THÔNG TIN CHUNG CỦA BÁO CÁO

• Họ và Tên: Nguyễn Thành Tâm • MSSV: 230201053 	• Lớp: CS2205.APR2023 • Tự đánh giá (điểm tổng kết môn): 6/10 • Số buổi vắng: 0 • Số câu hỏi QT cá nhân: 0 • Link Github: https://github.com/nguyentam88/TamNT-CS2205.APR2023 • Link Youtube: https://youtu.be/w2dThVfun5E

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

PHÁT HIỆN TẤN CÔNG DDOS BẰNG MẠNG ĐỐI NGHỊCH CHUNG VỚI BỘ PHÂN BIỆT ĐỐI XỬ KÉP (GANDD)

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

DETECTION OF ADVERSARIAL DDOS ATTACKS USING GENERATIVE ADVERSARIAL NETWORKS WITH DUAL DISCRIMINATORS (GANDD)

TÓM TẮT *(Tối đa 400 từ)*

Các cuộc tấn công DDoS làm cạn kiệt băng thông mạng và tài nguyên máy của hệ thống, ngăn hệ thống cung cấp cho người dùng các dịch vụ thông thường. Việc phát hiện các cuộc tấn công tấn công DDoS truyền thống gặp nhiều hạn chế như độ chính xác thấp, khả năng thích ứng kém với các loại tấn công mới được gọi là tấn công DDoS với lưu lượng tấn công đối nghịch. Vấn đề này đặt ra nhiều câu hỏi cần có câu trả lời như phương pháp nào có thể phát hiện và phóng chống cuộc tấn công mới, phương pháp mới sẽ sử dụng dữ liệu được thu thập từ đâu, xây dựng tập huấn luyện ra sao? Nhiều đề tài đã được nghiên cứu và phát triển để giải quyết nhưng những phương pháp này vẫn còn nhiều hạn chế cần cải tiến. Để tìm hướng giải quyết, thông qua đề tài này đề xuất một hướng đi mới là mạng đối nghịch chung (GAN) với Bộ phân biệt đối xử kép (GANDD) dựa trên cải tiến mô hình GAN truyền thống với 2 bộ xử lý kép được thiết kế bổ sung có khả năng học hỏi từ các mẫu dữ liệu phức tạp và phân biệt lưu lượng truy cập bình thường khỏi lưu lượng truy cập tấn công DDoS một cách hiệu quả. Trong đề tài này, tập trung nghiên cứu và phát triển hệ thống phát hiện tấn công DDoS dựa trên GANDD để đánh giá độ chính xác cao, khả năng thích ứng tốt với các loại tấn công mới và thời gian phản hồi nhanh của hệ thống phát hiện tấn công DDoS dựa trên GANDD.

GIỚI THIỆU *(Tối đa 1 trang A4)*

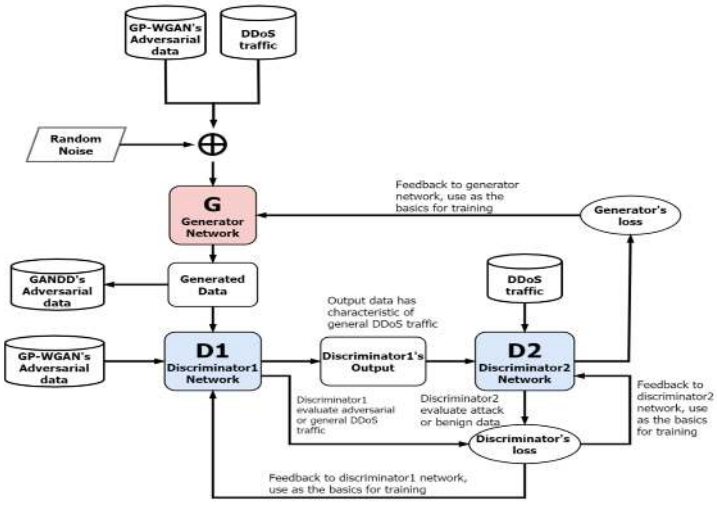
- Tấn công DDoS (Distributed Denial-of-Service) là một dạng tấn công mạng

nhằm làm quá tải hệ thống mạng bằng cách gửi một lượng lớn lưu lượng truy cập giả mạo, khiến hệ thống không thể đáp ứng các yêu cầu hợp pháp của người dùng

- Khác với tấn công DDoS truyền thống, tấn công DDoS với lưu lượng tấn công đối nghịch là một kỹ thuật sử dụng các gói tin giả mạo để đánh lừa hệ thống phòng chống DDoS. Các gói tin giả mạo này được gửi từ nhiều nguồn khác nhau, khiến hệ thống phòng chống DDoS khó có thể phân biệt được đâu là lưu lượng truy cập hợp pháp và đâu là lưu lượng truy cập tấn công.

- GAN (Generative Adversarial Networks) [1] là một phương pháp học máy tiên tiến có thể học hỏi các mẫu dữ liệu phức tạp của lưu lượng tấn công đối nghịch, giúp phân biệt nó với lưu lượng truy cập hợp pháp một cách hiệu quả. Tuy nhiên, GAN phụ thuộc vào chất lượng và độ đa dạng của dữ liệu huấn luyện và quá trình huấn luyện mô hình GAN có thể tốn thời gian.

- Trong đè tài này sẽ tập trung nghiên cứu về GANDD (Generative Adversarial Networks with Double Discriminators) hay còn được gọi là Mạng đối nghịch chung với bộ phân biệt đối xử kép (GANDD) là một cải tiến dựa trên Mạng đối nghịch chung (GAN) như Hình 1 mô tả.



Hình 1: kiến trúc mô hình GANDD

- Trong kiến trúc GANDD [3] được thiết kế gồm hai mạng đối nghịch: mạng tạo (generator) tạo ra lưu lượng truy cập tấn công DDoS giả nhằm đánh lừa mạng phân

biệt và mạng phân biệt (discriminator) được chia thành 2 bộ phân biệt đối xử kép discriminator1 giúp phân biệt giữa lưu lượng truy cập bình thường và lưu lượng truy cập tấn công DDoS thực tế với discriminator2 có chức năng phân biệt giữa lưu lượng truy cập giả do mạng tạo ra và lưu lượng truy cập thực. Thông qua đó hai mạng đối nghịch liên tục học hỏi và cải thiện khả năng phân biệt DDoS.

MỤC TIÊU

- Nghiên cứu và phát triển mô hình GANDD phù hợp cho việc phát hiện tấn công DDoS trong hệ thống mạng.
- Đánh giá hiệu quả của mô hình GANDD trong việc phân biệt lưu lượng tấn công DDoS với lưu lượng bình thường.
- So sánh hiệu quả của mô hình GANDD với các phương pháp phát hiện tấn công DDoS truyền thống.

NỘI DUNG VÀ PHƯƠNG PHÁP

Nội dung :

- Tìm hiểu các phương pháp phát hiện tấn công DDoS truyền thống: Phân tích ưu và nhược điểm của các phương pháp phát hiện tấn công DDoS hiện có theo phát hiện dựa trên quy tắc, phát hiện dựa trên thống kê và phát hiện dựa trên máy học.
- Tìm hiểu mô hình GANDD: mô tả kiến trúc và nguyên tắc hoạt động của mô hình GANDD, bao gồm Generator (Bộ tạo), Discriminator 1 (Bộ phân biệt thứ nhất) và Discriminator 2 (Bộ phân biệt thứ hai).
- So sánh GANDD với các biến thể khác của mô hình GAN: GAN. GAN Wasserstein (Wasserstein GAN) và GP-WGAN.
- Ứng dụng GANDD vào phát hiện tấn công DDoS: Mô tả cách thức ứng dụng GANDD để phát hiện tấn công DDoS, thiết kế kiến trúc mô hình GAND, thiết kế **và** quá trình huấn luyện mô hình nhằm phân biệt lưu lượng tấn công và lưu lượng bình thường, và cách thức xử lý các cảnh báo tấn công.
- Đánh giá hiệu quả của GANDD: So sánh hiệu quả phát hiện tấn công DDoS

của GANDD với các phương pháp truyền thống thông qua các chỉ số như tỷ lệ phát hiện thực tế (TPR), tỷ lệ báo động giả (FPR) và thời gian phản ứng.

- Phân tích thách thức và giải pháp: Xác định các thách thức khi ứng dụng GANDD vào phát hiện tấn công DDoS, chẳng hạn như yêu cầu dữ liệu huấn luyện, khả năng thích ứng với các cuộc tấn công mới và tính toán. Đề xuất các giải pháp để giải quyết các thách thức này.

Phương Pháp:

- Thu thập dữ liệu: Thu thập dữ liệu lưu lượng truy cập mạng và dữ liệu tấn công DDoS từ các nguồn khác nhau như bộ thu thập lưu lượng truy cập mạng (NTC),các phần mềm giám sát hệ thống mạng hoặc log Firewall.
- Xử lý dữ liệu: sử dụng các kỹ thuật lấy mẫu như Kỹ thuật lấy mẫu. Kỹ thuật cân bằng dữ liệu, Kỹ thuật giảm nhiễu nhằm Loại bỏ dữ liệu lỗi, dữ liệu thiếu sót và dữ liệu bất thường, chuyển đổi dữ liệu sang định dạng phù hợp đảm bảo các thuộc tính dữ liệu có cùng thang đo và phân phối cho việc huấn luyện mô hình GANDD.
- Huấn luyện mô hình GANDD: Thiết lập cấu trúc mô hình, sử dụng thuật toán PReLU để cải thiện hiệu suất phát hiện tấn công DDoS. PReLU được áp dụng cho các lớp ẩn trong mạng tạo và các bộ phân biệt của GANDD. Việc sử dụng PReLU giúp GANDD học hỏi các mẫu dữ liệu phức tạp hơn và phân biệt chính xác hơn giữa lưu lượng truy cập bình thường và lưu lượng truy cập tấn công DDoS.
- Đánh giá mô hình GANDD: Sử dụng dữ liệu đánh giá để đánh giá hiệu quả phát hiện tấn công DDoS của mô hình GANDD, bao gồm tính chính xác, độ nhạy và độ đặc trưng.
- Triển khai mô hình GANDD: Triển khai mô hình GANDD đã được huấn luyện và điều chỉnh vào hệ thống phát hiện tấn công DDoS thực tế.

KẾT QUẢ MONG ĐỢI

Việc ứng dụng GANDD (Generative Adversarial Networks with Dual Discriminators)

vào phát hiện tấn công DDoS mang lại nhiều kết quả mong đợi bao gồm:

- Phát triển nhiều mô hình GANDD có khả năng phát hiện tấn công DDoS hiệu quả cao.
- Đánh giá chi tiết hiệu quả của mô hình GANDD trong việc phân biệt lưu lượng tấn công DDoS với lưu lượng bình thường.
- So sánh hiệu quả của mô hình GANDD với các phương pháp phát hiện tấn công DDoS truyền thống.
- Xác định các thách thức và đề xuất giải pháp khi ứng dụng.

1]. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 2, 2672–2680.

[2]. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of Wasserstein GANs. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; 5769–5779

[3]. Nguyen, T.D.; Le, T.; Vu, H.; Phung, D. Dual Discriminator Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; 2667–2677.

[4]. Zhang, X.; Zhao, Y.; Zhang, H. Dual-discriminator GAN: A GAN way of profile face recognition. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 27–29 June 2020; 162–166.