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

Đỗ Xuân Cường

KỸ THUẬT PHÂN CỤM DỮ LIỆU TRONG PHÁT HIỆNXÂM NHẬP TRÁI PHÉP

Chuyên ngành: Khoa học máy tính
Mã số: 60 48 0101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS LƯƠNG THẾ DŨNG

LỜI CẢM ƠN

Đầu tiên em xin gửi lời cảm ơn sâu sắc nhất tới TS Lương Thế Dũng, người hướng dẫn khoa học, đã tận tình chỉ bảo, giúp đỡ em thực hiện luận văn.

Em xin cảm ơn các thầy cô trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên đã giảng dạy và truyền đạt kiến thức cho em.

Em xin trân thành cảm ơn các đồng chí Lãnh đạo Sở Thông tin và Truyền thông và các đồng nghiệp đã tạo mọi điều kiện giúp đỡ em hoàn thành nhiệm vụ học tập.

Em cũng xin bày tỏ lòng biết ơn đối với gia đình, bạn bè và người thân đã động viên khuyến khích và giúp đỡ trong suốt quá trình hoàn thành luận văn này.

Mặc dù đã hết sức cố gắng hoàn thành luận văn với tất cả sự nỗ lực của bản thân, nhưng luận văn vẫn còn những thiếu sót. Kính mong nhận được những ý kiến đóng góp của quý Thầy, Cô và bạn bè đồng nghiệp.

Em xin trân thành cảm ơn!

LỜI CAM ĐOAN

Luận văn là kết quả nghiên cứu và tổng hợp các kiến thức mà bản thân đã thu thập được trong quá trình học tập tại trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên, dưới sự hướng dẫn, giúp đỡ của các thầy cô và bạn bè đồng nghiệp, đặc biệt là sự hướng dẫn của TS Lương Thế Dũng – Trưởng khoa An toàn thông tin, Học viện Kỹ thuật Mật mã.

Em xin cam đoan luận văn không phải là sản phẩm sao chép của bất kỳ công trình khoa học nào.

Thái Nguyên, ngày tháng năm 2015

HỌC VIÊN

Đỗ Xuân Cường

MỤC LỤC

DANH MỤC CÁC TỪ VIẾT TẮT

TT	Viết tắt	Nội dung
1.	CNTT	Công nghệ thông tin
2.	ATTT	An toàn thông tin
3.	CSDL	Cơ sở dữ liệu
4.	IDS	Hệ thống phát hiện xâm nhập
5.	PHXN	Phát hiện xâm nhập
6.	KDD	Khám phá tri thức trong cơ sở dữ liệu
7.	KPDL	Khai phá dữ liệu
8.	PCDL	Phân cụm dữ liệu
9.	PAM	Thuật toán phân cụm phân hoạch

DANH MỤC CÁC BẢNG

DANH MỤC HÌNH VẼ

LỜI NÓI ĐẦU

Công nghệ thông tin liên tục phát triển và thay đổi, nhiều phần mềm mới ra đời mang đến cho con người nhiều tiện ích hơn, lưu trữ được nhiều dữ liệu hơn, tính toán tốt hơn, sao chép và truyền dữ liệu giữa các máy tính nhanh chóng thuận tiện hơn,... Hệ thống mạng máy tính của các đơn vị được trang bị nhưng vẫn tồn tại nhiều lỗ hổng và các nguy cơ về mất an toàn thông tin. Các vụ xâm nhập mạng lấy cắp thông tin nhạy cảm cũng như phá hủy thông tin diễn ra ngày càng nhiều, thủ đoạn của kẻ phá hoại ngày càng tinh vi.

Công nghệ phát hiện xâm nhập trái phép hiện nay hầu hết dựa trên phương pháp đối sánh mẫu, phương pháp này cho kết quả phát hiện khá tốt, tuy nhiên nó đòi hỏi các hệ thống phát hiện xâm nhập trái phép phải xây dựng được một cơ sở dữ liệu mẫu khổng lồ và liên tục phải cập nhật. Vì vậy hiện nay lĩnh vực nghiên cứu để tìm ra các phương pháp phát hiện xâm nhập trái phép hiệu quả hơn đang được rất nhiều người quan tâm. Trong đó, một hướng quan trọng trong lĩnh vực này dựa trên các kỹ thuật khai phá dữ liệu [1].

Hiện nay hầu hết các cơ quan, tổ chức, doanh nghiệp đều có hệ thống mạng máy tính riêng kết nối với mạng Internet và ứng dụng nhiều các chương trình, phần mềm CNTT vào các hoạt động sản xuất kinh doanh. Việc làm này đã góp phần tích cực trong quản lý, điều hành, kết nối, quảng bá và là chìa khoá thành công cho sự phát triển chung của họ và cộng đồng. Trong các hệ thống mạng máy tính đó có chứa rất nhiều các dữ liệu, các thông tin quan trọng liên quan đến hoạt động của các cơ quan, tổ chức, doanh nghiệp.

Sự phát triển mạnh của hệ thống mạng máy tính cũng là một vùng đất cónhiềuthuận lợi cho việc theo dõi và đánh cắp thông tin của các nhóm tội phạm tin học,việc xâm nhập bất hợp pháp và đánh cắp thông tin của các tổ

chức, đơn vị đangđặt ra cho thế giới vấn đề làm thế nào để có thể bảo mật được thông tin của tổ chức, đơn vị mình. Phát hiện xâm nhập bảo đảm an toàn an ninh mạng là những yếutố được quan tâm hàng đầu trong các các tổ chức, đơn vị. Đã có những đơn vị thực hiện việc thuê một đối tác thứ 3 với việc chuyên đảm bảo cho hệ thống mạng và đảm bảo an toàn thông tin cho đơn vị mình, cũng có những đơn vị đưa ra các kế hoạch tính toán chi phí cho việc mua sản phẩm phần cứng, phần mềmđể nhằm đáp ứng việc đảm bảo an toàn an ninh thông tin. Tuy nhiên đối với những giảipháp đó các tổ chức, đơn vị đều phải thực hiện cân đối về chính sách tài chínhhằng năm với mục đích làm sao cho giải pháp an toàn thông tin là tối ưu và cóđược chi phí rẻ nhất và đảm bảo thông tin trao đổi được an toàn, bảo vệ thôngtin của đơn vị mình trước những tấn công của tội phạm công nghệ từ bên ngoàido vậy mà đề tài Kỹ thuật phân cụm dữ liệu trong phát hiện xâm nhập trái phép dựa trên mã nguồn mở đượcphát triển giúp được phần nào yêu cầu của các tổ chức, đơn vị về an toàn thông tinvà đảm bảo an toàn cho hệ thống mạng.

Đề tài “Kỹ thuật phân cụm dữ liệu trong phát hiện xâm nhập trái phép” học viên thực hiện với mong muốn xây dựng một cách hệ thống về các nguy cơ tiềm ẩn về xâm nhập trái phép vào mạng máy tính, các phương pháp phân cụm dữ liệu và cụ thể cách thức để ứng dụng kỹ thuật phân cụm dữ liệu trong phát hiện xâm nhập trái phép, đảm bảo an toàn an ninh thông tin cho tổ chức, đơn vị.