

**ĐẠI HỌC THÁI NGUYÊN**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**VŨ MINH HUY**

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG**  
**GIÁM SÁT SỰ THAY ĐỔI NỘI DUNG WEBSITE**

**LUẬN VĂN THẠC SĨ**  
**CHUYÊN NGÀNH KHOA HỌC MÁY TÍNH**

**THÁI NGUYÊN - 2017**

**ĐẠI HỌC THÁI NGUYÊN**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**VŨ MINH HUY**

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG GIÁM SÁT SỰ THAY ĐỔI NỘI  
DUNG WEBSITE**

**LUẬN VĂN THẠC SĨ**  
**CHUYÊN NGÀNH KHOA HỌC MÁY TÍNH**

**THÁI NGUYÊN - 2017**

## LỜI CẢM ƠN

Trước hết tôi xin gửi lời cảm ơn sâu sắc nhất đến thầy hướng dẫn khoa học TS. Lê Quang Minh về những chỉ dẫn khoa học, định hướng nghiên cứu và tận tình hướng dẫn tôi trong suốt quá trình làm luận văn.

Tôi cũng xin cảm ơn các Thầy trong Viện Công nghệ thông tin và các Thầy Cô trong trường Đại học Công nghệ thông tin và Truyền thông – Đại học Thái Nguyên đã quan tâm chỉ bảo và trực tiếp giảng dạy, giúp đỡ trong suốt quá trình học tập và nghiên cứu.

Tôi xin chân thành cảm ơn lãnh đạo các phòng ban, đồng nghiệp đơn vị công tác cũ là Chi nhánh Viettel Thái Nguyên và đơn vị công tác mới là Công ty cổ phần VETC đã tạo điều kiện cho tôi học tập và nghiên cứu đề tài này.

Cuối cùng tôi xin cảm ơn gia đình và bạn bè, những người đã luôn ủng hộ và động viên tôi trong suốt quá trình học tập và nghiên cứu luận văn này.

*Thái Nguyên, tháng 05 năm 2017*

**Học viên**

**Vũ Minh Huy**

**LỜI CAM ĐOAN*****Tôi xin cam đoan:***

Những kết quả nghiên cứu được trình bày trong luận văn là hoàn toàn trung thực của tôi, không vi phạm bất cứ điều gì trong luật sở hữu trí tuệ và pháp luật Việt Nam. Nếu sai, tôi xin chịu hoàn toàn trách nhiệm trước pháp luật.

*Thái Nguyên, tháng 05 năm 2017*

**Học viên**

**Vũ Minh Huy**

## MỤC LỤC

LỜI CẢM ƠN.....	1
LỜI CAM ĐOAN.....	2
MỤC LỤC .....	3
DANH MỤC CÁC KÝ HIỆU .....	6
DANH MỤC HÌNH VẼ .....	8
LỜI MỞ ĐẦU.....	9
<b>CHƯƠNG 1: TỔNG QUAN VỀ WEBSITE VÀ MỘT SỐ VẤN ĐỀ AN NINH</b> .....	<b>12</b>
1.1. Một số khái niệm cơ bản về Website .....	12
1.1.1. Khái niệm Web .....	12
1.1.2. Hoạt động của Web.....	13
1.1.3. Trình duyệt Web .....	13
1.1.4. Địa chỉ URL .....	13
1.1.5. Tên miền.....	14
1.1.6. Giao thức HTTP và HTTPS.....	15
1.1.7. Web Server.....	16
1.1.8. Database Server.....	17
1.1.9. Các dịch vụ và ứng dụng trên nền Web.....	17
1.2. Thực trạng an ninh Web tại Việt Nam và trên thế giới.....	18
1.3. Vai trò và mục đích của việc theo dõi sự thay đổi nội dung trang web.....	19
1.4. Một số lỗ hổng an ninh trên Website.....	20
1.5. Một số kỹ thuật tấn công Website .....	21
1.5.1. Tấn công SQL Injection.....	21
1.5.2. Tấn công XSS (Cross Site Scripting).....	30
1.5.3. Tấn công từ chối dịch vụ DOS (Denial of Service).....	34

1.6. Kết luận chương 1.....	41
<b>CHƯƠNG 2: CÁC THUẬT TOÁN GIÁM SÁT VÀ MỘT SỐ PHƯƠNG PHÁP ĐẢM BẢO AN NINH WEBSITE .....</b>	<b>42</b>
2.1. Một số phương pháp đảm bảo an ninh Website .....	42
2.1.1. Đảm bảo an ninh hệ điều hành Webserver .....	42
2.1.2. Đảm bảo an ninh Webserver.....	43
2.1.3. Quản trị Webserver .....	47
2.1.3.1. Logging .....	47
2.1.3.2. Thủ tục sao lưu Webserver .....	47
2.1.3.3. Phục hồi từ một sự làm hại an ninh .....	48
2.1.3.4. Kiểm tra bảo mật Webserver .....	48
2.1.4. Bảo đảm an ninh nội dung Web.....	49
2.1.5. Sử dụng kỹ thuật xác thực và mã hóa .....	51
2.1.5.1. Xác định các yêu cầu xác thực và mã hóa .....	51
2.1.5.2. Xác thực dựa trên địa chỉ .....	52
2.1.5.3. Xác thực cơ bản.....	52
2.1.5.4. Xác thực Digest.....	53
2.1.5.5. SSL/TLS.....	53
2.1.6. Triển khai cơ sở hạ tầng mạng an ninh .....	55
2.1.6.1. Thành phần và cấu trúc mạng .....	55
2.1.6.2. Cấu hình phần tử mạng .....	55
2.2. Các thuật toán giám sát.....	56
2.2.1. Hàm băm mật mã .....	56
2.2.1.1. Giới thiệu hàm băm mật mã.....	56
2.2.1.2. Tính một chiều của hàm băm.....	57
2.2.1.3. Cấu trúc hàm băm .....	58
2.2.2. Thuật toán đối sánh chuỗi .....	60

2.2.3.1. Khái niệm đối sánh chuỗi .....	60
2.2.2.2. Phân loại thuật toán đối sánh chuỗi .....	61
2.2.3. Dấu vân tay tài liệu .....	61
2.2.4. Thuật toán Rabin Fingerprint.....	62
2.2.5. Thuật toán Rabin Fingerprint cải tiến .....	64
2.3 Kết luận chương 2.....	66
<b>CHƯƠNG 3: ỨNG DỤNG THUẬT TOÁN RABIN FINGERPRINT CẢI TIẾN XÂY DỰNG HỆ THỐNG GIÁM SÁT SỰ THAY ĐỔI NỘI DUNG WEBSITE.....</b>	<b>67</b>
3.1 Giới thiệu .....	67
3.2. Hệ thống giám sát nội dung Website.....	68
3.2.1. Hệ thống Builder .....	70
3.2.2. Hoạt động Multi-checker .....	70
3.2.3. Hệ thống Recover.....	72
3.2.4. Hệ thống Self-watcher .....	73
3.2.5. Hệ thống Admin.....	74
3.3. Hoạt động hệ thống dựa trên các phát hiện thay đổi bất thường.....	74
3.3.1. Tổng quan hệ thống.....	74
3.3.2. Kiến trúc hệ thống.....	75
3.4. Phân tích hoạt động của hệ thống.....	76
3.5 Cài đặt và thử nghiệm chương trình.....	79
3.5.1. Cài đặt chương trình.....	79
3.5.2. Thử nghiệm chương trình .....	80
3.5.3. Nhận xét kết quả.....	82
3.6. Kết luận chương 3.....	83
<b>KẾT LUẬN.....</b>	<b>84</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>85</b>

## DANH MỤC CÁC KÝ HIỆU

STT	Kí hiệu	Tiếng việt	Tiếng anh
1	ACK	Là một báo nhận	Acknowledgment
2	CNTT	Công nghệ thông tin	
3	CSDL	Cơ sở dữ liệu	Database
4	CSS	Ngôn ngữ quy định cách trình bày của các thẻ html trên trang web	Cascading Style Sheet
5	DNS	Hệ thống tên miền	Domain name System
6	DOS	Từ chối dịch vụ	Dinal of Services
7	Firewall	Tường lửa	
8	Hosting	Nơi cung cấp dịch vụ và lưu trữ website	
9	HTML	Ngôn ngữ đánh dấu siêu văn bản	Hypertext Markup Language
10	HTTP	Giao thức truyền siêu văn bản	Hypertext Transfer Protocol
11	HTTPS	Kết hợp giữa giao thức HTTP và giao thức bảo mật SSL hay TLS	
12	Hyperlink	Liên kết siêu văn bản	
13	Hypertext	Hệ thống siêu văn bản	
14	ICMP	Giao thức thông điệp vận chuyển Internet	Internet Control Message Protocol
15	Internet	Mạng toàn cầu	
16	IP	Địa chỉ IP	Internet protocol
17	IPS	Hệ thống chống xâm nhập	Intrusion Prevention System
18	OSI	Mô hình hệ thống mở	Open Systems Interconnection



19	SQL	Ngôn ngữ truy vấn mang tính cấu trúc	Structured Query Language
20	SSL	Giao thức bảo mật	Secure Sockets Layer
21	SYN	Ký tự đồng bộ hoá	The Synchronous Idle Character
22	TLS	Giao thức bảo mật	Transport Layer Security
23	UDP	Giao thức dữ liệu người dùng	User Datagram Protocol
24	URL	Định vị tài nguyên thống nhất	Uniform Resource Locator
25	Web browser	Trình duyệt web	
26	Web server	Máy chủ phục vụ web	
27	XML	Ngôn ngữ Đánh dấu Mở rộng	Xtensible Markup Language
28	XSS	Là một kiểu tấn công cho phép kẻ tấn công chèn những đoạn script độc hại (thông thường là javascript hoặc HTML) vào website (Cross Site Scripting)	

## DANH MỤC HÌNH VẼ

Hình 1.1. Một số Webservice thông dụng .....	17
Hình 1.2. Mô hình tấn công SQL Injection.....	21
Hình 1.3. Mô hình tấn công XXS (Cross-Site Scripting).....	33
Hình 1.4. Một mô hình tấn công từ chối dịch vụ DOS.....	37
Hình 1.5. Mô hình tấn công SYN Flood Attack.....	39
Hình 2.1. SSL/TLS đặt trong Internet Protocol Stack.....	55
Hình 3.1. Sơ đồ Merkel-Damgard .....	60
Hình 3.2. Mô tả thuật toán Rabin Fingerprint .....	65
Hình 3.3. Minh họa cải tiến giải thuật .....	67
Hình 3.4. Sơ đồ kiến trúc hệ thống giám sát website .....	71
Hình 3.5. Các thuộc tính cơ bản của tập tin .....	73
Hình 3.6. Kiến trúc hệ thống phát hiện thay đổi bất thường.....	77
Hình 3.7. Sơ đồ phương pháp phát hiện sự thay đổi nội dung trang web.....	79
Hình 3.8. Giao diện chính của chương trình giám sát website .....	82
Hình 3.9. Chương trình giám sát website theo dõi 4 website đồng thời .....	84
Hình 3.10. Hoạt động của chức năng Advanced Mode.....	82