm, đánh giá, so sánh với các phương pháp khác.

Tài liệu tham khảo

[1] Siniosoglou, I., Efstathopoulos, G., Pliatsios, D., Moscholios, I. D., Sarigiannidis, A., Sakellari, G., … Sarigiannidis, P. (2020). NeuralPot: An Industrial Honeypot Implementation Based On Deep Neural Networks. 2020 IEEE Symposium on Computers and Communications (ISCC). doi:10.1109/iscc50000.2020.9219712

[2] Fan, W., Du, Z., Smith-Creasey, M., & Fernandez, D. (2019). HoneyDOC: An Efficient Honeypot Architecture Enabling All-Round Design. IEEE Journal on Selected Areas in Communications, 1–1. doi:10.1109/jsac.2019.2894307

[3] Jiang, K., & Zheng, H. (2020). Design and Implementation of A Machine Learning Enhanced Web Honeypot System. 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). doi:10.1109/cisp-bmei51763.2020.9263640

[4] Mehta, V., Bahadur, P., Kapoor, M., Singh, P., & Rajpoot, S. (2015). Threat prediction using honeypot and machine learning. 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE). doi:10.1109/ablaze.2015.7155011

[5] Matin, I. M. M., & Rahardjo, B. (2020). The Use of Honeypot in Machine Learning Based on Malware Detection: A Review. 2020 8th International Conference on Cyber and IT Service Management (CITSM). doi:10.1109/citsm50537.2020.9268794