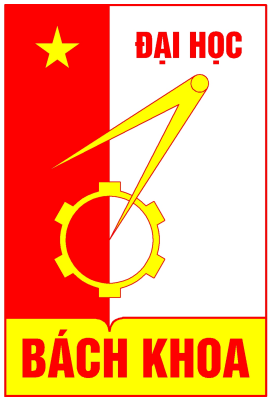
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

_____*******_____



ĐỒ ÁN TỐT NGHIỆP
NGÀNH KHOA HỌC MÁY TÍNH

Cài đặt bảo mật cho các thiết bị IoT

Sinh viên thực hiện:
Đặng Quang Trung 20134145 CNTT2.03-K58

Giảng viên:
TS. Trần Vĩnh Đức

Hà Nội 21-04-2018

Lời cảm ơn

Trước khi trình bày nội dung của đồ án này, tôi xin được gửi lời cảm ơn sâu sắc và chân thành nhất đến TS. Trần Vĩnh Đức, người đã tận tình hướng đẫn tôi trong suốt quá trình thực hiện đồ án này cũng như những năm tháng học tại trường Đại học Bách Khoa Hà Nội. Đồng thời tôi cũng xin bày tỏ lòng biết ơn đến các thầy cô trường Đại học Bách Khoa Hà Nội, Viện công nghệ thông tin và truyền thông, đặc biệt là các thầy cô bộ môn Khoa học máy tính đã tận tình chỉ dạy cho tôi trong những năm tháng học tập ở trường.

Đồng thời tôi xin gửi lời cảm ơn đến gia đình, bạn bè đã luôn ở bên tôi, đông viên và giúp đỡ tôi trong suốt quá trình học tập và thực hiện đồ án tốt nghiệp.

Mục lục

Bảng chữ viết tắt

Danh sách hình vẽ

Giới thiệu chung

Mật mã học là một lĩnh vực liên quan với các kỹ thuật ngôn ngữ và toán học để đảm bảo an toàn thông tin, cụ thể là trong thông tin liên lạc. Về phương diện lịch sử, mật mã học gắn liền với quá trình mã hóa, điều này có nghĩa là nó gắn với các cách thức để chuyển đổi thông tin từ dạng này sang dạng khác nhưng ở đây là từ dạng thông thường có thể nhận thức được thành dạng không thể nhận thức được, làm cho thông tin trở thành dạng không thể đọc được nếu như không có các kiến thức bí mật. Quá trình mã hóa được sử dụng chủ yếu để đảm bảo tính bí mật của các thông tin quan trọng, chẳng hạn trong công tác tình báo, quân sự hay ngoại giao cũng như các bí mật về kinh tế, thương mại. Trong những năm gần đây, lĩnh vực hoạt động của mật mã hóa đã được mở rộng: mật mã hóa hiện đại cung cấp cơ chế cho nhiều hoạt động hơn là chỉ duy nhất việc giữ bí mật và có một loạt các ứng dụng như: chứng thực khóa công khai, chữ ký số, bầu cử điện tử hay tiền điện tử. Ngoài ra, những người không có nhu cầu thiết yếu đặc biệt về tính bí mật cũng sử dụng các công nghệ mật mã hóa, thông thường được thiết kế và tạo lập sẵn trong các cơ sở hạ tầng của công nghệ tính toán và liên lạc viễn thông.

Mật mã học là một lĩnh vực liên ngành, được tạo ra từ một số lĩnh vực khác. Các dạng cổ nhất của mật mã hóa chủ yếu liên quan với các kiểu mẫu trong ngôn ngữ. Gần đây thì tầm quan trọng đã thay đổi và mật mã hóa sử dụng và gắn liền nhiều hơn với toán học, cụ thể là toán học rời rạc, bao gồm các vấn đề liên quan đến lý thuyết số, lý thuyết thông tin, độ phức tạp tính toán, thống kê và tổ hợp. Mật mã hóa cũng được coi là một nhánh của công nghệ, nhưng nó được coi là không bình thường vì nó liên quan đến các sự chống đối ngầm (xem công nghệ mật mã hóa và công nghệ an ninh). Mật mã hóa là công cụ được sử dụng trong an ninh máy tính và mạng.

Hiện nay, cùng với sự phát triển của tính toán khắp nơi, các hệ thống vạn vật kết nối (internet of things - IoT) ngày càng thu hút được sự quan tâm của các chuyên gia cũng như các nhà ứng dụng. Vấn đề an toàn thông tin trong hệ thống IoT với các thiết bị nhỏ gọn, năng lực tính toán thấp, trở thành một chủ đề nóng hiện nay. Với khả năng tính nhanh, an toàn chi phí thấp, mật mã tính toán trên đường công Elliptic tiêu biểu cho hệ mã trao đổi khóa của các tehiết bị IoT.

Nội dung chính của đồ án bao gồm 4 chương:

- **Chương 1:** Cơ sở lý thuyết
- **Chương 2:** Mật mã dựa trên đường cong Elliptic
- **Chương 3:** Chứng thư số ẩn dựa trên đường cong Elliptic
- **Chương 4:** Cài đặt cho thiết bị IoT

Chương 1

Cơ sở lý thuyết

Trong chương này sẽ trình bày cơ sở lý thuyết chung bao quát cho các hệ mật mã, nó sẽ cung cấp một số khái niệm và các kiến thức qúy giá, hỗ trợ đắc lực cho việc làm quen với lĩnh vực này.

1.1 Cơ sở mật mã

Mật mã là một lĩnh vực khoa học chuyên nghiên cứu về các phương pháp và kỹ thuật đảm bảo an toàn và bảo mật trong truyền tin liên lạc với giả thiết sự tồn tại của các thế lực thù địch, những kẻ muốn ăn cắp thông tin để lợi dụng và phá hoại. Tên gọi trong tiếng Anh, Cryptology đợưc dẫn giải nguồn gốc từ tiếng Hy lạp, trong đó kryptos nghĩa là "che dấu", logos nghĩa là "từ ngữ".

Các nhà nghiên cứu lĩnh vực này quan tâm xây dựng hoặc phân tích (để chỉ ra điểm yếu) các giao thức mật mã (cryptographic protocols), tức là các phương thức giao dịch có đảm bảo mục tiêu an toàn cho các bên tham gia (với giả thiết môi trường có kẻ đối địch, phá hoại).

Ngành Mật mã (cryptology) thường được quan niệm như sự kết hợp của 2 lĩnh vực con:

1. Sinh, chế mã mật (cryptography): nghiên cứu các kỹ thuật toán học nhằm cung cấp các công cụ hay dịch vụ đảm bảo an toàn thông tin.
2. Phá giải mã (cryptanalysis): nghiên cứu các kỹ thuật toán học phục vụ phân tích phá mật mã và/hoặc tạo ra các đoạn mã giản nhằm đánh lừa bên nhận tin

Hai lĩnh vực con này tồn tại như hai mặt đối lập, "đấu tranh để cùng phát triển" của một thể thống nhất là ngành khoa học mật mã (cryptology).

Mặc dù mật mã có thể coi là một ngành toán học phát triển cao, đòi hỏi tư duy cao để nắm được các thành tựu hiện đại của nó, nhưng cơ sở xuất phát ban đầu của nó lại là một mô hình thực tiễn khá đơn giản như sau:
hình vẽ

Như vậy trong một hệ thống mật mã khái quát sẽ có các thành phần sau:

- **Văn bản trơn**(plaintext X): tức là thông điệp nguyên gốc chưa được mã hóa.

- **Văn bản mã hóa**(ciphertext Y): tức là thông điệp đã được mã hóa.
- **Thuật toán mã hóa**(enciphering algorithm $E_z(X)$): là các giao thức hoặc hướng dẫn có tác dụng chuyển đổi văn bản trơn thành văn bản mã hóa. Đối với các hệ thống mật mã truyền thống, chỉ có người gửi thông điệp biết được thuật toán mã hóa, tuy nhiên đối với các hệ thống dùng mật mã hóa khóa công khai (Public key code - PKC), tất cả mọi người đều có thể biết thuật toán mã hóa mà không ảnh hưởng tiêu cực đến an ninh của hệ thống.
- **Khóa mã hóa**(enciphering key Z): là một hoặc nhiều đối tượng (thường là các con số hay là các hướng dẫn quan trọng nào đó) được dùng trong việc mã hóa văn bản trơn. Ngoại trừ trong hệ thống PKC, để đảm bảo bí mật an toàn thì khóa mã hóa thường chỉ được người gửi biết.
- **Thuật toán giải mã**(deciphering algorithm $D_z(Y)$): là các giao thức hoặc hướng dẫn có tác dụng chuyển đổi văn bản mã hóa trở về văn bản trơn. Để đảm bảo bí mật, chỉ có người nhận thông điệp biết được thuật toán giải mã.
- **Khóa giải mã**(deciphering key Z'): là một hoặc nhiều đối tượng (thường là các con số hay là các hướng dẫn quan trọng nào đó) được dùng trong việc giải mã văn bản bị mã hóa. Để đảm bảo bí mật, chỉ có người nhận thông điệp biết được khóa giải mã.
- **Sản phẩm mật mã**(Cryptography Product): bao gồm các hệ thống thiết bị, module, mạch tích hợp và các chương trình phần mềm mã hoá chuyên dụng có tích hợp các thuật toán mật mã, được thiết kế, chế tạo để bảo vệ thông tin giao dịch điện tử và lưu trữ dưới dạng số hoá, trong đó sử dụng "Thuật toán mã đối xứng" hoặc "Thuật toán mã không đối xứng".

1.2 Mật mã khóa đối xứng

Trong mật mã học, các thuật toán khóa đối xứng (symmetric-key algorithms) là một lớp các thuật toán mật mã hóa trong đó các khóa dùng cho việc mật mã hóa và giải mã có quan hệ rõ ràng với nhau (có thể dễ dàng tìm được một khóa nếu biết khóa kia). Mã khóa loại này không công khai.

Khóa dùng để mã hóa có liên hệ một cách rõ ràng với khóa dùng để giải mã có nghĩa chúng có thể hoàn toàn giống nhau, hoặc chỉ khác nhau nhờ một biến đổi đơn giản giữa hai khóa. Trên thực tế, các khóa này đại diện cho một bí mật được phân hưởng bởi hai bên hoặc nhiều hơn và được sử dụng để giữ gìn sự bí mật trong kênh truyền thông tin.

1.3 Mật mã khóa công khai

Giới thiệu

Mật mã hóa khóa công khai là một dạng mật mã hóa cho phép người sử dụng trao đổi các thông tin mật mà không cần phải trao đổi các khóa chung bí mật trước đó. Điều này được thực hiện bằng cách sử dụng một cặp khóa có quan hệ toán học với nhau là khóa công khai và khóa cá nhân (hay khóa bí mật).

Trong hệ mã khóa công khai, mỗi người sử dụng có hai khoá, một được gọi là khoá bí mật (secret key hay private key) và một được gọi là khoá công khai (public key). Khoá thứ nhất chỉ mình user biết và giữ bí mật, còn khoá thứ hai thì anh ta có thể tự do phổ biến công khai. Khoá thứ nhất thường đi liền với thuật toán giải mã, còn khoá thứ hai thường đi liền với thuật toán sinh mã, tuy nhiên điều đó không phải là bắt buộc. Ta hãy ký hiệu chúng là z (khóa riêng) và Z (khóa công khai)

Hoạt động của chúng là đối xứng

$$X = D(z, E(Z, X)) \tag{1.1}$$
$$X = E(Z, D(z, X)) \tag{1.2}$$

Trong đó hệ thức (1.1) biểu tượng cho bài toán truyền tin mật: bất kỳ người sử dụng nào khác như B,C,D ... muốn gửi tin cho A chỉ việc mã hoá thông tin với khoá công khai (Z_A) của A rồi gửi đi. Chỉ có A mới có thể khoá riêng để giải mã (z_A) và đọc được tin, kẻ nghe trộm Eve không thể giải mã để lấy được tin vì không có khoá z_A.

Còn hệ thức (1.2) sẽ được sử dụng để xây dựng các hệ chữ ký điện tử trong đó thao tác Ký chính là thực hiện $E(Z_A)$ còn kiểm định chữ ký là thông qua gọi $D(z_A)$.

Hệ mật mã theo nguyên tắc nói trên được gọi là hệ mã với khoá công khai (public key cryptosystems) hay còn được gọi là mã khóa phi đối xứng (asymmetric key cryptosystems). Ta sẽ viết tắt hệ thống kiểu này bằng PKC.

Hệ thống mật mã hóa khóa công khai có thể sử dụng với các mục đích:

- **Mã hóa**: giữ bí mật thông tin và chỉ có người có khóa bí mật mới giải mã được
- **Tạo chữ ký số**: cho phép kiểm tra một văn bản có phải đã được tạo với một khóa bí mật nào đó hay không.
- **Thỏa thuận khóa**: cho phép thiết lập khóa dùng để trao đổi thông tin mật giữa 2 bên.

Thông thường, các kỹ thuật mật mã hóa khóa công khai đòi hỏi khối lượng tính toán nhiều hơn các kỹ thuật mã hóa khóa đối xứng nhưng những lợi điểm mà chúng mang lại khiến cho chúng được áp dụng trong nhiều ứng dụng.

Hệ thống khóa công khai RSA

RSA là hệ mật mã khóa công khai phổ biến và cũng đa năng nhất trong thực tế, phát minh bởi Rivest, Shamir & Adleman (1977). Nó là chuẩn mật mã bất thành văn đối với PKC, cung cấp đảm bảo tính mật, xác thực và chữ ký điện tử.

Cơ sở thuật toán RSA dựa trên tính khó của bài toán phân tích các số lớn ra thừa số nguyên tố: không tồn tại thuật toán thời gian đa thức (theo độ dài của biểu diễn nhị phân của số đó) cho bài toán này. Chẳng hạn, việc phân tích một hợp số là tích của 2 số nguyên tố lớn hàng trăm chữ số sẽ mất hàng ngàn năm tính toán với một máy PC trung bình có CPU khoảng trên 2Ghz.

Thuật toán RSA

Xây dựng: Chọn các tham số

1. *Chọn hai số nguyên tố lớn p và q. Tính* $n = p \times q$ *và* $m = \phi(n) = (p-1) \times (q-1)$.
2. *Chọn e,* $1 \leq e \leq m-1$*, sao cho* $gcd(e, m) = 1$.
3. *Tìm d sao cho* $e * d = 1 \ (mod\ m)$*, tức là tính* $d = e^{-1} \ (mod\ m)$*, giải theo thuật toán gcd mở rộng.*

Khóa công khai (Public key) là (e, n)

Khoá dùng riêng (Private key) là d, p, q)

Giả sử X là một khối tin gốc (plaintext), Y là một khối mã tương ứng của X, và (z_A, Z_A) là các thành phần công khai và riêng của khoá của Alice

Sinh Mã. Nếu Bob muốn gửi một thông báo mã hoá cho Alice thì anh ta chỉ việc dùng khoá công khai của Alice để thực hiện:

$$Y = E_{Z_A}(X) = X^e \pm n$$

Giải mã. Khi Alice muốn giải mã Y, cô ta chỉ việc dùng khoá riêng $z_A = d$ để thực hiện như sau:

$$D_{z_A}(Y) = Y^d \pm n$$

Hệ Rabin

Hệ Rabin cũng xây dựng trên việc lấy $n = p \times q$ làm bí mật. N được coi là khoá công khai (PK) còn (p,q) là khoá bí mật (SK).

Mã hoá là việc thực hiện:

$$Y = X^2 (mod\ n)$$

còn giải mã là việc tính căn bậc hai:

$$X = \sqrt{Y} (mod\ n) \tag{1.3}$$

Có thể thấy, nếu biết $n = p \times q$ thì dễ dàng tìm được nghiệm cho phương trình này, còn nếu không thì việc tìm nghiệm là khó tương đương với bài toán PTTSNT số n.

Khi biết $N = p \times q$ thì 1.3 được giải ra có bốn nghiệm do đó để xác định được đâu là bản rõ gốc phải có mẹo để chọn được đúng giá trị cần thiết trong số 4 nghiệm đó

Hệ Rabin có một số ưu điểm so với RSA:

- Tính an toàn được chứng minh hoàn toàn tương đương với bài toán PTTSNT, nói cách khác tính ATBM của Rabin là có thể chứng minh được (provable)
- Ngoại trừ trường hợp RSA hoạt động với e nhỏ còn thuật toán sinh mã của Rabin nhanh hơn nhiều so với RSA là hệ đòi hỏi phải tính luỹ thừa. Thời gian giải mã thì tương đương nhau.

Nhược điểm: Vì phương trình giải mã cho 4 nghiệm nên làm khó dễ việc giải mã. Thông thường, bản rõ trước khi được mã hoá cần được nối thêm vào đuôi một chuỗi số xác định để làm dấu vết nhận dạng (chẳng hạn nối thêm 20 số 0 – như vậy trong số 4 nghiệm giải ra, chuỗi nào tận cùng bằng 20 con 0 thì đúng là bản rõ cần nhận). Vì lý do này nên Rabin thường được dùng chủ yếu cho chứng thực (chữ ký điện tử).

Hệ El-Gamal

Tạo khóa:

Alice chọn một số nguyên tố p và hai số nguyên ngẫu nhiên g và u, cả hai đều nhỏ hơn p. Sau đó tính

$$y = g^u (mod\ p)$$

Bây giờ khóa công khai của Alice được lấy là (p,g,y), khoá mật là u.

Sinh mã:

1. Nếu Bob muốn mã hoá một tin X để truyền cho Alice thì trước hết anh ta chọn một số ngẫu nhiên k sao cho $(k, p-1) = 1$
2. Tính
$$a = g^k (mod\ p) b = y^k X (mod\ p)$$

Mã là $Y = (a, b)$ và có độ dài gấp đôi bản rõ.

Giải mã: Alice nhận được $Y = (a, b)$ và giải ra X theo công thức sau:

$$X = \frac{b}{a^u} (mod\ p)$$

1.4 Hàm băm

1.5 Chữ ký điện tử

1.6 Chứng thư số

Chương 2

Mật mã dựa trên đường cong Elliptic

2.1 Đường cong Elliptic

2.2 Mật mã dựa trên đường cong Elliptic

Chương 3

Chứng thư số ẩn dựa trên đường cong Elliptic

Chương 4

Cài đặt cho thiết bị IoT

Kết luận và hướng phát triển

Tài liệu tham khảo

[1] http://www.csplib.org/Problems/prob030/

[2] Slide Trí tuệ nhân tạo thầy Nguyễn Nhật Quang. Chương 5 : Thỏa mãn ràng buộc

[3] Stuart Russell and Peter Norvig *Artificial Intelligence: A Modern Approach* 2nd edition, Prentice Hall, page 137, 2003.

[4] Ha Quang Minh and Deville Yves and Pham Quang Dung and Ha Minh Hoang, *On the min cost traveling salesman problem with drone*, arXiv preprint arXiv:1509.08764, 2015

[5] S. Banker, Amazon and drones – here is why it will work (dec 2013). URL http://www.forbes.com/sites/stevebanker/2013/12/19/amazon-drones-here-is-why-it-will-work/

[6] Nguyễn Đức Nghĩa, Nguyễn Tô Thành *Toán rời rạc* 3rd edition, Nhà xuất bản đại học quốc gia Hà Nội, page 107-108, 2006.

[7] Slide Tối ưu hóa tổ hợp thầy Nguyễn Đức Nghĩa. Chương mở đầu, Bài toán vận tải.

[8] Slide Tìm kiếm cục bộ dựa trên ràng buộc thầy Phạm Quang Dũng. Chương 5: Contraint-base local search applications.

[9] Slide Phân tích và thiết kế thuật toán thầy Nguyễn Đức Nghĩa. Chương 5: Quy hoạch động, Chương 3: Greedy Algorithms.

[10] Francesca Rossi, Peter van Beek, Toby Walsh *Handbook of Constraint Programming* 1st edition, Elsevier Science, page 107-108, 2006.

[11] Bài giảng môn tính toán tiến hóa cô Huỳnh Thị Thanh Bình. Thuật giải di truyền.

[12] Rupert Howell,Jonathon Wong *Apache OFBiz Development: The Beginner's Tutorial*, Paperback , page 53, 2008. truyền.

[13] Ruth Hoffman *Apache OFBiz Cookbook*, Paperback , page 28, 2010.

[14] Apache Ofbiz (wikipedia) https://en.wikipedia.org/wiki/Apache_OFBiz