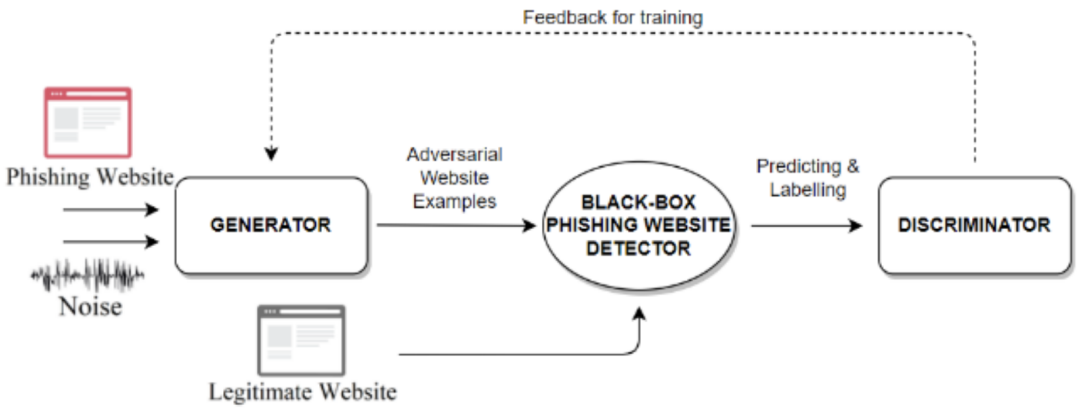
Họ và tên (IN HOA)	VĂN THIÊN LUÂN - CH2001008
Ảnh	
Số buổi vắng	2
Bonus	7
Tên đề tài (VN)	TĂNG CƯỜNG KHẢ NĂNG NHẬN DIỆN CỦA HỆ THỐNG PHÁT HIỆN XÂM NHẬP DỰA TRÊN MẠNG SINH ĐỐI KHÁNG
Tên đề tài (EN)	ENHANCE THE DETECTION CAPABILITIES OF INTRUSION DETECTION SYSTEM USING GAN
Giới thiệu	Với xu hướng ngày càng xuất hiện nhiều mối đe dọa bảo mật trên Internet, hệ thống phát hiện xâm nhập (Intrusion Detection System - IDS) trở thành công cụ thiết yếu để phát hiện và phòng tránh tấn công mạng được thể hiện dưới dạng lưu lượng mạng độc hại. IDS thực hiện giám sát lưu lượng mạng và đưa ra cảnh báo nếu lưu lượng không an toàn được xác định, phát

hiện bởi bộ phân tích lưu lượng. Mục đích chính của IDS là phân loại giữa bản ghi mạng bình thường và bất thường thông qua các dữ liệu mà hệ thống có được trước đó, hoặc thông qua các phương pháp dự đoán. Nhiều hệ thống phát hiện xâm nhập đã hiện thực các mô hình học máy, học sâu trong việc hỗ trợ nhận dạng, phân loại nguy cơ bảo mật hay dấu hiệu bị tấn công[1].

Phương pháp học máy (Machine Learning), học sâu (Deep Learning) là những kỹ thuật được áp dụng để phát triển các ứng dụng hiệu quả từ lượng dữ liệu thu thập được trong ngữ cảnh an toàn thông tin. Tuy nhiên, các giải pháp này đòi hỏi được huấn luyện với lượng dữ liệu lớn với nhiều loại hình khác nhau, trong khi việc chia sẻ các dữ liệu mạng như thế này lại gặp hạn chế do lo ngại về quyền riêng tư từ các nhân, tổ chức liên quan. Lúc này, mô hình mạng sinh đối kháng (Generative Adversarial Network - GAN) với khả năng ưu việc và tiên tiến của mình mang tiềm năng lớn trong việc liên tục phát sinh các mẫu mã độc mới bằng cách tái huấn luyện dựa trên các mẫu đã có[2][3][4]. Điều này giúp mở rộng bộ dữ liệu để huấn luyện IDS, giúp nâng cao khả năng phòng vệ và dự đoán trước các mẫu mã độc mới có thể được tạo ra[5][6].



Ứng dụng của GAN trong bài toán đánh lừa trình phát hiện lừa đảo (Nguồn: PWDGAN)

Input: Lưu lượng mạng

Output: Đưa ra kết quả dự đoán lưu lượng mạng bình thường hay độc hại

Mục tiêu	● Khảo sát các mô hình học máy được ứng dụng ở các bộ phát hiện xâm nhập (IDS) nhằm bảo vệ hạ tầng mạng. Khảo sát các tập dữ liệu huấn luyện dùng đánh giá IDS. ● Tìm hiểu và xây dựng mô hình mạng đối kháng sinh mẫu (GAN) để tự động sinh ra các dữ liệu mới có khả năng đánh lừa được bộ phát hiện xâm nhập (IDS). ● Thực nghiệm và đánh giá mức độ hiệu quả của mô hình triển khai dựa vào các tập dữ liệu lưu lượng mạng chuẩn và tự thu thập.
Nội dung và phương pháp thực hiện	**Nội dung 1: Nghiên cứu phương pháp học máy được ứng dụng ở các bộ phát hiện xâm nhập. Khảo sát các tập dữ liệu huấn luyện dùng đánh giá IDS** ● **Mục tiêu:** - Phân tích cơ chế giám sát, theo dõi lưu lượng mạng của bộ điều khiển trong việc hỗ trợ xây dựng giải pháp IDS. Từ đó, có cơ chế hỗ trợ giám sát, thu thập, chọn lựa, trích xuất các đặc trưng trong mạng dùng cho việc theo dõi trạng thái mạng tại một thời điểm. - Thực hiện mô hình hóa các thuộc tính trạng thái mạng với các tập dữ liệu phổ biến về các loại tấn công dùng cho các IDS. Tiếp đến, khảo sát một số giải pháp nổi bật ứng dụng các mô hình học máy, học sâu trong việc xây dựng các IDS. ● **Phương pháp:** - Nghiên cứu các tài liệu hướng dẫn và thực hiện triển khai môi trường, thực hiện bắt gói tin lưu lượng mạng. - Khảo sát các tập dữ liệu mẫu dành cho việc huấn luyện các IDS như CICIDS2017, CICIDS2018, NSL-KĐ, CAIDA, CTU-13, IoT-23,... - Khảo sát các IDS được nghiên cứu đã sử dụng các thuật toán máy học, hay các mô hình học sâu để nhận diện tấn công, xâm nhập.

	- Phân tích, đánh giá độ chính xác, hiệu quả khi triển khai. **Nội dung 2: Mô hình Mạng đối kháng sinh mẫu (GAN) và cơ chế phát sinh mẫu dữ liệu tấn công đối kháng dựa trên GAN** ● **Mục tiêu:** - Tìm hiểu mô hình mạng đối kháng sinh mẫu (GAN) và triển khai các biến thể của mô hình GAN phù hợp (CycleGAN, DCGAN, Wasserstein GAN,...) có thể ứng dụng trong lĩnh vực nhận biết dấu hiệu các cuộc tấn công mạng. - Chú ý tới yếu tố giữ nguyên các thuộc tính đặc trưng của lưu lượng tấn công, không làm vô hiệu hóa khả năng hoạt động của nó trong môi trường thực tế. - Nghiên cứu, áp dụng mô hình GAN mô phỏng cách hoạt động của IDS để phát sinh ra các dữ liệu tấn công đối kháng từ tập dữ liệu tấn công nhằm qua mặt các IDS. - Sử dụng tập dữ liệu đã khảo sát cho Generator / Discriminator trong mô hình GAN để sinh dữ liệu dị thường (dữ liệu mạng thuộc hành vi tấn công). - Kiểm soát, tối ưu hàm mất mát (loss function) trong quá trình huấn luyện Generator và Discriminator trong mô hình GAN phù hợp. - Xác định, định nghĩa các tiêu chí đánh giá độ hiệu quả của các mẫu đối kháng được sinh ra bởi mô hình GAN phù hợp. ● **Phương pháp:** - Khảo sát, thu thập tài liệu, đánh giá sự tương thích với dữ liệu tấn công mạng. - Thiết kế chi tiết hệ thống, xác định tiêu chí đánh giá khả năng sinh mẫu đối kháng. **Nội dung 3: Thực nghiệm và đánh giá kết quả**

	● **Mục tiêu:** - Xây dựng và triển khai mô-đun NIDS từ một số mô hình học máy; ứng dụng GAN để phát sinh mẫu lưu lượng tấn công qua mặt IDS dựa trên việc học liên tục cơ chế phát hiện IDS hộp đen. - Đánh giá được khả năng phát hiện tấn công của IDS ở các mẫu dữ liệu phát sinh từ các mô hình GAN, đồng thời xem xét hiệu năng của cơ chế trong môi trường mạng. - Quá trình phát sinh dữ liệu tấn công đối kháng, kiểm tra, tăng cường và tái triển khai được thực hiện xuyên suốt trong quá trình vận hành của hệ thống. ● **Phương pháp:** - Triển khai mô hình GAN trên tập dữ liệu huấn luyện IDS để tạo ra mẫu lưu lượng đối kháng mới nhắm vào IDS (giữ lại các thuộc tính bản chất đặc trưng của các loại tấn công) của mạng (sử dụng Python). - Xây dựng mô-đun NIDS từ mô hình đã thiết kế tương tác với bộ điều khiển. - Thực hiện các thử nghiệm khác nhau để đánh giá hiệu năng và độ chính xác, tỷ lệ phát hiện của IDS khi gặp mẫu dữ liệu đối kháng. - Phân tích và diễn giải số liệu thu được.
Kết quả dự kiến	● Triển khai hệ thống phát hiện xâm nhập trong hạ tầng mạng để phát hiện các lưu lượng mạng độc hại. Trong đó tích hợp GAN để tự động sinh ra các mẫu dữ liệu tấn công liên tục và cập nhật cho IDS để tăng cường khả năng phát hiện. ● Tập luật (rules) của IDS để phát hiện xâm nhập khi quá trình sinh mẫu đối kháng đạt trạng thái bão hòa.

Tài liệu tham khảo	[1] O. Ibitoye, R. Abou-Khamis, A. Matrawy e M. O. Shafiq, “The Threat of Adversarial Attacks on Machine Learning in Network Security : A Survey”, in arXiv:1911.02621, 2020. [2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz e Z. Wang, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”, in Computer Vision and Pattern Recognitio, 2017. [3] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang e Y.-H. and Yang, “ulti-track sequential generative adversarial networks for symbolic music generation and accompaniment”, in In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, 2018. [4] H. Su, X. Shen, P. Hu, W. Li e Y. and Chen, “Dialogue generation with gan”, in In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana:, 2018. [5] J.-Y. Kim, S.-J. Bu e S. and Cho, “Malware detection using deep transferred generative adversarial networks”, in In Proceedings of International Conference on Neural Information Processing, Guangzhou, China, 2017. [6] W. a. T. Y. Hu, “Generating adversarial malware examples for black-box attacks based on GAN”, in arXiv preprint, 2017.