Kỷ yếu Hội nghị Khoa học Quốc gia lần thứ IX "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR'9)"; Cần Thơ, ngày 4-5/8/2016
DOI: 10.15625/vap.2016.00063

NGHIÊN CỨU VỀ CÁC CƠ CHẾ RAID VÀ ĐỀ XUẤT GIẢI PHÁP LƯU TRỮ DỮ LIỆU AN TOÀN TRÊN DỊCH VỤ ĐÁM MÂY

Lê Quang Minh[1], Nguyễn Anh Chuyên[2], Lê Khánh Dương[2], Phan Huy Anh[1], Trịnh Thị Thu[3]

[1] Viện Công nghệ thông tin, Đại học Quốc gia Hà Nội
[2] Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên
[3] Trường Đại học Nông lâm, Đại học Thái Nguyên

quangminh@vnu.edu.vn, nachuyen@ictu.edu.vn

TÓM TẮT— *Công nghệ điện toán đám mây hiện đang phát triển và được ứng dụng rộng rãi trong việc triển khai các dịch vụ mạng và lưu trữ dữ liệu. Người dùng được hưởng lợi từ các dịch vụ lưu trữ dữ liệu trực tuyến khi đăng ký các tài khoản miễn phí, dữ liệu được lưu trên Cloud và có cơ chế đồng bộ rất tiện lợi, trên nhiều nền tảng, thiết bị. Tuy nhiên, việc bảo vệ cho dữ liệu quan trọng này không bị mất mát khi dịch vụ lưu trữ xảy ra lỗi, hay tránh bị xâm phạm một cách bất hợp pháp là điều ngoài tầm kiểm soát của người dùng. Trong bài báo này, nhóm nghiên cứu đề xuất một giải pháp lưu trữ an toàn cho dữ liệu người dùng dựa trên cơ chế RAID, nhằm khắc phục các nguy cơ ở trên, đồng thời vẫn có thể sử dụng các tài khoản lưu trữ từ những dịch vụ miễn phí.*

Từ khóa— *Độ tin cậy, điện toán đám mây, lưu trữ RAID, Cloud-RAID, RBCS.*

I. GIỚI THIỆU

Điện toán đám mây là giải pháp công nghệ đang phát triển và ứng dụng rộng rãi trong việc lưu trữ, xử lý và truy cập dữ liệu từ xa trên môi trường internet. Theo NIST (Viện nghiên cứu Tiêu chuẩn và Công nghệ quốc gia Hoa Kỳ), "Điện toán đám mây là mô hình điện toán cho phép truy cập qua mạng để lựa chọn và sử dụng tài nguyên tính toán theo nhu cầu một cách thuận tiện và nhanh chóng; đồng thời cho phép kết thúc sử dụng dịch vụ, giải phóng tài nguyên dễ dàng, giảm thiểu các giao tiếp với nhà cung cấp". Với sự linh hoạt chính là khả năng phân phát tài nguyên theo yêu cầu của người dùng, tạo điều kiện thuận lợi cho việc sử dụng một cách hữu ích tài nguyên tích luỹ của hệ thống. Đồng thời cung cấp khả năng tính toán không giới hạn theo yêu cầu cho người sử dụng, mà không đòi hỏi đầu tư vốn lớn để đáp ứng nhu cầu của họ. Bên cạnh đó người sử dụng cũng có thể truy cập tới dữ liệu của họ từ bất cứ nơi nào thông qua các kết nối internet. Mặc dù, những lợi thế của điện toán đám mây là rất hấp dẫn, tuy nhiên có một số vấn đề liên quan đến an ninh đặc biệt về an toàn và bảo mật dữ liệu [1].

Ứng dụng phổ biến hiện nay của công nghệ điện toán đám mây đó là các dịch vụ lưu trữ dữ liệu công cộng, cho phép người dùng đăng ký các tài khoản để sử dụng. Một số dịch vụ lưu trữ phổ biến như: GDrive, Dropbox, Box, OneDrive, ICloud,... cho phép người dùng đăng ký tài khoản bằng địa chỉ email cá nhân. Dung lượng lưu trữ trên mỗi dịch vụ là khác nhau, có thể từ 2GB đến 50GB. Dữ liệu lưu trữ trên các dịch vụ đó có cơ chế đồng bộ, được thực hiện trên nhiều nền tảng như: ứng dụng Web, Mobile, Desktop. Tuy nhiên, vấn đề an toàn cho dữ liệu của người dùng chưa thực sự được đảm bảo, các nguy cơ như tài khoản bị đánh cắp, quên mật khẩu, nhà cung cấp ngừng dịch vụ, hệ thống bị tấn công, hacker có thể truy cập vào dữ liệu của người dùng, thậm chí dữ liệu nhạy cảm của người dùng bị truy cập bất hợp pháp từ chính nhà cung cấp do yêu cầu cung cấp thông tin từ phía chính phủ,… Trong nội dung bài báo này, nhóm nghiên cứu đề xuất một phương án lưu trữ dữ liệu an toàn và bảo mật trên các dịch vụ đám mây.

II. CƠ CHẾ RAID VÀ MỘT SỐ VẤN ĐỀ LƯU TRỮ DỮ LIỆU ĐÁM MÂY

A. Đặc điểm và vai trò của RAID

1. Định nghĩa và vai trò của RAID

RAID (viết tắt của *Redundant Array of Independent Disks*) là giải pháp lưu trữ dữ liệu sử dụng loạt các ổ đĩa cứng vật lý được ghép lại với nhau thành một hệ thống có chức năng tăng tốc độ truy xuất dữ liệu hoặc bổ sung cơ chế sao lưu, dự phòng dữ liệu cho hệ thống. RAID cho phép lưu trữ dữ liệu giống nhau ở những nơi khác nhau trên nhiều đĩa, do đó thao tác đọc/ghi có thể chồng lên nhau một cách cân bằng, nhằm cải thiện hiệu suất và tăng cường khả năng bảo vệ dữ liệu [16]. Hiện nay, cơ chế lưu trữ RAID có thể được triển khai ở 2 dạng:

- RAID cứng: Thường dùng cho các máy chủ sử dụng một thiết bị phần cứng gọi là RAID Controller card để điều khiển cơ chế đọc/ghi dữ liệu trên các ổ cứng. Card RAID này hoạt động như một máy tính chuyên dụng và được tích hợp trên máy chủ, cung cấp hiệu suất hoạt động cao, tuy nhiên đòi hỏi các ổ cứng vật lý phải có thông số như nhau và cấu hình phức tạp.
- RAID mềm: Dùng cho các máy tính yêu cầu nâng cao hiệu năng với chi phí thấp. Loại RAID này do hệ điều hành điều khiển nên hiệu suất hoạt động không cao, sử dụng chính các phân vùng của các ổ đĩa vật lý trên hệ thống, cấu hình loại này đơn giản hơn.

2. Một số loại RAID được dùng phổ biến

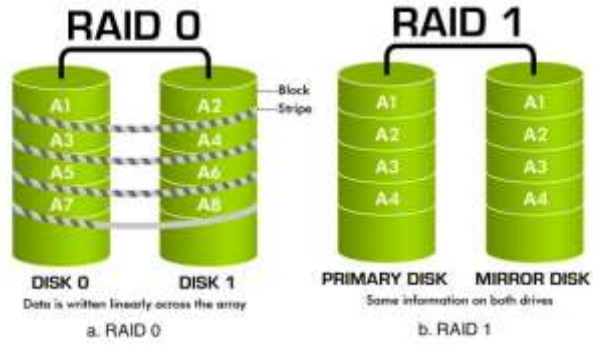
Sự ra đời của RAID đóng vai trò rất quan trọng trong ngành công nghiệp máy chủ. Tổ chức RAB (RAID Advisory Board - Hội đồng tư vấn phát triển RAID) đã phân ra các loại cấp độ (level) RAID, các tiêu chuẩn phần cứng sử dụng RAID. Một số loại RAID thường được sử dụng hiện nay:

a) RAID 0

Cơ chế lưu trữ kiểu RAID 0 cần tối thiểu 2 ổ đĩa (n >= 2) và các đĩa là cùng loại. Dữ liệu sẽ được chia ra nhiều phần bằng nhau để lưu trên từng đĩa, như vậy mỗi đĩa sẽ chứa 1/n dữ liệu. Dung lượng tổng sẽ được tính bằng công thức:

Array Capacity = Size of Smallest Drive * Number of Drives

Ưu điểm của cơ chế lưu trữ RAID 0 này là tăng tốc độ đọc/ghi đĩa, do mỗi đĩa chỉ cần phải đọc/ghi một lượng 1/n tổng dữ liệu được yêu cầu nên trên lý thuyết thì tốc độ sẽ tăng n lần. Tuy nhiên, nhược điểm đối với cơ chế này là tính an toàn thấp, do dữ liệu được phân mảnh để lưu trữ nên trong trường hợp nếu một đĩa bị hỏng thì dữ liệu trên tất cả các đĩa còn lại sẽ không sử dụng được. Xác suất hỏng của hệ thống sẽ tăng n lần so với dùng ổ đĩa đơn [16].



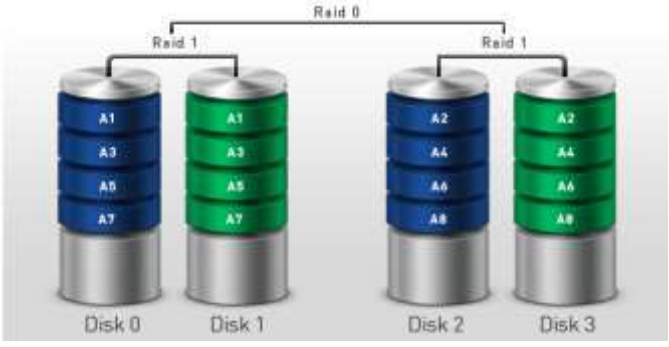
Hình 1. Cơ chế lưu trữ kiểu RAID 0 và RAID 1 (nguồn vinahost.vn)

b) RAID 1

Cơ chế RAID 1 cải thiện vấn đề đảm bảo an toàn dữ liệu hơn so với RAID 0, yêu cầu ít nhất hai đĩa cứng để làm việc. Trong quá trình lưu trữ, dữ liệu được ghi vào 2 ổ giống hệt nhau (cơ chế Mirroring) [16]. Khi một ổ đĩa bị mất dữ liệu, toàn bộ dữ liệu vẫn có thể khôi phục từ ổ còn lại. Đối với những hệ thống cần lưu trữ và quản lý nhiều thông tin quan trọng thì hệ thống RAID 1 là thứ không thể thiếu mặc dù hiệu năng không phải là yếu tố hàng đầu. Dung lượng cuối cùng của hệ thống RAID 1 bằng dung lượng của ổ đơn (Ví dụ với hai ổ 80GB chạy RAID 1 sẽ cho hệ thống nhìn thấy duy nhất một ổ RAID 80GB).

c) RAID 10

RAID 10 là sự kết hợp giữa 2 loại RAID 1 và RAID 0, để thiết lập cơ chế này cần tối thiểu 4 ổ đĩa cứng như Hình 2. Đối với RAID 10 dữ liệu sẽ được lưu đồng thời vào 4 ổ cứng, trong đó 2 ổ dạng Striping (Raid 0) và 2 ổ (Mirroring) RAID 1. Đặc điểm của cơ chế này tốc độ lưu trữ dữ liệu nhanh và an toàn, vừa nâng cao hiệu suất hoạt động mà có thể đảm bảo tính dự phòng cho dữ liệu khi 1 trong số 4 ổ cứng bị hỏng. Tuy nhiên nhược điểm của cơ chế này là chi phí đầu tư cao do dung lượng sẵn sàng sử dụng chỉ bằng ½ dung lượng của 4 ổ (giống như RAID 1) [12].



Hình 2. Cơ chế lưu trữ kiểu RAID 10 (nguồn vinahost.vn)

Ngoài ra, RAB còn giới thiệu một số cơ chế RAID khác như: RAID 3, RAID 4, RAID 5 RAID 6 với cách thức lưu trữ dữ liệu khác nhau, tuy nhiên đa phần đều dựa trên 2 cơ chế lưu trữ cơ bản là RAID 0 và RAID 1. Bên cạnh đó, với việc kết hợp đặc điểm của 2 hay nhiều loại RAID khác nhau lại để hình thành cơ chế mới, gọi là *Hybrid RAID*. Một số cơ chế RAID thuộc dạng này như: RAID 01, RAID 100, RAID 50, RAID 60,…

B. Một số vấn đề về an toàn dữ liệu trong lưu trữ trên Cloud

Thách thức lớn nhất trong việc triển khai thành công giải pháp dựa trên công nghệ điện toán đám mây chính là đảm bảo về vấn đề an ninh cho hệ thống. Khi các ứng dụng được cài đặt và chạy trên tài nguyên của máy ảo, hay khi dữ liệu quan trọng của người dùng được di chuyển và lưu trữ trên các kho dữ liệu đám mây, sẽ có rất nhiều vấn đề về an ninh và an toàn dữ liệu xảy ra. [1]

Theo một thống kê trên trang cnet.com, hàng loạt dịch vụ lưu trữ dữ liệu trực tuyến với hàng triệu tài khoản đang hoạt động có thể đã bị khai thác và hacker đã truy cập vào dữ liệu cá nhân của người dùng một cách bất hợp pháp. Dịch vụ Dropbox đã bị hacker tấn công và lấy cắp thông tin đăng nhập của hơn 7 triệu tài khoản người dùng, các thông tin nhạy cảm của một số tài khoản bị yêu cầu nộp tiền chuộc qua Bitcoin. Cùng với đó là sự đe doạ các dữ liệu cá nhân như: ảnh, video, tài liệu,… trên các tài khoản Dropbox của người dùng có thể bị công khai trên mạng [15].

Tháng 5/2014, một công ty về công nghệ Intralinks phát hiện ra lỗ hổng bảo mật trên dịch vụ lưu trữ dữ liệu của Box và Dropbox cho phép dữ liệu cá nhân để được đọc bởi các bên thứ ba hoặc được index bởi công cụ tìm kiếm. Intralinks phát hiện ra rằng nếu người dùng chia sẻ file qua các liên kết URL và các URL này được dán vào hộp tìm kiếm của trình duyệt thay vì thanh URL, các liên kết có thể sau đó được lập chỉ mục của công cụ tìm kiếm và có thể được đọc bởi các bên thứ ba. Từ đó họ cũng khuyến cáo người dùng nên sử dụng một dịch vụ mã hóa bên thứ 3 để bảo vệ các dữ liệu trên dịch vụ lưu trữ đám mây.

Một dịch vụ lưu trữ đám mây khác cũng rất phổ biến là GDrive của Google, các tài khoản Gmail đều được cung cấp kho lưu trữ với dung lượng 10GB trên GDrive. Tháng 7/2014, dịch vụ GDrive cũng bị thông báo có lỗ hổng về bảo mật liên quan tới việc chia sẻ các liên kết trên GDrive giống như của Dropbox [8,15].

Theo Lucas Mearian, trong bài phân tích của mình về vấn đề bảo mật trên các dịch vụ lưu trữ đám mây, tác giả đã đưa ra các dẫn chứng cho thấy dữ liệu của người dùng có nguy cơ rất cao bị xâm nhập bất hợp pháp. Trong năm 2012, Google nhận được hơn 21.000 yêu cầu từ phía chính phủ về việc cung cấp thông tin của hơn 33.000 tài khoản người dùng [11]. Các công ty công nghệ khác như Microsoft cũng nhận được hơn 70.000 yêu cầu về 122.000 tài khoản người dùng trên hệ thống lưu trữ của công ty. Một dẫn chứng nữa cho thấy dữ liệu riêng tư của người dùng có thể bị truy cập, hệ thống iMessage hay iCloud của Apple cho phép người dùng lưu trữ dữ liệu cá nhân và tin nhắn, từ đó đồng bộ trên các thiết bị như Iphone, Ipad, Macbook,... Tuy nhiên, hệ thống này là hoàn toàn đóng và không phải mã nguồn mở, do đó các nhà nghiên cứu cũng như người dùng cũng không thể biết được lời cam đoan của nhà cung cấp dịch vụ là chính xác hay không.

Theo Monjur Ahmed, tất cả các nguy hại và hình thức tấn công được áp dụng đối với mạng máy tính và dữ liệu đều có ảnh hưởng lên các hệ thống dựa trên dịch vụ điện toán đám mây, một số mối đe dọa thường gặp như: tấn công MITM, phishing, nghe trộm, sniffing,... Ngoài ra các cuộc tấn công DDoS (Distributed Denial of Service) cũng là nguy cơ ảnh hưởng cho cơ sở hạ tầng điện toán đám mây, mặc dù không có bất kỳ ngoại lệ nào để giảm thiểu tình trạng này [3]. Do đó, sự an toàn của máy ảo sẽ xác định tính toàn vẹn và mức độ an ninh của hệ thống dựa trên điện toán đám mây. Dựa trên các nghiên cứu, Cloud Security Alliance (CSA) đã đưa ra những vấn đề có mức độ nguy hại cao nhất trong điện toán đám mây gồm [4]:

- Sử dụng bất hợp pháp dịch vụ: Kẻ tấn công sẽ khai thác lỗ hổng trên các dịch vụ public cloud để phát tán mã độc tới người dùng và lây lan ra hệ thống máy tính, từ đó khai thác sức mạnh của dịch vụ đám mây để tấn công các máy tính khác.
- API (Application Programming Interfaces) không bảo mật: Đây là giao diện lập trình phần mềm để tương tác với các dịch vụ cloud. Khi các hãng thứ 3 sử dụng các API thiếu bảo mật này để tạo các phần mềm, tài khoản và dữ liệu của người dùng có thể bị ảnh hưởng thông qua các ứng dụng đó.
- Các lỗ hổng trong chia sẻ dữ liệu: Do sử dụng cùng một nền tảng dịch vụ trên cloud, nên việc rò rỉ thông tin có thể phát sinh khi chia sẻ thông tin từ một khách hàng cho những người khác.
- Mất dữ liệu: Mất dữ liệu là một vấn đề phổ biến trong điện toán đám mây. Nếu nhà cung cấp dịch vụ điện toán đám mây buộc phải đóng dịch vụ của mình do một số vấn đề tài chính hay pháp lý, khi đó tất cả dữ liệu của người dùng sẽ bị mất.
- Tấn công luồng dữ liệu: Đây là vấn đề mà những người sử dụng dịch vụ lưu trữ cloud cần lưu ý tới, chủ yếu là các thao tác mà hacker sử dụng để tấn công như MITM, spam, tấn công từ chối dịch vụ, virus, malware,…
- Những nguy hại từ bên trong: Các mối đe dọa này bao gồm gian lận, phá hỏng dữ liệu, đánh cắp hoặc mất thông tin bí mật do chính người trong cuộc được tin tưởng gây ra. Những người này có thể có khả năng xâm nhập vào bên trong tổ chức và truy cập dữ liệu bất hợp pháp nhằm phá hoại, gây tổn thất tài chính, hiệu suất công việc, thiệt hại thương hiệu.

Từ những vấn đề liên quan tới an toàn và bảo mật cho dữ liệu của người dùng khi lưu trữ trên các dịch vụ cloud miễn phí hiện nay, kết hợp với ý tưởng của cơ chế lưu trữ của RAID trên các thiết bị đĩa cứng. Nhóm nghiên cứu chúng tôi đề xuất một phương án lưu trữ dữ liệu online kiểu RAID, nhằm tăng khả năng bảo mật cho dữ liệu được lưu trữ của người dùng trên cloud, đồng thời vẫn sử dụng được các dịch vụ lưu trữ miễn phí.

III. GIẢI PHÁP LƯU TRỮ DỮ LIỆU AN TOÀN TRÊN CLOUD – RBCS

A. Đề xuất giải pháp lưu trữ an toàn RBCS

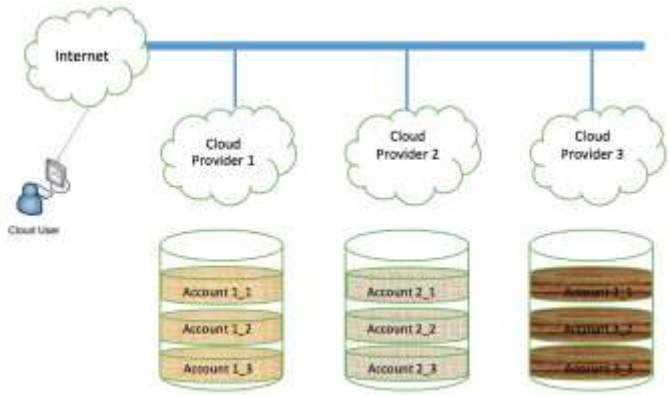
1. Giải pháp RBCS

RBCS (RAID Based Cloud Storage) là cơ chế lưu trữ dữ liệu trên các dịch vụ cloud do nhóm nghiên cứu đề xuất, sử dụng các tài khoản miễn phí của các nhà cung cấp dịch vụ như GDrive, Dropbox, Box, OneDrive,… RBCS kết hợp giữa cơ chế lưu trữ an toàn có dự phòng của RAID 0,1 đồng thời tận dụng được khả năng linh động của dịch vụ lưu trữ cloud. Khi đó, giải pháp này giải quyết được 2 vấn đề chính đối với dữ liệu được lưu trữ trên cloud đó là:

- Tính toàn vẹn: Dữ liệu được lưu trữ phân bố trên nhiều tài khoản khác nhau, không phụ thuộc hoàn toàn vào bất cứ nhà cung cấp dịch vụ lưu trữ cloud nào, do đó khả năng chịu lỗi có thể là toàn bộ các tài khoản của một nhà cung cấp dịch vụ bị mất hoặc không truy cập được. Trong trường hợp đó, dữ liệu sẽ vẫn được khôi phục dựa trên các mảnh được phân phối trên các tài khoản khác.
- Tính bảo mật: Các mảnh dữ liệu được phân chia sẽ là riêng rẽ và độc lập, ngay cả khi tài khoản bị tấn công hay bị xâm nhập bất hợp pháp từ chính nhà cung cấp dịch vụ, cũng không thể xem dữ liệu nhạy cảm của người dùng. Chỉ khi đọc dữ liệu, các mảnh ghép đó sẽ được tải về đồng bộ và khôi phục lại trên máy của người dùng.

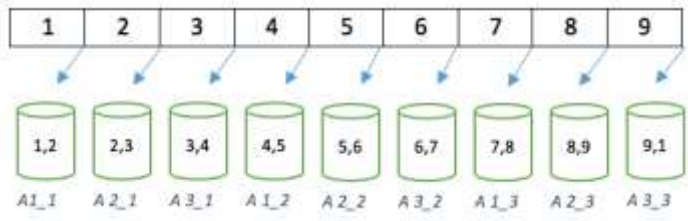
2. Cơ chế lưu trữ dữ liệu của RBCS

Giải pháp này sử dụng các tài khoản trên các nhà cung cấp dịch vụ cloud hiện nay như: Gdrive, OneDrive, Dropbox, Box,… để lưu trữ dữ liệu. Những tài khoản miễn phí này có thể được tạo ra đơn giản với địa chỉ email của người dùng. Để đảm bảo tính toàn vẹn cho dữ liệu khi lưu trữ, RBCS sẽ sử dụng tối thiểu 3 nhà cung cấp dịch vụ cloud và tối thiểu n (n>=2) tài khoản trên mỗi dịch vụ, do đó số tài khoản dùng để lưu trữ sẽ là 3*n tài khoản.



Hình 3. Cơ chế lưu trữ dữ liệu của RBCS

Quá trình lưu trữ dữ liệu trên các tài khoản cloud được thực hiện như sau: với mỗi tập tin người dùng cần lưu trữ, RBCS sẽ phân mảnh thành các phần và tiến hành lưu trữ các phần đó trên các tài khoản giống như cơ chế RAID 10. Lấy ví dụ một tập tin được phân thành 9 mảnh và sử dụng 3 tài khoản cloud trên mỗi dịch vụ (tổng có 9 tài khoản):



Hình 4. Phân mảnh dữ liệu và lưu trữ trên các kho dữ liệu cloud

Trên Hình 4, dữ liệu tập tin được lưu vào các tài khoản cloud theo quy tắc:

- Các tài khoản của cùng 1 nhà cung cấp dịch vụ được đặt xen kẽ nhau, theo quy tắc n*i+m (trong đó n là số tài khoản trên cùng 1 dịch vụ, i là số lượt, m là thứ tự tài khoản).
- Trên mỗi tài khoản sẽ lưu trữ 2 mảnh dữ liệu kề nhau theo thứ tự đã phân mảnh.
- Mảnh đầu tiên và cuối cùng sẽ được lưu trên cùng 1 tài khoản.

Với cách phân chia các mảnh vào các tài khoản và thứ tự sắp xếp các tài khoản như vậy sẽ có các ưu điểm là:

- Khi 1 tài khoản bất kì bị mất hoặc không truy cập được, dữ liệu có thể được lấy từ 2 tài khoản lân cận.

- Khi tất cả các tài khoản của cùng một nhà cung cấp dịch vụ bị mất (trường hợp này hiếm xảy ra hơn), dữ liệu của các mảnh vẫn khôi phục được từ các tài khoản khác trên các dịch vụ khác.
- Nếu 2 tài khoản liên tiếp trong danh sách bị mất dữ liệu (trường hợp này có thể xảy ra), dữ liệu không khôi phục được.
- Nếu 2 nhà cung cấp dịch vụ cùng ngừng hoạt động, dữ liệu cũng không khôi phục lại được.

Vấn đề tiếp theo là quản lý danh sách thứ tự các tài khoản khi lưu trữ và thứ tự các mảnh dữ liệu. Do thứ tự các tài khoản này có thể không cố định để tăng tính phức tạp và khó đoán khi bị hack. Hiện nay các nhà cung cấp dịch vụ thường quy định dung lượng tối đa cho mỗi tài khoản và kích thước tối đa cho mỗi tập tin khi được tải lên. Dung lượng này có thể khác nhau tuỳ từng nhà cung cấp dịch vụ cloud: Dropbox là 2GB, Box là 5GB, OneDrive là 5GB, Google Drive là 15GB (gồm cả email, photos, files), Mega là 50GB,… Kích thước tập tin tối đa có thể tải lên cũng khác nhau ở mỗi dịch vụ, tuy nhiên do còn các yếu tố như tốc độ đường truyền internet, hạ tầng công nghệ, độ an toàn cho dữ liệu,… nên với RBCS, chúng tôi khuyến khích để dung lượng tối đa cho tập tin tải lên là 200MB.

Do kích thước tập tin tải lên là khác nhau, tuy nhiên để đảm bảo vấn đề an toàn cho dữ liệu khi lưu trữ trên các tài khoản cloud, RBCS sẽ tiến hành phân mảnh dữ liệu theo số lượng tài khoản hoặc số lượng dịch vụ, để đảm bảo tối ưu khi lưu trữ các tập tin có dung lượng nhỏ. Sau khi phân mảnh, RBCS sẽ thêm vào các mảnh dữ liệu này phần header chứa các thông tin để quản lý như sau:

Total package	Order Package	Next Storage	Filesize	Data…

Hình 5. Cấu trúc header của mỗi phần

Trong đó:

- Total package: Tổng số mảnh mà tập tin này được phân mảnh.
- Order package: Số thứ tự của mảnh trong cấu trúc.
- Next storage: Lưu mã của kho dữ liệu chứa mảnh tiếp theo.
- Filesize: Kích thước tập tin, dùng kiểm tra khi ghép mảnh lại.
- Data: Dữ liệu của mảnh.

Do được phân mảnh và được lưu trữ phân tán trên các tài khoản của các kho dữ liệu khác nhau, nên dữ liệu của mỗi mảnh trong trường hợp bị truy cập trái phép cũng không thể hiện được nội dung của toàn bộ tài liệu. Tuy nhiên, với các tập tin đơn giản không có cấu trúc header như tập tin txt, thì dữ liệu từng mảnh cũng có thể được khai thác, do vậy thao tác mã hoá dữ liệu của từng mảnh cũng sẽ được quan tâm nghiên cứu tiếp.

B. Đánh giá và so sánh RBCS với giải pháp khác

Làm sao có thể ngăn chặn truy cập bất hợp pháp tới dữ liệu của người dùng khi mật khẩu của họ đang bị đánh cắp? Mã hóa có thể là một giải pháp cho vấn đề này, vì đơn giản chỉ cần mã hóa các tập tin trước khi gửi lên các dịch vụ cloud sẽ ngăn chặn thông tin rò rỉ từ các tập tin bị đánh cắp. Khi đó nếu mật khẩu bị đánh cắp, bên thứ 3 vẫn sẽ có quyền truy cập đến dữ liệu, nhưng họ sẽ không có khả năng giải mã để xem dữ liệu [16]. Hiện nay một số phần mềm đã được phát triển dựa trên nguyên lý mã hoá dữ liệu của người dùng trước khi đưa lên cloud:

Credeoncp là một ứng dụng mã hoá phía client cho các dịch vụ lưu trữ trên cloud [15], phần mềm có thể làm việc với tất cả các nhà cung cấp dịch vụ lưu trữ cloud phổ biến hiện nay, cho phép mã hoá các tập tin dữ liệu của người dùng, bảo vệ dữ liệu trước những truy cập trái phép bên ngoài và đặc biệt hơn, ứng dụng này cam kết bảo vệ dữ liệu người dùng khỏi sự can thiệp của cả chính quyền, cung cấp mã hoá AES 256 và FIPS 140-2.

Một ứng dụng khác là Spideroak, dịch vụ này cho phép người dùng lưu trữ dữ liệu trên cloud và các tập tin sẽ được mã hoá bởi mật khẩu của chính họ trước khi được chuyển lên server. Thông tin về mật khẩu người dùng sẽ được giữ an toàn tại chính máy tính của họ và không lưu trên máy chủ của nhà cung cấp dịch vụ. Do đó, vấn đề về an toàn dữ liệu có thể đảm bảo khi chính nhà cung cấp cũng không thể truy cập trái phép các tập tin của người dùng khi không có mật khẩu.

BoxCryptor là dịch vụ trung gian giữa người sử dụng và các dịch vụ lưu trữ cloud như Dropbox, Google Drive, OneDrive,… dịch vụ này sẽ thực hiện cơ chế mã hoá các dữ liệu của người dùng trước khi tiến hành lưu trữ chúng trên các kho dữ liệu trên cloud. Dữ liệu có thể được truy cập trên các nền tảng khác nhau như mobile, desktop và các hệ điều hành như Windows, MAC, Linux.

Các giải pháp để nâng cao tính bảo mật cho các dịch vụ lưu trữ cloud hiện nay đa phần đều ứng dụng cơ chế mã hoá dữ liệu, điều này hạn chế được việc lộ dữ liệu bí mật và truy cập bất hợp pháp. Tuy nhiên, cần nhận định rằng, những điều cam kết về quyền riêng tư của người dùng từ các nhà cung cấp dịch vụ chỉ là tương đối và chúng ta chưa thể khẳng định được do hạ tầng và giải pháp của họ là hoàn toàn đóng. Bên cạnh đó, yếu tố đảm bảo tính toàn vẹn dữ

liệu chưa được đề cập nhiều, dịch vụ cloud có thể dừng bất cứ khi nào do nhiều nguyên nhân, khi đó dữ liệu của người dùng sẽ không thể khôi phục được. Với đề xuất về giải pháp RBCS, nhóm nghiên cứu đã tính đến yếu tố bảo mật dữ liệu và tính dự phòng cho việc khôi phục trong trường hợp bị mất mát.

IV. KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU

Công nghệ điện toán đám mây đang phát triển nhanh chóng và trở thành một nền tảng được sử dụng rộng rãi cho các ứng dụng tính toán phức tạp và hình thành cụm lưu trữ dữ liệu. Vấn đề an ninh và an toàn dữ liệu luôn là điều được quan tâm và thu hút nhiều nghiên cứu của các nhà khoa học. Trong nội dung bài báo, nhóm nghiên cứu đã đưa ra những lập luận và dẫn chứng về sự mất an toàn và chính sách người dùng ở khía cạnh là các dịch vụ lưu trữ đám mây công cộng. Từ đó đề xuất một giải pháp lưu trữ dữ liệu an toàn dựa trên cơ chế của RAID. Giải pháp này đã phần nào giải quyết được 2 vấn đề được quan tâm hiện nay trên dịch vụ cloud: tính bảo mật và toàn vẹn cho dữ liệu người dùng.

Hướng nghiên cứu tiếp theo sẽ tập trung vào xây dựng mô hình hoá cho giải pháp RBCS để cài đặt thử nghiệm, đánh giá mức độ tin cậy và khả năng sẵn sàng cũng như mức độ chịu lỗi của hệ thống lưu trữ. Đồng thời tìm cơ chế mã hoá cho dữ liệu khi lưu trữ trên cloud, cũng như cơ chế đồng bộ dữ liệu trên các thiết bị hay hệ điều hành như một số dịch vụ lưu trữ hiện nay cung cấp.

Qua đây, nhóm tác giả bài báo xin bày tỏ lòng cảm ơn đối với Dự án sản phẩm công nghệ cao của Bộ Công Thương "Ứng dụng công nghệ bảo đảm an ninh, an toàn mạng và bí mật thông tin ở mức cao để phát triển bộ giải pháp an toàn an ninh mạng LAN cho cơ quan Nhà nước và doanh nghiệp".

V. TÀI LIỆU THAM KHẢO

[1] Anju Chhibber, Dr. Sunil Batra, "*Security Analysis of Cloud Computing*", International Journal of Advanced Research in Engineering and Applied Sciences, ISSN: 2278-6252. Vol. 2, No. 3, pp.49-53, March 2013.

[2] Jaydip Sen, "*Security and Privacy Issues in Cloud Computing*", Innovation Labs, Tata Consultancy Services Ltd., Kolkata, INDIA, 2013.

[3] Monjur Ahmed, Mohammad Ashraf Hossain, *"Cloud Computing and Security Issues in The Cloud",* International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.1, January 2014.

[4] Cloud Security Alliance, *"Top Threats to Cloud Computing"*, 2010.

[5] Dimitrios Zissis, "*Addressing cloud computing security issues*", Future Generation Computer Systems, 28 (3), 583-592, 2012.

[6] Nir Kshetri, "*Privacy and security issues in cloud computing: The role of institutions and institutional evolution*", Telecommunications Policy, Volume 37, Issues 4–5, Pages 372–386, 2013.

[7] Daniel Fitch, Haiping Xu, "*A Raid-Based Secure and Fault-Tolerant Model for Cloud Information Storage*", International Journal of Software Engineering and Knowledge Engineering, 2013.

[8] A. Cruz, *Update on Today's Gmail Outage, Google*, February 24, 2009, retrieved on September 20, 2010 from http://gmailblog. blogspot.com/2009/02/update-on-todays-gmail- outage.html.

[9] J. Mintz, *Microsoft Dumps Windows Live Spaces for WordPress.com*, Huffington Post, September 27, 2010, retrieved on April 25, 2011 from http://www.huffingtonpost.com/ 2010/09/27/microsoft-dumps-windows-l_n_741023.html.

[10] Fahmida Y. Rashid, *Introducing the 'Treacherous 12,' the top security threats organizations face when using cloud services*, from http://www.infoworld.com/article/3041078/security/the-dirty-dozen-12-cloud-security-threats.html.

[11] Claire Reilly, *Hackers hold 7 million Dropbox passwords ransom*, from http://www.cnet.com/news/hackers-hold-7-million-dropbox-passwords-ransom/.

[12] RAID Levels and SQL Server, https://technet.microsoft.com/en-us/library/ms190764(v=sql.105).aspx.

[13] Lucas Mearian, "*No, your data isn't secure in the cloud*", from http://www.computerworld.com/article/2483552/ cloud-security/no--your-data-isn-t-secure-in-the-cloud.html, 2013.

[14] Credeoncp Application, https://credeoncp.hitachisolutions-us.com.

[15] Hector Salcedo, *Google Drive, Dropbox, Box and iCloud Reach the Top 5 Cloud Storage Security Breaches List*, from https://psg.hitachi-solutions.com/credeon/blog/google-drive-dropbox-box-and-icloud-reach-the-top-5-cloud-storage-security-breaches-list.

[16] Glenn Berry, *SQL Server Hardware*, ISBN: 978-1-906434-62-5, Simple Talk Publishing 2011.

RESEARCH ON THE MECHANISM OF RAID AND PROPOSE A SOLUTION FOR SAFETY DATA STORAGE ON CLOUD SERVICES

Le Quang Minh, Nguyen Anh Chuyen, Le Khanh Duong, Phan Huy Anh, Trinh Thi Thu

ABSTRACT*— Cloud technology is developed and widely used in the deployment of network and data storage services. Users benefit from online data storage service when registering for free accounts, the data is stored on the Cloud and has a convenient synchronization mechanism, that across multiple platforms, devices. However, the protection for this important data without lost when storage services occur errors, or to avoid infringement of an illegal manner is beyond the control of the user. In this paper, our team propose a solution for safety storage of user's data based on RAID mechanism, in order to overcome the above risks, while still able to use the storage accounts from free services.*