TÌM HIỂU MD5 VÀ CÁC GIẢI THUẬT MÃ HÓA

MD5 (Message-Digest algorithm 5) là một hàm băm để mã hóa với giá trị băm là 128bit. Từng được xem là một chuẩn trên Internet, MD5 đã được sữ dụng rông rải trong các chương trình an ninh mạng, và cũng thường được dùng để kiểm tra tính nguyên vẹn của tập tin.

MD5 được thiết kế bởi Ronald Rivest vào năm 1991 để thay thế cho hàm băm trước đó, MD4 (cũng do ông thiết kế, trước đó nữa là MD2).

MD5 có 2 ứng dụng quan trọng:

1/ MD5 được sử dụng rộng rải trong thế giới phần mềm để đảm bảo rằng tập tin tải về không bị hỏng. Người sử dụng có thể so sánh giữa thông số kiểm tra phần mềm bằng MD5 được công bố với thông số kiểm tra phần mềm tải về bằng MD5. Hệ điều hành Unix sử dụng MD5 để kiểm tra các gói mà nó phân phối, trong khi hệ điều hành Windows sử dụng phần mềm của hãng thứ ba.

2/ MD5 được dùng để mã hóa mật khẩu. Mục đích của việc mã hóa này là biến đổi một chuỗi mật khẩu thành một đoạn mã khác, sao cho từ đoạn mã đó không thể nào lần trở lại mật khẩu. Có nghĩa là việc giải mã là không thể hoặc phải mất một khoãng thời gian vô tận (đủ để làm nản lòng các hacker).

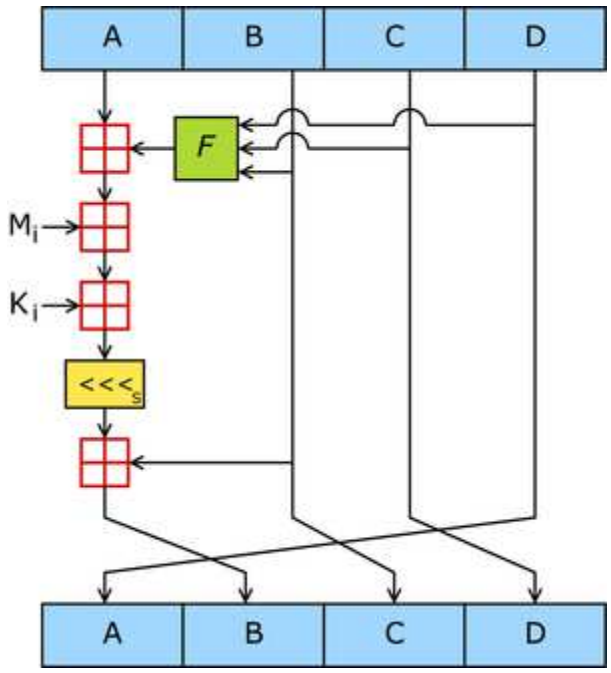
Thuật giải

MD5 biến đổi một thông điệp có chiều dài bất kì thành một khối có kích thước cố định 128 bits. Thông điệp đưa vào sẻ được cắt thành các khối 512 bits. Thông điệp được đưa vào bộ đệm để chiều dài của nó sẻ chia hết cho 512. Bộ đệm hoạt động như sau:
- Trước tiên nó sẻ chèn bit 1 vào cuối thông điệp.
- Tiếp đó là hàng loạt bit Zero cho tới khi chiều dài của nó nhỏ hơn bội số của 512 một khoảng 64 bit .
- Phần còn lại sẻ được lấp đầy bởi một số nguyên 64 bit biểu diễn chiều dài ban đầu của thông điệp.

Thuật toán chính của MD5 hoạt động trên một bộ 128 bit. Chia nhỏ nó ra thành 4 từ 32 bit, kí hiệu là A,B,C và D. Các giá trị này là các hằng số cố định.

Sau đó thuật toán chính sẻ luân phiên hoạt động trên các khối 512 bit. Mỗi khối sẻ phối hợp với một bộ. Quá trình xữ lý một khối thông điệp bao gồm 4 bước tương tự nhau, gọi là vòng ("round"). Mỗi vòng lại gồm 16 quá trình tương tự nhau dựa trên hàm một chiều F, phép cộng module và phép xoay trái…

Hình bên dưới mô tả một quá trình trong một vòng. Có 4 hàm một chiều F có thể sử dụng. Mỗi vòng sử dụng một hàm khác nhau.



Hàm băm MD5 (còn được gọi là hàm tóm tắt thông điệp - message degests) sẻ trả về một chuỗi số thập lục phân gồm 32 số liên tiếp. Dưới đây là các ví dụ mô tả các kết quả thu được sau khi băm.

MD5("The quick brown fox jumps over the lazy dog") = 9e107d9d372bb6826bd81d3542a419d6

Thậm chỉ chỉ cần một tahy đổi nhỏ cũng làm thay đổi hoàn toàn kết quả trả về:

MD5("The quick brown fox jumps over the lazy cog") = 1055d3e698d289f2af8663725127bd4b

Ngay cả một chuỗi rổng cũng cho ra một kết quả phức tạp:

MD5("") = d41d8cd98f00b204e9800998ecf8427e

Những Lỗ Hổng

Bất cứ thuật toán mã hóa nào rồi cũng bị giải mã. Với MD5, ngay từ năm 1996, người ta đã tìm thấy lỗ hổng của nó. Mặc dù lúc đó còn chưa rõ ràng lắm nhưng các chuyên gia mã hóa đã nghĩ đến việc phải đưa ra một thuật giải khác, như là SHA-1…

Và rồi gần đây, giới mã hoá đã xôn xao với thông tin các thuật toán bên trong nhiều ứng dụng bảo mật thông dụng, như chữ ký điện tử, cũng... có lỗ hổng (trong đó có MD5).

Mọi chuyện bắt đầu từ năm, khi nhà khoa học máy tính người Pháp Antoine Joux phát hiện ra một lỗ hổng trong thuật toán phổ biến MD5, thường dùng trong công nghệ chữ ký điện tử. Ngay sau đó, bốn nhà nghiên cứu người Trung Quốc lại phát hành công trình nghiên cứu chỉ ra cách xuyên phá thuật toán thứ hai có tên SHA-0.

Tuy chỉ mới ở giai đoạn nghiên cứu sơ bộ song những phát hiện này có thể tạo điều kiện để kẻ xấu cài những chương trình cửa sau (backdoor) bí mật vào trong mã máy tính, hoặc giả mạo chữ ký điện tử. Trừ phi một thuật toán mới, bảo mật hơn được xây dựng và đưa vào sử dụng!

Một phát hiện thứ ba, được đón đợi và đánh giá rất cao được công bố trong hội thảo Crypto. Hai nhà nghiên cứu Eli Biham và Rafi Chen của Viện Công nghệ Israel đã diễn thuyết về cách nhận dạng các hình thức tấn công vào chức năng bảo mật của thuật toán SHA-0, một thuật toán có sơ hở.

Những lỗ hổng bảo mật được cho là "nghiêm trọng" bên trong thuật toán SHA-0 và SHA-1, tuỳ thuộc vào mức độ chi tiết của phần trình bày, có thể làm chấn động cả ngành bảo mật. Từ trước tới nay, SHA-1 vẫn được coi là chuẩn mực "vàng" về thuật toán. Nó được tích hợp bên trong rất nhiều chương trình thông dụng như PGP và SSL, được chứng thực bởi Viện Chuẩn Công nghệ Quốc gia và là thuật toán chữ ký điện tử duy nhất được Cơ quan Chuẩn Chữ ký Số của chính phủ Mỹ phê chuẩn.

Cả ba thuật toán MD5, SHA-0 và SHA-1 đều được giới khoa học máy tính coi là "đa chức năng". Chúng có thể nhận mọi dạng dữ liệu đầu vào, từ tin nhắn email cho đến hạt nhân (kernel) của hệ điều hành, cũng như tạo ra một dấu vân tay số duy nhất. Chỉ thay đổi một ký tự bất kỳ bên trong file đầu vào cũng tạo ra những dấu vân tay hoàn toàn khác nhau. Các ứng dụng bảo mật đều dựa vào tính năng "dấu vân tay duy nhất" này làm nền. Tuy nhiên, nếu kẻ tấn công có thể tạo ra một

dấu vân tay "Dolly" với một dòng dữ liệu đầu vào khác, dấu vân tay "sinh sản vô tính" này sẽ khiến phần mềm bị gài backdoor nhận dạng nhầm. Kết quả là chúng có thể tạo ra chữ ký giả để vét sạch tài khoản ngân hàng của người sử dụng không may.

Tất nhiên, từ rất lâu, giới nghiên cứu đã hiểu rằng không có thuật toán mã hoá thực tiễn nào là tuyệt đối an toàn và bảo mật. Tuy vậy, họ vẫn nỗ lực thiết kế ra những thuật toán mà thời gian cần để tạo ra một dấu vân tay "Dolly" là vô tận, với hy vọng kẻ tấn công sẽ nản lòng. Thế nhưng nếu những sơ hở tương tự như của SHA-0 cũng được tìm thấy trong SHA-1, điều này đồng nghĩa với việc ***tốc độ giả mạo một dấu vân tay sẽ được đẩy nhanh lên... 500 triệu lần***, ***hoàn toàn trong tầm tay của một mạng máy tính tốc độ cao***.

Tuy mức độ tác hại ít trầm trọng hơn song sơ hở bảo mật trong thuật toán MD5 có lẽ lại gây hậu quả ngay tức thì. Sản phẩm máy chủ Apache Web nguồn mở đang sử dụng MD5 để kiểm duyệt những website có mã nguồn chưa bị chỉnh sửa, từ đó sẽ an toàn khi chạy trong máy. Tương tự, sẽ là cơ sở dữ liệu Solaris của Sun Microsystems, với khả năng mà theo hãng tự nhận là "xác minh một file đích thực chứ không phải phiên bản bị điều chỉnh để hạ gục hệ thông bảo mật".

Lỗ hổng mới phát hiện trong MD5 sẽ cho phép kẻ tấn công tạo ra file giả mạo chỉ trong vài giờ với một máy tính đạt chuẩn. "Giờ đây, người ta đã chứng minh được các thuật toán này có lỗ hổng. Trước khi kẻ tấn công lợi dụng khai thác được, đã đến lúc phải thôi dần việc sử dụng MD5." - nhà phân tích Hughes của Viện Chuẩn Công nghệ Quốc gia nhận định.