BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Huỳnh Nguyên Chính

GIẢI PHÁP PHÁT HIỆN NHANH CÁC HOT-IP TRONG HỆ THỐNG MẠNG VÀ ỨNG DỤNG

Chuyên ngành: Hệ thống thông tin
Mã số: 62.48.01.04

TÓM TẮT LUẬN ÁN TIẾN SĨ KỸ THUẬT

Hà Nội – 2017

Công trình được hoàn thành tại: Học viện Công nghệ Bưu chính Viễn thông

Người hướng dẫn khoa học: 1. PGS.TS. Nguyễn Đình Thúc

2. TS. Tân Hạnh

Phản biện 1: .....

.....

Phản biện 2: .....

.....

Phản biện 3: .....

.....

Luận án được bảo vệ trước Hội đồng chấm luận án cấp Học viện

họp tại: Học viện Công nghệ Bưu chính Viễn thông

Vào lúc:……. giờ……. ngày…… tháng…… năm……

Có thể tìm hiểu luận án tại thư viện:.....

MỞ ĐẦU

1. Giới thiệu

Các giải pháp phát hiện sớm các đối tượng có khả năng gây nguy hại trên mạng, nhất là hệ thống mạng trung gian ở phía các nhà cung cấp dịch vụ, có ý nghĩa quan trọng trong việc giúp giảm thiểu các ảnh hưởng xấu cho các máy chủ của khách hàng và các dịch vụ trên mạng Internet. Phát hiện sớm các đối tượng này để tiến hành các giải pháp ứng phó, ngăn chặn kịp thời là vấn đề quan trọng trong bài toán an ninh mạng.

Các gói tin lưu thông trên mạng IP có gắn thông tin về địa chỉ IP để xác định thiết bị gửi và nhận trong phần IP-header. Dựa trên thông tin các địa chỉ IP này, bài toán phát hiện các đối tượng hoạt động với tần suất xuất hiện cao trong một khoảng thời gian ngắn được đưa về bài toán phát hiện các Hot-IP.

Luận án nghiên cứu và đề xuất giải pháp phát hiện các Hot-IP trên mạng nhằm mục đích phát hiện sớm các đối tượng có khả năng gây hại. Các Hot-IP có thể là các mục tiêu trong tấn công từ chối dịch vụ, các máy phát động tấn công từ chối dịch vụ, các máy đang tiến hành quét mạng để tìm kiếm lỗ hổng nhằm phát tán sâu Internet, các thiết bị hoạt động bất thường trong hệ thống mạng. Phát hiện sớm các Hot-IP là bước cơ bản, quan trọng đầu tiên, từ đó giúp người quản trị tiến hành các giải pháp phòng chống hiệu quả, kịp thời.

2. Lý do chọn đề tài

Hai bài toán quan trọng trong lĩnh vực an ninh mạng là tấn công từ chối dịch vụ và phát tán sâu Internet. Đặc trưng quan trọng của các dạng tấn công này là số lượng gói tin mang các đối tượng tấn công xuất hiện rất lớn trong khoảng thời gian rất ngắn. Các giải pháp hiện tại ở bước phát hiện và phòng chống tấn công mới chỉ tập trung giải quyết vấn đề phát hiện có luồng lưu lượng tấn công vào hệ thống hay không mà không chỉ ra được các đối tượng gây nên tấn công đó. Các kỹ thuật phát hiện các đối tượng phát tán tấn công thực hiện ở bước hậu tấn công. Để có thể vừa phát hiện nguy cơ tấn công đồng thời có thể chỉ ra các đối tượng gây ra nguy cơ đó trong dòng gói tin IP

thời gian thực là vấn đề quan trọng đặt ra nhưng chưa có giải pháp, nhằm cảnh báo sớm để có giải pháp ứng phó kịp thời.

Phát hiện sớm các Hot-IP trên mạng và ứng dụng để phát hiện các đối tượng có khả năng là nguy cơ gây nên các cuộc tấn công từ chối dịch vụ, mục tiêu trong các cuộc tấn công này hay phát hiện các máy đang tiến hành quét mạng tìm kiếm lỗ hổng để phát tán sâu Internet là vấn đề luận án tập trung nghiên cứu. Trong các phương pháp mà luận án đã khảo sát thì phương pháp thử nhóm bất ứng biến là thích hợp nhất để triển khai áp dụng.

3. Mục tiêu nghiên cứu

3.1. Mục tiêu tổng quát

Mục tiêu của luận án là xây dựng giải pháp phát hiện các Hot-IP trên mạng máy tính bằng phương pháp thử nhóm bất ứng biến; sử dụng một số kỹ thuật và công cụ toán học kết hợp nhằm nâng cao hiệu quả phát hiện Hot-IP như xây dựng thuật toán và ma trận phân cách phù hợp với vị trí triển khai, xử lý song song, kiến trúc phân tán; áp dụng giải pháp này cho một số bài toán an ninh mạng như phát hiện các đối tượng có khả năng là mục tiêu hay nguồn phát trong tấn công từ chối dịch vụ hay tấn công từ chối dịch vụ phân tán, các thiết bị hoạt động bất thường, nguồn phát tán sâu Internet và giám sát các Hot-IP trên mạng.

3.2. Các mục tiêu cụ thể

- Nghiên cứu lý thuyết thử nhóm bất ứng biến để đề xuất giải pháp phát hiện các Hot-IP trên mạng.
- Đề xuất xây dựng ma trận phân cách tường minh nhằm giảm chi phí tính toán và bộ nhớ khi sử dụng ma trận phân cách.
- Đề xuất cải tiến thuật toán thử nhóm bất ứng biến để giảm thời gian tính toán và phát hiện Hot-IP trên dòng gói tin IP thời gian thực.
- Áp dụng kỹ thuật xử lý song song, kiến trúc phân tán để đề xuất giải pháp kết hợp nhằm phát hiện nhanh các Hot-IP trên mạng.

- Mô hình hóa các bài toán ứng dụng: phát hiện các đối tượng có khả năng là nạn nhân trong các cuộc tấn công từ chối dịch vụ, nguồn phát tấn công từ chối dịch vụ, nguồn phát tán sâu Internet, các thiết bị hoạt động bất thường về bài toán phát hiện các Hot-IP trên mạng.
- Giám sát các Hot-IP kết hợp với theo dõi tài nguyên hệ thống để điều phối lưu lượng mạng, giảm thiểu các nguy hại trên hệ thống.

4. Đối tượng, phạm vi nghiên cứu

Nghiên cứu lý thuyết thử nhóm bất ứng biến và áp dụng vào bài toán phát hiện các Hot-IP trên mạng; đồng thời sử dụng kết hợp với kỹ thuật xử lý song song, kiến trúc phân tán để nâng cao hiệu quả của giải pháp phát hiện và cảnh báo sớm các Hot-IP trên mạng.

5. Phương pháp nghiên cứu

Nghiên cứu lý thuyết thử nhóm bất ứng biến:

- Hệ thống hóa các khái niệm
- Phân tích và cải tiến các thuật toán trong thử nhóm bất ứng biến

Triển khai thực nghiệm các giải pháp phát hiện Hot-IP:

- Thực nghiệm nhằm xác định các tham số thích hợp
- Phân tích thực nghiệm

6. Những đóng góp chính của luận án

(i) **Đề xuất giải pháp phát hiện các Hot-IP trên mạng dựa trên thử nhóm bất ứng biến và một số kỹ thuật kết hợp để nâng cao hiệu quả của giải pháp.** Trong đó, kỹ thuật tính toán song song được sử dụng trong bước tính vector kết quả của các nhóm thử, sử dụng kiến trúc phân tán trong các hệ thống mạng đa vùng để cảnh báo sớm từ các vùng phát hiện được Hot-IP và lựa chọn kích thước ma trận phù hợp ở vị trí triển khai nhằm giảm áp lực tính toán. Lý thuyết nền tảng cho phương pháp đề xuất là sử dụng phương pháp nối mã để xây dựng tường minh ma trận phân cách. Nhờ đó, không gian lưu trữ được tối

ưu thay vì phải lưu trữ toàn ma trận có kích thước lớn trên trường hữu hạn.

(ii) **Đề xuất cải tiến thuật toán thử nhóm bất ứng biến để giảm thời gian tính toán phát hiện các Hot-IP trực tuyến.** Đặc điểm khác biệt của cải tiến là các nhóm thử đến ngưỡng không cần phải cập nhật tiếp tục, danh sách các địa chỉ IP nghi ngờ được xác định và khởi tạo bộ đếm tương ứng, các cập nhật đối với các IP có mặt trong danh sách nghi ngờ được thực hiện thay vì phải cập nhật trong tập tất cả các bộ đếm của các nhóm thử dựa vào ma trận phân cách.

(iii) **Mô hình hóa 4 bài toán ứng dụng**: (1) phát hiện các đối tượng có khả năng là các nguồn phát tán sâu Internet, (2) phát hiện các thiết bị có khả năng đang hoạt động bất thường, (3) phát hiện các đối tượng có khả năng là mục tiêu hay nguồn phát trong tấn công từ chối dịch vụ về bài toán phát hiện Hot-IP và (4) giám sát hoạt động của các Hot-IP kết hợp với theo dõi tài nguyên mạng để điều phối hay hạn chế hoạt động của các luồng dữ liệu chứa các Hot-IP này. Kỹ thuật được sử dụng là phân tích luồng dữ liệu dựa vào địa chỉ IP nguồn và đích trong các gói tin kết hợp với theo dõi tài nguyên hệ thống làm dữ liệu đầu vào trong thuật toán phát hiện các Hot-IP.

7. Giới thiệu tổng quan về nội dung luận án

Nội dung của luận án tập trung vào nghiên cứu phương pháp thử nhóm bất ứng biến và áp dụng vào bài toán phát hiện các Hot-IP trên mạng; đề xuất thuật toán cải tiến; đề xuất một số kỹ thuật kết hợp để tăng hiệu quả tính toán của giải pháp và đề xuất ứng dụng phát hiện các Hot-IP trong một số bài toán an ninh mạng.

Cấu trúc của luận án được tổ chức thành 4 chương. Chương 1 trình bày tổng quan về bài toán Hot-IP, một số khái niệm, khảo sát các nghiên cứu liên quan. Trên cơ sở đó, luận án đề xuất giải pháp phát hiện các Hot-IP trên mạng dùng phương pháp thử nhóm bất ứng biến.

Chương 2 trình bày phương pháp thử nhóm bất ứng biến, một số khái niệm liên quan, phương pháp xây dựng tường minh ma trận d-phân-cách bằng phép nối mã và áp dụng phương pháp thử nhóm bất ứng biến vào việc phát hiện các Hot-IP trên mạng. Trong chương này, luận án đề xuất hai thuật toán cải tiến *"Online Hot-IP Detecting"* và *"Online Hot-IP Preventing"* để giảm thời gian tính toán, tăng mức độ chính xác và đảm bảo hệ thống hoạt động ổn định, thông suốt khi phát hiện trực tuyến trên cơ sở sử dụng danh sách các địa chỉ IP nghi ngờ.

Chương 3 trình bày một số kỹ thuật kết hợp nhằm nâng cao khả năng phát hiện các Hot-IP trên mạng. Trong đó, luận án đề xuất kết hợp với kỹ thuật xử lý song song, kiến trúc phân tán trong các hệ thống mạng đa vùng, ý nghĩa và một số căn cứ để lựa chọn các tham số quan trọng trong giải pháp đề xuất áp dụng tại vị trí triển khai.

Chương 4 trình bày mô hình hóa một số ứng dụng trong lĩnh vực an ninh mạng như phát hiện các đối tượng có khả năng là các nguồn phát hay mục tiêu trong các cuộc tấn công từ chối dịch vụ, phát hiện các thiết bị có khả năng đang hoạt động bất thường trong hệ thống mạng, phát hiện các đối tượng có khả năng là nguồn phát tán sâu Internet về bài toán phát hiện các Hot-IP trên mạng, giám sát các Hot-IP trong một vài chu kỳ thuật toán kết hợp với theo dõi tài nguyên mạng để hạn chế hoạt động của chúng, nhằm giúp các nhà quản trị mạng theo dõi và ứng phó kịp thời, đảm bảo hệ thống mạng hoạt động ổn định, thông suốt.

Trong phần kết luận, luận án tổng kết những kết quả đạt được và bài toán mở cho nghiên cứu tương lai khi áp dụng kết quả luận án vào thực tiễn.

CHƯƠNG 1. TỔNG QUAN VỀ HOT-IP TRÊN MẠNG

1.1. Giới thiệu

Các cuộc tấn công từ chối dịch vụ, đặc biệt là tấn công từ chối dịch vụ phân tán, phát tán sâu trên Internet ngày càng dễ thực hiện nhưng tác hại của

nó là đặc biệt nghiêm trọng. Đặc điểm quan trọng trong các cuộc tấn công này là tốc độ và thời gian tiến hành rất ngắn. Chính vì nhanh và ngắn như vậy làm cho các nhà quản trị không thể kịp thời chống đỡ, hệ thống mạng bị cạn kiệt tài nguyên, băng thông, dẫn đến tình trạng các dịch vụ mạng bị ngưng trệ không thể đáp ứng tốt cho những người dùng hợp lệ.

Các nghiên cứu về phòng chống tấn công từ chối dịch vụ, phát tán sâu Internet ở giai đoạn phát hiện tấn công chủ yếu xem xét trong luồng dữ liệu có chứa đựng khả năng tấn công hay không mà không chỉ ra đối tượng hay mục tiêu trong các tấn công này.

Trên mạng tốc độ cao như ở phía nhà cung cấp dịch vụ hay các hệ thống cung cấp dịch vụ trên Internet rất cần có giải pháp thực hiện nhanh chóng, đơn giản và hiệu quả nhằm phát hiện nhanh các đối tượng có khả năng là nguy cơ gây nên các cuộc tấn công này để có thể kịp thời hạn chế ảnh hưởng xấu của chúng.

Dựa vào dòng dữ liệu lưu thông qua các thiết bị mạng, các thông tin về địa chỉ IP nguồn và địa chỉ IP đích xuất hiện với tần suất cao trong một khoảng thời gian rất ngắn (Hot-IP) dẫn đến khả năng các máy chủ có thể đang bị tấn công từ chối dịch vụ, các đối tượng đang phát tán sâu mạng hay đang thực hiện tấn công từ chối dịch vụ. Do đó, việc xác định các đối tượng có khả năng là mục tiêu trong tấn công từ chối dịch vụ, các máy đang phát động tấn công từ chối dịch vụ hay các máy đang phát tán sâu Internet có thể đưa về dạng bài toán phát hiện các Hot-IP trên mạng. Ở đây ta đã sử dụng nhận xét: ”Có tấn công dạng từ chối dịch vụ hay phát tán sâu Internet dạng quét không gian địa chỉ IP thì xuất hiện Hot-IP, nhưng xuất hiện Hot-IP chưa chắc bị tấn công”. Do đó, luận án nghiên cứu giải pháp dung hòa giữa phòng chống tấn công và tính sẵn sàng của mạng.

1.2. Một số khái niệm và định nghĩa

Khái niệm 1: Địa chỉ IP là chuỗi các ký hiệu dùng để định danh cho các thiết bị trên mạng.

Khái niệm 2: Gói tin IP là gói tin ở tầng mạng trong mô hình OSI, trong đó có phần IP-header mô tả thông tin ở tầng này. Trong cấu trúc của IP-header chứa thông số về địa chỉ IP nguồn và IP đích. Các giá trị địa chỉ này được sử dụng làm tham số đầu vào trong bài toán phát hiện các Hot-IP.

Khái niệm 3: Dòng gói tin IP là một dãy liên tiếp các gói tin IP $(a_1, a_2, ..., a_m)$ luân chuyển trên một đường truyền xác định. Trong đó, mỗi gói tin a_i có địa chỉ IP cần phân tích là s_i (s_i có thể là IP nguồn hay IP đích cần xem xét tùy vào ứng dụng cụ thể).

Định nghĩa 1: Hot-IP trong dòng gói tin IP trên mạng máy tính là những IP xuất hiện với tần suất cao trong khoảng thời gian ngắn xác định trước. Cho dòng gói tin IP có địa chỉ IP tương ứng $S = (IP_1, IP_2, ..., IP_m)$, ký hiệu N là số IP khác nhau trong m IP thuộc S ($0 \le N \le m$). Gọi $f_i = \left| \left\{ j \middle| IP_i = IP_j; i \ne j; IP_i, IP_j \in S \right\} \right|$, thì Hot-IP = $\left\{ IP_i \in S \middle| f_i \ge \phi \times m, 0 \le \phi \le 1 \right\}$.

1.3. Vị trí thu thập và xử lý dữ liệu

Vị trí triển khai thu thập dữ liệu có thể triển khai theo dạng inline hoặc promiscuous.

1.4. Các nghiên cứu liên quan

Các nghiên cứu liên quan đến Hot-IP chủ yếu được đề cập trong các công trình nghiên cứu về phát hiện và phòng chống tấn công từ chối dịch vụ, các nghiên cứu về một số loại sâu Internet dạng quét không gian địa chỉ để tìm kiếm lỗ hổng và phát tán trên môi trường Internet. Do đó, luận án tập trung phân tích các nghiên cứu liên quan này. Để mở rộng phạm vi so sánh và lựa chọn giải pháp thích hợp, luận án khảo sát các thuật toán phát hiện phần tử tần suất cao trong dòng dữ liệu. Từ đó, luận án có cơ sở lựa chọn giải pháp phù hợp để áp dụng vào bài toán phát hiện các Hot-IP trên mạng.

1.4.1. Các nghiên cứu về tấn công DoS/DDoS

Tấn công từ chối dịch vụ, đặc biệt là tấn công từ chối dịch vụ phân tán là dạng tấn công nguy hiểm trên mạng, gây nhiều hậu quả nghiêm trọng và thiệt hại lớn. Mục tiêu của kẻ tấn công là làm tê liệt các ứng dụng, máy chủ, gián

đoạn các kết nối, ngăn cản người dùng hợp lệ truy cập vào một dịch vụ nào đó trên mạng. Thông thường trong các cuộc tấn công này, các máy chủ sẽ bị "tràn ngập" bởi hàng loạt các truy vấn trong một khoảng thời gian rất ngắn, dẫn đến quá tải và mất khả năng phục vụ. Tấn công từ chối dịch vụ phân tán hiện nay đã phát triển một cách đáng lo ngại và là mối đe dọa thường trực đối với các hệ thống mạng.

Các giải pháp phát hiện và phòng chống tấn công từ chối dịch vụ được phân làm 4 loại chính: *đề phòng*, *phát hiện tấn công*, *phản ứng lại tấn công* và *xác định nguồn phát tấn công*. Trong đó, các nghiên cứu về phát hiện tấn công và phát hiện các nguồn phát tấn công là hai vấn đề quan tâm trong luận án này.

Các nghiên cứu về phát hiện và phòng chống xâm nhập có thể kể đến hai nhóm giải pháp chính: dựa vào dấu hiệu được định nghĩa sẵn và thiết lập ngưỡng tần suất. Giải pháp dựa vào dấu hiệu thực hiện việc so khớp các dấu hiệu được định nghĩa sẵn và thông tin nội dung trong các gói tin trong dòng dữ liệu thu thập được. Việc thiết lập ngưỡng dựa trên các chế độ bình thường được định nghĩa sẵn và sử dụng trong phương pháp thống kê, phương pháp học máy, khai phá dữ liệu. Các phương pháp này gặp khó khăn trong việc định nghĩa các trạng thái bình thường của hệ thống.

Mặc dầu có nhiều giải pháp phát hiện và phòng chống tấn công từ chối dịch vụ đã được nghiên cứu và đề xuất; tuy nhiên, cho đến nay vẫn chưa có giải pháp nào có khả năng phòng chống tấn công từ chối dịch vụ một cách toàn diện và hiệu quả do tính chất phức tạp, quy mô lớn và khả năng phân tán rất cao của các dạng tấn công này. Do vậy, phát hiện sớm các đối tượng có khả năng là nguồn phát tấn công hoặc mục tiêu trong các tấn công này có vai trò quan trọng trong bài toán an ninh mạng.

Có ba vị trí triển khai giải pháp phát hiện và phòng chống tấn công từ chối dịch vụ: *phía mạng của các máy chủ nạn nhân*, *vị trí mạng trung gian*, *vị trí mạng nguồn phát tấn công*. Trong các mạng trung gian như mạng ở các nhà

cung cấp dịch vụ, việc phát hiện các đối tượng có khả năng là mục tiêu hay nguồn phát trong các cuộc tấn công từ chối dịch vụ dựa vào phân tích lưu lượng đi qua nó có ý nghĩa quan trọng. Từ việc phát hiện này có thể giúp cảnh báo sớm cho khách hàng để tiến hành các biện pháp ứng phó kịp thời hoặc loại bỏ các nguy cơ này để đảm bảo hệ thống hoạt động ổn định.

Phương pháp thử nhóm bất ứng biến có thể xác định các đối tượng có khả năng là nguồn phát tấn công, các đối tượng có khả năng là mục tiêu trong tấn công từ chối dịch, các đối tượng có khả năng là nguồn phát tán sâu đang tiến hành quét mạng ngay ở giai đoạn phát hiện tấn công. Đồng thời phương pháp này cho kết quả tốt ở khía cạnh thời gian, độ chính xác cao và mức độ đơn giản của giải pháp.

1.4.2. Các nghiên cứu về sâu Internet

Trong các loại sâu Internet, *"scanning worm"*, *"routing worm"* và *"hit-list worm"* là những sâu nguy hiểm, phát tán dựa vào thông tin trong bảng định tuyến và danh sách địa chỉ IP được thiết lập trước với tốc độ cao.

Hoạt động lây nhiễm sâu gồm các giai đoạn: *phát hiện mục tiêu, truyền sâu, kích hoạt* và *lây nhiễm*. Quá trình hoạt động lây nhiễm sâu Internet ở hai giai đoạn đầu ảnh hưởng đến hoạt động của mạng, nên các hành vi của chúng ở hai giai đoạn này rất quan trọng để tiến hành triển khai các giải pháp phát hiện. Đặc điểm quan trọng cần lưu ý để tạo thuận lợi cho việc phát hiện chúng là ở bước phát hiện mục tiêu, phương pháp đơn giản nhất các sâu hay sử dụng là *"quét mù"*. Phương pháp này có tính cơ hội và tỷ lệ thất bại cao.

Như vậy, qua các phân tích về hai bài toán ứng dụng liên quan đến các địa chỉ IP xuất hiện tần suất cao trong khoảng thời gian ngắn là bài toán phát hiện các đối tượng là nguồn phát hay mục tiêu trong tấn công từ chối dịch vụ và bài toán phát tán sâu Internet đối với một số loại sâu dạng quét không gian địa chỉ IP cho thấy các giải pháp hiện tại chỉ mới tập trung vào việc phát hiện có tồn tại tấn công hay không trong bước phát hiện và phòng chống tấn công.

Việc xác định các đối tượng gây ra tấn công được thực hiện ở bước hậu tấn công.

Do vậy, cần một giải pháp có thể cân bằng điều này, nghĩa là có thể nhanh chóng phát hiện các nguy cơ tấn công và đồng thời chỉ ra được các đối tượng này là những IP nào ở giai đoạn xảy ra tấn công trong tấn công từ chối dịch vụ hay ở giai đoạn quét không gian địa chỉ IP để phát tán sâu Internet. Giải pháp đặt ra được đưa về giải bài toán phát hiện Hot-IP trên mạng mà luận án nghiên cứu giải quyết.

1.4.3. Các nghiên cứu về thuật toán phát hiện phần tử tần suất cao

Các thuật toán tìm phần tử tần suất cao trong dòng dữ liệu được chia thành hai nhóm chính: các thuật toán "counter-based" và các thuật toán "Sketch". Các thuật toán loại "counter-based" giám sát một tập các phần tử từ dòng dữ liệu đầu vào cùng với một biến đếm tương ứng với mỗi phần tử được giám sát, sau đó một tập các luật tương ứng cho mỗi thuật toán sẽ được áp dụng trên danh sách các phần tử này để tìm ra các phần tử tần suất cao. Các thuật toán loại "Sketch" không giám sát một tập các phần tử từ dòng dữ liệu mà xem dòng dữ liệu đầu vào như một vector với mỗi tọa độ của vector là tần suất xuất hiện của một phần tử tương ứng trong dòng dữ liệu, dựa trên các tần số ước lượng này sẽ tính toán ra các phần tử tần suất cao trong dòng dữ liệu.

Các thuật toán "counter-based" lưu trữ mỗi đối tượng bằng một bộ đếm nên tốn nhiều không gian lưu trữ với số lượng rất lớn các đối tượng trên mạng, đặc biệt trên mạng ở các nhà cung cấp dịch vụ, không thích hợp cho bài toán phát hiện các Hot-IP được thiết lập trên môi trường mạng với các thiết bị có tài nguyên hạn chế.

1.4.4. Phương pháp thử nhóm

Phương pháp thử nhóm bất ứng biến có nhiều ưu điểm trong bài toán tìm phần tử tần suất cao trong dòng dữ liệu lớn đã được đề cập trong một số nghiên cứu như thực hiện đơn giản, tốc độ nhanh và độ chính xác cao, tuy nhiên còn hạn chế là chiếm nhiều không gian lưu trữ.

Luận án cũng đã tiến hành thực nghiệm so sánh giữa một số thuật toán tiêu biểu của phương pháp "counter-based" và phương pháp thử nhóm bất ứng biến. Từ kết quả thực nghiệm cho thấy rằng phương pháp "counter-based" cho kết quả tốt hơn phương pháp thử nhóm bất ứng biến trong trường hợp số lượng phần tử nhỏ. Tuy nhiên với số lượng phần tử lớn, phương pháp thử nhóm bất ứng biến cho kết quả tốt hơn.

1.5. Giải pháp phát hiện các Hot-IP

Xuất phát từ hai bài toán ứng dụng thực tế là bài toán phát hiện các đối tượng có khả năng là nguồn phát hay mục tiêu trong tấn công từ chối dịch vụ và bài toán phát hiện các đối tượng tán sâu trên Internet có thể tổng quát thành bài toán phát hiện các Hot-IP trên mạng. Trên cơ sở phân tích các nghiên cứu liên quan và các thuật toán phát hiện phần tử tần suất cao trên dòng dữ liệu cho thấy rằng phương pháp thử nhóm bất ứng biến có nhiều lợi thế để áp dụng vào việc phát hiện các Hot-IP trực tuyến trên mạng.

Mục tiêu của luận án là đưa ra giải pháp phát hiện các Hot-IP trực tuyến với dòng dữ liệu lớn. Một số vấn đề cần xem xét là: không gian lưu trữ, thời gian tính toán, phương pháp bố trí bộ phát hiện Hot-IP phân tán cho các hệ thống mạng đa vùng, lựa chọn các tham số cho giải pháp phù hợp theo vị trí triển khai và khả năng của hệ thống

CHƯƠNG 2. PHÁT HIỆN CÁC HOT-IP SỬ DỤNG THỬ NHÓM BẤT ỨNG BIẾN

2.1. Giới thiệu về thử nhóm

Phương pháp thử nhóm được chia thành 2 loại là *thử nhóm ứng biến* và *thử nhóm bất ứng biến*. Trong thử nhóm ứng biến, phép thử sau được thiết kế dựa vào kết quả của phép thử trước đó, thuật toán thử nhóm ứng biến có bản chất tuần tự. Trong thử nhóm bất ứng biến, tất cả các phép thử phải được xác định trước mà không phụ thuộc vào bất kỳ phép thử nào.

2.2. Thử nhóm bất ứng biến

Trong một số ứng dụng cho các bài toán trên dòng dữ liệu yêu cầu phải sử dụng phương pháp thử nhóm bất ứng biến vì dữ liệu trên dòng dữ liệu đi qua thuật toán và cho ra kết quả ngay. Do đó, luận án chỉ tập trung nghiên cứu về phương pháp thử nhóm bất ứng biến để áp dụng vào bài toán phát hiện các Hot-IP trực tuyến trên mạng.

Mô hình hóa bài toán phát hiện các Hot-IP trên dòng gói tin IP về bài toán thử nhóm bất ứng biến như sau: cho dòng gói tin IP, trong đó có *N* địa chỉ IP phân biệt. Giả sử có tối đa *d* phần tử là Hot-IP, thiết kế *t* nhóm thử cho *N* địa chỉ IP này. Xây dựng một ma trận nhị phân M_{txN}, trong đó các cột của ma trận đại diện cho các địa chỉ IP và các hàng của ma trận đại diện cho các nhóm thử.

Nếu *M* là ma trận d-phân-cách thì chúng ta có thể chỉ ra rằng có nhiều nhất *d* phần tử là Hot-IP,với $d \ll N, t \ll N$, nghĩa là tổng không gian sử dụng để lưu trữ trong phương pháp thử nhóm bất ứng biến nhỏ hơn rất nhiều so với phương pháp dùng mỗi bộ đếm cho mỗi IP. Để chỉ ra các Hot-IP trong dòng gói tin IP, từ ma trận d-phân-cách và vector kết quả của phép thử, thuật toán giải mã sẽ chỉ ra những địa chỉ IP nào là Hot-IP mà không cần bất kỳ một cấu trúc dữ liệu nào khác.

2.3. Ma trận d-phân-cách

Định nghĩa 2. *Ma trận nhị phân* $M_{t\times N}$ *được gọi là d-phân-cách khi và chỉ khi hội của d cột bất kỳ không chứa bất kỳ một cột nào khác. Với d+1 cột* C_0, $C_1,...,C_d$ *bất kỳ của M, ta có* $C_0 \subsetneq C_1 \cup ... \cup C_d$.

2.4. Phát hiện Hot-IP dùng thử nhóm bất ứng biến

2.4.1. Phát biểu bài toán:

Cho một dòng *m* gói IP với địa chỉ tương ứng S=(IP_1, IP_2,…,IP_m), với *m* rất lớn. Mỗi gói tin IP có địa chỉ IP trong tập [N], N cũng rất lớn ($N=2^{32}$ với IPv4, $N=2^{128}$ với IPv6). Gọi $f_i = \left|\{ j \left| IP_i = IP_j; i \neq j; IP_i, IP_j \in S \}\right.\right|$, thì Hot-IP = $\{IP_i \in S \left| f_i \geq \phi \times m, 0 \leq \phi \leq 1\right.\}$. Giả sử có đối đa *d* Hot-IP trong dòng gói tin IP. Xác định các Hot-IP trong S.

Bài toán có thể giải bằng phương pháp thử nhóm bất ứng biến được mô hình hóa như sau: cho trước ma trận nhị phân $M_{t\times N}$ với *t* là hàm phụ thuộc *d* và *N*. Trong đó, *t* là số hàng của ma trận tương ứng với các nhóm thử trong thử nhóm và N là số cột của ma trận tương ứng với N địa chỉ IP phân biệt. Gọi m_{ij} là phần tử của ma trận ở hàng i, cột j; các phần tử của ma trận có giá trị như sau:

$$m_{ij}=\begin{cases}1 & \text{nếu } IP_j \text{ thuộc nhóm thử i}\\ 0 & \text{ngược lại}\end{cases}$$

Giả sử có vector kết quả $r_{t\times 1}$ sau khi đếm và xét ngưỡng, các r_i có giá trị như sau:

$$r_i=\begin{cases}1 & \textit{(nhóm thử có chứa Hot-IP)}\\ 0 & \textit{(nhóm thử không chứa Hot-IP)}\end{cases}$$

Ta cần xác định xem những IP nào là Hot-IP.

2.4.2. Giải pháp phát hiện các Hot-IP

Sử dụng *t* bộ đếm $c_1, c_2, ..., c_t$ tương ứng với số dòng của ma trận nhị phân *M*, khi một gói tin có địa chỉ IP $j \in [N]$ tới thì tăng tất cả các bộ đếm c_i nếu $m_{ij} = 1$. Từ bộ đếm này và một ngưỡng cho trước, một vector kết quả được tạo ra $r \in \{0,1\}^t$. Trong đó, kết quả nhóm thử có chứa Hot-IP là 1 và kết quả của nhóm thử không chứa Hot-IP là 0. Từ vector kết quả và ma trận M, xác định được các Hot-IP.

2.5. Đề xuất thuật toán cải tiến

Ý tưởng chính cho thuật toán cải tiến phương pháp thử nhóm bất ứng biến trong bài toán phát hiện các Hot-IP là:

(1) Việc cập nhật các bộ đếm khi một IP đến cho từng nhóm sẽ dừng lại nếu nó vượt ngưỡng.

(2) Xác định các IP làm vượt ngưỡng trong nhóm này, đưa vào danh sách nghi ngờ và thiết lập bộ đếm tương ứng.

(3) Nếu một IP đến có trong danh sách nghi ngờ thì tăng bộ đếm tương ứng cho IP đó mà không cập nhật các bộ đếm trong các nhóm thử chứa địa chỉ IP này.

(4) Xác định Hot-IP bằng cách so sánh bộ đếm của các IP trong danh sách nghi ngờ với ngưỡng.

2.5.1. Thuật toán cải tiến 1 "Online Hot-IP detecting"

Thuật toán cải tiến 1: *Online Hot-IP Detecting*	
	Input: Ma trận d-phân-cách, dòng gói tin IP trong chu kỳ ***Output:*** các Hot-IP
1:	Hot-List={}
2:	*For each IP* $j \in S_\Delta$ *// đối với mỗi gói tin IP đến*
3:	*If(current_timestamp-reference_timestamp<*Δ*)then*
4:	*If IP j* $\in$ *Hot-List then*
5:	*Hot-List[j].count++*
6:	*Else*
7:	*For i = 1 to N*
8:	*If* m_{ij} *= 1 and* $c_i < \delta$ *then* c_i*++*
9:	*If* $c_i \geq \delta$ *then*
10:	*Hot-List = Hot-List* $\cup$ *{j}*
11:	*Hot-List[j].count = min{*c_i *|* m_{ij}*=1}*
12:	*EndIf*
13:	*EndFor*
14:	*Else*
15:	*Return* $\{j \mid \text{Hot-List[j].count} \geq \delta, 1 \leq j \leq \mid \text{Hot-List} \mid\}$
16:	*//xuất ra các IP trong Hot-List có bộ đếm tương ứng vượt ngưỡng*
17:	*Reference_timestamp=current_timestamp*
18:	*Reset Hot-List*
19:	*EndIf*

Thuật toán cải tiến 1 "Online Hot-IP Detecting" thực hiện việc theo dõi các gói tin trực tuyến và xuất các Hot-IP phát hiện được trong một chu kỳ thuật toán. Khi một địa chỉ IP được trích ra từ gói tin IP đến, nó sẽ được kiểm tra trong danh sách IP nghi ngờ (Hot-List), nếu tồn tại trong danh sách này thì tăng bộ đếm tương ứng cho IP này. Nếu chưa tồn tại trong Hot-List thì việc

cập nhật cho các nhóm thử chứa IP này được thực hiện bình thường như trong thuật toán thử nhóm bất ứng biến truyền thống.

Khi bất kỳ một nhóm nào trong quá trình cập nhật IP mới vào làm vượt ngưỡng, địa chỉ IP đó được đưa vào danh sách nghi ngờ, khởi tạo bộ đếm tương ứng bằng cách lấy giá trị nhỏ nhất trong các nhóm mà IP này thuộc về, các nhóm vượt ngưỡng sẽ dừng việc cập nhật.

Qua các phân tích và thực nghiệm cho thấy thuật toán thử nhóm bất ứng biến cải tiến có nhiều ưu điểm hơn phương pháp thử nhóm bất ứng biến truyền thống. Thứ nhất là không cập nhật các nhóm thử đã đến ngưỡng, thứ hai là thay vì cập nhật IP trong dòng dữ liệu cho tất cả các nhóm thử chứa IP đó thì chỉ cần cập nhật trong danh sách nghi ngờ. Những ưu điểm này làm giải thời gian tính toán trong chương trình. Đồng thời, thuật toán cải tiến này cho kết quả chính xác hơn trong trường hợp số lượng Hot-IP thực tế nhiều hơn số Hot-IP tối đa định trước trong thử nhóm bất ứng biến truyền thống.

2.5.2. Thuật toán cải tiến 2 "Online Hot-IP preventing"

Thuật toán cải tiến 2: *Online Hot-IP Preventing*

Input: Ma trận nhị phân d-phân-cách, dòng gói tin IP, ngưỡng δ
Δ: chu kỳ thuật toán

Xử lý:

```
  T=Systemtime + Δ,
     Khởi tạo:cho phép tất cả các IP đi qua
 For each IP j đến và (Systemtime <T)
   j=get(IP)
   if IP j ∈ Hot-List then
      Hot-List[j].count++
      If Hot-List[j].count > δ then drop(IP j)
   Else
         Cập nhật các bộ đếm c_i với m_ij=1,
         không cập nhật cho các c_i vượt ngưỡng
    If c_i > δ then
     Đưa IP j vào Hot-List
     Khởi tạo bộ đếm cho IP j = min{c_i với m_ij=1}
   endIf
   EndIf
 EndFor
```

Thuật toán cải tiến 2 "Online Hot-IP Preventing" thực hiện hạn chế các đối tượng có khả năng là nguy cơ ngay khi chúng được phát hiện bằng cách ngăn chặn các Hot-IP trong một chu kỳ thời gian của thuật toán nhằm hạn chế nhanh chóng các nguy cơ và đảm bảo hệ thống hoạt động ổn định, thông suốt.

Thuật toán cải tiến này có thể dùng để triển khai ở các router biên trước các máy chủ cung cấp dịch vụ hoặc ở các router trung gian trong các mạng trung gian đáp ứng mục tiêu đảm bảo cho hệ thống mạng hoạt động ổn định, thông suốt.

CHƯƠNG 3. NÂNG CAO HIỆU QUẢ PHÁT HIỆN HOT-IP BẰNG MỘT SỐ KỸ THUẬT KẾT HỢP

3.1. Giới thiệu

Để triển khai giải pháp phát hiện Hot-IP trên mạng ở một vị trí cụ thể cần có những phân tích để tối ưu tính toán. Một số kỹ thuật có thể kết hợp để nâng cao khả năng của giải pháp trong việc phát hiện nhanh các Hot-IP như: (i) lựa chọn kích thước của ma trận d-phân-cách phù hợp để giảm thời gian và không gian xử lý dựa vào khả năng của vị trí triển khai giải pháp, (ii) sử dụng kỹ thuật xử lý song song để nâng cao khả năng tính toán và (iii) sử dụng kiến trúc phân tán để tổ chức triển khai ở các khu vực và cảnh báo sớm đến các khu vực khác trong các hệ thống mạng đa vùng.

3.2. Vấn đề kích thước ma trận phân cách

3.2.1. Sự ảnh hưởng của kích thước ma trận

Việc lựa chọn kích thước của ma trận d-phân-cách có ý nghĩa quan trọng để áp dụng vào thực tế có hiệu quả. Kích thước ma trận ảnh hưởng đến thời gian cập nhật các gói dữ liệu trong dòng dữ liệu đầu vào và thời gian thực hiện thuật toán để phát hiện ra các Hot-IP một cách đáng kể.

Kích thước ma trận cần được xem xét để lựa chọn ma trận phù hợp ở từng vị trí triển khai cụ thể dựa vào khả năng của hệ thống mạng tại các vị trí đó.

3.2.2. Lựa chọn các tham số

❖ **Xác định N**

Giá trị N đại diện cho số lượng địa chỉ IP phân biệt. Hai trường hợp áp dụng có thể xem xét để tính toán giá trị N được đề xuất như sau:

- *Trường hợp 1:* Xem xét các địa chỉ IP là như nhau. Trong trường hợp này, dựa vào khả năng của hệ thống tại vị trí triển khai và kinh nghiệm của người quản trị để xác định N trong một chu kỳ thuật toán.
- *Trường hợp 2:* Phân biệt các IP đăng ký và IP không đăng ký sử dụng dịch vụ. Số lượng người dùng đăng ký sử dụng dịch vụ là N1 và số lượng người dùng dịch vụ không đăng ký là N2. Đối với những người dùng không đăng ký, N2 có thể dùng với số lượng nhỏ bằng các địa chỉ đại diện, ta có N=N1+N2.

❖ **Xác định d**

Giá trị d trong ma trận d-phân-cách là số lượng Hot-IP tối đa có thể phát hiện được bằng phương pháp thử nhóm bất ứng biến. Tùy vào ứng dụng mà tham số này có ý nghĩa khác nhau. Trong tấn công từ chối dịch vụ phân tán, *d* có ý nghĩa là số lượng nguồn phát tấn công (giám sát dựa vào địa chỉ IP nguồn) hoặc là số lượng tối đa các server có khả năng bị tấn công (giám sát dựa vào địa chỉ IP đích). Trong ứng dụng phát hiện nguồn phát tán sâu Internet, giá trị *d* có ý nghĩa là số lượng tối đa các máy tính có khả năng đang quét mạng để phát tán sâu.

Bài toán thử nhóm bất ứng biến truyền thống phụ thuộc rất nhiều vào việc chọn *d*. Để giảm sự phụ thuộc vào *d*, luận án đề xuất phương án sử dụng danh sách IP nghi ngờ (Hot-List) sử dụng trong các thuật toán cải tiến.

❖ **Xác định m**

Tham số m_Δ là tổng số gói tin bắt được trong một chu kỳ thuật toán.

❖ **Ngưỡng tần suất xuất hiện cao**

Giá trị ngưỡng được dùng để xác định kết quả của một nhóm thử có chứa phần tử là Hot-IP hay không. Giá trị ngưỡng được xác định $\delta = \frac{m_\Delta}{|\text{Hot-List}|}$.

❖ **Giá trị t**

Giá trị *t* là số hàng của ma trận hay số nhóm thử được thiết kế trước theo phương pháp thử nhóm bất ứng biến. Trong xây dựng ma trận d-phân-cách bằng phương pháp nối mã, giá trị *t* được xác định như sau: $t = n_1 \times q,$ với $C_{out} : [n_1, k_1]_q$ - RS và $C_{in} : I_q$.

3.3. Kiến trúc phân tán

3.3.1. Giới thiệu

Các cuộc tấn công mạng ngày càng có tính phối hợp cao, phân tán rộng trên Internet, để phát hiện và phòng chống một cách hiệu quả thì chiến lược phát hiện và phòng chống cũng cần được triển khai phân tán và hợp tác giữa các thành phần.

3.3.2. Kiến trúc phân tán phát hiện Hot-IP

Trong các hệ thống mạng được tổ chức đa vùng hay hệ thống mạng của các nhà cung cấp dịch vụ được tổ chức thành các khu vực để cung cấp dịch vụ cho các khách hàng trong khu vực đó. Ở mỗi khu vực, hệ thống các bộ phát hiện Hot-IP được thiết lập để phát hiện các Hot-IP. Hệ thống bộ phát hiện Hot-IP ở các vị trí triển khai giải pháp được thiết kế để kết nối với nhau theo dạng ngoài luồng dữ liệu hoặc tích hợp vào các router biên mạng. Mục đích của việc thiết lập này để tăng khả năng phát hiện sớm các Hot-IP trong mạng và trách tắc nghẽn khi có tấn công xảy ra. Khi một bộ phát hiện Hot-IP phát hiện có Hot-IP sẽ phát cảnh báo cho các bộ phát hiện Hot-IP khác có thiết lập kế nối với nó.

3.3.3. Kịch bản thực nghiệm và kết quả

Dòng gói tin IP được thu thập vào các bộ phát hiện Hot-IP, địa chỉ IP nguồn và IP đích được trích ra từ IP-header ở mỗi gói tin để xử lý. Ở mỗi vị trí triển khai sử dụng ma trận phù hợp với lượng IP quản lý cộng với một số lượng nhỏ IP đại diện cho các đối tượng khác như các ISP khác, các quốc gia hay khu vực.

Các thực nghiệm cài đặt thuật toán cải tiến "Online Hot-IP Preventing", kết quả thực nghiệm cho thấy giải pháp có khả năng ứng dụng để ngăn ngừa sớm các nguy cơ có thể xảy ra các tấn công mạng dạng từ chối dịch vụ DoS/DDoS hay hạn chế các máy quét không gian địa chỉ trên mạng giúp hệ thống hoạt động ổn định, thông suốt.

3.4. Song song hóa giải pháp

3.4.1. Giới thiệu

Trong các ứng dụng thời gian thực, việc xác định nhanh các đối tượng mục tiêu là vô cùng quan trọng. Các cuộc tấn công từ chối dịch vụ có quy mô ngày càng lớn và gây hậu quả nghiêm trọng, cũng như các cuộc quét mạng tìm kiếm lỗ hổng để phát tán sâu Internet diễn ra với tốc độ rất nhanh, thời gian thực hiện quét mạng ngắn.

Áp dụng kỹ thuật xử lý song song để tăng tốc độ tính toán xác định các Hot-IP là một trong những giải pháp hữu hiệu để nâng cao khả năng áp dụng của giải pháp vào thực tế.

3.4.2. Xử lý song song trong bài toán thử nhóm

❖ **Xử lý ở bước thu thập dữ liệu đầu vào:**

Giải pháp phân tán xử lý các dữ liệu đầu vào thời gian thực được sử dụng để giải quyết bài toán xử lý với luồng dữ liệu lớn để tăng thời gian đáp ứng của hệ thống cho các yêu cầu truy xuất bên ngoài hệ thống.

❖ **Xử lý ở bước tính vector kết quả cho các nhóm thử:**

Các nhóm thử trong phương pháp thử nhóm bất ứng biến là độc lập nhau. Như vậy, ở bước xác định các kết quả của các nhóm thử có thể sử dụng kỹ thuật xử lý song song để tối ưu thời gian tính toán kết quả của các nhóm thử.

3.4.3. Kịch bản thực nghiệm và kết quả

❖ **Thực nghiệm xử lý song song dữ liệu đầu vào**

Để xử lý nhanh các luồng dữ liệu rất lớn đối với việc xử lý dữ liệu đầu vào, phần thực nghiệm sử dụng công cụ MapReduce (Hadoop) để thu thập

thông tin IP trên các dữ liệu đầu vào, luồng dữ liệu tổng hợp xử lý theo thuật toán cải tiến "Online Hot-IP Preventing".

❖ **Thực nghiệm xử lý song song ở bước tính toán kết quả**

Kết quả thực nghiệm cho thấy phương pháp xử lý song song cho kết quả giải mã nhanh hơn nhiều so với xử lý tuần tự. Từ đó cho thấy rằng với việc xây dựng giải pháp phát hiện nhanh các Hot-IP trên mạng dùng phương pháp thử nhóm bất ứng biến kết hợp với kỹ thuật xử lý song song cho kết quả rất tốt, có thể áp dụng hiệu quả trong triển khai thực tế trên các mạng tốc độ cao.

CHƯƠNG 4. MỘT SỐ ỨNG DỤNG PHÁT HIỆN CÁC HOT-IP

4.1. Giới thiệu

Bài toán phát hiện các Hot-IP trực tuyến trên mạng là bài toán có tính tổng quát, có thể ứng dụng vào một số bài toán an ninh mạng. Xác định các Hot-IP chính là xác định các đối tượng trên mạng hoạt động với tần suất cao trong một khoảng thời gian rất ngắn. Các đối tượng này có khả năng là nguy cơ ảnh hưởng đến hoạt động của hệ thống mạng, có thể là nguồn phát tấn công hay mục tiêu trong tấn công từ chối dịch vụ, các máy tính đang quét mạng để tìm kiếm lỗ hổng nhằm phát tán sâu mạng của một số loại sâu quét không gian địa chỉ IP.

4.2. Phát hiện các đối tượng có khả năng là mục tiêu, nguồn phát trong tấn công từ chối dịch vụ

4.2.1. Ý nghĩa thực tiễn

Tấn công DoS/DDoS là các cuộc tấn công rất nguy hiểm trên mạng bởi tính đơn giản trong việc thực hiện cuộc tấn công và hậu quả để lại rất nghiêm trọng của nó.

4.2.2. Vấn đề nghiên cứu đặt ra

Giải pháp cân bằng giữa phát hiện khả năng tấn công và phát hiện các đối tượng có khả năng là nguồn phát tấn công có ý nghĩa quan trọng.

Hai bài toán đặt ra: bài toán thứ nhất là phát hiện trực tuyến các khả năng là nguồn phát tấn công từ chối dịch hay nạn nhân trong các cuộc tấn công và bài toán thứ hai là đảm bảo hệ thống mạng hoạt động ổn định, thông suốt bằng cách ngăn chặn các Hot-IP trong một chu kỳ thuật toán.

4.2.3. Mô hình hóa về bài toán phát hiện Hot-IP

Xét dòng gói IP lưu thông qua bộ phát hiện Hot-IP, các gói tin được trích ra địa chỉ IP đích trong IP-header để phân tích. Nếu quan sát các gói dữ liệu đi qua bộ phát hiện Hot-IP mà trong đó có rất nhiều gói có cùng đích đến thì có khả năng địa chỉ IP đó đang bị tấn công từ chối dịch vụ.

Gọi Δ là thời gian của một chu kỳ thuật toán, $N_\Delta (N_\Delta \le N)$ là số lượng địa chỉ IP phân biệt trong khoảng thời gian Δ, m_Δ là số lượng gói tin hệ thống có thể nhận được trong khoảng thời gian Δ. Gọi Hot-List là danh sách lưu các địa chỉ IP nghi ngờ trong quá trình thực thi thuật toán, kích thước của Hot-List được lựa chọn là số lượng mục tiêu trong tấn công DoS/DDoS hoặc số lượng nguồn phát tấn công DoS/DDoS giải pháp có thể phát hiện được, ngưỡng tần suất cao $\delta = \dfrac{m_\Delta}{|\text{Hot-List}|}$.

Áp dụng thuật toán cải tiến 1 "Online Hot-IP detecting" để phát hiện các đối tượng có khả năng là các mục tiêu hoặc các nguồn phát tấn công trong tấn công DoS/DDoS với giải pháp được cài đặt chỉ mang tính chất phát cảnh báo.

Áp dụng thuật toán cải tiến 2 "Online Hot-IP Preventing" để giữ ổn định cho hoạt động của hệ thống bằng cách ngắt kết nối đối với các Hot-IP phát hiện được trong một chu kỳ thực hiện thuật toán.

4.2.4. Kịch bản thực nghiệm và kết quả

Mô hình thực nghiệm 1: Phát hiện các đối tượng có khả năng là nguồn phát tấn công DoS/DDoS

Mô hình thực nghiệm 2: Phòng chống tấn công DDoS

4.3. Phát hiện các đối tượng có khả năng là nguồn phát tán sâu Internet

4.3.1. Ý nghĩa thực tiễn

Sâu mạng hay sâu Internet là những chương trình máy tính độc hại tự nhân bản và phát tán bằng cách khai thác các lỗ hổng của các máy tính trên mạng. Một trong những bước đầu tiên của việc phát tán sâu là quét mạng với tốc độ cao. Chúng tập hợp những danh sách với hàng ngàn địa chỉ IP và quét mạng với tốc độ rất nhanh để tìm kiếm các máy bị lỗ hổng để khai thác.

4.3.2. Vấn đề nghiên cứu đặt ra

Quan sát trên luồng dữ liệu trên mạng, nếu có quá nhiều gói có cùng địa chỉ nguồn thì máy nguồn này có thể đang bị nhiễm sâu và nó đang quét mạng. Có thể mô hình hóa bằng phương pháp thử nhóm để phát hiện nhanh và cảnh báo sớm các sâu đang hoạt động trên mạng bằng cách phân tích dựa vào các gói tin IP và phát hiện các Hot-IP chính là các máy nhiễm sâu đang tiến hành quét mạng.

4.3.3. Mô hình hóa về bài toán phát hiện các Hot-IP

Gọi Δ là chu kỳ thực hiện thuật toán, m_Δ là số lượng gói tin tối đa mà hệ thống có thể nhận được trong khoảng thời gian Δ, Hot-List là danh sách chứa các địa chỉ IP nghi ngờ là nguồn phát tán sâu mạng (Hot-IP) trong quá trình thực hiện thuật toán, kích thước của Hot-List lớn hơn *d* và lựa chọn tùy thuộc vào thực tế và kinh nghiệm của người quản trị.

Ngưỡng tần suất cao δ, ma trận nhị phân được sinh ra dựa vào giá trị N trong khoảng thời gian Δ, suy ra các tham số trong phép nối mã C_{in} và C_{out}. Xác định các địa chỉ IP có khả năng là nguồn phát tán sâu mạng.

4.3.4. Kịch bản thực nghiệm và kết quả

- *Kịch bản 1:* Thực nghiệm để đo thời gian giải mã phát hiện các đối tượng là nguồn quét mạng tìm kiếm lỗ hổng để truyền sâu với khả năng lên đến 700.000 IP phân biệt.
- *Kịch bản 2:* Xử lý dữ liệu thời gian thực, hạn chế tốc độ lây lan bằng cách ngắt kết nối đối với các Hot-IP này trong một chu kỳ thuật toán (bằng thuật toán cải tiến "Online Hot-IP Preventing").

Qua kết quả thực nghiệm thấy rằng thời gian phát hiện các đối tượng có khả năng là nguồn phát tán sâu mạng nhanh, hệ thống hoạt động ổn định do các nguồn phát tán bị ngăn chặn trong một chu kỳ thuật toán, giải pháp có thể áp dụng triển khai vào hệ thống mạng thực tế.

4.4. Phát hiện khả năng các thiết bị hoạt động bất thường

4.4.1. Ý nghĩa thực tiễn

Các bất thường có thể là tình trạng hoạt động của các máy chủ trong hệ thống hoạt động quá mức có thể đang bị tấn công hoặc mức độ phục vụ chập chờn dưới mức bình thường có thể do hư hỏng.

4.4.2. Vấn đề nghiên cứu đặt ra

Bài toán đặt ra là trong môi trường mạng ở phía các nhà cung cấp dịch vụ làm sao có thể giám sát để cảnh báo sớm các server đang hoạt động bất thường về mặt truy cập như các hệ thống máy chủ chia tải cho các ứng dụng trên mạng.

4.4.3. Mô hình hóa về bài toán phát hiện Hot-IP

Giả sử tổng tần suất xuất hiện của *N* IP trên dòng dữ liệu là *S* và có nhiều nhất là *d* IP có tần suất hoạt động bất thường (*Hot-IP* hoặc *Low-IP*). Một IP được coi là bình thường nếu tần suất xuất hiện của nó nhỏ hơn $\frac{S}{d+1}$ và lớn hơn $\frac{S}{(N-2d+1)(d+1)}$. Giả sử chúng ta có nhiều nhất là *d Hot-IP* và *d Low-IP* nên các IP bình thường trong dòng dữ liệu là $N-2d$. Tổng các tần suất xuất hiện của chúng lớn hơn $(N-2d).\frac{S}{(N-2d+1)(d+1)}$. Do đó, tổng các tần suất của các IP *Low-IP* nhỏ hơn $\frac{S}{(N-2d+1)(d+1)}$.

4.4.4. Kịch bản thực nghiệm và kết quả

Trong phần thực nghiệm phát hiện các bất thường trên dòng dữ liệu, Thuật toán bắt gói được cài đặt bằng ngôn ngữ C, sử dụng thư viện *pcap* để

phân tích các gói tin. Khi một gói tin gửi đến, phần IP-header được phân tích và rút trích thông tin về địa chỉ IP nguồn trong đó.

Kết quả thực nghiệm cho thấy rằng giải pháp phát hiện các Hot-IP có thể triển khai áp dụng trong hệ thống mạng thực tế để phát hiện các thiết bị hoạt động bất thường.

4.5. Giám sát các Hot-IP

4.5.1. Ý nghĩa thực tiễn

Giám sát hoạt động của các thiết bị quan trọng trên mạng là một trong những nhiệm vụ trọng tâm đối với người quản trị mạng. Mục đích của công việc này là theo dõi tình trạng hoạt động của chúng trong hệ thống.

4.5.2. Vấn đề nghiên cứu đặt ra

Vấn đề đặt ra cho việc giám sát ở các nhà cung cấp dịch vụ theo dõi và phát cảnh báo cho khả năng về tình trạng các máy chủ hay các thiết bị hoạt động bất thường trong hệ thống mà không cần cài đặt chế độ giám sát từ máy chủ giám sát đến các thiết bị và server này.

4.5.3. Kịch bản thực nghiệm và kết quả

Các Hot-IP được phát hiện và được đưa vào trạng thái giám sát, xác định tần suất xuất hiện của chúng trong một khoảng thời gian định trước, thể hiện trên đồ thị trực quan giúp người quản trị trong việc theo dõi hoạt động của chúng. Người quản trị có thể đặt ngưỡng để phát cảnh báo hay điều phối các luồng lưu lượng của dòng gói IP chứa các Hot-IP này.

Qua việc theo dõi trạng thái hoạt động của các thiết bị định tuyến, chuyển mạch trên mạng, hệ thống giám sát có thể kích hoạt tường lửa để hạn chế hay ngăn chặn các Hot-IP này.

KẾT LUẬN

Luận án trình bày giải pháp phát hiện các Hot-IP trên mạng trực tuyến dựa trên phương pháp thử nhóm bất ứng biến. Mục tiêu của chính của luận án là đề xuất giải pháp phát hiện nhanh các Hot-IP trên mạng, ứng dụng trên mạng trung gian ở các nhà cung cấp dịch vụ hoặc các mạng cung cấp dịch vụ

trên Internet nhằm giúp người quản trị phát hiện nhanh, ứng phó kịp thời với các khả năng là nguy cơ ảnh hưởng xấu đến hoạt động của mạng và đảm bảo hệ thống hoạt động ổn định, thông suốt.

Trong các giải pháp đã được khảo sát, giải pháp phát hiện các Hot-IP trên mạng sử dụng thử nhóm bất ứng biến là giải pháp hữu hiệu để triển khai áp dụng nhằm phát hiện trực tuyến và hạn chế tác hại của các đối tượng là nguy cơ gây hại trên mạng.

Đây là giải pháp phù hợp nhất trong bài toán phát hiện các đối tượng có khả năng là các nạn nhân trong tấn công từ chối dịch vụ; các đối tượng đang phát động tấn công DoS/DDoS; các đối tượng đang phát tán sâu mạng, giải pháp này cân bằng giữa thời gian tính toán và phát hiện các đối tượng ngay ở bước phát hiện tấn công. Bên cạnh đó, giải pháp này có khả năng tối ưu năng lực xử lý về bộ nhớ, đơn giản, chính xác, tính toán nhanh, phù hợp để áp dụng trên môi trường mạng có số lượng người dùng rất lớn như môi trường mạng trung gian ở phía nhà cung cấp dịch vụ.

1. Các kết quả đạt được

(1) Luận án đã mô hình hóa bài toán phát hiện các Hot-IP sang bài toán thử nhóm bất ứng biến và đề xuất kết hợp một số kỹ thuật để nâng cao hiệu quả của giải pháp. Trong đó, phương pháp nối mã được áp dụng vào việc phát sinh ma trận d-phân-cách tường minh để phát sinh chính xác ma trận và tối ưu không gian lưu trữ khi thực thi chương trình. Phương pháp này cho phép phát sinh từng cột của ma trận trong quá trình xử lý và tính toán. Do đó, ma trận không cần được lưu trữ toàn bộ trong khi thực thi chương trình. Để nâng cao hiệu quả của giải pháp như lựa chọn các tham số, kích thước ma trận dựa vào khả năng của vị trí triển khai; đề xuất kết hợp với kỹ thuật xử lý song song để giảm thời gian giải mã phát hiện các Hot-IP dựa vào tính chất của phương pháp thử nhóm bất ứng biến là các phép thử được xác định trước và độc lập nhau; đề xuất kết hợp với kiến trúc phân tán để phát hiện và cảnh báo sớm các

Hot-IP trong hệ thống mạng tổ chức đa vùng, thích hợp áp dụng trong các mạng trung gian ở phía nhà cung cấp dịch vụ.

(2) Luận án đã đề xuất cải tiến phương pháp thử nhóm bất ứng biến trong việc phát hiện các Hot-IP với hai thuật toán cải tiến. Thuật toán cải tiến thứ nhất *"Online Hot-IP Detecting"* cho phép tối ưu về mặt tính toán và độ chính xác khi số lượng Hot-IP thực tế cao hơn giá trị cho trước khi xây dựng ma trận. Thuật toán cải tiến thứ hai *"Online Hot-IP Preventing"* đảm bảo hệ thống mạng hoạt động ổn định, thông suốt bằng cách ngắt kết nối đối với các Hot-IP trong một chu kỳ thực hiện thuật toán.

(3) Luận án đã mô hình hóa một số bài toán an ninh mạng như phát hiện nguồn phát tán sâu mạng, phát hiện các nạn nhân và nguồn phát tấn công trong các cuộc tấn công từ chối dịch vụ, phát hiện các thiết bị hoạt động bất thường trên mạng về bài toán tìm Hot-IP. Bên cạnh đó, luận án đề xuất giám sát các Hot-IP này kết hợp với việc theo dõi tài nguyên hệ thống để điều phối hoạt động của luồng lưu lượng chứa Hot-IP, giảm ảnh hưởng xấu đến hoạt động chung của toàn hệ thống mạng.

Các kết quả nghiên cứu và thực nghiệm cho thấy rằng giải pháp cho kết quả tốt, thời gian thực hiện để phát hiện các Hot-IP nhanh, có thể áp dụng triển khai vào môi trường thực tế ở phía các nhà cung cấp dịch vụ và các hệ thống mạng cung cấp dịch vụ trên môi trường Internet.

2. Hướng phát triển

Luận án đã trình bày một giải pháp hoàn chỉnh về phát hiện các Hot-IP trên mạng và một số ứng dụng trong lĩnh vực an ninh mạng. Bên cạnh việc áp dụng giải pháp vào thực tiễn, đặc biệt triển khai trên phần cứng, hướng nghiên cứu mở tiếp theo là kết hợp phân tích một số yếu tố khác trong dòng dữ liệu để nhận dạng, phân loại nguy cơ từ bài toán phát hiện các Hot-IP này.

CÁC CÔNG TRÌNH NGHIÊN CỨU CỦA TÁC GIẢ

TẠP CHÍ KHOA HỌC

1. **Huynh Nguyen Chinh**, Nguyen Dinh Thuc, Tan Hanh. Finding Hot-IPs in network using group testing method – A review. *Journal of Engineering Technology and Education – Kuas,Taiwan,* pp.374-379, 2013.
2. **Huynh Nguyen Chinh**, Nguyen Dinh Thuc, Tan Hanh. Group testing for detecting worms in computer networks. *Tạp chí Khoa học và Công nghệ - chuyên san các công trình nghiên cứu về Điện tử, Viễn thông và CNTT*, pp.12-19, 2013.
3. **Huynh Nguyen Chinh**, Tan Hanh, and Nguyen Dinh Thuc. Fast detection of DDoS attacks using Non-Adaptive group testing. *International Journal of Network Security and Its Applications (IJNSA), Vol.5 (5),* pp. 63–71, India, 2013.
4. **Huynh Nguyen Chinh**. Fast detecting Hot-IPs in high speed networks. Tạp chí Phát Triển KH-CN, chuyên san KHTN, ĐHQG Tp.HCM, Vol 18, pp.242-253, 2015.

HỘI NGHỊ KHOA HỌC QUỐC TẾ

5. Thach V. Bui, **Chinh N. Huynh**, Thuc D. Nguyen. Early detection for networking anomalies using Non-Adaptive Group testing. *International Conference on ICT Convergence 2013 (ICTC 2013), Korea,* pp. 984-987, IEEE, 2013.
6. **Huynh Nguyen Chinh**, Nguyen Dinh Thuc, Tan Hanh. A distributed architecture and Non-adaptive Group testing approach to fast detect Hot-IPs in ISP networks. *International Conference on Green and Human Information Technology (ICGHIT 2014),* pp.232-236, IEEE, 2014.

7. **Huynh Nguyen Chinh**, Nguyen Dinh Thuc, Tan Hanh. Early detection and limitation Hot-IPs using Non-adaptive group testing and dynamic firewall rules. *International Conference on Computing, Management and Telecommunications (ComManTel 2014)*, pp. 286-290, IEEE, 2014.
8. **Huynh Nguyen Chinh**, Nguyen Dinh Thuc, Tan Hanh. Monitoring Hot-IPs in high speed networks. *The 2014 International Conference on Advanced Technologies for Communications (ATC'14),* pp. 430-434, IEEE, 2014.