MỤC LỤC

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

CHÚ THÍCH VIẾT TẮT THUẬT NGỮ TIẾNG ANH

RAID	Redundant Array of Independent Disks
RBCS	RAID Based Cloud Storage
SaaS	Software as a Service
PaaS	Platform as a Service
IaaS	Infrastructure as a Service

MỞ ĐẦU

1. Lý do chọn đề tài

Ngày nay, cùng với sự phát triển không ngừng của internet , các dịch vụ lưu trữ đám mây như Google Drive, Dropbox, SugarSync, Amazon Cloud Drive,box, Mimedia (m) Drive, Skydrive, SpidekOak… cũng đang được sử dụng ngày càng rộng rãi bởi những tính năng sao lưu, lưu trữ dữ liệu trực tuyến với khả năng đồng bộ theo thời gian thực và tự động thực hiện sao lưu chia sẻ toàn bộ thư mục mà mình muốn, nó còn cho phép người sử dụng quay trở lại quá khứ để khôi phục những dữ liệu bị xóa hoặc bị thay đổi… Thêm vào đó, nhà cung cấp thường cho người dùng một số gói miễn phí hoặc với chi phí giá rất rẻ, thuận tiện trong việc cài đặt và sử dụng đối với các cá nhân và đơn vị nhỏ. Vì vậy số lượng sử dụng dịch vụ ngày càng tăng. Điều này đòi hỏi các dịch vụ trên phải tạo lập được uy tín, đảm bảo độ bảo mật và an toàn cho dữ liệu sử lưu trữ trên đó. Tuy nhiên, đây là chương trình lưu trữ tự động trên một máy chủ, tính bảo mật dữ liệu chưa thể khẳng định được, không thể chắc chắn thông tin có bị đánh cắp hoặc lộ bí mật hay không.

Chính vì vậy đề tài "***Nghiên cứu bảo vệ an toàn dữ liệu khi sử dụng dịch vụ lưu trữ điện toán đám mây***" được lựa chọn với mong muốn có thể là một tài liệu bổ ích để có thể giúp người phát triển hiểu kỹ hơn về khái niệm, lợi ích và những vấn đề liên quan đến lưu trữ đám mây. Ngoài ra đề tài cũng sẽ nghiên cứu và xây dựng một giải pháp nhằm nâng cao tính an toàn bảo mật cho dữ liệu lưu trữ đám mây.

Trên cơ sở các nghiên cứu đã có, luận văn đã tập trung vào các mục tiêu và các vấn đề cần giải quyết sau:

2. Mục tiêu và phạm vi nghiên cứu

Luận văn tập trung nghiên cứu về điện toán đám mây, các vấn đề lưu trữ dữ liệu, an toàn dữ liệu trên điện toán đám mây; chỉ ra, phân tích những mặt ưu

nhược điểm của các giải pháp đã được đưa vào sử dụng trong việc bảo vệ dữ liệu đám mây để làm rõ tính cấp thiết của đề tài. Đồng thời trình bày các phương pháp dự phòng nâng cao độ tin cậy của hệ thống. Sau đó, trình bày tổng hợp, phân tích kiến thức xoay quanh cơ chế RAID. RAID đối với bài toán an toàn dữ liệu cho hệ thống máy. Từ đó đề xuất chi tiết giải pháp RBCS, chứng minh độ tin cậy của giải pháp và vận dụng cụ thể vào doanh nghiệp Việt Nam hiện nay.

3. Phương pháp nghiên cứu

Trong luận văn này tôi đã kết hợp nhiều phương pháp nghiên cứu khác nhau phù hợp với yêu cầu của đề tài, bao gồm các phương pháp nghiên cứu sau:

- ***Phương pháp phân tích, tổng hợp***: nghiên cứu các tài liệu có liên quan tới vấn đề lưu trữ dữ liệu, thực tế vấn đề an toàn dữ liệu trong lưu trữ trên đám mây, các giải pháp đã sử dụng, các kiến thức cơ bản về cơ chế RAID, phân tích để rút ra các vấn đề cốt lõi, sau đó tổng hợp và xâu chuỗi lại để có cái nhìn tổng thể về vấn đề đang nghiên cứu.
- ***Phương pháp nghiên cứu thực tiễn***: tìm hiểu thực trạng của việc đảm bảo an toàn bảo mật dữ liệu được lưu trữ trên đám mây hiện nay, mức độ hiệu quả của các giải pháp mà các nhà chuyên gia đã đưa vào sử dụng và từ đó so sánh đánh giá hiệu quả của giải pháp được đề xuất.
- ***Phương pháp thực nghiệm***: thực nghiệm đối với cả giải pháp đã được các chuyên gia đưa ra và giải pháp đề xuất, so sánh đánh giá nhằm xác định tính khả thi, hiệu quả, mức độ giải quyết những vấn đề đang gặp phải trong việc đảm bảo an toàn dữ liệu lưu trữ trên đám mây.

4. Kết quả đạt được

Từ mục tiêu nghiên cứu giải pháp đảm bảo an toàn dữ liệu lưu trữ trên điện toán đám mây, luận văn đã tập trung làm rõ được những lý thuyết cơ bản về điện toán đám mây, vấn đề bảo vệ dữ liệu trên điện toán đám mây hiện nay, chỉ ra những giải pháp đã được sử dụng trước đó và phân tích rõ những ưu điểm, hạn chế cần phải khắc phục; các phương pháp dự phòng nâng cao độ tin cậy của hệ

thống. Sau đó, trình bày tổng hợp, phân tích kiến thức xoay quanh cơ chế RAID. RAID đối với bài toán an toàn dữ liệu cho hệ thống máy.

Đồng thời vận dụng cơ sở lý thuyết RAID vào việc giải quyết bài toán an toàn dữ liệu lưu trữ trên đám mây. Kết quả cuối cùng là luận văn đã đề xuất thành công giải pháp mới RBCS (viết tắt của RAID Based Cloud Storage), chứng minh thành công tính đúng đắn và hiệu quả và tính khả thi của giải pháp; đưa ra được quy trình cụ thể của việc ứng dụng giải pháp vào thực tiễn vào doanh nghiệp Việt Nam hiện nay.

5. Cấu trúc luận văn

Ngoài phần mở đầu, luận văn được trình bày trong ba chương, với nội dung chính của mỗi chương như sau:

Chương 1: *Tổng quan về lưu trữ dữ liệu và bảo mật dữ liệu điện toán đám mây.*

Chương này trình bày cơ sở lý thuyết về điện toán đám mây.

Lập luận dẫn chứng về những vấn đề mất mát dữ liệu, an toàn dữ liệu trong lưu trữ trên dịch vụ đám mây.

Trình bày và phân tích giải pháp đặc trưng mã hóa dữ liệu trong lưu trữ dữ liệu đám mây, ưu nhược điểm, những nhược điểm cần phải khắc phục để đảm bảo độ an toàn bảo mật cho dữ liệu đám mây.

Từ đó rút ra kết luận về tính cấp thiết, ý nghĩa thực tiễn khoa học của luận văn là giải quyết vấn đề bài toán đặt ra: "***Nghiên cứu bảo vệ an toàn dữ liệu khi sử dụng dịch vụ lưu trữ điện toán đám mây***"

Chương 2: *Các phương pháp bảo vệ an toàn khi sử dụng dịch vụ lưu trữ điện toán đám mây.*

Nêu các phương pháp dự phòng nâng cao độ tin cậy của hệ thống. Sau đó, trình bày tổng hợp, phân tích kiến thức xoay quanh cơ chế RAID. RAID đối với bài toán an toàn dữ liệu cho hệ thống máy.

Chương 3- *Đề xuất giải pháp lưu trữ dữ liệu trên đám mây – RBCS và ứng dụng vào thực tế doanh nghiệp*

Trình bày chi tiết giải pháp RBCS, phát biểu bài toán xác định và mô tả quy trình bài toán thực tế và đưa ra lập luận chứng minh độ tin cậy của giải pháp. Thực tế ứng dụng giải pháp vào doanh nghiệp.

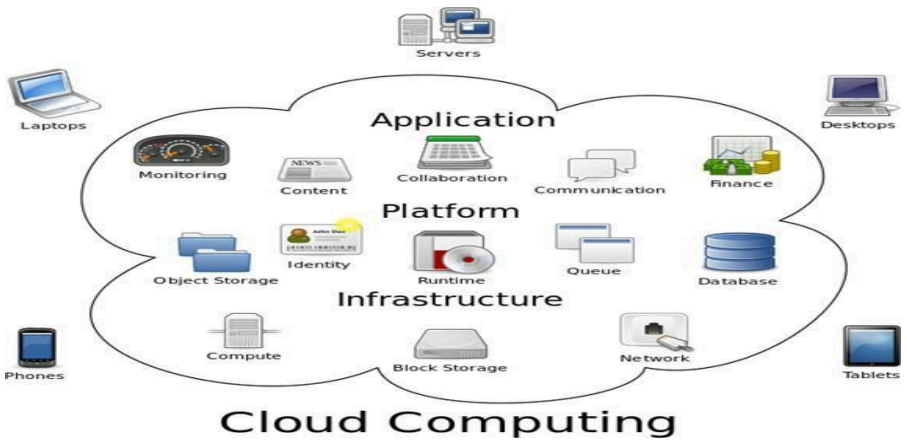
CHƯƠNG 1:
TỔNG QUAN VỀ LƯU TRỮ VÀ BẢO MẬT DỮ LIỆU ĐIỆN TOÁN ĐÁM MÂY

Chương này tập trung làm rõ những lý thuyết cơ bản về điện toán đám mây, vấn đề lưu trữ dữ liệu trên đám mây, bao gồm khái niệm, vai trò, kiến trúc, mô hình dịch vụ, mô hình triển khai điện toán đám mây, những nhà cung cấp dịch vụ và những vấn đề lưu trữ dữ liệu đám mây: mã hóa dữ liệu, bảo mật truy cập...

1.1. Khái quát về điện toán đám mây

1.1.1. Khái niệm

Điện toán đám mây (*Cloud Computing*), còn gọi là **điện toán máy chủ ảo**, là mô hình điện toán sử dụng các công nghệ máy tính và phát triển dựa vào mạng Internet. Thuật ngữ "đám mây" ở đây là lối nói ẩn dụ chỉ mạng Internet (dựa vào cách được bố trí của nó trong sơ đồ mạng máy tính) và như một liên tưởng về độ phức tạp của các cơ sở hạ tầng chứa trong nó. Ở mô hình điện toán này, mọi khả năng liên quan đến công nghệ thông tin đều được cung cấp dưới dạng các "dịch vụ", cho phép người sử dụng truy cập các dịch vụ công nghệ từ một nhà cung cấp nào đó "trong đám mây" mà không cần phải có các kiến thức, kinh nghiệm về công nghệ đó, cũng như không cần quan tâm đến các cơ sở hạ tầng phục vụ công nghệ đó. [2][3]



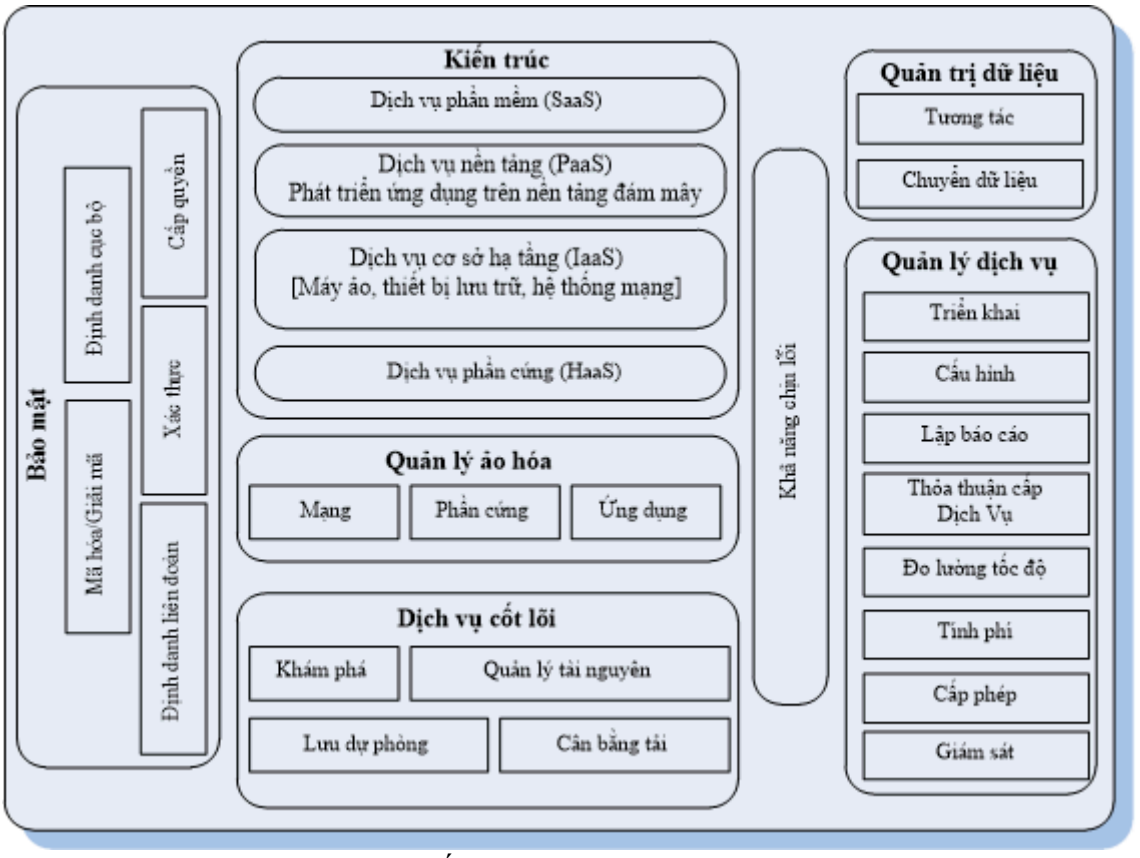
Hình 1. 1. Mô hình điện toán đám mây

1.1.2. Đặc điểm của điện toán đám mây

Những ưu điểm và thế mạnh dưới đây đã góp phần giúp "điện toán đám mây" trở thành mô hình điện toán được áp dụng rộng rãi trên toàn thế giới:

- Tốc độ xử lý nhanh, cung cấp cho người dùng những dịch vụ nhanh chóng và giá thành rẻ dựa trên nền tảng cơ sở hạ tầng tập trung (đám mây).
- Chi phí đầu tư ban đầu về cơ sở hạ tầng, máy móc và nguồn nhân lực của người sử dụng điện toán đám mây được giảm đến mức thấp nhất.
- Không còn phụ thuộc vào thiết bị và vị trí địa lý, cho phép người dùng truy cập và sử dụng hệ thống thông qua trình duyệt web ở bất kỳ đâu và trên bất kỳ thiết bị nào mà họ sử dụng (chẳng hạn là PC hoặc là điện thoại di động…).
- Chia sẻ tài nguyên và chi phí trên một địa bàn rộng lớn, mang lại các lợi ích cho người dùng.
- Với độ tin cậy cao, không chỉ dành cho người dùng phổ thông, điện toán đám mây còn phù hợp với các yêu cầu cao và liên tục của các công ty kinh doanh và các nghiên cứu khoa học. Tuy nhiên, một vài dịch vụ lớn của điện toán đám mây đôi khi rơi vào trạng thái quá tải, khiến hoạt động bị ngưng trệ. Khi rơi vào trạng thái này, người dùng không có khả năng để xử lý các sự cố mà phải nhờ vào các chuyên gia từ “đám mây” tiến hành xử lý.
- Khả năng mở rộng được, giúp cải thiện chất lượng các dịch vụ được cung cấp trên “đám mây”.
- Khả năng bảo mật được cải thiện do sự tập trung về dữ liệu.
- Các ứng dụng của điện toán đám mây dễ dàng để sửa chữa và cải thiện về tính năng bởi lẽ chúng không được cài đặt cố định trên một máy tính nào.
- Tài nguyên sử dụng của điện toán đám mây luôn được quản lý và thống kê trên từng khách hàng và ứng dụng, theo từng ngày, từng tuần, từng tháng. Điều này đảm bảo cho việc định lượng giá cả của mỗi dịch vụ do điện toán đám mây cung cấp để người dùng có thể lựa chọn phù hợp.

1.1.3. Kiến trúc của điện toán đám mây



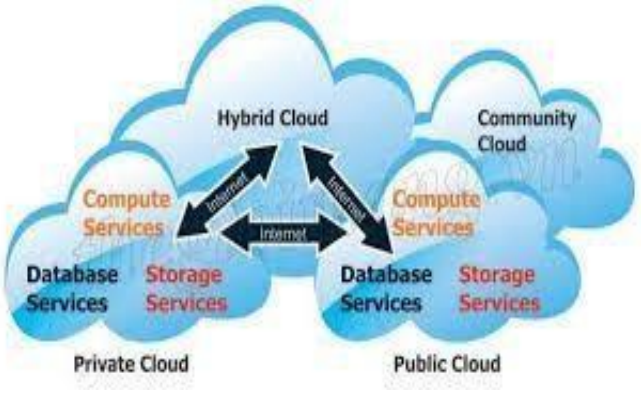
Hình 1. 2. Mô tả kiến trúc của điện toán đám mây

Điện toán đám mây bao gồm 6 thành phần cơ bản:

- Cơ sở hạ tầng (Infrastructure)
- Lưu trữ đám mây (Cloud Storage)
- Nền tảng đám mây (Cloud Platform)
- Ứng dụng (Application)
- Dịch vụ (Services)
- Khách hàng (Client)

Các mô hình triển khai điện toán đám mây:

- Đám mây công cộng (Public cloud)
- Đám mây riêng (Private cloud)
- Đám mây cộng đồng (Community cloud)
- Đám mây lai (Hybird cloud)



Hình 1. 3. Các mô hình triển khai điện toán đám mây

1.2. Các nhà cung cấp dịch vụ điện toán đám mây

Các nhà cung cấp dịch vụ điện toán đám mây cung cấp các dịch vụ của họ theo ba mô hình cơ bản: Cơ sở hạ tầng như một dịch vụ (IaaS – Infrastructure as a Service), nền tảng như một dịch vụ (PaaS – Platform as a Service) và phần mềm như một dịch vụ (SaaS – Software as a Service)

Hình 1. 4. Các mô hình dịch vụ của điện toán đám mây

- IaaS: Cho phép bạn truy cập đến phần cứng hệ thống mạng máy tính. Cung cấp nhiều nguồn tài nguyên như là firewalls, load balancers, các địa chỉ IP nhưng hệ điều hành và các ứng dụng sẽ do bạn cài đặt và cập nhật. Điều này giúp

bạn linh hoạt hơn trong việc sử dụng tài nguyên vào mục đích gì. IaaS xuất hiện rộng rãi bởi các nhà cung cấp Amazon, Memset, Google, Windows…. Một cách giúp quản lý IaaS dễ dàng hơn là phát triển các templates cho các dịch vụ đám mây nhằm tạo ra 1 bản kế hoạch chi tiết để xây dựng hệ thống ready-to-use, và tránh tình trạng di chuyển giữa các đám mây khác nhau.

Với loại mô hình này hiện nay có các nhà cung cấp dịch vụ lớn như là Amazon Web services và Microsoft Azure... Amazon Web service hiện đang là nhà cung cấp dịch vụ cloud IaaS giàu tiềm năng nhất, tuy nhiện hiện nay họ đang phải cạnh tranh về thì phần với 2 ông lớn công nghệ là Microsoft và Google.

Amazon Web Services là tập hợp các dịch vụ cung cấp cho người lập trình có khả năng truy cập tới hạ tầng kiến trúc tính toán kiểu sẵn sàng-để-sử dụng (ready-to-use) của Amazon. Các máy tính có nền tảng vững chắc đã được xây dựng và tinh chế qua nhiều năm của Amazon bây giờ là có thể cho phép bất cứ ai cũng có quyền cập tới Internet. Amazon cung cấp một số dịch vụ Web nhưng trong loạt bài viết này chỉ tập trung vào các dịch vụ khối hợp nhất (building-block) cơ bản, cái mà đáp ứng được một số yêu cầu cốt lõi của hầu hết các hệ thống như: lưu trữ, tính toán, truyền thông điệp và tập dữ liệu. Bạn có thể xây dựng các ứng dụng phức tạp và gồm nhiều phần khác nhau bằng cách sử dụng các chức năng phân tầng với các dịch vụ đáng tin cậy, hiệu quả khối hợp nhất được cung cấp bởi Amazon. Các dịch vụ Web mà tồn tại bên trong đám mây phía bên ngoài môi trường của bạn và có khả năng thực hiện là rất cao. Bạn sẽ trả chỉ dựa trên những cái bạn sử dụng mà không cần phải trả trước các chi phí và vốn đầu tư ban đầu. Bạn không cần phải mất chi phí cho bảo trì bởi vì phần cứng được duy trì và phục vụ bởi Amazon.

Trong ngành dịch vụ IaaS này Microsoft Azure thật sự là một một đối thủ nặng kí của AWS. Với thế mạnh về phân tích, lưu trữ cá nhân và đặc biệt là giải quyết được các thảm họa như phục hồi dữ liệu, khắc phục lỗi ứng dụng với các gói dịch vụ mở rộng của họ. Microsoft Azure là nền tảng điện toán đám mây mở

và linh hoạt cho phép bạn nhanh chóng xây dựng, triển khai và quản lý các ứng dụng thông qua mạng lưới toàn cầu của trung tâm dữ liệu Microsoft.

Microsoft Azure luôn đảm bảo tính sẵn sàng và có thiết kế tải cân bằng và có khả năng tự phục hồi khi phần cứng có sự cố. Bạn có thể sử dụng bất kỳ ngôn ngữ, công cụ hay nền tảng nào để xây dựng các ứng dụng. Và bạn có thể tích hợp các ứng dụng trên đám mây công cộng của bạn với môi trường IT có sẵn.

- **SaaS:** cho bạn truy cập đến các phần mềm trên nền tảng đám mây mà không cần quản lý cơ sở hạ tầng và nền tảng nó đang chạy phù hợp khi bạn muốn tập trung vào người dùng cuối. Điều này có nghĩa là nó dễ dàng truy cập và có khả năng mở rộng. Có rất nhiều ví dụ về SaaS gồm email, phần mềm văn phòng và các công cụ kiểm toán từ Google, Microsoft, Freshbooks … Các "as a service" khác.

Khi mô tả về điện toán đám mây, người ta hay thêm vào "as a service" phía sau để định nghĩa nó là 1 hệ thống mạng toàn cầu hơn là ngồi trên máy tính trong văn phòng. Từ "Storage as a service" (StaaS), "Data as a service" (DaaS) đến "Security as a service" (SECaaS), có rất nhiều biến thể từ 3 dạng gốc nói *trên.*

Salesforce.com và insightly là hai nhà cung cấp dịch vụ điện toán đám đám mây SaaS lớn hiện nay. Salesforce một nhà cung cấp giải pháp SaaS CRM với các Gartner nêu rõ rằng công ty đang chiếm lĩnh thị trường này. Insightly cung cấp SaaS CRM tích hợp với Gmail và Google Apps của Google, cũng như Outlook 2013 và Office 365. Với các ứng dụng dịch vụ của Insightly giúp khách hàng theo dõi và nghiên cứu được các khách hàng tiềm năng của họ. Tất cả các ứng dụng có thể truy cập từ iOS va Android.

- **PaaS:** hỗ trợ người sử dụng điện toán đám mây bằng các hệ điều hành, cơ sở dữ liệu, máy chủ web và môi trường thực thi lập trình. Hơn nữa, nó cho phép bạn tập trung vào các ứng dụng cụ thể, cho phép các nhà cung cấp đám mây quản lý và đo đạc tài nguyên 1 cách tự động.

Vậy PaaS có thể cho phép bạn tập trung hơn vào ứng dụng và dịch vụ đầu cuối hơn là phí thời gian cho hệ điều hành. Các nhà cung cấp IaaS cũng cung cấp PaaS, giúp bạn giảm tải công việc. Với loại hình công nghệ loại này có 2 nhà đầu tư phát triển nổi bật là *Red Hat OpenShift*Phần mềm chạy dịch vụ là mã nguồn mở và có sẵn trên GitHub với tên "OpenShift Origin". Người phát triển phần mềm có thể sử dụng Git để triển khai ứng dụng bằng các ngôn ngữ khác nhau trên nền tảng.

Đặc biệt, OpenShift cũng hỗ trợ các ứng dụng web dạng phần mềm mã nhị phân, miễn là nó có thể chạy trên RHEL Linux. Điều này làm tăng tính tùy biến của hệ thống, hỗ trợ nhiều ngôn ngữ và frameworks. OpenShift bảo trì dịch vụ bên dưới ứng dụng và thống kê ứng dụng nếu cần thiết.

Các dịch vụ đám mây phổ biến như: EC2 của Amazon, Azure của Microsoft, IBM cung cấp Smart Cloud Enterprise, Google cung cấp App Engine, Redhat cung cấp Redhat's Openshift, Vmware có Cloud Foundry, Viện Công nghiệp Phần mềm và Nội dung số Việt Nam có iDragon Clouds... Trong đó Google Cloud, Redhat's Openshift, Vmware Cloud Foundry và NISCI iDragon Clouds là những PaaS mã nguồn mở, cho phép thực thi trên một nền hạ tầng với chi phí thấp và dễ dàng thay thế. Theo nhiều chuyên gia đánh giá, số người sử dụng đám mây công cộng sẽ lên đến 1 tỷ trước năm 2020. Người ta cho rằng, năm 2012 trên thế giới có khoảng 1 tỷ người đang sử dụng theo các truyền thống như Microsoft Office, OpenOffice or LibreOffice, Microsoft Exchange or Sharepoint, IBM Lotus Notes, thì đến năm 2020 tất cả mọi người sẽ chuyển sang đám mây công cộng.

Với công nghệ lưu trữ đám mây chúng ta có thể xử lí dữ liệu dưới dạng cấu trúc và phi cấu trúc, tài liệu đến hình ảnh từ nhiều nguồn khác nhau. Ở đây ta xét 3 nhà cung cấp dịch vụ lưu trữ dữ liệu đám mây lớn là Google drive, dropbox và box. Cụ thể:

Hình 1. 5. Một số biểu tượng nhà cung cấp dịch vụ đám mây

- Với **Google Drive**: Dịch vụ lưu trữ của google cung cấp một không gian lưu trữ trên nền tảng cloud. Miễn phí lưu lượng lên đến 15GB. cho phép người dùng lưu trữ nhiều dạng dữ liệu như văn bản, video, âm thanh, PDF… trên nền tảng "đám mây". Google Drive với khả năng hỗ trợ Google Docs và Google+ cao cấp giúp người dùng dễ dàng truy cập và chỉnh sửa tài liệu ở bất cứ đâu hay chia sẻ tập tin chung với bạn bè. Với Google Drive, bạn có thể truy cập đến tài liệu của mình bất cứ đâu và bất cứ thiết bị iPhone, iPad, SmartPhone Android, Laptop hay máy để bàn.

- **Dropbox** là dịch vụ lưu trữ dữ liệu trực tuyến miễn phí cho phép bạn mang theo tất cả tài liệu, ảnh và video tới bất cứ nơi nào. Điều này có nghĩa là tập tin bạn đã lưu vào Dropbox sẽ tự động lưu vào máy tính, điện thoại của bạn và cả trên website Dropbox.

Ứng dụng này cũng giúp bạn dễ dàng chia sẻ tài liệu cho nhiều người. Thậm chí trong trường hợp ổ cứng máy tính bị hỏng, dữ liệu trên điện thoại mất hoàn toàn thì bạn vẫn có thể yên tâm vì đã có một bản sao lưu nội dung trên Dropbox.

- **Box:**là dịch vụ sao lưu, lưu trữ dữ liệu trực tuyến với khả năng đồng bộ theo thời gian thực và tự động thực hiện sao lưu, hỗ trợ miễn phí có thể lên đến 10GB dung lượng lưu trữ trực tuyến trên máy chủ của họ.

Để phân biệt với các nhà cung cấp khác, Box cho biết quy trình làm việc, một công cụ tự động hóaviệc định tuyến các tài liệu và tập tin cũng như các hành động của người cần phải thực hiện trên chúng. Các tập tin dữ liệu trên Box cho phép người dùng có thể chia sẽ và làm việc cùng nhau trên cùng một tập tin.

1.3. Phương pháp bảo vệ dữ liệu lưu trữ trên đám mây

1.3.1. Một số vấn đề thực tế về an toàn dữ liệu trong lưu trữ trên đám mây hiện nay

Công nghệ điện toán đám mây đang được sử dụng ngày càng phổ biến rộng rãi nhờ những vai trò hấp dẫn, tính tiện ích của dịch vụ mang lại cho người dùng. Tuy nhiên với thực tế nhu cầu sử dụng quá lớn, yêu cầu của mỗi đối tượng sử dụng ngày càng đa dạng, lượng dữ liệu cần lưu trữ tăng chóng mặt, những dữ liệu mật quan trọng của cá nhân hay tổ chức cũng được lưu trũ trên đám mây. Nếu những dữ liệu đó bị mất mát hay sao chép thì hậu quả sẽ vô cùng nghiêm trọng không thể ước tính được. Điều này đặt ra cho nhà cung cấp dịch vụ bài toán làm sao để đảm bảo tính an toàn và bảo mật dữ liệu và có được sự tin cậy của người dùng. Đã có những giải pháp được đưa ra và thực tiễn trong thời gian qua giải quyết được phần nào vấn đề bảo vệ dữ liệu trên đám mây. Một câu hỏi đã đặt ra: Liệu dữ liệu được lưu trữ trên đám mây có được đảm bảo tuyệt đối tính an toàn bảo mật hay không?

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 khidữ 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 vềan ninh và an toàn dữ liệu xảy ra.[4]

Theo một thống kê trên trang cnet.com, hàng loạt dịch vụ lưu trữ dữ liệu trực tuyến với hàng triệu tài khoảnđang 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[18].

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.

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 21000 yêu cầu từ phía chính phủ về việc cung cấp thông tin của hơn 33000 tài khoản người dùng [13]. Các công ty công nghệ khác như Microsoft cũng nhận được hơn 70000 yêu cầu về 122000 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 vụlà chính xác haykhông.

Theo [6], 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 này. 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. Trên thực tế lập luận và dẫn chứng, những giải pháp đã được đưa vào sử dụng chưa đảm bảo được tuyệt đối tính an toàn, toàn vẹn dữ liệu lưu trữ trên 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[7]:

- 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 tấncông các máy tínhkhác.
- 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 điệntoán 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.

1.3.2. Các biện pháp bảo vệ dữ liệu lưu trữ trên đám mây được sử dụng hiện nay

Mã hóa dữ liệu:

Mã hóa dữ liệu là công nghệ chuyển hóa dữ liệu này thành 1 dạng dữ liệu mới mà người dùng không thể đọc được hoặc hiểu được nó. Bằng cách sử dụng các thuật toán lồng vào nhau, thường dựa trên 1 khóa (key) để mã hóa dữ liệu.

Hầu hết các hình thức mã hóa đều yêu cầu bạn thiết lập mật khẩu, cho phép bạn mã hóa tập tin và sau đó giải mã nó khi bạn muốn xem lại. Nếu bạn sử dụng mật khẩu yếu, tin tặc có thể phá mã hóa và truy cập tập tin, làm thất bại mục đích của mã hóa.

Một mật khẩu mạnh nên có từ 10-12 ký tự, nên là sự kết hợp của chữ hoa và chữ thường, số và ký hiệu. Nếu bạn thấy mật khẩu chỉ chứa chữ cái sẽ dễ nhớ hơn, thì một mật khẩu vẫn có thể an toàn nếu nó dài hơn đáng kể, ví dụ như gồm 20 ký tự hoặc nhiều hơn.

Mã hóa nhằm đảm bảo các yêu cầu sau:

- Tính bí mật (Confidentiality): dữ liệu không bị xem bởi bên thứ ba.
- Tính toàn vẹn (Integrity): dữ liệu không bị thay đổi trong quá trình truyền

- Tính không khước từ (Non-repudiation): là cơ chế người thực hiện hành động không thể chối bỏ việc mình đã làm. Có thể kiểm chứng được nguồn gốc hoặc người đưa tin

Độ khó của mã khóa được đo bằng thời gian và vật chất được yêu cầu để giải mã. Mã hóa/giải mã được thực hiện với một khóa (key).

Key = một từ, số, câu,...

Cùng một thông tin được mã hóa với các khóa khác nhau sẽ cho ra kết quả mã hóa khác nhau. Tính an toàn của thông tin phụ thuộc vào độ khó của giải thuật và độ bí mật của khóa.

Hệ thống mã hóa (crytosytem) = giải thuật + khóa + quy trình.

Vậy thuật toán mã hóa bao gồm những loại nào, cách phân loại và đặc điểm của mỗi loại?

Thuật toán mã hóa có thể phân theo 2 cách: theo phương pháp và theo khóa.

Thứ nhất, phân theo phương pháp bao gồm: mã hóa cổ điển, đối xứng, bất đối xứng và mã hóa băm.

Trong đó:

- Mã hóa cổ điển thuật toán sử dụng khóa đơn giản dễ hiểu. Là phương pháp mà từng ký tự hay nhóm ký tự được thay bằng một ký tự hay nhóm ký tự khác. Bên nhận chỉ cần đảo lại trình tự thay thế trên thì sẽ nhận dc bản ban đầu.

Mã hóa cổ điển có hai phương pháp nổi bật là: Mã thay thế và mã hóa hoán vị. Các hệ mã hóa thường được sử dụng tring lịch sử là hệ mã hóa Ceasar, Vigenere, Hill...

- Mã hóa đối xứng hay mã hóa chia sẻ khóa là mô hình 2 chiều, tiến trình mã hóa và giải mã dùng chung một khóa. Khóa này được chuyển giao bí mật giữa

hai đối tượng tham gia giao tiếp. Mã đối xứng thực hiện nhanh nhưng có thể gặp rủi ro nếu bị đánh cắp.

Một số thuật toán mã hóa đối xứng nổi tiếng như DES, AES, RC4, RC2, RC 5, RC6... Ngoài ra còn một số thuật toán như: Skipjact, Blowfish, CATS-128

- Mã hóa bất đối xứng là mô hình mã hóa hai chiều sử dụng một cặp khóa là khóa chung (Public Key) và khóa riêng (Prtvate key). Trong đó khóa chung được công bố rông rãi. Người nhận thông tin sẽ dùng khóa Private Key để giải mã. Khóa Private Key chỉ do một người dữ nên do đó các phương pháp mã hóa bất đối xứng đảm bảo tính bí mật hơn.

Thuật toàn mã hóa bất đối xứng nổi tiếng và được sử dụng nhiều nhất ở RSA. Ngoài ra còn có một số thuật khác như: Hellman, Elgamal...

- Mã hóa hàm băm: Là cách thức mã hóa một chiều tiến hành biến đổi rõ bả rõ thành bản mà không bao giờ giải mã được. Trong xử lý hàm băm, dữ liệu đầu vào có thể khác nhau về độ dài nhưng độ dài của cử lý băm luôn xác định.

Một số thuật toán mã hóa hàm băm thường dùng như: MD4, MD5, SHA...

Thứ hai, với cách chia theo khóa, mã hóa được chia làm 2 loại là khóa công khai và khóa bí mật

- Khóa công khai: các thuật toán sử dụng mã hóa và khóa giải mã hoàn toàn khác nhau. Hơn nữa khóa hóa giải không thể tính toán được từ khóa mã hóa.Một người bất kỳ có thể sử dụng khóa công khai để mã hóa thông tin nhưng chỉ có người nhận thông tin có khóa giải mã phù hợp với khóa mã hóa để giải mã thông tin đó.

- Khóa bí mật: là thuật toán mà tại đó khóa giải mã có thể được tính toán từ khóa mã hóa. Trong rất nhiều trường hợp khóa mã hóa và khóa giải mã giống nhau. Thuật toán này yêu cầu người gửi và nhận thỏa thuận một khóa trước khi thông tin được gửi đi và khóa này được giữ bí mật. Độ an toàn của thuật toán này

phụ thuộc nhiều vào độ bí mật của khóa, nếu để lộ khóa thì bất kỳ người nào cũng có thể dễ dàng mã hóa và giải mã thông tin.

Có thể thấy hiện nay, mã hóa dữ liệu trước khi đưa lên lưu trữ trên cloud hay bảo mật truy cập phân quyền đang được sử dụng phổ biến để nâng cao tính an toàn bảo mật.

Hầu hết các admin cho rằng, dữ liệu khi lưu trữ trên đám mây, hay truyền đi trên kênh đều cần được mã hóa. Việc mã hóa dữ liệu trên kênh truyền giờ đây gần như không còn là vấn đề phức tạp nữa, bởi các nhà cung cấp dịch vụ thường hỗ trợ giao thức HTTPS để truyền dữ liệu. Vấn đề còn lại là mã hóa dữ liệu khi lưu trữ trên đám mây. Để giải quyết điều này có thể sử dụng nhiều công nghệ khác nhau. Một trong số đó là LUKS (The Linux Unified Key Setup, http://code.google.com/p/cryptsetup). LUKS được triển khai trong dm-crypt, sử dụng trong Linux để mã hóa các ổ đĩa logic. LUKS cũng hỗ trợ sơ đồ lập khóa TKS1 (Template Key Setup 1). TKS1 cho phép thay đổi khóa mà không cần phải mã hóa lại đĩa, có thể sử dụng nhiều khóa, có thể chia sẻ bí mật bằng cách yêu cầu nhập hai khóa.Trong Windows, công nghệ này được triển khai tại FreeOTFE (http://freeotfe.org). FreeOTFE tương thích với các ổ đĩa được mã hóa trong Linux (bởi cryptoloop hay dm- crypt) và hỗ trợ xác thực hai yếu tố bằng thẻ thông minh (smart card) và HSM (Hardware Security Module) theo chuẩn PKCS#11. Ngoài ra, các phiên bản Windows gần đây (kể từ Windows Vista) đều tích hợp công cụ mã hóa ổ đĩa BitLocker. Trước đó, EFS (Encrypted File System) cho phép mã hóa từng tập tin hay thư mục.

Thực ra, không nhất thiết phải mã hóa toàn bộ đĩa. Trong trường hợp dữ liệu được lưu trong các hệ quản trị cơ sở dữ liệu (DataBase Management System- DBMS) thì chỉ cần mã hóa các bảng chứa dữ liệu nhạy cảm. Đa phần các DBMS ngày nay đều cung cấp những cơ chế bảo vệ đáng tin cậy. Ví dụ, MySQL hỗ trợ đến 15 hàm mật mã và nén dữ liệu.

Giải pháp mã hóa dữ liệu được đưa vào sử dụng và đã phát triển mạnh. Dựa vào thực tế lập luận dẫn chứng, có thể đưa ra những điểm ưu và nhược của phương pháp này như sau:

Ưu điểm:

Với bài toán bảo vệ an toàn dữ liệu lưu trữ trên đám mây đặt ra, giải pháp mã hóa dữ liệu có những mặt tích cực, giải quyết được những vấn đề sau:

- Các hệ mã hóa che dấu được nội dung của văn bản rõ để đảm bảo cho chỉ người chủ hợp pháp của thông tin mới có quyền truy cập thông tin, hay nói cách khác là chống truy cập không đúng quyền hạn.Khi tệp tin bị tin tặc đánh cắp, tin tặc không thể đọc được dữ liệu trong tệp tin đó, điều đó hạn chế được việc tin tặc sử dụng trái phép gây ra những hậu quả không đáng có.
- Tạo các yếu tố xác thực thông tin, đảm bảo thông tin lưu hành trên hệ thống đến người nhận hợp pháp xác thực.
- Tổ chức các sơ đồ chữ ký điện tử, đảm bảo không có hiện tượng giả mạo để gửi thông tin trên mạng.

Nhược điểm:

Không thể phủ nhận những điểm tốt điểm mạnh của việc mã hóa dữ liệu trên đám mây. Tuy nhiên,mã hóa không thể giải quyết hết được vấn đề bảo mật, bởi nguyên nhân của rò rỉ thông tin bao gồm cả việc tồn tại các lỗ hổng như XSS hay SQL injection, cũng như là việc sử dụng mật khẩu khẩu quá ngắn hoặc dễ đoán... Bản thân việc mã hóa không ngăn chặn được việc thông tin bị đánh cắp, việc mất mật khẩu sẽ dẫn tới mất gói dữ liệu, việc này dẫn đến việc không đảm bảo được tính an toàn, toàn vẹn về mặt dự phòng dữ liệu. Bên cạnh đó cũng phải nhấn mạnh rằng, đa phần các công nghệ quản lý khóa mã được sử dụng rộng rãi hiện nay đều tiềm ẩn những rủi ro. Chưa có câu trả lời hoàn hảo cho các câu hỏi như “lưu trữ khóa ở đâu”, “phải bảo vệ khóa ra sao”, “nhập khóa như thế nào”.

Nếu số lượng máy ảo trong hệ thống là lớn thì bản thân hệ thống mật mã có thể là nguồn căn của vấn đề. Đó là vì cần có sự phân quyền xem ai được truy cập tới đối tượng nào và tương ứng với nó là việc phân phối khóa.

Để đánh giá một hệ mã hóa ta cần xét đến độ an toàn của thuật toán, tốc độ mã hóa, giải mã và việc phân phối khóa. Có thể thấy "thuật toán nào cũng có thể bị phá vỡ". Các thuật toán khác nhau cung cấp mức độ an toàn khác nhau, phụ thuộc vào độ phức tạp để phá vỡ chúng. Ở đây tôi muốn nhắc đến mức độ phức tạp của hệ mã hóa được sử dụng thấp thì rủi ro gói tin bị phá sẽ cao. Bên cạnh đó, tốc độ xử lý mã hóa, giải mã gói dữ liệu nhanh hay chậm cũng phụ thuộc vào hệ mã hóa có tốt không.

Việc bảo mật truy cập, dùng mật khẩu chia quyền cho người sở hữu và người được quyền sử dụng. Ở đây tùy theo người sở hữu có cách chia quyền khác nhau (ví dụ chia quyền cho người sử dụng này chỉ được đọc dữ liệu mà không được chỉnh sửa...), người dùng hoàn toàn tin tưởng vào nhà cung cấp. Đồng nghĩa với việc không thể đảm bảo tính an toàn bảo mật dữ liệu do những nguyên nhân chủ quan hoặc khách quan xuất phát từ nhà cung cấp dịch vụ mà bạn đang tin tưởng.

Vậy vấn đề cấp thiết đặt ra là tìm ra giải pháp tối ưu làm thế nào giải quyết được vấn đề an toàn bảo mật dữ liệu lưu trữ trên đám mây, khắc phục những hạn chế của giải pháp mã hóa dữ liệu đã áp dụng trước đó mà vẫn đảm bảo được hiệu quả sử dụng cho người dùng.

CHƯƠNG 2:

CƠ SỞ LÝ THUYẾT ĐÁNH GIÁ VÀ NÂNG CAO KHẢ NĂNG DỰ PHÒNG ĐẢM BẢO AN TOÀN DỮ LIỆU.

2.1. Tổng quan về phương pháp nâng cao độ tin cậy hệ thống

2.1.1. Một số khái niệm

Hệ thống là một tập hợp gồm nhiều phần tử tương tác, có các mối quan hệ ràng buộc lẫn nhau, tương hỗ nhau và cùng thực hiện hướng tới một mục tiêu nhất định"[3].

"Phần tử là một đối tượng có độ tin cậy độc lập, một bộ phận tạo thành hệ thống mà trong quá trình nghiên cứu độ tin cậy nó được xem như là một đơn vị không chia nhỏ hơn nữa trong hệ thống" [3].

Độ tin cậy của phần tử hoặc hệ thống là xác suất để trong suốt khoảng thời gian khảo sát t, phần tử đó hoặc hệ thống đó vận hành an toàn [3,4].

Giả sử gọi P(t) là độ tin cậy của phần tử, được định nghĩa như biểu thức sau:

$$P(t) = P\{\tau \geq t\} \qquad (2.1)$$

Trong đó: τ là thời gian liên tục vận hành an toàn của phần tử.

Theo công thức (1.1) phần tử chỉ vận hành an toàn với một xác suất nào đó ($0 \leq P \leq 1$) trong suốt khoảng thời gian t.

Khi bắt đầu vận hành nghĩa là ở thời điểm t = 0, phần tử bao giờ cũng hoạt động tốt nên P(0)= 1. Ngược lại thời gian càng kéo dài, khả năng vận hành an toàn của phần tử càng giảm đi và tới khi $t \rightarrow \infty$ thì theo quy luật phát triển của vật chất trong tác động bào mòn của thời gian, phần tử đó sẽ hỏng nên $P(\infty) = 0$. Vì phần tử bị hư hỏng là một sự kiện ngẫu nhiên xảy ra ở các thời điểm khác nhau nên các chỉ tiêu độ tin cậy cũng thường tính dưới dạng xác suất.

Theo định nghĩa xác suất [2] thì xác suất không an toàn Q(t) hay còn gọi là xác suất hỏng của hệ thống sẽ là:

$$Q(t) = 1 - P(t) \qquad (2.2)$$

2.1.2. Phương pháp đánh giá độ tin cậy của hệ thống qua cấu trúc hệ thống

Vậy làm thế nào để đánh giá độ tin cậy của một hệ thống?

Phần tử bị hư hỏng là một sự kiện ngẫu nhiên xảy ra ở các thời điểm khác nhau nên các chỉ số về độ tin cậy cũng thường tính dưới dạng xác suất.

Độ tin cậy của phần tử giảm dần theo thời gian, để tăng độ tin cậy của hệ thống thì phải thiết kế tăng độ tin cậy của phần tử.

Độ tin cậy hay xác suất vận hành an toàn của hệ thống cấu trúc các phần tử song song luôn cao hơn hệ thống cấu trúc các phần tử nối tiếp.

Cấu trúc của một hệ thống dù phức tạp đến đâu thì cũng chỉ quy về 2 dạng là cấu trúc nối tiếp và cấu trúc song song[6]. Phương pháp tính độ tin cậy của hệ thống qua cấu trúc nối tiếp và song song hay còn được biết đến với tên gọi khác là: phương pháp tính độ tin cậy của hệ thống không dự phòng và hệ thống có dự phòng [3].

Sơ đồ khối độ tin cậy [22] có thể được xem xét một cách độc lập bởi các thành phần của hệ thống có thể được ước tính độ tin cậy và khả năng sẵn sàng (hoặc không). Việc xây dựng sơ đồ khối độ tin cậy có thể khó khăn đối với hệ thống lớn và phức tạp.

Sơ đồ khối độ tin cậy bao gồm:

- Các nút: nút nguồn, nút tải và các nút trung gian.
- Các nhánh: Được vẽ bằng các khối hình chữ nhật mô tả trạng thái tốt của phần tử. Phần tử bị hỏng tương ứng với việc xóa khối của phần tử đó ra khỏi sơ đồ.

Nhánh và nút tạo thành mạng lưới nối liền nút phát và nút tải của sơ đồ. Có thể có nhiều đường nối từ nút phát đến nút tải, mỗi đường gồm nhiều nhánh nối tiếp, vì vậy số đường đi từ nút phát đến nút tải là rất lớn đối với các hệ thống phức tạp.

Theo sơ đồ mô tả:

- Trạng thái tốt của hệ thống là trạng thái trong đó có ít nhất một đường có thể đi từ nút phát đến nút tải.

- Trạng thái hỏng của hệ thống khi nút phát bị tách rời với nút tải do hỏng hóc của phần tử trung gian.

Với hệ thống không dự phòng (hệ thống các phần tử nối tiếp) và hệ thống dự phòng (hệ thống các phần tử song song) có sơ đồ khối tin cậy và công thức tính độ tin cậy riêng cho từng loại hệ thống.

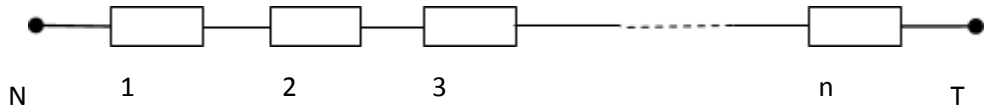
Cụ thể:

- Hệ thống các phần tử nối tiếp:

Hệ thống các phần tử nối tiếp [3, 5] (hệ thống không dự phòng) là hệ thống trong đó sự hỏng của toàn hệ thống xảy ra khi và chỉ khi một phần tử bị hỏng.

Cấu trúc đơn giản nhất là cấu trúc không có dự phòng của một hệ thống được tạo nên bởi n phần tử, mỗi trở ngại của một phần tử riêng biệt đều dẫn đến trở ngại của cả hệ thống [16].

Xét sơ đồ tin cậy của hệ thống gồm n phần tử nối tiếp như hình vẽ:



Hình 2.1. Sơ đồ của hệ các phần tử nối tiếp

Trong đó N là nút nguồn (nút phát) và T là nút tải.

$P_i(t)$ là xác suất không hỏng hay hàm tin cậy của phần tử thứ i, ở thời điểm xác định t và $P_s(t)$ của hệ thống [3, 14, 17].

$$P_s(t) = P_1(t).P_2(t) \dots P_n(t) = \prod P_i(t) \quad (2.3)$$

Vì $P_i(t) \leq 1$ nên

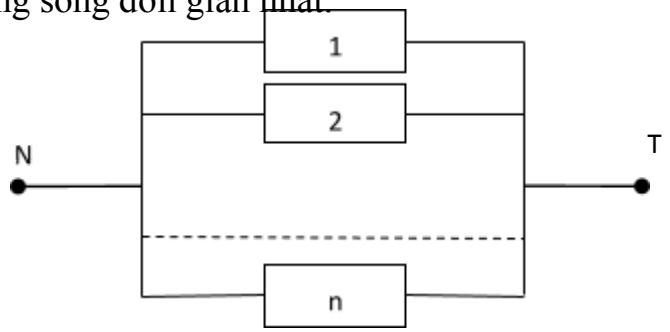
$$P_s(t) \leq min\, P_i(t); \quad i = 1, \dots, n \quad (2.4)$$

Như vậy, độ tin cậy của hệ nối tiếp không lớn hơn độ tin cậy của phần tử kém

tin cậy nhất trong hệ thống.

- **Hệ thống các phần tử song song**

Trong hệ thống các phần tử song song [3, 17,18] (hệ thống có dự phòng), sự cố của một phần tử nào đó không nhất định là sẽ dẫn đến sự cố cho toàn hệ thống, hệ thống sẽ gặp sự cố khi tất cả các phần tử gặp sự cố. Hình 1.9 thể hiện sơ đồ các phần tử song song đơn giản nhất.



Hình2.2. Sơ đồ hệ các phần tử song son

Xác suất sự cố $Q_s(t)$ của toàn hệ thống, hệ thống có sự cố khi toàn bộ n phần tử bị sự cố [15, 20]:

$$Q_s(t) = Q_1(t).Q_2(t) \dots Q_n(t) = \prod_{i=1}^{n} Q_i(t) \quad (2.5)$$

Trong đó $Q_i(t)$ với i = 1... n là xác suất sự cố của phần tử thứ i trong khoảng thời gian t khảo sát

Giả thiết độ tin cậy tuân theo quy luật hàm số mũ:

$$P_i(t) = e^{-\lambda_{i.t}} \quad (2.6)$$

Thì ta có xác suất sự cố của toàn hệ thống là [7]:

$$Q_s(t) = \prod_{i=1}^{n}(1 - e^{-\lambda_{i.t}}) \quad (2.7)$$

Độ tin cậy của hệ thống:

$$P_S(t) = 1 - Q_S(t) = 1 - \prod_{i=1}^{n}(1 - e^{-\lambda_{i.t}})$$

$$= 1 - (1 - P_1)(1 - P_2) \dots (1 - P_n) \qquad (2.8)$$

Từ công thức tính độ tin cậy của hệ thống (2.8) so sánh với công thức (2.9) ở trên ta thấy rõ ràng xác suất làm việc không có sự cố của hệ thống song song luôn cao hơn xác suất làm việc không có sự cố của hệ thống nối tiếp.

Thời gian hoạt động an toàn trung bình của hệ thống là:

$$T_s = \frac{1}{\lambda_s} \quad (2.9)$$

2.1.3. Ý nghĩa

Cùng với sự phát triển mạnh mẽ của khoa học kỹ thuật trong lĩnh vực công nghệ, các hệ thống tính toán được tạo ra giúp thay thế hoặc hỗ trợ con người, mang lại rất nhiều ứng dụng và lợi ích cho kinh tế, cuộc sống toàn cầu.Hệ thống được đặc trưng bởi một số lượng lớn các yếu tố thành phần, có cấu trúc phức tạp với các chương trình tính toán, điều khiển các hoạt động của nó. Đây chính là những hệ thốngcó tính ứng dụng cao, tham gia vào trong tất cả các lĩnh vực của đời sống, là toàn bộ cơ sở hạ tầng của xã hội hiện đại.

Tuy nhiên, nếu không đảm bảo được độ tin cậy thì hệ thống đó coi như không tồn tại, mà từ thực tế cho thấy việc thao tác sai hay sai lầm trong thiết kế chế tạo thiết bị có thể xảy ra bất kỳ lúc nào, điều này dẫn đến những nguy cơ tiềm tàng xảy ra đối với mỗi hệ thống như cấu trúc hệ thống bị phá vỡ, hệ thống hoạt động không chính xác... Như vậy, việc tìm ra những phương pháp hiệu quả nhằm nâng cao độ tin cậy của hệ thống cần được hết sức chú trọng và cần thiết, từ khâu thiết kế, đưa vào thử nghiệm, sản xuất và hoạt động.

Để đưa ra các phương pháp nhằm nâng cao độ tin cậy của các hệ thống thì chúng ta phải dựa trên các thông số đánh giá độ tin cậy của hệ thống. Các phương pháp đánh giá độ tin cậy của hệ thống liên quan đến các vấn đề về sản xuất, lập trình, dự toán, chi phí bảo trì và các chi phí tối thiểu cấu hình hệ thống. Khi biết được các thông tin về độ tin cậy của hệ thống có thể giúp chúng ta có được kế hoạch bảo trì, lập kế hoạch dự phòng để nâng cao độ tin cậy của hệ thống tránh được các lỗi sự cố có thể xảy ra.

Phương pháp dự phòng nâng cao độ tin cậy hệ thống bằng cách đưa ra các đối tượng dư thừa là nguồn lực bổ sung và cơ hội cần thiết tối thiểu để các đối tượng có thể thực hiện chức năng, nhiệm vụ của mình. Qua đó đảm bảo hệ thống vẫn hoạt động bình thường khi xuất hiện từ chối của hệ thống trong các thành phần của nó. Có nhiều phương pháp dự phòng: dự phòng cấu trúc, dự phòng thông tin và dự phòng thời gian.

- *Dự phòng cấu trúc:* còn gọi là dự phòng phần cứng, cung cấp cho việc sử dụng quá mức các thành phần trong hệ thống; thành phần chính trong yêu cầu tối thiểu tùy chọn của hệ thống được cung cấp thêm các yếu tố, thành phần bổ sung, thiết bị hoặc thay vì có một thì hệ thống được cung cấp thêm một hệ thống khác giống như nhau.

- *Dự phòng thông tin:* đó là kỹ thuật sử dụng phổ biến trong các phần tử, hệ thống được cung cấp sử dụng thông tin dư thừa. Ví dụ cơ bản của nó là chuyển lặp đi lặp lại của cùng một thông điệp trên một kênh truyền thông. Trang thiết bị kỹ thuật số mã tự điều chỉnh được áp dụng, việc tìm kiếm và sửa lỗi xuất hiện trong kết quả thất bại của thông điệp và thất bại của phần cứng do đó trong trường hợp này hoạt động của các thiết bị bình thường không bị hỏng.

- *Dự phòng thời gian:* là phương pháp làm cho hệ thống hiện thời sử dụng thời gian hệ thống dự phòng, mở rộng lớp phần cứng, thiết bị tự động và thông tin thiết bị đo được sử dụng trong phần cứng. Thời gian dự trữ có thể được sử dụng không chỉ sửa chữa và chuyển đổi cung cấp phần cứng mà còn trên những thất bại dịch vụ, loại bỏ các hậu quả của thất bại bằng cách lặp lại một số công việc và chờ đợi xử lý những thất bại để hệ thống có thể hoạt động.

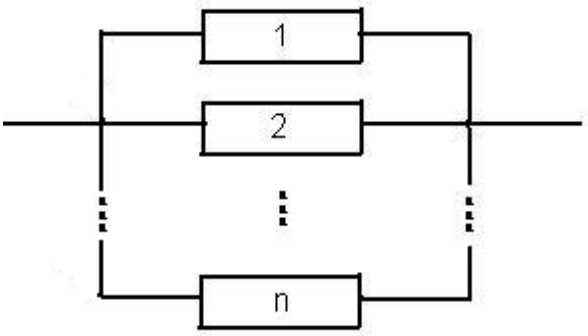
Dự phòng đóng một vai trò quan trọng trong việc tăng cường hệ thống đáng tin cậy. Dự phòng cho phép một hệ thống được hoạt động ngay cả khi một số phần tử và mối liên kết đã thất bại, do đó làm tăng độ tin cậy và tính sẵn sàng của nó. Các phần tử dự phòng có thể được phân loại là ở dạng hoạt động hoặc chế độ chờ. Các phần tử dự phòng tích cực hoạt động đồng thời thực hiện chức năng tương tự. Yếu tố trong dự phòng được thiết kế để hoạt động dự phòng hoặc có thể

được chuyển thành dịch vụ khi một phần tử hoạt động không thành công. Độ tin cậy của các hệ thống tăng lên cùng với số lượng phần tử dự phòng (giả sử rằng các cảm biến và các thiết bị chuyển tiếp của các phần tử dự phòng đang làm việc một cách hoàn hảo và các thành phần không dự phòng được thay thế trước khi hệ thống bị thất bại).

Trong thiết kế các thành phần của hệ thống có thể thấy rằng sử dụng các phần tử dự phòng là cách nhanh nhất để cải thiện độ tin cậy của hệ thống nếu không có đủ thời gian để tìm hiểu và lựa chọn thay thế hoặc nếu một phần hệ thống đã được thiết kế. Sử dụng các giải pháp dự phòng có chí phí thấp hơn nhiều so với việc chi phí thiết kế lại để cải thiện độ tin cậy.

Một trong những hình thức thường được sử dụng dự phòng là chế độ chờ dự phòng, các phương pháp dự phòng cấu trúc bao gồm: hệ thống dự phòng cấu trúc có tải (còn gọi là dự phòng nóng), hệ thống dự phòng cấu trúc không tải (còn gọi là dự phòng lạnh), hệ thống dự phòng cấu trúc nhẹ tải (còn gọi là hệ thống dự phòng ấm) và hệ thống dự phòng bảo vệ tích cực.

Hệ thống dự phòng có tải:Hệ thống pháp dự phòng có tải [3, 5, 22] là hệ thống dự phòng nóng, trong đó các phần tử cơ bản và phần tử dự phòng chịu tải như nhaucả trước và sau khi chúng bước vào trạng thái làm việc.



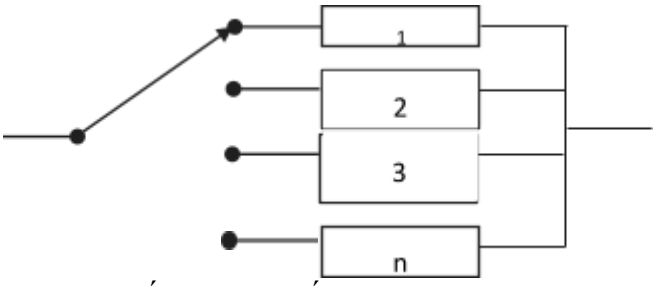
Hình 2. 1. Cấu trúc hệ thống dự phòng song song (dự phòng nóng).

Khi một phần tử hỏng, phần tử khác đang ở trạng thái dự phòng được đưa vào trạng thái làm việc nhờ một bộ chuyển tiếp. Giả thiết rằng bộ chuyển tiếp hoạt động tin cậy tuyệt đối và khoảng thời gian chuyển tiếp không đáng kể. Một phần tử dự phòng bắt đầu làm việc thay thế phần tử bị hỏng vẫn giữ nguyên chế độ tải trọng của mình. Do đó độ tin cậy của nó không phụ thuộc vào thời điểm chuyển tiếp nó từ trạng thái dự phòng sang trạng thái làm việc.

Hệ thống dự phòng có tải chỉ bị hỏng khi tất cả các phần tử đều bị hỏng cho tới phần tử cuối cùng.

- Hệ thống dự phòng không tải:

Hệ thống dự phòng không tải [3, 5, 22] còn gọi là hệ thống dự phòng lạnh, trong đó gồm các phần tử dự phòng không chịu tải chừng nào chúng chưa chuyển vào trạng thái làm việc thay cho phần tử bị hỏng.



Hình 2. 2. Cấu trúc hệ thống dự phòng không tải (dự phòng nguội).

Giả sử thừa nhận phần tử không bị hỏng khi nó còn ở trạng thái dự phòng không ảnh hưởng tới độ tin cậy của phần tử trong trạng thái làm việc. Vì vậy, ta thừa nhận rằng phần tử không bị hỏng khi nó còn ở trạng thái dự phòng và trạng thái dự phòng không ảnh hưởng tới độ tin cậy của phần tử trong trạng thái làm việc.

- Hệ thống dự phòng nhẹ tải:

Hệ thống dự phòng nhẹ tải [3] còn gọi là hệ thống dự phòng ấm trong đó các phần tử chỉ chịu một phần tải trong khi nó ở trạng thái dự phòng và bắt đầu chịu tải hoàn toàn khi nó bước vào trạng thái làm việc thay cho phần tử bị hỏng. Trong trạng thái dự phòng phần tử có thể bị hỏng tuy nhiên với xác suất nói chung nhỏ hơn trong trạng thái làm việc.

Để nâng cao độ tin cậy của hệ thống thì kết hợp giữa các phương pháp dự phòng cấu trúc hay còn gọi là dự phòng phần cứng và phương pháp nâng cao độ tin cậy của phần mềm thì hệ thống sẽ hoạt động sẽ đảm bảo được độ tin cậy.Độ tin cậy phần mềm là một phần quan trọng trong chất lượng phần mềm. Tuy nhiên nâng cao độ tin cậy phần mềm là khó khăn, khó khăn bắt nguồn từ sự hiểu biết và nắm bắt đầy đủ những thông tin về độ tin cậy phần mềm nói chung và đặc điểm của từng phần mềm nói riêng.

2.2. Khái quát về cơ chế RAID vàRAID đối với bài toán an toàn dữ liệu cho hệ thống máy

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 bảo vệ.

RAID là hệ thống đĩa được tạo ra nhằm mục đích tăng cường tốc độ truy cập dữ liệu hệ thống lưu trữ, tăng cường độ tin cậy về mặt dữ liệu. Tốc độ chuyển tải dữ liệu tăng lên khi các dữ liệu được chia đều cho các đĩa cứng hoạt động đồng thời.

RAID có những lợi thế như sau:

- Tính dự phòng

- Tính hiệu quả cao

- Giá thành thấp

RAID có tính dự phòng là nhân tố quan trọng nhất trong quá trình phát triển RAID cho môi trường máy chủ (Đặc điểm không có ở giải pháp mã hóa). Dự phòng cho phép sao lưu dữ liệu bộ nhớ khi xảy ra sự cố. Khi một ổ cứng trong dãy bị trục trặc, nó cho phép hoán đổi sang ổ cứng khác mà không cần

tắt cả hệ thống hoặc có thể sử dụng ổ cứng dự phòng. Phương pháp dự phòng phụ thuộc vào phiên bản RAID sử dụng. Về cơ bản khi hiệu quả sẽ tăng cao rõ khi sử dụng các phiên bản RAID mạnh. Hiệu quả cũng tùy thuộc vào số lượng ổ cứng được liên kết với nhau và các mạch điều khiển.

RAID giá thành thấp với mục tiêu là cung cấp bộ nhớ tốt hơn cho hệ thống so với việc sử dụng riêng biệt các ổ đĩa có dung lượng lớn.

Sử dụng phần lý thuyết ở mục trên 2.1. **Tổng quan về các phương pháp nâng cao độ tin cậy hệ thống** ta có thể thấy rõ được tầm quan trọng của yếu tố dự phòng đối với việc nâng cao độ tin cậy của một hệ thống bất kỳ. Mặt khác, phương pháp đánh giá độ tin cậy của hệ thống mang tính kinh tế rất cao, nó liên quan đến các vấn đề về sản xuất, lập trình, dự toán, chi phí bảo trì và các chi phí tối thiểu cấu hình hệ thống.

Như vậy, RAID có tính dự phòng và RAID có giá thành hợp lý, có thể sử dụng RAID vào việc giải quyết các bài toán về độ tin cậy, cụ thể xét ở khía cạnh luận văn này, RAID có những đặc điểm rất phù hợp để sử dụng vào giải pháp nâng cao an toàn bảo mật cho dữ liệu được lưu trữ trên hệ thống dịch vụ đám mây.

2.2.1. Các loại RAID

Lần đầu tiên RAID được phát triển năm 1987 tại trường Đại học California ở Berkeley với những đặc điểm chỉ ghép các phần đĩa cứng nhỏ hơn thông qua phần mềm để tạo ra một hệ thống đĩa dung lượng lớn hơn thay thế cho các ổ cứng dung lượng lớn giá đắt thời bấy giờ.

Mặc dù hiện nay không tồn tại nữa, nhưng Hội đồng tư vấn phát triển RAID (RAID Advisory Board: Viết tắt là RAB) đã ra thành lập tháng 7 năm 1992 để định hướng, lập ra các tiêu chuẩn, định dạng cho RAID. RAB đã phân ra các loại cấp độ RAID (level), các tiêu chuẩn phần cứng sử dụng RAID. Cấp độ ở đây không được hiểu rằng cứ cấp độ cao là cao cấp hoặc là "đời sau", mà chúng chỉ phân biệt rằng giữa loại RAID này và loại RAID khác.

RAID có rất nhiều chuẩn như RAID 0 (striping), RAID 1 (Mirror), RAID 0+1, RAID 10, RAID 5(Parity), RAID 50 … Với các dòng mainboard desktop cao cấp và các main server 1 way và 2 way có hỗ trợ RAID thì thường dùng RAID 0, 1, 5 RAID 0+1, RAID 10.

Các loại Raid thường dùng:

- ***Raid 0***

+ RAID 0 sử dụng kỹ thuật "striping": Phân chia khối dữ liệu đơn và trải chúng qua các ổ cứng, cụ thể chia những file ghi trên kĩa thành nhiều mẫu (ghi xen kẽ) và ghi mỗi mẫu trên những ổ cứng khác nhau, làm tăng hiệu quả thực thi. Có thể ghi được hai khối dữ liệu trên cùng ổ cứng, hơn hẳn so với một ổ cứng như trước.

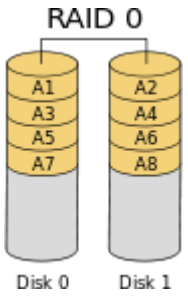
+ Thiết bị: cần ít nhất 2 ỗ đĩa cứng vật lý.Tổng quát: n đĩa(n >= 2) đĩa cứng cùng loại.

+ Hoạt động: Dữ liệu sẽ chia làm nhiều phần bằng nhau và được lưu trữ trên từng đĩa cứng. Mỗi đĩa sẽ chứa 1/n dữ liệu.

+ *Ưu điểm:* Dung lương tăng n lần ỗ đĩa đơn vật lý. Tăng tốc độ đọc ghi dữ liệu. Mỗi đĩa chỉ cần đọc/ghi 1/n dữ liệu được yêu cầu. Theo lý thuyết tốc độ sẽ tăng n lần.

+ *Khuyết điểm:* Tính an toàn dữ liệu thấp, rủi ro cao. Nếu 1 trong bất kỳ ổ đĩa đơn vật lý bị hư thì dữ liệu còn lại ở các ỗ đĩa vật lý khác cũng không còn sử dụng được nữa. Xác suất mất dữ liệu sẽ tăng n lần so với dùng ổ đĩa đơn.

+ Ứng dụng đề nghị: Chỉnh sửa và sản xuất phim, chỉnh sửa hình ảnh và một vài ứng dụng khác đòi hỏi băng thông đọc/ghi cao.



Hình 2. 3. ***RAID 0***

- ***RAID 1***

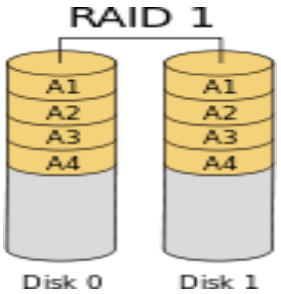
+ Thiết bị: như RAID 0

+ Hoạt động: (n=2)Dữ liệu được ghi vào 2 ổ giống hệt nhau (Mirroring). Trong trường hợp một ổ cứng đơn bị trục trặc, ổ cứng đơn còn lại còn lại sẽ tiếp tục hoạt động bình thường. Bạn có thể thay thế ổ đĩa cứng đơn bị hỏng mà không phải lo lắng đến vấn đề thông tin thất lạc.

+ *Ưu điểm:* Tốc độ đọc/ghi và dung lượng lưu trữ bằng ổ cứng đơn. Đảm bảo an toàn dữ liệu. Raid 1 là chuẩn RAID không thể thiếu đối với người quản trị mạng hoặc những ai phải quản lý nhiều thông tin quan trọng.

+ *Khuyết điểm:* Không phải là lựa chọn cho người say mê tốc độ. Hiệu năng không phải là yếu tố hàng đầu. Không gia tăng dung lượng lưu trữ.

+ Ứng dụng đề nghị: ứng dụng về tài chính và các ứng dụng dành cho văn phòng.



Hình 2. 4. RAID 1

- Raid 0+1

Hệ thống dùng RAID0 và RAID1 cùng một lúc. Nó cần 04 ổ cứng giống hệt nhau . Nếu một ổ cứng hỏng thì hệ thống trở thành RAID0

- RAID 2, RAID 3 : các chuẩn này giống như RAID1 nhưng nó dùng thêm một đĩa để ghi nhận và sửa các lỗi trong việc phân bồ dữ liệu ở các đĩa (Error checking and correction), làm nâng cao hơn độ tin cậy của hệ thống.

- RAID 4, RAID 5: cũng giống như RAID 2,3 nhưng thông tin về lỗi phân bố trên được chia đều trên các đĩa thành viên của các dãy. Điều này sẽ làm tiết kiệm thời gian truy cập các đĩa hơn so với hai chuẩn RAID 2 , 4.

- ***Raid 5:***

+ Thiết bị: cần ít nhất 3 ổ đĩa cứng vật lý riêng biệt.

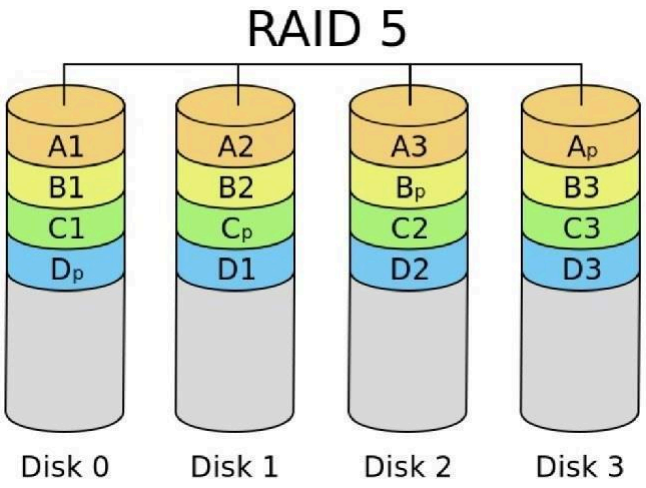
+ Hoạt động: Dữ liệu và bản sao lưu được lưu trữ trên tất cả các ổ đĩa cứng vật lý. Nguyên tắc này khá phức tạp. Với n = 3(Xem hình minh họa bên dưới). Ví dụ: ta có 8 đoạn dữ liệu(6 Block) tương ứng: Block 1, Block 2, Block 3, Block 4, Block 5, Block 6. Block 1 và Block 2 sẽ được lưu vào ổ đĩa cứng vật lý(1, 2) và bản sau lưu dữ liệu của chúng được ghi vào đĩa cứng vật lý 3. Block 3 và Block 4 sẽ được lưu vào ổ đĩa cứng vật lý(1, 3) và bản sau lưu dữ liệu của chúng được ghi vào đĩa cứng vật lý 2. Block 5 và Block 6 sẽ được lưu vào ổ đĩa cứng vật lý(2, 3) và bản sau lưu dữ liệu của chúng được ghi vào đĩa cứng vật lý 1.

+ *Ưu điểm:* Tốc độ được cải thiện và tính an toàn dữ liệu được đảm bảo. Thay thế 1 ổ đĩa cứng bị hư dễ dàng. Dung lượng lưu trữ = n – 1. Tức là nếu bạn sử dụng 3 ổ đĩa vật lý 500GB thì dung lượng cuối cùng được sử dụng là 1TB.

+ *Khuyết điểm:* Hư 2 trong 3 ổ thì xin chia buồn cùng bạn. Mặc dù điều đó khó xảy ra nhưng không phải là không xảy ra. Chẳng hạn: nguồn dỏm bị sốc điện.

+ Ứng dụng đề nghị: ứng dụng về tài chính và các ứng dụng dành cho văn

phòng.



Hình 2. 5. RAID 5

- ***Raid 6:***

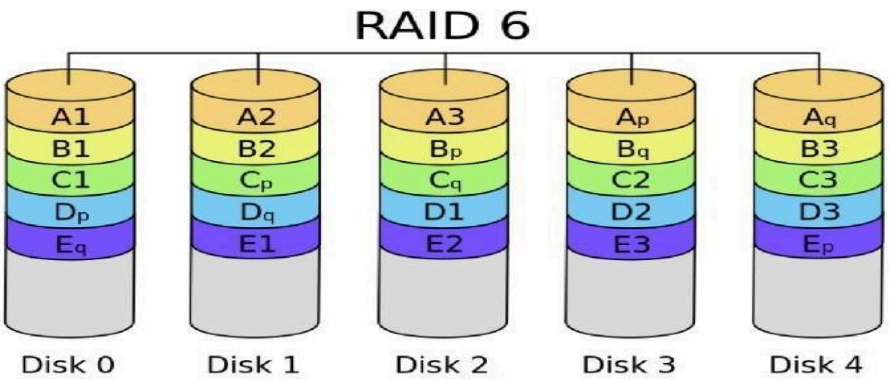
+ Thiết bị: cần ít nhất 4 ổ đĩa cứng vật lý riêng biệt.

+ Hoạt động: Raid 6 được phát triển từ Raid 5. Cơ chế hoạt động phần nào giống Raid 5 nhưng lập lại nhiều hơn số lần phân tách dữ liệu.

+ *Ưu điểm:* như ta đã biết đối với Raid 5 nếu 2 trong 3 ổ đĩa vật lý bị hư thì ta sẽ mất trắng dữ liệu còn Raid 6(2 trong 4 ổ đĩa vật lý bị hư) thì vẫn đảm bảo an toàn dữ liệu.

+ *Khuyết điểm:* chi phí cao. Chưa cải thiện nhiều về tốc độ đọc/ghi

+ Ứng dụng đề nghị: nhà cung cấp nơi đặt website, máy chủ lưu trữ dữ liệu, database server,…



Hình 2. 6. RAID 6

- ***Raid 10:***

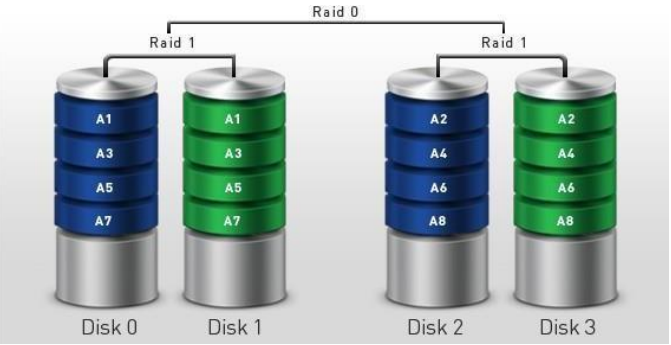
+ Thiết bị: cần ít nhất 4 ổ đĩa cứng vật lý riêng biệt.

+ Hoạt động: Là sự kết hợp Raid 0 + Raid 1. Tổng hợp các ưu điểm của đàn anh. Dữ liệu được ghi đồng thời lên 4 đĩa cứng với 2 ổ dạng Striping tăng tốc và 2 ổ dạng Mirroring sao lưu.

+ *Ưu điểm:* Tăng tốc đọc/ghi. Performance > Raid 5 và Raid 6. Tính an toàn dữ liệu cao hơn Raid 0, Raid 1 và Raid 5. Cho phép tối đa 2 ổ đĩa cùng lúc bị hỏng ở 2 pair khác nhau.

+ *Khuyết điểm:* chi phí đầu tư cao.

+ Ứng dụng đề nghị: Các ứng dụng về Database, ứng dụng tài chính và các ứng dụng đòi hỏi tính an toàn dữ liệu và tốc độ cao.



Hình 2. 7. RAID 10

Mỗi dạng RAID đều có ưu, nhược điểm riêng. Ngoài việc phụ thuộc giới hạn của từng loại thiết bị *(vd: thiết bị hỗ trợ tối đa 2 ổ cứng thì không thể chạy RAID 5, 6, 10)*, việc lựa chọn RAID nào hoàn toàn phụ thuộc vào kinh nghiệm, đôi khi là…sở thích của người quản trị. Bạn chưa biết phải chọn RAID nào phù hợp với hệ thống của mình, có 3 tiêu chí chính bạn cần quan tâm:

- **Tốc độ:** Bạn có nhu cầu tăng tốc độ truy xuất dữ liệu?
- **Độ an toàn:** hệ thống được phép hư bao nhiêu HDD, mà vẫn chạy bình thường không mất dữ liệu?

- **Dung lượng lưu trữ:** Bạn sẵn sàng đánh đổi bao nhiêu dung lượng lưu trữ để đổi lấy sự an toàn cho dữ liệu?

RAID hỗ trợ các hệ thống với nhiều tiện ích khác nhau tùy vào phiên bản được áp dụng. Đa số khách hàng sử dụng sẽ lựa chọn RAID 0 để tăng tốc độc thực thi mà không làm giảm không gian bộ nhớ. Chủ yếu là do dư thừa chưa phải là vấn đề chính cho người sử dụng trung bình. Thật ra, hầu hết các hệ thống máy tính chỉ cung cấp RAID 0 hay RAID 1. Chi phí để thực hiện RAID 0+1 hay RAID 5 là quá đắt đối với những khách hàng trung bình và chỉ được áp dụng cho các trạm làm việc hay các hệ thống máy chủ cấp cao.

2.2.2. Đánh giá độ tin cậy của các hệ thống RAID

Mỗi cấp độ RAID có mẫu khác nhau dẫn đến sự khác nhau về hiệu suất và khả năng dự phòng của từng hệ thống RAID.

RAID có thể nâng cao được tính tin cậy, tốc độ truy cập và dung lượng hệ thống vì RAID sử dụng 2 quy trình để chế tạo là kỹ thuật tạo lát đĩa và kỹ thuật soi gương đĩa. Kỹ thuật tạo lát đĩa điều khiển RAID cung cấp khả năng ghi và đọc song song các khối của cùng một đơn vị dữ liệu nhờ vậy tăng được tốc độ đọc ghi. Các dữ liệu cần ghi được chia thành các khối cùng kích thước và được ghi đồng thời vào các ổ đĩa độc lập. Tương tự, quá trình đọc, các khối dữ liệu cần đọc được đọc đồng thời từ các đĩa cứng độc lập, giúp giảm được thời gian đọc. Kỹ thuật soi gương đĩa giúp đạt độ tin cậy cao cho hệ thống lưu trữ. Theo đó, dữ liệu được chia thành các khối và mỗi khối được ghi đồng thời lên hai hay nhiều ổ đĩa độc lập. Như vậy, tại mọi thời điểm ta đều có nhiều bản sao dữ liệu trên các đĩa cứng độc lập, đảm bảo tính an toàn cao. Và giúp tăng dung lượng bằng các đĩa độ lập tham gia tạo RAID.

Với RAID 10 có nhiều ưu điểm, cấu hình phù hợp với máy chủ CSDL, nó là sự kết hợp giữa RAID 1 và 0, giữa tốc độ cao và tính tin cậy cao, vì thế rất phù hợp với các hệ thống máy chủ đòi hỏi tính an toàn cao, hiệu năng lớn như máy chủ CSDL.

2.2.3. Triển khai RAID

RAID có thể được triển khai ở 2 dạng:

- RAID cứng (Hardware RAID): Nếu các mã phần mềm được thực thi trên bộ xử lý on-board là Hardwar RAID.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 phứctạp.
- RAID mềm (Software RAID): Nếu các mã phần mềm được thực thi hoặc trên bộ xử lý (CPU) máy chủ thì là software RAID. 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 giảnhơn.

Để có thể RAID cho máy, cần tối thiểu một card điều khiển hay gọi là card raid máy chủ và hai ổ đĩa cứng giống nhau.

*** RAID mềm:**

Với RAID mềm sau khi cài xong HĐH, tiến hành dùng luôn Windows để thiết lập RAID - Windows based RAID. Nếu sử dụng Linux thì có sẵn mdadm utility. Ngày nay, đã và đang có khá nhiều software RAID được viết trên nền Linux, các mã nguồn mở và ngày càng chứng tỏ khả năng vượt trội.

Các software RAID dựa trên phần mềm chủ yếu được sử dụng với các máy lưu trữ gia đình, các máy chủ entry-level . Điểm chủ yếu để nhận diện là nó thực hiện tất cả các lệnh I / O và các thuật toán toán học RAID chuyên sâu trực tiếp trên các CPU của máy chủ lưu trữ. Chính điều này làm chậm hiệu suất hệ thống bằng cách tăng lưu lượng truy cập máy chủ qua PCI bus , sử dụng vào ngay luôn tài nguyên của hệ thống CPU, memory, Ưu điểm chính của software RAID là

giá thành rẻ hơn (nhiều software RAID cho free luôn) so với các lựa chọn thay thế RAID khác như hardware RAID có mức giá cao hơn nhiều.

*** RAID cứng:**

RAID cứng là một dạng card add-in. Loại card RAID controller này cắm vào một khe cắm bus chủ PCI. Giảm tải hệ thống máy chủ trong một số hoặc tất cả các lệnh I /O, dành các hoạt động tính toán RAID cho một hoặc nhiều bộ vi xử lý thứ cấp mà nó có.

Ngoài việc cung cấp những lợi ích chịu lỗi của một RAID thông thường , bộ điều khiển RAID cứng còn thực hiện các chức năng kết nối tương tự như bộ điều khiển trên máy chủ tiêu chuẩn. Và cũng bởi nhờ nó có riêng cho mình tài nguyên (CPU, memory,...) nên chúng thường cung cấp hiệu suất cao nhất cho tất cả các loại RAID, nó cung cấp tính năng chịu lỗi mạnh mẽ hơn đa dạng hơn RAID mềm.

Hình 2. 8. Ví dụ về RAID cứng

Với những kiến thức được nêu ở trên ta nhận thấy rõ tiềm năng của RAID, đặc biệt là tính dự phòng, tính sẵn sàng của dữ liệu. Qua quá trình tìm tòi, nghiên cứu, phân tích và lập luận, tôi đưa ra đề xuất giải pháp RBCS sẽ được trình bày chi tiết ở chương sau. Với giải pháp này, ta có thể dụng tuyệt đối những ưu điểm của giải pháp mã hóa dữ liệu và ứng dụng cơ chế RAID vào kết hợp để giải quyết những vấn đề hạn chế còn tồn đọng trong giải pháp mã hóa đám mây, bảo mật truy cập phân quyền.

CHƯƠNG 3:
GIẢI PHÁP LƯU TRỮ DỮ LIỆU TRÊN ĐÁM MÂY – RBCS VÀ ỨNG DỤNG VÀO THỰC TẾ DOANH NGHIỆP

3.1. Giải pháp RBCS

3.1.1. Giải pháp RBCS

RBCS (RAID Based Cloud Storage) là cơ chế lưu trữ dữ liệu trên các dịch vụ cloud do nhóm nghiên cứu đềxuất, 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 côngnhay 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.

3.1.2. Xây dựng quy trình bài toán thực tế doanh nghiệp:

*** Phát biểu bài toán**

Thực tế thấy rằng, điện toán đám mây được sử dụng rất nhiều, mang lại vai trò tiện ích to lớn đối với các doanh nghiệp lớn nhỏ hiện nay. Cùng với sự phát triển nâng cấp của dịch vụ điện toán đám mây thì các doanh nghiệp

lớn nhỏ cũng đang phát triển mạnh mẽ, nhu cầu sử dụng dịch vụ càng nhiều, những yêu cầu đặt ra ngày càng cao và tất yếu là việc bảo vệ dữ liệu lưu trữ, an toàn hệ thống được các nhà cung cấp dịch vụ hết sức coi trọng. Độ tin cậy cao, toàn vẹn là tiền đề cho sự phát triển bền vững của dịch vụ đám mây.

Một bài toán được đặt ra là: Một doanh nghiệp A ước lượng có một khối lượng thông tin dữ liệu rất lớn cần lưu trữ trên dịch vụ đám mây là n (Gb). Yêu cầu đặt ra là xây dưng quy trình cụ thể để giúp doanh nghiệp A có thể lưu trữ dữ liệu nhanh vàhiệu quả nhất mà vẫn đảm bảo được tính an toàn và tính kinh tế.

*** Xây dựng quy trình bài toán thực tế doanh nghiệp**

- Quy trình:

+ Bước 1: Với đầu vào của bài toán kích thước của tệp dữ liệu cần lưu trữ trên đám mây là n (Gb). Với khoảng kích thước dữ liệu ước lượng đó xác định ra sẽ băm ra bao nhiêu mảnh cho hợp lý (đặt là N mảnh).

+ Bước 2: Với N mảnh được băm ra, xác định sẽ có bao nhiêu tài khoản accout trên nhà cung cấp dịch vụ.

+ Bước 3: Dữ liệu đã được phân mảnh, mỗi mảnh sẽ được đưa lên lưu trữ trên các tài khoản khác nhau của các nhà cung cấp dịch vụ.

Đẩy lên đám mây để lưu trữ thì sẽ có nhu cầu lấy dữ liệu về máy để sử dụng. Lúc này các mảnh dữ liệu cần dùng sẽ được gộp lại và tải về máy người dùng.

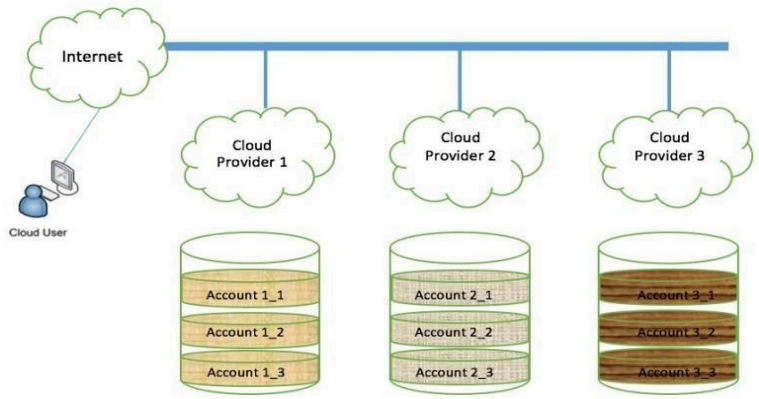
Với công nghệ phân mảnh, gộp dữ liệu này, hiện nay có nhiều phần mềm được sử dụng rất tiện ích. Ví dụ như: Tsplitter, winzip...

Trong phần 3.2 tiếp theo của luận văn sẽ chỉ rõ tường minh về cơ chế lưu trữ dữ liệu của RBCS giải quyết vấn đề theo quy trình nêu trên.

3.2. 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.

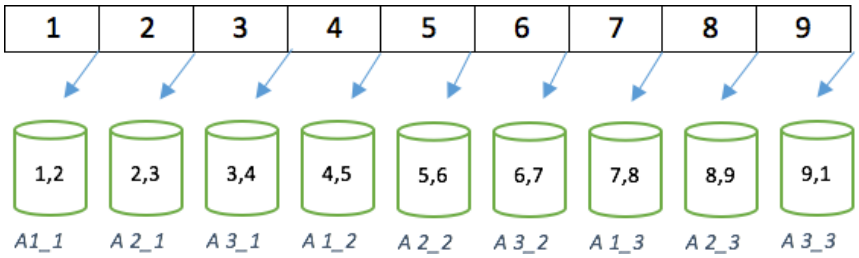
Dữ liệu sẽ được lưu trữ trong các dịch vụ đám mây bằng cách làm theo các quy trình sau: RBSC phân chia dữ liệu của người dùng và mã hóa một phần chúng sau đó lưu trữ dữ liệu vào các tài khoản khác nhau bằng phương pháp tương tự như mô hình RAID 5 hoặc RAID 6... ví dụ: Đối với mô hình lưu trữ RAID 5 dữ liệu được chia thành 9 khối và được lưu trữ trong 3 tài khoản trên 3 dịch vụ đám mây.



Hình 3. 1. 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.

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 3. 2. Phân mảnh dữ liệu và lưu trữ trên các kho dữ liệu đám mây

Trên Hình 3.2, 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 tàikhoản.

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.

Tương tự như cách lưu trữ RAID 5 nếu ta sử dựng mô hình lưu trữ RAID 6. Tập tin dữ liệu được chia thành 16 phần và được lưu trữ trên 16 account của 4 nhà cung cấp dịch vụ đám mây khác nhau. Các tệp tin được lưu trữ vào các tài khoản cloud theo quy tắc.

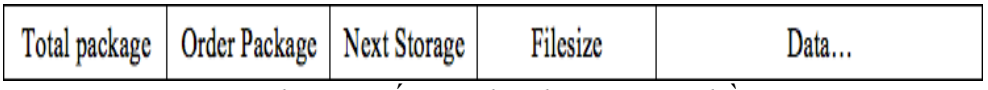
- Các tài khoản của cùng 1 nhà cung cấp dịch vụ được đặt xen kẽ nhau, theo quy tắc $n*i+m$ (trong đó n là số tài khoản trên cùng 1 dịch vụ, i là số lượt, m là thứ tự tài khoản).
- Trên mỗi tài khoản sẽ lưu trữ 3 mảnh dữ liệu kề nhau theo thứ tự đã phân mảnh.
- Mảnh đầu tiên và cuối cùng sẽ được lưu trên 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 2 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 vẫn khôi phục được.
- Nếu 2 nhà cung cấp dịch vụ cùng ngừng hoạt động, dữ liệu cũng vẫn khôi phục lại được.
- Dữ liệu chỉ không khôi phục được lại khi có 3 tài khoản liên tiếp hoặc 3 nhà cung cấp dịch của hệ thống ngừng cung cấp dịch vụ củng một lúc.

Như vậy từ 2 cách lưu trữ theo mô hình RAID 5 và RAID 6 ta thấy cách lưu trữ thông tin theo mô hình RIAD 5 và RAID 6 không khác nhau. Nhưng độ an toàn của thông tin khi lưu trữ theo mô hình RAID 6 được tăng lên đáng kể. Làm rõ hơn độ an toàn của bài toán lưu trữ thông tin sẽ sử dụng lý thuyết xác suất và độ tin cậy của hệ thống (tại mục 3.3 chương 3).

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 nếu muốn 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:



Hình 3. 3. Cấu trúc header của các 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.

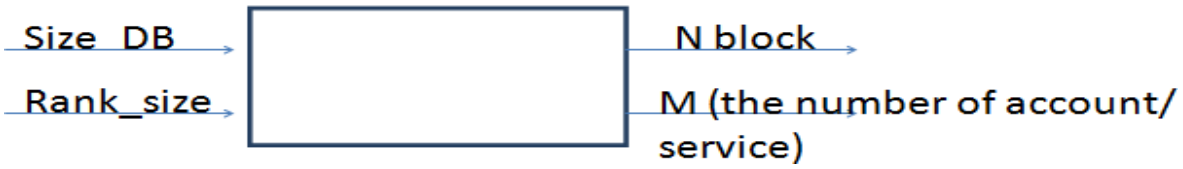
Do được phân mảnh và được lưu trữ phân tán trên các tài khoản của các kho dữ liệu khác nhau, nên dữ liệu của mỗi mảnh trong trường hợp bị truy cập trái phép cũng không thể hiện được nội dung của toàn bộ tài liệu. Tuy nhiên với các tập tin đơn giản không có cấu trúc header như tập tin txt, thì dữ liệu

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.

3.3. Mô hình bài toán dựa trên lý thuyết xác suất và độ tin cậy của hệ thống

Cho dữ liệu vào là tệp của một người dùng được lưu trữ trong dịch vụ lưu trữ đám mây với kích thước cụ thể.

Mô hình hoạt động của RBCS



Hình 3. 4. Mô hình hoạt động của RBCS

Kích thước của tệp dữ liệu vào là Size DB. RBCS sử dụng các tài khoản khác nhau trên nhà cung cấp và mỗi nhà cung cấp có chính sách riêng về giới hạn kích thước tệp, ta lấy kích thước tệp tối đa cho mỗi khối dữ liệu là giới hạn kích thước tệp nhỏ nhất giữa các nhà cung cấp dịch vụ.

Dữ liệu được chia thành các khối *N* và phân phối đều cho số tài khoản trên mỗi dịch vụ (*M*). Giả sử *M=N*

Gọi *P* là độ tin cậy ban đầu của hệ thống (P_s).

Giả sử một tập tin được phân thành 9 mảnh và sử dụng 3 tài khoản cloud trên mỗi dịch vụ

$\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3$: là độ tin cậy của mỗi mảnh. Khi đó độ tin cậy của hệ thống là:

$$P_{ss.} = \alpha_1 * \alpha_2 * \alpha_3 * \beta_1 * \beta_2 * \beta_3 * \gamma_1 * \gamma_2 * \gamma_3 . \quad (3.1)$$

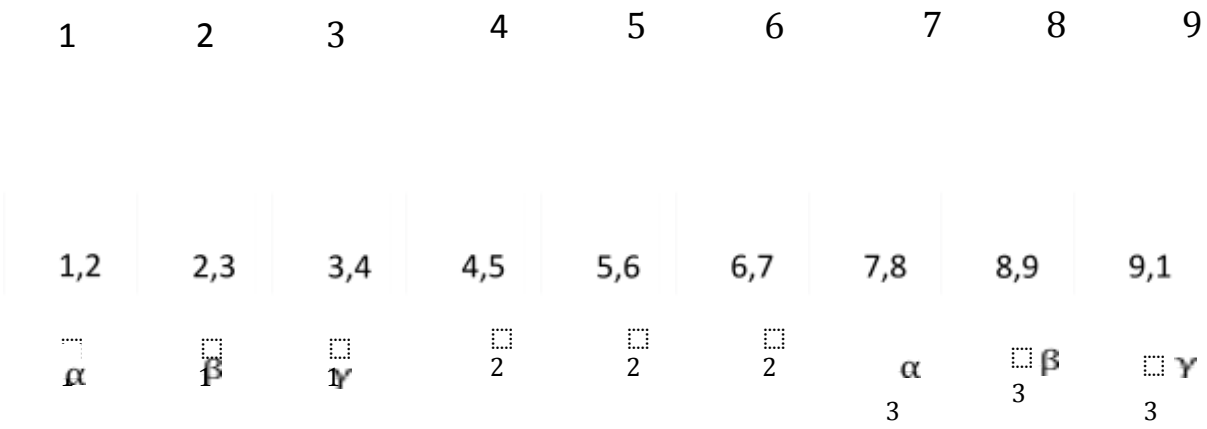
Giả sử: $\alpha_1 = \alpha_2 = \alpha_3$

$\beta_1 = \beta_2 = \beta_3 \gamma_1$

$= \gamma_2 = \gamma_3$

Thì độ tin cậy ban đầu của hệ thống là: $\boldsymbol{P_{ss}} = \alpha^3 * \beta^3 * y^3$ (3.2)

- Trường hợp 1: Mỗi tài khoản lưu trữ 2 mảnh dữ liệu, độ tin cậy ban đầu của hệ thống là:



Hình 3. 5. Mô hình hoạt động của RBCS

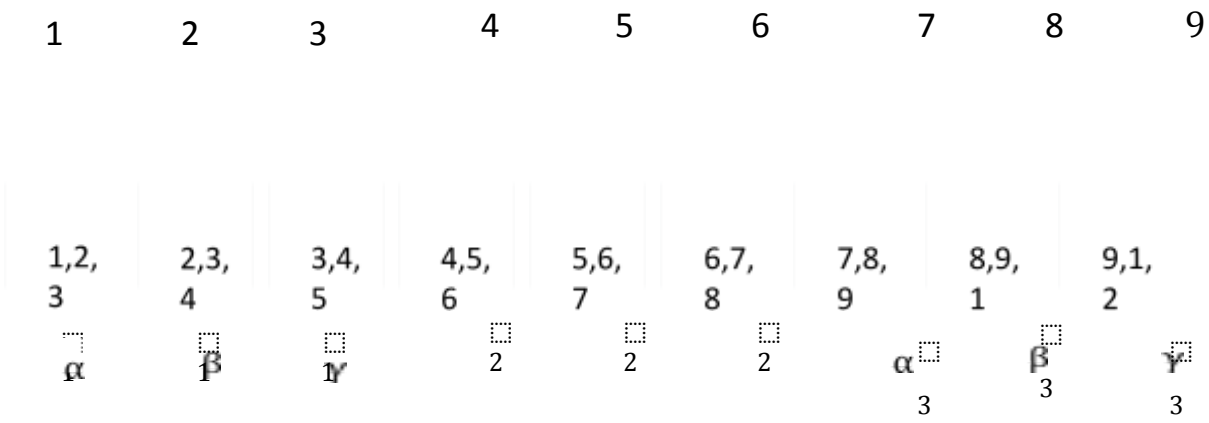
Trong trường hợp này, khi các tài khoản xen kẽ bị mất, dữ liệu có thể được khôi phục từ các tài khoản xung quanh. Nếu nhà cung cấp chấm dứt dịch vụ, dữ liệu sẽ được an toàn nhờ các tài khoản lân cận.

Nếu (*Pss*1) là độ tin cậy trong trường hợp 1, khi đó:

$$P_{ss1} = \alpha^3 * \beta^3 * \gamma^3 + (1-\alpha)^3 * \beta^3 * \gamma^3 + \alpha^3 * (1-\beta)^3 * \gamma^3 + \alpha^3 * \beta^3 * (1-\gamma)^3 + \alpha * \beta^2 * \gamma^2 * (1-\alpha)^2 * (1-\beta) * (1-\gamma) + \alpha^2 * \beta^2 * \gamma * (1-\alpha) * (1-\beta) * (1-\gamma)^2 + \alpha^2 * \beta * \gamma^2 * (1-\alpha) * (1-\beta)^2 * (1-\gamma) \quad (3.3)$$

- Trường hợp 2: Mỗi tài khoản lưu trữ 3 mảnh dữ liệu, độ tin cậy ban đầu của hệ thống là:

Trong Trường hợp 2, nếu có 2 tài khoản liền kề bị mất hoặc không thể tiếp cận được thì có thể lấy dữ liệu từ các tài khoản lân cận. Nếu một nhà cung cấp chấm dứt dịch vụ, bạn có thể sử dụng dữ liệu từ các tài khoản lân cận. Nếu (*Pss*2) là độ tin cậy trong trường hợp 2, khi đó:



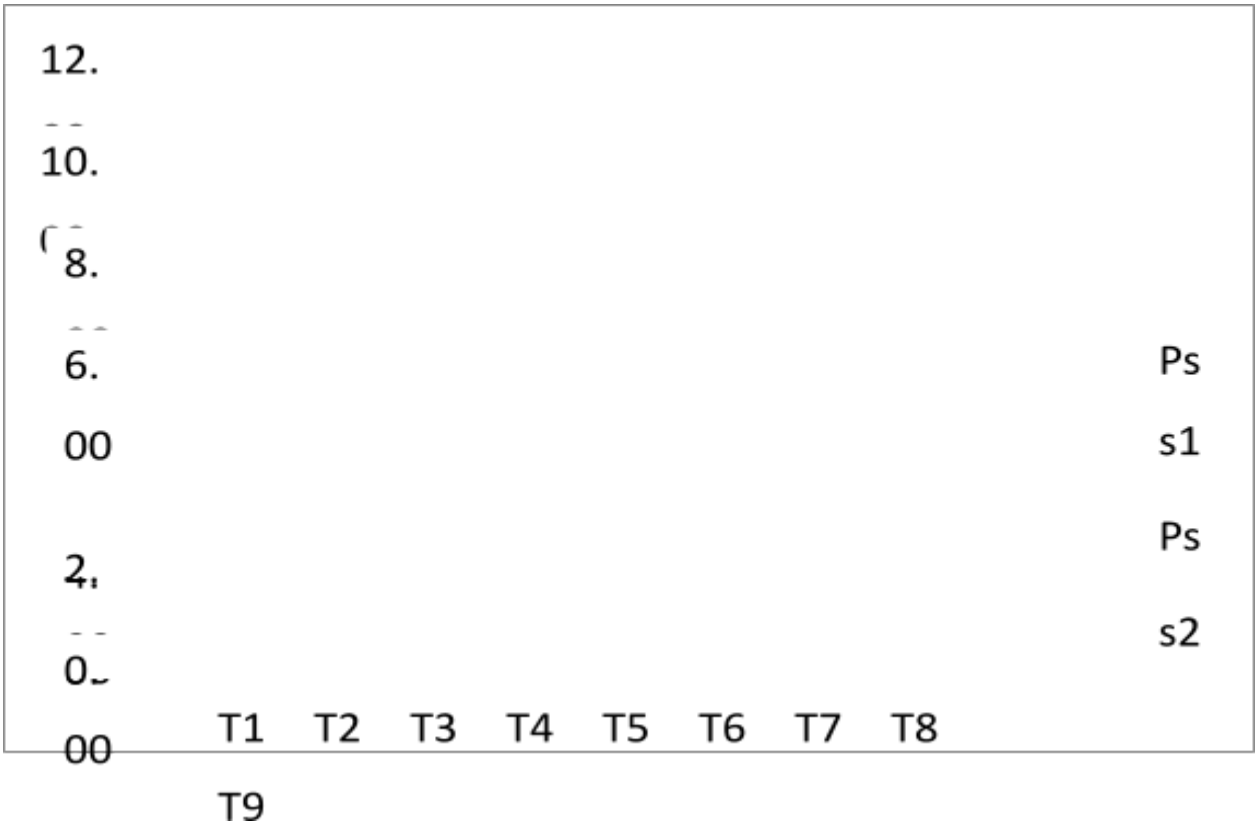
Hình 3. 6. Độ tin cậy của hệ thống trong trường hợp 2

$$\boldsymbol{P_{ss2}} = a^3 * \beta^3 * y^3 + (1-a)^3 * \beta^3 * y^3 + a^3 * (1-\beta)^3 * y^3 + a^3 * \beta^3 * (1-y)^3 + (1-a)^3 * (1-\beta)^3 * y^3 + a^3 * (1-\beta)^3 * (1-y)^3 + (1-a)^3 * \beta^3 * (1-y)^3 \quad (3.4)$$

Khi một nhà cung cấp gặp sự cố

A	*b*	*C*	*P*	*P1*	*P2*	*Độ tin cậy tăng %(TH1)*	*Độ tin cậy tăng % (TH2)*
0.9999	*0.9999*	*0.8*	*0.51*	*0.52*	*0.52*	*1.563*	*1.563*
0.99999	*0.99999*	*0.75*	*0.42*	*0.44*	*0.44*	*3.704*	*3.704*
0.99999	*0.99999*	*0.74*	*0.41*	*0.42*	*0.42*	*4.337*	*4.337*
0.99999	*0.99999*	*0.73*	*0.39*	*0.41*	*0.41*	*5.060*	*5.060*
0.99999	*0.99999*	*0.72*	*0.37*	*0.40*	*0.41*	*5.881*	*5.881*
0.99999	*0.99999*	*0.71*	*0.36*	*0.38*	*0.38*	*6.814*	*6.814*
0.99999	*0.99999*	*0.7*	*0.34*	*0.37*	*0.37*	*7.872*	*7.872*
0.99999	*0.99999*	*0.69*	*0.33*	*0.36*	*0.36*	*9.069*	*9.069*
0.99999	*0.99999*	*0.68*	*0.31*	*0.35*	*0.35*	*10.421*	*10.421*

Bảng 3. 1. Bảng so sánh độ tăng độ tin cậy của trường hợp 1

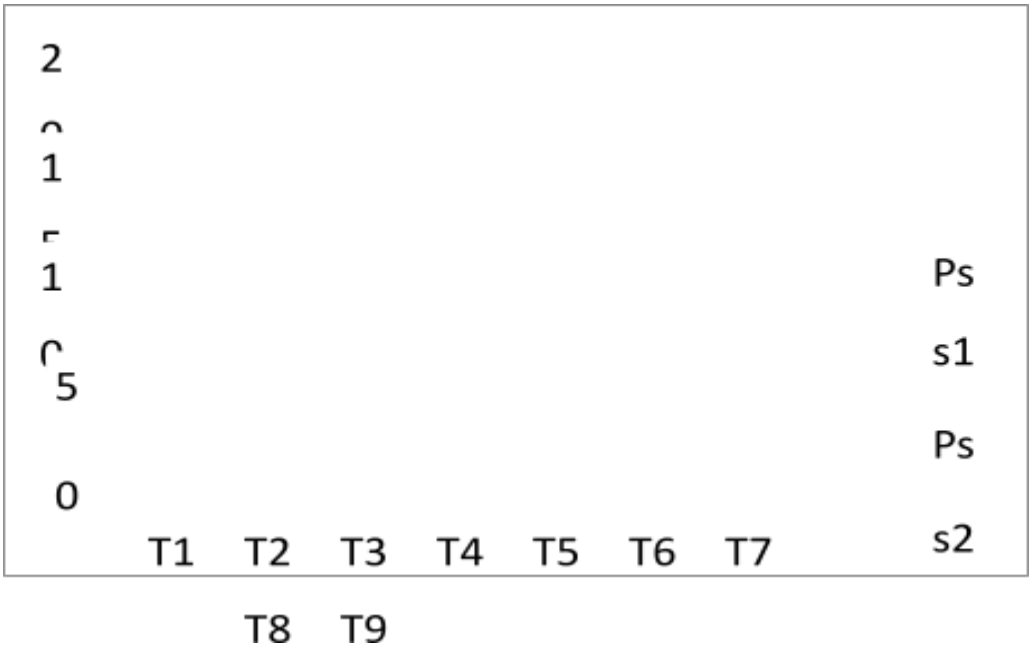


Hình 3. 7. Biểu đồ hiển thị độ tăng của độ tin cậy ở trường hợp 1

Khi hai nhà cung cấp gặp sự cố:

A	*b*	*c*	*P*	*P1*	*P2*	*Độ tin cậy tăng %(TH1)*	*Độ tin cậy tăng % (TH2)*
0.99999	*0.699*	*0.79*	*0.17*	*0.18*	*0.19*	*9.863*	*10.013*
0.99999	*0.698*	*0.78*	*0.16*	*0.18*	*0.18*	*10.343*	*10.525*
0.99999	*0.697*	*0.77*	*0.15*	*0.17*	*0.17*	*10.881*	*11.099*
0.99999	*0.696*	*0.76*	*0.15*	*0.16*	*0.17*	*11.482*	*11.744*
0.99999	*0.695*	*0.75*	*0.14*	*0.16*	*0.16*	*12.156*	*12.468*
0.99999	*0.694*	*0.74*	*0.14*	*0.15*	*0.15*	*12.910*	*13.281*
0.99999	*0.693*	*0.3*	*0.13*	*0.15*	*0.15*	*13.754*	*14.193*
0.99999	*0.692*	*0.72*	*0.12*	*0.14*	*0.14*	*14.699*	*15.217*
0.99999	*0.691*	*0.71*	*0.12*	*0.14*	*0.14*	*15.757*	*16.366*

Bảng 3. 2. Bảng so sánh độ tăng độ tin cậy của trường hợp 2



Hình 3. 8. Biểu đồ hiển thị độ tăng của độ tin cậy ở trường hợp 2

Giả sử hai trường hợp này có độ tin cậy cao, nhưng nếu một dịch vụ thay đổi chính sách bảo mật của mình hoặc vào thời điểm đó các hacker đang khai thác lỗ hổng, độ tin cậy của hệ thống sẽ giảm đáng kể. Sự thừa của hai trường hợp này sẽ trở nên khả thi hơn bằng cách nâng cao độ tin cậy lên 1,56% và 3,7%.

Ví dụ: Với α = 0.9999 tỷ lệ lỗi là 1/10000, β = 0.9999 tỷ lệ lỗi là 1/10000, γ = 0.8 tỷ lệ lỗi là 2000/10000, độ tin cậy của cả 2 trường hợp được cải thiện bằng 1.56% .

Với α = 0.99999 tỷ lệ lỗi là 1/100000, β = 0.99999 tỷ lệ lỗi là 1/100000, γ = 0.75 tỷ lệ lỗi là 25000/100000, độ tin cậy của cả 2 trường hợp được cải thiện 3.7%.

Khi sử dụng các mô hình này, độ tin cậy được cải thiện, vì vậy tính toàn vẹn dữ liệu được đảm bảo.

3.4. Ứng dụng bài toán thực tế tại Phòng Giáo dục và Đào tạo thị xã Đông Triều.

Do Phòng Giáo dục và Đào tạo Đông Triều nơi học viên công tác cần phải lưu trữ nhiều loại dữ liệu như: thông tin cán bộ giáo viên nhân viên toàn trong ngành, thông tin học sinh, thông tin cơ sở vật chất… Những dữ liệu này thường được lưu trữ trên các đĩa CD cứng nhưng thường chỉ được một đến 2 năm đĩa CD có thể bị hỏng gây mất dữ liệu. Học viên đã thử đưa dữ liệu lưu trữ cloud nhưng gặp phải gói dữ liệu quá lớn không đưa lên được và có thể một

ngày nào đó tên

truy cập có thể bị tin tặc cướp quyền quản lý, hoặc nhà cung cấp dịch vụ cloud ngừng hoạt động… gói dữ liệu có thể bị mất hoặc lộ thông tin. Vì vậy học viên đã giải quyết bài toán theo các bước sau:

- *Bước 1:* Nén tất cả dữ liệu cần lưu trữ thành 1 file;

- *Bước 2:* Chia file thành 9 file nhỏ nhờ phần mềm Splitter sau đó đặt tên file các file nhỏ lần lượt là DLPGD2016 (viết tắt của dữ liệu phòng giáo dục 2016). Đây là một mã khóa riêng mà nhà cung cấp dịch vụ cloud không biết và can thiệp được.

- *Bước 3:* Đưa lần lượt 9 file nhỏ đó lên 3 nhà cung cấp dịch vụ cloud

+ Nhà cung cấp dịch vụ google Driver lưu trữ bằng 3 account: account 1 lưu 2 mảnh DL;account 2 lưu 2 mảnh GD, account 3 lưu 2 mảnh 01)

+ Nhà cung cấp dịch vụ Drop Boxlưu trữ bằng 3 account: account 1 lưu 2 mảnh LP; account 2 lưu 2 mảnh D2, account 3 lưu 2 mảnh 16)

+ Nhà cung cấp dịch vụ Amazon Cloudlưu trữ bằng 3 account: account 1 lưu 2 mảnh PG; account 2 lưu 2 mảnh 20, account 3 lưu 2 mảnh 6D).

Khi 1 trong các account bị chiếm quyền ta vẫn truy cập và 2 account liền kề để khôi phục được dữ liệu. và dữ liệu tại các account đó không đủ để lấy được thông tin.

Khi một nhà cung cấp ngừng cung cấp dịch vụ ta vẫn khôi phục lại dữ liệu được nhờ hai nhà cung cấp còn lại được.

*** 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 dịchvụ 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. Hiện nay một số phần mềm đã được phát triển dựa trên nguyên lý mã hoá dữ liệu của người dùng trước khi đưa lêncloud:

Credeoncp là một ứng dụng mã hoá phía client cho các dịch vụ lưu trữ trên cloud, 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 FIPS140-2.

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 hệđiều 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.

KẾT LUẬN

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 luận văn tôi.

Sau thời gian tìm hiểu, nghiên cứu tài liệu và làm luận văn dưới sự hướng dẫn của thầy TS. Lê Quang Minh tôi đã hoàn thành luận văn với đề tài "***Nghiên cứu bảo vệ an toàn dữ liệu khi sử dụng dịch vụ lưu trữ điện toán đám mây***" Luận văn đã đạt được kết quả sau:

- Tìm hiểu, nghiên cứu đã đưa ra được những lý thuyết tổng quan xoay quanh dịch vụ lưu trữ đám mâybao gồm những kiến thức cơ bản về kiến trúc, các thành phần, các mô hình triển khai, mô hình dịch vụ của dịch vụ đám mây... Chỉ ra và phân tích những nhà cung cấp dịch vụ lớn hiện nay, đặc biệt đối với dịch vụ lưu trữ dữ liệu.

- Đưa ra luận điểm những vấn đề còn tồn tại, những lập luận và dẫn chứng về sự thiếu an toàn mất mát dữ liệu. Trình bày những vấn đề có mức độ nguy hại cao nhất trong điện toán đám mây.

- Trình bày chi tiết và phân tích ưu nhược điểm của giải pháp mã hóa dữ liệu, bảo mật truy cập phân quyền. Qua đó làm nổi bật lên tính cấp thiết, ý nghĩa thực tiễn của chủ đề luận văn thực hiện.

- Tìm hiểu về các phương pháp dự phòng nâng cao độ tin cậy của hệ thống. Sau đó, trình bày tổng hợp, phân tích kiến thức xoay quanh cơ chế RAID, triển khai RAID. RAID đối với bài toán an toàn dữ liệu cho hệ thống máy.

- Dựa trên những cơ sở lý thuyết nêu trên, luận văn đã đưa ra giải pháp mới RBCS, giải pháp này đã giải quyết vấn đề chính còn tồn tại ở những dịch vụ lưu trữ trên cloud hiện nay đó là: tính bảo mật và toàn vẹn cho dữ liệu ngườidùng.

Cụ thể:

+ Phát biểu bài toán. Xây dựng thành công quy trình giải quyết bài toán thực tế (đặc biệt sử dụng vào việc lưu trữ dữ liệu cho cá nhân tổ chức doanh nghiệp).

+ Mô hình hóa bài toán

+ Sử dụng toán học vào chứng minh được độ tin cậy của giải pháp, đưa ra bảng biểu và đồ thị để so sánh làm rõ mức độ cải thiện lớn về độ tin cậy của hệ thống khi sử dụng giải pháp.

TÀI LIỆU THAMKHẢO

Tiếng Việt

[1] Lê Quang Minh, Nguyễn Anh Chuyên, Lê Khánh Dương, Phan Huy Anh, Trịnh Thị Thu, "*Nghiên cứu về các cơ chế RAID và đề xuất giải pháp lưu trữ dữ liệu an toàn trên dịch vụ đám*"-Kỷ yếu Hội nghị Quốc gia lần thứ 9 về Nghiên cứu cơ bản và ứng dụng Công Nghệ thông tin (FAIR), Cần Thơ 2016.

[2] Trần Diên Hiển, Vũ Viết Yên (2005), *Nhập môn lý thuyết xác suất và thống kê toán*, Nhà xuất bản Đại học Sư phạm, Hà Nội, tr16, 31.

[3] PGS.TS Phan Văn Khôi (2001), *Cơ sở đánh giá độ tin cậy*, Nhà xuất bản Khoa học và Kỹ thuật, tr169-174, tr188-195

[4] Nguyễn Anh Khiêm, "*Nghiên cứu các phương pháp nâng cao độ tin cậy cho hệ thống tính toán qua cấu trúc hệ thống*". Luận văn Thạc sĩ , Trường Đại học Công nghệ, ĐHQGHN, 2014.

Tiếng Anh

[5] 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.

[6] Hassan, Qusay (2011). *"Demystifying Cloud Computing" (PDF)*. The Journal of Defense Software Engineering. CrossTalk. **2011** (Jan/Feb): 16–21. *Retrieved 11 December 2014*.

[7] Peter Mell and Timothy Grance (September 2011). *The NIST Definition of Cloud Computing* (Technical report). National Institute of Standards and Technology.Special publication 800-145.

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

[9] 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.

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

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

[12] 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.

[13] 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.

[14] 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.htm.

[15] 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.

[16] 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.

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

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

[19] 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.

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

[21] 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

Tiếng Nga

[22] Александр Майер, *Разработка методов повышения надежности процесса эксплуатации вычислительных систем*, 2008. - 31c, (Xây dựng các phương pháp nâng cao độ tin cậy của quá trình vận hành hệ thống máy tính).

[23] Шубин, Р.А, *Надёжность технических систем и техногенный риск*, 2012. -15c, (Độ tin cậy của hệ thống kỹ thuật và các nguy cơ công nghệ).

[24] LeQuangMinh (2007), "Анализ методов обеспечения отказоустойчивости и живучести вычислительных систем", Естественные науки и технологии- №5. (Phân tích các phương pháp bảo đảm độ tin cậy và độ hoạt động của hệ thống tính toán, Tạp chí "Khoa học tự nhiên và công nghệ", số 5-2007).

[25]. LeQuangMinh (2007),"Анализ эффективности применения методов повышения отказоустойчивости ИВС реального времени", Микроэлектроники и информатики, Тез. докл. Всероссийской конференции. (Phân tích hiệu quả của việc ứng dụng các phương pháp nâng cao độ tin cậy cho hệ thống thời gian thực có cấu trúc dạng cây. Hội thảo khoa học toàn LBNga, Mátxcơva).