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

ĐẶNG THÀNH CÔNG

NGHIÊN CỨU KỸ THUẬT RAINBOW- CRACK THÁM KHÓA MÃ RC4 VÀ ỨNG DỤNG

Chuyên ngành: Khoa học máy tính
Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC
TS. NGUYỄN NGỌC CƯƠNG

Thái Nguyên, năm 2015

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Qua đây em xin chân thành cảm ơn toàn thể các thầy cô trong khoa đào tạo sau đại học trường Đại học Công nghệ Thông tin và Truyền thông và đặc biệt là Thầy TS. Nguyễn Ngọc Cương, đã tạo điều kiện thuận lợi và hướng dẫn em để hoàn thành luận văn này.

MỤC LỤC

DANH MỤC BẢNG BIỂU

DANH MỤC HÌNH ẢNH

LỜI MỞ ĐẦU

1.Tính cấp thiết của đề tài

RC4 là tên của thuật toán mã hóa được sử dụng trong WEP, MS-OFFICE... Một thuật toán mã hóa là một tập hợp các hoạt động mà chúng ta sử dụng để biến đổi văn bản chưa mã hóa thành mật mã. Nó sẽ hữu ích, trừ khi có một thuật toán giải mã tương ứng. Trong trường hợp của RC4, cùng một thuật toán được sử dụng để mã hóa và giải mã. Giá trị của một thuật toán mã hóa là ở khả năng bảo mật cao và dễ dàng trong sử dụng. Sức mạnh của một thuật toán được đo bằng độ khó để crack các bản mã được mã hóa bằng thuật toán đó. Chắc chắn là có các phương pháp mạnh hơn RC4. Tuy nhiên, RC4 là khá đơn giản để thực hiện và được coi là rất mạnh, nếu được sử dụng đúng cách.

Kỹ thuật đánh đổi bộ nhớ-thời gian (Time Memory Trade –Off) còn có tên gọi khác là đánh đổi không gian-thời gian dùng để chỉ việc sử dụng bộ nhớ lưu trữ dữ liệu tính toán trước với mục đích giảm thời gian tính toán đối với một thao tác cụ thể. Đây là kỹ thuật được áp dụng trong một số bài toán có thể chia các thao tác tính toán thành hai phần: tính toán trước và tra cứu dữ liệu đã chuẩn bị trước. Nếu tính toán và lưu trữ trước được càng nhiều thì thời gian giải một bài toán cụ thể sẽ chỉ tương đương với thời gian tra cứu.

Các phương tiện lưu trữ máy tính ngày một lớn hơn làm cho khả năng ứng dụng kỹ thuật TMTO ngày càng hiện thực. Đã có nhiều ứng dụng sử dụng kỹ thuật TMTO để giải quyết các vấn đề về tốc độ và bộ nhớ lưu trữ. Chẳng hạn, các bài toán liên quan đến tra cứu bảng dữ liệu, bài toán lưu trữ dữ liệu dạng nén, bài toán lưu trữ thuật toán, lưu trữ kết quả hình ảnh trong hiển thị công thức toán học trên trang HTML,...

Kỹ thuật mật mã cần làm việc với một không gian dữ liệu lớn (không gian khóa). Tuy nhiên, ở một số chế độ làm việc, có thể tổ chức tính toán sẵn các bản mã có thể của một bản rõ để thành lập một từ điển tra cứu cho phép mã hóa và giải mã nhanh. Mã thám có thể lợi dụng tính chất này để tấn công mật mã (kiểu tấn công Brute-Force) nếu có đủ bộ nhớ.

Đề tài luận văn này lựa chọn mật mã RC4 với độ dài khoá 40 bit để nghiên cứu. Đây là dạng RC4 ứng dụng trong nhiều phần mềm. Độ dài khóa là tương

đương với một số kết quả nghiên cứu ứng dụng kỹ thuật TMTO đã công bố đối với một số thuật toán mật mã khác. Kết quả nghiên cứu của đề tài sẽ là hướng mở cho nghiên cứu ứng dụng tấn công mật khẩu bảo vệ tệp văn bản soạn trên một số phần mềm xử lý văn bản. Đồng thời là kinh nghiệm cho mã thám viên đối với kỹ thuật TMTO có thể áp dụng cho tấn công nhiều dạng mật mã ứng dụng khác.

Đề tài luận văn tìm hiểu lý thuyết cơ bản về kỹ thuật TMTO, những vấn đề cần quan tâm trong ứng dụng, những cải tiến đã công bố gần đây cho kỹ thuật TMTO. Đề tài tiến hành áp dụng kỹ thuật TMTO vào thực tế tấn công một giải thuật mật mã cụ thể có khả năng triển khai ứng dụng thực tế.

2. Mục tiêu nghiên cứu:

- Nghiên cứu kỹ thuật Time Memory Trade-Off (TMTO) đánh đổi không gian lưu trữ với thời gian tấn công mật mã. Nghiên cứu về các cải tiến "Điểm phân biệt" và "Bảng cầu vồng" đã được công bố của kỹ thuật này.

- Sử dụng kỹ thuật TMTO được Oechslin áp dụng với cải tiến "Rainbow Crack" tấn công thám khoá mã RC4 ứng dụng trong phần mềm soạn thảo văn bản MS-WORD phiên bản 2007 của MicroSoft.

3. Nội dung nghiên cứu:

Luận văn được trình bày trong 3 chương, có phần mở đầu, phần kết luận, phần mục lục, phần tài liệu tham khảo. Các nội dung cơ bản của luận văn được trình bày theo cấu trúc như sau:

Chương 1: Mật mã RC4 và kỹ thuật Time-Memory Trade-Off áp dụng trong bài toán tấn công mật

Chương 2: Kỹ thuật tấn công Rainbow đối với RC4

Chương 3: Xây dựng chương trình tính toán tham số tấn công Rainbow đối với RC4

Bằng sự cố gắng nỗ lực của bản thân và đặc biệt là sự giúp đỡ tận tình, chu đáo của thầy giáo TS. Nguyễn Ngọc Cương , em đã hoàn thành luận văn đúng thời hạn. Do thời gian làm đồ án có hạn và trình độ còn nhiều hạn chế nên không thể tránh khỏi những thiếu sót. Em rất mong nhận được sự đóng góp ý kiến của các thầy cô cũng như là của các bạn sinh viên để bài luận văn này hoàn thiện hơn nữa.

Chương 1: MẬT MÃ RC4 VÀ KỸ THUẬT TIME-MEMORY TRADE –OFF ÁP DỤNG TRONG BÀI TOÁN TẤN CÔNG MẬT MÃ

Trong chương này là trình bày tập hợp các thông tin cơ sở về kỹ thuật TMTO; các cải tiến "điểm phân biệt" của Rivest và "bảng cầu vồng" của Oechslin. Nội dung chương làm rõ phương thức chia không gian tìm kiếm thành các bộ phận và tổ chức lưu trữ hiệu quả từng bộ phận không gian tìm kiếm. Đặc biệt là phương pháp tổ chức các "bảng cầu vồng" của Oechslin, phương pháp được ứng dụng hiệu quả trong phần mềm OPH-Crack. Thuật toán mật mã RC4 đóng vai trò trung tâm trong lược đồ xác thực mật khẩu. Bên cạnh đó là những phương pháp thám mã khác cũng đang được áp dụng nhiều trong thực tế.

1.1 Tổng quan về RC4

RC4 là tên của thuật toán mã hóa được sử dụng trong WEP, MS-OFFICE... Một thuật toán mã hóa là một tập hợp các hoạt động mà chúng ta sử dụng để biến đổi văn bản chưa mã hóa thành mật mã. Nó sẽ hữu ích, trừ khi có một thuật toán giải mã tương ứng. Trong trường hợp của RC4, cùng một thuật toán được sử dụng để mã hóa và giải mã. Giá trị của một thuật toán mã hóa là ở khả năng bảo mật cao và dễ dàng trong sử dụng. Sức mạnh của một thuật toán được đo bằng độ khó để crack các bản mã được mã hóa bằng thuật toán đó. Chắc chắn là có các phương pháp mạnh hơn RC4. Tuy nhiên, RC4 là khá đơn giản để thực hiện và được coi là rất mạnh, nếu được sử dụng đúng cách. Thật may mắn là RC4 khá đơn giản để thực hiện và mô tả. Ý tưởng cơ bản mã hóa RC4 là tạo ra một chuỗi các trình tự giả ngẫu nhiên (giả ngẫu nhiên) của các byte được gọi là khóa dòng, sau đó được kết hợp với các dữ liệu bằng cách sử dụng toán tử OR (XOR). Toán tử XOR kết hợp hai byte và tạo ra một byte duy nhất. Nó làm điều này bằng cách so sánh các bit tương ứng trong từng byte. Nếu chúng bằng nhau, kết quả là 0, nếu chúng khác nhau, kết quả là 1. Về mặt lý thuyết, RC4 không phải là một hệ thống mã hóa hoàn toàn an toàn bởi vì nó tạo ra một dòng giả ngẫu nhiên chính, không phải byte thực sự ngẫu nhiên. Nhưng nó đủ chắc chắn an toàn cho các ứng dụng, nếu được áp dụng đúng.

RC4 là mật mã có cỡ của khóa biến đổi do Ron Rivest phát triển vào những năm 1987 cho liên hợp an ninh dữ liệu RSA. Trong bảy năm nó là sở hữu độc

quyền và các chi tiết của thuật toán ta chỉ có được sau khi ký thỏa thuận không tiết lộ bí mật.

Vào tháng 9 năm 1994, một người lạc danh đã gửi mã nguồn qua bưu điện vào danh sách thư tín Cypherpunks, Nó nhanh chóng lan tỏa đến nhóm Usenet và qua Internet đến các site ftp trên thế giới. Liên hiệp an ninh dữ liệu RSA tuyên bố rằng nó vẫn còn là một bí mật thương mại mặc dù nó đã được công bố, nhưng việc này đã quá muộn. Bởi nó đã được thảo luận và phân tích kỹ trên Usenet, đọc phân phát ở các hội nghị và được đưa vào các giáo trình mật mã.

RC4 có địa vị xuất khẩu đặc biệt nếu độ dài khóa của nó là 40 bít hoặc ít hơn. Địa vị xuất khẩu đặc biệt này sẽ dẫn đến việc không có gì để làm đối với độ an toàn của thuật toán, mặc dù liên hợp an ninh dữ liệu RSA đã nói bóng gió trong nhiều năm rằng vẫn có. Tên thuật toán này ðýợc thýõng mại hóa do ðó bất kỳ ngýời nào viết mã riêng của mình ðều phải gọi nó bằng một cái tên khác. Các tài liệu bên trong khác của liên hợp an ninh dữ liệu RSA vẫn chưa được công bố.

RC4 là một phần mềm trong các sản phẩm mật mã thương mại, bao gồm Lotus Notes, Apple Computer's AOCE và ORACLE Security SQL. Nó là một bộ phận của bản chỉ dẫn kỹ thuật Cellular Digital Packed Data.

RC4 là một họ các thuật toán phụ thuộc vào các tham số nguyên dương, mà điển hình là trường hợp n= 8. Ở thời điểm t, trạng thái bên trong của RC4 gồm bảng

$S_1=(S_1(l))_{l=0}^{2^n-1}$ có từ n-bít và 2 con trỏ n-bít là it và jt. Do đó cỡ bộ nhớ trong la $M=n2^n+2n$(bít). Gọi Zt là từ ra n-bít của RC4 ở thời điểm t. Bít có nghĩa thấp nhất của một từ là bít ở bên trái nhất của nó.

1.2. Các kỹ thuật thám mã

1.2.1 WEP

WEP (Wired Equivalent Privacy) là một thuật toán nhằm bảo vệ sự trao đổi thông tin chống lại nghe trộm, chống lại những kết nối mạng không được cho phép cũng như chống lại việc thay đổi hoặc làm nhiễu thông tin truyền. WEP sử dụng stream cipher RC4 cùng với một mã 40 bit và một số ngẫu nhiên 24 bit (initialization vector - IV) để mã hóa thông tin. Thông tin mã hóa và IV sẽ được gửi đến người nhận. Người nhận sẽ giải mã thông tin dựa vào khóa WEP đã biết trước.