**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG THÀNH PHỐ HỒ CHÍ MINH**

**KHOA HỆ THỐNG THÔNG TIN VÀ VIỄN THÁM**



**TIỂU LUẬN**

**MÔN : HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU**

**\*\*\***

**ĐỀ TÀI: SURVEY ON SECURING DATA STORAGE IN THE CLOUD**

Sinh viên thực hiện**: [NHÓM 12]**

**\*Nguyễn Đức Gia Hân**

**\*Nguyễn Thành Hậu**

Lớp**: 09\_ĐH\_CNTT03**

Giảng viên hướng dẫn**:** **ThS. Phạm Trọng Huynh**

***TP. Hồ Chí Minh, tháng 11 năm 2022***

LỜI CAM ĐOAN

Tôi xin cam đoan tất cả số liệu và dẫn chứng từ báo cáo đề tài tiểu luận SURVEY ON SECURING DATA STORAGE IN THE CLOUD là trên cơ sở tìm tòi và nghiên cứu của nhóm chúng em, không có bất kỳ sự sao chép hay gian lận nào. Mọi sự giúp đỡ cho việc xây dựng cơ sở lý thuyết đều được trình bày rõ ràng và chi tiết trong phần tài liệu trích dẫn hợp lệ.

Người viết lời cam đoan

Tp. HCM, ngày 12 tháng 12 năm 2022

*Nguyễn Thành Hậu*

*Nguyễn Đức Gia Hân*

MỤC LỤC

[**LỜI MỞ ĐẦU** 4](#_Toc119961574)

[**I - Lời Giới Thiệu** 5](#_Toc119961575)

[**II- Tổng Quan Về Bảo Mật Trong Lưu Trữ Đám Mây** 8](#_Toc119961576)

[**III-Tính Toàn Vẹn Dữ Liệu** 12](#_Toc119961577)

[**A.Giới thiệu về Bằng chứng về khả năng truy xuất và khả năng sở hữu dữ liệu** 15](#_Toc119961578)

[**B. Cải thiện khả năng xác minh công khai** 17](#_Toc119961579)

[**C. Cải thiện hiệu quả** 20](#_Toc119961580)

[**D.Cải thiện hỗ trợ dữ liệu động** 21](#_Toc119961581)

[**E. Tóm tắt** 22](#_Toc119961582)

[**IV. Bảo Mật Dữ Liệu** 23](#_Toc119961583)

[**A.Kiểm soát truy cập** 24](#_Toc119961584)

[***1)Kiểm soát truy cập bằng Mã hóa dựa trên thuộc tính***: 26](#_Toc119961585)

[***2)Kiểm soát truy cập dựa trên vai trò***: 28](#_Toc119961587)

[**B.Mã hóa có thể tìm kiếm.** 30](#_Toc119961588)

[**C.Mã hóa đồng hình hoàn toàn** 33](#_Toc119961589)

[**D.Cách tiếp cận bảo mật dữ liệu khác** 38](#_Toc119961590)

[**E. Tóm tắt** 39](#_Toc119961591)

[**V.Mức Độ Khả Dụng** 40](#_Toc119961592)

[**VI. KẾT LUẬN VÀ CÔNG VIỆC TƯƠNG LAI** 43](#_Toc119961593)

[**Tài liệu tham khảo** 44](#_Toc119961594)

# **LỜI MỞ ĐẦU**

Điện toán đám mây đã trở thành giải pháp tiên tiến nổi tiếng ngày nay; nhiều nhà nghiên cứu và công ty đang nắm bắt công nghệ hấp dẫn này với sự vội vàng. Trong thời gian chờ đợi, các thách thức về bảo mật và quyền riêng tư được đưa ra trong khi số lượng người dùng lưu trữ đám mây tăng lên nhanh chóng. Trong nghiên cứu này, chúng tôi tiến hành một cuộc khảo sát chuyên sâu về các hoạt động nghiên cứu gần đây về bảo mật lưu trữ đám mây kết hợp với điện toán đám mây. Sau khi xem xét tổng quan về hệ thống lưu trữ đám mây và vấn đề bảo mật của nó, chúng tôi tập trung vào bộ ba yêu cầu bảo mật chính, tức là tính toàn vẹn của dữ liệu, tính bảo mật của dữ liệu và tính khả dụng của dữ liệu. Đối với mỗi mục tiêu trong số ba mục tiêu bảo mật, chúng tôi thảo luận về những thách thức độc đáo mới mà các dịch vụ lưu trữ đám mây phải đối mặt, tóm tắt các vấn đề chính được thảo luận trong tài liệu hiện tại, kiểm tra và so sánh các phương pháp hiện có và mới nổi được đề xuất để đáp ứng những thách thức mới đó và chỉ ra các phần mở rộng có thể và các cơ hội nghiên cứu tương lai. Mục tiêu bài báo của chúng tôi là cung cấp kiến thức tiên tiến cho các nhà nghiên cứu muốn tham gia vào lĩnh vực mới đầy thú vị này.

**SURVEY ON SECURING DATA STORAGE IN THE CLOUD**

**(KHẢO SÁT VỀ BẢO MẬT LƯU TRỮ DỮ LIỆU TRÊN ĐÁM MÂY)**

**\*\*\***

# **I - Lời Giới Thiệu**

Những tiến bộ nhanh chóng trong giao tiếp băng thông rộng và mạng chuyển mạch gói tốc độ cao đã làm cho việc chia sẻ tệp lớn hiệu quả hơn nhiều trong hai thập kỷ qua. Những tiến bộ nhanh chóng trong giao tiếp băng thông rộng và mạng chuyển mạch gói tốc độ cao đã làm cho việc chia sẻ tệp lớn hiệu quả hơn nhiều trong hai thập kỷ qua. Nhìn chung, nhu cầu về các ứng dụng đa phương tiện, chẳng hạn như thư đa phương tiện, bản trình bày được dàn dựng, au-dio và chia sẻ video chất lượng cao, tài liệu cộng tác, đã tăng lên rất nhiều. Số lượng dữ liệu và tài nguyên máy tính được sử dụng bởi các ứng dụng đó cũng đã tăng lên theo cấp số nhân. Vì vậy, chi phí dịch vụ và hỗ trợ CNTT, chẳng hạn như đầu tư vào phần cứng và phần mềm mới, nhân sự để cài đặt và bảo trì đang tăng liên tục cho cả doanh nghiệp và người dùng cá nhân. Do đó, điện toán đám mây đã trở thành một mô hình mới hấp dẫn về cung cấp dịch vụ CNTT và hỗ trợ được thúc đẩy bởi lợi thế kinh tế và năng suất. Thay vì đầu tư vào phần cứng và phần mềm mới, cũng như duy trì các tài nguyên đó, người dùng có thể sử dụng các ứng dụng, cơ sở hạ tầng, máy chủ, lưu trữ, mạng và các tài nguyên điện toán khác có sẵn trong ’đám mây’, là một nhóm tài nguyên điện toán chia sẻ có thể dễ dàng truy cập thông qua kết nối mạng băng thông rộng. Mô hình cung cấp dịch vụ CNTT mới này cung cấp cho người dùng các tài nguyên tính toán dường như không giới hạn mà không cần mua trước và/hoặc duy trì chi phí bảo trì. Hơn nữa, nó cung cấp tính đàn hồi và tính linh hoạt theo yêu cầu trong việc sử dụng tài nguyên máy tính. Mô hình định giá tiện ích cho phép người dùng chỉ trả tiền cho việc sử dụng thực tế của họ.

Lưu trữ, là một trong những tài nguyên điện toán có ảnh hưởng và đòi hỏi cao nhất trong kỷ nguyên kỹ thuật số hiện nay, là một trong những tài nguyên đầu tiên được chuyển vào trong đám mây. Loại dịch vụ điện toán đám mây này, được gọi là lưu trữ đám mây, đại diện cho một mô hình kinh doanh trong đó nhà cung cấp dịch vụ thuê không gian trong cơ sở hạ tầng lưu trữ quy mô lớn của họ cho các tổ chức và cá nhân. Nó luôn là một trong những dịch vụ phổ biến nhất trong ngành điện toán đám mây. Là một phần mở rộng của trung tâm dữ liệu hoặc dịch vụ lưu trữ tệp truyền thống vào đám mây, lưu trữ đám mây có các đặc điểm riêng biệt bao gồm tự phục vụ theo yêu cầu, truy cập mạng băng thông rộng, ghép nối tài nguyên, độ co giãn nhanh chóng và sử dụng được đo lường để lập hóa đơn tiện ích. Bên cạnh những ưu điểm chính về tiết kiệm chi phí, lưu trữ đám mây có thể tạo điều kiện thuận lợi cho việc chia sẻ thông tin và cộng tác tác vụ, thúc đẩy tính di động và khả năng truy cập phổ biến của dữ liệu, cũng như cung cấp các giải pháp dễ dàng và thuận tiện cho một số vấn đề khác. Ví dụ: với mục đích khắc phục thảm họa, các tổ chức nên duy trì các bản sao lưu dữ liệu thứ cấp ngoài cơ sở. Việc lưu trữ dữ liệu nhạy cảm, chẳng hạn như dữ liệu tài chính, cá nhân hoặc y tế phải tuân theo ngày càng nhiều quy định và ràng buộc pháp lý. Lưu trữ đám mây được cung cấp bởi một nhà cung cấp dịch vụ tuân thủ quy định có thể giúp chủ sở hữu dữ liệu giảm bớt quy trình phức tạp.

Tuy nhiên, mô hình mới đầy hứa hẹn của điện toán đám mây đặt ra những thách thức riêng về hiệu suất, khả năng sẵn có, bảo mật và khả năng mở rộng (còn được gọi là PASS). Trong số những thách thức này, các vấn đề bảo mật đã được báo cáo là mối quan tâm lớn nhất ngăn cản các doanh nghiệp và tổ chức áp dụng các dịch vụ đám mây theo các nghiên cứu gần đây [15]. Do đó, bắt buộc phải cung cấp các chiến lược, công cụ và cơ chế bảo mật đáp ứng yêu cầu của người dùng trên đám mây. Bảo mật trong điện toán đám mây là một vấn đề phức tạp bao gồm nhiều khía cạnh bao gồm bảo mật vật lý, bảo mật cơ sở hạ tầng (máy tính phân tán, máy chủ và phần cứng khác), bảo mật dữ liệu, bảo mật mạng, bảo mật phần mềm. Hơn nữa, nó liên quan đến trách nhiệm và nghĩa vụ chung giữa các thành phần của dịch vụ đám mây. Thực thi bảo mật sẽ không thành công nếu không có thỏa thuận, lòng tin, quy định và sự phối hợp giữa các nhà cung cấp dịch vụ và người dùng đám mây.

Vì lưu trữ là một trong những cơ sở hạ tầng cốt lõi cần thiết trong đám mây, bảo mật dữ liệu trong lưu trữ là một trong những mối quan tâm chính của bất kỳ hệ thống điện toán đám mây nào, đặc biệt là trong các dịch vụ lưu trữ đám mây. Hậu quả của việc vi phạm bảo mật trong lưu trữ đám mây có thể gây tổn hại nghiêm trọng cho cả nhà cung cấp dịch vụ và người dùng. Nếu không có sự tin tưởng từ người dùng, nhà cung cấp dịch vụ có thể mất khách hàng của họ. Mặt khác, người dùng có dữ liệu quý giá bị mất hoặc thông tin nhạy cảm bị tấn công có thể bị mất hoặc hư hỏng không thể phục hồi. Đã có nhiều trường hợp được báo cáo là mối đe dọa về bảo mật lưu trữ đám mây. Nhiều nhà cung cấp dịch vụ hàng đầu, bao gồm Amazon, Window và Google, đã gặp phải sự ngắt kết nối các dịch vụ đám mây dựa trên web của họ vì những lý do khác nhau như mất điện, lỗi phần cứng và phần mềm. Ví dụ, máy chủ của Amazon Web Service đã bị sét đánh, gây ra sự phá hủy trên máy phát điện. Mặc dù Amazon đã chuyển dữ liệu thành công sang máy chủ dự phòng, dịch vụ vẫn dừng lại sau khi nguồn điện liên tục (UPS ) của họ bị mất. Đã xảy ra việc xóa email hàng loạt trong Gmail vào năm 2006; nhiều người dùng nhận thấy rằng họ đã mất email và thông tin liên hệ mà không có thông báo nào từ Google. Google không thể khôi phục các tài khoản sau khi người dùng phản hồi sự cố. Một sự cố khác xảy ra gần đây vào tháng 7 năm 2012. Vì một lỗ hổng bảo mật trên hệ thống kiểm soát truy cập, Dropbox, một dịch vụ lưu trữ đám mây phổ biến, đã bị tin tặc tấn công. Một số người dùng báo cáo rằng họ đã nhận được hàng tấn email spam và mật khẩu của một số người dùng thậm chí còn bị rò rỉ.

Mặc dù các yêu cầu bảo mật cho lưu trữ đám mây khác nhau với các ứng dụng và người dùng khác nhau, chúng có chung ba mục tiêu cơ bản như bất kỳ hệ thống thông tin máy tính nào [31]: tính toàn vẹn, tính bảo mật và tính khả dụng. Nhiều công cụ khác nhau đã được phát triển để đạt được những mục tiêu này, chẳng hạn như xác thực, kiểm soát truy cập, mã hóa, chứng nhận, thính giác, chữ ký số. Bài báo này nhằm cung cấp một nghiên cứu kỹ lưỡng về các cơ chế bảo mật dữ liệu gần đây được phát triển cho lưu trữ đám mây. Dựa trên kết quả của nghiên cứu, chúng tôi đưa ra những hiểu biết và đề xuất về các hướng nghiên cứu trong tương lai trong việc đạt được các mục tiêu an ninh của mỗi quốc gia.

Phần còn lại của bài báo này được tổ chức như sau: các mô hình của hệ thống lưu trữ đám mây và các tác động bảo mật liên quan được giới thiệu đầu tiên trong Phần II. Một mô hình kiến trúc hệ thống khái niệm chung của lưu trữ đám mây cũng được đề xuất để giải quyết các vấn đề bảo mật trong các lớp khác nhau. Trong Phần III, các nghiên cứu gần đây về bảo vệ tính toàn vẹn dữ liệu như bằng chứng về khả năng truy xuất và thử nghiệm của bên thứ ba được xem xét và so sánh. Trong Phần IV, bảo mật dữ liệu và công việc nghiên cứu liên quan được thảo luận. Một kỹ thuật mã hóa mới đầy hứa hẹn, mã hóa đồng hình hoàn toàn, cho phép các hoạt động hình học được thực hiện trên dữ liệu được mã hóa, được kiểm tra chi tiết trước, sau đó là thảo luận về kiểm soát truy cập và mã hóa có thể tìm kiếm. Trong Phần V, chúng tôi thăm dò các phương pháp để đảm bảo tính khả dụng của dữ liệu trong các hệ thống lưu trữ đám mây phân tán, chẳng hạn như đồng bộ hóa dữ liệu, khôi phục dữ liệu và thuật toán phân tán thông tin. Cuối cùng, kết luận được đưa ra trong Phần VI.

# **II- Tổng Quan Về Bảo Mật Trong Lưu Trữ Đám Mây**

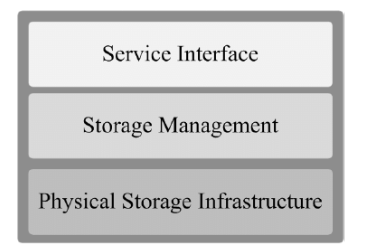
Phạm vi và yêu cầu về bảo mật đám mây khác nhau đáng kể với mô hình triển khai đám mây khác nhau. Viện Tiêu chuẩn và Công nghệ Quốc gia (NIST) đã xác định [51] 4 mô hình triển khai của điện toán đám mây. Đám mây riêng được cung cấp để sử dụng độc quyền bởi một tổ chức duy nhất bao gồm nhiều người dùng. Đám mây cộng đồng được cung cấp để sử dụng độc quyền bởi một cộng đồng người dùng cụ thể từ nhiều tổ chức có chung mối quan tâm. Đám mây công cộng được cung cấp để sử dụng mở bởi công chúng. Đám mây lai là sự kết hợp của ít nhất hai trong số ba loại trên. Rõ ràng, giá triển khai giảm từ đám mây riêng sang đám mây công cộng với chi phí tăng lên do lo ngại về bảo mật.

Lưu trữ đám mây có thể được triển khai trong bất kỳ mô hình nào trong bốn mô hình triển khai. Các dịch vụ lưu trữ đám mây công cộng dựa trên Internet đang phát triển nhanh chóng vì chúng có thể cung cấp cho người dùng mức tiết kiệm chi phí lớn nhất và độ co giãn cao nhất. Nhiều nhà cung cấp dịch vụ lưu trữ, bao gồm Amazon, IBM, Google, Microsoft, EMC, HP, Symantec, Rackspace, chỉ một số ít, đang cạnh tranh trên thị trường khổng lồ này. Tuy nhiên, các dịch vụ lưu trữ đám mây công cộng này cũng phải đối mặt với rủi ro tiềm ẩn cao nhất về vi phạm bảo mật vì cơ sở hạ tầng dùng chung mở cửa cho công chúng. Trên thực tế, các hệ thống lưu trữ đám mây được triển khai dưới các hình thức đám mây đa năng khác, bao gồm đám mây lai và đám mây cộng đồng, cũng phải đối mặt với rủi ro cao hơn so với đám mây riêng tư. Ngay cả trong đám mây cá nhân, rất có thể lưu trữ đám mây được quản lý và vận hành bởi một nhà cung cấp dịch vụ ngoài cơ sở, để tận dụng tối đa lợi thế của điện toán đám mây. Trên thực tế, một số sử dụng lưu trữ đám mây, chẳng hạn như sao lưu phục hồi sau thảm họa, yêu cầu lưu trữ ngoài hiện trường. Khi dữ liệu không còn được lưu trữ và quản lý bởi chủ sở hữu dữ liệu trên cơ sở riêng của mình, chủ sở hữu dữ liệu có ít quyền kiểm soát hơn đối với dữ liệu của họ. Do đó, bảo mật lưu trữ đám mây là thách thức nếu các nhà cung cấp dịch vụ không đáng tin cậy, bất kể mô hình triển khai là gì.

NIST cũng đã xác định ba mô hình dịch vụ đám mây chính. Phần mềm như một dịch vụ (SaaS) ngụ ý người tiêu dùng sử dụng ứng dụng của nhà cung cấp dịch vụ chạy trên cơ sở hạ tầng đám mây, chẳng hạn như SalesForce CRM, YouTube, Google Apps (Gmail, Google Document). Nền tảng như một dịch vụ (PaaS), có nghĩa là nhà cung cấp dịch vụ xây dựng một môi trường cho người tiêu dùng để thiết lập các ứng dụng có được với các ngôn ngữ lập trình, thư viện và công cụ đã được hỗ trợ trong nền tảng. PaaS nổi tiếng bao gồm Google App Engine, Microsoft Azure và Cloud Foundry từ VMware. Cơ sở hạ tầng như một Dịch vụ (IaaS), đại diện cho người tiêu dùng có thể triển khai và quản lý ứng dụng, hệ điều hành với mạng và thiết bị lưu trữ được cung cấp. Amazon 's Elastic Compute Cloud (EC2) là một ví dụ hàng đầu, với các dịch vụ khác như Rackspaces Mosso và GoGrid' s ServePath [51]. Từ SaaS đến PaaS và đến IaaS, người dùng dần dần có quyền kiểm soát sâu hơn đối với chồng kiến trúc đám mây, do đó chia sẻ nhiều trách nhiệm hơn về thực thi bảo mật.

Các dịch vụ lưu trữ đám mây cơ bản được phân loại là mô hình dịch vụ IaaS, mặc dù nhiều nhà cung cấp dịch vụ lưu trữ đám mây đang cung cấp các dịch vụ PaaS và SaaS giá trị gia tăng được xây dựng dựa trên các dịch vụ IaaS cơ bản của họ. Là một IaaS, lưu trữ đám mây cho phép người dùng tăng cường biện pháp bảo mật bằng cách sử dụng các cơ chế bảo vệ bảo mật của riêng họ. Ví dụ: người dùng có thể mã hóa dữ liệu của họ trước khi chuyển chúng vào lưu trữ đám mây bằng khóa riêng do chính họ quản lý. Trong trường hợp này, ngay cả khi dữ liệu được các bên không được phép truy cập, thông tin nhạy cảm sẽ không được tiết lộ mà không có được khóa. Tuy nhiên, người dùng dịch vụ SaaS chỉ có thể dựa vào các biện pháp an toàn.

Kiến trúc cơ bản của hệ thống lưu trữ đám mây được đặt ra bởi một nhóm tài nguyên lưu trữ, bao gồm hệ thống tệp phân tán, Thỏa thuận cấp độ dịch vụ (SLA) và giao diện dịch vụ [37], [94]. Để hiểu khái niệm về các hệ thống lưu trữ đám mây và cách bảo vệ an ninh có thể được tích hợp và triển khai trong hệ thống, chúng tôi phân hủy kiến trúc hệ thống thành một mô hình tham chiếu ba lớp dựa trên ranh giới chức năng logic như trong Hình 1.



*Hình 1: Kiến trúc lưu trữ đám mây [94]*

Trong lớp cơ sở hạ tầng lưu trữ vật lý, có các mạng có dây và không dây phân tán kết nối mạng thiết bị lưu trữ phân tán. Lớp thứ hai là lớp quản lý lưu trữ, xử lý các hoạt động cần thiết, chẳng hạn như đặt dữ liệu, sao chép và giảm, trên dữ liệu được lưu trữ trong lớp thứ nhất. Bằng công nghệ ảo hóa, lớp này trở thành lớp trừu tượng thông minh che giấu sự phức tạp của lớp bên dưới. Lớp giao diện dịch vụ cung cấp giao diện cho người dùng truy cập dữ liệu của họ được lưu trữ trong lưu trữ đám mây. Các hệ thống lưu trữ đám mây cơ bản chủ yếu cung cấp phần mềm phía máy khách hoặc giao diện trình duyệt web hoặc đôi khi cả hai. Phần mềm phía máy khách phải được cài đặt trên các thiết bị của người dùng được sử dụng để truy cập dữ liệu, trong khi giao diện trình duyệt web cho phép truy cập dữ liệu từ bất kỳ nơi nào mà không cần cài đặt cục bộ. Một số hệ thống lưu trữ đám mây tiên tiến cũng cung cấp Giao diện lập trình ứng dụng (API), có thể được sử dụng để tích hợp trực tiếp quyền truy cập dữ liệu được lưu trữ vào các ứng dụng khác. Hầu hết các ứng dụng đó thuộc về PaaS hoặc SaaS dựa trên cơ sở hạ tầng lưu trữ đám mây.

Vì các lớp khác nhau có chức năng khác nhau, các mối quan tâm khác nhau trong mỗi lớp có sự nhấn mạnh khác nhau. Lớp cơ sở hạ tầng lưu trữ vật lý giải quyết bảo mật vật lý và bảo mật vật lý. Lớp quản lý lưu trữ cần kiểm soát hiệu quả việc phân bổ tài nguyên và thực hiện quản lý dữ liệu một cách đáng tin cậy. Trong lớp giao diện dịch vụ, làm thế nào để tránh sự xâm phạm quyền của cả khách hàng và nhà cung cấp dịch vụ bằng cách sử dụng giao diện an toàn và API đã được phổ biến rộng rãi. Trong mỗi lớp, có thể có rủi ro, xâm nhập và tấn công chống lại tính toàn vẹn, bảo mật và/hoặc tính khả dụng của dữ liệu. Do đó, các cơ chế bảo vệ an ninh lưu trữ nên được tích hợp vào mọi lớp và các mục tiêu bảo mật không thể đạt được nếu không có các nỗ lực hợp tác trên cả ba lớp. Ví dụ: để đảm bảo tính khả dụng của dữ liệu trong mọi trường hợp như lỗi phần cứng hoặc thảm họa, lớp cơ sở hạ tầng lưu trữ vật lý thường có dữ liệu trùng lặp được lưu trữ ở các vị trí khác nhau. Trong trường hợp dữ liệu được lưu trữ ở một vị trí bị mất, lớp quản lý lưu trữ phải có khả năng định vị dữ liệu có sẵn ở một vị trí khác và định tuyến dữ liệu đó cho người dùng theo yêu cầu của họ. Lớp giao diện dịch vụ phải có khả năng nhận hiệu quả các yêu cầu đến từ bất cứ nơi nào và cung cấp phương thức truy cập đáng tin cậy cho người dùng hợp pháp. Với kiến trúc tổng quan về lưu trữ đám mây và các tác động bảo mật của nó, chúng ta sẽ thảo luận về những nỗ lực nghiên cứu gần đây trong việc đạt được ba mục tiêu bảo mật chính, cụ thể là tính toàn vẹn dữ liệu, tính bảo mật dữ liệu và tính khả dụng của dữ liệu, trong ba phần sau đây.

# **III-Tính Toàn Vẹn Dữ Liệu**

Tính toàn vẹn của dữ liệu đề cập đến tài sản mà dữ liệu chưa bị thay đổi hoặc phá hủy theo cách không được cho phép [91]. Trong lưu trữ đám mây, do người dùng không còn sở hữu lưu trữ vật lý dữ liệu của họ, làm thế nào để xác minh hiệu quả tính đúng đắn của dữ liệu có nguồn gốc ngoài được lưu trữ trong máy chủ đám mây đã trở thành một thách thức cũng như một chủ đề nghiên cứu đầy hứa hẹn cho bảo mật lưu trữ dữ liệu.

Trong các mạng truyền thông dữ liệu truyền thống, tính toàn vẹn dữ liệu thường chỉ bị đe dọa bởi những kẻ tấn công độc hại. Cả người gửi và người nhận dữ liệu đều đáng tin cậy và hợp tác trong việc phát hiện và bảo vệ tính toàn vẹn của dữ liệu. Tuy nhiên, trong lưu trữ đám mây, các máy chủ lưu trữ đám mây không phải lúc nào cũng đáng tin cậy. Nhà cung cấp dịch vụ lưu trữ đám mây có động lực để giúp người dùng dịch vụ tránh tình trạng dữ liệu được lưu trữ. Ví dụ: nhà cung cấp dịch vụ có thể xóa dữ liệu hiếm khi được truy cập để tiết kiệm việc sử dụng dung lượng hoặc ẩn các sự cố mất dữ liệu để duy trì danh tiếng của mình. Hơn nữa, máy chủ độc hại có thể thay đổi hoặc thay thế dữ liệu được lưu trữ. Để ngăn chặn các trường hợp trên, sẽ có giá trị hơn khi có quy trình xác minh tính toàn vẹn dữ liệu và thường xuyên truy vấn tính chính xác của dữ liệu trong các máy chủ lưu trữ. Một cơ chế xác minh hiệu quả cũng có thể cho phép người dùng phát hiện các mối đe dọa về tính toàn vẹn dữ liệu trong lưu trữ đám mây sớm hơn và thực hiện các hành động cần thiết để giảm thiểu thiệt hại hoặc khôi phục thiệt hại gây ra.

Có ba yêu cầu cơ bản đối với quá trình xác minh tính toàn vẹn dữ liệu, đó là, hiệu quả, sử dụng không bị ràng buộc và cơ chế tự bảo vệ. Hiệu quả ngụ ý chiều rộng băng thông mạng tối thiểu và dung lượng lưu trữ khách hàng là cần thiết cho quá trình xác minh. Khách hàng không cần truy cập toàn bộ dữ liệu cho mục đích xác minh. Việc sử dụng không giới hạn thể hiện quá trình xác minh nên hỗ trợ số lượng truy vấn không giới hạn. Cơ chế tự bảo vệ có nghĩa là bản thân quy trình phải được bảo mật chống lại máy chủ độc hại vượt qua bài kiểm tra tính toàn vẹn mà không cần truy cập dữ liệu.

Một số kỹ thuật và cơ chế khác nhau đã được đề xuất và thiết kế cho quá trình kiểm tra tính toàn vẹn của dữ liệu đám mây. Chủ đạo của nghiên cứu trong lĩnh vực này thuộc về Bằng chứng về Khả năng Truy xuất (POR) và Sở hữu Dữ liệu Khả thi (PDP), cả hai đều được thiết kế theo ba yêu cầu trên. Hai phương pháp ban đầu xuất hiện với một khái niệm tương tự nhưng cách tiếp cận khác nhau. Kể từ đó, mỗi người đã trải qua quá trình phát triển hơn nữa theo các hướng khác nhau như hỗ trợ dữ liệu động, xác minh công khai và quyền riêng tư đối với người xác minh. Hỗ trợ dữ liệu động cho phép khách hàng tự động cập nhật một phần dữ liệu của họ sau khi tải lên dữ liệu. Khả năng xác minh công khai cho phép mọi người, không chỉ chủ sở hữu dữ liệu hoặc người xác minh, thực hiện quá trình xác minh. Quyền riêng tư đối với người xác minh đảm bảo rằng quá trình xác minh không chứa bất kỳ thông tin cá nhân nào của chủ sở hữu dữ liệu. Các chương trình POR và PDP với sự phát triển của chúng sẽ được thảo luận và so sánh chi tiết hơn ở phần sau của phần này.

Bên cạnh hai cách tiếp cận đó, có một số phương pháp được nghiên cứu để giải quyết vấn đề toàn vẹn dữ liệu lưu trữ do chèn, sửa đổi và xóa dữ liệu ở cấp khối. Trong năm 2010, bằng chứng về khả năng xóa bỏ (POE) đã được đề xuất bởi Paul và Saxena [58]. POE giải quyết nhu cầu của khách hàng để đảm bảo phá hủy toàn diện dữ liệu được lưu trữ trong bộ lưu trữ khi họ rút dữ liệu và liên kết với nhà cung cấp dịch vụ lưu trữ. Mô hình này đóng vai trò như kỹ thuật thăm dò hoặc phá hủy, có thể đảm bảo dữ liệu được lưu trữ được cắt nhỏ một phần hoặc hoàn toàn dựa trên các quy tắc lưu trữ dữ liệu. Tuy nhiên, chương trình này chỉ cho phép chủ sở hữu dữ liệu biết dữ liệu đang bị phá hủy. Một chương trình song song khác được gọi là Proofs of Secure Erasure (PoSE-s) cũng có chức năng tương tự về chứng thực từ xa [59].

Trong các phần phụ sau đây, trước tiên chúng tôi sẽ giới thiệu POR và PDP, sau đó là sự phát triển của họ để cải thiện hiệu suất, hỗ trợ dữ liệu động và khả năng xác minh công khai.

**A.Giới thiệu về Bằng chứng về khả năng truy xuất và khả năng sở hữu dữ liệu**

Diagram

Description automatically generated

*Hình 2: Sơ đồ của một hệ thống POR [38]*

Là cơ chế được nghiên cứu rộng rãi để đảm bảo tính toàn vẹn dữ liệu, POR lần đầu tiên được đề xuất bởi Juels và Kaliski vào năm 2007 [38]. Hình 2 mô tả sơ đồ chung của hệ thống POR được đề xuất, đảm bảo máy chủ (kiểm chứng) cho khách hàng (thẩm tra viên) rằng dữ liệu được lưu trữ là nguyên vẹn trong quá trình lưu trữ và truy xuất của khách hàng. Đầu tiên, khách hàng mã hóa một tệp thô F thông qua một thuật toán mã hóa vào một tệp được mã hóa F’ và sau đó lưu trữ nó trong câu tục ngữ. Một thuật toán tạo khóa tạo khóa K được lưu trữ trong bộ xác minh và nó được sử dụng để mã hóa. Đối với quy trình kiểm tra, người xác minh có thể thực hiện quy trình thử thách-phản hồi với người xác thực để kiểm tra xem F có thể được truy xuất hay không.

Chương trình POR đầu tiên được giới thiệu bởi Juels và Kaliski đã sử dụng một chương trình trọng điểm. Giao thức POR mã hóa F và chèn ngẫu nhiên một số khối canh vào các khối dữ liệu tệp khác sau khi mã hóa. Những lính canh này đóng một vai trò quan trọng để xác minh. Người xác minh có thể thách thức người xác minh bằng cách chỉ ra vị trí của một tập hợp các lính canh và người xác minh sẽ trả về các giá trị của các lính canh. Nếu các giá trị khác với dữ liệu của người xác minh, thì điều đó cho thấy rằng người xác minh đã xóa hoặc sửa đổi F. POR cũng bao gồm mã sửa lỗi để khôi phục một phần nhỏ F nếu bị hỏng. Tuy nhiên, sơ đồ này yêu cầu xử lý trước và mã hóa F trước khi lưu trữ vào bộ lưu trữ dữ liệu và nó được sử dụng có giới hạn - số lượng lính canh có thể được sử dụng cho các truy vấn hạn chế. Do đó, Juels và Kaliski đã đề xuất một kỹ thuật khác từ Lillibridge et al. [47], Naor và Rothblum [53]. Nó lưu trữ các khối dữ liệu mã hóa dư thừa với mã xác thực tin nhắn (MAC) để thay thế các canh gác, và các MAC được lưu trữ cùng với các khối dữ liệu. Trong trường hợp này, thuật toán xác minh có thể kiểm tra tính toàn vẹn của dữ liệu và đảm bảo khả năng truy xuất bằng cách yêu cầu số lượng vị trí khối ngẫu nhiên với MAC của chúng. Cách tiếp cận này giải quyết vấn đề sử dụng bị giới hạn của kế hoạch trước đó, nhưng với chi phí cao hơn độ phức tạp truyền thông của cuộc kiểm toán.

Diagram

Description automatically generated

*Hình 3: Sơ đồ của một hệ thống PDP [4]*

Mặt khác, PDP ra đời đồng thời với kế hoạch của Juels- Kaliski. Nó được đề xuất bởi Ateniese et al. [4] và được xây dựng dựa trên mật mã khóa đối xứng. PDP đầu tiên chọn các thẻ có thể xác minh đồng hình dựa trên RSA [36] để kết hợp nhiều khối tệp thành một giá trị duy nhất. Một cách tiếp cận tương tự cũng đã được Shacham và Waters áp dụng sau đó [68] trong đề án POR năm 2008. Chương trình PDP cũng cung cấp tính độc lập về định dạng dữ liệu và nó không đặt ra hạn chế đối với định dạng dữ liệu. Nói cách khác, PDP cho phép bất kỳ người xác minh nào (không chỉ khách hàng) truy vấn máy chủ. POR và PDP đều sử dụng mã xóa, là một Hiệu chỉnh Lỗi Chuyển tiếp (FEC) cho kênh xóa nhị phân, giúp khôi phục tin nhắn gốc từ dữ liệu bị hỏng nhẹ. Sự khác biệt lớn giữa POR ban đầu và PDP là POR đảm bảo không chỉ toàn vẹn dữ liệu ở cuối máy chủ mà còn khả năng truy xuất, trong khi PDP chỉ đảm bảo toàn vẹn dữ liệu ở lưu trữ dữ liệu đám mây. Tuy nhiên, PDP hiệu quả hơn so với POR của Juels-Kaliski, vì nó không yêu cầu bất kỳ mã hóa hàng loạt nào và PDP yêu cầu không gian lưu trữ nhỏ hơn ở phía máy khách và ít băng thông hơn cho các thử thách và phản hồi. Tuy nhiên, cả hai chương trình chỉ hoạt động trên dữ liệu tĩnh, mặc dù Ateniese et al. [5] đã đề xuất một phiên bản động sau đó vào năm 2008, nhưng nó bị hạn chế bởi số lượng truy vấn và các hoạt động khối cơ bản.

**B. Cải thiện khả năng xác minh công khai**

Vì kế hoạch POR ban đầu của Juels-Kaliski đã được đặt ra mà không cần thực hiện khả năng xác minh công khai, và độ phức tạp của nó vẫn còn cao để giao tiếp và lưu trữ khách hàng, nó đã trở thành một chủ đề phổ biến cho các nhà nghiên cứu để cải thiện khả năng xác minh công khai và hiệu quả (thảo luận trong phần phụ tiếp theo). Năm 2008, Shacham và Waters [68] đã đề xuất hai cấu trúc hệ thống POR mới dựa trên khái niệm POR của Juels-Kaliski. Cả hai giải pháp chỉ cho phép một giá trị xác thực cho mục đích xác minh. Cái đầu tiên có thể được xác minh riêng bằng cách sử dụng các hàm giả ngẫu nhiên (PRF); cái thứ hai có thể được xác minh công khai, và nó được xây dựng dựa trên sơ đồ chữ ký của Boneh, Lynn và Shacham trong một nhóm tuyến tính [11]. Kể từ khi chữ ký BLS được thông qua, khả năng truy xuất công khai đã đạt được và các bằng chứng được giảm xuống thành một giá trị xác thực duy nhất, do đó giảm độ phức tạp liên lạc từ O(t) đếnO(1), trong đó t là số lượng vị trí khối. Tuy nhiên, chương trình này vẫn chỉ hoạt động trên dữ liệu tĩnh, mà không hỗ trợ cập nhật dữ liệu động. Bên cạnh đó, thông số bảo mật dựa vào Random Oracles, có nghĩa là kích thước thử thách của khách hàng tăng lên đến O(t2).

Diagram

Description automatically generated

*Hình 4: Cơ cấu kiểm toán viên bên thứ ba (TPA) [86]*

Một mô hình hệ thống mới, như được mô tả trong Hình 4, nhằm thiết lập một cơ chế đáng tin cậy giữa khách hàng và Máy chủ lưu trữ đám mây (CSS) bằng cách giới thiệu Kiểm toán viên bên thứ ba (TPA), đã được đề xuất vào năm 2009 [84]. Bằng cách sử dụng giao thức kiểm toán riêng tư của bên thứ ba, TPA được tin cậy để giám sát dữ liệu được lưu trữ trên đám mây và các giao dịch giữa khách hàng và CSS, cũng như đánh giá và phơi bày rủi ro của các dịch vụ đám mây. Đề án mới này đã được tiếp tục phát triển dựa trên các POR hiện có và các nguyên mẫu mật mã mới được phát triển [81]– [83], [85], [86].

TPA thường áp dụng một bộ đồng định nghĩa dựa trên khóa công khai với mặt nạ ngẫu nhiên để thực hiện kiểm toán lưu lượng truy cập mà không có bản sao dữ liệu cục bộ để kiểm tra tính toàn vẹn. Hệ thống kiểm toán công khai này có thể được xây dựng từ giai đoạn thiết lập, cho phép người dùng khởi tạo các thông số bí mật của hệ thống, gửi siêu dữ liệu xác minh đến TPA và kiểm tra kết quả tương ứng. Trong quá trình này, TPA sẽ phát hành một thông báo kiểm toán cho máy chủ để kiểm tra dữ liệu của người dùng.

Trình xác thực đồng hình được sử dụng để xác minh siêu dữ liệu được tạo ra từ các khối dữ liệu riêng lẻ trong khi các trình biên dịch tổng hợp có thể biện minh cho sự kết hợp tuyến tính của các khối dữ liệu. Như một mô hình, người ta có thể sử dụng một mã thông báo đồng hình với xác minh phân tán để kiểm tra tính toàn vẹn của dữ liệu được mã hóa xóa. Các mã hiệu chỉnh xóa đóng một vai trò quan trọng trong việc chuẩn bị các tệp để phân phối sao cho các tệp được phân phối có vectơ bằng thừa số và thuộc tính độ tin cậy dữ liệu. Tuy nhiên, sự kết hợp tuyến tính của các khối dữ liệu có thể tiết lộ quyền riêng tư của người dùng. Với mặt nạ ngẫu nhiên, TPA không thể thu được nội dung dữ liệu của người dùng bằng cách xây dựng một nhóm chính xác các phương trình tuyến tính.

Mô hình trên đã được cải thiện hơn nữa trong [87] bằng cách hỗ trợ dữ liệu động lưới vào năm 2011. Zhu et al. [102] cũng đề xuất xây dựng các dịch vụ kiểm toán động cho lưu trữ không đáng tin cậy và thuê ngoài. Nó có thể phát hiện hành vi bất thường bằng cách sử dụng cấu trúc đoạn, lấy mẫu ngẫu nhiên và bảng chỉ mục - dấu gạch ngang.

Mặc dù các chương trình dựa trên TPA cho phép xác minh công khai về kiểm tra tính toàn vẹn của dữ liệu, nhưng họ có một trở ngại tiềm ẩn đòi hỏi một cử tri bổ sung, là kiểm toán viên bên thứ ba, được thêm vào toàn bộ chương trình lưu trữ dữ liệu hiện có. Việc thực hiện các chương trình như vậy có thể là một gánh nặng cho các nhà cung cấp dịch vụ vì chi phí bổ sung. Để giải quyết mối lo ngại này, Han và Xin [32] đã đề xuất một chương trình mới cung cấp các chức năng TPA truyền thống do CSP cung cấp một cách đáng tin cậy. Chương trình này sử dụng các kỹ thuật RSA và Bilinear Diffie-Hellman, tạo tiêu đề tin nhắn và các cơ chế để đạt được quá trình xác thực, đồng thời giảm độ phức tạp của điện toán đám mây. Một công việc khác cung cấp khả năng xác minh công khai mà không có sự trợ giúp từ kiểm toán viên bên thứ ba đã được kiểm tra trong[33] dựa trên công việc của Sebe et al. [67], và nó đã được chứng minh là an toàn từ một máy chủ không đáng tin cậy.

**C. Cải thiện hiệu quả**

Hiệu quả của POR và PDP đã được cải thiện từ các khía cạnh khác nhau của quá trình xác minh. Ví dụ, Curtmola et al. [21] cho thấy làm thế nào để tích hợp mã sửa lỗi với PDP và một xây dựng mã sửa lỗi đối nghịch tương tự như POR. Nó cũng cho phép chương trình PDP bảo mật nhiều bản sao trên hệ thống phân tán mà không cần mã hóa từng bản sao riêng biệt.

Bên cạnh đó, Dodis et al. [24] đã cung cấp các giải pháp khác nhau của các chương trình POR được tối ưu hóa theo các ràng buộc khác nhau, chẳng hạn như sử dụng có giới hạn hoặc không sử dụng, âm thanh tri thức hoặc âm thanh thông tin. Họ đã phân tích sự đánh đổi về tham số và tính bảo mật giữa các chương trình sử dụng có giới hạn và không giới hạn, đồng thời họ cũng so sánh chi tiết các POR trong các tình huống khác nhau. Nó cũng cải thiện hệ thống Shacham-Waters POR bằng cách tránh sử dụng Oracles ngẫu nhiên, làm giảm kích thước thử thách xuống tuyến tính trong thông số bảo mật, từ O(t2) đến O(t).

Ngoài ra, một khung lý thuyết về cải thiện PORS đã được đề xuất đồng thời bởi Bowers et al. [12]. Mô hình cung cấp một cải tiến so với các giao thức của Juels-Kaliski[38] và Shacham-Waters [69] bằng cách đề xuất một biến thể mới để đạt được chi phí lưu trữ thấp hơn và chịu được tỷ lệ lỗi cao hơn. Một kế hoạch POR khác đã được đề xuất bởi Kumar và Saxena vào năm 2011 [74]. Nó nhắm vào việc đơn giản hóa sơ đồ trọng điểm của Juels-Kaliski, làm cho nó phù hợp với sức mạnh tính toán hạn chế hoặc lưu trữ nhỏ ở cuối bộ xác minh. Đối với PDP, Ateniese và Burns et al. đã kết luận sự phát triển và triển khai nghiên cứu trước đây của PDP trong năm 2011 [3], và đề xuất hai chương trình PDP được cải thiện an toàn với hiệu quả cao hơn so với trước đó.

**D.Cải thiện hỗ trợ dữ liệu động**

Việc hỗ trợ cập nhật dữ liệu động trong các sơ đồ xác minh tính toàn vẹn của dữ liệu đặc biệt khó khăn. Ateniese et al. [5] đã đề xuất sơ đồ PDP động một phần đầu tiên vào năm 2008. Chương trình này hiệu quả hơn trong giai đoạn thiết lập và xác minh so với phiên bản trước đó trong [4], vì nó chỉ dựa trên mật mã khóa đối xứng. Mặt khác, nó chỉ cho phép một số lượng hạn chế các truy vấn và các hoạt động khối cơ bản với chức năng hạn chế. Ví dụ: chèn khối không được hỗ trợ. Hơn nữa, khả năng xác minh công khai cũng không được hỗ trợ.

Năm 2009, Erway et al. đã đề xuất một cải tiến về PDP, được gọi là sở hữu dữ liệu có thể chứng minh động (DPDP) [25]. Để hỗ trợ các cập nhật có thể chứng minh được trên dữ liệu được lưu trữ, mô hình mới này sử dụng các thư mục được xác thực dựa trên thông tin thứ hạng và nó định nghĩa cập nhật là chèn khối, sửa đổi hoặc xóa để đạt được chương trình PDP động. Tuy nhiên, chương trình này duy trì danh sách bỏ qua [57] cho các thẻ và lưu trữ siêu dữ liệu gốc ở phía khách hàng để ngăn chặn tấn công phát lại, vì vậy độ phức tạp tính toán và giao tiếp của nó có thể lên đến O(logt).

Các cập nhật dữ liệu động về POR lần đầu tiên được xem xét vào năm 2009. Wang et al. [84], [86] đã đề xuất sơ đồ đầu tiên đạt được động lực dữ liệu hiệu quả của mô hình POR bằng cách sử dụng mã thông báo đồng hình với xác minh phân tán của dữ liệu được mã hóa xóa và thao tác Cây băm Merkle (MHT) [52] tương ứng. Chương trình đầu tiên chỉ hỗ trợ cập nhật khối, xóa và nối các hoạt động, trong khi chương trình thứ hai cung cấp cả khả năng xác minh công khai và động lực dữ liệu để kiểm tra tính toàn vẹn dữ liệu từ xa, nhưng độ phức tạp xác minh tăng lên O(log n) từ O(1) như một sự đánh đổi, và nó đạt được một phần động thay vì hoàn toàn động. Cả hai dự án cũng cho thấy một mô hình hệ thống mới liên quan đến Kiểm toán viên bên thứ ba (TPA).

Ngoài ra, Zheng và Xu đã trình bày một kế hoạch POR mới với một tài sản mới, cụ thể là, công bằng, để đối phó với dữ liệu động [98]. Thuộc tính này ngăn khách hàng vô đạo đức buộc tội máy chủ hợp pháp về việc sửa đổi dữ liệu được lưu trữ của họ. Vấn đề này phát sinh do tính năng của dữ liệu động. POR cho lưu trữ dữ liệu tĩnh có thể giải quyết vấn đề này đơn giản bằng cách yêu cầu người xác minh phê duyệt và ký số khi dữ liệu chưa được lưu trữ vào bộ nhớ. Bằng chứng về khả năng truy xuất công bằng và năng động (FDPOR) được đề xuất chủ yếu bao gồm hai phần, một cấu trúc dữ liệu được xác thực mới: cây 2-3 dựa trên phạm vi (rb23Tree) và sơ đồ chữ ký gia tăng mới được gọi là băm-nén-và-ký. Tuy nhiên, FDPOR không hỗ trợ khả năng xác minh công khai, và độ phức tạp cho cả người xác minh và người chứng minh cao hơn so với các POR trước đó.

**E. Tóm tắt**

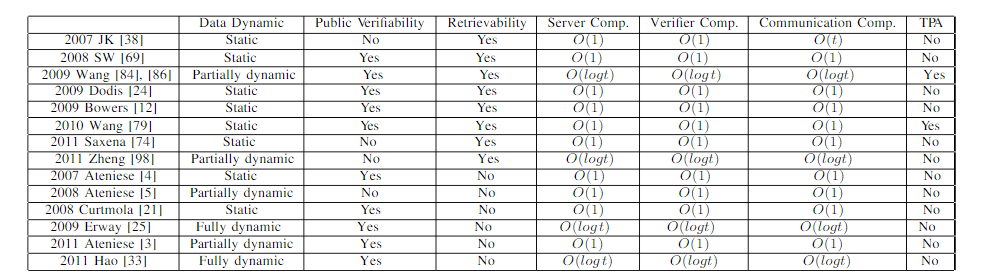
POR và PDP là các giao thức kiểm tra tính toàn vẹn dữ liệu từ xa chính được đề xuất trong các hệ thống lưu trữ đám mây. Các giao thức POR và PDP ban đầu khác nhau ở nhiều khía cạnh. POR được coi là an toàn hơn so với PDP, bởi vì nó yêu cầu mã hóa dữ liệu gốc và mã sửa lỗi để khôi phục dữ liệu bị hỏng, trong khi PDP được biết đến với hiệu quả cao hơn và khả năng áp dụng cho các cơ sở dữ liệu công cộng quy mô lớn, chẳng hạn như thư viện kỹ thuật số. Với sự cải thiện hơn nữa của mỗi chương trình, hai chương trình đã hội tụ về cùng một mục tiêu. Ví dụ, mặc dù khả năng xác minh công khai và thẻ có thể xác minh đồng hình lần đầu tiên được biết đến cho PDP, những đặc điểm này cũng áp dụng cho POR. Mặt khác, một số biến thể PDP cũng có thể áp dụng các công cụ mã hóa và/hoặc mã hóa lỗi cor- rection để tăng cường đo lường bảo mật của chúng. Do đó, nó là tất cả về việc đánh đổi giữa các chức năng an ninh và hiệu quả.

Trong Bảng I, chúng tôi tóm tắt các sơ đồ POR và PDP được xem xét ở trên bằng cách so sánh kỹ lưỡng hiệu suất của chúng. Điều đáng chú ý là các chương trình hỗ trợ dữ liệu động có độ phức tạp cao hơn so với các chương trình tương tự. Các hướng nghiên cứu trong tương lai bao gồm cải thiện hơn nữa về hiệu quả và hỗ trợ dữ liệu hoàn toàn động. Để cải thiện hiệu quả của các chương trình đó, giảm chi phí truyền thông và chi phí lưu trữ là những cân nhắc chính đáng. Tuy nhiên, hỗ trợ dữ liệu hoàn toàn động là một mục tiêu đầy thách thức, bởi vì nó làm tăng độ phức tạp nhưng làm giảm thông tin cập nhật ở đầu máy chủ.

# **IV. Bảo Mật Dữ Liệu**

Bảo mật dữ liệu trong bảo mật lưu trữ đám mây đề cập đến thuộc tính mà thông tin được lưu trữ trong bộ lưu trữ đám mây là không được cung cấp hoặc tiết lộ cho các cá nhân trái phép, thực thể, hoặc quá trình. Kiểm soát truy cập và mã hóa dữ liệu đã được triển khai rộng rãi để bảo vệ bí mật dữ liệu trong các mạng truyền thông dữ liệu truyền thống.Việc mở rộng triển khai của họ trong các hệ thống lưu trữ đám mây là điều tự nhiên. Đối với trường hợp, mã hóa Lớp cổng bảo mật (SSL) và AES-256 bit được áp dụng trong Dropbox để đảm bảo an toàn dữ liệu. Tuy nhiên, bảo mật dữ liệu trong các hệ thống lưu trữ đám mây phải đối mặt với những rủi ro và thách thức mới, do đó cần có các kỹ thuật mới hoặc cơ chế cải tiến. Trong phần này, chúng tôi thảo luận về những thách thức mới mà cơ chế mã hóa dữ liệu và kiểm soát truy cập phải đối mặt, cũng như những phát triển gần đây để đáp ứng những thách thức đó về bảo vệ tính bảo mật dữ liệu trong điện toán đám mây.

Mặc dù các kỹ thuật mã hóa truyền thống có thể ẩn thông tin dữ liệu khỏi máy chủ đám mây, nhưng nó sẽ không cung cấp giải pháp thỏa đáng nếu người dùng yêu cầu tính toán trên dữ liệu được lưu trữ của họ. Do tính toán không thể được thực hiện theo chức năng trên bản mã, người dùng sẽ phải giải mã dữ liệu trước khi thực hiện bất kỳ tính toán nào và mã hóa lại sau khi tính toán. Trong quá trình này, thông tin nhạy cảm có thể đã bị rò rỉ đến máy chủ tò mò. Nếu không, người dùng sẽ buộc phải thỏa hiệp với nhà cung cấp dịch vụ bằng cách tải lên văn bản gốc và ký SLA, điều này khiến dữ liệu của họ gặp rủi ro cao hơn. Để giải quyết vấn đề này, đã có nhiều nghiên cứu tập trung vào một phương pháp mã hóa nguyên thủy mới được đề xuất, đó là Mã hóa đồng hình hoàn toàn, cho phép tính toán văn bản mật mã mà không ảnh hưởng đến quá trình giải mã.



*BẢNG I: So sánh hiệu suất cho các lược đồ xác minh tính toàn vẹn của dữ liệu*

Trong phần tiếp theo của phần này, trước tiên chúng tôi kiểm tra các cơ chế kiểm soát truy cập mới với hiệu quả cao hơn và kiểm soát người dùng chi tiết phù hợp với lưu trữ đám mây. Sau đó, giới thiệu một số khái niệm mới về lược đồ mã hóa dữ liệu, chẳng hạn như mã hóa có thể tìm kiếm và FHE, đồng thời thảo luận về các ứng dụng tiềm năng của chúng trong việc bảo vệ tính bảo mật của dữ liệu trong điện toán đám mây. Sau đó, các phương pháp bảo mật dữ liệu khác cũng được thảo luận ngắn gọn. Chúng tôi cung cấp những hiểu biết của chúng tôi về những nỗ lực nghiên cứu hiện tại và các định hướng trong tương lai về bảo mật dữ liệu để tóm tắt chủ đề này.

## **A.Kiểm soát truy cập**

Như đã đề cập ở trên, kiểm soát truy cập là một trong những cơ chế chính để bảo vệ tính bí mật của dữ liệu trong các mạng dữ liệu truyền thống. Nó được thiết kế để chặn người dùng trái phép và tin tặc độc hại truy cập dữ liệu. Mặc dù mục tiêu kiểm soát truy cập trong lưu trữ đám mây không khác với mục tiêu trong mạng dữ liệu truyền thống, nhưng yêu cầu sẽ thay đổi. Kiểm soát truy cập truyền thống do nhà cung cấp dịch vụ thực thi không thể ngăn nhà cung cấp dịch vụ đám mây tò mò truy cập dữ liệu nhạy cảm của người dùng, được lưu trữ trong cơ sở hạ tầng của nhà cung cấp dịch vụ và do nhà cung cấp dịch vụ quản lý. Một máy chủ lưu trữ đám mây tò mò đang cố lấy thông tin nhạy cảm từ dữ liệu được lưu trữ của nó hoặc từ các hoạt động dữ liệu được thực hiện bởi chủ sở hữu dữ liệu và người dùng được ủy quyền, là một mô hình mối đe dọa mới đối với tính bảo mật của dữ liệu trong dịch vụ lưu trữ đám mây. Ngoài ra, nhà cung cấp dịch vụ độc hại có thể cố ý làm rò rỉ dữ liệu cho các bên trái phép để kiếm lợi hoặc kẻ tấn công độc hại có thể xâm phạm nhà cung cấp dịch vụ và có quyền truy cập trái phép vào dữ liệu.

Để giải quyết thách thức này, các lược đồ kiểm soát truy cập bằng mật mã đã chuyển cơ quan kiểm soát truy cập từ nhà cung cấp dịch vụ sang người dùng đã được đề xuất. Thay vì dựa vào nhà cung cấp dịch vụ không đáng tin cậy để cấp quyền kiểm soát truy cập, người dùng có thể thực thi quyền kiểm soát truy cập của chính họ bằng cách cấp có chọn lọc quyền truy cập giải mã khác nhau cho một phần nhất định của dữ liệu được mã hóa. Bằng phương tiện mã hóa, chủ sở hữu dữ liệu, tức là người dùng lưu trữ đám mây đã mất quyền kiểm soát vật lý đối với dữ liệu của chính họ có thể lấy lại quyền kiểm soát của họ ở cấp độ ngữ nghĩa.

Plutus [39] và SiRiUS [35] là những ví dụ về việc sử dụng mã hóa để bảo mật việc chia sẻ tệp trên bộ lưu trữ không đáng tin cậy từ xa. Các sơ đồ này đã mã hóa các tệp khác nhau bằng các khóa khác nhau, do đó thay đổi vấn đề truy cập tệp thành vấn đề quản lý khóa. Tuy nhiên, cách tiếp cận này không thể mở rộng khi áp dụng cho lưu trữ đám mây, vì độ phức tạp của quản lý khóa tăng theo số lượng tệp và/hoặc số lượng người dùng, cả hai đều có thể rất lớn trong hệ thống lưu trữ đám mây. Vì một số lượng lớn người dùng đang chia sẻ cùng một cơ sở hạ tầng trong bộ lưu trữ đám mây công cộng được xây dựng trên quy mô mạng phức tạp, điều quan trọng là phải có cơ chế kiểm soát truy cập hiệu quả, có thể mở rộng và đáng tin cậy.

Trong phần tiếp theo, chúng tôi xem xét nghiên cứu gần đây về kiểm soát truy cập hiệu quả hơn bằng cách sử dụng các kỹ thuật mã hóa được phát triển cho hệ thống lưu trữ đám mây.

### ***1)Kiểm soát truy cập bằng Mã hóa dựa trên thuộc tính***:

### Trong mô hình kiểm soát truy cập dựa trên thuộc tính, quyền truy cập được cấp dựa trên các thuộc tính của người dùng. Khi được áp dụng cho lưu trữ đám mây, kiểm soát truy cập được thực thi trên dữ liệu được mã hóa bằng cách sử dụng sơ đồ mã hóa dựa trên thuộc tính (ABE). Trong hệ thống ABE, khóa và bản mã của người dùng được gắn nhãn bằng tập hợp các thuộc tính mô tả. Một khóa cụ thể chỉ có thể giải mã một bản mã cụ thể nếu có sự trùng khớp giữa các thuộc tính của bản mã và khóa của người dùng.

Khái niệm về ABE được giới thiệu bởi Sahai và Waters [64]. Kiểm soát truy cập của họ cho phép giải mã khi số lượng thuộc tính chồng chéo giữa bản mã và khóa riêng vượt quá ngưỡng k được chỉ định. Bản chất mờ của lược đồ này ban đầu được thiết kế cho lược đồ mã hóa dựa trên nhận dạng chịu lỗi có thể sử dụng nhận dạng sinh trắc học. Tuy nhiên, với cấu trúc truy cập phẳng dựa trên ngưỡng, nó không thể được khái quát hóa cho các ứng dụng khác. Hai sơ đồ ABE nổi bật với cấu trúc truy cập cây tổng quát hơn, đó là Mã hóa dựa trên thuộc tính chính sách khóa (KP-ABE) [29] và Mã hóa dựa trên thuộc tính chính sách mã hóa (CP-ABE) [7], được đề xuất vào năm 2006 và 2007, tương ứng. Cả hai thuật toán đều liên kết một tập hợp các thuộc tính mô tả rõ ràng với cấu trúc truy cập dạng cây để thực thi kiểm soát truy cập trên dữ liệu được mã hóa, nhưng chúng hoạt động theo cách ngược lại. Trong KP-ABE, mỗi bản mã được gắn nhãn với một tập hợp các thuộc tính trong quá trình mã hóa, trong khi khóa riêng của người dùng được liên kết với một cây truy cập chỉ định bản mã nào mà khóa có thể giải mã. Ngược lại, trong CP-ABE, các khóa riêng của Người dùng dựa trên một tập hợp các thuộc tính của họ trong khi các bản mã được liên kết với một cây truy cập trên các thuộc tính trong quá trình mã hóa. Kết quả là, trong lược đồ KP-ABE, chính nhà phân phối khóa (thường là nhà cung cấp dịch vụ), người quyết định chính sách truy cập, trong khi trong lược đồ CP-ABE, chính người mã hóa (thường là chủ sở hữu dữ liệu) kiểm soát quyền truy cập. các dữ liệu được mã hóa.

Trong các lược đồ ABE đã đề cập ở trên, chính sách truy cập chỉ có thể chứa công thức logic “và” và “hoặc”, và các cổng ngưỡng. Lược đồ KP-ABE đã được giới thiệu trong [55] cho phép các ràng buộc “tiêu cực” được thể hiện trong các chính sách truy cập. Ngoài ra, nhiều lược đồ CP-ABE đã được đề xuất chẳng hạn như [18], [45], [88] đạt được sự an toàn khi tấn công bản mã (CCA) đã chọn hoặc được xây dựng trên các giả định bảo mật khác nhau. Mặc dù KP-ABE và CP-ABE hoạt động ngược lại, Goyal et al. [28] cung cấp một cách tiếp cận chung để chuyển đổi lược đồ KP-ABE thành CP-ABE. Malek và Miri đã kết hợp hai sơ đồ ABE thành một hệ thống và đề xuất kiểm soát truy cập cân bằng cho phép cả nhà cung cấp dịch vụ thiết lập chính sách truy cập toàn hệ thống và chủ sở hữu dữ liệu thiết lập cấu trúc truy cập vào dữ liệu của riêng họ [50]. Hơn nữa nghiên cứu về ABE cũng được thảo luận trong [92], [97]. Trong một hệ thống động, các chính sách truy cập có thể khác nhau theo thời gian và trình độ của người dùng cũng có thể thay đổi. Do đó, khả năng thu hồi các thuộc tính từ người dùng là mong muốn trong các hệ thống ABE. Một số lược đồ ABE có thể thu hồi [63], [95] đã được đề xuất khi hệ thống ABE có thể thu hồi người dùng truy cập dữ liệu được mã hóa mà họ từng có quyền truy cập trong hệ thống.

Khi sử dụng ABE trong một hệ thống có một số lượng lớn các thuộc tính, việc đánh giá trình độ của người dùng và tạo các khóa giải mã bởi một cơ quan trung ương trở nên không thực tế. Mã hóa dựa trên nhiều thuộc TÍNH (MA- ABE) lần đầu tiên được đề xuất để giải quyết vấn đề này vào năm 2007 [16]. Trong một chương trình MA-ABE, các thuộc tính được chia thành các tập hợp khác nhau và mỗi tập hợp có thể được quản lý bởi một cơ quan thuộc tính độc lập. Các khóa thuộc tính tương ứng để giải mã được phát hành bởi nhiều cơ quan thuộc tính và các trình mã hóa có thể chỉ định chính sách truy cập yêu cầu người dùng lấy khóa giải mã cho các thuộc tính thích hợp từ các cơ quan khác nhau để giải mã một thông báo. Sau đó, một số cấu trúc MA-ABE khác đã được đề xuất trong [17], [46].

### ***2)Kiểm soát truy cập dựa trên vai trò***:

Một mô hình kiểm soát truy cập khác được gọi là Kiểm soát truy cập dựa trên vai trò (RBAC) [1], [66], cũng đã được áp dụng phổ biến trong hệ thống lưu trữ truyền thống để đơn giản hóa việc quản lý các quyền. Chính sách truy cập của nó được xác định dựa trên các vai trò khác nhau được hệ thống chỉ định cho người dùng, trong khi chủ sở hữu dữ liệu có thể chỉ định một tập hợp các quyền của dữ liệu của họ cho các vai trò khác nhau. Bằng cách tách các nhiệm vụ phân công vai trò và phân công quyền, RBAC là hiệu quả và có khả năng mở rộng hơn nhiều so với kiểm soát truy cập khác dựa trên người dùng cá nhân, bởi vì số lượng vai trò thường ít hơn đáng kể so với số lượng người dùng. Hơn nữa, nó giúp kiểm soát truy cập động dễ dàng hơn. Ví dụ: trong các ứng dụng mà quyền đối với vai trò thay đổi chậm, trong khi người dùng có thể nhập, rời khỏi hoặc thay đổi vai trò nhanh chóng, người quản lý vai trò có thể chỉ định vai trò mới cho người dùng hoặc thu hồi vai trò từ người dùng. Mặt khác, chủ sở hữu dữ liệu cũng có thể thêm quyền cho một vai trò hoặc thu hồi quyền từ một vai trò. Các tác giả của [54] đã đề xuất đưa RBAC vào mô hình kiểm soát truy cập mới cho hệ thống chăm sóc sức khỏe có thể cung cấp quyền truy cập linh hoạt, bởi vì nó có thể được sửa đổi linh hoạt khi nhiệm vụ thay đổi. Tuy nhiên, một trong những lời chỉ trích chính đối với các chương trình RBAC là quy trình phức tạp khi thiết lập cấu trúc vai trò. Để làm cho RBAC hiệu quả hơn, các vai trò có thể được cấu trúc theo thứ bậc để một số vai trò kế thừa quyền từ những vai trò khác. Để thực thi các chính sách kiểm soát truy cập dựa trên vai trò, một cách tiếp cận là biến vấn đề kiểm soát truy cập thành vấn đề quản lý chính. Trong tài liệu, tồn tại nhiều lược đồ kiểm soát truy cập phân cấp [2], [23], [65] đã được xây dựng dựa trên lược đồ quản lý khóa phân cấp (HKM). Do sự giống nhau về cấu trúc giữa kiểm soát truy cập phân cấp và RBAC, sơ đồ kiểm soát truy cập phân cấp có thể dễ dàng được sử dụng để thực thi các chính sách truy cập RBAC trong môi trường đám mây. Năm 2010, lược đồ mã hóa dựa trên vai trò (RBE) [100] được xây dựng trực tiếp trên các chính sách RBAC. Tính bảo mật của lược đồ kiểm soát truy cập phân cấp phụ thuộc vào việc thực hiện đúng quy trình gán khóa, trong khi tính bảo mật của RBE dựa trên tính bảo mật của thuật toán đồ họa mật mã. Cụ thể hơn, khi người dùng được gán cho một vai trò trong RBE, khóa giải mã được tính toán thông qua thuật toán mật mã bằng cách lấy giá trị bí mật và danh tính của người dùng và vai trò làm đầu vào. Trong sơ đồ kiểm soát truy cập phân cấp, khóa cho người dùng được tạo dựa trên các chính sách kiểm soát truy cập của toàn bộ hệ thống. Năm 2011, Zhu et al. [101] đã đề xuất một chương trình RBE có thể thu hồi cho phép người dùng được cấp hoặc thu hồi tư cách thành viên vai trò một cách linh hoạt.

Trong các chương trình trên để thực thi các chính sách RBAC, tư cách thành viên người dùng của mỗi vai trò và hệ thống phân cấp vai trò được quản lý bởi một cơ quan trung ương. Tuy nhiên, trong các hệ thống RBAC quy mô lớn có hàng trăm hoặc thậm chí hàng nghìn vai trò và hàng trăm nghìn người dùng và quyền, việc tập trung nhiệm vụ quản lý những người dùng và quyền này cũng như mối quan hệ của họ với các vai trò trong một nhóm bảo mật nhỏ là không thực tế. quản trị viên. Zhou et al. [99] đã đề xuất một chương trình RBE mới sử dụng thuật toán mã hóa phát sóng dựa trên danh tính (IBBE) [22], cho phép các thành viên người dùng được quản lý bởi các vai trò riêng lẻ. Trong chương trình RBE mới, văn bản thuần có thể được mã hóa cho một vai trò cụ thể và chỉ những người dùng trong vai trò đó và vai trò tiền nhiệm của nó mới có thể giải mã dữ liệu bằng bí mật vai trò và khóa giải mã của họ. Việc sử dụng thuật toán mã hóa phát sóng cho phép tự động thêm người dùng mới vào vai trò mà không cần mã hóa lại, cũng như thu hồi người dùng hiện tại khỏi vai trò mà không ảnh hưởng đến bất kỳ người dùng hiện tại nào khác. Ngoài ra, sơ đồ này còn có các tính năng khác như: khóa có kích thước không đổi và văn bản mật mã.

Cũng đã có các chương trình kiểm soát truy cập dựa trên thuộc tính (ABAC) và RBAC kết hợp được đề xuất để tận dụng lợi thế của cả hai để cung cấp kiểm soát truy cập hiệu quả cho các ứng dụng phân tán và thay đổi nhanh chóng [43]. Hong et al. [34] đã triển khai hệ thống RBAC cho lưu trữ đám mây thông qua CP- ABE. Trong công việc của họ, việc chỉ định quyền được xử lý bởi chủ sở hữu dữ liệu trong khi chỉ định vai trò được xử lý bởi người dùng khác thông qua tuyên truyền..

## **B.Mã hóa có thể tìm kiếm.**

Với việc ngày càng có nhiều dữ liệu được chuyển sang lưu trữ đám mây, việc cho phép tìm kiếm trên lượng dữ liệu khổng lồ cho nhiều ứng dụng người dùng là điều bắt buộc. Để duy trì tính thống nhất và toàn vẹn của dữ liệu, cần phải lưu trữ dữ liệu được mã hóa trong các máy chủ lưu trữ đám mây. Để thực hiện tìm kiếm dữ liệu, người dùng phải lưu trữ một chỉ mục cục bộ hoặc tải xuống tất cả dữ liệu được mã hóa, giải mã nó và tìm kiếm cục bộ. Cách tiếp cận này cũng không hiệu quả khi kích thước dữ liệu phát triển trong đám mây. Khi người dùng tìm kiếm và tải xuống các tệp có liên quan từ hệ thống lưu trữ đám mây, Nhà cung cấp dịch vụ lưu trữ (SSP) thường mong muốn lưu trữ dịch vụ tìm kiếm, bởi vì nó có thể giảm thiểu lưu lượng mạng và giảm độ phức tạp quản lý cho người dùng. Do đó, làm thế nào để thực hiện tìm kiếm trên cơ sở dữ liệu được mã hóa mà không cần giải mã đã trở thành một chủ đề ngày càng hấp dẫn trong các hệ thống lưu trữ đám mây. Tái trung tâm, đã có những nguyên tắc mật mã mới, được gọi là các sơ đồ mã hóa có thể tìm kiếm [?], [10], được đề xuất để giải quyết vấn đề này.

Ý tưởng cơ bản của các chương trình mã hóa có thể tìm kiếm là mã hóa một chỉ mục tìm kiếm được tạo ra trên một bộ sưu tập dữ liệu theo cách mà nội dung của nó được ẩn mà không có mã thông báo thích hợp, chỉ có thể được tạo bằng khóa bí mật. Được cung cấp mã thông báo cho một từ khóa, người ta có thể truy xuất các con trỏ đến các tệp dữ liệu được mã hóa có chứa từ khóa. Trong quá trình truy xuất, không có nội dung nào của các tệp dữ liệu hoặc từ khóa được tiết lộ, ngoài việc tất cả các tệp dữ liệu được truy xuất chứa chung một từ khóa.

Các chương trình mã hóa có thể tìm kiếm, bao gồm Mã hóa Tìm kiếm Đối xứng (SSE) [73], Mã hóa Tìm kiếm Đối xứng (ASE) [10] và các cải tiến khác trên cả hai chương trình được xem xét trong [40]. Do đó, SSE sử dụng các thuật toán mật mã đối xứng, chẳng hạn như mã hóa khối hoặc hàm băm, phù hợp khi bên thực hiện tìm kiếm trên dữ liệu cũng là người tạo ra nó, trong khi ASE sử dụng các thuật toán mật mã bất đối xứng như đường cong elliptic, do đó cũng phù hợp khi bên thực hiện tìm kiếm trên dữ liệu khác với bên tạo ra nó. Do đó, ASE có các ứng dụng rộng hơn SSE trong lưu trữ đám mây so với SSE. Trong khi đó, so với các chương trình SSE, ASE có thể đạt được các truy vấn tìm kiếm phức tạp hơn, chẳng hạn như kết hợp các thuật ngữ, nhưng với chi phí của độ phức tạp cao hơn và bảo đảm bảo mật yếu hơn. ASE hiệu quả, hoặc kế hoạch ESE đã được giới thiệu trong [6] để cải thiện hiệu quả khi các từ khóa khó đoán. Tuy nhiên, nó dễ bị tấn công từ điển hơn.

Vì SSE đạt được hiệu quả cao hơn và bảo mật mạnh hơn, nó đã được phát triển hơn nữa gần đây. Ví dụ, SSE động [41], [42] mở rộng cách tiếp cận chỉ mục đảo ngược [20] để cho phép cập nhật chỉ mục và tệp dữ liệu được mã hóa và để đạt được bảo mật thích ứng chống lại các cuộc tấn công bằng từ khóa được chọn. Hơn nữa, SSE song song và động [41] cho phép xây dựng hiệu quả hơn và có thể mở rộng dựa trên cấu trúc dữ liệu đa bản đồ dựa trên cây màu đỏ và đen. Mặt khác, các chương trình SSE với các chức năng được cải thiện nhưng bảo mật bị tổn hại đã được đề xuất. Kuzu và cộng sự. [44] đã sử dụng băm nhạy cảm với độ chính xác cao (LSH), được sử dụng rộng rãi để tìm kiếm sự tương đồng nhanh chóng trong không gian chiều cao cho dữ liệu thuần túy, và đề xuất một sơ đồ tìm kiếm để cho phép tìm kiếm sự tương đồng nhanh chóng trong bối cảnh dữ liệu được mã hóa. Một cách tiếp cận khác, được đề xuất bởi Wang và Cao và cộng sự [80] để bảo mật tìm kiếm từ khóa được xếp hạng trong dữ liệu đám mây được mã hóa. Phương pháp này sử dụng Mã hóa đối xứng bảo vệ đơn hàng (OPSE) [8], [9], đạt được cả bảo mật và bảo mật bằng cách bảo vệ thông tin có trọng số nhạy cảm.

Đối với lưu trữ đám mây có thể truy cập được với nhiều người dùng, cách thực thi các đặc quyền và kiểm soát truy cập trong khi tìm kiếm thông qua lưu trữ đám mây đã thu hút sự chú ý của các nhà nghiên cứu. Một cách tiếp cận đã được đề xuất bởi Singh và Srivatsa et al. [71] vào năm 2009, trong đó thực hiện lập chỉ mục trong lĩnh vực doanh nghiệp đáng tin cậy và sử dụng các chỉ số kết quả một cách có hệ thống với Thùng Kiểm soát Truy cập (ACB) [70] nguyên thủy và các khái niệm về phân cấp truy cập của người dùng. Giải pháp này cải thiện hiệu quả lập chỉ mục và cho phép chuyển các chỉ mục sang SSP để lưu trữ và nó có thể được phát triển dựa trên tính toàn vẹn của kết quả tìm kiếm được SSP trả về trong tương lai.

Ngoài các thuật toán tìm kiếm trên cơ sở dữ liệu được mã hóa, tính toán tổng quát hơn trên cơ sở dữ liệu được mã hóa là một chủ đề liên quan. Tính toán bảo mật trên cơ sở dữ liệu được mã hóa (SCONEDB) [90] đã được đề xuất để giải quyết việc tính toán k-Nearest Neighbor (kNN) trong một cơ sở dữ liệu được mã hóa sử dụng sản phẩm quy mô nhỏ không đối xứng để bảo tồn mã hóa (ASPE). Bên cạnh đó, SCONEDB có thể kết hợp các kỹ thuật hiện có khác, chẳng hạn như OPSE cho truy vấn phạm vi và mã hóa đồng hình cho các truy vấn tổng hợp. CryptDB [60] đã triển khai một hệ thống tích hợp hỗ trợ các hoạt động truy vấn SQL tổng quát hơn trên cơ sở dữ liệu mã hóa, bằng cách điều chỉnh một số nguyên tắc mã hóa SQL hiện có và mới với các thuộc tính và chức năng bảo mật khác nhau. CryptDB tự động điều chỉnh các chiến lược mã hóa bằng cách sử dụng các cấu trúc lớp củ hành, trong đó mỗi dữ liệu được mã hóa ngày càng mạnh hơn, sao cho lớp ngoài cùng cung cấp bảo mật tối đa, trong khi các lớp bên trong cung cấp nhiều chức năng hơn. Một proxy đáng tin cậy chấm dứt cho dù các lớp mã hóa cần phải được loại bỏ khi nhận được một truy vấn từ ứng dụng người dùng.

## **C.Mã hóa đồng hình hoàn toàn**

Mã hóa đồng hình cho phép các toán tử đại số cụ thể được thao tác trên văn bản mật mã, vì vậy nó có thể tạo ra kết quả được mã hóa giống như văn bản mật mã của kết quả của các hoạt động giống nhau (hoặc khác nhau nhưng đã biết) được thực hiện trên văn bản thuần túy. Nói cách khác, các hoạt động được thực hiện trên dữ liệu gốc bây giờ nó có thể được thực hiện trên văn bản mã hóa mà không cần biết dữ liệu gốc. Mã hóa đồng hình có thể được phân loại thành hai loại: mã hóa đồng hình một phần (PHE) và mã hóa đồng hình hoàn toàn (FHE). PHE chỉ cho phép một thao tác đồng hình, bổ sung (ví dụ: Paillier [56]) hoặc nhân (ví dụ: RSA chưa được thêm vào), trong khi FHE hỗ trợ cả thao tác cộng và nhân. Kể từ khi thuật toán RSA không đệm ban đầu được xuất bản năm 1977, đã có nhiều thuật toán PHE được phát triển. Tuy nhiên, tính chất đồng hình par của một thuật toán mã hóa hiếm khi được coi là có lợi, nhưng khá dễ bị tổn thương trước các cuộc tấn công văn bản được chọn thích ứng. Do đó, các thuật toán PHE đã được tìm thấy chỉ hữu ích trong các ứng dụng bảo mật hạn chế như hệ thống bỏ phiếu điện tử. Mặt khác, kể từ khi thuật toán FHE đầu tiên được công bố vào năm 2009 [26], nó đã được công nhận là một bước đột phá lớn trong lĩnh vực bảo mật điện toán. Ứng dụng thực tế của hệ thống mã hóa FHE sẽ có khả năng cho phép phát triển các chương trình máy tính, chạy trên dữ liệu đầu vào được mã hóa để tạo ra đầu ra được mã hóa. Do đó, các chương trình này có thể được chạy bởi các thực thể không đáng tin cậy mà không tiết lộ bất kỳ thông tin nhạy cảm nào trong quá trình tính toán.

Một hệ thống mật mã đồng hình ε bao gồm bốn thuật toán, KeyGenε, Encryptε, Decryptε, và một algo- rithm Assessε. Ba thuật toán đầu tiên được định nghĩa giống như những thuật toán trong bất kỳ hệ thống mật mã khóa công khai nào. KeyGenε (λ) tạo ra cặp khóa (pk, sk) với tham số bảo mật λ. Thuật toán Encryptε lấy pk và một văn bản thuần túy π làm đầu vào, và nó xuất ra một văn bản mật mã φ. Giải mãε lấy sk và φ làm đầu vào, và xuất văn bản thuần túy π. Ngoài ra, thuật toán Đánh giáε lấy đầu vào pk, một mạch C từ một tập hợp cho phép Cε và một tập hợp các văn bản mật mã φ = (φ1, φt...), do đó xuất ra một văn bản mật mã φ. Hệ thống mật mã đồng hình ε đúng với Cε nếu đối với bất kỳ cặp khóa nào (pk, sk) được tạo bởi KeyGenε (λ), bất kỳ mạch C , Cε,bất kỳ văn bảnthuần túy nào π1, ..., πt, và bất kỳ văn bản mậtmãnàoφ=(φ1, ... → φt) với φi Encryptε (pk, πi), đó là trường hợp

Ifφ ← Đánhgiá ε (pk, C, ϕ),

ThenDecryptε (sk, φ) → C(π1, ...,πt)

Độ phức tạp tính toán của tất cả các thuật toán trên phải là đa thức trong kích thước của C và tham số mức độ bảo mật λ, mà được định nghĩa là tất cả các cuộc tấn công đã biết chống lại các lược đồ mất thời gian ít nhất2λ. ε là hoàn toàn đồng hình nếu nó là đồng hình cho tất cả các mạch [26].

Một họ các lược đồ ε(d) : d ,Z +được san bằng hoàn toàn đồng hình nếu tất cả chúng sử dụng cùng một mạch giải mã, ε(d) là đồng hình cho tất cả các mạch có độ sâu nhiều nhất làd (sử dụng một số tập hợp cổngđược chỉ định) và độ phức tạp tính toán của các thuật toán của ε(d) là đa thức trong λ, d, và (trong trường hợp Đánhgiáε(d)) kích thước của C.

Đề án FHE đầu tiên được Craig Gentry đề xuất vào năm 2009 [26] áp dụng mật mã dựa trên mạng để xây dựng đề án, trong đó mạng L là một tập hợp các điểm trong không gian Euclide không gian n chiều Rn với đặc tính tuần hoàn mạnh. Đề án được đề xuất bắt đầu từ một sơ đồ mã hóa đồng dạng sử dụng các mạng lý tưởng, được giới hạn ở đa thức đánh giá "mức độ thấp" trên dữ liệu được mã hóa do sự gia tăng của tiếng ồn trong văn bản mật mã trong quá trình đánh giá. Sau giai đoạn "xây dựng ban đầu" này, một "squash mạch giải mã" đã được sử dụng để sửa đổi sơ đồ để làm cho nó "bootstrappable".Chương trình mã hóa được sửa đổi có thể đánh giá mạch giải mã của chính nó và làm mới hiệu quả văn bản mật mã để giảm tiếng ồn tăng cường, điều này loại bỏ giới hạn về độ sâu của mạch được đánh giá qua văn bản mật mã. Tóm lại, Craig Gentry đã sửa đổi nhẹ một số - mã hóa đồng hình bằng cách tự nhúng đệ quy. Chương trình kết quả có thể làm giảm tiếng ồn tích lũy gây ra bởi nhiều hoạt động đại số, do đó làm cho nó có thể nhận ra FHE trong chiều sâu tùy ý.

Tuy nhiên, sơ đồ FHE đầu tiên này là không thực tế vì độ phức tạp tính toán và kích thước văn bản mật mã là đa thức bậc cao trong tham số mức độ bảo mật λ, có nghĩa là chúng tăng mạnh để đạt được mức độ bảo mật thực tế đủ cao. Điều này cấm ứng dụng thực tế của FHE, đặc biệt là trong bối cảnh điện toán đám mây, nơi mức độ bảo mật cao là rất quan trọng. Một mối quan tâm chính của chương trình này là bảo mật của nó dựa trên hai giả định tương đối mới, cụ thể là, độ cứng của bài toán giải mã khoảng cách giới hạn trong trường hợp xấu nhất (BDD) trên mạng lý tưởng và độ cứng của bài toán tổng tập con nhỏ trung bình (SSSP) của bước vắt. Cả hai đều là những giả định mật mã tương đối chưa được kiểm chứng.

Gần đây hơn, đã có những nỗ lực nghiên cứu ngày càng tăng trong việc tìm kiếm các thuật toán FHE thực tế, hiệu quả hơn và/hoặc dựa trên các giả định bảo mật đáng tin cậy hơn. Một phiên bản thứ hai của kế hoạch FHE, được gọi là DGHV, đã được đề xuất bởi Marten van Dijk, Craig Gentry, Shai Halevi, và Vinod Vaikuntanathan vào năm 2010 [77]. DGHV sử dụng các kỹ thuật của Gentry với chỉ số học mô-đun sơ cấp trên các số nguyên để chuyển đổi một sơ đồ phần nào đồng hình thành một sơ đồ FHE có thể khởi động. Sơ đồ này đạt được sự đơn giản về khái niệm vì tất cả các phép tính được thực hiện trên các số nguyên thay vì mạng lý tưởng. Nó cũng làm giảm giả định bảo mật đến độ cứng của vấn đề bộ chia phổ biến nhất (GCD). Tuy nhiên, giá của sự đánh đổi này là kích thước khổng lồ của khóa công khai, điều này có thể không thực tế đối với các hệ thống hiện tại. Stehle và Steinfeld [75] đã trình bày một mã hóa đồng hình nhanh hơn để cải thiện sơ đồ của Gentry bằng cách phân tích tích tích cực hơn về giả định SSSP, và giới thiệu một thuật toán giải mã xác suất được thực hiện bởi một mạch đại số có cấp số nhân thấp. Với hai cải tiến này, chương trình này có được O(λ3.5) độ phức tạp bit để làm mới một văn bản mật mã, trong khi chương trình trước đó tuyên bố O(λ6) cho cùng một nhiệm vụ, trong đó λ là tham số bảo mật. Tuy nhiên, có một xác suất khác không của lỗi giải mã liên quan đến sơ đồ này.

Bên cạnh đó, Zvika Brakerski và Vinod Vaikuntanathan đã cải tiến kế hoạch của Gentry [14] bằng cách thay đổi hai giả định bảo mật được đưa ra trong [26]. Đầu tiên, mã hóa đồng hình phần nào được dựa trên học vòng với các lỗi (RLWE) giả định từ Lyubashevsky, Peikert và L. Regev [49] thay vì vấn đề BDD mạng lý tưởng. Thứ hai, để làm cho hệ thống mã hóa hơi đồng hình dạng khởi động, nó sử dụng một kỹ thuật giảm kích thước-mod moode thay vì kỹ thuật squashing của Gentry, do đó loại bỏ giả định của SSSP. Kỹ thuật bootstrapping mới này cũng rút ngắn văn bản mật mã và giảm độ phức tạp. Dựa trên sự cải thiện trên, Brakerski, Gentry, và Vaikuntanathan đã làm việc cùng nhau để đề xuất một kế hoạch FHE mới mà không cần thủ tục khởi động của Gentry vào năm 2011 [13]. Bằng cách áp dụng RLWE, sơ đồ FHE này có tính toánO (λ · L3) cho mỗi cổng cho các mạch số học cấpL. Như một cách tiếp cận tùy chọn, họ cũng đề xuất một kế hoạch FHE được san bằng sử dụng bootstrapping như tối ưu hóa để tiếp tục giảm tính toán mỗi cổng xuống O(λ2), không phụ thuộc vào L.

Tiếp theo vào năm 2011, Coron et al. đã đề xuất một cải tiến của FHE so với các số nguyên được mô tả bởi van Dijk et al... Lược đồ mới được đề xuất đã rút ngắn kích thước khóa công khai từ O(λ10) thành O(λ7) [19]. Thủ tục này được thực hiện bằng cách sử dụng dạng bậc hai thay vì dạng tuyến tính trong các phần tử khóa công khai, để khóa công khai có độ dài đầy đủ được nén thành một tập hợp con nhỏ hơn của khóa ban đầu. Thay vì đề xuất bất kỳ cải tiến nào hơn nữa trên FHE, Ron Rothblum thể hiện cách chuyển đổi bất kỳ sơ đồ mã hóa khóa riêng đồng nhất nào thành sơ đồ mã hóa khóa công khai [62]. Để xây dựng quá trình này, sơ đồ này phát triển một định lý mà bất kỳ phụ gia nhỏ gọn đồng hình đối với bổ sung modulo hai có thể được biến đổi thành sơ đồ ngữ nghĩa. Kết quả là, sơ đồ mã hóa khóa công khai lưu một bước đồng hình liên quan đến cùng một tập hợp các hoạt động với mã hóa khóa riêng tư, là các sơ đồ FHE trước đó.

Với tất cả các phát triển lý thuyết của FHE al- gorithms khác nhau, nó là cần thiết để điều tra thực hiện thực tế của họ mentation. Có một số triển khai FHE của Gentry vào năm 2010, và nỗ lực đầu tiên đã được thực hiện bởi Smart và Vercauteren [72]. Họ đã có thể thực hiện sơ đồ đồng hình phần nào bằng cách sử dụng"mạngnguyên lý" của yếu tố quyết định chính, chỉ có thể được ngụ ý bởi hai số nguyên. Tuy nhiên, họ không thể thực hiện chức năng khởi động để có được sơ đồ đồng hình hoàn toàn. Cái cổ chai của việc thực hiện này là sự thất bại trong việc hỗ trợ một lượng lớn các thông số.

Dựa trên công việc này, vào năm 2011, Gentry và Halevi đã phát triển một loạt các đơn giản hóa và tối ưu hóa mà làm cho việc thực hiện dây buộc ủng có thể thực hiện được. Kết quả là độ phức tạp toàn phần được giảm từ O(n2.5) của Smart và Vercauteren xuốngO(n1.5). Các mức tối ưu hóa từ bài báo này cũng được sử dụng trong [19] để thực hiện sơ đồ DGHV đồng hình hoàn toàn theo biến thể mới. Với kết quả là có hiệu suất tương tự, Coron et al. đã cho thấy thành công rằng FHE có thể được thực hiện với một sơ đồ số học đơn giản.

Trong Bảng II, chúng tôi cung cấp một so sánh hiệu suất cho các chương trình FHE khác nhau. BDD và SSSP là những vấn đề đã được nêu trong sơ đồ FHE đầu tiên.

## **D.Cách tiếp cận bảo mật dữ liệu khác**

Có một số phương pháp bảo mật dữ liệu khác ở trên. Ví dụ: Việc áp dụng các thuật toán mật mã cho các khối dữ liệu trong lưu trữ đám mây là một phương pháp phổ biến được sử dụng để đảm bảo tính bảo mật của dữ liệu được lưu trữ. Một chương trình bảo mật dữ liệu trong CoreFS, là một hệ thống tệp mạng cấp độ người dùng, đã được đề xuất vào năm 2009 [93]. Sơ đồ này được xây dựng dựa trên MAC phổ quát-hash stateful mới. Nó có chi phí tính toán nhỏ hơn của các hoạt động mật mã so với MHT. Bên cạnh đó, nó cho phép khả năng giao tiếp tốt hơn. Tuy nhiên, việc lựa chọn chiến lược bộ nhớ cache, lịch cập nhật cây MAC và phương pháp lưu trữ cây có thể ảnh hưởng đến hiệu suất của lược đồ này.

Một chương trình bảo mật dữ liệu khác đã khai thác mô hình chứng minh an toàn (SP) mới được đề xuất dựa trên các kỹ thuật ghép nối song tuyến tính vào năm 2010 [48]. Sơ đồ này ghi lại quyền sở hữu và lịch sử quá trình của các đối tượng dữ liệu trong lưu trữ đám mây để tăng niềm tin từ người dùng công cộng. Mô hình SP bao gồm các mô-đun sau: thiết lập hệ thống, tạo khóa, xác thực ẩn danh, truy cập được ủy quyền và theo dõi nguồn gốc. Kỹ thuật bảo mật được chứng minh đã được thử nghiệm trên hệ thống này theo mô hình SP tiêu chuẩn. Nó đòi hỏi một số cân nhắc thực tế trong các ứng dụng thực tế và cải thiện hơn nữa theo khuôn khổ hiện tại.

Khác với các kế hoạch trên, một tuyên bố thuyết phục đã được đề xuất bởi Dijk và Juel vào năm 2010 [78], tuyên bố rằng không có giao thức mật mã nào, ngay cả bao gồm các nguyên thủy quyền lực như FHE, có thể thực thi quyền riêng tư được yêu cầu bởi các dịch vụ đám mây chung một mình. Bài báo cũng chứng minh rằng thay vào đó, nhu cầu trên có thể đạt được bằng các biện pháp thực thi khác, chẳng hạn như phần cứng chống giả mạo, điện toán phân tán và hệ sinh thái ủy thác phức tạp.

## **E. Tóm tắt**

Bảo mật dữ liệu là một trong những vấn đề quan trọng nhất đối với các ứng dụng có dữ liệu nhạy cảm, chẳng hạn như thông tin cá nhân, thông tin tài khoản của khách hàng, tài chính và chăm sóc sức khỏe đang hình thành. Thách thức mới của việc lưu trữ những dữ liệu này trên đám mây là làm thế nào để ngăn chặn rò rỉ dữ liệu vô tình hoặc cố ý cho nhà cung cấp dịch vụ lưu trữ đám mây, sao cho ngay cả khi nhà cung cấp dịch vụ bị xâm phạm, thông tin vẫn có thể được giữ bí mật. Giải pháp cơ bản cho vấn đề này vẫn là mã hóa dữ liệu. Người dùng phải mã hóa dữ liệu trước khi chúng được chuyển đến máy chủ đám mây và giữ nó được mã hóa trong toàn bộ khoảng thời gian dữ liệu ở trên đám mây. Khi dữ liệu cần được truy cập hoặc xử lý bởi chủ sở hữu dữ liệu hoặc người dùng hợp pháp khác có khóa để giải mã, việc truy xuất dữ liệu được mã hóa, giải mã, xử lý và mã hóa lại dữ liệu trước khi gửi trở lại máy chủ là không hiệu quả. Do đó, các cơ chế mã hóa mới cho phép xử lý văn bản mật mã trực tiếp mà không tiết lộ thông tin gốc trong văn bản thuần túy sẽ có tiềm năng đáng kể trong việc lưu trữ dữ liệu nhạy cảm trên đám mây. Mã hóa WIth, chủ sở hữu dữ liệu hoặc người dùng lấy lại quyền kiểm soát đối với dữ liệu của họ không được lưu trữ vật lý bởi chính họ.

Graphical user interface, text, application

Description automatically generated

*BẢNG II: So sánh hiệu suất của các chương trình FHE*

FHE là một ví dụ lý tưởng về các thuật toán mã hóa này. Tuy nhiên, các ứng dụng đầy hứa hẹn của FHE hiện tại algo- rithms bị cản trở bởi độ phức tạp tính toán của nó và các khó khăn thực hiện khác. Cải tiến phải được thực hiện- trước khi nó có thể được đưa vào ứng dụng thực tế. Ngoài ra, nhiều triển khai cải tiến khác nhau đang chờ đợi để được đánh giá trên các nền tảng hiện tại.

Không giống như FHE, có mục tiêu tham vọng là tính toán tùy ý trên văn bản mật mã, các chương trình mã hóa đám mây khác nhằm vào loại kiểm soát cụ thể đối với dữ liệu được mã hóa. Ví dụ, ABE cho phép kiểm soát truy cập được thực thi trên dữ liệu được mã hóa bằng cách kết hợp cấu trúc truy cập dựa trên thuộc tính vào văn bản mật mã hoặc khóa giải mã. Các lược đồ mã hóa có thể tìm kiếm cung cấp một cách để tìm kiếm văn bản mật mã cho mã thông báo từ khóa mà không tiết lộ nội dung thực của dữ liệu hoặc từ khóa.

# **V.Mức Độ Khả Dụng**

Là một biện pháp bảo mật khác, tính khả dụng trong lưu trữ đám mây đề cập đến việc dữ liệu có thể truy cập và sử dụng được khi người dùng tự động yêu cầu chúng từ bất kỳ máy nào vào bất kỳ lúc nào. Trong giai đoạn trước đó của điện toán đám mây, tính sẵn có là mối quan tâm bảo mật nhiều hơn do thiếu sự trưởng thành và đáng tin cậy trong cấu trúc. Nhiều sự cố không có dịch vụ xảy ra do lỗi phần cứng và dẫn đến hậu quả nghiêm trọng. Có cơ sở hạ tầng tốt hơn và đáng tin cậy hơn, thách thức phải đối mặt với sự sẵn có của dịch vụ lưu trữ đám mây là làm thế nào để bảo tồn dữ liệu của người dùng trong trường hợp khẩn cấp, chẳng hạn như thiên tai.

Giải pháp đơn giản nhất là giữ các bản sao lưu dữ liệu ở nhiều vị trí vật lý. Amazon EC2 và S3 cung cấp một ví dụ hoàn hảo dựa trên khu vực có sẵn, nằm trong các khu vực địa lý được phân chia, ví dụ: Chúng tôi - phía Tây và Mỹ - phía Đông. Mỗi khu vực chứa một số trường hợp có cùng dữ liệu. Khi xảy ra tai nạn, Amazon EC2 và S3 có thể dễ dàng khôi phục dữ liệu bị hỏng hoặc mất từ các vùng có sẵn khác trong cùng khu vực để tiết kiệm điện năng và thời gian. Cách tiếp cận này không hiệu quả về việc sử dụng tài nguyên lưu trữ.

Đã có các chương trình quản lý lưu trữ sao lưu, chẳng hạn như sao lưu gia tăng và sao chép dữ liệu, được phát triển để cải thiện việc sử dụng lưu trữ. Sao lưu tăng dần đã được sử dụng rộng rãi trong các dịch vụ sao lưu tệp. Nó khai thác mối tương quan giữa các tệp hiện tại với phiên bản sao lưu trước đó và chỉ lưu trữ sự khác biệt. Khi sao lưu gia tăng được triển khai ở mức khối dữ liệu hoặc thậm chí mức byte dữ liệu, nó trở nên hiệu quả hơn trong việc sử dụng lưu trữ, nhưng với chi phí xử lý cao hơn. Mã hóa Delta là một ví dụ sao lưu gia tăng nổi tiếng được áp dụng bởi Dropbox. Sự sao chép dữ liệu là một kỹ thuật nén dữ liệu chuyên dụng để xác định các phân đoạn dữ liệu phổ biến trong và trên các tệp khác nhau và chỉ lưu trữ chúng một lần để cải thiện việc sử dụng lưu trữ. Thật không may, sự sao chép dữ liệu có khả năng làm suy yếu tính bảo mật dữ liệu về cả tính toàn vẹn và bảo mật dữ liệu. Đầu tiên, theo định nghĩa, sự sao chép dữ liệu làm thay đổi dữ liệu gốc từ người dùng và lưu trữ chúng ở một hình thức khác trong lưu trữ đám mây, do đó dẫn đến những lo ngại về tính toàn vẹn của dữ liệu.Thứ hai, dữ liệu dedupli- cation cố gắng xác định và khai thác các đoạn dữ liệu giống hệt nhau, trong khi các thuật toán mã hóa thường cố gắng ngẫu nhiên hóa chúng để che giấu nội dung thực. Văn bản mã hóa cho cùng một văn bản thuần túy có thể sẽ cực kỳ khác nhau. Để giải quyết những vấn đề này, việc sao chép dữ liệu hiệu quả và an toàn cho phép sao chép dữ liệu được thực hiện trên văn bản mã hóa đã được phát triển [76]. Kỹ thuật này sử dụng mã hóa hội tụ [], trong đó khóa mã hóa được tạo bằng cách sử dụng hàm băm của văn bản thuần túy của đoạn dữ liệu. Do đó, cùng một đoạn dữ liệu văn bản thuần túy sẽ được mã hóa bằng cùng một khóa, bất kể khi nào và bởi ai nó được mã hóa. Điều này dẫn đến cùng một đoạn dữ liệu văn bản mật mã cho cùng một văn bản thuần túy. Chương trình lưu trữ các mảng dữ liệu hoặc byte duy nhất trong quá trình phân tích dữ liệu và sau đó so sánh các mảng khác với dữ liệu được lưu trữ. Nếu kết quả được so sánh trùng khớp, thì phần dư thừa sẽ được thay thế bằng một con trỏ nhỏ trỏ đến vị trí của dữ liệu được lưu trữ trùng khớp.

Một cách tiếp cận chủ động khác là dự đoán khả năng xảy ra sự cố trong tương lai để có thể thực hiện các hành động sớm hơn nhằm tránh gián đoạn dịch vụ. Guan và cộng sự đã đề xuất hai phương pháp tiếp cận học hỏi để dự đoán động lực thất bại trong các hệ thống điện toán đám mây bằng cách sử dụng các phương pháp Bayes và cây quyết định [30]. Một giai đoạn ban đầu là cần thiết cho dữ liệu giám sát, và sau đó một tập hợp các phương pháp Bayes dán nhãn dữ liệu có các hành vi bất thường. Sau khi tất cả các bất thường được xác định, mô hình có thể dự đoán các trường hợp thất bại trong tương lai dựa trên các bộ phân loại cây quyết định.

Khi sự cố đã xảy ra, các chương trình khôi phục dữ liệu là cần thiết để giảm hoặc loại bỏ tổn thất. Zhang [96] đã trình bày một phương pháp khôi phục dữ liệu kiểm tra thiệt hại trong cơ sở dữ liệu đám mây hạt mịn và cho phép chủ sở hữu cơ sở dữ liệu đám mây biết và xác định vị trí thiệt hại chính xác cho mục đích khôi phục. Thuật toán phân tán thông tin [61] được sử dụng để cho phép tính khả dụng cao hơn của dữ liệu khi gặp phải sự cố vật lý và mất mạng.

Bên cạnh các kỹ thuật trên, khôi phục dữ liệu cũng có thể đạt được bằng khung dịch vụ mới. Chi-won Song và cộng sự đã đề xuất Parity Cloud Service (PCS) vào năm 2011 [89]. Nó tạo ra đĩa ảo trong hệ thống người dùng để sao lưu riêng tư và tạo ra một nhóm ngang hàng của nhiều người dùng. Dữ liệu tương tự giữa những người dùng trong nhóm bình đẳng được lưu trữ ở đầu máy chủ. Do đó, khi người dùng phát hiện ra rằng tệp gốc yêu cầu khôi phục, họ có thể yêu cầu dữ liệu từ máy chủ cuối cùng mà không vi phạm quyền riêng tư vì sao lưu riêng tư được lưu trữ trên đĩa ảo của mỗi người dùng. Cách tiếp cận này đơn giản và an toàn, nhưng mỗi người dùng phải xây dựng một đĩa ảo, tốn thêm chi phí cho người dùng.

Tóm lại, đảm bảo tính khả dụng của dữ liệu của người dùng bất cứ khi nào người dùng yêu cầu là yêu cầu cơ bản và chính trong lưu trữ đám mây. Thách thức chính nảy sinh khi xem xét các mối quan tâm an ninh và chính trị khác. Sự cân bằng giữa hiệu quả và độ tin cậy phải được thực hiện để cân bằng lợi ích của nhà cung cấp dịch vụ và người sử dụng. Hơn nữa, tính toàn vẹn dữ liệu và tính bảo mật dữ liệu không nên bị xâm phạm bởi sự cải thiện tính sẵn có của dữ liệu.

## **VI. KẾT LUẬN VÀ CÔNG VIỆC TƯƠNG LAI**

Với xu hướng triển khai nhanh chóng của lưu trữ đám mây và điện toán hiện nay, việc các hệ thống lưu trữ đám mây được trang bị các giải pháp bảo mật được chứng minh là đáng tin cậy và đáng tin cậy là điều cần thiết. Trong công việc này, chúng tôi đã tiến hành một cuộc khảo sát về các nguyên tắc được đề xuất hoặc phát triển gần đây nhất để đảm bảo ba trong số các phép đo bảo mật quan trọng nhất, cụ thể là tính toàn vẹn dữ liệu, tính bảo mật dữ liệu và tính khả dụng của các hệ thống lưu trữ đám mây. Đối với mỗi khía cạnh, chúng tôi đã xác định những thách thức duy nhất khác với những thách thức trong mạng dữ liệu hoặc hệ thống lưu trữ tệp truyền thống, tóm tắt tiến trình phát triển hiện có cho đến nay và cung cấp thông tin chi tiết về các hướng nghiên cứu trong tương lai. Nhìn chung, chúng tôi cảm thấy rằng bảo mật lưu trữ đám mây vẫn còn sơ khai và kỳ vọng sẽ có bước đột phá nổi bật hơn trong tương lai gần. Ví dụ, mặc dù các giải pháp bảo mật lưu trữ đám mây đã được phát triển nhanh chóng trong những năm gần đây, chúng tôi vẫn chưa thấy một mô hình được chấp nhận rộng rãi để triển khai. Bên cạnh thiết kế hệ thống, hệ thống bảo mật lưu trữ đám mây cần đủ linh hoạt để có thể được cải thiện bằng các thuật toán mật mã mới.

**Sự Xác Nhận**

Công trình này được hỗ trợ bởi Phòng thí nghiệm trọng điểm về mạng và an toàn thông tin tỉnh Giang Tô với số tài trợ BM2003201, và Phòng thí nghiệm trọng điểm An ninh mạng thông tin, Bộ Công an.

# **Tài liệu tham khảo**

Công trình này được hỗ trợ bởi Phòng thí nghiệm trọng điểm về mạng và an toàn thông tin tỉnh Giang Tô với số tài trợ.

**THAM KHẢO**

*[1] Gail-Joon Ahn and Ravi Sandhu. Role-based authorization constraints specification. ACM Trans. Inf. Syst. Secur., 3:207–226, November 2000.*

*[2] Mikhail J. Atallah, Marina Blanton, Nelly Fazio, and Keith B. Frikken. Dynamic and efficient key management for access hierarchies. ACM Trans. Inf. Syst. Secur., 12(3):18:1–18:43, January 2009.*

*[3] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Osama Khan, Lea Kissner, Zachary Peterson, and Dawn Song. Remote data checking using provable data possession. ACM Trans. Inf. Syst. Secur., 14(1):12:1–12:34, June 2011.*

*[4] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson, and Dawn Song. Provable data possession at untrusted stores. In Proceedings of the 14th ACM conference on Computer and communications security, CCS ’07, pages 598–609, New*

*York, NY, USA, 2007. ACM.*

*[5] Giuseppe Ateniese, Roberto Di Pietro, Luigi V. Mancini, and Gene Tsudik. Scalable and efficient provable data possession. In Proceedings of the 4th international conference on Security and privacy in communication netowrks, SecureComm ’08, pages 9:1–9:10, New York, NY, USA, 2008. ACM.*

*[6] Mihir Bellare, Alexandra Boldyreva, and Adam O’Neill. Deterministic and efficiently searchable encryption. In Proceedings of the 27th annual international cryptology conference on Advances in cryptology, CRYPTO’07, pages 535–552, Berlin, Heidelberg, 2007. SpringerVerlag.*

*[7] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In Proceedings of the 2007 IEEE Symposium on Security and Privacy, SP ’07, pages 321–334, Washington, DC, USA, 2007. IEEE Computer Society.*

*[8] Alexandra Boldyreva, Nathan Chenette, Younho Lee, and Adam O’Neill. Order-preserving symmetric encryption. In Proceedings of the 28th Annual International Conference on Advances in Cryptology: the Theory and Applications of Cryptographic Techniques, EUROCRYPT ’09, pages 224–241, Berlin, Heidelberg, 2009. Springer-Verlag.*

*[9] Alexandra Boldyreva, Nathan Chenette, and Adam O’Neill. Orderpreserving encryption revisited: Improved security analysis and alternative solutions. In Proceedings of the 31st Annual Conference on Advances in Cryptology, CRYPTO’11, pages 578–595, Berlin, Heidelberg, 2011. Springer-Verlag.*

*[10] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano. Public key encryption with keyword search. In Christian Cachin and Jan Camenisch, editors, Advances in Cryptology - EUROCRYPT 2004, volume 3027 of Lecture Notes in Computer Science, pages 506–522. Springer Berlin / Heidelberg, 2004.*

*[11] Dan Boneh, Ben Lynn, and Hovav Shacham. Short signatures from the weil pairing. In Colin Boyd, editor, Advances in Cryptology ASIACRYPT 2001, volume 2248 of Lecture Notes in Computer Science, pages 514–532. Springer Berlin / Heidelberg, 2001.*

*[12] Kevin D. Bowers, Ari Juels, and Alina Oprea. Proofs of retrievability: theory and implementation. In Proceedings of the 2009 ACM workshop on Cloud computing security, CCSW ’09, pages 43–54, New York, NY, USA, 2009. ACM.*

*[13] Zvika Brakerski, Craig Gentry, and Vinod Vaikuntanathan. Fully homomorphic encryption without bootstrapping. Cryptology ePrint Archive, Report 2011/277, 2011.*

*[14] Zvika Brakerski and Vinod Vaikuntanathan. Fully homomorphic encryption from ring-lwe and security for key dependent messages. In Proceedings of the 31st annual conference on Advances in cryptology, CRYPTO’11, pages 505–524, Berlin, Heidelberg, 2011. SpringerVerlag.*

*[15] Darryl Chantry. Mapping applications to the cloud. Technical report, January 2009.*

*[16] Melissa Chase. Multi-authority attribute based encryption. In Theory of Cryptography, 4th Theory of Cryptography Conference, volume 4392 of Lecture Notes in Computer Science, pages 515–534. Springer, 2007.*

*[17] Melissa Chase and Sherman S. M. Chow. Improving privacy and security in multi-authority attribute-based encryption. In Proceedings of the 2009 ACM Conference on Computer and Communications Security, pages 121–130. ACM, 2009.*

*[18] Ling Cheung and Calvin Newport. Provably secure ciphertext policy abe. In Proceedings of the 14th ACM conference on Computer and communications security, CCS ’07, pages 456–465, New York, NY, USA, 2007. ACM.*

*[19] Jean-Sebastien Coron, Avradip Mandal, David Naccache, and Mehdi Tibouchi. Fully homomorphic encryption over the integers with shorter public keys. In Proceedings of the 31st annual conference on Advances in cryptology, CRYPTO’11, pages 487–504, Berlin, Heidelberg, 2011. Springer-Verlag.*

*[20] Reza Curtmola, Juan Garay, Seny Kamara, and Rafail Ostrovsky. Searchable symmetric encryption: Improved definitions and efficientconstructions. J. of Computer Security, 19(5):895–934, 2011.*

*[21] Reza Curtmola, Osama Khan, and Randal Burns. Robust remote data checking. In Proceedings of the 4th ACM international workshop on Storage security and survivability, StorageSS ’08, pages 63–68, New York, NY, USA, 2008. ACM.*

*[22] Cecile Delerabl ´ ee. Identity-based broadcast encryption with constant size ciphertexts and private keys. In Proceedings of the Advances in Crypotology 13th international conference on Theory and application of cryptology and information security, ASIACRYPT’07, pages 200–215, Berlin, Heidelberg, 2007. Springer-Verlag.*

*[23] Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. Encryption policies for regulating access to outsourced data. ACM Trans. Database Syst., 35(2), 2010.*

*[24] Yevgeniy Dodis, Salil Vadhan, and Daniel Wichs. Proofs of retrievability via hardness amplification. In Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography, TCC ’09, pages 109–127, Berlin, Heidelberg, 2009. Springer-Verlag.*

*[25] Chris Erway, Alptekin Kupc¸ ¨ u, Charalampos Papamanthou, and Roberto Tamassia. Dynamic provable data possession. In Proceedings of the 16th ACM conference on Computer and communications security, CCS 09, pages 213–222, New York, NY, USA, 2009. ACM.*

*[26] Craig Gentry. Fully homomorphic encryption using ideal lattices. In Proceedings of the 41st annual ACM symposium on Theory of computing, STOC ’09, pages 169–178, New York, NY, USA, 2009. ACM.*

*[27] Craig Gentry and Shai Halevi. Implementing gentry’s fullyhomomorphic encryption scheme. In Proceedings of the 30th Annual international conference on Theory and applications of cryptographic techniques: advances in cryptology, EUROCRYPT’11, pages 129–148, Berlin, Heidelberg, 2011. Springer-Verlag.*

*[28] Vipul Goyal, Abhishek Jain, Omkant Pandey, and Amit Sahai. Bounded ciphertext policy attribute based encryption. In Automata, Languages and Programming, 35th International Colloquium, 2008, volume 5126 of Lecture Notes in Computer Science, pages 579–591. Springer, 2008.*

*[29] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attributebased encryption for fine-grained access control of encrypted data. In Proceedings of the 13th ACM conference on Computer and communications security, CCS ’06, pages 89–98, New York, NY, USA, 2006. ACM.*

*[30] Qiang Guan, Ziming Zhang, and Song Fu. Proactive failure management by integrated unsupervised and semi-supervised learning for dependable cloud systems. In Proceedings of the 2011 Sixth International Conference on Availability, Reliability and Security, ARES ’11, pages 83–90, Washington, DC, USA, 2011. IEEE Computer Society.*

*[31] Barbara Guttman and Edward A. Roback. Sp 800-12. an introduction to computer security: the nist handbook. Technical report, Gaithersburg, MD, United States, 1995.*

*[32] Shuai Han and Jianchuan Xing. Ensuring data storage security through a novel third party auditor scheme in cloud computing. In Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on, pages 264 –268, sept. 2011.*

*[33] Zhuo Hao, Sheng Zhong, and Nenghai Yu. A privacy-preserving remote data integrity checking protocol with data dynamics and public verifiability. Knowledge and Data Engineering, IEEE Transactions on, 23(9):1432 –1437, sept. 2011.*

*[34] Cheng Hong, Zhiquan lv, Min Zhang, and Dengguo Feng. A secure and efficient role-based access policy towards cryptographic cloud storage. In Proceedings of the 12th international conference on Web-age information management, WAIM’11, pages 264–276, Berlin, Heidelberg, 2011. Springer-Verlag.*

*[35] Eu jin Goh, Hovav Shacham, Nagendra Modadugu, and Dan Boneh. Sirius: Securing remote untrusted storage. In in Proc. Network and Distributed Systems Security (NDSS) Symposium 2003, pages 131–145, 2003.*

*[36] Robert Johnson, David Molnar, Dawn Song, and David Wagner. Homomorphic signature schemes. In Bart Preneel, editor, Topics in Cryptology CT-RSA 2002, volume 2271 of Lecture Notes in Computer Science, pages 204–245. Springer Berlin / Heidelberg, 2002.*

*[37] M. Tim Jones. Anatomy of a cloud storage infrastructure. Technical report, IBM, 2010.*

*[38] Ari Juels and Burton S. Kaliski, Jr. Pors: proofs of retrievability for large files. In Proceedings of the 14th ACM conference on Computer and communications security, CCS ’07, pages 584–597, New York, NY, USA, 2007. ACM.*

*[39] Mahesh Kallahalla, Erik Riedel, Ram Swaminathan, Qian Wang, and Kevin Fu. Plutus: Scalable secure file sharing on untrusted storage. In Proceedings of the 2nd USENIX Conference on File and Storage echnologies, pages 29–42, Berkeley, CA, USA, 2003. USENIX Association.*

*[40] Seny Kamara and Kristin Lauter. Cryptographic cloud storage. In Radu Sion, Reza Curtmola, Sven Dietrich, Aggelos Kiayias, Josep Miret, Kazue Sako, and Francesc Seb?, editors, Financial Cryptography and Data Security, volume 6054 of Lecture Notes in Computer Science, pages 136–149. Springer Berlin / Heidelberg, 2010.*

*[41] Seny Kamara and Charalampos Papamanthou. Parallel and dynamic searchable symmetric encryption. In Financial Cryptography, pages 258–274, 2013.*

*[42] Seny Kamara, Charalampos Papamanthou, and Tom Roeder. Dynamic searchable symmetric encryption. In ACM Conference on Computer and Communications Security, pages 965–976, 2012.*

*[43] D. Richard Kuhn, Edward J. Coyne, and Timothy R. Weil. Adding attributes to role-based access control. Computer, 43(6):79–81, June 2010.*

*[44] Mehmet Kuzu, Mohammad Saiful Islam, and Murat Kantarcioglu. Efficient similarity search over encrypted data. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, ICDE’12, pages 1156–1167, Washington, DC, USA, 2012. IEEE Computer Society.*

*[45] Allison B. Lewko, Tatsuaki Okamoto, Amit Sahai, Katsuyuki Takashima, and Brent Waters. Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. In Henri Gilbert, editor, Advances in Cryptology EUROCRYPT 2010, volume 6110 of Lecture Notes in Computer Science, pages 62–91. Springer Berlin / Heidelberg, 2010.*

*[46] Allison B. Lewko and Brent Waters. Decentralizing attribute-based encryption. In Advances in Cryptology - EUROCRYPT 2011, volume 6632 of Lecture Notes in Computer Science, pages 568–588. Springer, 2011.*

*[47] Mark Lillibridge, Sameh Elnikety, Andrew Birrell, Mike Burrows, and Michael Isard. A cooperative internet backup scheme. In Proceedings of the annual conference on USENIX Annual Technical Conference, ATEC ’03, pages 3–3, Berkeley, CA, USA, 2003. USENIX Association.*

*[48] Rongxing Lu, Xiaodong Lin, Xiaohui Liang, and Xuemin Sherman Shen. Secure provenance: the essential of bread and butter of data forensics in cloud computing, pages 282–292. ACM, 2010.*

*[49] Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. J. ACM, 60(6):43:1–43:35, November 2013.*

*[50] B. Malek and A. Miri. Combining attribute-based and access systems. In Computational Science and Engineering, 2009. CSE ’09. International Conference on, volume 3, pages 305 –312, aug. 2009.*

*[51] Peter Mell and Tim Grance. The nist definition of cloud computing. Technical report, July 2009.*

*[52] Ralph C Merkle. Protocols for Public Key Cryptosystems, pages 122– 134. IEEE Computer Society Press, 1980.*

*[53] Moni Naor and Guy Rothblum. The complexity of online memory checking. Cryptology ePrint Archive, Report 2006/091, 2006.*

*[54] H.A.J. Narayanan and M.H. Gunes. Ensuring access control in cloud provisioned healthcare systems. In Consumer Communications and Networking Conference (CCNC), 2011 IEEE, pages 247 –251, jan. 2011.*

*[55] Rafail Ostrovsky, Amit Sahai, and Brent Waters. Attribute-based encryption with non-monotonic access structures. In Proceedings of the 14th ACM conference on Computer and communications security, CCS ’07, pages 195–203, New York, NY, USA, 2007. ACM.*

*[56] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In Jacques Stern, editor, Advances in Cryptology EUROCRYPT 99, volume 1592 of Lecture Notes in Computer Science, pages 223–238. Springer Berlin / Heidelberg, 1999.*

*[57] Charalampos Papamanthou, Roberto Tamassia, and Nikos Triandopoulos. Authenticated hash tables. In Proceedings of the 15th ACM conference on Computer and communications security, CCS ’08, pages 437–448, New York, NY, USA, 2008. ACM.*

*[58] Mithun Paul and Ashutosh Saxena. Proof of erasability for ensuring comprehensive data deletion in cloud computing. In Natarajan Meghanathan, Selma Boumerdassi, Nabendu Chaki, and Dhinaharan Nagamalai, editors, Recent Trends in Network Security and Applications, volume 89 of Communications in Computer and Information Science, pages 340–348. Springer Berlin Heidelberg, 2010.*

*[59] Daniele Perito and Gene Tsudik. Secure code update for embedded devices via proofs of secure erasure. In Proceedings of the 15th European conference on Research in computer security, ESORICS’10, pages 643–662, Berlin, Heidelberg, 2010. Springer-Verlag.*

*[60] Raluca Ada Popa, Catherine M. S. Redfield, Nickolai Zeldovich, and Hari Balakrishnan. Cryptdb: Processing queries on an encrypted database. Commun. ACM, 55(9):103–111, September 2012.*

*[61] Michael O. Rabin. Efficient dispersal of information for security, load balancing, and fault tolerance. J. ACM, 36(2):335–348, 1989.*

*[62] Ron Rothblum. Homomorphic encryption: From private-key to publickey. In Yuval Ishai, editor, Theory of Cryptography, volume 6597 of Lecture Notes in Computer Science, pages 219–234. Springer Berlin /Heidelberg, 2011.*

*[63] Amit Sahai, Hakan Seyalioglu, and Brent Waters. Dynamic credentials and ciphertext delegation for attribute-based encryption. In Advances in Cryptology - CRYPTO 2012, volume 7417 of Lecture Notes in Computer Science, pages 199–217. Springer, 2012.*

*[64] Amit Sahai and Brent Waters. Fuzzy identity-based encryption. In Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings, volume 3494 of Lecture Notes in Computer Science, pages 457–473. Springer, 2005.*

*[65] Pierangela Samarati and Sabrina De Capitani di Vimercati. Data protection in outsourcing scenarios: issues and directions. In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, 2010, pages 1–14. ACM, 2010.*

*[66] R.S. Sandhu, E.J. Coyne, H.L. Feinstein, and C.E. Youman. Role-based access control models. Computer, 29(2):38 –47, Feb 1996.*

*[67] Francesc Sebe, Josep Domingo-Ferrer, Antoni Martinez-Balleste, Yves ´ Deswarte, and Jean-Jacques Quisquater. Efficient remote data possession checking in critical information infrastructures. IEEE Trans. On Knowl. and Data Eng., 20:1034–1038, August 2008.*

*[68] Hovav Shacham and Brent Waters. Compact proofs of retrievability. In Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology, ASIACRYPT ’08, pages 90–107, Berlin, Heidelberg, 2008. Springer-Verlag.*

*[69] Hovav Shacham and Brent Waters. Compact proofs of retrievability. J. Cryptology, 26(3):442–83, July 2013.*

*[70] A. Singh, M. Srivatsa, and Ling Liu. Efficient and secure search of enterprise file systems. In Web Services, 2007. ICWS 2007. IEEE International Conference on, pages 18 –25, july 2007.*

*[71] Aameek Singh, Mudhakar Srivatsa, and Ling Liu. Search-as-a-service: Outsourced search over outsourced storage. ACM Trans. Web, 3:13:1–13:33, September 2009.*

*[72] N. Smart and F. Vercauteren. Fully homomorphic encryption with relatively small key and ciphertext sizes. In Phong Nguyen and David Pointcheval, editors, Public Key Cryptography PKC 2010, volume 6056 of Lecture Notes in Computer Science, pages 420–443. Springer Berlin / Heidelberg, 2010.*

*[73] Dawn Xiaoding Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In Security and Privacy, 2000. S P 2000. Proceedings. 2000 IEEE Symposium on, pages 44 –55, 2000.*

*[74] R. Sravan Kumar and A. Saxena. Data integrity proofs in cloud storage. In Communication Systems and Networks (COMSNETS), 2011 Third International Conference on, pages 1 –4, jan. 2011.*

*[75] Damien Stehle and Ron Steinfeld. Faster fully homomorphic encryption. In Masayuki Abe, editor, Advances in Cryptology – ASIACRYPT 2010, volume 6477 of Lecture Notes in Computer Science, pages 377–394. Springer Berlin / Heidelberg, 2010.*

*[76] Mark W. Storer, Kevin Greenan, Darrell D.E. Long, and Ethan L. Miller. Secure data deduplication. In Proceedings of the 4th ACM international workshop on Storage security and survivability, StorageSS ’08, pages 1–10, New York, NY, USA, 2008. ACM.*

*[77] Marten Van Dijk, Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan. Fully homomorphic encryption over the integers. In Henri Gilbert, editor, Advances in Cryptology EUROCRYPT 2010, volume 6110 of Lecture Notes in Computer Science, pages 24–43. Springer Berlin / Heidelberg, 2010.*

*[78] Marten Van Dijk and Ari Juels. On the impossibility of cryptography alone for privacy-preserving cloud computing. In Proceedings of the 5th USENIX conference on Hot topics in security, HotSec’10, pages 1–8, Berkeley, CA, USA, 2010. USENIX Association.*

*79] Cong Wang, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou. Secure ranked keyword search over encrypted cloud data. In Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on, pages 253 –262, june 2010.*

*[80] Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Transactions on Parallel and Distributed Systems, 23(8):1467–1479, 2012.*

*[81] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, and Wenjing Lou. Privacy-preserving public auditing for secure cloud storage. IEEE Transactions on Computers, 62(2):362–375, 2013.*

*[82] Cong Wang, Kui Ren, Wenjing Lou, and Jin Li. Toward publicly auditable secure cloud data storage services. Network, IEEE, 24(4):19–24, july-august 2010.*

*[83] Cong Wang, Qian Wang, Kui Ren, Ning Cao, and Wenjing Lou. Toward secure and dependable storage services in cloud computing. IEEE Trans. Serv. Comput., 5(2):220–232, January 2012.*

*[84] Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou. Ensuring data storage security in cloud computing. In Quality of Service, 2009. IWQoS. 17th International Workshop on, pages 1 –9, july 2009.*

*[85] Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou. Privacypreserving public auditing for data storage security in cloud computing. In Proceedings of the 29th conference on Information communications, INFOCOM’10, pages 525–533, Piscataway, NJ, USA, 2010. IEEE Press.*

*[86] Qian Wang, Cong Wang, Jin Li, Kui Ren, and Wenjing Lou. Enabling public verifiability and data dynamics for storage security in cloud computing. In Proceedings of the 14th European conference on Research in computer security, ESORICS’09, pages 355–370, Berlin, Heidelberg, 2009. Springer-Verlag.*

*[87] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, and Jin Li. Enabling public auditability and data dynamics for storage security in cloud computing. Parallel and Distributed Systems, IEEE Transactions on, 22(5):847 –859, may 2011.*

*[88] Brent Waters. Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In Public Key Cryptography PKC 2010, volume 6571 of Lecture Notes in Computer Science, pages 53–70. Springer Berlin / Heidelberg, 2011.*

*[89] Chi won Song, Sungmin Park, Dong wook Kim, and Sooyong Kang. Parity cloud service: A privacy-protected personal data recovery service. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on, pages 812 –817, nov. 2011.*

*[90] Wai Kit Wong, David Wai-lok Cheung, Ben Kao, and Nikos Mamoulis. Secure knn computation on encrypted databases. In Proceedings of the 35th SIGMOD international conference on Management of data, SIGMOD ’09, pages 139–152, New York, NY, USA, 2009. ACM.*

*[91] CCITT Recommendation X.800. Security architecture for open systems interconnection for ccitt applications. Technical report, March 1991.*

*[92] Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Achieving secure, scalable, and fine-grained data access control in cloud computing. In INFOCOM, 2010 Proceedings IEEE, pages 1 –9, march 2010.*

*[93] Aaram Yun, Chunhui Shi, and Yongdae Kim. On protecting integrity and confidentiality of cryptographic file system for outsourced storage. In Proceedings of the 2009 ACM workshop on Cloud computing security, CCSW ’09, pages 67–76, New York, NY, USA, 2009. ACM.*

*[94] Wenying Zeng, Yuelong Zhao, Kairi Ou, and Wei Song. Research on cloud storage architecture and key technologies. In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, ICIS ’09, pages 1044–1048, New York, NY, USA, 2009. ACM.*

*[95] Fengli Zhang, Qinyi Li, and Hu Xiong. Efficient revocable key-policy attribute based encryption with full security. In Eighth International Conference on Computational Intelligence and Security 2012, pages 477–481. IEEE, 2012.*

*[96] Min Zhang, Ke Cai, and Dengguo Feng. Fine-grained cloud db damage examination based on bloom filters. In Proceedings of the 11th international conference on Web-age information management, WAIM’10, pages 157–168, Berlin, Heidelberg, 2010. Springer-Verlag.*

*[97] Fangming Zhao, Takashi Nishide, and Kouichi Sakurai. Realizing finegrained and flexible access control to outsourced data with attributebased cryptosystems. In Proceedings of the 7th international conference on Information security practice and experience, ISPEC’11, pages*

*83–97, Berlin, Heidelberg, 2011. Springer-Verlag.*

*[98] Qingji Zheng and Shouhuai Xu. Fair and dynamic proofs of retrievability. In Proceedings of the first ACM conference on Data and application security and privacy, CODASPY ’11, pages 237–248, New York, NY, USA, 2011. ACM.*

*[99] Lan Zhou, Vijay Varadharajan, and Michael Hitchens. Enforcing rolebased access control for secure data storage in the cloud. Comput. J., 54(10):1675–1687, October 2011.*

*[100] Yan Zhu, Gail-Joon Ahn, Hongxin Hu, and Huaixi Wang. Cryptographic role-based security mechanisms based on role-key hierarchy. In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, 2010, pages 314–319. ACM, 2010.*

*[101] Yan Zhu, Hongxin Hu, Gail-Joon Ahn, Huaixi Wang, and ShanBiao Wang. Provably secure role-based encryption with revocation mechanism. J. Comput. Sci. Technol., 26(4):697–710, 2011.*

*[102] Yan Zhu, Huaixi Wang, Zexing Hu, Gail-Joon Ahn, Hongxin Hu, and Stephen S. Yau. Dynamic audit services for integrity verification of outsourced storages in clouds. In Proceedings of the 2011 ACM Symposium on Applied Computing, SAC ’11, pages 1550–1557, New York, NY, USA, 2011. ACM*

.