ĐẢM BẢO TOÀN VẸN DỮ LIỆU TRONG HOẠT ĐỘNG THƯ VIỆN ĐIỆN TỬ

ThS Nguyễn Văn Hiệp, ThS Nguyễn Tấn Công
Trường Đại học KHXH&NV, Tp. Hồ Chí Minh

Tóm tắt: *Trình bày khái quát về an toàn bảo mật thông tin, an toàn bảo mật thông tin trong hoạt động thư viện điện tử. Giới thiệu hàm băm (hash), các ứng dụng của hàm băm trong việc đảm bảo toàn vẹn dữ liệu và việc ứng dụng thuật toán này trong hoạt động thư viện điện tử.*

Từ khóa: An toàn bảo mật thông tin; thư viện điện tử.

Ensuring data integrity in e-library

Abstract: *The article provides overview of information security, information security in e-library activities. It introduces hash and its application in ensuring data integrity and in e-library activities.*

Keywords: Information security; e-library.

1. Lời nói đầu

Hoạt động thông tin-thư viện (TT-TV) Việt Nam nói chung và thư viện điện tử nói riêng đang trong quá trình phát triển mạnh mẽ, hoàn thiện và nâng cao chất lượng để hòa nhập với thế giới. Thư viện ngày một đáp ứng tốt hơn nhu cầu của người sử dụng. Ngày nay, để tham khảo một cuốn sách, giáo trình, hay một cuốn luận văn, luận án,… người sử dụng không nhất thiết phải đến trực tiếp thư viện, mà chỉ cần ngồi tại nhà hoặc bất kỳ nơi nào, thông qua kết nối Internet đã có thể sử dụng được các tài liệu này một cách nhanh chóng và dễ dàng. Để làm được điều đó các thư viện đã không ngừng nỗ lực trong việc hoàn thiện cơ sở vật chất, trang thiết bị, công nghệ hay xây dựng cho mình các bộ sưu tập tài liệu số nội sinh nhằm phục vụ tốt nhất cho người sử dụng.

Tuy nhiên, việc mở rộng khả năng truy cập cũng đưa tới nhiều thách thức đối với các cơ quan TT-TV, như: vấn đề bản quyền, quyền tác giả, vấn đề phân cấp, phân quyền người sử dụng, vấn đề kinh phí, đường truyền… đặc biệt là vấn đề đảm bảo an toàn thông tin trước nguy cơ tấn công từ các đối tượng xấu, trong đó vấn đề đảm bảo toàn vẹn dữ liệu, ngăn ngừa việc sửa đổi nội dung thông tin được đặt lên hàng đầu.

Toàn vẹn dữ liệu trong hoạt động thư viện là việc đảm bảo các file dữ liệu không bị thay đổi trong quá trình lưu trữ, chuyển giao và sử dụng. Để đảm bảo dữ liệu được toàn vẹn người ta có thể sử dụng nhiều cách khác nhau, trong đó cách làm được coi là tối ưu nhất là sử dụng các mô hình mật mã một chiều (One Way Hash cripto system) để tính giá trị băm (Message

Digest-MD) của văn bản. Giá trị băm này được coi như "dấu tay" của văn bản. Mỗi văn bản có một MD duy nhất. Hai văn bản khác nhau dù chỉ 1 bit cũng sẽ cho hai MD khác nhau. Dựa vào tính chất này người ta dùng MD để kiểm tra tính toàn vẹn của dữ liệu. Tính toàn vẹn là một thuộc tính rất quan trọng đối với hệ thống thông tin (HTTT) nói chung và HTTT thư viện nói riêng. Nó đảm bảo tính chính xác, không thay đổi của dữ liệu trong mọi tình huống.

Đây là cách làm không mới đối với rất nhiều lĩnh vực khác nhau trong cuộc sống ngày nay như giao dịch điện tử, ngân hàng, quân sự,… Tuy nhiên. nó còn khá mới mẻ trong hoạt động của các cơ quan TT-TV, và trong bối cảnh hiện nay, việc áp dụng hàm băm (Hash) trong hoạt động các thư viện điện tử là việc làm vô cùng cần thiết.

2. An toàn thông tin trong thư viện điện tử

2.1. Khái niệm an toàn thông tin

Công nghệ thông tin và truyền thông ngày càng phát triển, những khái niệm như: an ninh mạng, bảo mật, an toàn thông tin,… không còn xa lạ đối với mỗi chúng ta. An toàn thông tin (ATTT) giờ đây không chỉ còn là mối quan tâm của các công ty, tổ chức liên quan đến tài chính, ngân hàng mà nó cũng là mối quan tâm của các thư viện. Đặc biệt là các thư viện điện tử, thư viện số nơi mà các hoạt động thư viện đang dần được tự động hóa, mục lục truyền thống được thay thế bằng mục lục điện tử, cùng với đó là các dịch vụ trực tuyến dựa trên web được cung cấp cho người sử dụng.

Thông tin có thể tồn tại dưới nhiều dạng. Thông tin có thể được in hoặc được viết trên giấy, được lưu trữ dưới dạng điện tử như các thư viện hiện nay đang thực hiện, được trình diễn trên các bộ phim, hoặc được nói trên các cuộc đàm thoại. Nhưng cho dù tồn tại dưới dạng nào đi chăng nữa, thông tin được đưa ra với hai mục đích chính là chia sẻ và lưu trữ, nó luôn cần sự bảo vệ thích hợp. Vậy an toàn thông tin là gì?

An toàn thông tin bao gồm các hoạt động quản lý, nghiệp vụ và kỹ thuật đối với HTTT nhằm bảo vệ, khôi phục các hệ thống, các dịch vụ và nội dung thông tin đối với nguy cơ tự nhiên hoặc do con người gây ra. Việc bảo vệ thông tin, tài sản và con người trong HTTT nhằm bảo đảm cho các hệ thống thực hiện đúng chức năng, phục vụ đúng đối tượng một cách sẵn sàng, chính xác và tin cậy. An toàn thông tin bao hàm các nội dung bảo vệ và bảo mật thông tin, an toàn dữ liệu, an toàn máy tính và an toàn mạng [1].

Theo ISO 17799/27001 [6], ATTT là khả năng bảo vệ đối với môi trường thông tin kinh tế xã hội, đảm bảo cho việc hình thành, sử dụng và phát triển vì lợi ích của mọi công dân, mọi tổ chức và của quốc gia. Thông qua các chính sách về ATTT, lãnh đạo thể hiện ý chí và năng lực của mình trong việc quản lý HTTT. ATTT được xây dựng trên nền tảng một hệ thống các chính sách, quy tắc, quy trình và các giải pháp kỹ thuật nhằm mục đích đảm bảo an toàn tài nguyên thông tin mà tổ chức đó sở hữu cũng như các tài nguyên thông tin của các đối tác, các khách hàng trong một môi trường thông tin toàn cầu.

An toàn thông tin là sự duy trì tính bảo mật, tính toàn vẹn và tính sẵn sàng của thông tin; ngoài ra còn có thể bao hàm một

số tính chất khác như tính xác thực, kiểm soát được, không từ chối và tin cậy [3].

2.2. Các mục tiêu cơ bản của an toàn thông tin trong thư viện điện tử

Như chúng ta đều biết, đối với các cơ quan TT-TV, thông tin/dữ liệu đóng một vai trò hết sức quan trọng, chúng ảnh hưởng trực tiếp tới sự tồn tại và phát triển của các thư viện. Vì vậy, việc bảo mật những thông tin và dữ liệu là điều vô cùng cần thiết, nhất là trong bối cảnh hiện nay khi các HTTT thư viện ngày càng được mở rộng về quy mô và khả năng truy cập.

Khi phân tích một hệ thống bảo mật, chúng ta cần xuất phát từ những tính chất cơ bản của ATTT. Có vùng dữ liệu yêu cầu *tính bảo mật* của thông tin, có vùng dữ liệu cần *tính toàn vẹn*, tất cả các dữ liệu đó đều phải được đáp ứng khi yêu cầu đó là *tính sẵn sàng* của hệ thống. Trong đó:

- Tính bảo mật (confidentiality): đảm bảo thông tin chỉ được truy cập bởi người dùng hợp pháp. Giảm thiểu tối đa mọi hành vi ăn cắp, khai thác thông tin bất hợp pháp.

- Tính sẵn sàng (availability): đảm bảo những người dùng hợp pháp mới được truy cập các thông tin và tài sản liên quan khi có yêu cầu. Hệ thống cần được sẵn sàng phục vụ và đứng vững trước mọi rủi ro khách quan và chủ quan [7].

- Tính toàn vẹn (integrity): bảo vệ tính chính xác, đầy đủ của thông tin cũng như các phương pháp xử lý. Ngăn ngừa các hành vi sửa đổi, giả mạo thông tin

2.2.1. Tính bảo mật

Dữ liệu trong HTTT của thư viện điện tử rất đa dạng, chúng khác nhau về nội dung, mục đích và đối tượng sử dụng. Xét về khía cạnh ATTT, các thông tin trên cần được phân cấp theo mức độ bảo mật, như: thông tin dùng chung (public), các thông tin dùng riêng cho một số đối tượng (private) và thông tin mật (secret)- những thông tin, hồ sơ chưa được bạch hóa.

Để đảm bảo tính bảo mật của thông tin, ngoài việc phân cấp bảo mật thông tin, người ta sử dụng các hệ thống mật mã đối xứng và bất đối xứng để mã hóa thông tin. Tùy theo mức độ mật và môi trường sử dụng thông tin, ta có thể sử dụng các thuật toán mật mã phù hợp. Hiện nay trong các hệ điều hành, như: WINDOWS, LINUX có tích hợp sẵn các công cụ mật mã như: DES, 3DES, RSA…. Ngoài ra, ta cũng có thể xây dựng các phần mềm mật mã dùng riêng, phục vụ nhu cầu của từng thư viện.

Tính mật của thông tin được đại diện bởi quyền READ (đọc).

2.2.2. Tính sẵn sàng

Khả năng đáp ứng thông tin là điều rất quan trọng, điều này thể hiện *tính sẵn sàng* phục vụ của các dịch vụ. Khả năng đáp ứng của hệ thống chịu ảnh hưởng bởi khá nhiều thành phần như phần cứng, phần mềm hay hệ thống Backup.

Khả năng đáp ứng của hệ thống cần được tính đến dựa trên số người truy cập và mức độ quan trọng của dữ liệu.

2.2.3. Tính Toàn vẹn

Tính toàn vẹn dữ liệu trước hết liên quan đến an ninh vật lý. Nếu các thiết bị vật lý này bị hư hỏng thì tính toàn vẹn của thư viện điện tử sẽ bị phá hủy. Tiếp theo là việc đảm bảo an toàn phần mềm, ta biết rằng hoạt động của HTTT thư viện đều được xây

dựng trên các phần mềm hệ thống, nếu hệ thống bị nhiễm virus, tài liệu bị phá hủy thì tính toàn vẹn của HTTT thư viện điện tử cũng bị phá hủy. Bên cạnh đó, cần quan tâm tới các vấn đề phòng chống sự phá hoại của các hacker đối với hệ thống và trong bài viết này, tác giả chỉ đề cập tới việc chống lại sự phá hoại của các hacker đối với các dữ liệu trong thư viện điện tử.

Tính toàn vẹn là một thuộc tính rất quan trọng đối với các HTTT nói chung và HTTT thư viện nói riêng. Nó đảm bảo tính chính xác, không thay đổi của dữ liệu trong mọi tình huống.

Tính toàn vẹn của thông tin được đại diện đặc trưng bởi quyền sửa đổi. Và như đã nói ở phần trên, để đảm bảo tính toàn vẹn phương pháp được sử dụng phổ biến nhất đó là sử dụng các mô hình mật mã một chiều (One Way Hash cripto system) để tính MD của văn bản.

3. Hàm băm và ứng dụng

3.1. Khái niệm hàm băm

Hàm băm (Hash) là thuật toán dùng để "tóm tắt" (băm) tài liệu, bản tin hoặc thông điệp và cho kết quả là một giá trị "băm" có kích thước cố định. Giá trị băm này còn được gọi là "đại diện tài liệu", "đại diện bản tin" hay "đại diện thông điệp" [6].

Hàm băm là hàm một chiều mà nếu đưa một lượng dữ liệu bất kỳ qua hàm này sẽ cho ra một chuỗi có độ dài cố định ở đầu ra.

3.2. Tính chất cơ bản của hàm băm

- *Tính một chiều*: không thể suy ra dữ liệu ban đầu từ kết quả, điều này tương tự như việc bạn không thể chỉ dựa vào một dấu vân tay lạ mà suy ra ai là chủ của nó được.
- *Tính duy nhất*: xác suất để có một vụ va chạm (hash collision), tức là hai thông điệp khác nhau có cùng một kết quả hash là cực kỳ nhỏ.

3.3. Các ứng dụng của hàm băm

- *Xác thực mật khẩu:* Mật khẩu thường không được lưu dưới dạng văn bản rõ (clear text), mà ở dạng tóm tắt. Để xác thực một người dùng, mật khẩu do người đó nhập vào được băm ra bằng hàm Hash và so sánh với kết quả băm được lưu trữ.
- *Xác thực thông điệp (Message authentication-Thông điệp tóm tắt-message digests):* Giá trị đầu vào (tin nhắn, dữ liệu...) bị thay đổi tương ứng giá trị băm cũng bị thay đổi. Do vậy, nếu một kẻ tấn công phá hoại, chỉnh sửa dữ liệu thì server có thể biết ngay lập tức.
- *Bảo vệ tính toàn vẹn của tập tin, thông điệp được gửi qua mạng:* Hàm băm mật mã có tính chất là hàm một chiều. Từ khối dữ liệu hay giá trị đầu vào chỉ có thể đưa ra một giá trị băm duy nhất. Như chúng ta đã biết đối với tính chất của hàm một chiều, một người nào đó dù bắt được giá trị băm họ cũng không thể suy ngược lại giá trị, đoạn tin nhắn băm khởi điểm. Ví dụ, việc xác định xem một file hay một thông điệp có bị sửa đổi hay không có thể thực hiện bằng cách so sánh tóm tắt được tính trước và sau khi gửi (hoặc một sự kiện bất kỳ nào đó), hoặc có thể dùng tóm tắt thông

điệp làm một phương tiện đáng tin cậy cho việc nhận dạng file. Hàm băm thường được dùng trong bảng băm nhằm giảm chi phí tính toán khi tìm một khối dữ liệu trong một tập hợp. Giá trị băm đóng vai trò gần như một khóa để phân biệt các khối dữ liệu.

- *Tạo chữ ký điện tử (Digital signatures):* Chữ ký số có được bằng cách đem mã hoá bản tóm tắt của thông điệp bằng khoá bí mật của người ký.

Nếu kết quả băm giống nhau, Thông điệp được xác thực. Tại sao? Vì nếu bất kỳ BIT nào của M hay SIG bị thay đổi, kết quả băm sẽ khác.

3.4. Ứng dụng hàm băm đảm bảo toàn vẹn dữ liệu trong hoạt động các thư viện điện tử

Thư viện thế kỷ XXI không chỉ là một trung tâm tri thức, mà còn trở thành một trung tâm thông tin, ở đó không chỉ có sách, báo, tạp chí in trên giấy mà còn có các xuất bản phẩm dưới dạng điện tử. Vì vậy, hình thức tổ chức và phương pháp hoạt động của các thư viện cũng có nhiều thay đổi; trong đó việc sử dụng máy tính để lưu giữ, khai thác thông tin và xây dựng các bộ sưu tập số là xu hướng quan trọng nhất trong việc phát triển tự động hoá các thư viện. Không những thế, để đáp ứng tốt nhất nhu cầu ngày một cao của người dùng, các thư viện đã sử dụng mạng Internet và "mở kết nối" nhằm tạo điều kiện tối đa cho người sử dụng có thể truy cập từ xa tới các nguồn dữ liệu này.

Những thay đổi trong phương thức phục vụ trên đem lại rất nhiều lợi ích cho người sử dụng cũng như nâng cao vị thế của thư viện trong mắt người dùng và xã hội. Tuy nhiên, việc mở rộng khả năng truy cập tới các nguồn tài nguyên đó lại đem tới nhiều rủi ro cho các thư viện, như: tăng khả năng bị hacker tấn công, dễ dẫn tới mất mát dữ liệu,… đặc biệt các đối tượng xấu có thể lợi dụng thư viện làm nơi để thực hiện các âm mưu đen tối của mình.

Đảm bảo toàn vẹn dữ liệu trong các thư viện điện tử bao gồm rất nhiều công việc khác nhau, đó là đảm bảo toàn vẹn trong việc vận hành hệ thống quản lý thư viện, đảm bảo việc lưu trữ mật khẩu, thông tin hệ thống đều được giữ bí mật và đảm bảo các tài liệu điện tử trong các bộ sưu tập số luôn luôn toàn vẹn trong khi lưu trữ và truyền trên mạng Internet.

Với việc sử dụng hàm băm, các thư viện có thể xác định xem một file hay một thông điệp có bị sửa đổi hay không và thực hiện bằng cách so sánh tóm tắt được tính trước và sau khi gửi (hoặc một sự kiện bất kỳ nào đó). Hoặc có thể dùng tóm tắt thông điệp làm một phương tiện đáng tin cậy cho việc nhận dạng file. Một ứng dụng nữa các thư viện có thể áp dụng là kiểm tra mật khẩu như đã trình bày ở phần trên. Các hàm băm có thể được dùng để tạo các bit giả ngẫu nhiên (pseudorandom) có thể kể tới như: SHA-1, MD5,…

3.4.1. Ứng dụng trong lưu trữ mật khẩu các tài khoản trong hệ thống thư viện

Hầu hết các phần mềm quản lý thư viện ngày nay đều có chứng thực người sử dụng. Nghĩa là để sử dụng ứng dụng, người sử dụng phải qua một cơ chế chứng thực username và mật khẩu, và từ đó được cung cấp các quyền sử dụng tương

ứng đã được thư viện cung cấp. Do đó, vấn đề bảo mật mật khẩu là vấn đề vô cùng quan trọng đối với các thư viện điện tử.

Mật khẩu người sử dụng thường gồm các chữ cái thường và hoa, cộng thêm các chữ số. Giả sử mật khẩu được lưu trữ dưới dạng thường, không mã hóa trên máy chủ, trong một file dữ liệu hay trong hệ quản trị thư viện tích hợp. Như vậy, sẽ xuất hiện một nguy cơ là có một người khác, hoặc là người quản trị (administrator), hoặc là hacker, có thể mở được file dữ liệu hoặc cơ sở dữ liệu, và xem trộm được mật khẩu. Do vậy, mật khẩu không thể được giữ bí mật tuyệt đối.

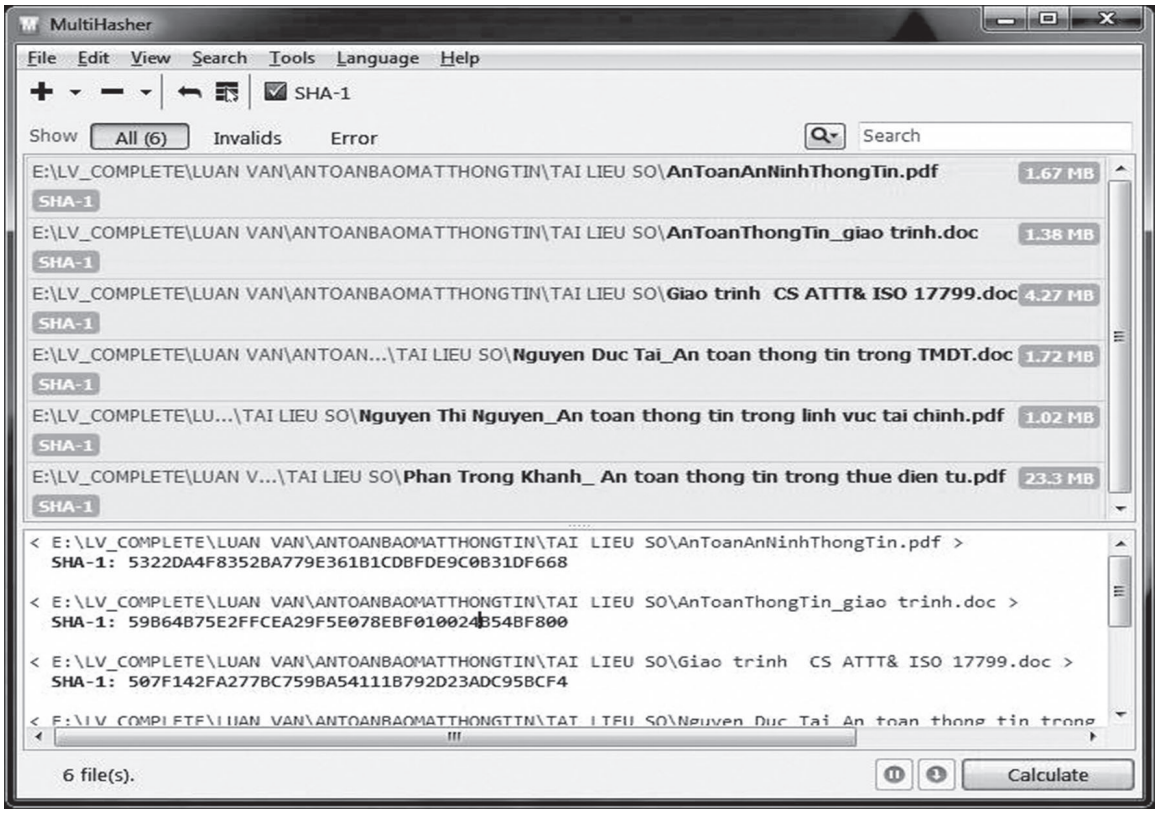
Một phương pháp để bảo vệ mật khẩu là dùng mã hóa, chương trình phần mềm sẽ dùng một khóa bí mật để mã hóa mật khẩu trước khi lưu mật khẩu xuống file hay cơ sở dữ liệu. Do đó, tránh được vấn đề xem trộm mật khẩu. Tuy nhiên, phương pháp này có nhược điểm là lại phải lo bảo vệ khóa bí mật này. Nếu khóa bí mật bị lộ thì việc mã hóa không còn ý nghĩa.

Phương pháp bảo vệ mật khẩu hiệu quả nhất là dùng hàm băm. Khi người sử dụng đăng ký mật khẩu, giá trị băm của mật khẩu được tính bằng một hàm băm nào đó (MD5 hay SHA-1,…). Giá trị băm được lưu trữ vào file hay cơ sở dữ liệu. Vì hàm băm là một chiều, nên dù biết được giá trị băm và loại hàm băm, hacker cũng không thể suy ra được mật khẩu. Khi người sử dụng đăng nhập, mật khẩu đăng nhập được tính giá trị băm và so sánh với giá trị băm đang được lưu trữ. Do tính chống trùng, chỉ có một mật khẩu duy nhất có giá trị băm tương ứng, nên không ai khác ngoài người sử dụng có mật khẩu đó mới có thể đăng nhập ứng dụng.

3.4.2. Ứng dụng hàm băm để kiểm tra sự toàn vẹn của dữ liệu lưu trữ trong các bộ sưu tập số

Hiện nay, các thư viện đang xây dựng cho mình các bộ sưu tập số và cung cấp khả năng truy cập từ xa tới các bộ sưu tập này cho người sử dụng. Tuy nhiên, câu hỏi đặt ra là làm cách nào thư viện có thể kiểm tra được rằng, các tài liệu số do mình lưu trữ và cung cấp cho người sử dụng có bị thay đổi nội dung hay không? Đây thật sự là một bài toán khó đối với các thư viện điện tử. Một số cách các thư viện hiện nay đang thực hiện đó là: đặt mật khẩu cho các tài liệu điện tử, để tài liệu ở chế độ chỉ xem,… Tuy nhiên, tất cả các cách trên đều tồn tại nhược điểm và cũng rất dễ để hacker có thể phá bỏ hoặc tìm ra mật khẩu của các tập tin đó.

Một cách đơn giản để các thư viện có thể đảm bảo được tính toàn vẹn đối với các dữ liệu số của mình đó là sử dụng hàm băm. Như ta đã biết, hai văn bản dù chỉ khác nhau một ký tự thì cũng cho ta hai giá trị băm khác nhau, do đó, trước khi lưu trữ hoặc gửi dữ liệu cho người sử dụng, cán bộ thư viện sẽ tiến hành quá trình băm và gắn giá trị băm tìm được vào tài liệu, người sử dụng sau khi nhận được tài liệu sẽ dùng một phần mềm bất kỳ để kiểm tra giá trị băm từ tài liệu nhận được, so sánh hai giá trị này, nếu trùng khớp thì tài liệu không thay đổi, ngược lại, nếu không trùng khớp tức là tài liệu đã bị thay đổi so với nội dung ban đầu.



Tạo ra giá trị SHA - 1 với phần mềm Mutihasher

Kết luận

Mở rộng khả năng truy cập từ xa đến thư viện nói chung, đến các bộ sưu tập số và dịch vụ thư viện nói riêng là xu thế tất yếu trong hoạt động của các thư viện ngày nay. Đây là việc làm cần thiết đối với các thư viện với mong muốn ngày một đáp ứng tốt hơn nhu cầu của người sử dụng. Tuy nhiên, để đảm bảo hoạt động thư viện luôn được ổn định và xuyên suốt, các thư viện nên có lộ trình thực hiện cụ thể và có những bước chuẩn bị kỹ càng cả về trình độ của cán bộ thư viện, cơ sở vật chất cũng như các giải pháp công nghệ về an toàn bảo mật. Các giải pháp nên bắt đầu từ việc xây dựng chính sách an toàn bảo mật, tới việc trang bị về máy móc, trang thiết bị và cuối cùng là các giải pháp công nghệ. Việc sử dụng hàm băm là một giải pháp tối ưu cho các thư viện trong việc xây dựng các giải pháp công nghệ nhằm đảm bảo an toàn, toàn vẹn dữ liệu cho hoạt động của mình.

TÀI LIỆU THAM KHẢO

1. Cryptography and Network Security Principles and Practices, 4th Edition - William Stallings- Prentice Hall - 2005.

2. Nguyễn Đình Vinh (2011). *Giáo trình cơ sở an toàn thông tin*. H. Ban Cơ yếu chính phủ, Học viện Kỹ thuật Mật mã, 2011.

3. Phan Đình Diệu (2002). *Lý thuyết mật mã & An toàn thông tin, Đại học quốc gia Hà Nội*, Hà Nội, 2002.

4. Trần Minh Văn (2008). *Giáo trình An toàn và bảo mật thông tin*, Đại học Nha Trang, Nha Trang.

5. Trịnh Nhật Tiến (2008). *Giáo trình An toàn dữ liệu*, NXB ĐHQGHN. H. 2008.

6. Truy cập từ: http://vnexperts.net/bai-viet-ky-thuat/security/661-chun-bo-mt-iso-17799-toan-tp.html(ngày 24/10/2016)

(Ngày Tòa soạn nhận được bài: 4-5-2017; Ngày phản biện đánh giá: 12-10-2017; Ngày chấp nhận đăng: 20-10-2017).