ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

NGUYỄN VĂN DIỄN

NGHIÊN CỨU GIẢI PHÁP PHÁT HIỆN XÂM NHẬP MẠNG MÁY TÍNH BẤT THƯỜNG DỰA TRÊN KHAI PHÁ DỮ LIỆU

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2014

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài "Nghiên cứu Giải pháp phát hiện xâm nhập mạng máy tính bất thường dựa trên Khai phá dữ liệu" là công trình nghiên cứu của riêng tôi. Đề tài được hoàn thành dưới sự hướng dẫn của Thầy TS. Nguyễn Ngọc Cương. Những kết quả nghiên cứu, thử nghiệm được thực hiện hoàn toàn khách quan và trung thực. Các số liệu, kết quả trình bày trong luận văn là hoàn toàn trung thực và chưa từng được công bố trong bất cứ công trình nào.

Các tài liệu tham khảo sử dụng trong luận văn đều được dẫn nguồn (có bảng thống kê các tài liệu tham khảo) hoặc được sự đồng ý trực tiếp của tác giả.

Nếu xảy ra bất cứ điều gì không đúng như những lời cam đoan trên, tôi xin chịu hoàn toàn trách nhiệm.

Hà Nội, ngày 18 tháng 07 năm 2014

TÁC GIẢ

Nguyễn Văn Diễn

LỜI CẢM ƠN

Em xin chân thành cảm ơn Thầy TS. Nguyễn Ngọc Cương người đã trực tiếp hướng dẫn tận tình em trong suốt quá trình thực hiện Luận văn tốt nghiệp.

Em xin chân thành cảm ơn Quý thầy, cô Trường Đại học Công nghệ thông tin & Truyền thông Thái Nguyên, Viện Công nghệ Thông Tin, những người đã nhiệt tình giảng dạy và truyền đạt những kiến thức quí báu trong suốt thời gian em học tập và nghiên cứu tại trường. Với vốn kiến thức tiếp thu được trong quá trình học tập và nghiên cứu không chỉ là nền tảng cho quá trình nghiên cứu luận văn mà còn là hành trang quí báu trong quá trình hoạt động chuyên môn của em.

Cuối cùng, em xin kính chúc Quý thầy cô, đồng nghiệp, gia đình dồi dào sức khỏe và thành công.

Trân trọng cảm ơn!

MỤC LỤC

DANH MỤC TỪ VIẾT TẮT

ADAM	Audit Data Analysis Mining
CSDL	Cơ sở dữ liệu
DdoS	Distributed Daniel of Servies
DOS	Daniel of Services
HIDS	Host Instrucsion Detection System
HTTP	Hypertext Markup Languge
ICMP	Internet Control Message Protocol
IDS	Intrucsion Detection System
IDDM	Intrucsion Detection Data Mining
IPS	Intrucsion Prevention System
IP	Internet Protocol
KPDL	Khai phá dữ liệu
LOF	Local Outlier Partor
LSC	Local Sparsity Ratio
NIDS	Networks Instrusion Detection System
MAC	Media Accsess Controllers
SQL	Structured Query Language
VPN	Virtual Private Network
TCP	Transmission Control Protocol
UDP	User Datagram Protocol

DANH MỤC BẢNG

DANH MỤC HÌNH

MỞ ĐẦU

Ngày nay, Công nghệ thông tin nói chung và Ngành mạng máy tính nói riêng đã được ứng dụng trong hầu hết các lĩnh vực quan trọng của đời sống, nó tác động trực tiếp đến sự tồn tại và phát triển của nền kinh tế tri thức và công nghệ. Chính vì vậy, việc áp dụng Công nghệ thông tin đã trở thành một yêu cầu không thể thiếu cho tất cả các tổ chức, doanh nghiệp. Với tầm quan trọng như vậy, cần phải có một hệ thống mạng doanh nghiệp ổn định, hoạt động liên tục, đảm bảo tính tin cậy, nguyên vẹn, sẵn sàng và không thể từ chối để đáp ứng được mọi yêu cầu kết nối và xử lý của công việc.

Tuy nhiên, bên cạnh yêu cầu cấp thiết đó thì mạng máy tính luôn phải đối diện với rất nhiều nguy cơ mất an toàn như các cuộc "viếng thăm" bất hợp pháp hoặc các cuộc tấn công từ bên ngoài mạng luôn luôn có thể xảy ra với mức độ ngày càng phức tạp và tinh vi hơn. Do đó, yêu cầu phải có một hệ thống có thể phát hiện tự động những hành vi thâm nhập không được phép để cảnh báo nguy cơ và ngăn chặn đã trở nên cấp thiết.

Đã có nhiều hướng nghiên cứu và xây dựng hệ thống cảnh báo và thâm nhập dựa trên các phương pháp thâm nhập như: phát hiện thâm nhập dựa vào luật; kỹ thuật phân biệt ý định người dùng, phân tích trạng thái phiên, phương pháp phân tích thống kê … Tuy nhiên đây là các phương pháp phát hiện xâm nhập dựa trên các dấu hiệu bất thường. Tức là dựa trên các dấu hiệu của các vụ tấn công đã biết, các phương pháp này phát hiện ra xâm nhập mạng bằng cách so sánh các giá trị đặc tả với một dãy các ký tự tấn công được cung cấp bởi chuyên gia và được cập nhật lại trong cơ sở dữ liệu. Điểm hạn chế của các phương pháp trên là chúng không thể phát hiện ra các cuộc tấn công mới không có trong cơ sở dữ liệu. So với các phương pháp trên thì phương pháp phân tích dựa trên kỹ thuật khai phá dữ liệu có nhiều ưu điểm rõ rệt hơn. Phương pháp này có thể sử dụng với cơ sở dữ liệu chứa nhiều nhiễu, dữ liệu không đầy đủ, biến đổi liên tục, đặc biệt phương pháp này đòi hỏi mức độ sử dụng các chuyên gia không quá thường xuyên. Các ưu điểm này đem lại