BẢO MẬT DỮ LIỆU VỚI KỸ THUẬT AES – DCT WATERMARKING

Dương Anh Đức[+] – Nguyễn Thanh Sơn* – Trần Minh Triết[+]
[+] Khoa Công Nghệ Thông Tin – Trường Đại học Khoa học Tự Nhiên – ĐHQG-HCM
* Trường Phổ Thông Năng Khiếu – ĐHQG-HCM
(Bài nhận ngày 10 tháng 7 năm 2003)

TÓM TẮT: *Trong bài viết này chúng tôi giới thiệu kỹ thuật bảo mật dữ liệu thông qua sự kết hợp giữa kỹ thuật chuẩn mã hóa nâng cao AES và kỹ thuật che dấu dữ liệu sử dụng phép biến đổi cosin rời rạc DCT Watermarking. Kỹ thuật này cho phép nâng cao khả năng bảo mật của khối lượng lớn dữ liệu nhờ tận dụng các ưu thế của cả 2 kỹ thuật: mã hóa và che dấu thông tin.*

1. Mở đầu

Trong những năm gần đây, công nghệ thông tin đã có những bước phát triển vượt bậc, theo đó, việc trao đổi và lưu trữ thông tin cũng phát triển hơn bao giờ hết. Một trong những vấn đề muôn thuở của việc lưu trữ và trao đổi thông tin là vấn đề bảo mật. Bài toán đặt ra là làm sao bảo mật được một khối lượng thông tin cực lớn với độ an toàn cao. Một số phương pháp đã chọn cách mã hóa thông tin để chống lại sự xâm nhập bất hợp pháp. Số khác chọn cách che dấu dữ liệu bằng cách tạo một vỏ bọc dữ liệu giả để tránh việc bị tấn công. Mỗi nhóm phương pháp đều có những mặt mạnh và yếu khác nhau.

Khai thác những điểm mạnh của mỗi nhóm phương pháp, kỹ thuật AES – DCT Watermarking được trình bày trong bài viết này có khả năng bảo mật dữ liệu rất cao: lẩn tránh được các cuộc tấn công và vô hiệu hóa các xâm nhập bất hợp pháp.

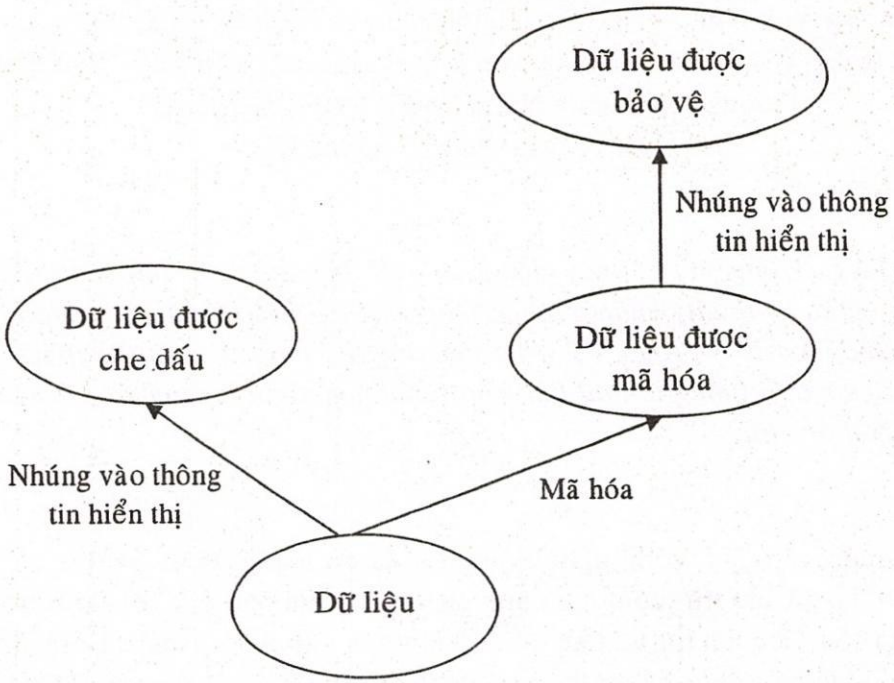
2. Các mức bảo mật thông tin

a) Dữ liệu: Ở trạng thái nguyên thủy, dữ liệu thường không được bảo vệ. Trong trạng thái này bất cứ đối tượng nào cũng có thể xâm nhập, sử dụng hay thậm chí có thể sửa đổi. Trạng thái này thường không tồn tại lâu dài trong các hệ thống thông tin.

b) Che dấu dữ liệu: Để tránh việc truy xuất dữ liệu tùy tiện, một trong kỹ thuật thường được sử dụng là nhúng dữ liệu vào một thông tin nền sau đó lưu trữ hay trao đổi cả gói tin – thông tin nền có vai trò như công cụ che dấu dữ liệu. Các thông tin nền thường được chọn như âm thanh, hình ảnh [1, 3, 4]. Kỹ thuật này giúp dữ liệu thoát khỏi các "con mắt tò mò" của những người không được phép vì các chương trình hiển thị thông thường chỉ thể hiện gói tin này ở rất giống như thông tin nền. Tuy nhiên, việc phát hiện tự động với các chương trình máy tính là điều có thể làm được dễ dàng. Kỹ thuật này chỉ giúp dữ liệu né tránh để không bị phát hiện. Một khi đã bị phát hiện, việc truy xuất dữ liệu là điều khó tránh khỏi.

c) Mã hoá dữ liệu: Một kỹ thuật rất thường được sử dụng để bảo vệ dữ liệu là mã hóa chúng trước khi trao đổi. Chỉ khi có được "chìa khóa" mới có thể giải mã và truy xuất được dữ liệu này [5-7]. Có nhiều phương pháp mã hóa dữ liệu, các phương pháp này đều có những ưu và nhược điểm khác nhau tuy nhiên chúng đều có một điểm yếu chung: đó là "chìa khóa", nếu "chìa khóa" bị phát hiện, việc bảo mật thông tin sẽ

không còn. Kỹ thuật này không che dấu dữ liệu mà lưu trữ, trao đổi dữ liệu ở dạng được mã hóa vì vậy thường xuyên bị tấn công.



Các cấp độ bảo vệ dữ liệu

d) Che dấu dữ liệu đã được mã hoá: Để hạn chế các nhược điểm của các kỹ thuật được giới thiệu trong mục (b) và (c), chúng tôi đề xuất kỹ thuật kết hợp: Mã hóa dữ liệu và sau đó che dấu chúng đi bằng một thông tin nền. Như vậy dữ liệu sẽ tránh được sự phát hiện và do đó sẽ không bị tấn công thường xuyên. Một khi bị phát hiện, dữ liệu cũng chưa thể sử dụng nếu không được giải mã. Vì vậy, có thể nói kỹ thuật này bảo vệ được sự bí mật và an toàn cho dữ liệu rất cao.

3. Che dấu dữ liệu đã được mã hoá

Việc mã hóa dữ liệu trong lưu trữ và trao đổi thông tin đã đạt được nhiều kết quả rất tốt. Rất nhiều phương pháp mã hóa đã được áp dụng đem lại độ tin cậy khá lớn trong lĩnh vực bảo mật dữ liệu. Tuy nhiên, như đã nói ở trên, do không có được kỹ thuật "tàng hình" nên dữ liệu bị mã hóa rất dễ bị tấn công, xác suất thành công của những kẻ tấn công bất hợp pháp cũng không phải nhỏ. Vì vậy nếu kết hợp thêm kỹ thuật "tàng hình" dữ liệu thì độ an toàn sẽ tăng lên rất nhiều do tránh được hầu hết các cuộc tấn công có chủ ý.

Việc kết hợp 2 kỹ thuật: mã hóa và che dấu dữ liệu đem lại độ an toàn cao cho dữ liệu nhưng lại gặp phải một số khó khăn như: chất lượng thông tin nền tỉ lệ nghịch với khối lượng dữ liệu cần che dấu, độ an toàn dữ liệu cần phải cao với một thời gian xử lý chấp nhận được.

Một trong các mục tiêu của phương pháp là che dấu dữ liệu nhưng nếu muốn ẩn một lượng dữ liệu khá lớn thì phải có một lượng thông tin nền lớn hơn rất nhiều (có thể gấp hàng chục lần) chiếm nhiều tài nguyên của hệ thống, làm chậm đường truyền, v.v.. Nếu giảm lượng thông tin nền thì độ trung thực không còn nữa, gói tin có chứa dữ liệu ẩn sẽ khác xa so với thông tin nền ban đầu, mục tiêu ẩn dấu dữ liệu không đáp ứng được.

Ảnh gốc

Ảnh có "nhúng" ít dữ liệu

Ảnh có "nhúng" nhiều dữ liệu

Chất lượng tỉ lệ nghịch với khối lượng

Một mục tiêu khác của kỹ thuật là yêu cầu bảo mật cao và tiết kiệm thời gian xử lý. Trong các phương pháp mã hóa dữ liệu, chúng tôi nhận thấy phương pháp AES có thể đáp được mục tiêu này.

Từ những nhận xét như trên, chúng tôi đề xuất một phương pháp cụ thể để bảo vệ dữ liệu: Mã hóa dữ liệu bằng kỹ thuật khóa bí mật AES và che dấu thông tin bằng kỹ thuật DCT Watermarking.

4. Bảo mật dữ liệu với kỹ thuật AES – DCT Watermarking

Với phương pháp này dữ liệu cần được bảo vệ (D) sẽ được mã hóa bằng kỹ thuật AES, sau đó "tàng hình" vào một ảnh nền (I) bằng kỹ thuật DCT Watermarking để tạo nên ảnh (I'). I' sẽ được dùng để lưu trữ và trao đổi dữ liệu. Để phát hiện được sự tấn công, trước khi nhúng dữ liệu vào ảnh nền, dữ liệu còn được thêm vào đó các ký hiệu nhận dạng để kiểm tra phát hiện các thay đổi.

a. Mã hóa dữ liệu bằng kỹ thuật AES:

Với tốc độ và khả năng xử lý ngày càng được nâng cao của các bộ vi xử lý hiện nay, phương pháp mã hóa chuẩn (Data Encryption Standard – DES) trở nên không an toàn trong bảo mật thông tin. Các tác giả Vincent Rijmen và Joan Daeman đã đề xuất thuật toán Rijndael và được chọn thành chuẩn mã hóa nâng cao (Advanced Encryption Standard - AES) từ tháng 10 năm 2000 [9, 10]. Phương pháp mã hóa Rijndael là phương pháp mã hóa theo khối (block cipher) có kích thước khối và mã khóa thay đổi linh hoạt với các giá trị 128, 192 hay 256 bits [9, 10]. Phương pháp này thích hợp ứng dụng trên nhiều hệ thống khác nhau từ các thẻ thông minh cho đến các máy tính cá nhân.

Quy trình mã hóa Rijndael có thể được tóm tắt như sau:

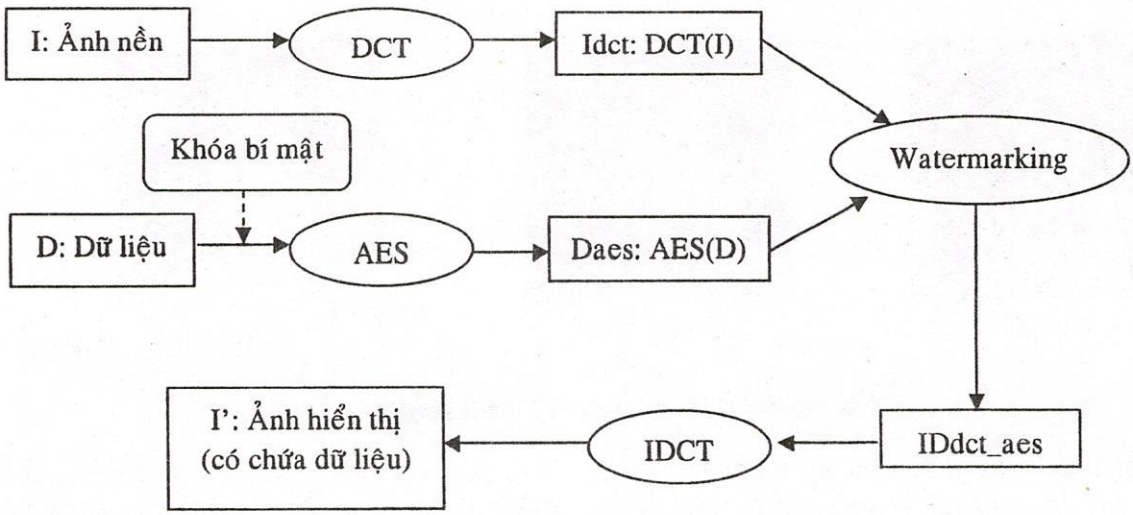
- Thực hiện thao tác AddRoundKey đầu tiên.
- Nr–1 chu kỳ mã hóa bình thường: mỗi chu kỳ bao gồm 4 bước biến đổi liên tiếp nhau: SubBytes, ShiftRows, MixColumns, và AddRoundKey.
- Thực hiện chu kỳ mã hóa cuối cùng: chu kỳ này bỏ qua thao tác MixColumns.

Trong đó:

- Nr : Số lượng chu kỳ phụ thuộc vào giá trị của độ dài khối và độ dài khóa theo công thức:

$Nr=\max\{Nb, Nk\}+6$ (Nb: độ dài khối /32, Nk: độ dài khóa /32)

- AddRoundKey: cộng (⊕) mã khóa của chu kỳ vào trạng thái hiện hành. Độ dài của mã khóa của chu kỳ bằng với kích thước của trạng thái.
- SubBytes : thay thế phi tuyến mỗi byte trong trạng thái hiện hành thông qua bảng thay thế (S-box).



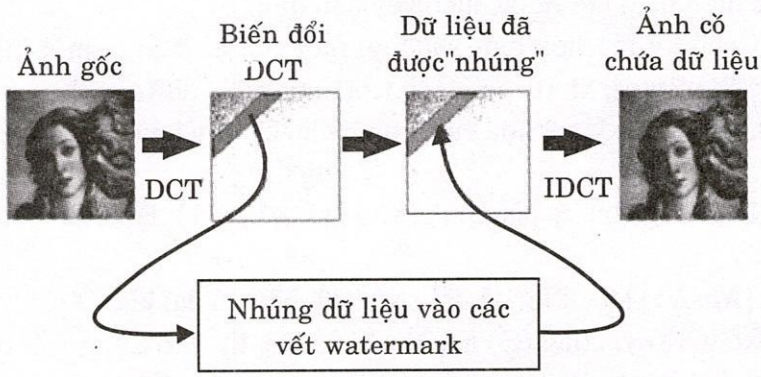
Qui trình mã hóa và che dấu dữ liệu

- MixColumns: trộn thông tin của từng cột trong trạng thái hiện hành. Mỗi cột được xử lý độc lập.
- ShiftRows: dịch chuyển xoay vòng từng dòng của trạng thái hiện hành với di số khác nhau.

Kỹ thuật mã hóa Rijndael và các phiên bản mở rộng của nó có rất nhiều tính năng ưu việt: tốc độ mã hóa rất cao, khả năng an toàn lớn (với các máy tính các nhân hiện nay, việc tấn công truy tìm mật mã là điều không khả thi, nhất là đối với các phiên bản mở rộng của thuật toán) [10-14].

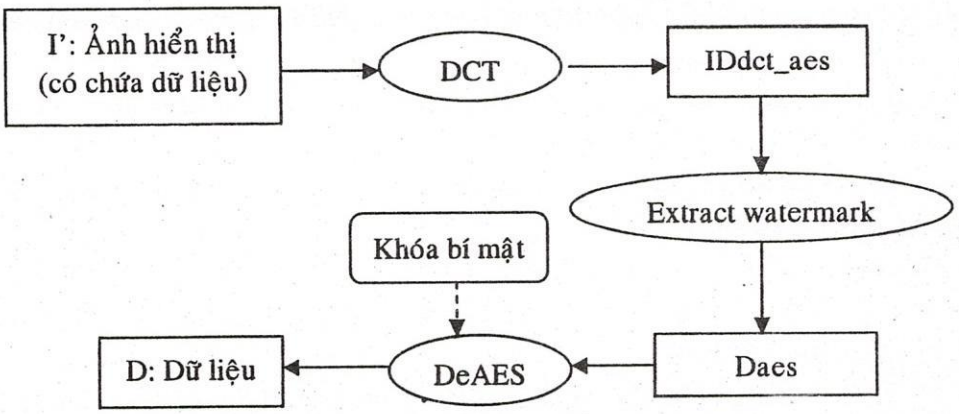
b. Che dấu dữ liệu với kỹ thuật DCT Watermarking (Watermarking dùng phép biến đổi cosin rời rạc).

Để che dấu dữ liệu, một trong những kỹ thuật hiện thường được dùng là Watermarking [1,4]. Kỹ thuật này dùng các phép biến đổi trên dữ liệu rời rạc như DFT, DCT [8] để tập trung năng lượng của ảnh nền về một miền có mật độ cao, tạo ra vùng không gian năng lượng thấp của ảnh khá lớn, từ đó các vết watermark sẽ được nhúng vào vùng không gian này [3]. Do khả năng tập trung năng lượng của của phép biến đổi DCT là rất lớn nên lượng dữ liệu được nhúng vào ảnh cũng lớn, giải quyết được tương đối bài toán về khối lượng dữ liệu lưu trữ [3].



c. Thu nhận lại dữ liệu từ gói tin:

Với ảnh có chứa dữ liệu I', việc trích xuất dữ liệu được thực hiện theo qui trình ngược lại. Việc tách các vết watermark có chứa dữ liệu đã mã hóa có thể thực hiện không cần phải có ảnh gốc [3]. Dữ liệu sau khi trích xuất được sẽ được kiểm tra các ký hiệu nhận dạng và chuyển đến bộ phận giải mã. Do kỹ thuật AES sử dụng bảng tra cho phép giải quyết bài toán giải mã cũng có thời gian nhanh như khi mã hóa vì vậy toàn bộ việc thu hồi dữ liệu từ gói tin được thực rất nhanh [10-14].



Qui trình thu hồi dữ liệu

KẾT LUẬN

Kỹ thuật AES – DCT Watermarking đã phát huy được khả năng tự bảo mật của dữ liệu với phương pháp mã hóa AES có độ an toàn cao, đồng thời cũng phát huy được khả năng lẩn tránh các cuộc tấn công với kỹ thuật Watermarking.

Ngoài ra, việc áp dụng phép biến đổi cosin rời rạc đã nâng cao khả năng "chứa" của ảnh nền lên rất cao, có thể che dấu được một lượng dữ liệu lớn trong một ảnh nền kích thước không lớn lắm.

Những điều này cho thấy kỹ thuật AES – DCT Watermarking hoàn toàn có thể áp dụng để bảo mật khối lượng lớn dữ liệu với độ an toàn rất cao.

APPLYING AES - DCT WATERMARKING IN DATA SECURITY

Duong Anh Duc, Nguyen Thanh Son, Tran Minh Triet

ABSTRACT: *In this paper we present a new technique used to protect secured data embed in images. This technique is the combination of the Advanced Encryption Standard (AES) and the DCT-based data hiding technique. It provides better capabilities of hiding a large amount of secured data in images taking advantages of these two techniques: data encryption and data hiding.*

TÀI LIỆU THAM KHẢO

1. Dương Anh Đức, Trần Minh Triết, Đặng Tuân, Hồ Ngọc Lâm, *Watermarking – Tổng quan và ứng dụng trong các hệ thống quản lý và bảo vệ sản phẩm trí tuệ*. Kỷ yếu hội nghị khoa học trường Đại học Khoa học Tự nhiên – ĐHQG TP.HCM lần 3, 10-2002.

2. Dương Anh Đức, Trần Minh Triết, *Kỹ thuật watermarking an toàn đối với các phép biến đổi hình học*. Kỷ yếu hội nghị khoa học trường Đại học Khoa học Tự nhiên – ĐHQG TP.HCM lần 3, 10-2002.
3. A. Piva, *A DCT-Domain Watermarking System for Copyright Protection of Digital Images*, Ciclo XI, 1998
4. J.F.Delaigle, *Protection of Intellectual Property of Images by Preceptual Watermarking*, Ph.D. Thesis, Sep 2000
5. Douglas R. Stison, *Cryptography – Theory and Practice*, CRC Press, 1995.
6. Richard Demillo, *Applied Cryptography, Cryptographic Protocols, and Computer Security Model*, American Mathematical Society, 1983.
7. Bruce Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, 2nd Edition, John Wiley & Sons, Inc., 1995.
8. Nguyễn Thanh Sơn, *Xây dựng một số công cụ trợ giúp lưu trữ và xử lý ảnh lớn*, Luận văn tốt nghiệp cử nhân, 1995.
9. Dương Anh Đức, Trần Minh Triết, Lương Hán Cơ, *Chuẩn mã hóa mới AES*, kỷ yếu Hội thảo Quốc gia "Một số vấn đề chọn lọc của Công Nghệ Thông Tin" lần 4, Hải Phòng, Việt Nam, tháng 6 năm 2001
10. Trần Minh Triết, Lương Hán Cơ, *Nghiên cứu các phương pháp mã hóa và ứng dụng*, Luận văn tốt nghiệp Cử nhân Tin học, 2001
11. Dương Anh Đức, Trần Minh Triết, Lương Hán Cơ, *The 256/384/512-bit version of the Rijndael Block Cipher*, Tạp chí Tin học và Điều khiển, Việt Nam, tập 17, số 4, tháng 12 năm 2001.
12. Duong Anh Duc, Tran Minh Triet, Luong Han Co, *The extended Rijndael-like Block Ciphers*, International Conference on Information Technology: Coding and Computing – 2002, The Orleans, Las Vegas, Nevada, USA, tháng 4 năm 2002
13. Duong Anh Duc, Tran Minh Triet, Luong Han Co, *The extended versions of the Advanced Encryption Standard*, Workshop on Applied Cryptology, Coding Theory and Data Integrity, Singapore, tháng 12 năm 2001
14. Duong Anh Duc, Tran Minh Triet, Luong Han Co, *The extended version of the Rijndael Block Cipher*, Journal of Institute of Mathematics and Computer Sciences, India, Vol. 12, No. 2, tháng 12 năm 2001.