BỘ GIÁO DỤC VÀ ĐÀO TẠO

ĐẠI HỌC ĐÀ NẴNG

NGUYỄN VĂN MINH

NGHIÊN CỨU GIẢI THUẬT ADVANCES ENCRYPTION STANDARD MÃ HÓA VĂN BẢN MẬT TẠI TRƯỜNG CAO ĐẲNG CÔNG KỸ NGHỆ ĐÔNG Á

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2011

Công trình được hoàn thành tại

ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **TS. NGUYỄN THANH BÌNH**

Phản biện 1: **PGS.TS. VÕ TRUNG HÙNG**

Phản biện 2: **TS. TRƯƠNG CÔNG TUẤN**

Luận văn được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 15 tháng 10 năm 2011

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng

- Trung tâm Học liệu, Đại học Đà Nẵng

MỞ ĐẦU

1. Lý do chọn đề tài

Ngày nay với sự xuất hiện của máy tính, các tài liệu văn bản giấy tờ và các thông tin quan trong đều được số hóa và xử lý trên máy tính, được truyền đi trên một môi trường mà mặc định là không an toàn. Do đó yêu cầu về việc có một cơ chế, giải pháp để bảo vệ sự an toàn và bí mật của các thông tin nhạy cảm, quan trọng ngày càng trở nên cấp thiết.Việc bảo vệ dữ liệu là vấn đề mà tất cả những ai sử dụng máy tính quan tâm. Dữ liệu của mỗi người dùng có thể khác nhau từ các tập tin văn bản đến các chương trình máy tính hoặc các dữ liệu rất quan trọng như tài khoản trong ngân hàng, bí mật quốc gia. Tùy theo tầm quan trọng của tài liệu mà người dùng lựa chọn các phương pháp khác nhau. Trong đó, đảm bảo tính bảo mật nội dung của tập tin là cần thiết. Vì vậy, các tập tin sau khi tạo cần được mã hóa nhằm tăng tính bảo mật và độ an toàn của dữ liệu. Mã hóa được xem là mức bảo vệ tối ưu nhất đối với dữ liệu giúp thông tin không bị lộ và nâng cao độ an toàn trong các giao dịch truyền tải thông tin.

Mật mã học – Cryptography (hay crypto) – ngành khoa học nghiên cứu về việc giấu thông tin. Cụ thể hơn, mật mã học là ngành học nghiên cứu về những cách chuyển đổi thông tin từ dạng "có thể hiểu được" (bản rõ – plaintext) thành dạng "không thể hiểu được" (bản mã – ciphertext) và ngược lại. Mật mã học giúp đảm bảo tính bí mật của thông tin – thông tin chỉ tiết lộ cho những ai được phép; tính toàn vẹn – thông tin không thể bị thay đổi mà không phát hiện; tính xác thực – người gởi hoặc người nhận có thể chứng minh đúng họ; đảm bảo tính không chối bỏ, tức là người gởi hoặc nhận sau này không thể chối bỏ việc đã gởi hoặc nhận thông tin.

Trên thực tế đã có nhiều phương pháp mã hóa và các thuật toán tương ứng với mỗi phương pháp được ứng dụng để mã hóa thông tin như phương pháp mã hóa đối xứng với các thuật toán Data Encryption Standard – DES,

Triple DES, Advanced Encryption Standard – AES; phương pháp mã hóa bất đối xứng với khóa công khai gồm các thuật toán nổi tiếng: Diffle-Hellman, RSA... Trong các thuật toán kể trên, AES còn được biết với một tên gọi khác là thuật toán Rijndael do hai tác giả Vincent Rijmen, Joan Daeman đề nghị và được Viện tiêu chuẩn và công nghệ quốc gia Hoa Kỳ - NIST (National Institue of Standards and Technology) chọn làm chuẩn mã hóa nâng cao trong công báo FIPS PUB 197. AES được đánh giá là một thuật toán tiên tiến áp dụng rộng rãi trong các hệ thống khác nhau từ các thẻ thông minh cho đến các máy tính cá nhân kể cả phần cứng và phần mềm.

Trong phạm vi của trường Cao đẳng công kỹ nghệ Đông Á có rất nhiều tài liệu quan trọng cần được lưu giữ bí mật. Điển hình là các đề thi tốt nghiệp cuối khóa kèm đáp án, hồ sơ sinh viên, các công trình nghiên cứu khoa học, các đề án chưa được công bố. Công tác đảm bảo bí mật cho các tài liệu của nhà trường hiện chưa được quan tâm đúng mức. Do đó, với những kỹ thuật tinh vi, lợi dụng các lỗ hổng của hệ thống, tội phạm có thể dễ dàng tiếp cận các thông tin mật đó.

2. Mục đích nghiên cứu và kết quả đạt được

Xuất phát từ cơ sở phân tích, tôi chọn đề tài **"Nghiên cứu giải thuật Advances Encryption Standard mã hóa văn bản mật tại trường Cao đẳng công kỹ nghệ Đông Á"** nhằm nghiên cứu giải thuật mã hóa đối xứng Advanced Encryption Standard - AES để đưa ra một phương pháp mã hóa tập tin tối ưu và xây dựng hệ thống mã hóa các tài liệu mật đảm bảo độ an toàn của dữ liệu, tin cậy, bí mật của tài liệu.

Mục đích của tác giả là tìm hiểu và nghiên cứu sâu các lý thuyết liên quan đến mã hóa dữ liệu, so sánh các ưu điểm, nhược điểm của một số thuật toán hiện tại, đồng thời nêu ra các ứng dụng điển hình liên quan đến các thuật toán trên. Trong đó, tôi tập trung nghiên cứu sâu giải mã hóa dữ liệu tiên tiến nhất hiện hay, AES, và một số cơ chế mã hóa tập tin đã có để từ đó đưa ra một giải pháp mã hóa tập tin đảm bảo các tính chất, đặc điểm của vấn đề bảo mật thông tin.

3. Phương pháp và phạm vi nghiên cứu

Dựa trên cơ sở thu thập các tài liệu liên quan của các tác giả trong và ngoài nước để tìm hiểu và phân tích các thông tin liên quan đến luận văn. Nghiên cứu và cài đặt giải thuật AES trên môi trường Windows. Sau đó, kiểm thử và đánh giá kết quả.

Đề tài chỉ nghiên cứu ứng dụng mã hóa các văn bản quan trọng tại trường Cao đẳng Công Kỹ Nghệ Đông Á.

4. Cấu trúc của luận văn

Cấu trúc của đề tài được tổ chức thành 03 chương.

Chương 1. Tổng quan mã hóa thông tin

Chương 2. Thuật toán Advanced Encryption Standard

Chương 3. Ứng dụng thuật toán Advanced Encryption Standard mã hóa các văn bản mật trong trường Cao đẳng công kỹ nghệ Đông Á

CHƯƠNG 1. TỔNG QUAN MÃ HÓA THÔNG TIN

Chương 1 này tập trung nghiên cứu một số lý luận, các cơ sở lý thuyết về mật mã học đồng thời giới thiệu một số thuật toán phổ biến.

1.1. Các khái niệm

1.1.1. Mật mã học

Định nghĩa 1.1:

Mật mã học là nghiên cứu các thuật toán liên quan đến khía cạnh bảo mật thông tin như thông tin mật, ràng buộc dữ liệu, chứng thực thực thể, chứng thực nguồn gốc dữ liệu ***Error! Reference source not found.***.

Một số khái niệm trong mật mã học gồm: Mã hóa (encrypt hay encipher), Giải mã (Decrypt hay decipher), Bản rõ (Plaintext), Cipher (hay cypher), Khóa (Key).

1.1.2. Hệ mật mã (Crypto System):

Định nghĩa 1.2:

Một hệ mật mã là bộ 5 (P,C,K,E,D) thõa mãn các tính chất sau:

1. *P là không gian bản rõ: là tập hữu hạn các bản rõ có thể có.*

2. *C là không gian bản mã: là tập hữu hạn các bản mã có thể có.*
3. *K là không gian khóa: là tập hữu hạn các khóa có thể có.*
4. *Đối với mỗi $k \in K$, có một quy tắc mã hóa $e_k \in E$ và một quy tắc giải mã tương ứng $d_k \in D$. Với mỗi e_k: $P \rightarrow C$ và d_k: $C \rightarrow P$ là những hàm mà $d_k(e_k(x))=x$ cho mọi bản rõ $x \in P$. Hàm giải mã d_k chính là ánh xạ ngược của hàm mã hóa e_k..*

1.1.3. Nguyên tắc Kerckhoffs

Một hệ mật mã sẽ được an toàn ngay cả khi tất cả mọi thứ trên hệ thống đó là công khai ngoại trừ khóa (key).

"Thuật toán mã hóa được tạo ra không cần phải giữ bí mật, có thể được công bố công khai, rơi vào tay quân địch mà không có bất kỳ sự phiền phức nào cả".

1.2. Tính chất của mã hóa thông tin

Mã hóa thông tin phải đảm bảo các tính chất sau: Tính bí mật (Confidentiality), tính xác thực (Authentication), tính toàn vẹn (Integrity), tính không thể chối bỏ (Non-repudation).

1.3. Độ an toàn của hệ mật mã

Độ an toàn của thuật toán phụ thuộc vào độ phức tạp của nó. Các yếu tố xem xét thuật toán an toàn là chi phí hay phí tổn; thời gian cần thiết để phá vỡ; lượng dữ liệu để phá vỡ.

Có hai phương pháp được Claude Shannon trình bày vào năm 1949 để đánh giá độ an toàn của một hệ mật mã.

1.3.1. Độ an toàn tính toán:

1.3.2. Độ an toàn không điều kiện

1.4. Các phương pháp mã hóa

1.4.1. Mã hoá cổ điển (Classical cryptography)

Phương pháp này là tiền thân của các phương pháp mã hóa đối xứng ngày nay. Có hai phương pháp nổi bật đó là: *Mã hoá thay thế (Substitution Cipher), Mã hoá hoán vị (Transposition Cipher)*

1.4.2. Mã hoá đối xứng (Symetric cryptography)

Mã hoá đối xứng sử dụng cùng một khoá cho cả hai quá trình mã hoá và giải mã. Mã hoá đối xứng có thể tác động trên bản rõ theo từng nhóm bit hay theo từng bit một.

1.4.3. Mã hoá bất đối xứng (Asymetric cryptography)

Mã hóa bất đối xứng được thiết kế sao cho khoá sử dụng trong quá trình mã hoá khác biệt với khoá được sử dụng trong quá trình giải mã. Tất nhiên không thể suy luận khóa giải mã từ khóa mã và ngược lại. Khoá để mã hoá được gọi là khóa công khai (Public Key), khoá để giải mã được gọi là khóa bí mật (Private Key).

1.4.4. Hệ thống mã hoá khoá lai (Hybrid Cryptosystems)

Hệ thống mã hoá khoá lai ra đời là sự kết hợp giữa tốc độ và tính an toàn của hai hệ thống mã hoá ở trên.

1.5. Ứng dụng của mã hóa thông tin

Mã hóa thông tin được ứng dụng trong rất nhiều lĩnh vực cả về phần cứng và phần mềm.

1.6. Giới thiệu một số giải thuật mã hóa tiên tiến.

1.6.1. Các hệ mã khối

Các hệ mã khối dựa trên cơ sở làm việc với các khối dữ liệu là các chuỗi bit có kích thước khau nhau (tối thiểu là 64bit), khóa của hệ mã hóa cũng là một xâu bit có độ dài cố định.

Một số giải thuật được sử dụng khá phổ biến là DES, Triple DES (3DES), AES.

1.6.1.1. Mã hóa dữ DES

DES (Data Encryption Standard) là thuật toán mã hóa với dữ liệu đầu vào và đầu ra là một khối 64 bit với độ dài khóa 64 bit (trong đó 8 được dùng để kiểm tra tính chẵn lẻ). Thuật toán thực hiện 16 vòng với 16 khóa (48bit) được sinh ra trong mỗi vòng.

Quá trình giải mã được diễn ra tương tự nhưng với các khóa con ứng dụng vào các vòng trong theo thứ tự ngược lại

Thuật toán DES bộc lộ một số điểm yếu mà những kẻ lợi dụng nó để thám mã.

1.6.1.2. Triple DES (3DES)

Triple DES thật chất là mã hóa theo DES ba lần với khóa K_1, K_2, K_3 cho mỗi lần.

1.6.1.3. Chuẩn mã hóa cao cấp AES

AES là một chuẩn mã hóa cao cấp với khóa bí mật cho phép xử lý các khối dữ liệu đầu vào có kích thước 128 bit và sử dụng các khóa có độ dài 128, 192, 256 bit.

1.6.2. Các hệ mã hóa công khai

1.6.2.1. Khái niệm

Mã hoá bằng khoá công khai là phương thức được thực hiện trên hai khóa, một được dùng để mã hóa (được gọi là khóa công khai – public key) và một khóa được dùng trong quá trình giải mã (gọi là khóa bí mật – private key). Khóa giải mã không thể tính toán được từ khóa mã hóa.

1.6.2.2. Thuật toán RSA

1.6.2.3. Hệ mật mã dựa trên các đường cong Elliptic

Thuật toán mã hóa sử dụng đường cong elliptic dựa vào độ phức tạp của bài toán logarit rời rạc trong hệ thống đại số. Một được cong Elliptic là một đường cong tạo ra bởi một phương trình dạng mẫu $y^2 = x^3 + Ax + B$

Thuật toán mã hóa dựa trên đường cong Elliptic (ECC) thực hiện việc mã hoá và giải mã dựa trên tọa độ của các điểm dựa trên đường cong Elliptic.

1.6.3. Hàm băm

Hàm băm là hàm toán học chuyển đổi một thông điệp có độ dài bất kỳ thành một dãy bit có độ dài cố định. Mọi thay đổi dù là rất nhỏ trên thông điệp đầu vào đều làm thay đổi giá trị băm của nó.

1.7. Kết chương

Chương 1 đã nêu ra được một số lý luận về mật mã học cũng như các ứng dụng của nó trong các lĩnh vực khác nhau cũng như giới thiệu cơ bản một số phương pháp mã hóa đã và đang được các nhà khoa học nghiên cứu

đồng thời trình bày nguyên tắc hoạt động cũng như nêu ra các ưu và nhược điểm của mỗi giải thuật.

CHƯƠNG 2. THUẬT TOÁN ADVANCED ENCRYPTTION STANDARD

Chuẩn mã hóa AES cho phép xử lý các khối dữ liệu đầu vào có kích thước 128 bit sử dụng các khóa có độ dài 128, 192 hoặc 256 bit với các tên gọi là AES-128, AES-192 hay AES-256. Chương này tập trung đặc tả chi tiết thuật toán.

2.1. Các thuật ngữ

2.2. Các ký hiệu và qui ước.

2.2.1. Các hàm, ký hiệu và tham số của thuật toán

2.2.1.1. Các ký hiệu phép toán:

Phép nghịch đảo bit (⊕-XOR), phép nhân đa thức (⊗), phép nhân trên trường hữu hạn (•).

2.2.1.2. Các tham số của thuật toán:

2.2.2. Các quy ước

2.2.2.1. Đầu vào, đầu ra

Đầu vào (Input) và đầu ra (Output) của thuật toán AES là các dãy nhị phân 128 bit được đánh số thứ tự từ 0, còn được gọi là các khối (block). Khóa mã dùng tho thuật toán AES là một dãy nhị phân có độ dài 128, 192 hay 256 bit.

2.2.2.2. Đơn vị Byte:

Byte là một dãy 8 bit được biểu diễn dưới dạng các bit nhị phân theo thứ tự $\{b_7, b_6, b_5, b_4, b_3, b_2, b_1, b_0\}$ hoặc biểu diễn trên trường hữu hạn bằng đa thức: $\sum_{i=0} b_i x_i$ hoặc bằng 2 ký tự trong hệ Hexa. Một số phép toán trên trường hữu hạn còn đòi hỏi thêm một bit (b_8) vào bên trái của một byte 8 bit, ký hiệu là {01}.

2.2.2.3. Mảng Byte

Các mảng byte được biểu diễn theo dạng sau: $a_0\ a_1\ a_2\ \ldots\ a_{n-1}$ (với n=16, 24 hay 32 tùy thuộc vào độ dài của dữ liệu đầu vào là 128 bit, 192 bit và 256 bit).

2.2.2.4. Mảng trạng thái (State)

Mảng trạng thái là một mảng hai chiều gồm 4 hàng, Nb cột, ký hiệu là s được dùng để lưu trữ giá trị trung gian trong mỗi bước của quá trình xử lý

Bắt đầu của phép mã hóa hay giải mã là việc sao chép mảng các byte $in_0, in_1, in_2, \ldots, in_{15}$ đầu vào vào mảng trạng thái s theo công thức sau:

$$s[r,c] = in[r+4c], \text{ với } 0 \le r, c \le 4$$

Vào cuối quá trình mã hóa hay giải mã, Mảng trạng thái sẽ được sao chép vào mảng byte đầu ra theo công thức: $out_0, out_1, out_2, \ldots, out_{15}$

2.2.2.5. Mảng các từ

Bốn Byte trên mỗi cột của Mảng trạng thái được tạo thành một từ 32 bit là một mảng gồm 4 byte được đánh chỉ mục theo hàng r. Từ đó ta có thể coi Mạng trạng thái là một mảng môt chiều gồm các từ w_0, w_1, w_2, w_3, trong đó giá trị cột c dùng làm chỉ mục cho mảng.

2.3. Cơ sở toán học

Hai phép cộng và phép nhân trên trường Galoris GF (2^8) là cơ sở toán học của thuật toán AES.

2.3.1. Phép cộng

Phép "cộng" ở đây được hiểu là phép XOR trên hai bit tương ứng trong byte và có ký hiệu là ⊕

2.3.2. Phép nhân

Phép nhân trên trường GF(2^8), ký hiệu là □, tương ứng với phép nhân thông thường của hai đa thức đem chia lấy dư (modulo) cho một đa thức tối giản bậc 8. Trong thuật toán AES, đa thức tối giản được chọn là:

$$m(x) = x^8 + x^4 + x^3 + x + 1 \qquad (2.2)$$

hay {01}{1b} nếu biểu diễn dưới dạng hexa.

Kết quả nhận được của phép rút gọn là một đa thức có bậc nhỏ hơn 8 nên có thể biểu diễn được dưới dạng 1 byte.

2.3.3. *Phép nhân với x*

Phép nhân với đa thức x (hay phần tử {00000010}∈GF(2^8)) có thể được thực hiện ở mức độ byte bằng một phép dịch trái và sau đó thực hiện tiếp phép XOR với giá trị {1b} nếu b_7=1. Thao tác được ký hiệu là *xtime()*. Phép nhân với các lũy thừa của x có thể được thực hiện bằng cách áp dụng nhiều lần thao tác *xtime()*. Kết quả phép nhân với một giá trị bất kỳ được xác định bằng phép cộng (⊕)các kết quả trung gian này lại với nhau.

2.3.4. *Đa thức với các hệ số trên trường GF(2^8)*

Phép nhân của hai đa thức bậc 4 với các hệ số trên GF(2^8) $a(x) \otimes b(x)$ được xác định bằng 4 hạng tử $d(x)$:

$$d(x) = d_3x^3 + d_2x^2 + d_1x + d_0$$

trong đó,

$$d_0 = (a_0 \bullet b_0) \oplus (a_3 \bullet b_1) \oplus (a_2 \bullet b_2) \oplus (a_1 \bullet b_3)$$

$$d_1 = (a_1 \bullet b_0) \oplus (a_0 \bullet b_1) \oplus (a_3 \bullet b_2) \oplus (a_2 \bullet b_3)$$

$$d_2 = (a_2 \bullet b_0) \oplus (a_1 \bullet b_1) \oplus (a_0 \bullet b_2) \oplus (a_3 \bullet b_3)$$

$$d_3 = (a_3 \bullet b_0) \oplus (a_2 \bullet b_1) \oplus (a_1 \bullet b_2) \oplus (a_0 \bullet b_3)$$

2.4. Đặc tả thuật toán

Thuật toán AES được thực hiện bởi tuần tự gồm nhiều bước biến đổi, kết quả đầu ra của phép biến đổi trước là đầu vào của phép biến đổi tiếp theo. Kết quả trung gian của các phép biến đổi chính là mảng trạng thái (state) như đã trình bày ở phần 2.2.2.4.

Độ dài của khối dữ liệu đầu vào của AES là cố định với Nb=4. Tùy vào độ dài khóa (Nk=4, 6, 8) ban đầu ta có số lần lặp (Nr) cho mỗi quá trình được xác định theo công thức Nr=max{Nb, Nk}+6.

2.4.1. *Qui trình mã hóa*

Bắt đầu quá trình mã hóa, bản rõ được sao chép vào mảng trạng thái theo phương pháp được mô tả ở phần 2.2.2.4. Sau khi thực hiện thao tác cộng với khóa mã đầu tiên, mảng trạng thái sẽ được biến đổi qua *Nr* vòng trong đó lần cuối cùng được thực hiện khác với *Nr-1* vòng trước đó. Nội dung của mảng trạng thái ở vòng cuối cùng sẽ là bãn mã của quá trình mã hóa.

Trong quy trình mã hóa của AES, tất cả các vòng lặp đều sử dụng 4 hàm (theo thứ tự*Subbytes(), ShiftRows(), MixColumns(), AddRoundKey().* Riêng vòng cuối cùng bỏ qua việc gọi hàm *MixColumns().*

2.4.1.1. *Phép biến đổi SubBytes()*

Phép biến đổi SubByte làm biến Mảng trạng thái hiện hành bằng cách sử dụng một bảng thay thế (S-Box) như sau:

1. Nhân nghịch đảo trên trường GF(2^8). Quy ước $\{00\}^{-1} = \{00\}$.
2. Áp dụng phép biến đổi Affine (trên GF(2)) sau:

$$b_i^{'} = b_i \oplus b_{(i+4) \bmod 8} \oplus b_{(i+5) \bmod 8} \oplus b_{(i+6) \bmod 8} \oplus b_{(i+7) \bmod 8} \oplus c_i$$

trong đó, 0 ≤ i ≤8 là bit thứ *i* bủa byte *b* tương ứng và c_i là bit thứ thứ *i* của byte *c* với giá trị {63} hay {01100011}.

2.4.1.2. *Phép biến đổi ShiftRows()*

Làm biến đổi các byte trên ba hàng cuối cùng mảng Trạng thái bằng cách dịch vòng thể hiện qua công thức:

$$s'_{r,c} = s_{r,(c+shift(r,Nb) \bmod Nb}, \text{ với } 0 < r \le 3 \text{ và } 0 \le c < Nb \qquad (2.13)$$

Số lần dịch vòng shift(r,Nb) phụ thuộc vào chỉ số vòng. Cụ thể shift(1,4)=1, shift(2,4)=1, shift(3,4)=3.

2.4.1.3. *Phép biến đổi MixColumns()*

MixColumns() là một phép biến đổi mã hóa được thực hiện bằng cách lấy tất cả các cột của Mảng trạng thái trộn với dữ liệu của chúng một cách độc lập nhau để tạo ra các cột mới.

2.4.1.4. Phép biến đổi AddRoundKey()

Phép AddRoundKey() nhằm bổ sung khóa riêng biệt cho mỗi vòng lặp trong quá trình mã hóa hay giải mã của AES. Các khóa này được sinh ra từ thủ tục sinh khóa. Hàm *AddRoundKey()* thực hiện bằng cách cộng một khóa vòng tại vòng đang xét với Mảng trạng thái thông qua phép toán XOR đơn giản trên bit.

2.4.2. Thuật toán sinh khóa

Tạo ra một bảng khóa từ khóa ban đầu để bổ sung khóa cho các vòng trong quá trình mã hóa hay giải mã của AES và được thực hiện trước tiên. Thủ tục này sẽ tạo ra tổng cộng *Nr+1* khóa được lưu trữ trong mảng một chiều, ký hiệu là **w**, đánh chỉ mục từ *0* đến *Nb*(Nr+1)-1*.

2.4.3. Qui trình giải mã

Quá trình giải mã được thực hiện theo chiều ngược lại với quy trình mã hóa, đồng thời các phép biến đổi trong quá trình này cũng được thực hiện đảo ngược. Ngoại trừ phép biến đổi AddRoundKey() không thay đổi vì chính bản thân nó là một phép biến đổi thuận ngịch do chỉ áp dụng một phép toán XOR.

2.4.3.1. Phép biến đổi InvShiftRows()

InvShiftRows() chính là phép biến đổi ngược của ShiftRows().

2.4.3.2. Phép biến đổi InvSubBytes()

Là phép biến đổi ngược của *SubBytes()* được thực hiện trên bảng thay thế S-Box là nghịch đảo của S-Box.

2.4.3.3. Phép biến đổi InvMixColumns()

InvMixColumns() là phép biến đổi ngược của *MixColumns()*.

2.4.3.4. Thuật toán giải mã tương đương

2.5. Các ưu điểm và hạn chế của giải thuật

Ưu điểm của thuật toán AES là cho phép thực thi hiệu quả bằng cả phần mềm và phần cứng mà không cần nhiều bộ nhớ và có thiết kế đơn giản, linh hoạt nhưng vẫn đảm bảo độ an toàn của thông tin.

Nhược điểm là không thích hợp trên các thiết bị thẻ thông minh vì cần nhiều mã và số vòng xử lý; không kế thừa lại mã lệnh mã hóa trong giải mã hoặc rất ít trong cả thiết kế phần cứng và phần mềm.

2.6. Kết chương

Ở chương 2, tác giả tập trung nghiên cứu và trình bày một cách chi tiết đặc tả giải thuật AES đồng thời nêu ra các ưu điểm và nhược điểm của nó.

CHƯƠNG 3. ỨNG DỤNG THUẬT TOÁN ADVANCED ENCRYPTION STANDARD MÃ HÓA VĂN BẢN MẬT TẠI TRƯỜNG CAO ĐẲNG CÔNG KỸ NGHỆ ĐÔNG Á

3.1. Đặt vấn đề

Các văn bản được hình thành tại các quan, tổ chức, đơn vị có thể chứa đựng bên trong nó các thông tin thuộc về bí mật Nhà nước mà mọi công dân có nhiệm vụ bảo vệ nhằm góp phần xây dựng và bảo vệ Tổ quốc. Việc lưu trữ, bảo vệ bí mật nhà nước được quy định tại Chương 3 của Pháp lệnh số 30/2000/PL-UBTVQH10.

Trường Cao đẳng Công Kỹ Nghệ Đông Á là một tổ chức giáo dục hoạt động trong lĩnh vực giáo dục và đào tạo nên các văn bản hình thành trong quá trình hoạt động thuộc danh mục được quy định tại điều 1 của quyết định số 160/2005/QĐ-BCA(A11) của Bộ Công An về danh mục bí mật nhà nước độ Mật trong ngành giáo dục và đào tạo phải được giữ bí mật cho đến khi được phép công bố.

3.2. Thực trạng quản lý văn bản tại trường Cao đẳng công kỹ nghệ Đông Á

Những văn bản quan trọng được hình thành trong quá trình hoạt động của trường Cao đẳng Công Kỹ Nghệ Đông Á chưa được chú trọng. Cụ thể:

- Đề thi chưa được bảo mật tốt
- Các văn bản lưu tại máy cá nhân không có cơ chế bảo mật.
- Nhà trường chưa lập được danh mục văn bản mật
- Quản lý văn bản chưa tuân theo một quy trình cụ thể

Các giải pháp và các quy trình tiếp nhận, xử lý và quản lý văn bản dựa trên các mô tả được trình bày trong phần tiếp theo.

3.3. Mô tả công tác tổ chức lưu trữ và quản lý văn bản tại trường Cao đẳng Công Kỹ Nghệ Đông Á

3.3.1. Đặc tả hệ thống

Ở đây, xây dựng một máy chủ để lưu trữ tập trung tất cả các văn bản của người dùng theo từng danh mục. Với mỗi tập tin văn bản được đánh nhãn theo các từ khóa khác nhau phục vụ cho việc tìm kiếm sau này, đồng thời cấp quyền để những người được phép mới có thể xem hoặc tải về. Thông tin của một tập tin văn bản và người dùng được lưu trữ tại một CSDL. Người dùng tương tác qua ứng dụng khách.

3.3.2. Xác định các yêu cầu của hệ thống

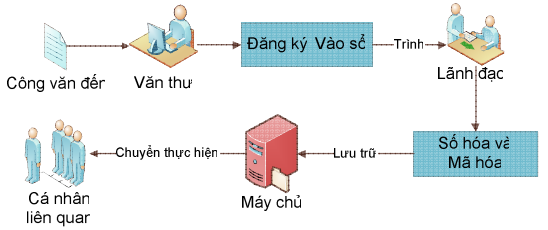
Hệ thống sau khi xây dựng phải đáp ứng được các yêu cầu:

- Cho phép lưu trữ, quản lý toàn bộ các tập tin văn bản do người dùng tải lên.
- Theo dõi và cập nhật thông tin của các văn bản cũng như "lịch sử" sử phát hành của nó.
- Cho phép người dùng tra cứu các văn bản theo các tiêu chí.
- Đảm bảo chỉ những người được phép mới có quyền truy cập đến văn bản chỉ định.
- Đảm bảo các văn bản quan trọng phải được mã hóa.
- Phải thường xuyên sao lưu dự phòng (Back-up).
- Cho phép thống kê, theo dõi tình hình thực hiện văn bản

3.3.3. Các quy trình xử lý văn bản

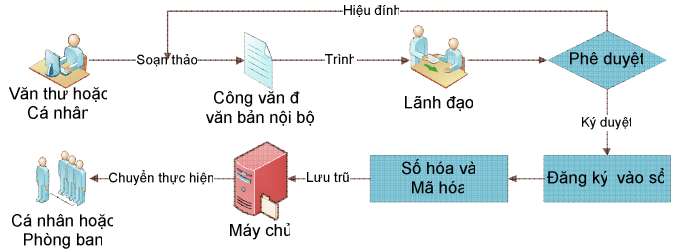
3.3.3.1. Quy trình xử lý văn bản hành chính

Đối với công văn đến:



Hình 3.1. Sơ đồ quy trình xử lý công văn đến

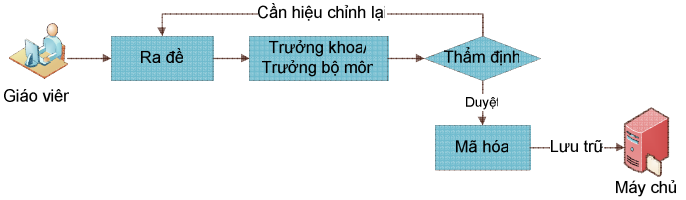
Đối với các công văn đi:



Hình 3.2. Sơ đồ quy trình xử lý văn bản nội bộ

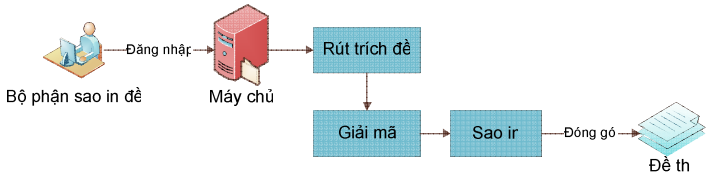
3.3.3.2. Quy trình ra đề và xử lý đề thi

Đối với giáo viên:



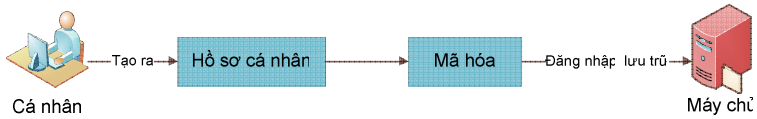
Hình 3.3. Sơ đồ quy trình ra đề thi

Đối với bộ phận sao in đề:



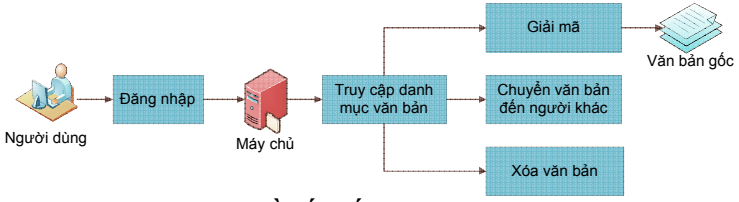
Hình 3.4. Sơ đồ quy trình rút trích, sao in đề thi

3.3.3.3. Quản lý các văn bản cá nhân



Hình 3.5. Sơ đồ quy trình đưa tài liệu cá nhân

3.3.3.4. Quá trình nhận và xử lý văn bản



Hình 3.6. Sơ đồ tiếp tiếp nhận và xử lý văn bản

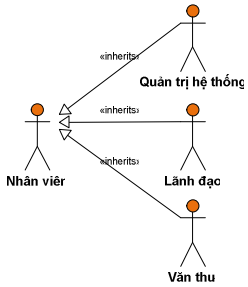
3.4. Phân tích và thiết kế hệ thống

3.4.1. Các tác nhân (actor)

Các tác nhân của hệ thống gồm: Quản trị hệ thống, nhân viên, cán bộ văn thư, lãnh đạo.

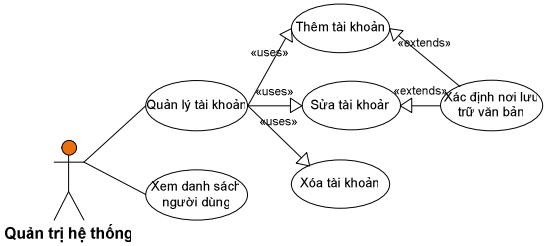
3.4.2. Sơ đồ ca sử dụng (use case)

a) Sơ đồ ca sử dụng tổng quát



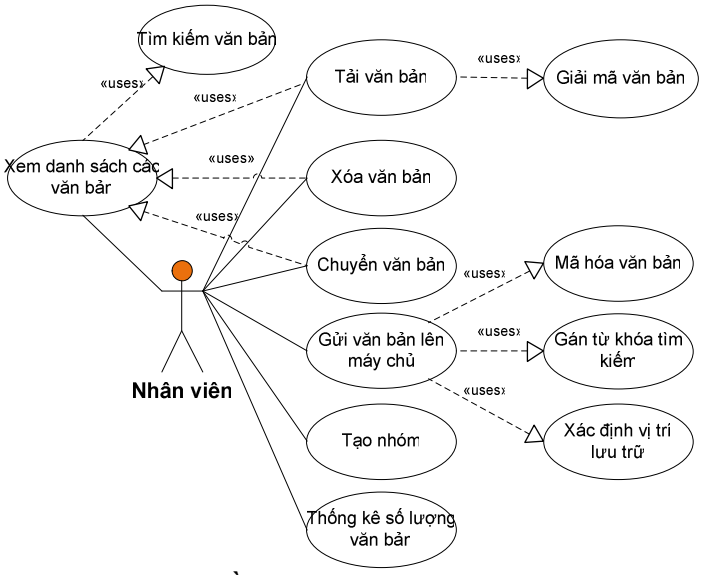
Hình 3.7. Sơ đồ ca sử dụng tổng quát

b) Sơ đồ ca sử dụng của tác nhân "quản trị hệ thống"



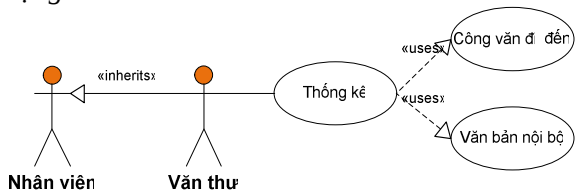
Hình 3.8. Sơ đồ ca sử dụng của tác nhân Quản trị hệ thống

c) Sơ đồ ca sử dụng của tác nhân "cán bộ"



Hình 3.9. Sơ đồ ca sử dụng của tác nhân Nhân viên

d) Sơ đồ ca sử dụng của tác nhân "Văn thư"



Hình 3.10. Sơ đồ ca sử dụng của tác nhân Văn thư

3.4.3. Đặc tả ca sử dụng

a) Ca sử dụng "Gửi văn bản lên máy chủ"

b) Ca sử dụng "Mã hóa văn bản"

c) Ca sử dụng "Giải mã văn bản"

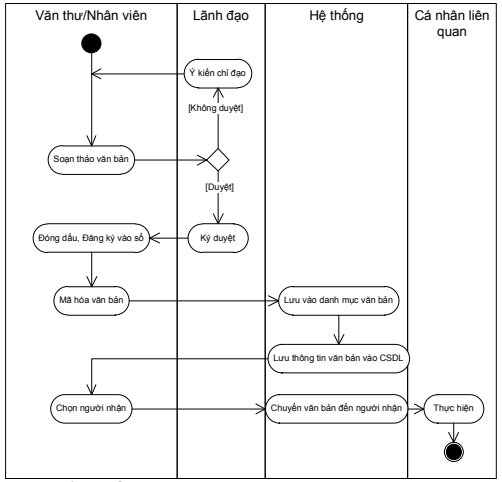
d) Ca sử dụng "Chuyển văn bản"

e) Ca sử dụng "Tải văn bản"

3.4.4. Biểu đồ lớp

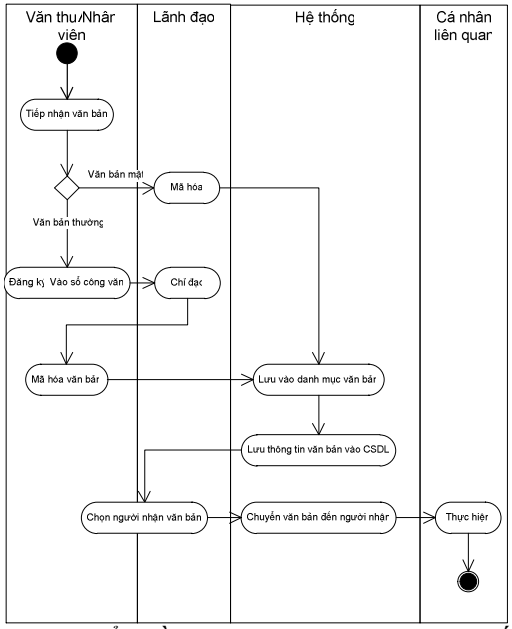
3.4.5. Biểu đồ hoạt động

a) Xử lý văn bản hành chính nội bộ



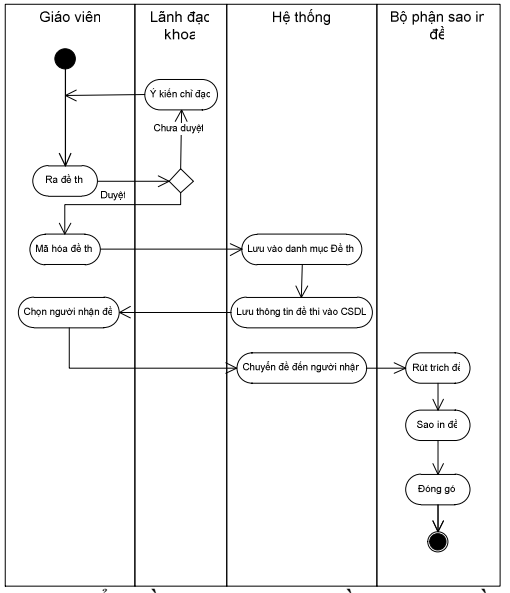
Hình 3.11. Biểu đồ hoạt động xử lý văn bản hành chính

b) Xử lý công văn đến



Hình 3.12. Biểu đồ hoạt động xử lý công văn đến

c) Hoạt động ra đề thi và xử lý đề thi

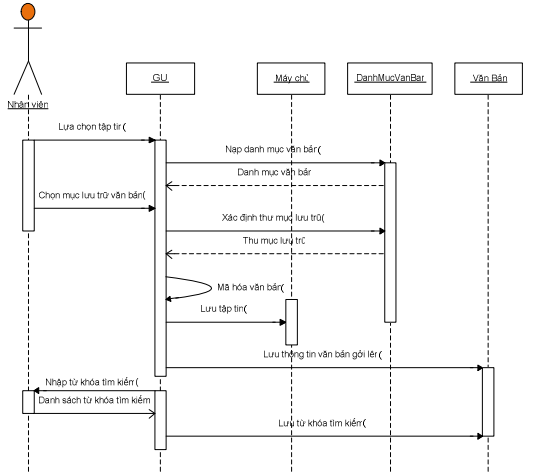


Hình 3.13. Biểu đồ hoạt động ra đề và xử lý đề thi

d) Hoạt động tiếp nhận và xử lý văn bản

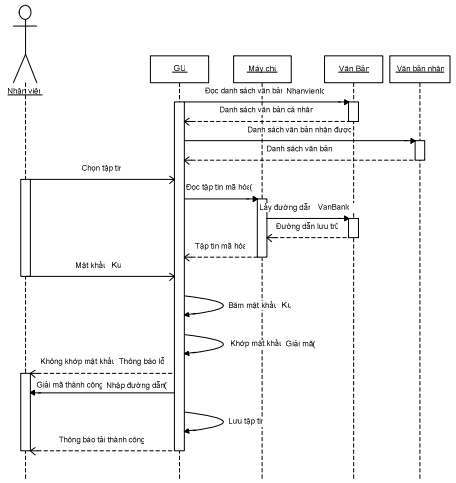
3.4.6. Biểu đồ tuần tự

a) Nhân viên gửi văn bản lên máy chủ



Hình 3.14. Biểu đồ tuần tự lưu trữ văn bản

b) Tải văn bản về



Hình 3.15. Biểu đồ tuần tự tải văn bản

c) Chuyển văn bản đến người dùng khác

3.5. Vấn đề mã hóa

3.5.1. Mã hóa và giải mã văn bản

Cả hai quá trình mã hóa và giải mã đều thực hiện dựa trên hai khóa bí mật. Một khóa được dùng để mã hóa nội dung tập tin gọi là khóa mã hóa, ký hiệu là K_e, được sinh ra từ bộ sinh khóa ngẫu nhiên của chương trình. Một khóa được dùng để mã hóa khóa K_e ở trên gọi là khóa người dùng, ký hiệu K_u, do người dùng nhập từ bàn phím.

Kết quả mã hóa của tập tin văn bản rõ theo khóa K_e là C_{Ke}. Kết quả mã hóa của khóa K_e theo khóa người dùng K_u là C_{Ku}. Và kết quả băm của khóa K_u là B_{Ku}. Tập hợp các dữ liệu B_{Ku}, C_{Ku}, C_{Ke} sẽ là kết quả của tập tin bản mã.

Quá trình giải mã được thực hiện ngược lại. Với khóa K_u, ta sẽ tiến hành so khớp giá trị băm của nó với B_{Ku} được đọc từ tập tin bản mã. Nếu phép so khớp thỏa mãn thì giải mã 32 bit tiếp theo (tương đương với giá trị C_{Ku}) bằng thuật toán AES với khóa giải mã K_u để nhận được khóa K_e. Cuối

cùng tiến hành giải mã phần nội dung còn lại bằng khóa K_e để thu được văn bản rõ ban đầu.

3.5.2. Tổ chức tập tin mã hóa

Với những phân tích ở trên, tập tin sau khi mã hóa được tổ chức làm hai phần cơ bản: phần Header chứa các thông tin về tập tin gốc và tập tin mã hóa; phần Content chứa các byte mã hóa của tập tin gốc ban đầu. Kết cấu của tập tin mã hóa được mô tả như ở bảng Bảng 3.1.

Bảng 3.1. Cấu tạo tập tin mã hóa

Thành phần	Độ dài	Nội dung
Tiêu đề	6 byte	Dấu hiệu nhận biết tập tin mã hóa bởi hệ thống
	128 byte	Tên tập tin bản rõ
	38 byte	Thời gian khởi tạo và chỉnh sửa lần cuối của tập tin bản rõ
	64 byte	Giá trị băm của khóa K_u
	32 byte	Giá trị khóa mã hóa K_e đã mã hóa theo khóa K_u
	4 byte	Giá trị số byte của khối cuối cùng
Nội dung mã hóa	Nx16 byte	Chứa N khối byte dữ liệu mã hóa tương ứng với N khối byte dữ liệu bản rõ ban đầu. Mỗi khối gồm 16 byte dữ liệu.

3.6. Cài đặt

3.6.1. Thiết kế hệ thống

Xây dựng một máy chủ FTP để truyền file thông qua môi trường mạng và cấp phép cho những người dùng được quyền truy cập đến những thư mục lưu trữ tương ứng. Việc đăng tập tin lên hay tải tập tin từ hệ thống về đều được thực hiện qua chương trình ứng dụng thiết kế riêng. Người dùng đăng nhập thông qua chương trình để được phép truy cập đến hệ thống. Việc chuyển và nhận văn bản giữa các thành viên được ghi nhận trên CSDL của hệ thống.

3.6.2. Công cụ

Hệ điều hành Windows Server 2003 với công cụ IIS; MS SQL Server 2008; Microsoft Visual C# trên nền dotNetFramework 4.0.

3.6.3. Cơ sở dữ liệu

3.6.4. Xây dựng thư viện mã hóa AESCrypto

Giải thuật AES được cài đặt trong một thư viện có tên AESCrypto gồm hai lớp: lớp AESImpClass khởi tạo các giá trị ban đầu cho giải thuật; lớp AESMainClass chứa các hàm mã hóa, giải mã cũng như các hàm khác như đã trình bày trong phần 2.4.

3.7. Thử nghiệm và đánh giá

Kết quả thử nghiệm cho thấy tốc độ xử lý trên các loại tập tin khác nhau đều cho cùng một kết quả do trong các xử lý mã hóa và giải mã được thực hiện trên từng byte của tập tin gốc.

Tất cả các tập tin sau khi mã hóa có độ lớn tăng thêm 448 byte so với tập tin gốc.

Do độ dài của các khóa khác nhau trong giải thuật nên số vòng thực hiện trong mỗi quá trình sẽ khác nhau. Điều này dẫn tới với chiều dài khóa của nó càng lớn thì tốc độ xử lý chậm hơn.

Hệ thống quản lý văn bản này qua thử nghiệm đã thể hiện một số ưu điểm là khả năng bảo mật thông tin rất cao; thực hiện tốt chức năng xuất bản tài liệu lên máy chủ, tải và giải mã văn bản, chuyển tiếp văn bản giữa các người dùng và đồng thời cho phép các cá nhân tìm kiếm các văn bản mong muốn theo chủ đề.

3.8. Kết chương

Ở chương 3 này thể hiện một số kết quả đạt được sau khi cài đặt thử nghiệm thành công giải thuật AES trên môi trường Windows. Qua đó, so sánh tốc độ mã hóa và giải mã của giải thuật với các độ dài khác nhau trên các hệ thống khác nhau. Đồng thời áp dụng để xây dựng thành công hệ thống quản lý văn bản chung cho toàn trường với cơ chế bảo mật tốt.

KẾT LUẬN

Qua quá trình nghiên cứu, đề tài đã đạt được một số kết quả nhất định. Bên cạnh đó cũng tồn tại những hạn chế ở một số mặt nào đó. Phần này sẽ đánh giá lại những kết quả trên đồng thời phân tích các khả năng ứng dụng của đề tài để từ đó có hướng phát triển cao hơn.

Về kết quả đạt được

Nội dung chính của đề tài làm rõ cách tổ chức, hoạt động của giải thuật AES cũng như khả năng ứng dụng của nó.

Kết quả nổi bật của đề tài là xây dựng được một hệ thống quản lý văn bản áp dụng tại trường Cao đẳng Công Kỹ Nghệ Đông Á dựa vào việc ứng dụng chuẩn mã hóa AES nói trên vào việc mã hóa tập tin nhằm đảm bảo tính an toàn, bí mật của thông tin mà nó chứa đựng. Qua đó, người dùng hoàn toàn yên tâm với những văn bản cá nhân được lưu trữ tại hệ thống vì tính bảo mật và độ an toàn cao. Ngoài ra hệ thống cũng thể hiện được một số chức năng ưu việt khác như cho phép người dùng chuyển văn bản đến những người dùng khác trong cùng hệ thống, và giải mã những văn bản được phép về máy cá nhân hay tìm kiếm các văn bản như ý muốn theo từ khóa.

Về ưu điểm và nhược điểm của đề tài

Như đã trình bày, đề tài thể hiện được ba ưu điểm chính:

Thứ nhất, đề tài đã làm rõ hoạt động của chuẩn mã hóa AES và ứng dụng thành công vào việc mã hóa văn bản trên hệ thống quản lý văn bản được xây dựng và áp dụng tại trường Cao đẳng Công Kỹ Nghệ Đông Á đảm bảo độ an toàn, tính bí mật của thông tin.

Thứ hai, hệ thống cũng đáp ứng được các chức năng quan trọng của một hệ thống quản lý văn bản.

Thứ ba, việc áp dụng hệ thống quản lý văn bản này vào công tác quản lý văn bản tại trường sẽ giúp tiết kiệm về không gian lưu trữ, tiết kiệm về thời gian và chi phí chuyển giao văn bản.

Ngoài những ưu điểm kể trên, do phạm vi nghiên cứu chỉ tập trung vào một số chức năng chính nên đề tài vẫn còn một số hạn chế như sau:

- Về vấn đề quản lý khóa mã hóa: Khó khăn nhất của vấn đề là làm sao vận chuyển khóa đến những người được nhận văn bản. Đây là một hạn chế của đề tài.
- Về chức năng của hệ thống: Dù hệ thống đã được phân tích kỹ lưỡng và được xây dựng thành công nhưng hệ thống chỉ mới đáp ứng một số chức năng cơ bản trong công tác quản lý văn bản.
- Về cơ chế mã hóa: Do chỉ tập trung vào việc cài đặt giải thuật và mã hóa tuần tự các khối dữ liệu của tập tin đầu vào và đánh giá khả năng thực thi của giải thuật theo nguyên bản của nó nên thời gian xử lý tập tin văn bản còn chậm.

Về khả năng ứng dụng của đề tài

Với những ưu điểm có được của hệ thống, nó có khả năng ứng dụng vào thực tiễn rất cao. Hầu hết trong các cơ quan, doanh nghiệp, tổ chức trong quá trình hoạt ứng dụng công nghệ thông tin vào trong công tác hành chính, lưu trữ văn bản. Trong khối lượng lớn văn bản được hình thành trong quá trình hoạt động chắc chắn rằng sẽ có những văn bản mang thông tin quan trọng cần được bảo mật. Tuy nhiên vấn đề đảm bảo an toàn cho những thông tin lưu trữ đó chưa được quan tâm hoặc có quan tâm nhưng phương pháp bảo vệ còn sơ sài, có khả năng bị tấn công và nguy cơ lộ bí mật là rất cao. Do đó việc áp dụng một hệ thống như trên là cần thiết nhưng những nhược điểm kể trên cần phải khắc phục để tối ưu hệ thống.

Hướng phát triển

Để hệ thống thực sự hiệu quả nếu được áp dụng thì ngoài việc khắc phục những nhược điểm được nêu trên nênphát triển thêm một số vấn đề:

- Nghiên cứu mở rộng giải thuật AES để có thể xử lý với khối dữ liệu đầu vào lớn hơn như 512 bit hay 1024 bit.
- Áp dụng cơ chế đa luồng trên các bộ xử lý hiện nay để tăng hiệu quả việc xử lý tập tin.
- Tiếp tục nghiên cứu vấn đề quản lý và vận chuyển khóa mã hóa.