ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

ĐẶNG HÙNG THẮNG

NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP PHÁT HIỆN THAY ĐỔI NỘI DUNG TRANG WEB

BÁO CÁO LUẬN VĂN THẠC SỸ

CHUYÊN NGÀNH KHOA HỌC MÁY TÍNH

THÁI NGUYÊN, NĂM 2015

LỜI CẢM ƠN

Tôi xin gửi lời cảm ơn sâu sắc đến Thầy TS. Vũ Duy Linh người đã giúp đỡ tôi trong suốt thời gian tôi thực hiện đề tài. Thầy đã định hướng, tạo những điều kiện thuận lợi và đã tận tình hướng dẫn để tôi hoàn thành đề tài này.

Tôi xin gửi lời cảm ơn chân thành đến gia đình, bạn bè đã luôn là nguồn động viên to lớn, giúp đỡ tôi trong suốt quá trình tôi thực hiện đề tài.

TÁC GIẢ LUẬN VĂN

Đặng Hùng Thắng

LỜI 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.

TÁC GIẢ LUẬN VĂN

Đặng Hùng Thắng

MỤC LỤC

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

STT	Kí hiệu	Tiếng việt	Tiếng anh
01	CERT		Computer Emegancy Response Team
02	PKI	Thuật toán mã hóa công khai	
03	CA	Nhà cung cấp dịch vụ chứng thực chữ ký số	Certification Aithority
04	SQL	Ngôn ngữ truy vấn mang tính cấu trúc	Structured Query Language
05	XSS	Là một kiểu tấn công cho phép kẻ tấn công chèn những đoạn script độc hại vào website	Cross Site Scripting
06	DOS	Từ chối dịch vụ	Dinal of Services
07	VPN	Mạng riêng ảo	Virtual Private Network
08	Firewall	Tường lửa	
09	HTML	Ngôn ngữ đánh dấu siêu văn bản	Hypertext Markup Language
10	HTTPS	Kết hợp giữa giao thức HTTP và giao thức bảo mật SSL hay TLS	
11	SSL	Giao thức bảo mật	Secure Sockets Layer
12	TLS	Giao thức bảo mật	Transport Layer Security
13	OSI	Mô hình hệ thống mở	Open Systems Interconnection

14	URL	Định vị tài nguyên thống nhất	Uniform Resource Locator
15	XML	Ngôn ngữ đánh dấu mở rộng	Xtensible Markup Language
16	IP	Địa chỉ IP	Internet protocol
17	DNS	Hệ thống tên miền	Domain name System
18	CSDL	Cơ sở dữ liệu	

DANH MỤC HÌNH VẼ

PHẦN MỞ ĐẦU: GIỚI THIỆU BÀI TOÁN THEO DÕI SỰ THAY ĐỔI NỘI DUNG TRANG WEB

1. Đặt vấn đề

Cùng với sự phát triển của công nghệ thông tin, công nghệ máy tính và mạng Internet ngày nay các dịch vụ mạng đã có mặt trong hầu hết các lĩnh vực đời sống xã hội. Các thông tin trên Internet cũng đa dạng và phong phú, có rất nhiều thông tin đòi hỏi yêu cầu cao về bảo mật bởi tính kinh tế, chính xác và tin cậy của thông tin đó. Bên cạnh đó các hình thức phá hoại trên Internet cũng ngày càng trở nên tinh vi và phức tạp hơn nhiệm vụ đặt ra cho người quản trị mạng là hết sức quan trọng và cần thiết.

Với sự phát triển nhanh của Web và các ứng dụng trên nền Web hiện nay, việc bảo mật an ninh thông tin cho các trang Web là hết sức quan trọng thế nhưng không phải nhà quản trị hay nhà phát triển nào cũng chú tâm tới việc bảo mật an ninh cho các sản phẩm của mình, chính vì vậy rất nhiều Website hiện tại tồn tại lỗ hổng về bảo mật an ninh cao gây nguy cơ bị tấn công, thiệt hại rất nhiều khi bị tấn công. Việc phát hiện các nguy cơ thủ công là khó, việc người quản trị hay người phát triển muốn kiểm tra mức độ an toàn của sản phẩm gặp phải nhiều khó khăn.

Đề tài nghiên cứu nghiên cứu xây dựng một hệ thống với công cụ phát hiện sự thay đổi nội dung trang Web và đưa ra những cảnh bảo cho người quản trị để có biện pháp xử lý kịp thời.

Chương 1. TỔNG QUAN VỀ AN TOÀN NỘI DUNG SỐ VÀ WEBSITE

1.1. Vấn đề đảm bảo tính an toàn của các nội dung trên internet

Ngày nay, cùng với sự phát triển mạnh mẽ của công nghệ thông tin là sự ra đời hàng loạt của các dịch vụ trên internet điều đó đã mang lại cho cá nhân, các tổ chức, các đơn vị, các doanh nghiệp rất nhiều tiện ích, thuận lợi.

Mọi người sử dụng internet như một công cụ bắt buộc trong hoạt động của cá nhân. Các đơn vị, tổ chức, doanh nghiệp sử dụng internet trong mọi hoạt động của đơn vị mình. Những gì internet mạng lại là rất lớn, nó là một phần không thể thiếu trong cuộc sống hiện địa ngày nay.

Ở bất kỳ đâu, khi một máy tính có nối mạng Internet, con người có thể thực hiện các chia sẻ, đăng tải các thông tin, mua bán, thanh toán... một cách thuận tiện và nhanh chóng. Mọi thông tin của người dùng đều được lưu trư trên internet. Mục tiêu chính của nối mạng là để con người có thể lưu trữ, sử dụng tài nguyên từ nhiều vị trí địa lý khác nhau. Tài nguyên bị phân tán dẫn tới nó dễ bị xâm phạm, gây mất mát dữ liệu cũng như các thông tin có giá trị. Điều đó vô hình chung khiến cho internet trở thành một thứ tốt nhưng rất nguy hiểm. Nếu như các tài nguyên đó không được bảo vệ tốt nó có thể gây nên những thiệt hại lớn cho cá nhân cũng như tập thể. Những hiểm họa và thiệt hại phải gánh chịu là không lường trước được.

1.2. Những nguy cơ tiềm ẩn và việc bảo mật nội dung số trên internet

Mọi nguy cơ trên mạng đều là những nguy cơ tiềm tàng, từ một lỗ hổng bảo mật nhỏ của hệ thống nếu bị khai thác và lợi dụng có thể trở thành một tai họa khôn lường.

Theo thống kê của CERT (Computer Emegancy Response Team) – Tổ chức bảo mật nổi tiếng thế giới, thì số vụ tấn công ngày càng tăng và sẽ còn tăng mạnh trong thời gian tới, thiết hại gây ra ngày càng nghiêm trọng. Điều này là dễ hiểu vì một thực thể luôn tồ tại hai mặt đối lập, công nghệ và kỹ thuật