HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Vũ Thị Quý

NGHIÊN CỨU PHÁT HIỆN TẤN CÔNG WEB CƠ BẢN DỰA TRÊN HỌC MÁY SỬ DỤNG WEB LOG

Chuyên ngành : KHOA HỌC MÁY TÍNH

Mã số : 8.48.01.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

HÀ NỘI – NĂM 2020

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. HOÀNG XUÂN DẬU

Phản biện 1: ……

Phản biện 2:……

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng ..... năm 2020

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

PHẦN MỞ ĐẦU

Trong thế giới hiện đại ngày nay, ứng dụng web ngày một trở nên quan trọng và là một phần không thể thiếu trên mạng Internet. Các ứng dụng web, website chiếm tỷ lệ áp đảo trong số các ứng dụng trên nền Internet. Cũng chính vì vậy mà vấn đề về bảo mật web ngày càng trở thành một vấn đề được quan tâm.

Theo số liệu thống kê của BKAV [11], năm 2019, thiệt hại do virus máy tính gây ra đối với người dùng Việt Nam đã lên tới 20.892 tỷ đồng (902 triệu USD), vượt xa con số 14.900 tỷ đồng của năm 2018. Tổng số lượt máy tính bị nhiễm mã độc được ghi nhận trong năm 2019 lên tới 85,2 triệu lượt, tăng 3,5% so với năm 2018. Năm này cũng tiếp tục chứng kiến sự hoành hành của các loại mã độc mã hóa dữ liệu tống tiền (ransomware). Số lượng máy tính bị mất dữ liệu trong năm 2019 lên tới 1,8 triệu lượt, tăng 12% so với năm 2018. Nghiêm trọng hơn, trong số này có rất nhiều máy chủ (server) chứa dữ liệu của các cơ quan. Không chỉ gây thiệt hại lớn, việc các máy chủ bị xóa dữ liệu cũng gây đình trệ hoạt động của cơ quan, doanh nghiệp trong nhiều ngày sau đó, thậm chí đến cả tháng.

Đối với các công ty lớn, nguy cơ bị tấn công vào hệ thống đồng nghĩa với việc họ sẽ bị thiệt hại hàng tỷ USD, uy tín trước khách hàng bị giảm sút. Với các cơ quan y tế và quốc phòng thì thiệt hại còn có thể thảm khốc hơn gấp nhiều lần.

Qua số liệu trên cho thấy tấn công web cơ bản là các dạng tấn công thường gặp lên các website, web portal và các ứng dụng trên nền web. Các dạng tấn công này có thể bao gồm: tấn công chèn mã SQL (SQLi hay SQL Injection), tấn công chèn mã XSS (Cross-Site Scripting), tấn công duyệt đường dẫn (Path traversal) và tấn công chèn dòng lệnh hệ điều hành (CMDi hay Command Injection). Trong số này, tấn công chèn mã SQL là một trong các dạng tấn công phổ biến và nguy hiểm nhất. Tùy vào mức độ tinh vi, tấn công chèn mã SQL có thể cho phép kẻ tấn công (1) vượt qua các khâu xác thực người dùng, (2) chèn, sửa đổi, hoặc xóa dữ liệu, (3) đánh cắp các thông tin trong cơ sở dữ liệu và (4) chiếm quyền điều khiển hệ thống máy chủ cơ sở dữ liệu [1]. Tấn công XSS có thể cho phép tin tặc đánh cắp dữ liệu người dùng lưu trong cookie của trình duyệt, từ đó kiểm soát tài khoản của người dùng trên máy chủ. Theo một hướng khác, tấn công duyệt đường dẫn cho phép tin tặc tải hoặc truy nhập vào các file chứa dữ liệu nhạy cảm trên các máy chủ và thông qua đó có thể xâm nhập sâu vào hệ thống. Tấn công chèn dòng lệnh hệ điều hành có thể cho phép tin tặc thực hiện các lệnh nguy hiểm cho phép xóa file, dữ liệu trên hệ thống nạn nhân.

Mặc dù các dạng tấn công thường gặp lên các website và các ứng dụng trên nền web đã được biết đến từ lâu và đã có nhiều biện pháp phòng chống được nghiên cứu, triển khai, như sử dụng các bộ lọc, tường lửa, các cơ chế kiểm soát truy nhập… Tuy nhiên, các dạng tấn công web cơ bản vẫn khá phổ biến và gây nhiều thiệt hại cho các trang web, các cổng thông tin điện tử, các trang thương mại điện tử của các cơ quan tổ chức. Nguyên nhân của điều này là do vẫn có nhiều website và các ứng dụng trên nền web không có, hoặc thiếu cơ chế lọc dữ liệu đầu vào thực sự hiệu quả, và/hoặc sử dụng các mã chương trình trộn lẫn với dữ liệu, tạo điều kiện cho tin tặc chèn mã độc tấn công hệ thống [1]. Việc xây dựng các bộ lọc dựa trên các mẫu cố định thực sự gặp khó khăn, khi các mẫu tấn công liên tục thay đổi và ngày càng tinh vi hơn. Việc xây dựng các bộ lọc phát hiện các dạng tấn công web cơ bản dựa trên học máy là một hướng giải quyết hiệu quả thay thế cho các bộ lọc mẫu truyền thống. Theo hướng nghiên cứu này, đề tài luận văn thạc sĩ của học viên có tên **"Nghiên cứu phát hiện tấn công web cơ bản dựa trên học máy sử dụng web log"** tập trung nghiên cứu vấn đề phát hiện tấn công web cơ bản dựa trên học máy sử dụng web log.

Do còn nhiều hạn chế về thời gian và tài liệu nên đề tài còn nhiều thiếu sót. Rất mong nhận được sự đóng góp của các thầy cô và các bạn để đề tài được hoàn thiện hơn.

Tôi xin chân thành cảm ơn!

Tổng quan về vấn đề nghiên cứu

Đã có nhiều giải pháp phòng chống các dạng tấn công web cơ bản được nghiên cứu và ứng dụng [1][3]. Các giải pháp thực tế có thể kể đến gồm:

- Sử dụng các bộ lọc để kiểm tra và lọc dữ liệu đầu vào. Các bộ lọc có thể sử dụng bao gồm, lọc dựa trên từ khóa, lọc dựa trên mẫu và lọc dựa trên biểu thức chính quy.
- Sử dụng các dạng tường lửa, hoặc proxy ở mức ứng dụng, như tường lửa ứng dụng web (WAF – Web Application Firewall). WAF được sử dụng để lọc tất cả truy vấn của người dùng. WAF có ưu điểm là có thể bảo vệ đồng thời nhiều website và không đòi hỏi chỉnh sửa mã nguồn của website.
- Kết hợp sử dụng các biện pháp kiểm soát truy nhập, phân quyền người dùng để giảm thiểu khả năng bị tấn công, khai thác.

- Sử dụng các công cụ theo dõi, giám sát website, ứng dụng web, như các bộ phát hiện xâm nhập (IDS).

Trên phương diện nghiên cứu học thuật, có thể chia các đề xuất nghiên cứu phát hiện tấn công, xâm nhập nói chung và tấn công web cơ bản nói riêng thành 2 nhóm dựa trên kỹ thuật phát hiện: (1) nhóm phát hiện dựa trên chữ ký, mẫu, hoặc luật và (2) nhóm phát hiện dựa trên bất thường.

Phát hiện dựa trên chữ ký (signature), mẫu (pattern), hoặc luật (rule) là phương pháp phát hiện tấn công dựa trên việc tìm hay so khớp tập chữ ký của các tấn công đã biết với các dữ liệu giám sát thu thập được. Một tấn công được phát hiện khi có ít nhất một so khớp chữ ký thành công. Kỹ thuật phát hiện tấn công, xâm nhập dựa trên chữ ký có ưu điểm là có khả năng phát hiện nhanh và chính xác các dạng tấn công đã biết. Tuy nhiên, kỹ thuật này có nhược điểm là không có khả năng phát hiện các dạng tấn công mới, hay tấn công khai thác lỗ hổng zero-day do chữ ký của chúng chưa tồn tại trong cơ sơ dữ liệu. Ngoài ra, việc xây dựng và cập nhật cơ sở dữ liệu chữ ký thường được thực hiện thủ công, nên tốn nhiều công sức.

Phát hiện tấn công, xâm nhập dựa trên bất thường dựa trên giả thiết: *các hành vi tấn công, xâm nhập thường có quan hệ chặt chẽ với các hành vi bất thường*. Quá trình xây dựng và triển khai một hệ thống phát hiện xâm nhập dựa trên bất thường gồm 2 giai đoạn: (1) huấn luyện và (2) phát hiện. Trong giai đoạn huấn luyện, hồ sơ (profile) của đối tượng trong chế độ làm việc bình thường được xây dựng. Để thực hiện giai đoạn huấn luyện này, cần giám sát đối tượng trong một khoảng thời gian đủ dài để thu thập được đầy đủ dữ liệu mô tả các hành vi của đối tượng trong điều kiện bình thường làm dữ liệu huấn luyện. Tiếp theo, thực hiện huấn luyện dữ liệu để xây dựng mô hình phát hiện, hay hồ sơ của đối tượng. Trong giai đoạn phát hiện, thực hiện giám sát hành vi hiện tại của hệ thống và cảnh báo nếu có khác biệt rõ nét giữa hành vi hiện tại và các hành vi lưu trong hồ sơ của đối tượng. Ưu điểm của phát hiện xâm nhập dựa trên bất thường là có tiềm năng phát hiện các loại tấn công, xâm nhập mới mà không yêu cầu biết trước thông tin về chúng. Tuy nhiên, phương pháp này có tỷ lệ cảnh báo sai tương đối cao so với phương pháp phát hiện dựa trên chữ ký. Điều này làm giảm khả năng ứng dụng thực tế của phát hiện xâm nhập dựa trên bất thường. Ngoài ra, nó cũng tiêu tốn nhiều tài nguyên hệ thống cho việc xây dựng hồ sơ đối tượng và phân tích hành vi hiện tại.

Phương pháp phát hiện tấn công web cơ bản dựa trên học máy sử dụng web log thực hiện trong luận văn thuộc nhóm kỹ thuật phát hiện dựa trên bất thường. Theo đó, các URI truy nhập được tách ra từ web log và được phân loại bởi một bộ phân loại đã được huấn luyện sử dụng tập dữ liệu đã được gán nhãn. Luận văn dự kiến sử dụng các thuật toán học máy có giám sát nên có thể giảm thời gian huấn luyện và phát hiện.

Mục đích nghiên cứu

Luận văn nghiên cứu một số thuật toán học máy có giám sát và ứng dụng cho việc phát hiện tấn công web cơ bản sử dụng web log.

Trên cơ sở đó tiến hành thực nghiệm để đánh giá hiệu quả trong việc phát hiện tấn công web cơ bản của một số thuật toán học máy.

Đối tượng và phạm vi nghiên cứu

Đối tượng và phạm vi nghiên cứu của luận văn bao gồm:

- Một số dạng tấn công web cơ bản

- Một số thuật toán học máy để phát hiện tấn công

- Web log

- Một số công cụ, phần mềm để thử nghiệm và đánh giá hiệu quả các thuật toán học máy trong phát hiện dựa trên Web log.

Phương pháp nghiên cứu

- Phương pháp lý thuyết: Khảo sát, phân tích các tài liệu khoa học liên quan đến các dạng tấn công và một số thuật toán học máy.

- Phương pháp thực nghiệm: Sử dụng các công cụ, phần mềm để thử nghiệm và đánh giá hiệu quả các thuật toán học máy trong phát hiện dựa trên web log đối với bộ dữ liệu được lựa chọn.

Trên cơ sở đó đánh giá được các ưu nhược điểm và từ đó định hướng xây dựng định hướng nghiên cứu, cải tiến mở rộng quy trình, phương pháp.

Cấu trúc luận văn

Ngoài phần mở đầu và kết luận, luận văn được chia thành 3 chương:

Chương 1: Tổng quan về các dạng tấn công vào Website, ứng dụng Web và các giải pháp phòng chống

Chương này trình bày về kiến trúc ứng dụng web, các yêu cầu về bảo mật, các hình thức tấn công vào ứng dụng web cũng như cách phòng chống bị tấn công của các hình thức tấn công phổ biến trong các năm gần đây dựa theo OWASP. Phần cuối của chương là các biện pháp bảo mật ứng dụng web, bao gồm nguyên tắc chung và một số biện pháp bảo mật cụ thể cho ứng dụng web.

Chương 2: Phát hiện tấn công WEB dựa trên học máy sử dụng Web

Trong chương 2, luận văn sẽ tiếp tục đi tìm hiểu về WEBLOG, khái quát và các dạng, đồng thời đi sâu vào việc giới thiệu học máy và các thuật toán học máy, đưa ra mô hình phát hiện tấn công website và chi tiết các khâu xử lý dữ liệu.

Chương 3: Cài đặt và thử nghiệm

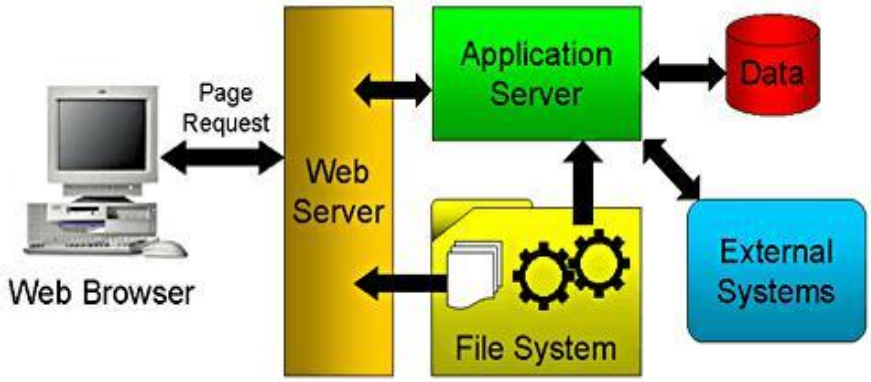
Trong chương 3, nội dung chủ yếu là giới thiệu tập dữ liệu được sử dụng để huấn luyện cho học máy, cách xử lý tiền dữ liệu, các bước làm trong pha huấn luyện và phân loại các dữ liệu đầu vào. Trình bày một số phương pháp để sử dụng huấn luyện và phát hiện, các kết quả sẽ được dùng để đánh giá mức độ hiệu quả khi sử dụng các phương pháp học máy khác nhau.

CHƯƠNG 1: TỔNG QUAN VỀ CÁC DẠNG TẤN CÔNG VÀO WEBSITE, ỨNG DỤNG WEB VÀ CÁC GIẢI PHÁP PHÒNG CHỐNG

1.1. *Kiến Trúc Ứng Dụng Web và Các Yêu Cầu Bảo Mật*

1.1.1 Kiến trúc ứng dụng web

Một ứng dụng web (Web application) có thể gồm các thành phần: Máy khách web/trình duyệt web (Web client/web browser), Máy chủ web (HTTP/web server), URL/URI, Web session và cookie, Bộ diễn dịch và thực hiện các server script, Các server script (CGI – Common Gateway Interface), Máy chủ cơ sở dữ liệu và Hạ tầng mạng TCP/IP kết nối giữa máy khách và máy chủ web.



Hình 1.1: Kiến trúc chuẩn của ứng dụng web [1]

Hình 1.1 biểu diễn kiến trúc chuẩn của hệ thống ứng dụng web (hay ngắn gọn là ứng dụng web), trong đó mô tả các thành phần của một ứng dụng web và giao tiếp giữa chúng. Theo đó, các thành phần của một ứng dụng web gồm Web Browser (Trình duyệt), Web Server (Máy chủ web), Application Server (Máy chủ ứng dụng), Data (Kho chứa dữ liệu – thường là cơ sở dữ liệu), File System (Hệ thống file trên máy chủ) và External System (Các hệ thống bên ngoài). Web Browser tạo và gửi yêu cầu về trang web (Page Request) đến Web Server. Nếu đó là yêu cầu trang web tĩnh, Web Server sẽ đọc nội dung trang từ File System và gửi trang web cho Web Browser. Nếu đó là yêu cầu trang web động, Web Server sẽ chuyển yêu cầu cho Application Server xử lý. Application Server sẽ dịch và thực hiện mã script trong trang web để tạo kết quả. Application Server có thể cần truy nhập Data, File System, hoặc External System để xử lý yêu cầu. Kết quả xử lý yêu cầu được chuyển lại cho Web Server để tạo trang web và gửi cho Web Browser.

Các máy chủ web phổ biến hiện này có thể kể tới là Apache, Nginx, IIS, Tomcat… Các ứng dụng web thì tùy thuộc vào yêu cầu triển khai mà có thể được tạo nên bởi các ngôn ngữ lập trình khác nhau như: C#, Java, Python, PHP, Ruby… Cơ sở dữ liệu (Data) sẽ đóng vai trò lưu trữ, cung cấp thông tin cho ứng dụng web trong quá trình xử lý request. Một số hệ quản trị cơ sở dữ liệu thường được sử dụng bao gồm: SQL Server, MySQL, MongoDB, Oracle. Ngoài ra, tùy thuộc vào độ phức tạp, quy mô, yêu cầu trong việc phát triển mà website có thể có thêm nhiều thành phần khác như Message Queue, Proxy, Cache.

- Giao thức HTTP
- Giao thức HTTPS

Máy khách web và máy chủ web giao tiếp với nhau bằng giao thức HTTP hoặc HTTPS thông qua phương thức yêu cầu/đáp ứng (Request/Response), trong đó yêu cầu là http request gửi từ máy khách web lên máy chủ web và đáp ứng hay phản hồi là http response gửi từ máy chủ web tới máy khách web.

```
POST
/gen_204?atyp=i&ei=4BTKXYUJ7NfPuw-pk6oY&ct=slh&v=2&m=HV&t=C&s=1&pv=0.348789283644
0,R,1,7,20,28,92,33:0,R,1,CAcQAA,166,172,600,315:0,R,1,CAcQAQ,166,172,600,95:6,B,
Host: www.google.com
User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:60.0) Gecko/20100101 Firefox/60.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: https://www.google.com/
Cookie: 1P_JAR=2019-11-12-02;
NID=191=VR0UnwUg5UXmMB8wNm-9wCgxlFOelFZY5rXJcDkFgt_1NwvXxr6REZ7oRfYLC72lVoPBY1aST
kc5ALiNhsRtkfq60f1VL2PGNdU; ANID=AHWqTUlwgbWb26DpDUzd_xlWeZM67Wpa4SHR1j_lHUQAAXWR
Connection: close
Content-Length: 0
```

Hình 1.2: Cấu trúc của http request

```
HTTP/1.1 204 No Content
Content-Type: text/html; charset=UTF-8
Date: Tue, 12 Nov 2019 02:11:54 GMT
Server: gws
Content-Length: 0
X-XSS-Protection: 0
X-Frame-Options: SAMEORIGIN
Set-Cookie: 1P_JAR=2019-11-12-02; expires=Thu, 12-Dec-2019 02:11:54 GMT; path=/;
Alt-Svc: quic=":443"; ma=2592000; v="46,43",h3-Q050=":443"; ma=2592000,h3-Q049="
ma=2592000,h3-Q043=":443"; ma=2592000
Connection: close
```

Hình 1.3: Cấu trúc của http reponse

1.1.2 Các yêu cầu bảo mật ứng dụng web, website

1.1.2.1. *Yêu cầu về cài đặt*

1.1.2.2. *Tắt/disable các thành phần mặc định*

1.1.2.3. *Thay đổi các thành phần mặc định*

1.1.2.4. *Giới hạn truy cập*

1.2. *Các Nguy Cơ và Các Dạng Tấn Công Lên Ứng Dụng Web*

1.2.1 Các nguy cơ và các lỗ hổng bảo mật trong website, ứng dụng web (TOP 10 OSWAP 2017)

OWASP Top 10 năm 2017 được phát hành công khai, dựa trên cuộc thăm dò, kiểm tra hơn 2,3 triệu lỗ hổng tác động đến 50000 ứng dụng, bao gồm 2 bản cập nhật lỗ hổng quy mô lớn và cập nhật các kịch bản tấn công mới. Phần tiếp theo mô tả danh sách Top 10 của năm 2017 [8][9][10].

A1 – Injection (Lỗi nhúng mã)

A2 - Broken Authentication and Session Management (Lỗi xác thực và quản phiên yếu)

A3 - Cross-Site Scripting (XSS)

A4 - Broken Access Control (Điều khiển truy nhập yếu)

A5 - Security Misconfiguration (Cấu hình thiếu an toàn)

A6 - Sensitive data exposure (Rò rỉ dữ liệu nhạy cảm)

A7 - Missing function level access control (Lỗi phân quyền)

A8 - Cross Site Request Forgery (CSRF)

A9 - Using component with known vulnerabilities (Sử dụng những thư viện, ứng dụng tồn tại lỗ hổng trước đó)

A10- Underprotected APIs (Các API không được bảo vệ)

1.2.2 Một số dạng tấn công web cơ bản

1.2.2.1. *Tấn công chèn mã SQLi*

Tấn công chèn mã SQL (SQL Injection - SQLi) là một kỹ thuật cho phép kẻ tấn công chèn mã SQL vào dữ liệu gửi đến máy chủ và cuối cùng được thực hiện trên máy chủ cơ sở dữ liệu. Tùy vào mức độ tinh vi, tấn công chèn mã SQL có thể cho phép kẻ tấn công (1) vượt qua các khâu xác thực người dùng, (2) chèn, sửa đổi, hoặc xóa dữ liệu, (3) đánh cắp các thông tin trong cơ sở dữ liệu và (4) chiếm quyền điều khiển hệ thống máy chủ cơ sở dữ liệu. Tấn công chèn mã SQL là dạng tấn công thường gặp ở các ứng dụng web, các trang web có kết nối đến cơ sở dữ liệu.

Có 2 nguyên nhân của lỗ hổng trong ứng dụng cho phép thực hiện tấn công chèn mã SQL:

- Dữ liệu đầu vào từ người dùng hoặc từ các nguồn khác không được kiểm tra hoặc kiểm tra không kỹ lưỡng;
- Sử dụng các câu lệnh SQL động trong ứng dụng, trong đó có thao tác nối dữ liệu người dùng với mã lệnh SQL gốc.

1.2.2.2. *Tấn công Cross-Site Scriting (XSS)*

Tấn công Cross-Site Scriting (XSS – Mã script liên site, liên miền) là một trong các dạng tấn công phổ biến nhất vào các ứng dụng web. XSS xuất hiện từ khi trình duyệt bắt đầu hỗ trợ ngôn ngữ JavaScript (ban đầu được gọi là LiveScript – trên trình duyệt Netscape). Mã tấn công XSS được nhúng trong trang web chạy trong lòng trình duyệt với quyền truy nhập của người dùng, có thể truy nhập các thông tin nhạy cảm của người dùng lưu trong trình duyệt. Do mã XSS chạy trong lòng trình duyệt nên nó miễn nhiễm với các trình quét các phần mềm độc hại và các công cụ bảo vệ hệ thống [6].

XSS có thể được xem là một dạng của chèn mã HTML (HTML Injection). Trên thực tế, có thể thực hiện tấn công bằng chèn mã HTML mà không cần mã JavaScript và cũng không cần liên site, hoặc liên miền. Kẻ tấn công khai thác các lỗ hổng bảo mật để chèn mã XSS vào trang web, trong đó dữ liệu web (như tên và địa chỉ email) và mã (cú pháp và các phần tử như <script>) của XSS được trộn lẫn vào mã gốc của trang web.

Tấn công XSS thường xuất hiện khi trang web cho phép người dùng nhập dữ liệu và sau đó hiển thị dữ liệu lên trang. Kẻ tấn công có thể khéo léo chèn mã script vào trang và mã script của kẻ tấn công được thực hiện khi người dùng khác thăm lại trang web đó.

Có thể chia tấn công XSS thành 3 loại chính: Stored XSS (XSS lưu trữ), Reflected XSS (XSS phản chiếu) và DOM-based/Local XSS (XSS dựa trên DOM hoặc cục bộ)

1.2.2.3. *Duyệt đường dẫn (Directory traversal)*

Directory Traversal là một dạng tấn công cho phép tin tặc truy cập đến những chỉ mục bị giới hạn, thực thi lệnh bên ngoài chỉ mục gốc của máy chủ web. Hình thức tấn công này không cần sử dụng một công cụ nào mà chỉ đơn thuần là thao tác với các biến ../ (dot-dot-slash) để truy cập đến các file, thư mục, bao gồm cả source code, những file hệ thống…

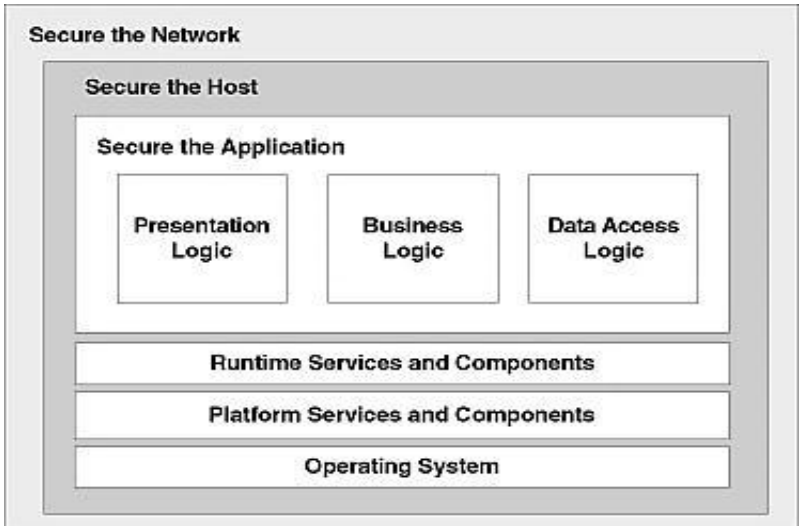
1.2.2.4. *Tấn công CMDi*

OS Command Injection (CMDi) là một lỗ hổng bảo mật web cho phép kẻ tấn công có thể thực thi các lệnh của hệ điều hành (OS) tùy ý trên máy chủ đang chạy service nào đó. Các lỗ hổng CMDi xảy ra khi phần mềm tích hợp dữ liệu do người dùng quản lý trong một lệnh, các dữ liệu này được xử lý trong trình thông dịch lệnh. Nếu dữ liệu không được kiểm tra, một hacker có thể sử dụng các ký tự đặc biệt để thay đổi lệnh đang được thực thi từ đó kẻ tấn công có thể khai thác, truy xuất thông tin, tấn công sang các hệ thống máy chủ khác trong cùng vùng mạng.

1.2.3 Các biện pháp bảo mật ứng dụng web, website

1.2.3.1. *Nguyên tắc chung*

Nguyên tắc bảo mật ứng dụng web tuân theo nguyên tắc chung của bảo mật an toàn hệ thống thông tin là phòng vệ nhiều lớp theo chiều sâu (Defense in depth). Hình 1.4 biểu diễn 3 lớp bảo mật ứng dụng web: Lớp bảo mật mạng (Network), Lớp bảo mật máy chủ (Host) và Lớp bảo mật ứng dụng (Application). Trong đó, lớp bảo mật mạng thực hiện bảo vệ ở vòng ngoài, lớp bảo mật máy chủ thực hiện bảo vệ nền tảng và lớp bảo mật ứng dụng thực hiện bảo vệ dữ liệu thông qua kiểm soát quyền truy nhập.



Hình 1.4: Các lớp bảo mật ứng dụng web [1]

Tiếp theo từng lớp bảo mật cũng phải đảm bảo từng nhiệm vụ cụ thể:

❖ Lớp bảo mật mạng đảm bảo cung cấp hạ tầng mạng an toàn cho giao tiếp giữa máy chủ và máy khách.

❖ Lớp bảo mật máy chủ (Host) có nhiệm vụ đảm bảo an toàn cho các thành phần nền tảng trong hệ thống.

❖ Lớp bảo mật ứng dụng có trách nhiệm đảm bảo an toàn cho người dùng và dữ liệu của người dùng lưu trong hệ thống ứng dụng web.

1.2.3.2. *Một số biện pháp bảo mật cụ thể*

1.2.3.2.1. Kiểm tra dữ liệu đầu vào

1.2.3.2.2. Giảm thiểu các giao diện có thể bị tấn công

1.2.3.2.3. Phòng vệ theo chiều sâu

1.3. *Kết luận Chương 1*

Chương 1 đã giới thiệu tổng quan về kiến trúc ứng dụng web, các yêu bảo mật đối với ứng dụng web, web server. Chương này cũng đã giới thiệu các lỗ hổng nằm trong TOP 10 OWASP 2017 và một số lỗ hổng tấn công web điển hình hiện nay như là SQLi, XSS, Duyệt đường dẫn *(Directory traversal),* CMDi cũng như cách phòng chống tương ứng đối với mỗi loại lỗ hổng cụ thể và đối với hệ thống web nói chung.

Trong chương 2, với nội dung là **PHÁT HIỆN TẤN CÔNG WEB DỰA TRÊN HỌC MÁY SỬ DỤNG WEB LOG,** luận văn sẽ tiếp tục đi tìm hiểu về WEBLOG, khái quát và các dạng, đồng thời đi sâu vào việc giới thiệu học máy và các thuật toán học máy, đưa ra mô hình phát hiện tấn công website và chi tiết các khâu xử lý dữ liệu.

CHƯƠNG 2: PHÁT HIỆN TẤN CÔNG WEB DỰA TRÊN HỌC MÁY SỬ DỤNG WEB LOG

2.1. Tìm hiểu về Web log

2.1.1. Khái quát về Web log

Web log hay nhật ký web là tệp nhật ký tự động được tạo và duy trì bởi một máy chủ web. Mỗi lần người dùng truy cập vào trang Web, bao gồm từng chế độ xem tài liệu HTML, hình ảnh hoặc đối tượng khác đều được máy chủ web ghi lại. Định dạng một bản ghi nhật ký web về cơ bản là một dòng văn bản cho mỗi lần truy cập vào trang web. Tài liệu này chứa thông tin về những người đã truy cập trang web, nơi họ đến và chính xác những gì họ đang làm trên trang web bao gồm một loạt các mục được sắp xếp theo thứ tự thời gian đảo ngược, thường được cập nhật thường xuyên với thông tin mới về các chủ đề cụ thể.

Các web server chuẩn như Apache và Microsoft IIS tạo thông điệp ghi nhật ký theo một chuẩn chung (CLF – common log format). Tệp nhật ký CLF chứa các dòng thông điệp cho mỗi một gói HTTP request theo định dạng như sau:

Host Ident Authuser Date Request Status Bytes

2.1.2. Các dạng web log

- ✓ Tệp nhật ký truy cập
- ✓ Tệp nhật ký đối tượng
- ✓ Tệp nhật ký lỗi
- ✓ Tệp nhật ký giới thiệu

Bảng 2.1: Các loại định dạng của tệp nhật ký máy chủ Web

Các loại tệp nhật ký	Hoạt động	Định dạng	Trích xuất kiến thức
Nhật ký Truy cập	1. Ghi lại tất cả các yêu cầu người dùng xử lý bởi máy chủ. 2. Ghi thông tin về người dùng.	[Wed Oct 11 14:32:52 2000] [error] [Client 127.0.0.1] máy khách bị từ chối bởi máy chủ cấu hình: /export/home/live/ap/htdocs /test	Người dùng hồ sơ cá nhân Các mẫu thường xuyên. Sử dụng băng thông.
Nhật ký tác nhân	1. Trình duyệt người dùng 2. Phiên bản trình duyệt	"Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)"	Phiên bản đại lý Hệ điều hành được sử dụng.
Tệp nhật ký lỗi	Danh sách lỗi cho người dùng yêu cầu được thực hiện bởi máy chủ.	[Wed Oct 11 14:32:52 2000] [error] [client 127.0.0.1] máy khách bị từ chối bởi máy chủ cấu hình: /export/home/live/ap/htdocs /test	Các loại lỗi Tạo địa chỉ lỗi IP Ngày và thời gian xảy ra lỗi.
Nhật ký giới thiệu	1.Thông tin về liên kết. 2. Chuyển hướng khách truy cập vào trang web.	"http://www.google.com/search?q=keyword", "/page.html"	Trình duyệt đã sử dụng Từ khóa. Chuyển hướng nội dung liên kết.

Định dạng tệp nhật ký: Có ba loại định dạng tệp nhật ký

Định dạng tệp nhật ký chung:

Định dạng tệp nhật ký kết hợp:

Nhiều nhật ký truy cập:

Thông số tệp nhật ký máy chủ:

Dưới đây sẽ minh họa trong Bảng 2.2, danh sách một số thông số hữu ích cho các quá trình phân tích.

Bảng 2.2: Một số trường của Web log

TT	Tên trường	Sự miêu tả
1	DATE	Ngày xử lý yêu cầu theo định dạng yyyy-mm-dd (năm-tháng-ngày)
2	TIME	Giờ xử lý yêu cầu theo định dạng hh:mm:ss (giờ:phút:giây)
3	CLIENT_IP	Địa chỉ IP của máy khách
4	HTTP_METHOD	Phương thức HTTP máy khách gửi yêu cầu
5	URI_STEM	Địa chỉ tương đối của trang, ví dụ /products/search.aspx
6	URI_QUERY	Chuỗi truy vấn của trang (HTTP query string). Ví dụ: category_id=100& category_desc=Science Fiction Books
7	HTTP_STATUS	Mã trạng thái xử lý yêu cầu. Ví dụ 200 là mã xử lý yêu cầu thành công
8	BYTE_RECEIVED	Số lượng Byte của yêu cầu (request) máy chủ nhận được từ máy khách
9	BYTE_SENT	Số lượng Byte của trả lời (response) máy chủ gửi được từ máy khách
10	TIME_TAKEN	Thời gian xử lý yêu cầu tính bằng giây

2.2. Khái quát về Học Máy và các thuật toán Học Máy

2.2.1. Khái quát về học máy

2.2.1.1. Khái niệm

Học máy (machine learning) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Cụ thể hơn, học máy là một phương pháp để tạo ra các chương trình máy tính bằng việc phân tích các tập dữ liệu (*là khả năng của chương trình máy tính sử dụng kinh nghiệm, quan sát, hoặc dữ liệu trong quá khứ để cải thiện công việc của mình trong tương lai thay vì chỉ thực hiện theo đúng các quy tắc đã được lập trình sẵn)*. Học máy có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, học máy tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. [3][13] Nhiều bài toán suy luận được xếp vào loại bài toán NP-khó, vì thế một phần của học máy là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lý được.

Quá trình học máy đơn giản được hiểu là ta cung cấp tập dữ liệu để cho thuật toán có thể tự học mà không cần phải cài đặt các luật quyết định. Để từ đó ta sẽ đưa các dữ liệu kiểm thử vào để hệ thống đưa ra các kết quả nhận định dựa vào quá trình học trên tập dữ liệu huấn luyện (quá trình học thông thường, một hệ thống học máy cần có khả năng ghi nhớ, thích nghi, và đặc biệt là tổng quát hóa. Tổng quát hóa là khả năng của hệ thống học máy ra quyết định chính xác trong các trường hợp mới, chưa gặp, dựa trên kinh nghiệm học được từ dữ liệu hoặc các quan sát trước đó).

Học máy có hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động rô-bốt (robot locomotion).

2.2.1.2. Phân loại kỹ thuật học máy

Xét theo phương thức học, các thuật toán ML được chia làm bốn nhóm, bao gồm “Học có giám sát” (Supervised Learning), “Học không giám sát” (Unsupervised Learning), “Học bán giám sát” (hay học kết hợp - Semi-supervised Learning) và “Học tăng cường” (Reinforcement Learning).

- Học có giám sát
- Học không giám sát
- Học bán giám sát

- Học tăng cường

2.2.2. Một số thuật toán học máy

2.2.2.1. Naive Bayes

Naive Bayes Classification (NBC) là một thuật toán dựa trên định lý Bayes về lý thuyết xác suất để đưa ra các phán đoán cũng như phân loại dữ liệu dựa trên các dữ liệu được quan sát và thống kê. Naive Bayes Classification là một trong những thuật toán được ứng dụng rất nhiều trong các lĩnh vực Machine learning dùng để đưa các dự đoán chính xác nhất dựa trên một tập dữ liệu đã được thu thập, vì nó khá dễ hiểu và độ chính xác cao. Nó thuộc vào nhóm Supervised Machine Learning Algorithms (thuật toán học có giám sát), tức là máy học từ các ví dụ từ các mẫu dữ liệu đã có.

Định luật Bayes được phát biểu như sau:

Định lý Bayes cho phép tính xác suất xảy ra của một sự kiện ngẫu nhiên A khi biết sự kiện liên quan B đã xảy ra. Xác suất này được ký hiệu là **P(A|B)**, và đọc là "xác suất của A nếu có B". Đại lượng này được gọi xác suất có điều kiện hay xác suất hậu nghiệm vì nó được rút ra từ giá trị được cho của B hoặc phụ thuộc vào giá trị đó.

Công thức của định luật Bayes [14] được phát biểu như sau:

$$P(H|x) = P(H) * P(x|H) / P(x)$$

Với:

- $P(H|x)$ là xác suất để xảy ra giả thuyết H với đầu vào là tập dữ liệu ngẫu nhiên cần dự đoán x.
- $P(H)$ là xác suất xảy ra của bản thân giả thuyết H mà không quan tâm đến x.
- $P(x|H)$ là xác suất xảy ra x khi biết H xảy ra, gọi là "xác suất của x nếu có H".
- $P(x)$ là xác suất xảy ra của riêng tập dữ liệu dự đoán x.

Tổng quát:

$$P(H|x1 \dots xn) = P(H)P(x1|H) \dots P(xn|H) / P(x1) \dots P(xn)$$

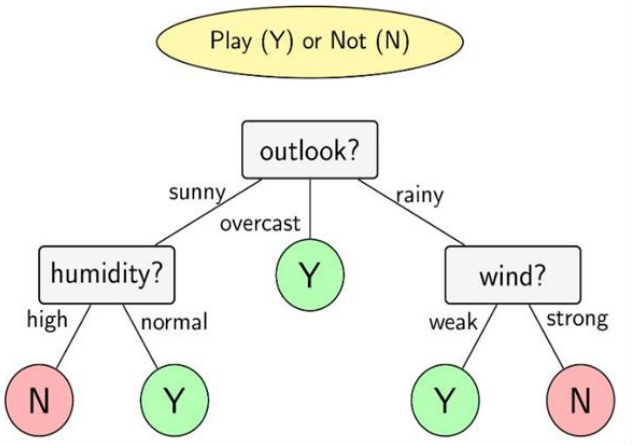
Ứng dụng:

1. **Real time Prediction:** Tốc độ của thuật toán phân loại có thể giúp nó sử dụng trong việc ra quyết định trong thời gian thực.
2. **Multi class Prediction**: Bản chất thuật toán là phân loại và dự đoán rồi chia thành nhiều lớp.
3. **Sentiment Analysis**: Naive Bayes được sử dụng trong phân loại ngôn ngữ tự nhiên và cho kết quả tốt hơn so với một số thuật toán khác. Bên cạnh đó còn phân loại được spam-mail và nhận định được bình luận tích cực hay không tích cực trong mạng xã hội.
4. **Recommendation System:** các hệ thống gợi ý hoạt động dựa trên dự đoán.

2.2.2.2. Cây quyết định

Cây quyết định (Decision Tree) là một đồ thị của các quyết định và các hậu quả có thể của nó (bao gồm rủi ro và hao phí tài nguyên). Cây quyết định được sử dụng để xây dựng một kế hoạch nhằm đạt được mục tiêu mong muốn. Các cây quyết định được dùng để hỗ trợ quá trình ra quyết định. Cây quyết định là một dạng đặc biệt của cấu trúc cây.

Học bằng cây quyết định cũng là một phương pháp thông dụng trong khai phá dữ liệu. Khi đó, cây quyết định mô tả một cấu trúc cây, trong đó, các lá đại diện cho các phân loại còn cành đại diện cho các kết hợp của các thuộc tính dẫn tới phân loại đó. Một cây quyết định có thể được học bằng cách chia tập hợp nguồn thành các tập con dựa theo một kiểm tra giá trị thuộc tính. Quá trình này được lặp lại một cách đệ qui cho mỗi tập con dẫn xuất. Quá trình đệ qui hoàn thành khi không thể tiếp tục thực hiện việc chia tách được nữa, hay khi một phân loại đơn có thể áp dụng cho từng phần tử của tập con dẫn xuất.



Hình 2.3: Mô hình thuật toán cây quyết định [13]

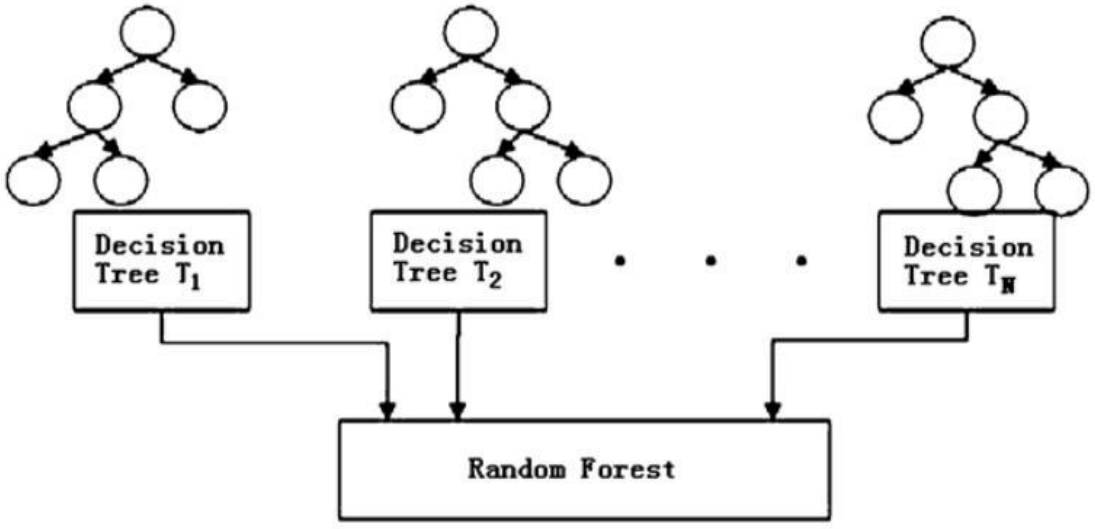
Cây quyết định có 2 loại:

Cây hồi quy (Regression tree): ước lượng các hàm có giá trị là số thực thay vì được sử dụng cho các nhiệm vụ phân loại (định giá, ước lượng giá trị của một căn nhà cần giao bán, khoảng thời gian nằm viện của 1 bệnh nhân).

Cây phân loại (Classification tree): được dùng trong các bài toán phân loại kết quả (phân biệt giới tính, kết quả trận đấu, …).

2.2.2.3. Rừng ngẫu nhiên

Rừng ngẫu nhiên là một thuật toán học có giám sát. Rừng ngẫu nhiên sử dụng các cây để làm nền tảng. Rừng ngẫu nhiên là một tập hợp của các cây quyết định, mà mỗi cây được chọn theo một thuật toán dựa vào ngẫu nhiên. Rừng ngẫu nhiên hoạt động bằng cách đánh giá nhiều cây quyết định ngẫu nhiên, và lấy ra kết quả được đánh giá tốt nhất trong số kết quả trả về. Mô hình rừng ngẫu nhiên rất hiệu quả cho các bài toán phân loại vì nó huy động cùng lúc hàng trăm mô hình nhỏ hơn bên trong với quy luật khác nhau để đưa ra quyết định cuối cùng. Mỗi mô hình con có thể mạnh yếu khác nhau, nhưng theo nguyên tắc "wisdom of the crowd", ta sẽ có cơ hội phân loại chính xác hơn so với khi sử dụng bất kì một mô hình đơn lẻ nào. Mô hình tiêu biểu cơ bản của thuật toán Random Forest được biểu diễn như hình sau:

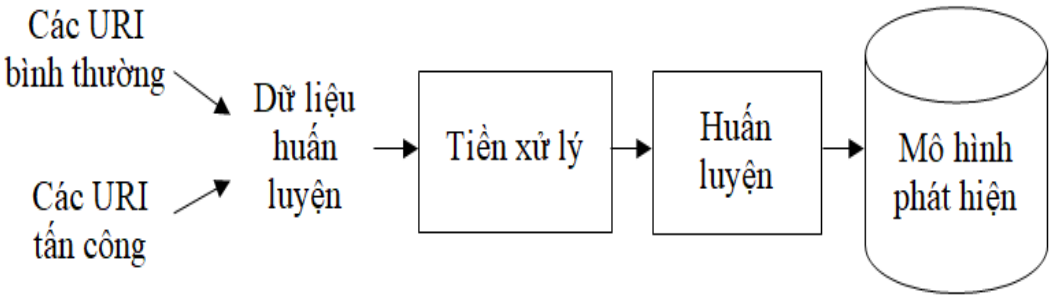


Hình 2.4: Mô hình thuật toán rừng ngẫu nhiên [13]

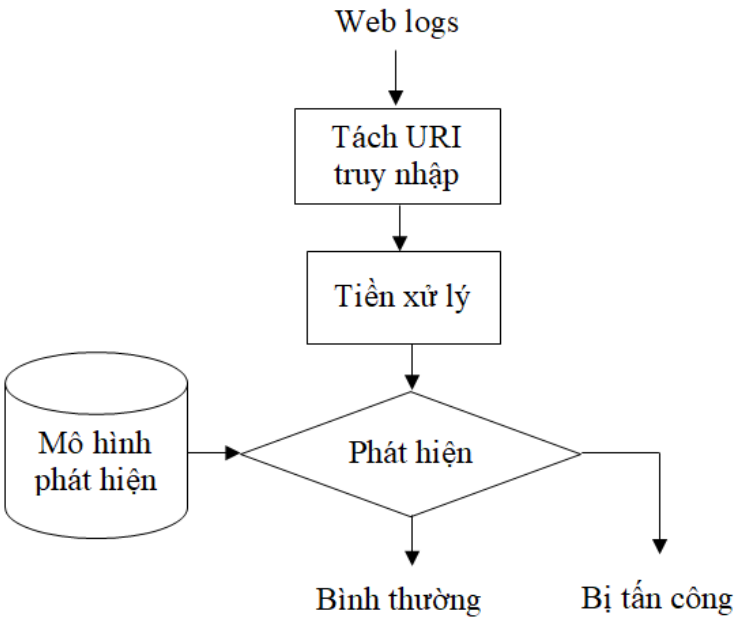
2.3. Phát hiện tấn công web dựa trên học máy sử dụng web log

2.3.1. Mô hình phát hiện

Mô hình phát hiện tấn công web cơ bản dựa trên học máy sử dụng web log trong luận văn này được triển khai theo 2 giai đoạn: (1) giai đoạn huấn luyện như biểu diễn trên Hình 2.5 và (2) giai đoạn phát hiện như biểu diễn trên Hình 2.6. Trong đó, các URI (Uniform Resource Indicator) là các chuỗi truy nhập được bóc tách từ web log. Các URI bình thường và URI tấn công dùng cho giai đoạn huấn luyện được lấy từ tập dữ liệu mẫu đã được gán nhãn.



Hình 2.5: Mô hình phát hiện tấn công web cơ bản: Giai đoạn huấn luyện [2]



Hình 2.6: Mô hình phát hiện tấn công web cơ bản: Giai đoạn phát hiện [2][20]

2.3.2. Các giai đoạn huấn luyện và phát hiện

2.3.2.1. Giai đoạn huấn luyện

Giai đoạn này thực hiện xây dựng mô hình phát hiện từ dữ liệu huấn luyện và gồm các bước sau:

- Thu thập tập dữ liệu huấn luyện
- Tiền xử lý
- Huấn luyện
- Kiểm thử mô hình phát hiện

2.3.2.2. Giai đoạn phát hiện

Giai đoạn này thực hiện phân tích các dòng web log nhằm phát hiện các dấu hiệu tấn công SQLi, XSS, duyệt đường dẫn và chèn dòng lệnh hệ điều hành và gồm các bước sau:

- Tách URI truy nhập
- Tiền xử lý
- Phát hiện

2.4. Kết luận Chương 2

Chương 2 đã giới thiệu những hiểu biết cơ bản về WEB LOG, các dạng WEB LOG, hoạt động cũng như định dạng của từng loại WEB LOG. Ngoài ra chương 2 cũng giới thiệu khái quát về học máy và đưa ra một số thuật toán học máy được sử dụng để phát hiện tấn công web như Naive Bayes, Cây quyết định (Decision Tree), Rừng ngẫu nhiên. Tuy nhiên mục đích của chương 2 chủ yếu là đi sâu vào việc trình bày mô hình phát hiện tấn công được sử dụng, các pha xử lý trong mô hình như là tiền xử lý, huấn luyện và phát hiện.

Trong chương 3, nội dung chủ yếu là giới thiệu tập dữ liệu được sử dụng để huấn luyện cho học máy, cách xử lý tiền dữ liệu, các bước làm trong pha huấn luyện và phân loại các dữ liệu đầu vào. Trình bày một số phương pháp để sử dụng huấn luyện và phát hiện, các kết quả sẽ được dùng để đánh giá mức độ hiệu quả khi sử dụng các phương pháp học máy khác nhau.

CHƯƠNG 3: CÀI ĐẶT VÀ THỬ NGHIỆM

3.1. Giới thiệu tập dữ liệu

3.1.1 Tập dữ liệu mẫu

Tập dữ liệu dùng cho thử nghiệm đánh giá mô hình phát hiện là HttpParamsDataset [19]. Tập này gồm các tham số truy vấn HTTP với 19.304 truy vấn bình thường được gán nhãn *norm* và 11.763 truy vấn bất thường được gán nhãn *anom*.

- 10.852 truy vấn tấn công chèn mã SQL được gán nhãn *sqli*
- 532 truy vấn tấn công XSS được gán nhãn *xss*
- 89 truy vấn tấn công chèn mã lệnh hệ điều hành được gán nhãn *cmdi*
- 290 truy vấn tấn công duyệt đường dẫn được gán nhãn *path-traversal*.

Tập dữ liệu HttpParamsDataset được chia thành 2 phần sử dụng cho huấn luyện và kiểm thử:

- Tập cho huấn luyện gồm 20.712 truy vấn, trong đó có 7.842 truy vấn bất thường;
- Tập cho kiểm thử gồm 10.355 truy vấn, trong đó có 3.921 truy vấn bất thường.

3.1.2 Dữ liệu web log thực

Tập dữ liệu web log thực là dữ liệu thu thập thực tế từ các máy chủ web. Luận văn sử dụng một phần dữ liệu web log thu thập bởi đề tài khoa học công nghệ cấp nhà nước, mã số KC.01.05/16-20 **Error! Reference source not found.** thực hiện tại Học viện Công nghệ Bưu chính Viễn thông. Web log được thu thập và chuẩn hóa theo định dạng W3C Extended phục vụ cho phân tích, xử lý.

3.2. Tiền xử lý dữ liệu

Khâu tiền xử lý dữ liệu nhằm trích chọn và số hóa các đặc trưng cho mỗi truy vấn HTTP được thực hiện theo các bước được mô tả trong phần giới thiệu mô hình phát hiện ở Chương 2.

Do ta sử dụng một bộ 3-gram chuẩn được xây dựng từ việc lấy tất cả các phần tử 3-gram khác nhau trong quá trình phân tách 3-gram của các bản ghi của tập huấn luyện gồm 20.712 truy vấn, chính vì vậy độ dài của bộ 3-gram chuẩn này rất lớn có thể lên tới vài chục nghìn phần tử. Nếu thực hiện lưu trữ, ánh xạ các phần tử trên tập chuẩn này thì sẽ gặp khó khăn trong quá trình cả về lưu trữ và tốc độ xử lý tính toán. Vì vậy, để quá trình huấn luyện

được diễn ra nhanh hơn mà không mất đi tính chính xác thì mô hình sẽ sử dụng một phương pháp để giảm chiều dữ liệu bộ 3-gram chuẩn đó là Principal Component Analysis (PCA).

3.3. Huấn luyện và kiểm thử mô hình phát hiện

Tập dữ liệu huấn luyện sau tiền xử lý được sử dụng để huấn luyện sử dụng thuật toán cây quyết định để sinh mô hình phân loại (cụ thể là thuật toán cây quyết định CART được hỗ trợ trong thư viện sk-learn của Python). Mô hình được lưu vào file cho khâu kiểm thử. Trong khâu kiểm thử, tập dữ liệu kiểm thử sau tiền xử lý được sử dụng để đánh giá độ chính xác phân loại.

3.4. Thử nghiệm, kết quả và nhận xét

3.4.1. Lựa chọn công cụ thử nghiệm

3.4.2. Kết quả thử nghiệm

3.4.3. Nhận xét

- ✓ Mô hình phát hiện tấn công web cơ bản đạt độ chính xác phát hiện trung bình khá cao, đạt 98.51%. Hầu hết các dạng tấn công và trạng thái bình thường đều có độ chính xác phát hiện cao, riêng độ chính xác phát hiện tấn công CMDi chỉ đạt 66.67% do lượng dữ liệu huấn luyện cho dạng tấn công này khá ít. Trên thực tế, tấn công CMDi ít gặp trên dịch vụ web hơn các dạng SQLi, XSS và duyệt đường dẫn.
- ✓ Kết quả phát hiện thử trên web log thực cho thấy mô hình phát hiện khá chính xác từng loại tấn công. Mô hình có khả năng phát hiện 4 dạng tấn công web cơ bản bao gồm SQLi và XSS, tấn công duyệt đường dẫn và CMDi.

3.5. Kết luận chương 3

Trong chương 3 của luận văn đã mô tả chi tiết dữ liệu được sử dụng cho mô hình phát hiện tấn công web sử dụng học máy, mô tả chi tiết các phương pháp huấn luyện và phát hiện, thống kê chi tiết các kết quả đạt được bằng nhiều kịch bản thử nghiệm khác nhau từ đó rút ra được những nhận xét ưu điểm và những hạn chế của phương pháp học máy sử dụng.

KẾT LUẬN

Kết quả đạt được:

Từ nội dung của 3 chương, luận văn đã đạt được những kết quả sau:

- Trình bày khái quát về ứng dụng web, các yêu cầu bảo mật đối với ứng dụng web, web server, các loại tấn công web cũng như đặc điểm cách khai thác của loại tấn công web phổ biến và các biện pháp bảo mật, cách phòng chống.
- Trình bày các phương pháp phát hiện tấn công web sử dụng học máy, các thuật toán học máy được áp dụng cho bài toán phát hiện tấn công web. Đưa ra mô hình phát hiện tấn công web và nguyên lý hoạt động của mô hình phát hiện tấn công. Trình bày quá trình xử lý dữ liệu, đưa dữ liệu vào huấn luyện và phát hiện kiểm tra.
- Thử nghiệm mô hình phát hiện tấn công web cơ bản dựa trên học máy với các kịch bản cụ thể.

Hướng phát triển trong tương lai

- Do hạn chế về thời gian và khả năng, luận văn mới chỉ thử nghiệm mô hình trên 1 thuật toán học máy là Cây quyết định. Trong tương lai sẽ sử dụng các thuật toán khác trong quá trình huấn luyện và phát hiện, như Naive Bayes, Rừng ngẫu nhiên và SVM, từ đó tìm ra thuật toán tối ưu.
- Cập nhật thêm dữ liệu để phát hiện được các loại tấn công mới hiện nay cũng như cập nhật được cách thức tấn công mới trên các lỗ hổng cũ.