THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút): https://youtu.be/4fKUBxvE3Mg
- Link slides (dạng .pdf đặt trên Github của nhóm): https://github.com/ltd-uit/CS2205.CH190/blob/main/%C4%90%E1%BA%A1t%20L%C3%AA%20Th%C3%A0nh%20-%20CS2205.FEB2025.DeCuong.FinalReport.Template.Slide.pdf

- Họ và Tên: Lê Thành Đạt
- MSHV: 240101041

- Lớp: CS2205.CH190
- Tự đánh giá (điểm tổng kết môn): 8.0/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 3
- Link Github: https://github.com/ltd-uit/CS2205.CH190

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

NGHIÊN CỨU VÀ XÂY DỰNG CÔNG CỤ KIỂM THỬ BẢO MẬT TỰ ĐỘNG SỬ DỤNG PHƯƠNG PHÁP HỌC TĂNG CƯỜNG

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

FULLY AUTOMATED PENETRATION TESTING TOOL USING DEEP REINFORCE LEARNING

TÓM TẮT *(Tối đa 400 từ)*

Trong bối cảnh xã hội hiện đại, sự phụ thuộc vào các hệ thống thông tin và mạng máy tính ngày càng gia tăng. Điều này đồng nghĩa với việc an ninh mạng trở thành một yếu tố sống còn khi mà các cuộc tấn công mạng ngày càng tinh vi và phức tạp. Kiểm thử xâm nhập (penetration testing) là một trong những phương pháp quan trọng để đánh giá và cải thiện an ninh của hệ thống. Tuy nhiên, phương pháp này đang phải đối mặt với những thách thức lớn như Sự gia tăng của các mối đe dọa mạng; Sự phức tạp của hệ thống công nghệ thông tin; Thiếu hụt nhân lực và kĩ năng …

Trước những thách thức này, kiểm thử xâm nhập tự động hóa trở thành một giải pháp thiết yếu. Kiểm thử tự động hóa giúp giảm tải công việc cho các chuyên gia an ninh, nâng cao độ chính xác và nhất quán trong việc phát hiện lỗ hổng, đồng thời cải thiện khả năng phòng thủ của hệ thống trước các cuộc tấn công mạng ngày càng phức tạp. Nhận thức được điều đó, nhóm nghiên cứu đã thực hiện khảo sát một số giải pháp và thấy rằng hầu hết các công trình mới chỉ dừng lại ở việc tạo một bộ khung tự động khai thác đơn giản, chưa phù hợp với thực tế. Điểm yếu của các nghiên cứu này chính là chưa tổng hợp thông tin chi tiết về lỗ hổng đang gặp trên mục tiêu và phụ thuộc quá nhiều vào các mẫu khai thác có sẵn mà không thể tự động sinh ra mẫu khai thác. Bên cạnh đó, sử dụng phương pháp lựa chọn mã khai thác một cách ngẫu nhiên, bộ khung nói trên gặp nhiều hạn chế và gây lãng phí tài nguyên và thời gian.

Từ những hạn chế đó, nhóm nghiên cứu đề xuất xây dựng một framework với sự hỗ

trợ từ mô hình học máy nhằm giúp tự động khai thác toàn diện thông qua cải thiện hiệu suất thu thập thông tin mục tiêu, tích hợp thêm các công cụ dò quét phát hiện lỗ hổng và nâng cao chất lượng các mẫu khai thác tự sinh.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Theo Báo cáo Rủi ro Toàn cầu năm 2023 của Diễn đàn Kinh tế Thế giới, đại dịch COVID-19 đã bắt buộc các quốc gia cần phải thúc đẩy việc áp dụng công nghệ vào mọi lĩnh vực của đời sống. Trong kỷ nguyên số hiện tại, sự phát triển nhanh chóng của công nghệ thông tin và mạng máy tính đã mang lại nhiều lợi ích nhưng cũng đặt ra không ít thách thức về an ninh mạng. Các tổ chức, từ doanh nghiệp đến các cơ quan chính phủ đều phải đối mặt với nguy cơ bị tấn công mạng, mất mát dữ liệu, và các thiệt hại nghiêm trọng khác. Trong bối cảnh này, kiểm thử xâm nhập (Penetration Testing) trở thành một công cụ không thể thiếu để đảm bảo an ninh và bảo mật hệ thống.

Kiểm thử thâm nhập là một loại kiểm tra bảo mật được thực hiện nhằm xác minh liệu một ứng dụng có được cài đặt bảo mật hay không thông qua việc phát hiện những lỗ hổng trong hệ thống. Các lỗ hổng có thể được kiểm tra và tìm ra trong mỗi quá trình triển khai hoặc vận hành hệ thống hoặc ứng dụng. Quan trọng nhất, kiểm thử thâm nhập tiến hành ở chế độ mũ trắng, trong điều kiện có sự đồng ý của chủ sở hữu hệ thống nhằm giả lập xâm nhập. Trong đề tài này, một cuộc trình diễn thử thâm nhập sẽ được tiến hành cùng với các dụng cụ và kỹ thuật.

Với tốc độ phát triển vượt bậc của công nghệ thì phương pháp kiểm thử thâm nhập truyền thống đã không còn phù hợp trong thời đại ngày nay. Chính vì thế, chúng tôi nhận thấy việc ứng dụng tự động hoá vào quá trình kiểm thử xâm nhập là một nhu cầu cấp thiết và cần phải xây dựng một công cụ có khả năng thực hiện quá trình này một cách tự động nhằm tiết kiệm tối đa chi phí về thời gian và tài nguyên, đáp ứng nhu cầu về nguồn nhân lực an toàn thông tin đang thiếu hụt hiện nay.

MỤC TIÊU *(Viết trong vòng 3 mục tiêu)*

● Nghiên cứu xây dựng công cụ kiểm thử bảo mật hệ thống có khả năng rà soát các lỗ hổng đang có và thông qua các lỗ hổng thực hiện tấn công xâm nhập sâu vào mạng nội bộ.

● Nghiên cứu phương pháp học tăng cường để tích hợp vào công cụ đã thiết kế nhằm tự động hóa quy trình khai thác cũng như tích lũy kinh nghiệm học thông qua mỗi lần thực thi. Cuối cùng, công cụ sẽ có khả năng tìm kiếm các lỗ hổng và thực hiện khai thác một cách chính xác, hiệu quả.

● Mở rộng khả năng khai thác các lỗ hổng phức tạp, yêu cầu nhiều bước khai thác hơn so với các công cụ đã có.

NỘI DUNG VÀ PHƯƠNG PHÁP

Nội dung, phương pháp nghiên cứu chính:

a. Nội dung 1: Tìm hiểu chức năng, cách thức hoạt động của một số công cụ hỗ trợ kiểm thử bảo mật phổ biến trong thực tế.

- *Mục tiêu:*
  - Nắm được chức năng, quy tắc hoạt động của một số công cụ.
  - Nắm được ngữ cảnh sử dụng công cụ.
- *Phương pháp:* Nghiên cứu khảo sát để biết những công cụ nào thường được dùng cho những mục đích nào. Sau đó tiến hành thử nghiệm.

b. Nội dung 2: Tìm hiểu và xây dựng công cụ kiểm thử bảo mật toàn diện sử dụng kết hợp các công cụ khác một cách có chiến lược.

- *Mục tiêu:*
  - Xây dựng thành công công cụ khai thác toàn diện.
  - Đảm bảo hiệu suất tốt nhất có thể cho quá trình khai thác lỗ hổng hệ thống.
- *Phương pháp:*
  - Xây dựng chiến lược khai thác hợp lý, tối ưu.
  - So sánh, lựa chọn những công cụ cho từng bước.
  - Kết hợp các công cụ lại với nhau theo chiến lược đã đề xuất.

c. **Nội dung 3**: Tìm hiểu về học tăng cường sâu (deep reinforcement learning) và tìm cách tích hợp vào công cụ đã xây dựng.

- *Mục tiêu:*
  - Nắm vững kiến thức về học tăng cường sâu, tổng quan về kiến trúc và các đặc điểm. Hiểu về các thành phần, cách hoạt động và các thư viện hỗ trợ.
  - Xây dựng mô hình học tăng cường sâu giúp tự động hóa các bước của công cụ kiểm thử bảo mật đã xây dựng sử dụng các thư viện như Keras, Tensorflow.
  - Đào tạo mô hình học tăng cường sâu để có thể tích lũy kinh nghiệm khai thác.
- *Phương pháp:*
  - Tham khảo cơ sở lý thuyết, thực nghiệm của nghiên cứu, các video mô tả, các ứng dụng liên quan đến học tăng cường sâu và phương pháp xây dựng các mô hình.
  - Tìm hiểu các tham số cần cho bài toán, tinh chỉnh và đào tạo để mô hình có hiệu quả tốt nhất.

d. **Nội dung 4**: Thực nghiệm và đánh giá kết quả.

- *Mục tiêu:*
  - Xây dựng được các kịch bản kiểm tra và đánh giá.
  - Đánh giá hiệu năng, độ chính xác và khả năng phát triển của hệ thống.
- *Phương pháp:*
  - Xây dựng mục tiêu là máy server chứa các lỗ hổng.
  - Thực hiện các kịch bản thử nghiệm khác nhau để đánh giá về hiệu năng, độ chính xác và hiệu quả của công cụ.
  - So sánh với các công cụ đã biết như DeepExploit, Shennina…

KẾT QUẢ MONG ĐỢI

- Có kiến thức khái quát tổng quan về cách thức hoạt động của một công cụ tự động kiểm thử dựa trên học tăng cường.

- Đề xuất được chiến lược thu thập thông tin và triển khai module dựa theo chiến lược.
- Xây dựng thành công module khai thác có khả năng đánh giá bảo mật mục tiêu và thực hiện các tấn công phức tạp.
- Triển khai thành công mô hình học tăng cường và tích hợp vào công cụ đã xây dựng.
- Có được kết quả thực nghiệm và đưa ra được báo cáo tổng quan về quá trình thực hiện đề tài.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1]. Shah, Mujahid, et al. "Penetration testing active reconnaissance phase–optimized port scanning with nmap tool." *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, 2019.

[2]. Li, Zegang, Qian Zhang, and Guangwen Yang. "EPPTA: Efficient partially observable reinforcement learning agent for penetration testing applications." *Engineering Reports* 7.1 (2025): e12818.

[3]. Ghanem, M. C., Chen, T. M., & Nepomuceno, E. G. (2023). Hierarchical reinforcement learning for efficient and effective automated penetration testing of large networks. *Journal of Intelligent Information Systems*, *60*(2), 281-303.

[4]. Ovidiu Valea, Ciprian Oprisa: Towards Pentesting Automation Using the Metasploit Framework. ICCP 2020: 171-178.

[5]. Fabio Massimo Zennaro, László Erdodi: Modelling penetration testing with reinforcement learning using capture-the-flag challenges: Trade-offs between model-free learning and a priori knowledge. IET Inf. Secur. 17(3): 441-457 (2023).

[6]. Tyler Cody: A Layered Reference Model for Penetration Testing with Reinforcement Learning and Attack Graphs. CoRR abs/2206.06934 (2022)

[7]. Li, Qianyu, et al. "An intelligent penetration testing method using human feedback." *IEEE Transactions on Industrial Informatics* (2024).