Nhóm 3 :Tú anh,Tuấn Đạt,Đức Anh

Báo cáo

**Chương 1: Tổng quan về phần mêm lưu trữ và bảo mật dữ liệu**

***Khái niệm:***

* **Bảo mật dữ liệu** là công việc bảo vệ thông tin số sao cho không bị truy cập trái phép, bị hỏng hay đánh cắp. Khái niệm này bao trùm mọi phương thức bảo mật thông tin, từ bảo mật phần cứng, kiểm soát quyền truy cập, bảo mật logic cho phần mềm hay thậm chí là cả chính sách và quy trình của doanh nghiệp.
* **Lưu trữ dữ liệu** (trong tiếng anh là **Data Storage**) là việc lưu trữ (ghi) thông tin (hay các dữ liệu) trong phương tiện lưu trữ nào đó. Hình thức lưu trữ điện tử yêu cầu cần có năng lượng điện để lưu trữ, truy xuất dữ liệu. Chúng ta có thể lưu trữ các thông tin bằng nhiều hình thức khác nhau bằng nhiều phương tiện khác nhau.

***Tầm quan trọng của bảo mật dữ liệu trong thời đại số***

* Có thể nói, các máy tính hiện nay đã phủ sóng toàn cầu. Do đó mà mạng lưới máy tính cũng tác động đến nhiều lĩnh vực của đời sống xã hội. Dù là dữ liệu cá nhân hay dữ liệu doanh nghiệp thì việc bảo mật để tránh bị rò rỉ, thất thoát dữ liệu vẫn là yếu tố cần được ưu tiên hàng đầu.
* Bảo mật dữ liệu trong thời đại kỹ thuật số hiện nay đã không còn là một khái niệm quá xa lạ. Theo đó, bảo mật dữ liệu là việc tiến hành áp dụng các biện pháp bảo vệ nhằm ngăn ngừa việc rò rỉ, mất mát thông tin cá nhân hoặc dữ liệu doanh nghiệp. Điều này đồng nghĩa với việc giúp các cá nhân, tổ chức có thể xác định xem dữ liệu của mình có bị chia sẻ hoặc đánh cắp bởi bên thứ ba nhằm chuộc lợi hay không.
* Hiện nay việc kết nối để truy cập, sử dụng dữ liệu đã trở nên vô cùng dễ dàng. Đối với những đối tượng xấu, việc đánh cắp dữ liệu không còn là thử thách quá khó khăn. Chính vì vậy mà người dùng cần thật cẩn thận và tỉnh táo áp dụng các giải pháp ngăn ngừa để tránh bị đánh cắp dữ liệu.

***Những loại dữ liệu bắt buộc phải sao lưu***

* Dữ liệu cá nhân: Đầu tiên và quan trọng nhất phải thực hiện sao lưu chính là dữ liệu cá nhân. Mọi dữ liệu cá nhân từ hình ảnh, video về bản thân, gia đình, bạn bè cho đến tài liệu quan trọng khác trên máy tính cần phải được sao lưu thường xuyên và bảo vệ cẩn thận. Bởi bột khi những dữ liệu này không may bị mất đi thì không có cách nào để khôi phục lại.
* Dữ liệu công việc: Những tài liệu quan trọng liên quan đến công việc, hợp đồng, biên bản, chứng từ… đều phải được sao lưu kỹ càng, để có thể xem lại khi cần thiết hoặc làm bằng chứng nếu không may có những tranh chấp về sau.
* Hệ điều hành và các chương trình hệ thống: Bên cạnh dữ liệu cá nhân và công việc, nếu cẩn thận hơn bạn cũng có thể sao lưu dữ liệu về hệ điều hành và các chương trình cài đặt trên máy. Nếu bạn là người thích tìm tòi về các tệp hệ thống hay thường xuyên cập nhật phần cứng, thì việc có một bản sao lưu toàn bộ dữ liệu hệ thống sẽ giúp bạn tiết kiệm thời gian hơn khi gặp sự cố.

***Phương pháp sao lưu dữ liệu đơn giản và an toàn:***

* **Sao lưu vào ổ cứng rời bên ngoài**

Đây là phương pháp sao lưu bảo vệ an toàn dữ liệu truyền thống. Bạn có thể sao lưu mọi dữ liệu từ máy tính sang chiếc ổ cứng rời bên ngoài bằng cách sử dụng các tính năng sao lưu tích hợp sẵn trong máy.

**+** Ưu điểm: Giá thành rẻ và tốc độ sao lưu nhanh

**+** Nhược điểm: Nếu máy tính và ổ cứng bị hư hại thì khả năng cao là sẽ mất toàn bộ dữ liệu.

* **Lưu trữ đám mây**
* Google Drive là dịch vụ lưu trữ dữ liệu trực tuyến, cho phép người dùng dễ dàng truy cập, sao lưu và chia sẻ thông tin lưu trữ giữa những người dùng với nhau, trên đa nền tảng. Bên cạnh đó, vì được phát triển bởi Google nên người dùng có thể quản lý dễ dàng dữ liệu cá nhân giữa ứng dụng Drive với dịch vụ khác do Google cung cấp.

+ Ưu điểm:

Miễn phí 15Gb lưu trữ ban đầu

Lưu trữ tệp an toàn và truy cập từ mọi nơi.

Dễ dàng chia sẻ tệp và thư mục.

Xem lại lịch sử, chỉnh sửa tệp nhanh, chính xác.

Quản lí tệp trực quan, dễ sử dụng.

* Dropbox: Giao diện đơn giản trực quan, sắp xếp thuận tiện cho người dùng, Dropbox là 1 trong số các ứng dụng cung cấp dịch vụ lưu trữ đám mây được yêu thích nhất hiện nay. Dropbox hỗ trợ không gian lưu trữ lớn, giúp người dùng sao lưu thông tin nếu được cài đặt, hay truy cập tài liệu kể cả khi không có Internet...

+Ưu điểm

Miễn phí 2GB lưu trữ ban đầu

Tự động tải dữ liệu từ thư viện ảnh của bạn lên lưu trữ ảnh đám mây.

Dễ dàng gửi các tệp lớn chỉ bằng cách sao chép và dán liên kết.

Truy cập bất kỳ tệp nào trong tài khoản không cần Internet.

Đồng bộ hóa các thư mục trên máy tính với Dropbox.

* OneDrive: Được phát triển bởi Microsoft, cài sẵn trên các thiết bị máy tính sử dụng hệ điều hành Windows.

+ Ưu điểm

Mở và lưu tệp OneDrive nhanh chóng.

Tìm kiếm ảnh dễ dàng nhờ tính năng gắn thẻ tự động.

Nhận thông báo khi tài liệu chung được chỉnh sửa.

Chia sẻ album ảnh và video yêu thích của bạn.

Truy nhập ngoại tuyến vào những tệp quan trọng nhất của bạn.

Miễn phí 5GB.

* iCloud: iCloud được công ty Apple phát triển riêng cho người dùng và chỉ sử dụng các thiết bị hệ điều hành iOS hay macOS. Vì vậy, về khả năng truy cập và chia sẻ dữ liệu giữa nhiều người dùng sẽ có nhiều hạn chế hơn.

+ Ưu điểm:

Bảo mật dữ liệu cao, an toàn.

Sử dụng tốt trong hệ sinh thái của Apple

***Lưu ý khi sao lưu dữ liệu***

Vậy phương pháp sao lưu nào là hiệu quả nhất?

* Câu trả lời là không có một phương pháp sao lưu dữ liệu nào hoàn hảo 100%. Cách tốt nhất để bảo vệ dữ liệu của bạn chính là kết hợp nhiều phương pháp và kết hợp giữa bản sao lưu Online (đám mây) và Offline (Ổ cứng máy tính).
* Bên cạnh đó, tránh các phương thức sao lưu thủ công (tự thực hiện quá trình sao lưu) và nên ưu tiên các phương pháp sao lưu tự động. Bởi việc sao lưu thủ công vừa tốn thời gian, vừa không đều đặn.

Việc sao lưu dữ liệu thường xuyên và nhiều bảo sao là điều vô cùng cần thiết giúp bạn giảm thiểu tối đa thiệt hại mất dữ liệu khi gặp sự cố không mong muốn. Hãy bảo vệ dữ liệu bằng nhiều cách để giảm thiểu tác động dữ liệu nếu có tình huống xấu nhất xảy ra.

**Chương 2 Nội dung thực hiện**

Trong lĩnh vực an ninh thông tin, thuật toán được sử dụng để bảo vệ dữ liệu và thông tin quan trọng, điều đó là rất cần thiết với xã hội hiện đại phát triển như bây giờ và để tránh khỏi các lỗ hổng bảo mật nhằm phòng ngừa các hacker tin tặc có thể đánh cắp các thông tin quan trọng của chúng ta thì sau đây là một vài phương án có thể thực hiện cùng với ưu nhược điểm của nó.

***Thứ nhất là thuật toán RSA***



**Khái niệm:**

* RSA là một thuật toán mã hóa công khai (asymmetric encryption) được đặt tên theo ba nhà khoa học Ron Rivest, Adi Shamir và Leonard Adleman, người đã phát triển nó vào năm 1977. RSA là một trong những thuật toán mã hóa công khai phổ biến nhất và được sử dụng rộng rãi trong lĩnh vực an ninh thông tin.
* Thuật toán RSA dựa trên việc sử dụng cặp khóa gồm khóa công khai (public key) và khóa bí mật (private key). Quá trình mã hóa bằng RSA sử dụng khóa công khai, trong khi quá trình giải mã sử dụng khóa bí mật. Khóa công khai được chia sẻ với mọi người, trong khi khóa bí mật chỉ được giữ bí mật bởi chủ sở hữu. Điều này tạo ra tính chất "không thể đoán được" của quá trình mã hóa và giải mã, đảm bảo tính an toàn và bảo mật của thông tin.

**Các bước chính của thuật toán RSA bao gồm:**

Tạo khóa:

a. Chọn hai số nguyên tố lớn ngẫu nhiên p và q.

b. Tính n = p \* q và ϕ(n) = (p-1) \* (q-1), trong đó ϕ(n) là hàm Euler.

c. Chọn số nguyên e sao cho 1 < e < ϕ(n) và e không có ước chung với ϕ(n).

d. Tính d sao cho (d \* e) % ϕ(n) = 1.

e. Khóa công khai là cặp (e, n) và khóa bí mật là cặp (d, n).

Mã hóa:

a. Chuyển đổi thông điệp cần mã hóa thành một số nguyên m.

b. Tính c = m^e mod n, trong đó ^ là phép mũ và mod là phép chia lấy dư.

c. Số c chính là thông điệp đã được mã hóa.

Giải mã:

a. Tính m = c^d mod n, trong đó c là thông điệp mã hóa.

b. Số m chính là thông điệp đã được giải mã.

\*Đặc điểm chính của thuật toán RSA bao gồm tính an toàn cao, khả năng mã hóa và giải mã nhanh cho thông điệp ngắn, nhưng hiệu suất giảm khi mã hóa/giải mã thông điệp dài. Thuật toán RSA cũng được sử dụng rộng rãi trong các ứng dụng khác nhau như chứng thực và xác thực người dùng, tạo chữ ký số và bảo vệ tính toàn vẹn của dữ liệu.

**Ưu và nhược điểm của thuật toán RSA**

**Ưu điểm**

* Bảo mật cao: RSA được coi là một thuật toán mã hóa mạnh và được sử dụng rộng rãi trong lĩnh vực an ninh thông tin. Việc sử dụng khóa công khai và khóa bí mật tạo ra tính chất "không thể đoán được" trong quá trình mã hóa và giải mã, đảm bảo tính bảo mật cao
* Mã hóa công khai: RSA sử dụng mô hình mã hóa công khai, trong đó khóa công khai được chia sẻ với mọi người và khóa bí mật chỉ được giữ bí mật bởi chủ sở hữu. Điều này giúp đơn giản hóa việc trao đổi khóa và bảo mật thông tin trong môi trường công cộng.
* Chữ ký số: RSA cũng được sử dụng để tạo chữ ký số, cho phép xác thực tính toàn vẹn và nguồn gốc của thông tin. Chữ ký số được tạo bằng cách sử dụng khóa bí mật để ký điện tử và khóa công khai để xác minh chữ ký.
* Sử dụng rộng rãi: RSA được sử dụng rộng rãi trong các ứng dụng an ninh và mật mã hiện đại như mạng máy tính, giao tiếp qua mạng, thanh toán trực tuyến và xác thực người dùng.

**Nhược điểm**

* Hiệu suất: Thuật toán RSA có hiệu suất thấp khi mã hóa và giải mã thông điệp dài. Quá trình tính toán với số nguyên lớn có thể mất nhiều thời gian và tài nguyên tính toán.
* Kích thước khóa: Để đảm bảo mức độ an toàn, kích thước khóa trong RSA phải đủ lớn. Điều này có nghĩa là kích thước khóa có thể lớn hơn so với các thuật toán mã hóa khác, làm tăng thời gian xử lý và lưu trữ.
* Sự phụ thuộc vào số nguyên tố: RSA dựa trên tính khó khăn của việc phân tích số nguyên lớn thành các thừa số nguyên tố. Nếu có sự tiến bộ trong việc tìm ra thuật toán phân tích số nguyên tố hiệu quả, sẽ ảnh hưởng đến tính bảo mật của RSA.

*Thứ hai là thuật toán AES*



**Khái niệm**:

* AES (Advanced Encryption Standard) là một thuật toán mã hóa đối xứng được sử dụng rộng rãi để bảo mật thông tin. Nó là một tiêu chuẩn mã hóa được chính phủ Hoa Kỳ chấp thuận và được sử dụng trong nhiều ứng dụng và hệ thống bảo mật trên toàn thế giới.

**Cách hoạt dộng của AES**:

* AES gồm ba mật mã khối AES-128, AES-192, AES-256 tương ứng với độ dài của key là 128 bit, 192 bit và 256 bit. Số vòng của key khác nhau, cụ thể 10 vòng cho 128 bit, 12 vòng cho 192 bit và 14 vòng cho 256 bit. Mỗi vòng đều thực hiện ba bước thay thế, biến đổi và hòa trộn khối plain text (văn bản thuần túy) đầu vào để biến nó thành Ciphertext (văn bản đã mã hóa).

**Thông tin được chính phủ phân loại theo ba cấp độ**:

* Bảo mật, bí mật, tối mật. Tất cả các độ dài của key từ 128, 192 và 256 bit đều được dùng ở cấp độ bảo mật, bí mật. Riêng với những thông tin tối mật để đảm bảo không xảy ra bất cứ sai sót nào phải cần đến key 192 hoặc 256 bit. Mật mã sẽ dùng một key riêng tư để mã hóa và giải mã dữ liệu và tất nhiên cả người gửi và người nhận đều phải nhận biết và sử dụng được key này.

**Các tiêu chí cần có của AES**

* NIST đưa ra yêu cầu đối với AES đó là phải sử dụng phương pháp mã hóa khối với độ dài của key là 128, 192 và 256 bit để mã hóa và giải mã dữ liệu. Ngoài ra AES phải đáp ứng được những tiêu chí sau:
* Bảo vệ: Đây là một trong những tính năng hàng đầu AES cần phải có để đánh bại các đối thủ khác. Nó phải có khả năng chống lại các cuộc tấn công mạnh, quy mô lớn.
* Chi phí: AES mở ra nhiều cơ hội cho người dùng bằng cách phát hành trên toàn cầu và miễn phí bản quyền.
* Khả năng thực hiện: Linh hoạt, phù hợp và đơn giản chính là 3 yếu tố quan trọng hội tụ ở AES để đáp ứng trọn vẹn nhu cầu của người dùng.

**Chọn thuật toán AES mới**

* Để có được một thuật toán đạt tới độ hoàn hảo vào tháng 8 năm 1999 NIST đã chọn MARS, RC6, Rijndael, Serpent, Twofish để phân tích. Tất cả 5 thuật toán trên đã được thử nghiệm trong ANSI, các ngôn ngữ như Java và C. Chúng được so sánh với nhau dựa vào các yêu tố như tốc độ mã hóa, mức độ tin cậy, thời gian thiết lập key và thuật toán, mức độ chống lại các cuộc công.
* Sau quá trình làm việc nghiêm túc, vào tháng 10 năm 2000 các thành viên của cộng đồng mật mã toàn cầu đã chọn Rijndael làm thuật toán đề xuất cho AES. Tháng 12 năm 2001 nó được Bộ trưởng Thương mại Hoa Kỳ chấp nhận và năm 2002 chính thức có hiệu lực như một tiêu chuẩn của chính phủ liên bang.
* Tháng 6 năm 2004, AES được chính phủ Mỹ thông báo với công chúng và được ứng dụng để bảo vệ thông tin đã phân loại. Kể từ đó AES phủ sóng ở nhiều lĩnh vực và nó được NSA (cơ quan an ninh Quốc gia Hoa Kỳ) chọn để bảo vệ hệ thống an ninh cho chính phủ và đất nước. Có thể khẳng định khi AES được chính phủ Hoa Kỳ sử dụng thành công đã tạo được tiếng vang rất lớn. Nó nhanh chóng được các tổ chức tư nhân săn đón để tạo “bức tường lửa” bảo vệ những dữ liệu mật.

**Mô tả thuật toán**

* Thuật toán AES (Advanced Encryption Standard) là một thuật toán mã hóa đối xứng sử dụng trong quá trình mã hóa và giải mã dữ liệu. Dưới đây là các bước cần thiết cho mã hóa và giải mã AES:

**Mã hóa:**

**1.Chuẩn bị:**

-Chọn một khóa (key) có độ dài phù hợp cho thuật toán AES (128-bit, 192-bit hoặc 256-bit).

-Chia dữ liệu cần mã hóa thành các khối có độ dài bằng nhau (thường là 128-bit) để xử lý.

**2. Khởi tạo:**

-Áp dụng phép XOR giữa khóa và khối dữ liệu ban đầu để tạo khối dữ liệu đầu tiên (plaintext).

**3.Vòng lặp:**

-Áp dụng một số vòng lặp (tùy thuộc vào độ dài khóa) để thay đổi trạng thái của khối dữ liệu theo các phép biến đổi đặcififnh của AES.

-Các phép biến đổi chính bao gồm: SubBytes, ShiftRows, MixColumns và AddRoundKey.

-Lặp lại quá trình trên cho đến khi đạt đến vòng cuối cùng.

**Kết quả cuối cùng:**

**4. Sau khi hoàn thành vòng lặp, bạn sẽ có được khối dữ liệu đã được mã hóa (ciphertext).**

**Giải mã:**

-Quá trình giải mã AES hoàn toàn tương tự quá trình mã hóa, nhưng sử dụng các phép biến đổi ngược lại. Dưới đây là các bước cần thiết:

**1.Chuẩn bị:**

-Sử dụng cùng một khóa mà bạn đã sử dụng để mã hóa để giải mã dữ liệu.

-Chia khối dữ liệu đã mã hóa thành các khối có độ dài bằng nhau (thường là 128-bit) để xử lý.

**2.Khởi tạo:**

-Áp dụng phép XOR giữa khóa và khối dữ liệu mã hóa để tạo khối dữ liệu đầu tiên.

**3. Vòng lặp**:

-Áp dụng các phép biến đổi ngược lại so với quá trình mã hóa: InvSubBytes, InvShiftRows, InvMixColumns và AddRoundKey.

-Lặp lại quá trình trên cho đến khi đạt đến vòng cuối cùng.

**4. Kết quả cuối cùng:**

-Sau khi hoàn thành vòng lặp, bạn sẽ có được khối dữ liệu đã được giải mã (plaintext).

\*Lưu ý: Mã hóa và giải mã AES đều sử dụng cùng một khóa, và quá trình giải mã chỉ là quá trình đảo ngược của quá trình mã hóa. Điều này chỉ áp dụng cho AES trong chế độ đơn giản nhất (ECB mode). Để đảm bảo tính an toàn cao hơn, cần áp dụng các chế độ mã hóa và các biến thể khác của AES như CBC (Cipher Block Chaining), CTR (Counter), hoặc GCM (Galois/Counter Mode) trong thực tế.

**Bảo mật**

\*Mặc dù mã hóa AES là thuật toán mạnh mẽ nhưng không tránh khỏi các cuộc tấn công. Nhiều nhà nghiên cứu đã công bố danh sách các cuộc tấn công, chẳng hạn như:

-Năm 2009 cryptanalysis (công cụ phân tích mật mã) đã tìm cách bẻ key của hệ thống AES.

-Năm 2009, AES-128 đã trở thành mục tiêu của một cuộc tấn công, tuy nhiên mối đe dọa này không gây thiệt hại gì.

-Đáng quan ngại nhất với mã hóa AES là phải đối mặt với side-channel attack (tấn công kênh phụ). Để xác định được key của mã hóa AES các cuộc tấn công này đã sử dụng những thông tin về thời gian, rò rỉ điện từ, manh mối âm thanh…

-AES nếu được triển khai đúng quy trình thì sẽ đảm bảo an toàn tuyệt đối. Thế nhưng một điều cần lưu ý đó là bất kỳ một hệ thống nào cũng có thể bị tấn công nếu hacker biết được key mã hóa. Do đó các key mã hóa AES phải được bảo vệ bằng nhiều cách khác nhau như dùng mật khẩu mạnh, xác thực, tường lửa hay phần mềm chống độc hại. Ngoài ra các tổ chức phải đào tạo nhân viên để nâng cao nhận thức của họ trong việc bảo mật thông tin, tránh bị kẻ xấu lợi dụng.

**Ưu và nhược điểm**

**Ưu điểm:**

-Bảo mật mạnh: AES được coi là một thuật toán mã hóa mạnh mẽ và an toàn. Hiện tại, không có phương pháp nhanh chóng và hiệu quả để phá vỡ AES bằng cách tấn công trực tiếp.

-Hiệu suất cao: AES được thiết kế để có hiệu suất cao trên các nền tảng phần cứng và phần mềm hiện đại. Nó có thể xử lý dữ liệu nhanh chóng và hiệu quả.

-Linh hoạt: AES hỗ trợ các khóa có độ dài 128, 192 và 256 bit, cho phép lựa chọn mức độ bảo mật phù hợp với yêu cầu cụ thể.

-Độ tin cậy cao: AES đã được sử dụng rộng rãi và kiểm tra bởi cộng đồng mật mã học và chính phủ. Nó đã chứng minh được tính tin cậy và khả năng chống lại các cuộc tấn công.

-Tiêu chuẩn quốc tế: AES là một tiêu chuẩn quốc tế được chính phủ Hoa Kỳ chấp thuận và được sử dụng rộng rãi trên toàn thế giới. Điều này đảm bảo tính tương thích và sự khả dụng rộng rãi của thuật toán.

**Nhược điểm**

-Cần tài nguyên tính toán: AES có thể đòi hỏi một lượng tài nguyên tính toán khá lớn, đặc biệt là khi thực hiện trên các thiết bị có tài nguyên hạn chế như thiết bị di động.

-Khả năng tấn công bên ngoài: Mặc dù AES được coi là an toàn, nó vẫn có thể bị tấn công thông qua các phương pháp như tấn công lực brute-force hoặc tấn công xung đột (collision attack) trên hàm băm.

-Sử dụng không đúng cách có thể gây lỗ hổng: Một điểm yếu của AES không phải là thuật toán mà là cách thực hiện và quản lý khóa. Nếu không sử dụng các khóa mạnh và các phương pháp thích hợp để quản lý khóa, có thể gây ra lỗ hổng bảo mật

\*Tuy nhiên, tổng thể, AES vẫn là một thuật toán mã hóa mạnh mẽ và được sử dụng rộng rãi trong nhiều ứng dụng bảo mật

***Thứ ba là thuật toán chính mà chúng em muốn đề cập đến là thuật toán MD5***



**1.Khái niệm:**

- Thuật toán MD5 (Message Digest Algorithm 5) là một thuật toán băm (hash) được sử dụng để tạo ra một giá trị băm duy nhất từ một đầu vào có độ dài bất kỳ.

-Và Chúng ta có thể hiểu như thế này : Thuật toán băm (hashing algorithm) là một phương pháp được sử dụng để biến đổi một đầu vào (dữ liệu) có độ dài bất kỳ thành một giá trị băm có độ dài cố định. Giá trị băm thường được gọi là mã băm (hash code) hoặc vân tay (fingerprint).

-MD5 tạo ra một chuỗi băm có độ dài cố định là 128 bit (16 byte) được phát triển vào năm 1991, bởi Giáo sư Ronald L. Rivest để thay thế cho hàm băm MD4 trước đó. MD5 không phải là thuật toán nhanh như thuật toán MD4, nhưng bảo mật dữ liệu trên MD5 an toàn hơn. MD5 được sử dụng trong nhiều ứng dụng bảo mật và để kiểm tra tính toàn vẹn của file dữ liệu, xác minh xem file đã bị thay đổi hay chưa.

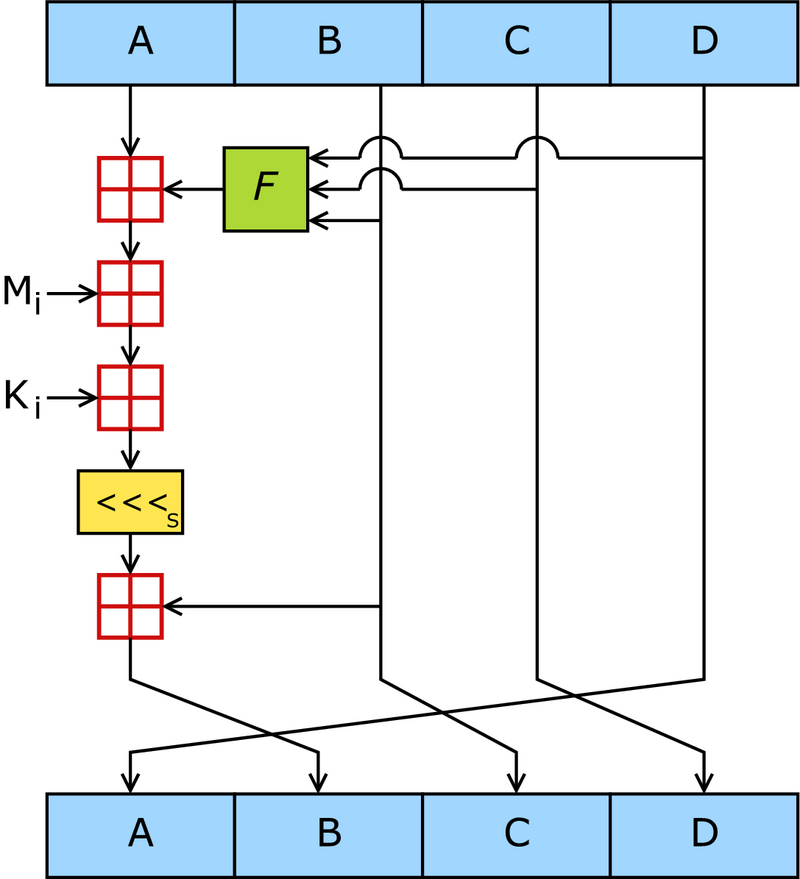
**2.Ứng dụng**

-Các đồng hóa MD5 được dùng rộng rãi trong các phần mềm trên toàn thế giới để đảm bảo việc truyền tập tin được nguyên vẹn. Ví dụ, máy chủ tập tin thường cung cấp một tổng kiểm (checksum) MD5 được tính toán trước cho tập tin, để người dùng có thể so sánh với tổng kiểm của tập tin đã tải về. Những hệ điều hành dựa trên nền tảng Unix luôn kèm theo tính năng MD5 sum trong các gói phân phối của họ, trong khi người dùng Windows sử dụng ứng dụng của hãng thứ ba.

-Tuy nhiên, hiện nay dễ dàng tạo ra xung đột MD5, một người có thể tạo ra một tập tin để tạo ra tập tin thứ hai với cùng một tổng kiểm, do đó kỹ thuật này không thể chống lại một vài dạng giả mạo nguy hiểm. Ngoài ra, trong một số trường hợp tổng kiểm không thể tin tưởng được (ví dụ, nếu nó được lấy từ một lệnh như tập tin đã tải về), trong trường hợp đó MD5 chỉ có thể có chức năng kiểm tra lỗi: nó sẽ nhận ra một lỗi hoặc tải về chưa xong, rất dễ xảy ra khi tải tập tin lớn.

-MD5 được dùng rộng rãi để lưu trữ mật khẩu. Để giảm bớt sự dễ thương tổn đề cập ở trên, ta có thể thêm muối (salt) vào mật khẩu trước khi băm chúng. Một vài hiện thực có thể áp dụng vào hàm băm hơn một lần- xem làm mạnh thêm khóa.

**3.Thuật toán**



Quá trình mã hóa bằng MD5 được thực hiện theo các bước sau:

1. Chuẩn bị dữ liệu đầu vào: Dữ liệu đầu vào có thể là bất kỳ loại dữ liệu nào, chẳng hạn như một văn bản, một tập tin, hoặc một chuỗi ký tự. Dữ liệu đầu vào này được chuyển đổi thành dạng chuỗi nhị phân (bit string) trước khi tiến hành quá trình băm.

2. Bước khởi tạo: Một số giá trị khởi tạo được sử dụng trong quá trình băm. Đối với MD5, các giá trị khởi tạo này được định nghĩa sẵn và được sử dụng để tạo ra các giá trị trung gian trong quá trình băm.

3. Chia dữ liệu thành các khối: Dữ liệu đầu vào được chia thành các khối có độ dài cố định (thông thường là 512 bit) để xử lý.

4. Băm từng khối dữ liệu: Quá trình băm áp dụng các phép biến đổi trên từng khối dữ liệu. MD5 sử dụng một loạt các phép biến đổi như xoay trái (left rotation), phép AND, OR, XOR, và cộng modulo 2.

5. Kết hợp các khối: Kết quả băm của từng khối dữ liệu được kết hợp để tạo ra giá trị băm cuối cùng. Quá trình này thường bao gồm việc thêm các giá trị băm trung gian vào giữa các khối dữ liệu và thực hiện các phép toán hỗn hợp để tạo ra giá trị băm cuối cùng.

6. Kết quả: Cuối cùng, thuật toán MD5 trả về giá trị băm cuối cùng, mà thông thường được biểu diễn dưới dạng một chuỗi số hexa hoặc một chuỗi ký tự.

\*Không có bước giải mã cụ thể cho thuật toán MD5 vì nó chỉ hoạt động theo chiều một chiều. Điều này có nghĩa là không thể khôi phục lại thông điệp gốc từ giá trị băm đã được tạo ra bởi MD5.

Vì vậy, MD5 thường được sử dụng để kiểm tra tính toàn vẹn dữ liệu, xác thực mậtkhẩu, hoặc làm vân tay cho các tệp tin. Nếu bạn quan tâm đến quá trình mã hóa và giải mã, bạn có thể xem xét các thuật toán mã hóa đối xứng như AES hoặc thuật toán mã hóa không đối xứng như RSA.

**4 Ưu và nhược điểm**

Thuật toán MD5 có những ưu điểm và nhược điểm sau:

**Ưu điểm của MD5:**

1. Tính nhanh: MD5 được thiết kế để thực hiện nhanh chóng trên các máy tính hiện đại. Nó có hiệu suất cao và có thể băm nhanh chóng các khối dữ liệu lớn.

2. Kích thước băm cố định: Giá trị băm MD5 có kích thước cố định là 128 bit (16 byte). Điều này giúp trong việc lưu trữ và so sánh giá trị băm.

3. Dễ sử dụng: Việc sử dụng MD5 rất đơn giản. Chỉ cần gọi hàm MD5 trên dữ liệu đầu vào, thuật toán sẽ trả về giá trị băm tương ứng.

**Nhược điểm của MD5:**

1. Độ tin cậy yếu: MD5 đã bị các lỗ hổng và tấn công nghiêm trọng. Nhiều phương pháp tấn công đã được phát triển để tìm ra các xung đột (collision) trong MD5, nghĩa là tìm được hai đầu vào khác nhau nhưng có cùng giá trị băm. Do đó, MD5 không còn đảm bảo tính chống xung đột.

2. Khả năng tìm kiếm ngược (reverse lookup): Do tính chất của MD5, nếu có một giá trị băm, người tấn công có thể thực hiện một cuộc tìm kiếm ngược để tìm ra đầu vào ban đầu. Điều này có thể đặt rủi ro cho các ứng dụng yêu cầu tính toàn vẹn dữ liệu.

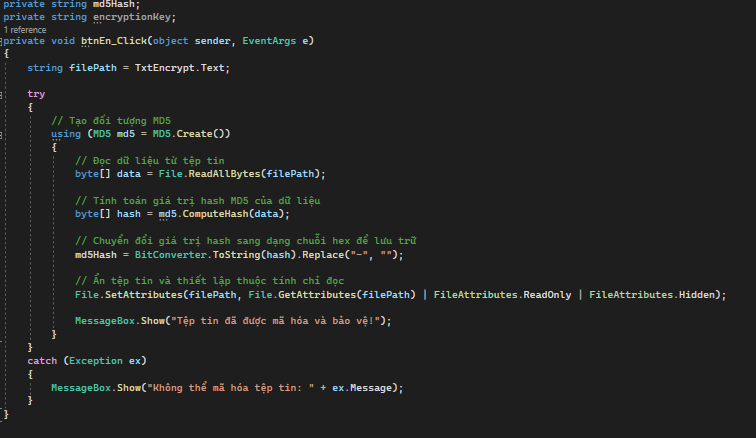
3. Sử dụng rộng rãi của thuật toán khác: Do các vấn đề bảo mật nêu trên, MD5 không nên được sử dụng trong các ứng dụng yêu cầu tính bảo mật cao. Thay vào đó, các thuật toán băm khác như SHA-256 được khuyến nghị

**Chương 3:**

Câu hỏi 1: Giải quyết từng vấn đề của bài toán như thế nào?

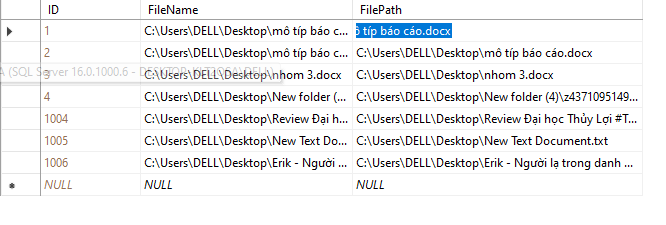
Để thực hiện đề tài của bạn về việc sử dụng thuật toán MD5 để tạo một phần mềm lưu trữ dữ liệu bảo mật, cần xem xét các vấn đề :

1) Lựa chọn ngôn ngữ và môi trường phát triển: Ngôn ngữ lựa chọn C#

2) Triển khai thuật toán MD5: 

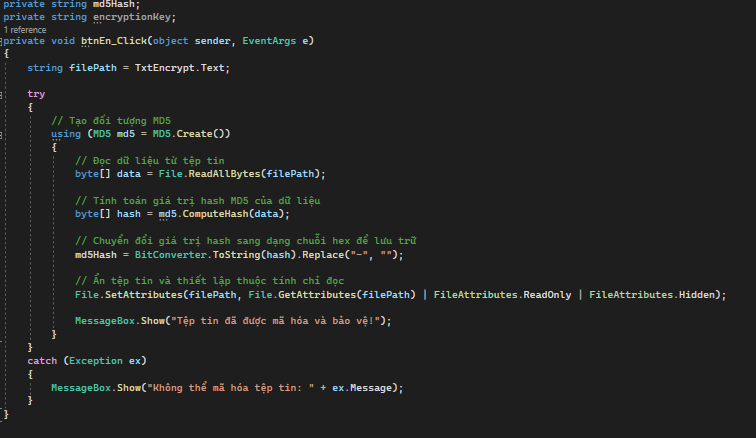
3) Thiết kế hệ thống lưu trữ dữ liệu:





4) Xây dựng chức năng mã hóa dữ liệu:

Trong đoạn mã bạn đã cung cấp, chức năng mã hóa dữ liệu được triển khai trong phương thức `btnEn\_Click`. Dưới đây là phần mã của chức năng mã hóa dữ liệu trong đoạn mã:



Trong phương thức `btnEn\_Click`, thực hiện các bước sau để mã hóa dữ liệu:

1. Đọc dữ liệu từ tệp tin được chọn bằng cách sử dụng phương thức `File.ReadAllBytes`.

2. Tạo một đối tượng `MD5` để tính toán giá trị hash MD5.

3. Sử dụng `md5.ComputeHash(data)` để tính toán giá trị hash của dữ liệu.

4. Chuyển đổi giá trị hash sang dạng chuỗi hex bằng cách sử dụng `BitConverter.ToString(hash).Replace("-", "")`.

5. Thiết lập thuộc tính của tệp tin để ẩn và chỉ đọc bằng cách sử dụng `File.SetAttributes`.

6. Hiển thị thông báo cho người dùng để xác nhận rằng tệp tin đã được mã hóa và bảo vệ.

Với phần mã này, khi người dùng nhấn vào nút "Mã hóa", tệp tin được chọn sẽ được mã hóa bằng thuật toán MD5 và thuộc tính của tệp tin sẽ được thay đổi để ẩn và chỉ đọc.

Câu hỏi 2: Kết quả ra sao?

Ứng dụng này có các chức năng sau:

1. Lưu đường dẫn của tệp tin vào cơ sở dữ liệu:

- Khi người dùng chọn một tệp tin và nhấp vào nút "Lưu", đường dẫn tệp tin được lưu vào cơ sở dữ liệu. Đường dẫn tệp tin được lưu trong bảng "Files" trong cơ sở dữ liệu.

- Khi thao tác thành công, ứng dụng hiển thị thông báo "Đường dẫn đã được lưu thành công vào cơ sở dữ liệu."

2. Mã hóa tệp tin sử dụng MD5:

- Khi người dùng chọn một tệp tin và nhấp vào nút "Mã hóa", tệp tin đó sẽ được mã hóa bằng thuật toán MD5.

- Giá trị hash MD5 của tệp tin sau mã hóa được tính toán và hiển thị trong hộp thoại thông báo.

- Thuộc tính của tệp tin được thay đổi để ẩn và chỉ đọc.

3. Giải mã tệp tin:

- Khi người dùng chọn một tệp tin đã được mã hóa và nhấp vào nút "Giải mã", tệp tin sẽ được giải mã.

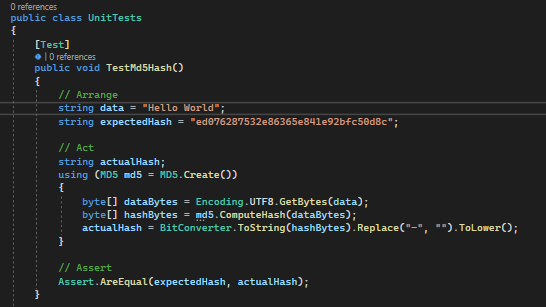
- Thuộc tính của tệp tin được thay đổi để bỏ chế độ chỉ đọc và ẩn.

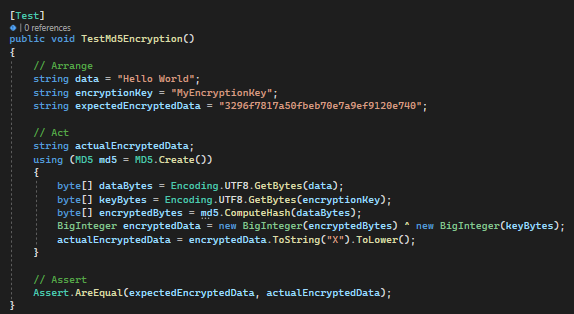
4. Chức năng khác:

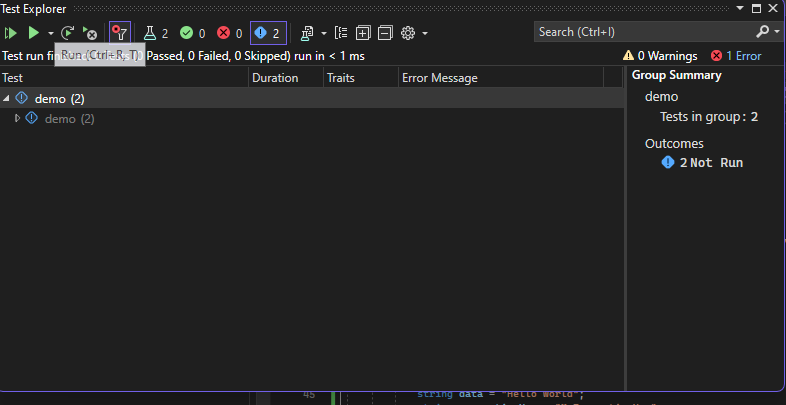
- Người dùng có thể mở một tệp tin văn bản, lưu trữ đường dẫn của tệp tin, đặt khóa mã hóa,

Câu hỏi 3: Kiểm thử thế nào?

1. Kiểm thử đơn vị (Unit testing): Viết các testcase để kiểm tra tính đúng đắn của các hàm mã hóa và kiểm tra tính toàn vẹn của mã hash.







Chương 4: Kết luận và kiến nghị

Câu hỏi 1: Bạn và các thành viên đã làm đc gì

* Đã tìm hiểu được cách tạo giao diện từ window form
* Đã biệt tạo csdl và liên kết với phần code
* Và đã biết hiểu rõ được cơ chế hoạt động của các thuật toán
* Đã tìm hiểu được thế giới đã tạo được các phần mềm nổi tiếng bảo mật và lưu trữ

Câu hỏi 2: còn tồn đọng gì?

* Giao diện chưa được dẹp
* Thuật toán MD5 về bảo mật vẫn còn yếu
* Vẫn chưa thể cho file lưu và phần mềm chỉ lưu được đường dẫn
* Phần hiện thị đường dẫn file vẫn còn sơ sài
* Chưa lược gọn được phần mềm

Câu hỏi 3: phương hướng triển khai nốt,nâng cấp Như thế nào

Phương hướng Nâng cấp :

* Cải thiện dao diện đẹp và bắt mắt hơn
* Nâng cấp MD5 lên AES để tăng khả năng bảo mật
* Cải thiện phần lưu file để có thể đưa file vào phần mềm và lưu trữ
* Cải thiện phần hiển thị đường dẫn thành phần xem file và có thể thao tác trong phần hiển thị
* Lược gọn phần mềm đê không cần nhiều thao tác

Chương 5:

Tài liệu tham khảo ( nguồn tham Khảo)

ChatGPT