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

ĐINH MẠNH CƯỜNG

PHÁT HIỆN XÂM NHẬP DỰA TRÊN THUẬT TOÁN K-MEANS

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên 2015

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

ĐINH MẠNH CƯỜNG

PHÁT HIỆN XÂM NHẬP DỰA TRÊN THUẬT TOÁN K-MEANS

Chuyên ngành: Khoa học máy tính

Mã số: 60.48.01.01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC
PGS.TS. NGUYỄN VĂN TAM

Thái Nguyên 2015

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn "**Phát hiện xâm nhập dựa trên thuật toán K-Means"** là công trình nghiên cứu do tôi thực hiện dưới sự hướng dẫn của PGS.TS. Nguyễn Văn Tam. Các nội dung được trình bày trong luận văn là những kết quả đạt được trong thời tôi gian thực đề tài dưới sự hướng của tập thể giáo viên hướng dẫn, tôi không sao chép nguyên bản lại kết quả của các nghiên cứu đã từng được công bố và đây cũng là kết quả của quá trình nghiên cứu, học tập và làm việc nghiêm túc của tôi trong quá trình học cao học. Bên cạch đó, trong một số nội dung luận văn là kết quả phân tích, nghiên cứu, tổng hợp từ nhiều nguồn tài liệu khác. Các thông tin tổng hợp hay các kết quả lấy từ nhiều nguồn tài liệu khác đã được tôi trích dẫn một cách đầy đủ và hợp lý. Nguồn tài tài liệu tham khảo có xuất xứ rõ ràng và được trích dẫn hợp pháp.

Các số liệu và thông tin sử dụng trong luận văn này là trung thực.

Thái Nguyên, ngày 20 tháng 07 năm 2015

Người cam đoan

Đinh Mạnh Cường

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn các thầy, cô trong Viện Công nghệ thông tin, Trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên đã tham gia giảng dạy, giúp đỡ tôi trong suốt quá trình học tập nâng cao trình độ kiến thức để phục vụ cho công tác giảng dạy của tôi hiện tại và sau này.

Tôi xin bày tỏ lòng biết ơn chân thành tới PGS.TS. Nguyễn Văn Tam, các Thầy đã tận tình hướng dẫn hướng dẫn tôi trong suốt thời gian thực hiện luận văn.

Vì điều kiện thời gian và trình độ có hạn nên luận văn cũng không thể tránh khỏi những thiếu sót. Tôi xin kính mong các Thầy, Cô giáo, các bạn đồng nghiệp đóng góp ý kiến để đề tài được hoàn thiện hơn.

Tôi xin chân thành cảm ơn!

MỤC LỤC

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG

BẢNG TỪ VIẾT TẮT

IDS	Intrusion Detection System
IPS	Intrusion Prevention Systems
HIDS	Host-based IDS
NIDS	Network-based IDS

MỞ ĐẦU

Ngày nay, hệ thống mạng máy tính đã trở nên rất phổ biến và được ứng dụng trong hầu hết các hoạt động kinh tế-xã hội của nước ta. Tuy nhiên, mạng máy tính cũng phải đương đầu với nhiều thách thức, đặc biệt là vấn đề an toàn và bảo mật dữ liệu trên mạng. Trong các mối đe dọa đối với an ninh mạng thì việc xâm nhập mạng để thay đổi thông tin, lấy cắp dữ liệu và phá hoại hạ tầng mạng là nghiêm trọng nhất. Chính vì vậy, việc phát hiện và ngăn chặn xâm nhập mạng máy tính là chủ đề đang được quan tâm nghiên cứu và phát triển ứng dụng mạnh mẽ hiện nay. Phát hiện và ngăn chặn được hiểu là xác định xâm nhập và ngăn chặn một cách nhanh nhất khi nó xảy ra. Hiện nay không có phương pháp phát hiện truy nhập trái phép nào là hoàn hảo bởi các kĩ thuật xâm nhập ngày càng tinh vi và luôn luôn được đổi mới. Khi phương pháp phát hiện xâm nhập được biết đến thì những kẻ xâm nhập sẽ sửa những chiến lược và thử một kiểu xâm nhập mới.

Trong quá trình học tập và nghiên cứu tôi đã tìm hiểu được một số kiến thức về phát hiện xâm nhập và tìm hiểu các thuật toán để phát hiện xâm nhâp đó nhằm đảm bảo thông tin trao đổi được đảm bảo an toàn. Chính vì thế tôi đã lựa chọn chủ đề "**Phát hiện xâm nhập dựa trên thuật toán K-means.**" là đề tài nghiên cứu cho luận văn của mình.

* Cấu trúc của luận văn bao gồm 3 chương như sau:

Chương 1: Chương này trình bày nhưng kiến thức cơ bản về phát hiện xâm nhập như: định nghĩa, các thành phần và chức năng của hệ thống, phân loại, và các phương pháp phát hiện xâm nhập.

Chương 2: Chương này trình bày về việc phát hiện xâm nhập dựa trên thuật toán K-means. Nội dung của thuật toán, ví dụ minh họa thuật toán, tập dữ liệu kiểm thử và mô hình phát hiện xâm nhập dựa trên thuật toán K-means.

Chương 3: Chương này là kết quả cài đặt bài toán phát hiện xâm nhập dựa trên thuật toán k-means.

Chương 1
KHÁI QUÁT BÀI TOÁN PHÁT HIỆN XÂM NHẬP

1.1. Định nghĩa về phát hiện xâm nhập

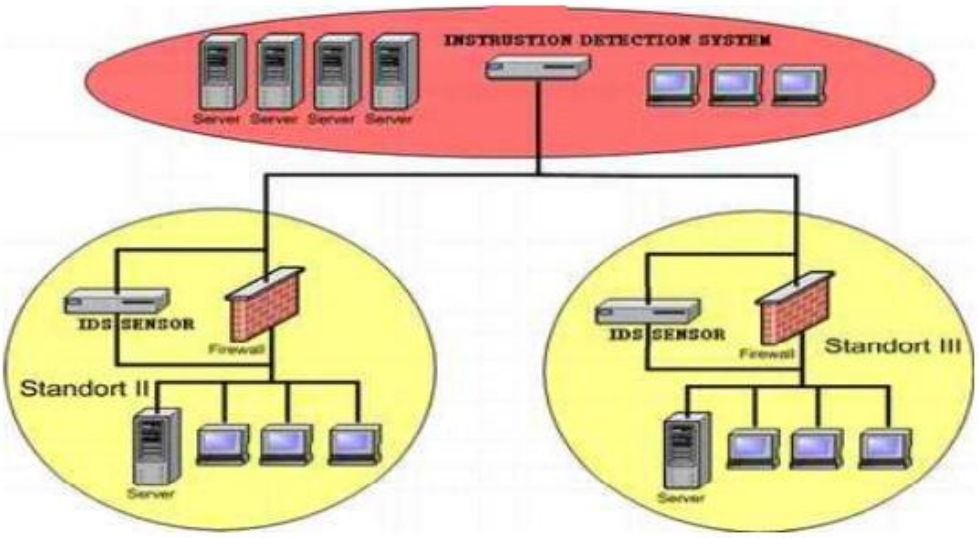
1.1.1. Định nghĩa

Hệ thống phát hiện xâm nhập (IDS) là hệ thống có nhiệm vụ theo dõi, phát hiện và (có thể) ngăn cản sự xâm nhập, cũng như các hành vi khai thác trái phép tài nguyên của hệ thống được bảo vệ mà có thể dẫn đến việc làm tổn hại đến tính bảo mật, tính toàn vẹn và tính sẵn sàng của hệ thống.[6]

Hệ thống IDS sẽ thu thập thông tin từ rất nhiều nguồn trong hệ thống được bảo vệ sau đó tiến hành phân tích những thông tin đó theo các cách khác nhau để phát hiện những xâm nhập trái phép.

Khi một hệ thống IDS có khả năng ngăn chặn các nguy cơ xâm nhập mà nó phát hiện được thì nó được gọi là một hệ thống phòng chống xâm nhập hay IPS.

Hình sau minh họa các vị trí thường cài đặt IDS trong mạng:



Hình 1.1: *Các vị trí đặt IDS trong mạng.*

1.1.2. Sự khác nhau giữa IDS/IPS

Có thể nhận thấy sự khác biệt giữa hai khái niệm ngay ở tên gọi: "phát hiện" và "ngăn chặn". Các hệ thống IDS được thiết kế với mục đích chủ yếu là phát hiện và cảnh báo các nguy cơ xâm nhập đối với mạng máy tính nó đang bảo vệ trong khi đó, một hệ thống IPS ngoài khả năng phát hiện còn có thể tự hành động chống lại