ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG

NGUYỄN HỒNG SƠN
NGUYỄN HẢI LONG

KHÓA LUẬN TỐT NGHIỆP

MÃ HÓA VÀ KIỂM SOÁT QUYỀN TRUY CẬP TRÊN HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU SQL

ENCRYPT AND ACCESS CONTROL ON SQL DATABASE MANAGEMENT SYSTEM

KỸ SƯ NGÀNH AN TOÀN THÔNG TIN

TP. HỒ CHÍ MINH, NĂM 2021

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG

NGUYỄN HỒNG SƠN - 17520988
NGUYỄN HẢI LONG - 17520712

KHÓA LUẬN TỐT NGHIỆP

MÃ HÓA VÀ KIỂM SOÁT QUYỀN TRUY CẬP TRÊN HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU SQL

ENCRYPT AND ACCESS CONTROL ON SQL DATABASE MANAGEMENT SYSTEM

KỸ SƯ NGÀNH AN TOÀN THÔNG TIN

GIẢNG VIÊN HƯỚNG DẪN
TS. NGUYỄN NGỌC TỰ

TP. HỒ CHÍ MINH, NĂM 2021

Thông tin hội đồng chấm khóa luận tốt nghiệp

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số 463/QĐĐHCNTT ngày 23 tháng 7 năm 2021 của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

1. TS. Nguyễn Tuấn Nam	Chủ tịch
2. ThS. Nguyễn Duy	Ủy viên
3. ThS. Trần Hồng Nghi	Thư ký

Lời cảm ơn

Không ai đạt được điều gì đó to lớn mà không nhờ sự giúp đỡ của những người xung quanh, cho dù là trực tiếp hay gián tiếp đi nữa. Để hoàn thành được khóa luận này, nhóm tác giả may mắn nhận được nhiều sự giúp đỡ và hỗ trợ từ quý thầy, cô, anh chị, bạn bè và người thân. Nhóm tác giả xin dành những trang đầu tiên này để bày tỏ lòng tri ân của mình tới tất cả mọi người, những người đã đồng hành cùng nhóm trong khoảng thời gian vừa qua.

Đầu tiên, nhóm tác giả xin gửi lời cảm ơn sâu sắc đến toàn thể các thầy cô của Trường Đại học Công nghệ Thông tin nói chung và các thầy cô khoa Mạng máy tính và Truyền thông nói riêng. Nhờ những kiến thức quý giá mà thầy cô đã truyền đạt, cũng như việc hỗ trợ tận tình trong suốt khoảng thời gian thực hiện, nhóm đã hoàn thành khóa luận và đạt được các kết quả đáng ghi nhận. Nhóm tác giả xin đặc biệt cảm ơn TS. Nguyễn Ngọc Tự là người đã truyền cảm hứng, tận tình hướng dẫn và hỗ trợ tận tình về kiến thức, tạo môi trường thuận lợi để nhóm có thể học hỏi, trao đổi với các bạn, các em trong nhóm nghiên cứu. Đây là những kiến thức, kinh nghiệm quý giá, không chỉ có tác dụng trong khóa luận tốt nghiệp này mà còn trong khoảng thời gian làm việc trong chặng đường tiếp theo.

Trong giai đoạn dịch bệnh khó khăn, tình hình ngày càng phức tạp, dù có khó khăn trong nhiều công việc, nhóm nhận được nhiều sự giúp đỡ và động viên từ thầy cô và bạn bè. Đây là động lực to lớn thúc đẩy nhóm làm việc trong suốt quá trình tìm hiểu và hoàn thành khóa luận này.

Cuối cùng, nhóm tác giả không quên bày tỏ lòng tri ân đến gia đình và người thân, những người đã luôn là những hậu phương vững chắc và luôn ủng hộ từng quyết định mà nhóm đưa ra.

Mặc dù đã nỗ lực rất nhiều để luận văn được hoàn thiện nhất, song khó có thể tránh khỏi thiếu sót và hạn chế. Kính mong nhận được sự thông cảm và ý kiến đóng góp từ quý thầy cô và các bạn.

TP. Hồ Chí Minh, ngày 12 tháng 7 năm 2021

Nhóm tác giả

Mục lục

Danh sách hình vẽ

Danh sách bảng

Danh sách từ viết tắt

ABAC	Attribute-based access control
BLP	Bell – La Padula
CSP	Cloud Service Provider
DAC	Discretionary access control
FsP	Forward search privacy
MAC	Mandatory access control
ORAM	oblivious RAM
PIR	private information retrieval
RBAC	Role-Based access control
SE	Searchable Encryption
SSE	Searchable Symmetric Encryption

Tóm tắt khóa luận

Vào năm 2008, Lizhe Wang, một trong những ngọn cờ tiên phong đã đưa ra khái niệm cơ bản về "Cloud Computing". Theo đó, điện toán đám mây là một tập hợp các dịch vụ hỗ trợ mạng, cung cấp khả năng mở rộng, chất lượng dịch vụ được đảm bảo, thường được cá nhân hóa, chi phí thấp theo yêu cầu và có thể truy cập một cách đơn giản và phổ biến [**wang2008scientific**].

Từ thời điểm đó, điện toán đám mây đã dần trở thành một trong những từ khóa tiếp theo của ngành công nghệ thông tin. Vào năm 2020, tổng giá trị của thị trường là 371,4 tỷ USD, tốc độ tăng trưởng kép hàng năm (CAGR) là 17,5%. Theo dự kiến, tới năm 2025, thị trường điện toán đám mây sẽ có giá trị tới 832,1 tỷ USD, sẽ có hơn 100 zettabytes dữ liệu được lưu trữ trên đám mây. Trong cùng khoảng thời gian đó, tổng dung lượng dữ liệu được lưu trữ sẽ vượt quá 200 zettabytes, nghĩa là có khoảng 50% dữ liệu được lưu trên đám mây, con số này là 25% vào năm 2015 [**vladimir2021cloudcomputing**].

Ở Việt Nam hiện nay, đón đầu xu thế trong cuộc cách mạng công nghệ 4.0, Chính phủ thúc đẩy mạnh mẽ và tạo điều kiện trên mọi phương diện nhằm thúc đẩy khởi nghiệp, nhóm ngành công nghệ thông tin là một trong những lĩnh vực được ưu tiên thúc đẩy phát triển. Nhiều công ty công nghệ thông tin được thành lập và có tiềm năng phát triển cao. Đi đôi với việc thành lập doanh nghiệp, các công ty xây dựng cơ sở dữ liệu của mình nhằm phục vụ các hoạt động nội bộ của công ty cũng như kinh doanh với đối tác, khách hàng. Xu hướng sử dụng cloud service để lưu và quản trị cơ sở dữ liệu được nhiều đơn vị lựa chọn.

Tuy có những ưu điểm vượt trội như trên, hệ thống này cũng đặt ra nhiều vấn đề về an toàn thông tin cho dữ liệu được lưu trữ trên đám mây. Môi trường đám mây được xem là không tin cậy cho những dữ liệu riêng tư có tính nhạy cảm: dữ liệu có thể bị truy cập trái phép từ nhà quản cung cấp dịch vụ, bị rò rỉ sang bên thứ ba không liên quan hoặc bị đánh cắp bởi quản trị viên, hacker. Các công ty cho thuê dịch vụ đám mây phục vụ nhiều khách hàng có thể dẫn tới việc không duy trì được sự tách biệt giữa những người cùng thuê dịch vụ. Dữ liệu của người dùng sau khi hết sử dụng có thể không được xóa an toàn do bị giảm khả năng hiển thị vào nơi dữ liệu được lưu trữ vật lý trên đám mây hoặc bị mất mát do nhà cung cấp dịch vụ vô tình xóa hoặc

thảm họa. Người dùng nội bộ của công ty cho thuê có thể lạm quyền và truy cập nhằm đánh cắp thông tin [**timothy2018risk**]. Do vậy, ngoài những cơ chế bảo mật sẵn có từ nhà cung cấp dịch vụ, chủ sở hữu dữ liệu cần có những cơ chế bảo mật riêng để đảm bảo thông tin lưu trữ, tương tác và trao đổi được an toàn. Một trong những giải pháp được sử dụng là mã hóa dữ liệu trước khi lưu trữ tại đám mây và thay đổi phương pháp tương tác truyền thống trên dữ liệu rõ bằng các phương pháp tương tác trên dữ liệu mã hóa.

Để thử nghiệm và tìm ra phương pháp mã hóa, kiểm soát truy cập với cơ sở dữ liệu được lưu trữ trên đám mây, nhóm tác giả đã tiến hành nghiên cứu và xây dựng mô hình nhằm mục đích đảm bảo sự an toàn của cơ sở dữ liệu trước những rủi ro của hệ thống điện toán đám mây. Từ đó, rút ra nhận xét và hướng phát triển trong tương lai.

Phần còn lại của khóa luận được chia thành 5 chương. Chương 1 giới thiệu những nét khái quát chung và những vấn đề cơ bản đặt ra khiến nghiên cứu này là cần thiết. Chương 2 và chương 3 trình bày một số nghiên cứu hiện tại về kiểm soát truy cập và mã hóa dữ liệu, cũng như đưa ra một số chi tiết về phương pháp của nhóm tác giả cho vấn đề được nêu ra ở chương 1. Chương 4 trình bày mô tả ứng dụng hiện thực về kiểm soát truy cập dựa trên mã hóa CP-ABE. Chương 5 tổng kết lại một số kết quả đạt được và hướng phát triển của nghiên cứu.

Chương 1

Giới thiệu

1.1 Khái quát chung

Trong kỷ nguyên mà sự phát triển của các hệ thống điện toán đám mây càng phát triển với độ nở lớn như hiện nay, lượng dữ liệu đổ lên các dịch vụ đám mây đang tăng theo cấp số nhân. Đại dịch Covid-19 vẫn đang lan rộng, thành tựu về vắc – xin tuy đã kiểm soát khá tốt tốc độ lây lan của dịch bệnh, nhưng vẫn chưa thể xóa sổ hoàn toàn nó ra khỏi đời sống kinh tế, xã hội. Dịch bệnh thúc đẩy phát triển những cách làm việc phi truyền thống, chuyển dần sang nền tảng trực tuyến, nơi mà dòng chảy dữ liệu vốn đã chật chội nay càng thêm khó kiểm soát. Vào năm 2020, tổng chi tiêu của người dùng cuối cho các dịch vụ đám mây đạt tổng cộng 270 tỷ USD, con số này dự kiến sẽ tăng với mức đáng kinh ngạc là 23,1% vào năm 2021, lên 332,3 tỷ USD. Trong khi đó, 48% doanh nghiệp chọn lưu trữ dữ liệu quan trọng của họ, bao gồm dữ liệu mã hóa và dữ liệu thông thường trên đám mây. Vì vậy, không có gì ngạc nhiên khi 75% doanh nghiệp coi các vấn đề bảo mật đám mây là mối quan tâm hàng đầu. Trong số đó, 33% người được hỏi cực kỳ quan tâm, 42% cực kỳ lo ngại và 25% không quan tâm hoặc quan tâm vừa phải [**vladimir2021cloudcomputing**].

Vì nhiều công nghệ điện toán đám mây đang ngày càng được sử dụng nhiều hơn, và do đó, giống như hầu hết các công nghệ mới, vấn đề bảo mật cho nó đã, đang và tiếp tục được đặt ra và ngày càng trở nên quan trọng. Dữ liệu là nguồn sống của doanh nghiệp, và khi nó càng lớn thì động lực bảo vệ tài sản quý giá này thúc đẩy họ tìm kiếm các giải pháp bảo mật an toàn hơn. Vấn đề chính đặt ra là tìm một nhà cung cấp dịch vụ đám mây (CSP) có uy tín và ổn định để các doanh nghiệp có thể ít bị tấn công hơn hay bản thân các công ty phải chủ động phát triển một công cụ như là chìa khóa két sắt để bảo vệ tài sản của mình khi giao cho người khác nắm giữ.

Cả hai hướng tiếp cận này đều có những ưu và nhược điểm. Sử dụng dịch vụ của một đối tác CSP uy tín giúp chủ sở hữu dữ liệu tiết kiệm được nhiều kinh phí, mức độ tin cậy cũng được đảm bảo ở mức tương đối. Tuy nhiên, hướng tiếp cận này chỉ giải quyết được một số vấn đề xảy

ra từ những đối tượng xấu có ý đồ tấn công mà không giải quyết được các vấn đề có thể xảy ra từ bên cung cấp dịch vụ. Ngược lại, để phát triển một giải pháp bảo mật hiệu quả, nguồn lực kinh phí và con người phải bỏ ra là rất lớn. Không phải công ty, doanh nghiệp nào cũng có thể đầu tư nguồn lực lớn vào vấn đề này.

Nhằm khắc phục điểm yếu và phát huy điểm mạnh của hai hướng trên, việc một bên thứ ba có đầy đủ năng lực chuyên môn và nguồn lực tài chính đứng ra phát triển dịch vụ bảo mật trên nền tảng đám mây giải quyết được cả hai vấn đề nếu trên: dịch vụ cung cấp cho mỗi doanh nghiệp sử dụng nên giá dịch vụ sẽ rẻ hơn so với tự phát triển; bên thứ ba được đảm bảo và chứng nhận độc lập với CSP hoàn toàn đã tránh được các rủi ro chủ quan phát sinh. Nói tóm lại, sự xuất hiện và phát triển của một bên thứ ba đảm nhiệm vai trò bảo mật cơ sở dữ liệu trên dịch vụ đám mây xuất phát từ nhu cầu thực tế của các bên liên quan, hứa hẹn sẽ mang đến những tiện ích và trải nghiệm tuyệt vời cho người dùng, nhất là các doanh nghiệp và là hướng đi tiềm năng phát triển trong tương lai.

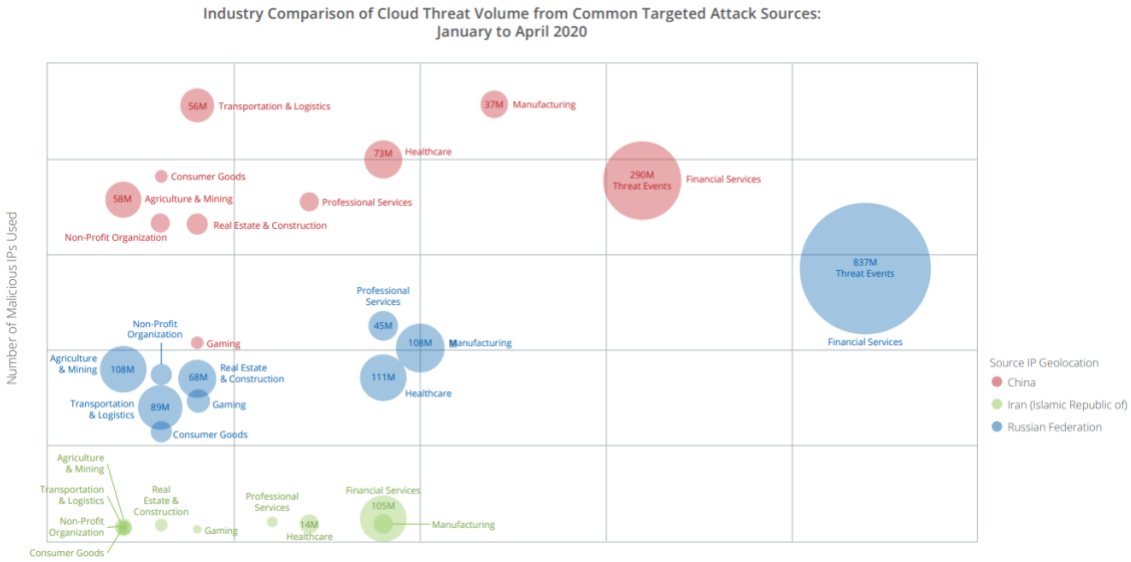
1.2 Báo cáo vấn đề

Trước sự phát triển mạnh mẽ của các dịch vụ lưu trữ đám mây, nguy cơ tấn công vào các cơ sở dữ liệu này đang tiềm ẩn trên hầu hết khu vực và đa dạng các đối tượng nạn nhân, kể cả các công ty lớn chuyên cung cấp dịch vụ lưu trữ đám mây như Google Cloud hay AWS.

McAfee đã tiến hành một nghiên cứu về các cuộc tấn công mạc vào các dịch vụ đám mây để xác định xem liệu có sự gia tăng các cuộc tấn công kể từ khi đại dịch Covid19 bắt đầu hay không. Kết quả cho thấy sự gia tăng lên đến 630% các cuộc tấn công mạng vào các dịch vụ đám mây kể từ tháng 1 đến tháng 4 năm 2020. Với sự mở rộng của các dịch vụ chăm sóc sức khỏe qua điện thoại, nhiều nhà cung cấp dịch vụ đã chuyển sang sử dụng dịch vụ đám mây. Trong quý đầu tiên năm 2020, đây là ngành bị nhắm mục tiêu đa số trong các cuộc tấn công với khoảng 198 triệu IP độc hại được phát hiện [**mcafee2020risk**].

1.3 Cơ sở lý thuyết và phương pháp luận

Trong khóa luận này, nhóm tác giả sử dụng kỹ thuật kiểm soát truy cập dựa trên thuộc tính, thông qua các phiên bản ứng dụng của mã hóa dựa trên thuộc tính. Đây là kỹ thuật được phát triển thêm từ kiểm soát truy cập dựa trên vai trò. Kiểm soát truy cập dựa trên thuộc tính không phân chia vai trò cho từng người dùng mà cấp quyền truy cập cho họ thông qua một tập thuộc tính biểu thị chính sách truy cập. Người dùng có những thuộc tính thỏa mãn chính sách được quy định có thể đọc được tài liệu mã hóa.



Hình 1.1: Số tổ chức bị tấn công [**mcafee2020risk**]

Bên cạnh đó tìm hiểu, nghiên cứu một số phương pháp thao tác trên cơ sở dữ liệu mã hóa để đảm bảo quyền riêng tư dữ liệu của người dùng trước các máy chủ đám mây bên thứ ba không đáng tin cậy.

1.4 Đối tượng và phạm vi thực hiện

1.4.1 Đối tượng nghiên cứu

- Kỹ thuật kiểm soát truy cập dựa trên thuộc tính, kỹ thuật mã hóa dựa trên thuộc tính, phương pháp mã hóa chính sách khóa dựa trên thuộc tính, phương pháp mã hóa chính sách bản mã dựa trên thuộc tính.
- Mã hóa đồng cấu, lược đồ searchable encryption.
- Hệ quản trị cơ sở dữ liệu quan hệ có cấu trúc SQL.

1.4.2 Phạm vi thực hiện

Phân tích và đánh giá mô hình kiểm soát truy cập dựa trên thuộc tính trên dữ liệu mã hóa; hỗ trợ các thao tác tìm kiếm, cập nhật trên dữ liệu mã hóa; kết hợp triển khai trên hệ quản trị cơ sở dữ liệu MySQL.

1.4.3 Mục tiêu nghiên cứu

Tìm hiểu về các kỹ thuật, phương pháp, thuật toán nhằm kiểm soát truy cập dựa trên thuộc tính; hỗ trợ các thao tác trên dữ liệu mã hóa; từ đó xây dựng một ứng dụng nhằm đánh giá mức độ hiệu quả của phương pháp đối với khả năng bảo mật dữ liệu trên dịch vụ đám mây.

Chương 2

Điều khiển truy cập trên dữ liệu mã hóa

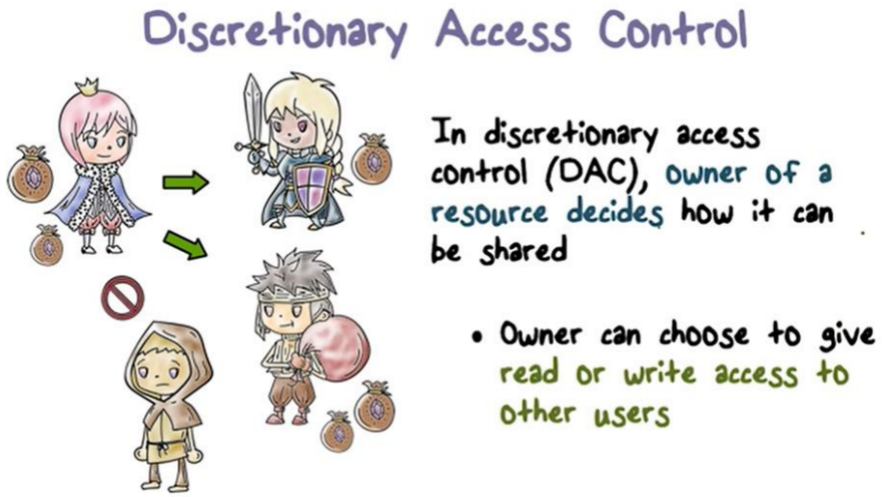
2.1 Tổng quan về kiểm soát truy cập trên cơ sở dữ liệu

Kiểm soát truy cập là vấn đề quan trọng trong bất kỳ cơ quan, tổ chức hay hệ thống nào. Từ thời cổ đại, con người đã phát minh ra các phương pháp kiểm soát truy cập khác nhau: sử dụng lệnh bài, giấy tờ hợp lệ... để ra vào các khu vực giới hạn. Trong lĩnh vực công nghệ thông tin hiện đại, kiểm soát truy cập là khả năng của hệ thống để xác định xem người dùng có thể truy cập vào một phần dữ liệu cụ thể được lưu giữ trong hệ thống máy tính và môi trường hoạt động liên quan của nó hay không [**emms1987definition**]. Nó bao gồm hai thành phần chính là xác thực *(authentication)* và ủy quyền *(authorization)*. Xác thực là phương pháp xác minh danh tính của người đang truy cập cơ sở dữ liệu. Ủy quyền là phương pháp xác định người dùng có được phép truy cập vào dữ liệu hoặc thực hiện các thao tác trên dữ liệu đó hay không. Nếu thiếu một trong hai yếu tố này, dữ liệu không được bảo vệ.

Các chính sách kiểm soát truy cập có thể chia thành ba nhóm chính chính: Điều khiển truy cập tùy ý (DAC), Điều khiển truy cập bắt buộc (MAC) và Điều khiển truy cập theo vai trò (RBAC) [**sandhu1996role**].

2.1.1 Điều khiển truy cập tùy ý (DAC).

Ở mô hình điều khiển truy cập DAC, chủ sở hữu kiểm soát quyền truy cập nhưng chỉ với bản gốc, không phải với bản sao. Mô hình gồm ba thành phần: chủ thể (subject), quyền truy cập (access) và đối tượng (object). Trong DAC, chủ sở hữu dữ liệu quyết định quyền truy cập của các chủ thể khác trên đối tượng do mình sở hữu, xem2.1.



Hình 2.1: Minh họa mô hình DAC

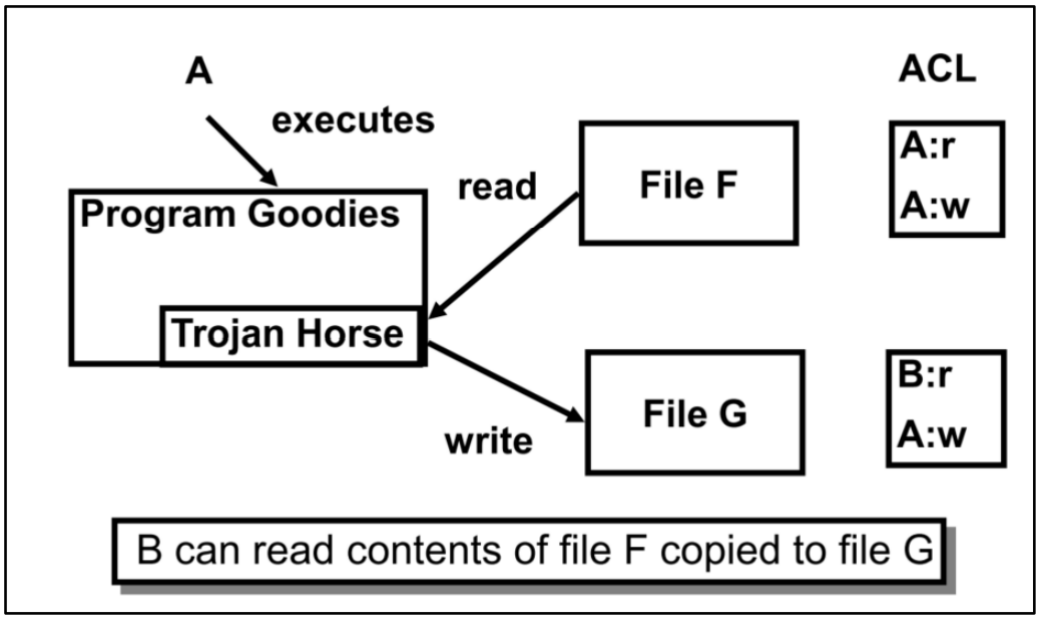
Tuy nhiên, mô hình này có nhược điểm lớn là việc trao quyền kiểm soát truy cập của đối tượng chủ sở hữu của nó có thể làm tiết lộ dữ liệu cho những người không có quyền truy cập hợp pháp với nó. Chúng ta không thể đảm bảo rằng chủ sở hữu dữ liệu là tuyệt đối tin tưởng được. Trong trường hợp chủ sở hữu có thể tin tưởng, khả năng chủ sở hữu quyết định sai các quyền truy cập trên các tài nguyên của mình có thể gây nên đổ vỡ hệ thống. Các chủ thể xấu tồn tại trong hệ thống có thể lợi dụng sự thiếu hiểu biết hoặc sai lầm của chủ sở hữu dữ liệu, kết hợp với nhau để chiếm đoạt tài nguyên bất hợp pháp.

Giả sử hệ thống chúng ta có các thành phần như mô tả trong bảng sau:

Bảng 2.1: Ví dụ về kiểm soát truy cập DAC

Chủ thể	**Quyền truy cập**	**Đối tượng**
A	R	File F
A	W	File F
A	W	File G
B	R	File G

Theo như bảng trên, B không thể đọc được các nội dung trong file F do không có quyền R trên file F. Tuy nhiên, tận dụng lỗ hổng của mô hình DAC, chủ thể A có thể đọc các nội dung trên file F, thực hiện ghi nó vào file G. Từ đó, B có thể đọc được nội dung của file F mà không cần có quyền gì trên file F (xem 2.2). Vì những rủi ro của nó, trong các hệ thống hiện đại ngày nay, các công ty, tổ chức hạn chế và hầu như không sử dụng mô hình kiểm soát truy cập DAC.

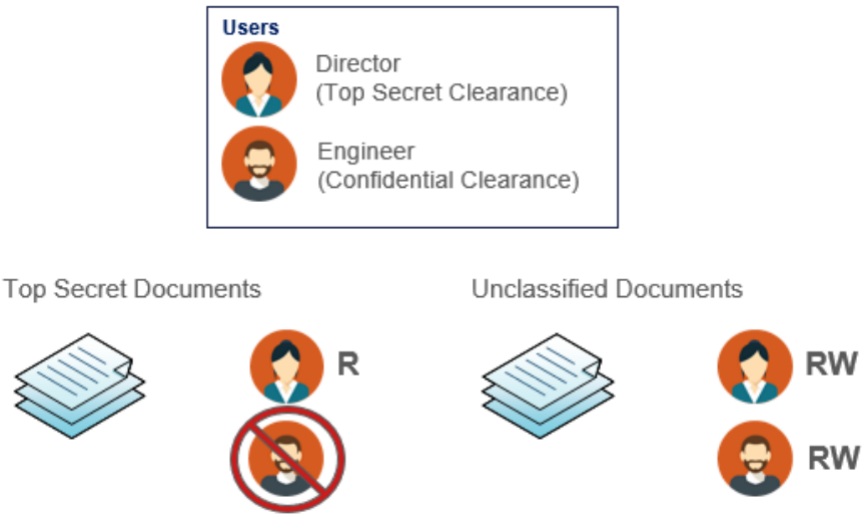


Hình 2.2: Lỗ hổng trojan horse trong mô hình DAC [**sandhu1996role**]

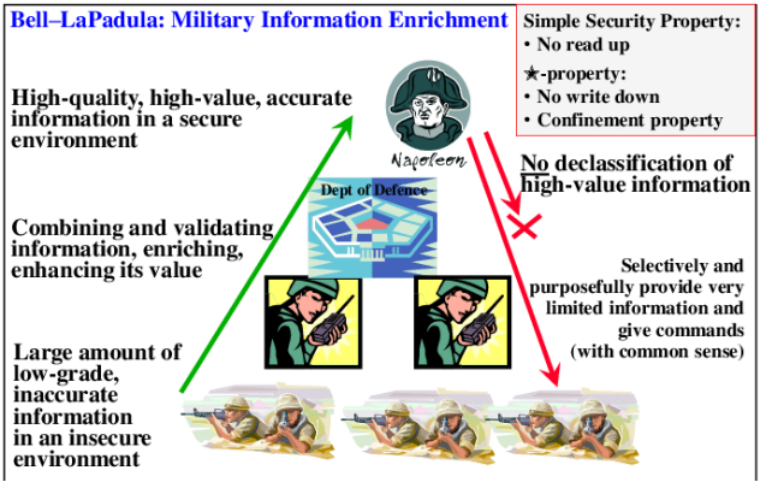
2.1.2 Điều khiển truy cập bắt buộc (MAC).

Trong mô hình MAC, quyền truy cập đối với các đối tượng không được quyết định bởi chủ thể mà được quyết định bởi một người (hoặc hệ thống) quản trị. Các chủ thể và đối tượng được gắn các nhãn tương ứng với mức độ "nhạy cảm" của họ và được sắp xếp một cách có trật tự (xem). MAC áp dụng hai quy tắc chính đó là no read-up và no write-down. Mối quan hệ giữa các luồng thông tin trong hệ thống này là phản xạ, bắc cầu và phản đối xứng. Điểm mạnh của MAC là đảm bảo bí mật của thông tin và cung cấp biện pháp chống lại phương pháp tấn công trojan horse. Điều này đạt được do các nhãn của đối tượng được truyền qua các bản sao và ngăn chặn việc hạ cấp nhãn của đối tượng, từ đó đảm bảo dữ liệu không thể chuyển từ mức độ "nhạy cảm" cao xuống thấp.

Khi nhắc tới MAC, mô hình Bell-La Padula thường được nhắc tới thường xuyên nhất. Nhóm tác giả phân tích rất khái quát về mô hình Bell-La Padula nhằm đại diện cho phương pháp MAC. Trong mô hình Bell-La Padula, có một tập hợp nhãn phân loại (top-secret, secret, confidential, unclassified) được xác định hoàn toàn. Bên cạnh đó, có một danh sách các danh mục không được phân loại. Sự kết hợp của một nhãn và một thành phần con của danh mục được gọi là mức bảo mật. Mức bảo mật được sắp xếp một phần và tạo thành một mạng lưới . Trong mạng lưới này, các chủ thể không được đọc các tài liệu có nhãn mức bảo mật cao hơn mình, đồng thời không được viết ra các tài liệu có nhãn bảo mật thấp hơn mình. Nghĩa là một chủ thể có mức bảo mật secret thì không thể đọc được tài liệu có nhãn top-secret nhưng có thể đọc được các tài



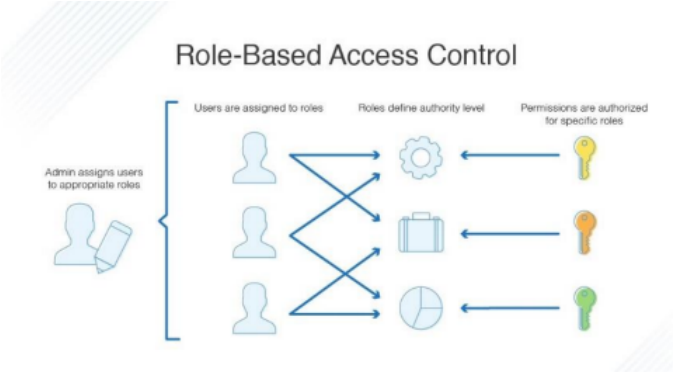
Hình 2.3: Minh họa mô hình MAC



Hình 2.4: Ví dụ về mô hình Bell-La Padula

liệu có nhãn bằng hoặc thấp hơn mình (secret, confidential, unclassified). Bên cạnh đó, chủ thể này cũng không thể tạo ra các tài liệu có nhãn bảo mật thấp hơn nhãn của mình. Phương pháp kiểm soát truy cập này rất phù hợp trong môi trường cần đề cao tính bí mật của thông tin như quân sự, ngoại giao (xem ví dụ hình 2.4).

Tuy nhiên, do sự đa dạng các yêu cầu, đặc biệt là mong môi trường điện toán đám mây với lượng truy cập lớn, các quy tắc và định lý bảo mật hiện tại trong BLP không thể thích ứng được. Các hạn chế của nó dần dần bộc lộ trong quá trình phát triển lý thuyết về bảo mật. Giả sử, trong hệ thống ban đầu chỉ có một một loại phân cấp, khi một chủ thể yêu cầu truy cập vào đối tượng. Trong trường hợp này, do tất cả chủ thể và đối tượng đều được phân cấp là thấp nhất, yêu cầu này đương nhiên được chấp thuận theo BLP, nhưng đó rõ ràng là rủi ro lớn.



Hình 2.5: Minh họa về RBAC

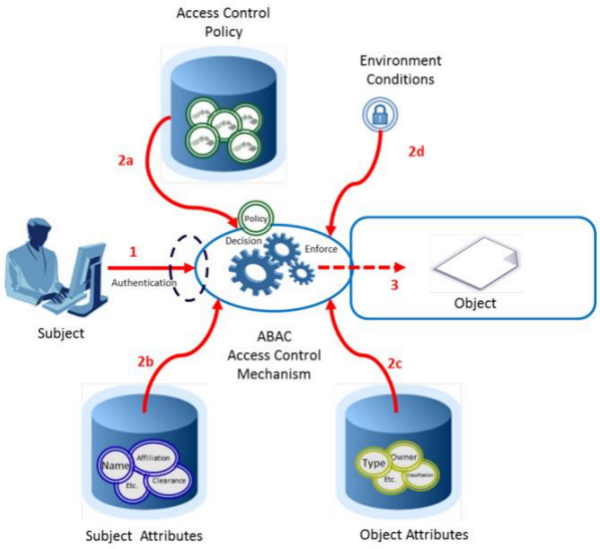
Nhiều giải pháp cải thiện mô hình BLP được đưa ra để phù hợp với sự phát triển của lý thuyết bảo mật: đề xuất một phương pháp có thể xác định các chủ thể một cách linh động bằng cách cung cấp các phương pháp chuyển đổi trạng thái trong BLP; định nghĩa các chủ thể, đối tượng và quy tắc bảo mật trong network domain, cung cấp các quy tắc để hạn chế hành vi của hệ thống.

Mặc dù các nghiên cứu đã có nhiều nỗ lực để cải thiện BLP và các ứng dụng của nó, nhưng mô hình chủ động điều chỉnh với những thay đổi ít khi được đề cập đến. Các nghiên cứu ít chú đến các hệ thống yêu cầu thời gian thực, thiếu khả năng thay đổi theo bối cảnh.

2.1.3 Điều khiển truy cập theo vai trò (RBAC)

Phương pháp điều khiển truy cập dựa theo vai trò thực hiện dựa trên chính sách quy định quyền truy cập của người dùng vào các tài nguyên trên cơ sở mà người dùng thực hiện. Nó yêu cầu xác định cụ thể từng vai trò trong hệ thống. Theo đó, vai trò là một tập hợp các quyền sử dụng các tài nguyên phù hợp với chức năng và công việc của một người, là một tập hợp các hành động và trách nhiệm liên quan đến một hoạt động cụ thể. Thay vì chỉ định tất cả các quyền mà người dùng được phép thực thi, các quyền truy cập được chỉ định cho các vai trò. Người dùng được thêm vào các tập hợp vai trò mà thông qua đó, được phép thực thi các quyền của vai trò đó.

Theo thời gian, các doanh nghiệp nhận thấy nhu cầu vượt ra ngoài các định nghĩa về nhóm người dùng và quyền của RBAC. Chúng cần bao gồm các thuộc tính linh động như thời gian trong ngày, vị trí của người dùng. Các hệ thống phân tán và thay đổi liên tục, kiểm soát truy cập dựa trên thuộc tính là một lựa chọn hoặc thay thế, hoặc bổ trợ cho RBAC. ABAC sử dụng các đối tượng được gắn nhãn và thuộc tính người dùng thay vì quyền để kiểm soát một cách linh hoạt.



Hình 2.6: Mô hình ABAC cơ bản

2.1.4 Điều khiển truy cập dựa theo thuộc tính (ABAC)

ABAC là một mô hình logic điều khiển truy cập vào các đối tượng bằng cách đánh giá các quy tắc dựa trên các thuộc tính của các đối tượng và chủ thể, các hoạt động và môi trường liên quan đến một ngữ cảnh yêu cầu cụ thể. Nó cho phép kiểm soát truy cập chính xác hơn bằng cách cho phép số lượng lớn đầu vào rời rạc tham gia vào quá trình quyết định quyền truy cập, và do đó cung cấp một lượng lớn hơn các kết hợp có thể có của các thuộc tính để phản anh một chính sách lớn hơn rất nhiều so với các phương pháp khác. Nó chỉ bị giới hạn ngôn bởi ngôn ngữ tính toán và sự phong phú của các thuộc tính có sẵn.

Mô hình điều khiển truy cập ABAC cơ bản có 3 bước: chủ thể yêu cầu một truy cập (đọc, ghi, thực thi...) tới một đối tượng cụ thể; hệ thống kiểm tra và đánh giá các thuộc tính mà chủ thể cung cấp; chủ thể được cho phép truy cập tài nguyên nếu là người dùng xác thực và bị cấm nếu không xác thực.

ABAC cho phép đưa ra các quyết định kiểm soát truy cập mà không cần chủ thể biết trước về đối tượng. Với phương pháp khái quát thuộc tính được xác định nhất quán giữa các tổ chức, nó tránh được nhu cầu xác minh quyền truy cập rõ ràng của các chủ thể với các đối tượng, nghĩa là chủ thể không biết rõ mình có quyền truy cập được đến đối tượng hay không khi chưa có nhu cầu truy cập tài nguyên đó, trong khi đối với RBAC, các chủ thể trong một tập vai trò biết rõ và tường minh các quyền của mình đối với đối tượng cụ thể. ABAC cho phép sự linh hoạt trong một doanh nghiệp lớn và rất lớn, nơi việc quản lý các danh sách vai trò và nhóm kiểm

soát truy cập tốn thời gian và phức tạp. Nếu các thuộc tính được xác định nhất quán, các thủ tục xác thực và cho phép truy cập có thể được thực thi và quản lý trong các cơ sở khác nhau, đồng thời duy trì mức độ bảo mật thích hợp trong tổ chức mới. Ví dụ một chủ thể có thể xác thực và truy cập tài nguyên trong bệnh viện này và sau đó được ủy quyền để truy cập các tài nguyên trên bệnh viện khác tương tự như đối với bệnh viện cũ, nhờ các giá trị thuộc tính được xác định nhất quán từ trước. Sau khi hoàn thành vai trò tại cơ sở tạm thời, việc loại bỏ quyền được cấp tạm thời được thực hiện dễ dàng mà không ảnh hưởng tới những người dùng khác trong hệ thống.

Kỹ thuật kiểm soát truy cập quản lý dựa trên từng giá trị thuộc tính do người sở hữu dữ liệu định nghĩa. Nó hỗ trợ quản lý các nhóm người dùng có cùng chính sách truy cập như kỹ thuật kiểm soát dựa trên vai trò, nhưng cá nhân hóa hơn nhờ áp dụng phương pháp dẫn xuất khóa nhóm. Người quản trị không cần phải thay đổi nhóm vai trò của từng người dùng mỗi khi họ thay đổi nhiệm vụ hoặc rời khỏi hệ thống mà việc phân bố lại nhóm trong hệ thống được thực hiện một cách tự động hoàn toàn. Khi mỗi người dùng có thay đổi về thuộc tính, gia nhập hoặc rời khỏi hệ thống, người quản trị chỉnh sửa tương ứng giá trị thuộc tính của họ, những người dùng khác trong hệ thống không bị ảnh hưởng. Thông thường, chính sách truy cập quy định một chủ thể s có thể truy cập vào tài nguyên r trong điều kiện bảo mật **e** được biểu diễn bằng biểu thức đại số bởi 3 biến **s, r, e** như sau:

$$can\_access(s,r,e) \leftarrow f(ATTR_s, ATTR_r, ATTR_e)$$ [**yuan2005attributed**]

với $ATTR_s$ là thuộc tính của chủ thể s.

Thông thường, một người dùng sẽ được chỉ định (hủy chỉ định) một cách tự động vào các nhóm nếu họ đáp ứng các điều kiện thành viên nhóm. Một điều quan trọng khác là cơ chế quản lý khóa nhóm, bởi mục tiêu các nhóm thường là chia sẻ dữ liệu. Do đó dữ liệu phải được mã hóa với các khóa chỉ được cung cấp cho các thành viên của nhóm. Việc quản lý những khóa, bao gồm các lựa chọn, phân phối, lưu trữ, cập nhật yêu cầu phương pháp quản lý khóa nhóm dựa trên thuộc tính. Theo đó các khóa nhóm được gán cho người dùng dựa trên các thuộc tính nhận dạng của họ.

Khi nhóm thay đổi, khóa nhóm mới phải được chia sẻ với các thành viên hiện có, để một thành viên nhóm mới không thể truy cập dữ liệu được truyền trước khi họ tham gia nhóm và một người dùng đã rời đi khỏi nhóm không thể truy cập dữ liệu của nhóm. Một vấn đề khác là bảo vệ chống lại các cuộc tấn công thông đồng mà qua đó, một nhóm người dùng gian lận thông đồng có thể có được khóa nhóm bằng cách tập hợp những khóa của nhau, sau đó lấy khóa nhóm

và đáng ra họ không được phép lấy riêng lẻ.

Mặc dù ABAC có thể hỗ trợ trong việc chia sẻ thông tin doanh nghiệp, nhưng khi triển khai trên quy mô doanh nghiệp, tập hợp các khả năng cần thiết để thực hiện trở nên phức tạp hơn. Ở cấp độ doanh nghiệp, quy mô lớn đòi hỏi khả năng quản lý phức tạp và đôi khi cần được thiết lập một cách độc lập. Hình 2.7 trình bày chi tiết các thành phần cần thiết khi triển khai ABAC trong doanh nghiệp. Hầu hết các doanh nghiệp hiện tại đều có thể đáp ứng được các thành phần trong sơ đồ như có hình thức quản lý và thông tin xác thực nhân viên. Tuy nhiên, các quy tắc thường được viết thành các loại văn bản mà con người có thể đọc được và thiết kế thành các phần mềm riêng lẻ. Để có thể triển khai ABAC trong doanh nghiệp được hiệu quả, các chính sách quản lý phải lưu dưới dạng máy có thể đọc được. Từ đó có thể triển khai quản lý các chủ thể và đối tượng được toàn diện.

Trong ABAC, chúng ta không cần định nghĩa các vai trò cụ thể và cấp quyền truy cập cho các vai trò đó như với RBAC. Trong ví dụ tiếp theo, một hệ thống quản lý xếp loại phim và độ tuổi người dùng được phép xem phim hàng tháng. Xếp loại phim và điều kiện xem phim quy định trong Bảng 2.2. Như vậy, chính sách quản lý truy cập trong hệ thống này quy định chỉ

Bảng 2.2: Ví dụ về loại hình ABAC

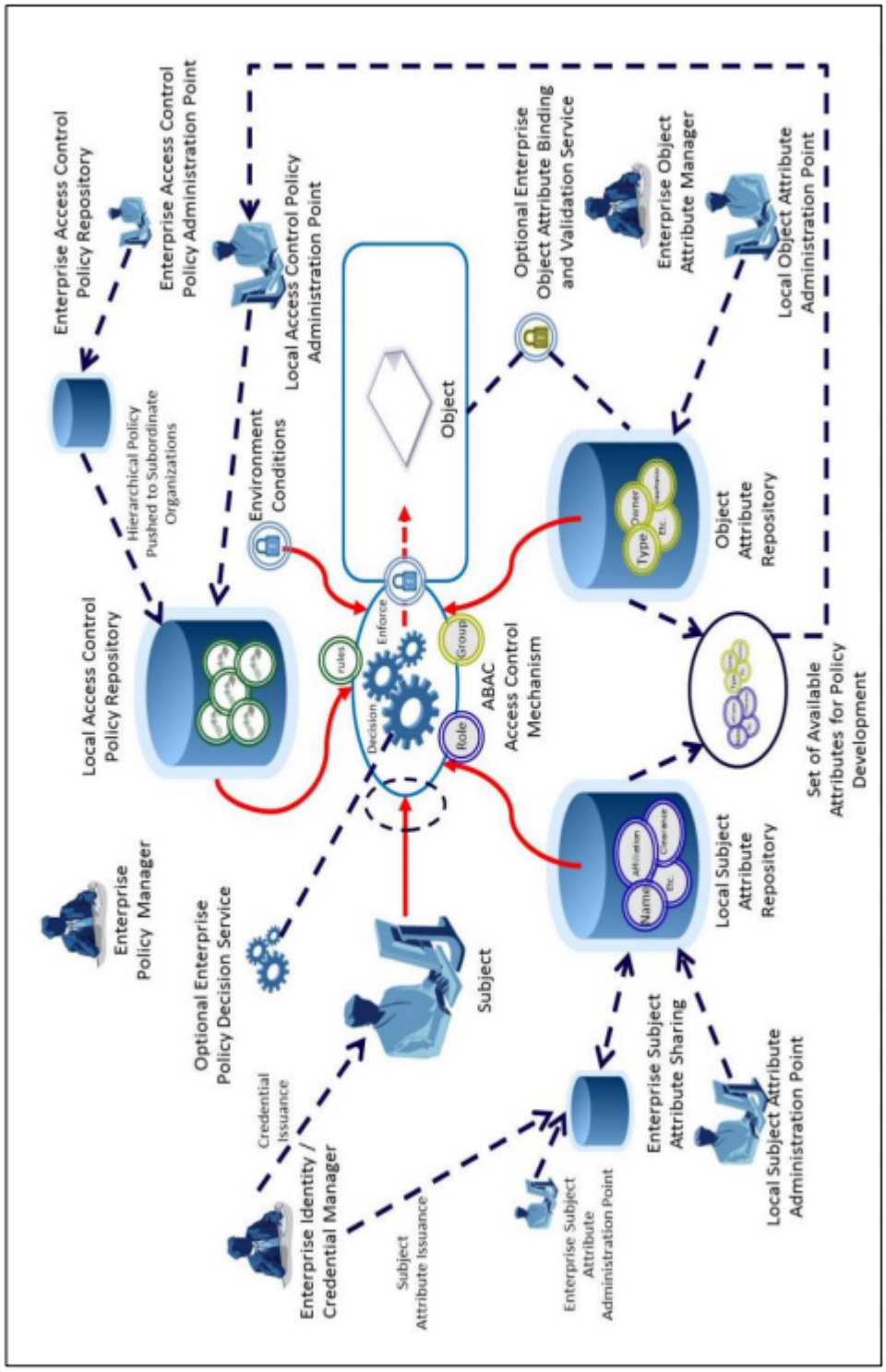
Loại phim	**Người dùng cho phép**
R	Từ 21 tuổi trở lên
PG-13	Từ 13 tuổi trở lên
G	Tất cả

người lớn mới được xem loại phim R, trẻ vị thành niên hoặc lớn hơn mới được được loại phim PG13 và phim G thì dành cho tất cả các đối tượng.

Đối với ABAC, hệ thống không quy định các vai trò người lớn, trẻ vị thành niên hay trẻ em em mà dựa vào các thuộc tính vốn có (ở ví dụ này là tuổi) để định nghĩa chính sách truy cập trong hệ thống. Một người sử dụng hệ thống u có thể truy cập vào và xem bộ phim m (trong điều kiện e mà ở đây không nhắc tới) phải thỏa mãn chính sách truy cập:

$$\textbf{R1: } can\_access(u,m,e) \leftarrow (age(u) \geq 21 \wedge rating(m) \in (R, PG-13, G)) \vee (age(u) \geq 13 \wedge rating(m) \in (PG-13, G)) \vee (age(u) < 13 \wedge rating(m) \in (G))$$

Các chính sách truy cập chi tiết hơn thường liên quan đến nhiều thuộc tính của các chủ thể và đối tượng. Trong các trường hợp như vậy, ABAC dễ dàng quản lý và mở rộng hơn. Để minh họa điều này, nhóm tác giả mở rộng ví dụ một chút: giả sử phim được phân loại thêm về chất lượng



Hình 2.7: Ví dụ về ABAC trong doanh nghiệp

phim được xem. Người dùng đã trả phí (premium) được xem loại phim có độ phân giải full-HD trong khi người dùng thường (nomal) chỉ được xem loại phim có độ phân giải HD. Trong ví dụ này, chính sách quy định R1 vẫn được áp dụng và bên cạnh đó, áp dụng thêm:

$$\textbf{R2: } can\_access(u,m,e) \leftarrow (membershipType(u) = Premium \vee (membershipType(u) = Nomal \wedge movieType(m) = HD)$$

và chính sách truy cập cuối cùng sẽ là:

$$\textbf{R3: } can\_access(u,m,e) \leftarrow \textbf{R1} \wedge \textbf{R2}$$

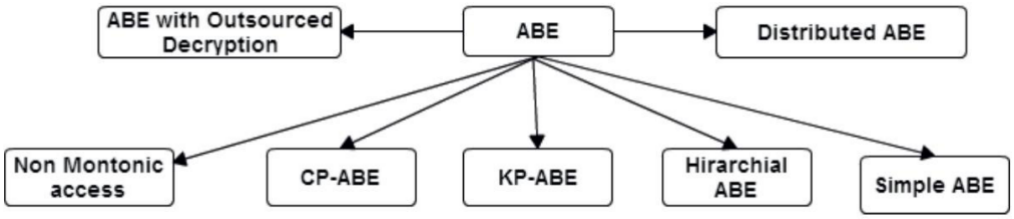
Bên cạnh đó, chính sách đối với điều kiện **e** cũng có thể áp dụng tương tự, như trẻ em chỉ được xem phim trong khung giờ cho phép hoặc các loại phim có yếu tố chính trị, tôn giáo có thể xem xét địa điểm truy cập của người dùng.

2.2 Quản lý truy cập dựa trên thuộc tính

Như đã đề cập ở phần trên, mô hình quản lý truy cập dựa trên thuộc tính có ưu điểm lớn so với các mô hình cũ. Hầu hết các doanh nghiệp ngày nay đang áp dụng các giải pháp quản lý định danh nhân viên, các thông tin về bộ phận làm việc, cấp bậc, chuyên môn... đều được lưu trữ và quản lý tập trung trên các phần mềm nhân sự (hoặc các phần mềm tùy biến do doanh nghiệp thiết kế). Điều quan trọng là làm sao tận dụng được các giải pháp này để quản lý các nhóm nhằm hiện thực ABAC.

Một yêu cầu quan trọng khác là cung cấp cơ chế quản lý khóa nhóm, vì mục tiêu của một nhóm là thường xuyên chia sẻ dữ liệu. Dữ liệu phải được mã hóa bằng các khóa chỉ thành viên trong nhóm đó biết. Việc quản lý các khóa này, bao gồm thêm mới, phân phối, lưu trữ và cập nhật phải hỗ trợ hiệu quả cho mô hình ABAC.

Không giống như việc quản lý theo mô hình RBAC, nơi mà hệ thống cấp quyền cho các vai trò (role), sau đó quản lý người dùng trong các vai trò, ABAC cung cấp quyền trực tiếp cho người dùng thông qua các thuộc tính mà người đó sở hữu. Thách thức lớn là phương pháp quản lý khóa hiệu quả dành cho nhóm người dùng có chung một số thuộc tính. Khi nhóm này thay đổi (thêm hoặc bớt người dùng) thì họ có/không truy cập hợp lệ tới các tài nguyên của nhóm, trong khi các thành viên cũ của nhóm không bị ảnh hưởng. Quá trình cấp khóa mới cho nhóm phải đúng với chính sách truy cập, đồng thời cũng không tác động lớn tới các thành viên cũ. Bên cạnh đó, việc chống lại các cuộc tấn công thông đồng của những người dùng xấu cũng đặt



Hình 2.8: Phân loại ABE

ra một thử thách.

Trong ví dụ ở phần 2.2, giả sử mỗi thành viên của hệ thống được cấp một khóa để có thể truy xuất vào bộ phim mà họ muốn xem. Khi một người dùng 'Premium' không tiếp tục trả phí, hệ thống cần loại bỏ quyền của họ đối với chất lượng full-HD của bộ phim M mà không cần tác động đến khóa của họ (thực tế là không thể tác động vì phải giao khóa cho người dùng), trong khi những người dùng khác đã có khóa được cấp hợp lệ vẫn xem được như cũ mà không cần cấp lại toàn bộ khóa mới. Việc chống lại các cuộc tấn công phối hợp như một người dùng nhỏ hơn 21 tuổi nhưng có hạng 'Premium' và một người dùng lớn hơn 21 tuổi có hạng "Normal' có thể phối hợp với nhau để xem bộ phim xếp loại R với chất lượng full-HD.

Nhằm hiện thực quản lý truy cập dựa trên thuộc tính, đồng thời giữ được tính bí mật của dữ liệu được lưu trên các dịch vụ đám mây, phương pháp mã hóa dựa trên thuộc tính, Attribute-based Encryption được phát triển với nhiều biến thể khác nhau được phân loại như hình 2.8. Trong phần tiếp theo, nhóm tác giả trình bày chi tiết thuật toán của hai phương pháp mã hóa dựa trên thuộc tính trong Hình 2.8 là CP-ABE và KP-ABE. Đây là hai phương pháp mã hóa nổi bật nhất. Các nghiên cứu về ABE theo hai hướng này ngày càng rộng và tăng nhanh.

Cả KP-ABE và CP-ABE đều sử dụng chung một cấu trúc truy cập, được gọi là cây truy cập. Gọi τ là một cây biểu thị cấu trúc truy cập. Mỗi điểm nút trong cây là một cổng truy cập. Một cổng có cấu trúc bao gồm các nút con của nó và một giá trị cổng. Giả sử $\boldsymbol{n}_x$ là số lượng nút của của nút $\mathbf{x}$ và $\boldsymbol{t}_x$ là giá trị cổng thì ta có $0 < t_x \leq n_x$. Nếu $t_x = 1$ thì cổng đó là cổng *OR*, và ngược lại, nếu $t_x = n_x$ thì đó là cổng *AND*. Mỗi nút lá (là nút không có nút nào là nút con của nó) diễn giải một thuộc tính và có giá trị cổng $t_x = 1$. Hàm ***att(x)*** lúc này trả về thuộc tính liên kết với nút lá x. Mỗi nút trong τ được đánh số thứ tự duy nhất. Hàm ***index(x)*** trả về thứ tự của nó, hàm ***parent(x)*** trả về nút cha của nó.

Giả sử τ là cây truy cập với nút gốc tại x, τ_x là cây con của τ có nút gốc tại x. Tập thuộc tính λ được gọi là thỏa mãn τ_x nếu $\tau_x(\lambda) = 1$. Việc tính toán $\tau_x(\lambda)$ được thực hiện như sau:

- Nếu x là nút lá, $\tau_x(\lambda) = 1 \iff att(x) \in \lambda$
- Nếu x không phải là nút lá, gọi các nút con của x là x'. Lần lượt tính toán các cây con $\tau_{x'}(\lambda)$. Nếu có ít nhất t_x cây con $\tau_{x'}(\lambda) = 1$ thì $\tau_x(\lambda) = 1$. Ngược lại, tập λ không thỏa mãn cấu trúc truy cập được quy định trong τ_x.