TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Nghiên cứu, triển khai hệ thống phát hiện bất thường dựa trên nghiên cứu dữ liệu lưu lượng mạng của hệ thống máy chủ tên miền (DNS)

DƯƠNG VĂN THANH
thanh.dv194174@sis.hust.edu.vn

Ngành: Công nghệ thông tin

Giảng viên hướng dẫn: TS. Vũ Tuyết Trinh _____

Chữ kí GVHD

Khoa: Khoa học máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 06/2024

LỜI CẢM ƠN

Lời đầu tiên, tôi muốn bày tỏ lòng tri ân đến các thầy cô giảng viên của trường Công Nghệ Thông Tin và Truyền Thông vì đã tận tình hướng dẫn, hỗ trợ và truyền đạt những kiến thức quý báu trong suốt quá trình học tập, cũng như toàn bộ giảng viên, nhân viên, viên chức của Đại học Bách Khoa Hà Nội vì đã tạo ra một môi trường học tập và rèn luyện tuyệt vời, nơi em không chỉ trau dồi kiến thức, phát triển kỹ năng mà còn được gặp gỡ những người thầy cô tận tâm, những người bạn đồng hành tuyệt vời, trải qua những thử thách góp phần tạo nên con người tôi hôm nay.

Đặc biệt, xin được gửi lời cảm ơn đến cô Vũ Tuyết Trinh và anh Trần Cảnh Toàn đã tận tình hướng dẫn tôi trong suốt quá trình thực hiện đồ án. Tôi cũng xin chân thành cảm ơn Trung tâm Internet Việt Nam và các anh chị phòng KTH - VNNIC đã cung cấp tài liệu, dữ liệu thực nghiệm và hỗ trợ kỹ thuật, tạo điều kiện thuận lợi để tôi triển khai và hoàn thiện đồ án này.

Cuối cùng, tôi xin bày tỏ lòng biết ơn đến gia đình và bạn bè, những người luôn động viên, cổ vũ tinh thần và đồng hành cùng tôi trong suốt quá trình học tập và nghiên cứu. Tôi hy vọng rằng kết quả nghiên cứu trong đồ án sẽ đóng góp một phần nhỏ vào sự phát triển của các phương pháp phát hiện bất thường trong hệ thống máy chủ tên miền DNS. Do hạn chế về năng lực bản thân, trong quá trình nghiên cứu không tránh khỏi việc mắc phải nhiều thiếu sót, tôi mong nhận được sự góp ý quý báu từ thầy cô và các bạn để hoàn thiện hơn.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Với chức năng chuyển chuyển đổi tên miền sang địa chỉ IP, DNS được ví như trái tim của Internet khi mà hầu hết các hoạt động trên Internet đều bắt đầu từ truy vấn DNS. Cũng vì lý do đó mà DNS đã trở thành một mục tiêu phổ biến của các cuộc tấn công mạng. DNS được cấu trúc theo dạng phân cấp, với tầng thứ hai là các nhà quản lý quản lý các tên miền như .com, .uk, .vn. Trung tâm Internet Việt Nam (VNNIC) quản lý các tên miền .vn và cung cấp cơ sở hạ tầng để trả lời truy vấn các tên miền này một cách tin cậy, an toàn và ổn định. Với yêu cầu đó, VNNIC đã thu thập dữ liệu các hoạt động truy vấn để phục vụ cho mục đích phân tích và phát hiện các bất thường sớm nhằm ngăn chặn các cuộc tấn công, giữ an toàn ổn định cho hệ thống DNS. Tuy nhiên với các phương pháp phát hiện bất thường truyền thống như dựa trên chữ ký hay đặt ngưỡng cho các nguồn truy vấn thì hiệu quả phát hiện bất thường đang rất hạn chế. Đồ án đã nghiên cứu các phương pháp phát hiện bất thường đang được phát triển ở các cơ quan quản lý tên miền cấp cao tương tự như VNNIC cũng như các phương pháp phát hiện bất thường trong lưu lượng mạng chung để áp dụng, xây dựng và tích hợp hệ thống phát hiện bất thường vào hệ thống lưu trữ và dahsboard hiện có của VNNIC. Kết hợp giữa hai phương pháp: QLAD-flow dựa trên công trình nghiên cứu của Dewaele et al và QLAD-global dựa trên công trình nghiên cứu của Pieter Robberechts et al, đồ án đã đưa ra một giải pháp phát hiện bất thường nhanh chóng, có chi phí tính toán thấp, xử lý dữ liệu gần thời gian thực và có thể phát hiện nhiều loại bất thường.

Sinh viên thực hiện

(Ký và ghi rõ họ tên)

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	**Ý nghĩa**
(D)DoS	(Distributed) Denial of Service
AuthNS	Authoritative Name Server
ccTLD	country code TLD
DNS	Domain Name System
DNSSEC	DNS Security Extensions
EDNS	Extension mechanisms for DNS
EMA	Exponential Moving Average
ENTRADA	ENhanced Top-level domain Resilience through Advanced Data Analysis
gTLD	generic TLD
HDFS	Hadoop File System
ISP	Internet Service Provider
NXDOMAIN	non-existent domain
RDNS	Recursive DNS resolver
RR	Resource Record
TLD	Top-Level Domain
TTL	Time To Live

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

1.1 Đặt vấn đề

Trong bối cảnh mạng Internet ngày càng trở thành hạ tầng quan trọng cho các hoạt động kinh tế, xã hội và công nghệ, việc đảm bảo an toàn và ổn định cho các hệ thống mạng là một yêu cầu cấp bách. Hệ thống máy chủ tên miền (DNS) đóng vai trò trọng yếu trong các hoạt động của mạng. Các dịch vụ trên Internet như web, email... được sử dụng thường xuyên hàng ngày, đều gắn với tên miền và cần có hoạt động của hệ thống máy chủ tên miền - DNS ... DNS được ví như "trái tim của Internet", quyết định đến hoạt động của mạng Internet: Trả lời các thông tin về địa chỉ IP và tên miền sử dụng trong hầu hết các giao dịch, kết nối, sử dụng dịch vụ trên không gian mạng. Với chức năng chuyển đổi giữa tên miền và địa chỉ IP, DNS không chỉ là mục tiêu chính của các cuộc tấn công mạng mà còn là một trong những thành phần dễ bị tổn thương nhất trong hạ tầng mạng.

Cụ thể, theo báo cáo năm 2023 của IDC, dựa trên khảo sát 1000 tổ chức lớn nhỏ ở các lĩnh vực khác nhau đến từ 10 quốc gia thì 90% các tổ chức đã từng bị chịu ảnh hưởng bởi các cuộc tấn công DNS, trung bình mỗi tổ chức phải hứng chịu 7,5 cuộc tấn công DNS trong suốt năm 2023 và thiệt hại trung bình mỗi vụ tấn công là 1,1 triệu đô [1]. Đặc biệt các cuộc tấn công dựa vào khai thác giao thức DNS đang ngày càng trở nên phổ biến khi mà theo báo cáo của cloudflare, số lượng các cuộc tấn công DNS trong quý I năm 2024 đã tăng 80% so với năm 2023 [2].

Tại Việt Nam, Trung tâm Internet Việt Nam - VNNIC chịu trách nhiệm cho quản lý tên miền ccTLD .vn cũng như hệ thống máy chủ tên miền (DNS) quốc gia [3]. Tính đến cuối năm 2023, có tổng cộng 607,758 tên miền đã được đăng ký và hệ thống máy chủ tên miền quốc gia .vn chịu trách nhiệm trả lời những truy vấn liên quan đến những tên miền này. Trung bình hệ thống phải xử lý xấp xỉ 1 tỷ truy vấn mỗi ngày, và sẽ còn tăng cùng với sự phát triển của Internet [4].

Một trong những sứ mệnh của VNNIC là đảm bảo tính an toàn, ổn định của hệ thống máy chủ tên miền quốc gia. Do đó, VNNIC quan tâm đến việc xác định các mục tiêu tấn công, lạm dụng cơ sở hạ tầng của hệ thống tên miền máy chủ (DNS). Với khối lượng dữ liệu lớn các truy vấn thu thập mỗi ngày thì phương pháp phát hiện bất thường truyền thống dựa trên các mẫu chữ ký (signature-based) hoặc các ngưỡng cố định (threshold-based) hoặc giám sát thủ công các biểu hiện bất thường của các truy vấn là rất hạn chế. Chính vì vậy, việc nghiên cứu, phát triển và triển khai một hệ thống phát hiện bất thường dựa trên phân tích dữ liệu lưu lượng mạng DNS trở thành một hướng tiếp cận khả thi và hiệu quả.

1.2 Mục tiêu và định hướng giải pháp

Mục tiêu của đồ này là xây dựng một hệ thống phát hiện bất thường bằng cách sử dụng các kỹ thuật phân tích dữ liệu và triển khai các thuật toán để khai thác các đặc điểm tiềm ẩn trong lưu lượng mạng DNS authoritative tại máy chủ ccTLD (country code Top-Level Domain) cụ thể là tên miền .vn. Các bất thường này là các mẫu lưu lượng DNS không giống như kỳ vọng thông thường chẳng hạn như các yêu cầu DNS với tần suất cao, yêu cầu đến các tên miền đáng ngờ hoặc các hành vi tiềm ẩn khác có thể liên quan đến các cuộc tấn công mạng như DDoS, mã độc, hoặc phishing.

Để giải quyết các vấn đề đó, đồ án nghiên cứu tham khảo nền tảng phân tích log truy vấn (Query Log Anomaly Detection - QLAD) của DNS Belgium, một cơ quan có chức năng tương tự như VNNIC ở Việt Nam, quản lý tên miền .be và phương pháp phát hiện bất thường dựa trên dựa đoán lưu mạng (Anomaly Detection based on Predictions - AD-BoP) của NIC Chile, quản lý tên miền .cl.

Bằng cách nghiên cứu, tìm hiểu và áp dụng các phương pháp phát hiện bất thường của các tổ chức, quốc gia khác, đồ án hy vọng tìm ra giải pháp giúp tự động giám sát, theo dõi và phát hiện ra các bất thường về truy vấn trên hệ thống DNS, từ đó cung cấp cơ sở cho người quản trị có thể đưa ra các chiến lược phù hợp nhằm ngăn chặn sớm các cuộc tấn công, giảm thiểu các rủi ro tiềm tàng đối với hệ thống DNS, đóng góp cho sự phát triển bền vững của nền kinh tế số quốc gia.

1.3 Đóng góp của đồ án

Đồ án này có những đóng góp chính như sau:

1. Đồ án cung cấp phân tích về các loại hành vi bất thường (các loại tấn công, lạm dụng) thường gặp trong lưu lượng mạng DNS, qua đó giúp làm sáng tỏ mối liên hệ giữa các thông số mạng và các hoạt động tấn công.
2. Phát triển và triển khai một hệ thống tự động, có khả năng phát hiện các hành vi bất thường trong thời gian thực, từ đó hỗ trợ quản trị viên mạng xử lý các mối đe dọa một cách nhanh chóng và chính xác

1.4 Bố cục đồ án

Phần còn lại của báo cáo đồ án tốt nghiệp này được tổ chức như sau.

Chương 2 cung cấp tổng quan về DNS, các loại tấn công phổ biến, phương pháp phát hiện bất thường, và tổng quan dữ liệu log. Nội dung cũng phân tích dữ liệu lưu lượng mạng DNS authoritative thu thập từ máy chủ tên miền .vn.

Chương 3 trình bày chi tiết phương pháp phát hiện được đề xuất cũng như thảo