Hệ thống quản lý danh tính (IM)

I. Khảo sát hiện trạng các hệ thống hiện thời

I.1. Tổng quan

Sự cần thiết của một hệ thống quản lý danh tính

Những vi phạm an ninh gần đây trên toàn thế giới - Từ Target đến SONY, Anthem Inc. và Home Depot - là bằng chứng cho thấy thị trường trực tuyến nếu không bảo đảm, có nguy cơ trở thành mối hiểm họa bất cứ lúc nào. Do tỷ lệ tội phạm mạng ngày càng gia tăng nhanh, nên ngày càng có nhiều công ty đang quan tâm đến dữ liệu của công ty, vì gần không có sự ngăn chặn hiệu quả nào đối với những phương thức tấn công mạo danh hoặc tấn công nội bộ.

Giả mạo danh tính là việc sử dụng trái phép thông tin cá nhân của một người khác để chiếm đoạt tài sản một cách bất hợp pháp. Số nạn nhân bị giả mạo danh tính đã tăng lên từ 500.000 người tới 13,1 triệu người chỉ trong năm 2013, hầu hết các nạn nhân nằm trong độ tuổi từ 35 tới 44. Cứ khoảng hai giây lại có một một nạn nhân bị giả mạo danh tính.

Phát hiện giả mạo

Chúng ta biết rằng bị đánh cắp dữ liệu là một trong những kịch bản đáng sợ nhất. Đánh cắp dữ liệu có thể gây thiệt hại nhiều hơn chiếm đoạt tài khoản. Một trong ba người nhận được thông báo bị đánh cắp dữ liệu là nạn nhân của giả mạo danh tính. Trên 40% người tiêu dùng bị đánh cắp thẻ ghi nợ bị giả mạo danh tính.

Theo một báo cáo về nghiên cứu gian lận nhận dạng năm 2018 được công bố với bởi nhà nghiên cứu Javelin, chỉ ra rằng vào năm 2017, tổng số nạn nhân đã tăng thêm tăng 8% so với trước đó lên 16,7 triệu. Báo cáo của Javelin cũng cho thấy 6.64% người tiêu dùng đã trở thành nạn nhân của những vụ gian lận danh tính năm ngoái, tăng gần 1 triệu nạn nhân so với năm trước đó. Sự gia tăng này được thúc đẩy bởi sự tăng trưởng trong các chương trình gian lận phi thẻ và kế hoạch tiếp quản tài khoản (ATO).

Một trong kịch bản thường thấy là kẻ gian lận sau khi chiếm tài khoản của một người, thực thi những hành vi phạm pháp và sau đó thay đổi tài khoản trước khi nhân viên an ninh hoặc nạn nhân biết điều gì đã xảy ra. Javelin phát hiện thấy tổng thiệt hại của ATO đạt 5,1 tỷ đô la Mỹ trong năm ngoái, tăng 120% so với năm 2016. Các nạn nhân phải trả trung bình 290 đô la tiền túi, và dành trung bình 16 giờ để giải quyết ATO. Al-Pascual, phó giám đốc nghiên cứu, giám đốc nghiên cứu và là người đứng đầu về gian lận và an ninh của Javelin nói: “Các tên tội phạm cũng đang truy cập vào điện thoại di

động và sử dụng thông tin họ tìm thấy trên điện thoại của nạn nhân để tiến hành giao dịch trên một tài khoản khác.”

Một nghiên cứu khác, 2017 Fraud Index được công bố bởi Radial, cho thấy có sự gia tăng gấp bốn lần trong các vụ tấn công thẻ quà tặng kỹ thuật số từ Lễ Tạ ơn đến Giáng sinh trong mùa giải vừa qua. Ngoài ra, vận chuyển qua đêm vẫn là một phương thức thường xuyên bị tấn công bởi những kẻ gian lận. KC Fox, phó chủ tịch phụ trách thanh toán, nói: “Tôi nghĩ rằng các công ty phải đối mặt với việc vượt qua gian lận này đã trở thành một công việc toàn thời gian, không phải là điều bạn có thể làm cho khách hàng của mình làm việc trong thời gian rảnh rỗi”.

Cứ khoảng hai giây lại có một một nạn nhân bị giả mạo danh tính

Tổng thiệt hại lcủa các vụ giả mạo danh tính trong năm 2013 là 18 tỷ USD

Quá nhiều dữ liệu, kiểm soát quá ít

Mối quan tâm hàng đầu và gây ra sự thất vọng giữa các doanh nghiệp là ngày càng có nhiều ứng dụng trực tuyến hoặc dịch vụ đòi hỏi nhân viên phải kết nối với nhiều tài khoản khác nhau, kết quả là rất nhiều dữ liệu và rắc rối về quản lý cho các công ty. Các doanh nghiệp cần một cách để duy trì một hệ thống kiểm tra và cân bằng để biết dữ liệu nhạy cảm của họ đang được xử lý và thao tác từ người này sang người như thế nào.

Một điểm đáng lo ngại khác đó là việc kiểm soát quyền truy cập của người dùng mà không cần theo dõi sự riêng tư cá nhân và cách quản lý quyền truy cập trong hệ thống, tổ chức. Điều này đặc biệt quan trọng đối với các doanh nghiệp có nhân viên cộng tác từ xa hoặc có tỷ lệ tiêu thay mới nhân viên cao.

Sự nguy hiểm của tin tức giả

Hiện nay, thông tin giả mạo là một vấn đề lớn trong không gian mạng và không có biện pháp giải quyết triệt để. Cũng bởi một phần những hệ thống lớn như google, facebook không đưa ra những biện pháp để thắt chặt ngay từ ban đầu ví dụ như từ việc xác thực thông tin người dùng.Do vậy, người sử dụng hệ thống không có ý thức trách nhiệm đối với những thông tin mà bản thân cung cấp, pháp luật các nước cũng có những biện pháp để hạn chế sự lây lan của tin tức giả mạo, nhưng không thể ngăn chặn trước sự gia tăng quá nhanh của thông tin và nhu cầu của người sử dụng.

Sự bùng nổ của tin tức giả mạo (tin bịa đặt, sai sự thật) mang lại không ít phiền toái cho người sử dụng mạng xã hội trong năm vừa qua. Tại Mỹ, tin tức giả mạo cũng tràn ngập Facebook, Google, Twitter… đặc biệt liên quan đến các sự kiện lớn. Tại Việt Nam, số liệu thống kê từ chương trình đánh giá an ninh mạng của Bkav cho thấy, 63% người dùng thường xuyên đọc được tin tức giả mạo trên Facebook, trong đó 40% là nạn nhân hằng ngày. Không chỉ khiến người đọc hoang mang, tin tức giả mạo còn tiềm ẩn nguy cơ gây bất ổn xã hội khi kẻ xấu cố tình đưa tin sai sự thật liên quan đến tình hình kinh tế, chính trị của đất nước.

Đa nhân dạng và sự bất tiện đối với người dùng mạng

Hiện nay khi mà các dịch vụ Web-based hoặc Web-supplement liên tục phát triển với tốc độ cao, do vậy mỗi người dùng Internet thường phải quản lý hàng chục "nhân dạng" khác nhau tương ứng với mỗi hệ thống. Họ cần tên và mật khẩu để đọc email trực tuyến của họ, đặt mua sách từ cửa hàng trực

tuyến, đăng lên blog của họ, v.v. Ngoài các loại nhân dạng được sử dụng thường xuyên này, người sử dụng Internet thường phải đăng ký các dịch vụ mà họ hiếm khi sử dụng, và phải ghi nhớ tài khoản cho mỗi dịch vụ đó, một điều khá bất tiện.Và sau khi đã cung cấp thông tin cá nhân cho hệ thống/dịch vụ để thực hiện một số quy trình nhất định, người dùng thường có xu hướng quên nhanh và những thông tin cá nhân đó có thể sẽ bị hệ thống sử dụng với mục đích không mong muốn.

Theo khía cạnh khác, việc cô lập các hệ thống cũng gây ra những trải nghiệm không "liền mạch" của người dùng. Ví dụ một số doanh nghiệp có phát triển các hệ thống web để giới thiệu sản phẩm của họ, mặt khác họ sử dụng facebook là một kênh để tiếp cận đến lượng người dùng khổng lồ. Những người nào cảm thấy có hứng thú sẽ bấm vào những đường dẫn dẫn đến hệ thống web của doanh nghiệp để tìm hiểu thông tin hoặc đọc sách online, mua sách, v.v... Vậy để tăng tính trải nghiệm của người dùng, ta có thể lợi dụng sự tương tác của facebook vốn đã được tối ưu hóa dựa trên sự phân tích hành vi con người, và hệ thống chạy ngầm để cung cấp nền tảng thanh toán, thông tin sách,… Tức là cả hai hệ thống có thể

Do những vấn đề đã nêu trên đây và một số vấn đề liên quan khác, quản lý nhận dạng mạng đã được chủ đề nghiên cứu tích cực từ đầu thiên niên kỷ. Nhiều hệ thống quản lý nhận dạng đã được giới thiệu, một số trong số họ đang cố gắng trở thành các tiêu chuẩn chính thức được chấp nhận hoặc thực tế.

Sự cồng kềnh của các hệ thống

Đối với sự phát triển hệ thống doanh nghiệp, trong quá trình triển khai có thể phải xây dựng những hướng đi, giải pháp bổ sung để hoàn thiện mô hình tổng thể, giải pháp ở đây không chỉ đề cập đến công nghệ mà còn là mô hình kinh doanh.ví dụ một cửa hàng bán hoa quả nhận thấy nhu cầu người dùng là muốn được cung cấp sản phẩm đến tận nhà và muốn trả tiền qua thẻ,...Doanh nghiệp bán hoa quả sẽ phải xây dựng một bộ phận chuyển phát nhanh có hiệu suất lớn vì nhu cầu thị trường cao, và song song là phát triển các nền tảng công nghệ để đáp ứng việc trả tiền thông qua ngân hàng , hoặc quản lý, tiếp nhận giao hàng, bộ phận chăm sóc khách hàng qua hotline,... Tùy theo quy mô, độ phức tạp mà chi phí để xây dựng và duy trì có thể tăng lên khá nhiều, và điều đó khó có thể phù hợp với những hệ thống vừa và nhỏ, không có tiềm lực mạnh. Về lâu dài, việc phát triển này sẽ còn kéo theo nhiều nhược điểm tiềm tàng của doanh nghiệp như

- Tốc độ phát triển bị giới hạn, tính đáp ứng chưa cao, nhiều xu hướng phát triển có thể đột ngột xuất hiện và lan rộng trong thời gian ngắn, để bắt kịp và tận dụng hoàn toàn thời cơ, những doanh nghiệp sẽ phải gia tăng khối lượng đầu tư để phát triển hệ thống. Tuy vậy, vì hệ thống được triển khai hoàn toàn độc lập, do đó vẫn bị hạn chế về nguồn lực xây dựng, và có thể sau

một khoảng thời gian khi xu hướng chìm xuống, hệ thống chưa kịp thu lợi đủ thì đã trở nên vô ích.

- Sự nặng nề của hệ thống sẽ gia tăng từng ngày, cùng với đó là những chi phí phát sinh như quản lý, bảo trì,… sẽ là mối lo ngại đối với những doanh nghiệp quy mô trung bình.
- Không thể đầu tư nguồn lực đủ lớn để cạnh tranh đối với những hệ thống đã có uy tín. Nguồn lực của một doanh nghiệp là có hạn, và phải đầu tư chủ yếu vào hướng đi xương sống của toàn bộ hệ thống, điều đó cũng hạn chế việc đầu tư xây dựng những ý tưởng phát triển tiềm năng, nhất là đối với những doanh nghiệp có nguồn lực trung bình. Nhưng dù kể cả đó là những doanh nghiệp khổng lồ, dư dả về nguồn lực sẽ vẫn còn những khó khăn khác phải đối mặt .Một ví dụ như facebook gần đây đã tung ra khá nhiều hướng đi để cạnh tranh với những ông lớn khác như
  - Search Engine - Trong thời gian một hai năm trở lại đây, để cạnh tranh với công cụ tìm kiếm google, facebook đã và đang cải tiến bộ lọc tìm kiếm của mình và có một số cơ chế "uốn" hành vi của người dùng đem lại một số điều bất tiện. Ví dụ nếu ngày trước, tìm kiếm tài khoản của người dùng sẽ chuyển hướng ngay đến tài khoản của người dùng sau thao tác nhập thì giờ sẽ chuyển hướng đến một trang lọc kết quả tìm kiếm khá gượng gạo, và một lượng lớn người sử dụng chưa tin dùng chức năng này của facebook.
  - Ngược lại Google đã cho ra google+ cách đây vài năm với tham vọng cạnh tranh trực tiếp với facebook, và thực tế đã rõ ràng là hầu như không ai sử dụng, mặc dù cộng đồng người sử dụng những tiện ích của google như gmail là cực kỳ lớn.
  - Thương mại điện tử - một mảng mà facebook cũng khá quan tâm, đã cho ra một số tiện ích giúp người bán đăng tin sản phẩm, hoặc một số chức năng tiếp cận khách hàng, nhưng những chức năng đó chưa thể hình thành nên một mạng lưới thương mại điện tử đúng nghĩa, vẫn dừng ở lại ở khía cạnh rất cơ bản của facebook là đăng tin tức. Nếu facebook có mong muốn đánh bại một tập đoàn cực lớn trong lĩnh vực này là Amazon thì có thể nói đây là những bước đi chưa sắc nét.

Vậy nếu có một viễn cảnh mà khi 3 hệ thống trên: Facebook, Google, Amazon không còn cạnh tranh mà cùng hợp tác để hình thành nên một mô hình kinh tế chia sẻ khổng lồ, thì có thể nói những twitter, alibaba ,...sẽ khó có thể cạnh tranh sòng phẳng như trước được nữa. Amazon có thể tiếp cận đến nguồn khách hàng khổng lồ được cung cấp bởi facebook hay như facebook tận dụng được những yếu tố công nghệ của google để tối ưu trải nghiệm người dùng,...Nhưng nói chung sự bắt tay giữa những tập đoàn cực lớn là khó có thể xảy ra vì một vấn đề rất cơ bản và khó giải quyết, đó là phân chia lợi ích mà mô hình cộng tác đem lại, đó có thể là những khoản lợi nhuận khổng lồ, hoặc những quyền lợi cấp cao trong mối quan hệ hợp tác,…

Tài sản ảo, cơ hội và rủi ro

Định nghĩa

Về tính pháp lý: Tài sản ảo là một khái niệm rất rộng như tên miền internet, địa chỉ hộp thư điện tử, các loại tài khoản Game Online,… phổ biến nhất là tài sản ảo trong trò chơi trực tuyến, tên miền, hoặc trong thời gian gần đây là tiền điện tử. Tiếp cận theo nghĩa hẹp, tài sản ảo là các đối tượng ảo trong thế giới ảo, còn theo nghĩa rộng thì tài sản ảo được hiểu là những tài nguyên trên mạng máy tính được xác định giá trị bằng tiền và có thể chuyển giao trong các giao dịch dân sự. Khái niệm này được tiếp cận thông qua tư duy lý luận về quyền tài sản.

Về bản chất: Tài sản ảo chỉ là hình ảnh thể hiện ra bên ngoài, mà bên trong chính là thông tin tồn tại dưới dạng các đoạn mã máy tính. Các đoạn mã khác nhau tạo nên những loại tài sản ảo khác nhau. Chính vì vậy, tài sản ảo cũng có sự thống nhất của tính chất nội tại và hình ảnh bên ngoài như bất kỳ tài sản thông thường nào khác. Tuy nhiên, do các đoạn mã máy tính không tồn tại độc lập hoàn toàn nên không thể thực hiện quyền chiếm hữu như tài sản thông thường mà chỉ có thể thực hiện được quyền này thông qua giá trị bằng tiền của tài sản ảo đó. Điều này về bản chất không khác với quyền sở hữu trí tuệ (có tính vô hình) đã được thừa nhận là một loại quyền tài sản. Tương tự như vậy, việc thừa nhận tài sản ảo là các đoạn mã ghi nhận quyền của người chơi sẽ có ý nghĩa quan trọng trong việc bảo hộ và khai thác các lợi ích của tài sản ảo, đồng thời giải quyết được nhiều vấn đề về tài nguyên mạng đã gây tranh chấp hiện nay.

Về giá trị: Tài sản ảo có giá trị kinh tế và giá trị sử dụng vì nó đáp ứng những nhu cầu của con người. Game online đáp ứng nhu cầu về giải trí; tên miền cung cấp một hình thức đại diện cho doanh nghiệp, cơ quan, thương hiệu,…

Nói chung tài sản ảo vẫn còn nhiều điểm thiếu sót, gây mơ hồ, vậy nên vẫn còn là một thị trường không chắc chắn trong những giao dịch. Không có một phương pháp quản lý, tập hợp và định quyền sở hữu nên những loại tài sản này gặp rất nhiều rủi ro. Ví dụ khi trao đổi một vật phẩm trong game, tiền thật sẽ là thứ được đem ra là phương thức trao đổi, và khi game đó bị đóng cửa, mặc nhiên vật phẩm đó hoàn toàn vô giá trị và không còn khả năng giao dịch, hoặc tự tái đầu tư. Khi sử dụng giá trị trong thế giới thật (tiền) để trao đổi giá trị ảo thì một cách gián tiếp, tài sản ảo có thể được định ra làm hai trạng thái : hoặc là dịch vụ để phục vụ mục đích giải trí, hoặc là sản phẩm hàng hóa có khả năng lưu thông. Nếu theo định nghĩa tài sản ảo là một trong những loại "dịch vụ" thì có thể chấp nhận tính "bốc hơi" vô điều kiện của nó, tức là khi tài sản đã hết giá trị sử dụng. Tuy vậy, tài sản ảo vốn không thể nghĩ theo một cách đơn giản vậy, nhưng định nghĩa ban đầu, tài sản ảo cũng có sự thống nhất của tính chất nội tại và hình ảnh bên ngoài

như bất kỳ tài sản thông thường nào khác. Nó có thể được lưu trữ, có thể được sở hữu và trao đổi. Vậy nên có thể nói tài sản ảo mang cả tính dịch vụ cũng như tính sản phẩm.

Những sự nổi lên gần đây của tài sản ảo, như tiền điện tử đã trở thành một trào lưu cực lớn trên toàn cầu và lôi kéo được hàng triệu người tham gia nhưng bên cạnh đó là những mối hiểm họa tiềm tàng điển hình như

- Các giao dịch sử dụng Bitcoin có tính ẩn danh cao cho nên Bitcoin có thể sẽ trở thành một công cụ cho tội phạm như là buôn bán ma túy, rửa tiền, giao dịch, thanh toán tài sản phi pháp, trốn thuế.
- Bitcoin là một loại tiền ảo được lưu giữ dưới dạng kỹ thuật số do đó nguy cơ bị tấn công, đánh cắp hay thay đổi dữ liệu hoặc sẽ bị ngừng giao dịch là rất lớn.
- Do giá trị của đồng Bitcoin biến động mạnh và rất phức tạp trong thời gian ngắn nên các hoạt động đầu tư vào Bitcoin tiềm ẩn nhiều nguy cơ về bong bóng, nguy cơ gây thiệt hại cho người đầu tư.
- Bitcoin không bị chi phối và kiểm soát giao dịch bởi bất kì cơ quan quản lý nhà nước nào, do đó,những người sở hữu Bitcoin sẽ chịu toàn bộ những rủi ro vì không có cơ chế bảo vệ quyền lợi.

Do vậy để hiện thực hóa thị trường tài sản ảo, cần có một phương thức quy đổi tin cậy giữa các loại tài sản ảo, vdu như tài sản trong game dota và lol có tỷ giá đối với nhau như thế nào, để tạo ra một phương thức giao dịch chuẩn hóa giữa nhà cung cấp và người dùng, cũng như người dùng với nhau. Việc định giá không chỉ thể dựa vào những nhà cung cấp, hoặc những người tạo ra mà không có một thang đo so sánh chung. Và giá trị tài sản cần được niêm yết hoặc thống nhất trong toàn bộ không gian ảo. Đây sẽ là bước đầu tiên cũng như quan trọng nhất trong việc hiện thực hóa quyền sở hữu tài sản ảo. Cũng như hàng hóa, tiền tệ được định giá thông qua vàng, thang đo cần giải quyết vấn đề chuyển đổi những giá trị ảo với nhau, và cả giá trị ảo với giá trị thật. Ngoài ra, cần một phương pháp quản lý tài sản ảo để giảm thiểu rủi ro giao dịch, thất thoát giá trị.

I.2. Mục tiêu

- Xây dựng nên một mô hình đạt được độ tin cậy cao về tính xác thực, là một môi trường an toàn và dự báo rủi ro hiệu quả đối với các bên liên quan.

- Thúc đẩy sự phát triển của nền kinh tế chia sẻ giữa các doanh nghiệp, dịch vụ, xây dựng nên những quy trình nghiệp vụ hoặc dịch vụ thông suốt giữa các hệ thống.
- Thay đổi cách nhìn về quyền lực của người sử dụng, gắn kết các cộng đồng người tiêu dùng thành một tổ chức thống nhất để thành một đối trọng đối với các doanh nghiệp, dịch vụ.
- Trở thành giải pháp để hiện thực hóa việc quản lý tài sản ảo, vdu tiền điện tử. Xây dựng ra một thước đo giá trị dành cho tài sản ảo, phương thức quản lý và minh bạch hóa tài sản ảo.

I.3. Những hệ thống đã phát triển

I.3.1. Thekey

I.3.1.1. Mô tả chung

Dự án ICO THEKEY được khởi xướng bởi một công ty Singapore, đã thành lập từ năm 2014, hoạt động trong lĩnh vực xác nhận thông tin cá nhân. Hiện tại công ty đã có 23 giấy chứng nhận quyền sở hữu và 15 bằng sang chế trong lĩnh vực liên quan. THEKEY đã thực hiện chạy thử dự án ở 2 thành phố và đạt được những kết quả khả quan.

Dự kiến THEKEY sẽ là ứng dụng hợp đồng thông minh trên nền tảng NEO, trở thành một phần của hệ sinh thái NEO gồm tài sản số, danh tính số và hợp đồng thông minh.

Token TKY sẽ là loại token được sử dụng trong hệ sinh thái THEKEY. Dự kiến hệ sinh thái này sẽ bao gồm 3 thành phần chính: người xác nhận, nhà cung cấp dịch vụ và các cá nhân. Token TKY sẽ được sử dụng để kích thích các cá nhân tham gia vào cộng đồng. Token TKY sẽ được sử dụng khi người dùng kí hợp đồng thông minh với hệ thống, và có thể dùng để mua các dịch vụ và các sản phẩm trong hệ sinh thái THEKEY

I.3.1.2. Ưu điểm

Dự án đã nhận được ốn giai đoạn seeding từ quỹ đầu tư mạo hiểm, và đã thu được hàng chục bản quyền và bằng sáng chế. Ngoài ra, hệ thống xác minh danh tính của công ty đã được triển khai trong hai dự án thí điểm ở các thành phố Kaifeng và Jiaxing, Trung Quốc. Thêm 41 thành phố đang được triển khai. Hơn nữa, nền tảng này đã thu thập dữ liệu cá nhân của hàng trăm triệu người. Nói cách khác, THEKEY không phải là ICO tầm thường để kiếm tiền nhanh. Nó đã thiết lập một sự hiện diện mạnh mẽ trên thị trường trước khi tung ra token của nó.

THEKEY là một số ít dự án trong năm nay có đội ngũ các chuyên gia rất mạnh về các lĩnh vực bảo mật, khoa học dữ liệu, lập trình và blockchain.

Công ty đã và đang hợp tác với các đối tác mạnh của Trung Quốc như China Unicom, PingAn Insurance, China Re và China Minsheng Bang.

Hardcap của dự án khá thấp chỉ khoảng 22 triệu USD, điều này rất hấp dẫn nhà đầu tư. Nhận được sự quan tâm lớn từ nhà đầu tư với hơn 7,000 người trên telegram.

Dự án sẽ được tích hợp hợp đồng thông minh NEO và công nghệ xác nhận danh tính đa chiều BDMI, điều này cho phép hệ thống xác nhận sẽ có ưu điểm sau:

- Tích hợp được nhiều ngôn ngữ lập trình, phù hợp với nền tảng NEO, tăng tốc độ xử lý và khả năng mở rộng dự án.
- Công nghệ BDMI cho phép tích hợp nhiều chỉ số sinh học cho một cá nhân như vân tay, đồng tử mắt, ngày sinh.
- Cho phép kiểm tra chéo giữa các thông tin khác nhau, đảm bảo tính chính xác và minh bạch của thông tin được cung cấp. Để minh họa mình có thể lấy ví dụ , trong trường hợp bạn chỉ cần dữ liệu về đồng tử mắt, bạn có thể xác nhận tất cả các thông tin về nơi sinh, nghề nghiệp…
- Tăng cường tính bảo mật dữ liệu cá nhân khi lưu trữ trên mạng blockchain, điều này tránh việc xóa sửa thông tin gây sai lệch hay việc tấn công ăn trộm dữ liệu.
- **Cho phép xác nhận thông tin ẩn danh mà không lộ danh tính, điều này cho phép bạn xác nhận thông tin công cần cung cấp các thông tin không cần thiết.**

I.3.1.3. Nhược điểm

Mặc dù token TKY có một triển vọng lớn, nhưng nó có thể có các trường hợp sử dụng hạn chế. THEKEY đang tìm kiếm để lấy thông tin nhận dạng cá nhân của người dùng và đặt nó vào blockchain nơi nó sẽ tồn tại mãi mãi. Tùy thuộc vào quan điểm ý thức hệ và quan điểm về sự riêng tư, đây có thể là một trở ngại của hệ thống. THEKEY dường như là một dự án lý tưởng cho các nền kinh tế tập trung hơn như Trung Quốc. Công ty đã thu thập dữ liệu nhận dạng của hàng trăm triệu người từ chính phủ và các nguồn tài chính. Một dự án có tính chất như vậy có thể sẽ gặp nhiều khó khăn hơn khi thực hiện ở các nước phương Tây với sự riêng tư, bảo mật được đặt lên cao hơn.

THEKEY đã thiết lập mối quan hệ chặt chẽ với các công ty hàng đầu thế giới, nhưng chưa đưa ra mô hình kinh doanh. Trong Whitepaper, công ty cho biết mô hình kinh doanh các sản phẩm IDV đã "đã được chứng minh phần nào" mà không đi sâu vào chi tiết cụ thể.

II. Các vấn đề của hệ thống hiện thời và cơ hội để đề xuất hệ thống mới

II.1. Những vấn đề của một hệ thống xác thực

1) **Di sản** - Đối với những hệ thống cũ, sử dụng những nền tảng xây dựng đã lỗi thời thì việc tham gia vào mô hình sẽ là một trở ngại lớn. Việc này có thể tiêu tốn nhiều chi phí,mà khi hệ thống IM chưa đủ mạnh và chứng minh được những lợi ích nó mang lại cho mô hình nhiều hơn những chi phí mà doanh nghiệp phải bỏ ra thì việc cộng tác là rất khó khăn.
2) **Đa dạng** - Các quy trình nghiệp vụ của các hệ thống sử dụng các ứng dụng được xây dựng bằng ngôn ngữ bất kỳ, trên bất kỳ nền tảng nào, sử dụng bất kỳ cơ sở dữ liệu nào, vậy nên sự kết nối công nghệ sẽ gặp nhiều khó khăn.
3) **Phân quyền nội bộ** - Không chỉ người dùng hệ thống và những hệ thống ghép nối là những đối tượng được phân quyền truy cập và thao tác các tài nguyên, mà sự phân quyền còn cần được thực hiện trên chính những cá nhân quản lý,duy trì hệ thống chung.
4) **Quyền lực tập trung** - Vì hệ thống lưu trữ một lượng thông tin không thuộc sở hữu và cung cấp cho những bên liên quan, nên không tránh khỏi những câu hỏi về tính khách quan,đúng đắn và hoàn toàn trung lập đối với những thông tin đó.
5) **Tài khoản người dùng** - Mỗi một hệ thống đều có một hệ sinh thái người dùng đã và đang duy trì,việc loại bỏ chúng mang lại nhiều tổn thất nhiều hơn là lợi ích,vdu như nếu việc hợp tác kết thúc,những hệ thống ghép nối cần phục hồi lại hệ sinh thái người dùng của họ,việc này có quá nhiều khó khăn và rủi ro.
6) **Cân bằng lợi ích** - Những bên tham gia mô hình có những nhu cầu riêng và có thể bị xung khắc với nhau.Vdu như người sử dụng muốn kiểm soát hoàn toàn việc phân phối thông tin cá nhân,và cũng có thể hạn chế đi việc chia sẻ những thông tin nhạy cảm.Trong khi đó các hệ thống ghép nối tham gia vào mô hình không chỉ muốn một nguồn thông tin "sạch" mà còn muốn khả năng khai thác tối đa lượng thông tin người dùng.Hoặc những bên như tổ chức,cơ quan chính phủ muốn kiểm soát hoàn toàn thông tin của người dùng, bằng cách trực tiếp hoặc gián tiếp sẽ yêu cầu quyền sở hữu, thao tác hoặc chi phối một phần hoặc toàn bộ hệ thống.
7) **Tính phức tạp** - Sự phân quyền có thể tạo ra những vòng tròn rối rắm giữa các bên liên quan khiến việc kiểm soát các nguồn tài nguyên trở nên khó khăn. Rủi ro này có thể xảy ra thường xuyên khi có nhiều người dùng cũng không ý thức được những rủi ro và lợi ích của những việc bản thân đang làm.

8) **Chi phí** - Chi phí để xây dựng hệ thống sẽ tốn kém tài chính và thời gian, trong khi đó lợi ích thu được có thể đến chậm hoặc với lượng ít.
9) **Con đường tiếp cận người dùng** - Việc người dùng tiếp cận đến hệ thống hoặc ngược lại là một vấn đề khó cần giải quyết, tiếp cận trực tiếp, hay gián tiếp qua kênh trung gian như chính phủ, doanh nghiệp? Nếu tiếp cận thông qua những bên trung gian sẽ gặp phải những chi phối một phần từ chúng, điều này cũng cần cân nhắc. Còn nếu tiếp cận trực tiếp tới người dùng, thì tiện ích nào của hệ thống có đủ khả năng hấp dẫn người sử dụng?

II.2. Cơ hội và thách thức để đề xuất hệ thống mới

II.2.1. Cơ hội

II.2.1.1. Thị trường

Lĩnh vực xác thực thông tin cá nhân là một thị trường tiềm năng có doanh thu trên 6.5 tỷ USD, một năm với tốc độ tăng trưởng trên 20% trong vòng năm tới, dự kiến sẽ đạt trên 34 tỷ USD trước năm 2025.

II.2.1.2. Công nghệ

Những nền tảng công nghệ nổi lên trong thời gian gần đây như blockchain trở thành một trong những giải pháp hàng đầu để giải quyết nhiều bài toán liên quan đến độ tin cậy và xác thực của thông tin.

II.2.2. Thách thức

Ngoài những vấn đề đã được nêu ra trong "**Những vấn đề của một hệ thống xác thực**", sẽ còn những trở ngại khi triển khai hệ thống

Chưa bổ sung

III. Xác định nhu cầu thực sự của các bên liên quan trong ngữ cảnh hệ thống mới

III.1. Các bên liên quan(mô tả chung)

Người dùng hệ thống là những cá nhân muốn tham gia trực tiếp với mục đích muốn lưu trữ thông tin cá nhân lên hệ thống (ít) hoặc tham gia hệ thống thông qua các hệ thống ghép nối với IM (chủ yếu).

Hệ thống hỗ trợ xác thực là những hệ thống công nghệ thông tin nắm giữ những thông tin quan trọng, có độ chính xác và tin cậy cao của những người dùng hệ thống, tham gia với vai trò xác thực thông tin mà người sử dụng cung cấp. Vdu như các hệ thống công nghệ thông tin mà các tổ chức chính phủ đã xây dựng.

Hệ thống IM là hệ thống trung tâm với vai trò lưu trữ dữ liệu người dùng,thực hiện các công việc xác thực,phân quyền sử dụng dữ liệu đối với các hệ thống ghép nối.

Hệ thống sử dụng IM là các hệ thống sử dụng thông tin người dùng hệ thống, có thể tham gia xác thực danh tính dựa trên sự cung cấp dữ liệu phân tích hành vi.

Hệ thống hỗ trợ tính toán (offchain) là các hệ thống tham gia vào mô hình với mục đích cung cấp năng lực tính toán để giảm tải những tính toán trung gian trên toàn hệ thống.

III.2. Nhu cầu thực sự

Nhu cầu của các hệ thống sử dụng IM

Nhu cầu xác thực danh tính người dùng đang ngày càng trở nên quan trọng trong những hệ thống lớn cần sự chính xác về danh tính của người dùng như giao dịch thương mại điện tử,…Các tổ chức tài chính cần biết và hiểu khách hàng cũng như các giao dịch tài chính của họ để quản lý rủi ro. Ví dụ Các ngân hàng hoặc công ty tài chính nhà ở phải hiểu rõ khách hàng bằng cách sử dụng chương trình nhận diện khách hàng (Customer Identification Program - CIP) sử dụng nguồn văn bản, dữ liệu hoặc thông tin đáng tin cậy và độc lập để xác minh danh tính của khách hàng, địa chỉ tạm trú và thường trú, bản chất của cá thể kinh doanh, tình hình tài chính và các thông tin tương tự khác.Một ví dụ nữa trong những vấn đề rủi ro có thể gặp phải là gian lận thẻ tín dụng. Người sử dụng thẻ tín dụng không cần thiết là chủ thẻ tín dụng hay không, chỉ đơn giản là nhập vào các thông tin cần thiết trên thẻ tín dụng sẽ cho phép các khoản thanh toán để tiến hành trực tuyến. Mặc dù ngày nay, các ngân hàng có thể cung cấp một cơ chế xác nhận đôi như gửi một mã ngẫu nhiên đến điện thoại di động đã đăng ký. Tuy nhiên, nếu điện thoại di động đã bị đánh cắp cùng một lúc thì khả năng rủi ro vẫn tồn tại.

Đối với những hệ thống muốn tham gia vào một dây chuyền xử lý dịch vụ chung,hoặc một mô hình hợp tác,dữ liệu khách hàng phải được xác định khách quan,độc lập, đồng bộ và được tất cả các bên thừa nhận tính đúng đắn.Đây cũng là nền tảng để xây dựng một nền kinh tế chia sẻ.

Nền kinh tế số đã và đang đứng trước những thay đổi sâu sắc trong cách chúng ta tương tác, kinh doanh và kết nối với nhau. Một trong những chủ đề nổi bật nhất xuất hiện từ mô hình này là nhu cầu phải xác nhận lẫn nhau một cách số hóa và để đảm bảo rằng dấu chân kỹ thuật số của chúng ta là an toàn và không thể thay đổi. Khi các ranh giới giữa kỹ thuật số và thực tế ngày càng được xóa nhòa, nhu cầu về các công cụ xác minh danh tính mạnh mẽ đang trở nên ngày một quan trọng hơn.

Nhu cầu của người dùng

Trong cuộc sống hàng ngày, con người đang phải liên tục thực hiện xác minh danh tính trong các công việc thông thường như thanh toán, đặt các chuyến bay, hay sử dụng bảo hiểm y tế. Tuy nhiên, công nghệ này hiện nay đang ẩn chứa rất nhiều rủi ro như bảo mật dữ liệu, và việc đánh cắp nhận dạng tiêu tốn của người dùng hàng tỷ đô mỗi năm.Do vậy cần một hệ thống với những chỉ số an toàn đủ để đáp ứng nhu cầu bảo mật.

Thời kỳ số hóa đang diễn ra mạnh mẽ,nhu cầu số hóa tài liệu cá nhân cũng là một yếu tố cần thiết đối với con người hiện đại,để giảm tải đi sự cồng kềnh của việc sử dụng giấy tờ truyền thống.Giả sử khi muốn xuất nhập cảnh để du lịch sẽ không cần chứng minh thư hoặc hộ chiếu khi dữ liệu người dùng đã được lưu trữ và xử lý trên hệ thống,người dùng chỉ cần xác nhận danh tính thông qua dữ liệu sinh trắc học hoặc thông qua những lớp bảo mật cần thiết.

Đối với việc chia sẻ thông tin cá nhân cho những hệ thống CNTT, cần cùng cấp một phương pháp kiểm soát dễ dàng,hiệu quả cho người dùng.

Đối với những tài sản ảo, như tiền ảo, người dùng cần giải pháp để thu thập và quản lý một cách thống nhất, an toàn và dễ dàng.

Nhu cầu của hệ thống hỗ trợ xác thực

Các hệ thống hỗ trợ việc xác thực,tiêu biểu như ngân hàng,các hệ thống lưu trữ dữ liệu công dân,… tham gia vào mô hình có thể với một số mục đích:

Gia tăng sự ảnh hưởng của tổ chức - đối với những hệ thống này, sự tham gia mô hình gần như không phải vì mục đích lợi nhuận mà là mục tiêu tiếp cận đến các hệ thống khác trong mô hình và cả những dữ liệu phong phú của người dùng. Sức chi phối của những hệ thống này trong mô hình khá lớn, do nắm giữ thông tin của người dùng đã thông qua những quy trình xác thực và đã được kiểm chứng, nên cần thiết lập những cơ chế hỗ trợ để cân bằng quyền lực của các hệ thống.

Nhu cầu của hệ thống hỗ trợ tính toán

Đúng như tên gọi,các hệ thống này tham gia vào mô hình với mục đích cung cấp khả năng tính toán cho hệ thống để chịu tải bớt khối lượng công việc. Điểm yếu của phương pháp là hạn chế tính chất phi tập trung của blockchain, và chịu sự ảnh hưởng gián tiếp từ các tổ chức chính phủ, hơn nữa sẽ trở nên phụ thuộc vào các công ty cung cấp. Nhu cầu chính của các hệ thống này là các nguồn lợi trực tiếp (tài chính) hoặc một số quyền lợi ngầm định.

Dựa trên những hệ thống và nền tảng đã xây dựng, hệ thống IM sẽ tận dụng phát huy những điểm mạnh của chúng,cụ thể là:

Nguồn lợi của hệ thống

Bổ sung sau

Tầm nhìn phát triển

Điều cần được xem xét là IM không chỉ tuyệt vời cho mục đích bảo mật, mà còn là một công cụ kinh doanh xuất sắc. Nó cho phép các doanh nghiệp chia sẻ các ứng dụng của họ với bất kỳ đối tác khác thông qua một kênh an toàn. Điều này có nghĩa là cả hai công ty và đối tác sẽ nằm trên cùng một kênh về an ninh và có thể dễ dàng lập kế hoạch và phát triển kinh doanh với nhau.

Như đã nêu ở phần đầu tiên, một số hệ thống sẽ có nhu cầu xây dựng những giải pháp, mô hình bổ sung để phát triển mô hình chung, có thể là do xu hướng thị trường. Nếu những giải pháp đó hoàn toàn mới, thì việc xây dựng hoàn toàn từ đầu là điều khó tránh khỏi, mặt khác nếu chúng đã được một số tổ chức , hệ thống xây dựng hoàn thiện và sự nhận được kiểm chứng của thị trường, thì việc hợp tác sẽ mang lại lợi ích cho cả đôi bên. Bên cần giải pháp sẽ mất bớt chi phí và thời gian hoàn thiện, bên cung cấp giải pháp thì được tiếp cận với đối tác mới, và có thể là cả thị trường người dùng tiềm năng mà sự cộng tác mang lại. Hơn nữa việc cộng tác còn giúp các bên không cần mất chi phí duy trì những chức năng cồng kềnh mà có thể không sử dụng nhiều, tận dụng được chất lượng và sự uy tín của các hệ thống cộng tác mang lại để củng cố độ tin cậy của dịch vụ/mô hình. Có thể nói, đối với những hệ thống vừa và nhỏ, mô hình kinh tế chia sẻ sẽ giải quyết vấn đề cạnh tranh đối với những hệ thống đã có tiềm lực lớn.

Với một cơ chế phân quyền rõ ràng, hệ thống có thể tiến xa hơn trong vấn đề quản lý các nguồn lực trong nhiều hệ thống khác nhau, cung cấp một giải pháp tổng quát để phân phối quyền lợi giữa các cá nhân, tổ chức trên những vấn đề trừu tượng.Ví dụ như những tài sản ảo hoặc những nguồn dữ liệu,những chức năng được quyền thao tác trên các hệ thống. Để hình thành nên một hệ sinh thái với các tiến trình thông suốt giữa các hệ thống khác nhau.

Đối với mỗi một người trong thế giới thực, sẽ có một và chỉ một định danh số hóa trên không gian mạng, tất cả những hệ thống khai thác thông tin của người dùng sẽ hoàn toàn rõ ràng và được kiểm soát dễ dàng bởi người sử dụng.

Nếu hệ thống đạt được một sự tin cậy nhất định và đủ lớn, dữ liệu người dùng hệ thống trở thành một nguồn thông tin chính quy và có được sự công nhận của các bên có đặc quyền pháp lý như chính phủ, sẽ mở ra những hướng đi tiềm năng để khai thác.Vdu như khi hai quốc gia cùng sử dụng hệ thống IM làm trung gian, thì việc xuất nhập cảnh hoàn toàn được kiểm soát dễ dàng mà không cần trải qua thủ tục rườm rà,hoặc những tài sản ảo giao dịch ngầm sẽ không là mối nguy ngại với chính phủ,hoặc hơn nữa IM có thể là hệ thống tiền đề để phát triển những dự án khác thuộc nhiều mảng khác nhau như y tế cộng đồng, quản lý học sinh sinh viên,...Tất nhiên không dễ gì để đạt được những yếu tố như vậy. Hầu hết các hệ thống hiện tại, chính phủ hoặc phi chính phủ, đều tạo ra được những hệ sinh thái người dùng của riêng mình nhưng lại đều đóng đối với những hệ thống khác.Đối với những thông tin mang tính chất nhạy cảm cao,thì việc này hoàn toàn hợp lý.Nhưng để đạt được những lợi ích khác, như tối đa hóa thủ tục, dịch vụ, cắt giảm khối lượng công việc xử lý, giảm sức người,… thì việc mở rộng quy mô dữ liệu là khó có thể tránh khỏi.

Một mục tiêu lớn hơn, hệ thống không chỉ sẽ xác thực một đối tượng duy nhất là người sử dụng, mà còn sẽ tiến tới việc kiểm soát tính xác thực của các hệ thống, dịch vụ. Người sử dụng không còn là đối tượng trung tâm duy nhất bị đánh giá nữa, mà các dịch vụ, hệ thống tham gia mô hình, hoặc không tham gia cũng trở thành đối tượng bị đánh giá, dữ liệu đánh giá có thể đến từ những người đã và đang sử dụng dịch vụ, hệ thống.

IV. Đề xuất những phương pháp để đáp ứng nhu cầu của các bên liên quan và những vấn đề của hệ thống

Về tổ chức mô hình

Sẽ có những thành phần sau đây tham gia mô hình

- Người dùng hệ thống
- Các hệ thống hỗ trợ xác thực
- Hệ thống IM
- Các hệ thống sử dụng IM
- Các hệ thống hỗ trợ tính toán (offchain)

Dưới đây là những giải pháp để đáp ứng những nhu cầu của các bên và những hạn chế của những hệ thống xác thực hiện nay đã được nêu ở trên. Phần này sẽ được chia ra làm 3 phần:

- Mô hình cộng tác
- Mô hình kinh doanh
- Mô hình công nghệ

IV.1. Mô hình cộng tác

IV.1.1. Cân bằng quyền lực giữa các bên

IV.1.1.1. Xác thực

Những hệ thống quản lý danh tính hiện nay hướng tới việc xác thực thông tin người dùng và dùng những thông tin đã được xác thực này cung cấp tới những hệ thống cần, nhưng thực tế mô hình như vậy là chưa đủ tính tổng quát khi lợi ích sẽ nghiêng về những hệ thống, dịch vụ hơn là người sử dụng IM. Hơn nữa, do những vấn đề lừa đảo dịch vụ không phải là hiếm và có chiều hướng gia tăng mạnh, những hệ thống, dịch vụ cũng phải là những đối tượng để thực hiện tính xác thực.

IV.1.1.2. Quản lý thông tin

Đối với việc xác thực danh tính của người dùng, sẽ có thể có những đối tượng tham gia sau đây

- Các hệ thống hỗ trợ xác thực (chính quyền)
- Các hệ thống sử dụng IM

Và tương ứng có các kiểu xác thực danh tính như sau:

- Xác thực dựa trên những hệ thống pháp lý có sự đầu tư về cơ sở hạ tầng công nghệ thông tin,vdu ở đây là các hệ thống công nghệ thông tin quản lý danh tính của chính quyền địa phương nơi mà người sử dụng sinh sống hoặc cơ sở dữ liệu của ngân hàng,...
- Xác thực dựa trên độ uy tín và tin tưởng đối với cộng đồng, các hệ thống tương tác.

Xác thực dựa trên những hệ thống pháp lý

Ưu điểm

Thông tin có độ tin cậy và chính xác cao, cả về mặt pháp lý.Là hướng đi tổng quát để giải quyết vấn đề xác thực danh tính.

Chi phí thu thập dữ liệu cũng thấp hơn vì bộ dữ liệu và cơ sở hạ tầng đã tồn tại

Nhược điểm

Để thực hiện trên phạm vi rộng,có thể là phạm vi toàn cầu,thì vướng phải một số rủi ro:

- Không phải chính phủ nào cũng có cơ sở hạ tầng công nghệ thông tin để quản lý công dân.
- Nếu có cơ sở hạ tầng,thì việc chỉ dựa hoàn toàn vào một hoặc một vài tổ chức cũng gây ra nguy cơ tạo ra một sự sụp đổ dây chuyền nếu hệ thống kết nối gặp sự cố,chưa kể là cần sự đồng ý hợp tác của các cơ quan chính quyền,điều này cũng là trở ngại khá lớn.Bên cạnh đó, nếu IM phụ thuộc hoàn toàn vào các cơ quan chính phủ nghĩa là trong bất cứ giai đoạn nào của dự án, nếu sự hợp tác này (giữa chính phủ và IM) thay đổi thì một phân mảnh của hệ thống sẽ bị ảnh hưởng hoặc bị ngừng hoạt động.
- Còn một mối lo ngại khác là cần bao nhiêu thông tin để thực hiện xác minh? Việc xác thực không chỉ cần CMND hay ID card,phải có đủ lượng thông tin nhất định để xác định một chủ

thể cá nhân,và từng khu vực lại có những loại thông tin công dân có thể giống hoặc khác nhau,hoặc mang cả tính chuyên biệt.

- Luật pháp của mỗi chính quyền mỗi nơi một khác,việc đàm phán để thiết lập một quan hệ với chính quyền cần tuân theo những quy định mỗi nơi,và những yêu cầu, điều khoản hợp tác có thể thay đổi theo từng nơi,khó có thể thiết lập một mô hình chung để xử lý trên toàn bộ hệ thống,hơn nữa,có thể có những yêu cầu mang tính chất rủi ro đối với hệ thống.

Xác thực dựa trên sự đánh giá của các hệ thống kết nối

Mô tả chung

Những hành vi của người sử dụng trên những hệ thống ghép nối trên mô hình sẽ được tổng hợp,sau đó đánh giá độ tin cậy đối về danh tính của người sử dụng và gửi đến hệ thống xử lý.

Dữ liệu sẽ được lưu trữ,đánh giá sau mỗi khoảng thời gian cụ thể, hoặc sau mỗi một giao dịch,hệ thống sẽ xem xét và cho ra những hình thức xử lý hợp lý.

Sau khi kết thúc quy trình,dữ liệu được lưu lại vào kho dữ liệu hệ thống để dùng cho việc đánh giá sau này(nếu có).

Ưu điểm

Triển khai nhanh hơn phương án xác thực đầu tiên,không một bên nào kiểm soát việc xác thực hoàn toàn,nên sẽ không xảy ra tình trạng gián đoạn nếu bên một bên cung cấp việc xác thực gặp sự cố.

Các hệ thống có thể cùng nhau hợp tác để ngăn chặn việc giả mạo thông tin,tạo ra một mô hình cộng tác với quy mô lớn, khi một cá nhân có dấu hiệu nghi vấn ở hệ thống A,thì toàn bộ những hệ thống con khác mà người dùng sử dụng đều nhận được thông báo để có những xử lý thích hợp.

Quyền lực không tập trung hoàn toàn tại IM,mà các hệ thống ghép nối có thể cùng tham gia phát triển,có một số quyền quyết định gián tiếp khi tham gia mô hình.

Nhược điểm

Thông tin xác thực mang tính tương đối,không thể chính xác hoàn toàn, dựa trên phần nhiều về phân tích hành vi người dùng.

Cơ bản việc đánh giá độ tin cậy tài khoản cũng khá trừu tượng,cần môt cơ chế, thước đo chung hoặc cần cả một hệ thống trung gian để thực hiện việc đánh giá sao cho công bằng.

Quá trình xác minh chỉ có thể thực hiện ở pha sau,tức là sau khi người sử dụng đã sử dụng hệ thống một thời gian,hoặc những hành vi,dữ liệu phân tích đã đủ để đánh giá.

Quyền lực trong hệ thống có thể bị tập trung ở một số hệ thống lớn,gây ảnh hưởng nhiều đến sự chính chính xác của thông tin trên toàn bộ hệ thống.

Giải quyết vấn đề

Đánh giá độ tin cậy

Đối với việc phân tích hành vi người dùng,sau mỗi khoảng thời gian, hoặc sau mỗi giao dịch,hệ thống ghép nối sẽ cung cấp dữ liệu hành vi người dùng cho một hệ thống trung gian xử lý và phân tích do IM cung cấp là IUBA. Dưới đây đề cập đến những thuộc tính có thể xét tới trong việc đánh giá.

Dữ liệu hành vi được tương quan và phân tích dựa trên hoạt động trong quá khứ và đang diễn ra.IUBA xác định hành vi bình thường là gì, và những gì cấu thành hoạt động bất thường. Nếu hành vi bất thường của một người được chia sẻ bởi những người khác trong nhóm đồng đẳng của họ, thì nó không còn được coi là rủi ro trung bình hoặc cao.

IUBA thực hiện mô hình phân tích rủi ro. Hành vi bất thường không phải là tự động được coi là một rủi ro. Trước hết nó phải được đánh giá dựa trên tác động tiềm ẩn của nó. Nếu hoạt động bất thường rõ ràng liên quan đến các tài nguyên không nhạy cảm, như thông tin lập lịch trình phòng họp, tác động tiềm ẩn là thấp. Tuy nhiên, các nỗ lực để truy cập các tệp tin nhạy cảm như sở hữu trí tuệ, mang lại một điểm số tác động cao hơn.

Do đó, rủi ro đối với hệ thống được đặt ra bởi một giao dịch cụ thể được xác định bằng các yếu tố: Likelihood, Impact. Trong đó:

- Likelihood liên quan đến xác suất mà hành vi người dùng được đề cập là bất thường. Nó được xác định bởi các thuật toán mô hình hóa hành vi.
- Trong khi đó, tác động (Impact) được dựa trên sự phân loại dựa trên tính nghiêm trọng của thông tin được truy cập, và những gì đã được kiểm soát đối với dữ liệu đó. Những thông tin về tính nghiêm trọng,...sẽ được xác định bởi hệ thống ghép nối.
- Các giao dịch và tính toán rủi ro sau đó có thể được kết hợp với người sử dụng đang thực hiện các giao dịch, để xác định mức độ rủi ro. Việc tính toán rủi ro của người dùng thường bao gồm các yếu tố bổ sung, chẳng hạn như phân loại tài sản, giấy phép, tiềm năng dễ bị tổn thương, chính sách, v.v .. Bất kỳ sự gia tăng nào trong các yếu tố này sẽ làm tăng điểm số rủi ro của người dùng đó.

- Thiết lập giá trị weight có thể được sử dụng cho tất cả các yếu tố trong các tính toán này, để tự động điều chỉnh mô hình tổng thể.
- IUBA thu thập và phân tích hàng trăm thuộc tính, bao gồm thông tin tình huống và thông tin về mối đe dọa bên thứ ba. Kết quả là tập dữ liệu có khối lượng petabyte theo ngữ cảnh. Các thuật toán học máy của IUBA phải được xây dựng để không chỉ loại trừ và loại bỏ các sai lệch và cung cấp dự báo rủi ro mà còn sửa đổi các định mức, dự đoán, và các quy trình đánh giá rủi ro tổng thể dựa trên thông tin thu thập.
- Thay đổi trong phân loại thông tin cũng như thay đổi hoạt động (chẳng hạn như các phòng ban mới, mã số công việc mới hoặc vị trí mới) được tích hợp tự động vào bộ dữ liệu của hệ thống. Ví dụ, nếu một quản trị viên CNTT tạm thời cấp quyền truy cập hệ thống cao hơn, điểm số rủi ro của họ sẽ bị thay đổi trong khoảng thời gian đó.
- IUBA cũng có thể, trong thời gian tự động, xác định những giá trị trọng số tùy chỉnh nào có ý nghĩa hoạt động lớn nhất trong việc giảm sai tích cực.
- Hành vi nguy hiểm ít rõ ràng hơn, như phá hoại, trộm cắp bí mật thương mại của doanh nghiệp, hoặc hoạt động lâu dài hơn như gian lận tài chính, cũng sẽ tạo ra những hành vi bất thường mà hệ thống IUBA có thể phát hiện.

  => IUBA đang chuyển đổi quản lý an ninh và gian lận vì nó cho phép các doanh nghiệp phát hiện khi các tài khoản người dùng / danh tính đã bị kẻ tấn công bên ngoài tấn công hoặc đang bị lạm dụng bởi những người bên trong vì các mục đích độc hại.

** Sau khi được tổng hợp và đánh giá độ tin cậy,đến một thời điểm nhất định,tất cả hệ số độ tin cậy đã được đánh giá từ các hệ thống được tổng hợp lại để cho ra một hệ số tin cậy "định kỳ".Mỗi tài khoản có một hoặc vài chục chỉ số đánh giá (có thể hơn) của những hệ thống mà người dùng sử dụng,cần phải xây dựng một phương trình tổng hợp các kết quả đánh giá,có thể là:*

$$Q = \frac{\sum_{i=1}^{n} w_i v_i}{\sum_{i=1}^{n} \quad w_i}$$

n: Số hệ thống tham gia đánh giá tài khoản

v_i: Chỉ số tin cậy được đánh giá trên hệ thống i

w_i: Trọng số của v_i

* w_i có thể được đánh giá dựa trên các yếu tố của hệ thống i:

- Độ tin cậy của dữ liệu, dựa trên tính an toàn hệ thống,quy mô của hệ thống

- Dữ liệu thuộc về mảng hoạt động nào,vdu: so với giải trí,dữ liệu giao dịch ngân hàng có mức độ quan trọng cao hơn,
- Tần suất dữ liệu,vì mỗi hệ thống có lượng dữ liệu cung cấp để đánh giá có thể khác nhau,vậy nên dữ liệu có tần suất ít thì khó phát hiện những sai phạm hơn.

** Dữ liệu đánh giá độ tin cậy của tài khoản bao trùm cả 2 khía cạnh: sự đánh giá về danh tính của người dùng và đánh giá tài khoản có phải là của người sử dụng độc hại hay không.Điều này cũng giảm được sự tấn công từ bên trong (Insider threat) vốn khó phát hiện và phức tạp.*

Mô hình kết hợp 2 loại hình xác thực

Ưu điểm

Hoàn toàn kế thừa từ phương án 2,và ngoài những hệ thống phi chính phủ hoạt động trong mô hình để góp phần xác thực người dùng,thì hệ thống chính phủ ở một số nơi có đủ điều kiện cũng có thể tham gia vào với vai trò tương đương.Nếu có những hệ thống thông tin của chính phủ tham gia thì càng củng cố sự đúng đắn của mô hình,nhưng nếu điều kiện không cho phép,thì độ tin cậy của mô hình vẫn có thể giữ ở một mức độ chấp nhận được,được công nhận giữa các hệ thống con trong mô hình.

Hệ thống của các chính phủ có thể tham gia vào việc xác thực ở ngay pha đầu tiên,là việc tạo tài khoản người sử dụng.Các hệ thống phi chính phủ tham gia vào việc xác thực trong quá trình sử dụng của người dùng hệ thống.

Xác thực hệ thống, dịch vụ

Mặt khác, để tạo nên một mô hình công bằng, người sử dụng các hệ thống, dịch vụ cũng có khả năng đánh giá tính xác thực, chất lượng,... của chúng. Do vậy, đây sẽ là một hướng tiếp cận trực tiếp đến tập khách hàng sử dụng hệ thống. Trong khi các hệ thống cùng có thể cộng tác với nhau để xác thực tài khoản người dùng, thì cộng đồng người dùng cũng có khả năng đánh giá những dịch vụ mà bản thân được cung cấp, điều này sẽ giúp cải thiện một số yếu tố sau trong mô hình

IV.2. Mô hình kinh doanh

Bổ sung sau

IV.3. Mô hình công nghệ

IV.3.1. Mô hình 1: Xác thực danh tính

C

IV.3.2. Mô hình 2: Cơ chế lưu trữ & Bảo mật

IV.3.2.1. Lưu trữ

Giới thiệu chung

Thông tin được lưu trong mạng của hệ thống được chia lưu trữ dưới 2 dạng:

- Dữ liệu thô,là những dữ liệu đã được chuẩn hóa để lưu trữ của người dùng và được mã hóa giữa hai bên, hệ thống và người sử dụng.Dữ liệu này được lưu trên mạng node IMP (identity management system's public ledger), có thể giải mã nếu có tối thiểu một số lượng bên tham gia (đối với phía hệ thống) và khóa của người dùng (có thể là dữ liệu sinh trắc học).Do đó, hệ thống không thể giải mã dữ liệu người dùng nếu không có sự chấp thuận của người dùng.Dữ liệu trên IMP có thể tham gia việc phục hồi dữ liệu trên IMS(identity management system's private ledger) trong trường hợp dữ liệu không còn đảm bảo tính toàn vẹn.
- Dữ liệu tính toán,những dữ liệu được mã hóa và không thể giải mã.Dữ liệu này được lưu trữ trên mạng node IMS,dùng chủ yếu để tính toán(có ánh xạ đến dữ liệu trên IMP).

Về loại hình thông tin người dùng cung cấp cho hệ thống,có thể là:

- Dữ liệu sinh trắc học, chủ yếu dùng để xác nhận giao dịch đối với hệ thống.
- Dữ liệu ,thông tin cá nhân.
- Dữ liệu về các lớp bảo mật do người dùng tự cài đặt vdu password,câu hỏi xác nhận,...

** Vậy sẽ tồn tại hai lớp mạng để xử lý và lưu trữ thông tin: IMS và IMP.Cả hai lớp mạng đều sử dụng cơ chế DHT (distributed hash-table) làm nền tảng,tuy nhiên có những thay đổi nhất định.*

Cấu trúc của DHT:

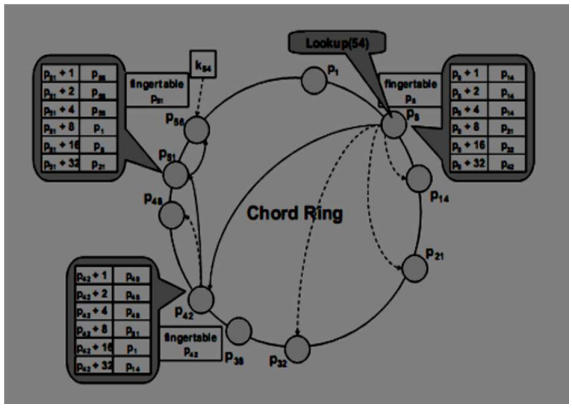
https://voer.edu.vn/c/distributed-hash-table-dht-va-chord/0b8f6b01/397bf60d

https://pdos.csail.mit.edu/~petar/papers/maymounkov-kademlia-lncs.pdf

Một số vấn đề:

- DHT có cơ chế phân một lượng vừa đủ dữ liệu đã hash cho một node,không node nào lưu dữ liệu thừa,điều đó giúp giảm tải công suất xử lý của node,nhưng khi dữ liệu của node bị mất,thì khả năng phục hồi chậm và cần sự can thiệp của người sử dụng (trong trường hợp phục hồi trên IMP), hơn nữa việc lưu trữ này cũng thiếu đi tính sẵn sàng của dữ liệu.

- Việc lưu trữ hiện tại là hoàn toàn ngẫu nhiên,dữ liệu tài khoản của người dùng có thể do bất kỳ node nào của hệ thống nắm giữ,chỉ phụ thuộc vào cơ chế hashing.Nên có thể nói là DHT coi tất cả dữ liệu được lưu trữ có độ ưu tiên ngang hàng nhau,và điều đó không đúng trong thực tế.



Phương án giải quyết

*** Cần giải quyết những vấn đề**

- Chi phí phục hồi: Khi bị mất dữ liệu tại một node,dữ liệu cần được phục hồi với chi phí tối thiểu.
- Tính nhất quán: Khi dữ liệu được thêm,cập nhật,xóa,cần cập nhật trên các node một cách nhanh,chính xác và tiết kiệm chi phí nhất.Cần đảm bảo tính thống nhất của các bản sao.
- Nơi lưu trữ: Giữa dữ liệu cần lưu trữ và nơi cần lưu trữ cần có một sự ràng buộc nhất định,để tăng hiệu suất xử lý và sự tin cậy của thuật toán.
- Số lượng bản sao: Số lượng bản sao dữ liệu có thể bị cố định,hoặc mở rộng dựa trên một vài yếu tố như sự nhạy cảm.
- Cân bằng tải: Một số mảng nội dung được thao tác với tần suất lớn sẽ ảnh hưởng đến hiệu năng của node.
- Độ nhạy cảm thông tin: Một số thông tin cần sự bảo mật và tính chính xác cao,vậy nên không thể lưu trữ và thao tác tùy tiện trên những node bất kỳ.Ví dụ những node có băng thông nhỏ,hiệu suất nhỏ,an toàn thấp,...
- Sự mở rộng của không gian lưu trữ(cực kỳ hiếm khi xảy ra): Việc mở rộng cần một cơ chế giúp giảm thiểu lượng thao tác phát sinh.Sự mở rộng có thể xuất hiện khi không gian tối đa dùng để hash dữ liệu chạm ngưỡng tối đa.Về cơ bản nếu không gian key đã đạt đến tối đa,nhiều ưu điểm của hashing sẽ bị hạn chế.Trường hợp này hiếm khi xảy ra,nhưng nếu xảy ra thì chi phí xử lý lớn.Cách giải quyết cơ bản là tăng không gian lưu trữ key bằng tăng số bit tối đa của một key,nhưng sẽ phải chuyển đổi dữ liệu trên toàn mạng lưới,sẽ mất một chi phí lớn mà chưa chắc đã đảm bảo tính đúng đắn dữ liệu.

Các cải tiến

1) Vấn đề lưu trữ

1.1) ID của node

Đầu tiên,như cơ chế của DHT,mỗi node cần một ID để định danh,liên lạc trên mạng lưới,vậy nên cần có một phương pháp thống nhất để tạo ra ID của node.Một cách đơn giản và đáng tin cậy là ID của node sẽ được sinh ra bằng cách hash địa chỉ IP của server,điều này vừa đảm bảo sự ngẫu nhiên và tính duy nhất của ID.

1.2) Phương pháp lưu trữ

Phương pháp lưu trữ của DHT có nhiều điểm yếu về tính sẵn sàng và khả năng phục hồi,vậy nên ta sẽ xem xét một số phương pháp khác để giải quyết những vấn đề này và mô tả khái quát từng phương pháp:

- Soft-State relication
- Successor Replication (Successor là những node nằm ở chiều kim đồng hồ của node đang được xét tới)
- Dynamic Replication

1.2.1) Soft-State replication

Soft-State duy trì những bản sao của dữ liệu gốc bằng việc thêm lại dữ liệu vào mạng lưới một cách định kỳ.Các dữ liệu trong mạng có một lượng thời gian duy trì nhất định và bị xóa sau khi "quá hạn",phiên bản mới của dữ liệu đã bị xóa sau đó sẽ được thêm lại.Vì các thuật toán bảo trì và cập nhật không đòi hỏi tính địa chỉ của bản sao, chúng ta có thể đặt bản sao ở bất cứ đâu theo yêu cầu. Điều này cung cấp cơ hội để khai thác mạng cục bộ và sử dụng bộ nhớ đệm để cung cấp tra cứu dữ liệu nhanh.

Tuy nhiên phương pháp Soft-State yêu cầu cả một hệ thống lưu trữ đáng tin cậy bên ngoài của DHT, và rất nhiều truyền thông mạng để nạp lại các dữ liệu trên các nút chính xác. Bởi vì bảo trì tốn kém, cần thời gian chờ lâu, nên không thể thực hiện thường xuyên. Điều này hạn chế tính hữu dụng thực tế của việc sao chép soft-state trong việc cung cấp lưu trữ có thể thay đổi.

Phương pháp lưu trữ Soft-State có thể hữu ích hơn khi được sử dụng kết hợp với một thuật toán khác. Việc định nghĩa giới hạn cho việc duy trì dữ liệu có thể hữu ích cho việc thu gom rác, trong khi một số kỹ thuật khác cung cấp độ tin cậy.

1.2.2) Successor Replication

Phương pháp duy trì các bản sao dữ liệu trên một lượng r successors của node đang xét.Nếu một node offline khỏi hệ thống,những node successor của nó sẽ có trách nhiệm duy trì dữ liệu mà nó đang lưu trữ.Dữ liệu sẽ mất hoàn toàn chỉ khi toàn bộ node successor này đều bị hỏng.

Các node kết nối mới sẽ kế thừa một khoảng khóa từ node nó đứng trước, và có thể được gửi dữ liệu mà họ chịu trách nhiệm trước khi tham gia hoặc chuyển tiếp các yêu cầu tới những người kế nhiệm cho đến khi thuật toán bảo trì tiếp theo chạy.

Phương pháp này sử dụng thuật toán duy trì DHASH chạy hai giao thức để ngăn chặn số lượng của các bản sao quá thấp hoặc quá cao:

Local Maintenance: Một node gửi một thông điệp đồng bộ bao gồm vùng không gian lưu trữ của nó đến r successor của nó. Những node này sau đó đồng bộ hóa những nội dung của cơ sử dữ liệu của chúng để key range được lưu cả ở node ban đầu lẫn những node successor.

Global Maintenance: Một node được kiểm tra một cách định kỳ cơ sở dữ liệu key của nó để phát hiện nếu bất kỳ key nào nó không còn được phép duy trì. Để làm điều này,nó tìm kiếm chủ sở hữu của mỗi key mà nó lưu trữ,sau đó kiểm tra danh sách successor của chủ sở hữu đó(Thực chất là kiểm tra key range mà nó đang lưu trữ,chứ không phải kiểm tra từng key).Nếu node đó không có trong danh sách r successor đầu tiên thì xóa những dữ liệu key không cần thiết.

Thuật toán bảo trì cục bộ có thể mở rộng vì nó chỉ giao tiếp với các nút lân cận. Merkle Tree băm có thể đảm bảo rằng việc đồng bộ hoá là một hoạt động hiệu quả, và chỉ các mục dữ liệu cần thiết được gửi đi. Thuật toán được trình bày làm giảm số lượng bản sao bị thiếu mỗi lần chạy, nhưng không sửa tất cả. Nếu tất cả các bản sao tồn tại trên tất cả các r successors sau một lần bảo trì, thuật toán **Local Maintenance** phải chạy hai lần, lần đầu tiên thu thập toàn bộ key range vào node gốc, lần thứ hai phân phối những thứ trên tất cả các successor.

Nếu các node mới tham gia vào khi mà thuật toán **Global Maintenance** vừa chạy lần cuối,một vấn đề có thể nảy sinh,các nút mới sẽ có nghĩa là một số nút cũ không còn trong vòng r của chủ sở hữu của tất cả các bản sao nó lưu trữ. Điều này có nghĩa là nó sẽ không được cập nhật.

Khoảng lưu trữ và độ tin cậy

Việc duy trì một hệ thống ổn định đòi hỏi phải có sự cân bằng giữa sử dụng băng thông, hay làm mới soft-state, và không gian lưu trữ được sử dụng bởi bản sao.

Đối với một dữ liệu bị mất khỏi hệ thống,điều đó tức là tất cả r successor lưu trữ dữ liệu này đều bị mất dữ liệu.Nếu một bản sao của dữ liệu biến mất khỏi hệ thống với xác suất p,thì xác suất để dữ liệu bị biến mất sẽ là p^r. Xác suất nếu một node và tất cả r successor của nó cùng bị mất dữ liệu ở một vài khoảng lưu trữ là một hàm xác suất của một bản sao dữ liệu không được lưu trên một node như dự định, p, kích thước mạng N,và số lượng bản sao r – $R_p(r, N)$.

Với N lớn,có thể coi chord ring như một đường thẳng ,và hàm $R_p(r, N)$ trở nên tương đương với *Run Problem,* xác suất đạt được r hoặc nhiều kết quả liên tiếp trong các thử nghiệm n bernoulli. Giải pháp tổng quát $R_p(r, N)$ được đưa ra dưới dạng:

$$F_p(r,s) = \frac{p^r s^r (1-ps)}{1-s+(1-p)p^r s^{r+1}} \equiv \sum_{i=r}^{\infty} c_i^p s^i$$

$$R_p(r,n) = \sum_{i=r}^{n} c_i^p$$

Để làm rõ đến khoảng bảo trì này, chúng ta cần phải xác định p. Chúng ta sẽ thực hiện điều này theo số lượng các lần bảo trì trong một *half life*, S. Một *half life* của hệ thống là thời gian tối thiểu cho N / 2 của các nút rời đi, và thời gian cần cho các nút N / 2 tham gia. Chúng ta sẽ chỉ xem xét các hệ thống trạng thái ổn định, nơi các nút tham gia và rời hệ thống với cùng tỷ lệ.Với giả định p = 1/2S, xác suất của việc mất dữ liệu trong một *half life* cho một hệ thống N node với r bản sao và S lần bảo trì mỗi *half-life*, được gọi là FAIL(N,r,S), với

$$FAIL(N,r,S) = SR_{\frac{1}{2S}}(r,N)$$

1.2.3) Dynamic Replication

Chiến lược Replica Enumeration dựa trên một hàm phân bổ, h(m, d). Đối với mỗi tài liệu có ID d, các bản sao được đặt tại địa chỉ xác định bởi h(m, d) trong đó m ≥ 1 là chỉ số của trường hợp đó. Chức năng phân bổ là để giả ngẫu nhiên, sao cho bản sao được phân bố đều trên không gian địa chỉ.

Số lượng bản sao biến đổi trong một dải cố định $1 \leq r \leq R_{MAX}$, cho phép sao chép lớn hơn cho các mục yêu cầu lớn hơn. Cơ chế được sử dụng để xác định chính xác số bản sao không được chỉ định, nhưng có thể được thiết kế để thích ứng với độ tin cậy của mạng và tải trên các nút cung cấp mỗi tài liệu.

Một số quy tắc được áp dụng:

1) Các bản sao của một đối tượng d chỉ được đặt ở các địa chỉ được cung cấp bởi hàm h(m,d).
2) Đối với tài liệu d bất kỳ trong hệ thống,luôn luôn tồn tại một bản sang ở h(1, d).
3) Bất kỳ bản sao nào với thứ tự > 1 chỉ có thể tồn tại nếu đang tồn tại một bản sao với thứ tự m – 1.
4) Không có nhiều hơn R_{MAX} bản sao.

Dynamic Replication có thể làm giảm bớt nút cổ chai tìm kiếm có thể ảnh hưởng đến Successor Replication. Successor Replication yêu cầu tất cả các tra cứu cho một chìa khóa được hướng đến chủ sở hữu của chìa khóa đó. Với chức năng phân bổ thích hợp, Dynamic Replication có thể đặt các bản sao ở các vị trí đồng đều.

1.2.3.1) Thuật toán tìm kiếm: Sử dụng đệ quy để lấy khóa ở location = h(m, key) với m bất kỳ

Algorithm 1 Recursive get for *key* at $location = h(m, key)$ for some m

```
if self.keyrange.containsReplica(key) and self.store.contains(key) then
  return (self.store.get(key))
end if
if self.keyrange.contains(location) then
  return NULL
end if
next = self.closestPrecedingNode^2(location)
return next.recursiveGet(key, location)
```

1.2.3.2) Thuật toán duy trì trong Dynamic replication

Khi một node mới tham gia,hoặc node tiền nhiệm của nó bị lỗi,nó sẽ đáp ứng việc tra cứu dữ liệu trong phạm vi nó vừa được thừa kế từ người tiền nhiệm với một thông điệp UNKNOWN đặc biệt, cho biết nút truy vấn tìm nơi khác để lấy dữ liệu hoặc quay lại và thử lại. Node sau đó cố gắng để phục hồi bằng việc tìm kiếm h(m + 1, d) để thay thế dữ liệu.

Một số định nghĩa sẽ sử dụng trong thuật toán:

1) Node tương ứng với h(1, d) là người sở hữu của d
2) "Nhóm bản sao" của một phần tử d là những node $\{h(m, d) : 1 \leq m \leq R_{MAX}\}$.
3) "Nhóm cơ bản" của một phần tử d là một tập bản sao được giữ ở các node $\{h(m, d) : 1 \leq m \leq R_M \text{ IN }\}$
4) "Nhóm ngoại vi" là những nhóm $\{h(m, d) : R_{MIN} < m < R_{MAX}\}$.

Vì chúng ta không thể dựa vào một node duy nhất h(1,d) luôn sẵn sàng,vậy nên chúng ta đảm bảo sự tồn tại của bản sao nằm trong một node trong "Nhóm cơ bản". Xác suất chính xác sẽ dựa vào tần suất duy trì.

- Các bản sao của một đối tượng d chỉ có thể được lấy ra từ các địa chỉ h(m,d).
- Đối với tài liệu d bất kỳ trong hệ thống, xác suất tồn tại một bản sao trong "Nhóm cơ bản" sẽ cao hơn.
- Với bất kỳ bản sao ngoại vi ($m > R_{MIN}$) chỉ có thể tồn tại cho một khoảng duy trì nếu không có bản sao nào đang tồn tại ở node m – 1.
- Không một đối tượng nào có thể được tồn tại ở nhiều hơn R_{MAX} node.

Bây giờ chúng ta có thể tạo ra một thuật toán bảo trì duy trì các biến bất biến này khi đối mặt với node rời đi.

Core Maintenance (Duy trì trong nhóm cơ bản) Chủ sở hữu của dữ liệu, node A, tính toán và tìm kiếm những node trong nhóm cơ bản khoảng dữ liệu mà node đó đang duy trì. Nó đồng bộ hóa cơ sở dữ liệu với mỗi người giữ bản sao trong phạm vi khóa mà các node cùng chịu trách nhiệm.

Peripheral Maintenance (Duy trì trong nhóm ngoại vi) Một node lưu trữ một bản sao ngoại vi của một đối dượng d với chỉ số $i > R_{MIN}$ phải kiểm tra rằng một bản sao của d cũng được lưu ở node với chỉ số i – 1

Điều này có thể được thực hiện bằng cách tìm kiếm h (i-1, d) và yêu cầu một Bloom filter tóm tắt các bản sao được giữ bởi chỉ số i-1 từ node đó, chúng ta lưu trữ trong suốt thời gian bảo trì. Bất kỳ mục nào không tìm thấy trong Bloom filter sẽ bị xóa.

Global Maintenance (Duy trì toàn cục) Mỗi node tính toán trên những nhóm cơ bản đối với một đối tượng mà nó nắm giữ.Bất kỳ những đối tượng không còn được quyền lưu trữ ở trong node đều cung cấp cho nhóm cơ bản, sau đó xóa đi dữ liệu đó.

Để thuật toán này tăng độ tin cậy, điều quan trọng là các dữ liệu của một node trong phạm vi h (m, d) thuộc sở hữu của các node khác nhau. Nếu nhiều bản sao ánh xạ đến không gian quyền sở hữu của một nút duy nhất, sự va chạm phân bổ thì tất cả các bản sao này sẽ bị mất khi một node đó không thành công.Để ngăn chặn điều này, thuật toán **Core Maintenance** phải kiểm tra liệu nó đã ánh xạ trước key range này vào node này hay không. Nếu sự giao thoa xảy ra, thuật toán sẽ kích hợp việc tạo ra node giữ bản sao ngoại vi.

Xác suất của của việc nhiều bản sao cùng ánh xạ đến cùng một node có thể giảm thiểu bởi việc sử dụng những hiểu biết về kích thước mạng lưới (và cả kích thước keyspace trung bình) trong hàm phân bổ

1.2.3.3) Thuật toán lấy dữ liệu

Thuật toán cần chọn những chỉ số nào cần để tìm kiếm và thứ tự của chúng

Algorithm 2 Dynamic Fetch for *key*

```
indexes ← [0 ... R_MAX]
item ← NULL
while ¬item do
  index ← indexes.popRandom()
  item ← recursiveGet(key, h(index, key))
  if ¬item and index > R_MIN then
    indexes.removeRange([index, R_MAX])
  end if
  if indexes=[] then
    indexes ← [0 ... R_MAX]
  end if
end while
```

Nếu việc bảo trì không thường xuyên, việc loại bỏ toàn bộ các bản sao thiết bị ngoại vi với các chỉ mục cao hơn nếu một bản sao ngoại vi được tìm thấy có sản phẩm nào có thể dẫn đến hiệu suất kém, và chỉ cần lấy ra bản sao được biết là trống rỗng.

1.2.3.4) Hàm cấp phát

1.2.3.4.1) Successor Allocation

$$h(m, d) = (d + (m \cdot avgKeyspace)) \bmod MaxID$$

Trong đó

- MaxID là số lượng key có thể sinh ra
- avgKeyspace là số lượng key trong bình được sở hữu bởi một node, giá trị này hoặc được ước lượng trong thời gian chạy hoặc được cung cấp bởi người dùng.

1.2.3.4.2) Predecessor Allocation

$$h(m, d) = (d - (m \cdot avgKeyspace)) \bmod MaxID$$

1.2.3.4.3) Finger Allocation

$$\delta = \log2(avgKeyspace)$$

$$h(m, d) = (d + 2(m+\delta)) \bmod MaxID$$

1.2.3.4.4) Block Allocation (Được khuyến khích)

$$h(m, d) = (d - (d \bmod blockSize) + (d \bmod avgKeyspace) + (m * avgKeyspace)) \bmod MaxID$$

Chức năng phân bổ này cố gắng làm cho các nhóm core replica chồng lên nhau hoàn toàn với các nhóm core replica khác.Phương pháp chia không gian khóa thành n/r phần có kích thước bằng nhau,là các block.Tất cả các node trong một block lưu tất cả đối tượng trong block đó.

Vậy từ những phân tích trên,hệ thống sẽ lựa chọn Dynamic Replication và Block Allocation để triển khai

1.3) Gia tăng không gian lưu trữ

* Đối với việc gia tăng không gian lưu trữ,phương pháp giải quyết:

Tăng thêm "chiều" hash dữ liệu bằng cách hash thêm một số lần nữa dựa trên một số Di được sử dụng tượng trưng có kích cỡ tối đa của "chiều i",vdu chiều thứ hai có thể tượng trưng cho số mạng lưới con có trong hệ thống.

* Cải tiến này có thể áp dụng cho một số việc:

- Phân chia mạng lưới tổng thành các mạng lưới con,với mỗi mạng lưới con là một lớp mới của dữ liệu.Mạng lưới mới có thể có một số tính chất chuyên biệt,ví dụ nếu thông tin nhạy cảm cần lưu trữ với mức độ bảo mật cao hoặc cần nhiều bản sao trên một mạng lưới chuyên biệt,thì sẽ thêm một số tham số vào key để xác định độ nhạy cảm.Hoặc nếu muốn phân bố lượng thông tin dữ liệu của một tập người dùng ở một cơ sở hạ tầng công nghệ thông tin nhất định,ví dụ quốc gia nào thì sở hữu danh tính của công dân có quốc tịch tương ứng.Có thể nói việc này có thể giúp cho việc phân cấp,phân tán dữ liệu một cách có chủ đích.

1.4) Định tuyến

Đối với việc định tuyến trong mạng lưới,mỗi node sẽ lưu ID/địa chỉ của hai loại node:

1) Những node có chung phân vùng lưu trữ với node đang xét, để thuận tiện cho những trường hợp cập nhật hay phục hồi dữ liệu.
2) Những finger node được xác định bằng địa chỉ của node đang xét + một số bằng lũy thừa của 2, độ phức tạp tìm kiếm sẽ là O(logN).

Những giao thức xử lý chi tiết sẽ được đề cập ở phần **Giao thức mạng**

1.4.1) Ưu điểm:

- Việc duy trì sự thống nhất dữ liệu trên các node có phân vùng dữ liệu chung sẽ dễ dàng hơn,một khi một node A thay đổi dữ liệu trên phân vùng B,thì các node A_k cũng nhận được tín hiệu thay đổi theo,mà không phải tổ chức tìm kiếm quy mô trên mạng node.Do đó giảm tải băng thông và chi phí xử lý.
- Khi có một node tham gia vào phân vùng,sẽ dễ dàng xác định những node có điều kiện thuận lợi nhất để thực hiện việc chuyển giao các phân vùng xác định.Vdu những node có công suất cao sẽ được ưu tiên xử lý nhiệm vụ này.

2) Cải tiến 2: Cân bằng tải

Cơ chế Consistent Hashing có một điểm yếu là cân bằng tải.Việc coi mọi node có trách nhiệm như nhau sẽ dẫn đến sự không cân đối việc xử lý yêu cầu,sẽ có một số node bị quá tải công suất,thông lượng,gây nghẽn mạng,hoặc một số node có thời gian rỗi quá nhiều.Xử lý vấn đề này có thể sử dụng thuật toán Bounded-Load,ý tưởng cơ bản của thuật toán là:

1) Xác định một yếu tố cân bằng, c > 1 (c = 1 + ε). c kiểm soát sự mất cân bằng giữa các máy chủ. Ví dụ: nếu c = 1,25, không máy chủ nào sẽ nhận được nhiều hơn 125% tải trung bình. Nếu c tăng lên ∞, thuật toán trở nên tương đương với consistent hashing, khi c giảm xuống gần 1, việc băm trở nên ít quan trọng hơn.
2) Khi yêu cầu đến, tính tải trung bình (số yêu cầu chưa hoàn thành, m, bao gồm cả yêu cầu vừa đến, chia cho số máy chủ có sẵn, n). Nhân tải trung bình với c để có được một "mục tiêu tải", t. Trong tài liệu gốc, công suất được gán cho các máy chủ sao cho mỗi máy chủ có công suất □t□ hoặc □t□, và tổng công suất là □cm□. Do đó công suất tối đa của một máy chủ là □cm / n□, lớn hơn c lần tải trung bình ít hơn 1 yêu cầu. Để hỗ trợ cho các máy chủ 'trọng lượng' khác nhau, như HAProxy, thuật toán phải thay đổi một chút, nhưng tinh thần là như nhau - không có máy chủ nào có thể vượt quá tỷ lệ tải trọng của nó nhiều hơn 1 yêu cầu.
3) Để gửi yêu cầu, vẫn thực hiện việc tính toán hàm băm và lựa chọn máy chủ vật lý gần nhất. Nếu máy chủ đó đang hoạt động dưới công suất tối đa thì gán yêu cầu cho máy chủ đó. Nếu không, yêu cầu được chuyển tới máy chủ tiếp theo trong vòng băm và tiếp tục việc kiểm tra, tiếp tục cho đến khi tìm thấy một máy chủ đủ đáp ứng yêu cầu. Thuật toán có một số ưu điểm:
    - Không có máy chủ nào được phép quá tải bởi hơn một yếu tố c cộng 1 yêu cầu.
    - Việc phân phối các yêu cầu giống cơ chế consistent hashing nếu máy chủ không bị quá tải.
    - Nếu máy chủ bị quá tải, danh sách các máy chủ dự phòng sẽ giống nhau cho cùng một yêu cầu - tức là cùng một máy chủ sẽ luôn là "sự lựa chọn thứ hai" cho một nội dung phổ biến. Điều này là tốt cho bộ nhớ đệm.
    - Nếu máy chủ quá tải, danh sách các máy chủ dự phòng thường sẽ khác nhau cho các yêu cầu khác nhau - tức là tải của máy chủ sẽ được phân phối giữa các máy chủ có sẵn thay vì tất cả hạ cánh trên một máy chủ. Điều này phụ thuộc vào mỗi máy chủ được gán nhiều điểm trong vòng băm hợp nhất.

Chi tiết thuật toán Bounded-Load: https://arxiv.org/pdf/1608.01350.pdf

Thuật toán cân bằng tải sẽ kết hợp tốt với thuật toán lưu trữ Dynamic Replication để đạt được hiệu suất tìm kiếm và xử lý yêu cầu hiệu quả.

3) Bảo mật và xác thực trao đổi trong mạng lưới

Mục tiêu áp dụng

Hệ thống có hai nền tảng lưu trữ, là IMP và IMS, trong đó IMS với vai trò là off-chain của hệ thống,chịu trách nhiệm về private part và những tính toán xử lý trên chúng. Off-chain được tạo ra bằng cách thuê các máy chủ ngoài để sử dụng trong việc tính toán. Trong quá trình tính toán, những máy chủ này có thể cố ý hoặc vô ý lưu lại dữ liệu được gửi đến hoặc dữ liệu đã qua tính toán,gây nên sơ hở bảo mật,do vậy cần một phương pháp để hạn chế rủi ro này.

Cơ sở công nghệ

Về vấn đề tính toán an toàn trên mạng lưới (SMPC), đã có nhiều mô hình sinh ra để giải quyết,tiêu biểu là:

- Circuit Garbling, với nhiều công trình nghiên cứu của Yao.Phương pháp sử dụng việc mã hóa và giải mã các khóa theo một thứ tự xác định để mô phỏng việc đánh giá mạch.
- Secret sharing, đây là phương pháp giúp chia sẻ một thông tin bí mật giữa một nhóm các server,mỗi server được cung cấp một phần thông tin của bí mật chung.Thông điệp bí mật có thể được xây dựng lại chỉ khi có một số lượng đủ lớn của lượng thông tin đã phân mảnh để cùng kết hợp. Trong số đó những giao thức đã được xây dựng,có một số giao thức được tin dùng là SPDZ,phiên bản cải thiện của nó là SPDZ2 và Mascot,hai phiên bản này giải quyết về một số vấn đề trong giai đoạn offline,và phần online hầu hết được giữ nguyên so với phiên bản cha,do vậy hệ thống sẽ sử dụng SPDZ cho giai đoạn online và Mascot cho giai đoạn offline.

** online là giai đoạn đánh giá mạch,offline là giai đoạn khởi tạo dữ liệu tiền xử lý và phân phát dữ liệu khởi tạo đến các bên tham gia việc tính toán mạch.*

2.2.1) SPDZ

The Preprocessing Model

Mô hình tiền xử lý là một nền tảng mà trong đó giả định có tồn tại một 'trusted dealer' sẽ phân phát các dữ liệu ban đầu cho các tổ chức trước khi bất kỳ sự đánh giá mạch được thực hiện, giúp cho việc đánh giá mạch trở nên hiệu quả hơn vì các tổ chức có được dữ liệu sẵn sàng để xử lý.

SPDZ(Speedz): SPDZ sử dụng việc chia sẻ thành phần bí mật trên một không gian F,kết hợp với lý thuyết MACs(message authentication code) để đảm bảo an toàn. Một giá trị bí mật x □ F được đại diện bởi:

$$[\![x]\!] = (x^{(1)}, \dots x^{(n)}, m^{(1)}, \dots, m^{(n)}, \Delta^{(1)}, \dots, \Delta^{(n)})$$

Ở đây mỗi tổ chức P_i giữ giá trị chia sẻ ngẫu nhiên $x^{(i)}$, MAC ngẫu nhiên được chia sẻ $m^{(i)}$ và MAC cố định chia sẻ $\Delta^{(i)}$, như vậy quan hệ MAC là$m = x \cdot \Delta$, trong đó

$$x = \sum_i \quad x^{(i)}, m = \sum_i \quad m^{(i)}, \Delta = \sum_i \quad \Delta^{(i)}$$

trên F

Khi tính toán một giá trị chia sẻ $[\![x]\!]$, các tổ chứa đầu tiên truyền đi sự giá trị chia sẻ của họ $x^{(i)}$ và tính toán x.Để đảm bảo x là chính xác,họ sau đó kiểm tra MAC bằng việc cam kết và tính toán $m^{(i)} = x \cdot \Delta^{(i)}$và kiểm tra tổng tất cả chia sẻ có tiến tới 0 hay không.Để tăng hiệu quả khi tính toán nhiều giá trị,mội kết hợp tuyến tính ngẫu nhiên của MACs có thể được kiểm tra thay thế.

Nhiệm vụ chính của giai đoạn tiền xử lý của SPDZ là để tạo ra những loại giá trị chia sẻ ngẫu nhiên ,được xác thực sau đây:

Input P_i: $([\![r]\!], i)$một giá trị ngẫu nhiên được chia sẻ r, chỉ P_i mới biết giá trị r
Tripple: $([\![a]\!], [\![b]\!], [\![c]\!])$cho ngẫu nhiên đều a,b,với c = ab

Trong giai đoạn offline, các bên giao tiếp và sử dụng các giá trị Input để tạo ra các đại giện được chia sẻ của những thông tin bảo mật của họ, và các giá trị Tripple để thực thi các phép nhân trên các giá trị được chia sẻ bí mật.Chú ý rằng thể hiện $[\![.]\!]$là tuyến tính,các phép cộng và các hàm tuyến tính có thể được tính toán cục bộ

2.2.2) Mascot

Mục đích

Mascot tạo ra phép nhân bộ ba, chủ yếu là các chia sẻ bí mât của các nhóm (a, b, a· b, a·Δ, b ·Δ, a· b ·Δ), a,b là các giá trị ngẫu nhiên và Δ là một khóa MAC ngẫu nhiên trên toàn cục được chia sẻ bí mật. Thông tin phân mảnh a,b và Δ có thể được tạo ra bởi mọi tổ chức sau khi mỗi tổ chức chọn lượng thông tin phân mảnh mà nó sẽ nắm giữ. .

Chiến lược

Để có được một giao thức bảo mật tốt,mascot sử dụng hai chiến lược khác nhau: một để đảm bảo tính đúng đắn của kết quả trong sự khởi tạo MAC, và một để đảm bảo tính đúng đắn và bảo mật của phép nhân bộ ba.

Đối với việc khởi tạo MAC, Mascot chỉ ra rằng các giao thức bảo mật thụ động là gần như đủ,chúng ta chỉ cần kiểm tra sự kết hợp tuyến tính ngẫu nhiên của MAC ngay sau khi tạo ra và cũng như các giá trị mở sau đó. Mascot mô hình hóa sự khởi tạo MAC và những yêu cầu mở trong một tính năng chuyên biệt $F_{[.]}$được coi là một sự tổng hợp của chia sẻ bí mật có thể kiểm chứng được đối với trường hợp ngưỡng đầy đủ (full-threshold) không chính xác.

Chi tiết triển khai

Giao thức bắt đầu bằng việc khởi tạo các thành phần a,b,c trong vector (a,b,c), trong đó b □ F và a, c □ F^{τ} với τ bất kỳ,sử dụng giao thức COPE, một sự mở rộng của OT,chi tiết về giao thức sẽ được mô tả trong tài liệu.Dưới đây là ví dụ minh họa của giao thức này:

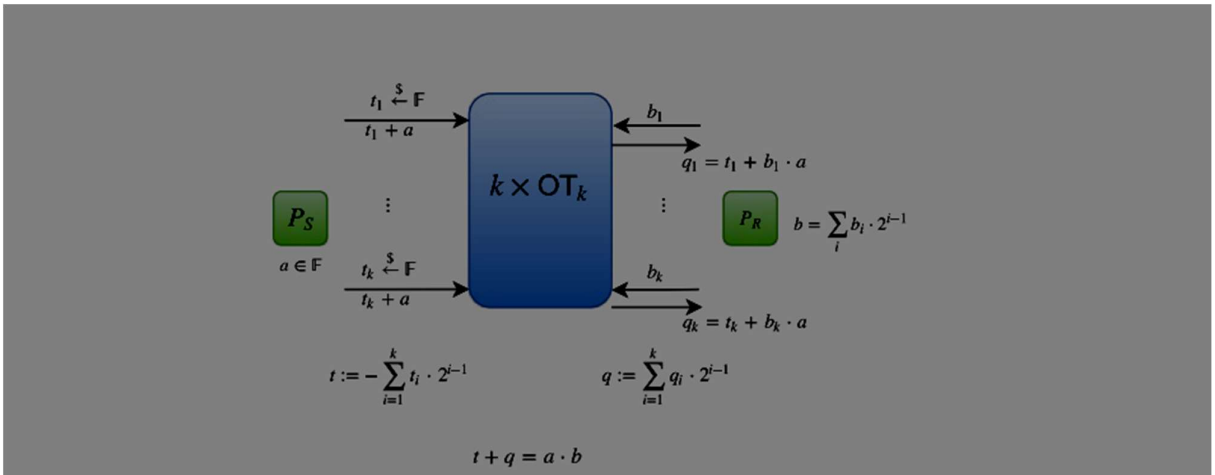


Figure 2: Two-party secret-shared multiplication in $\mathbb{F}_p$ using 1-out-of-2 OT

OT(Oblivious Transfer): *Một giao thức giữa người gửi và nhận,ở đó người gửi chuyển một trong vài thông điệp đến người nhận, trong khi không hề biết tin nhắn nào đã được gửi.*

Các bên tạo ra một mẫu vector ngẫu nhiên công khai r □ F^{τ} và có được bộ ba (a,b,c) bằng định nghĩa

$$a = \langle a, r \rangle, c = \langle c, r \rangle$$

Một cách trực quan,bất kỳ bit nào của a bị lộ đều được kết hợp một cách ngẫu nhiên với những bit không bị lộ,vì thế giá trị cuối cùng a xuất hiện một cách ngẫu nhiên đều.Trong sự chứng minh bảo mật,trình mô phỏng có thể xác định một cách chính xác bất kỳ sự rò rỉ thông tin nào của a và giới hạn min-entropy của nó bằng cách phân tích các dữ liệu đầu vào của kẻ phá hoại với OTs.Vì sản phẩm

bên trong định nghĩa hàm băm phổ quát, chúng ta có thể sử dụng hash lemma còn sót lại để chỉ ra rằng a là ngẫu nhiên thống nhất khi τ đủ lớn.

Tại thời điểm này, chúng ta có thể lặp lại quá trình để có được thêm một bộ ba, sau đó chứng thực cả hai cặp bộ ba và kiểm tra tính chính xác với một sự "hy sinh". Tuy nhiên, chúng ta thấy rằng giai đoạn này có thể được tối ưu hoá bằng cách sử dụng vector bộ ba (a, b, c) ban đầu để có được một thứ bộ ba thứ hai tương quan, với cùng một giá trị b và chi phí thấp hơn. Để làm điều này, mascot tạo một vector ngẫu nhiên$\hat{r}$ và tính $\hat{a}$, $\hat{c}$tương ứng. Một lần nữa, chúng ta có thể chỉ ra (cho τ thích hợp) rằng $\hat{a}$là ngẫu nhiên đều và độc lập với a. Sau đó chúng ta có thể sử dụng $(\hat{a}, \hat{b}, \hat{c})$để kiểm tra tính đúng đắn của (a, b, c) như sau. Sau khi thêm MACs vào cả hai bộ ba, các bên lấy mẫu một giá trị ngẫu nhiên s □ F và mở ρ = $s \cdot [\![a]\!] - [\![\hat{a}]\!]$, trong đó $[\![.]\!]$biểu thị sơ đồ chia sẻ bí mật được chứng thực tuyến tính. Bây giờ chúng tôi có:

$$s \cdot [\![c]\!] - [\![\hat{c}]\!] - [\![b]\!] \cdot \rho = [\![s \cdot (c - a \cdot b) + (\hat{a} \cdot b - \hat{c})]\!]$$

Biểu thức vế trái là tuyến tính trong những giá trị được chia sẻ,các bên có thể tính toán nó và kiểm tra rằng nó tiến tới 0.Nếu một hoặc cả hai bộ ba đều không chính xác thì nó khác 0 với xác suất gần bằng$1/|F|$, vì s là ngẫu nhiên đều và không rõ tại thời điểm xác minh.

Nó chỉ ra rằng đối với phương pháp tối ưu hóa này,bằng cách sử dụng τ = 4 là đủ để đưa một bộ ba chính xác và đảm bảo một lợi thế để phân biệt trong $O(1/|F|)$.Nếu chúng ta cho phép nó đạt đến$O(1/\sqrt{|F|})$ thì có thể có τ = 3. Cụ thể điều này có nghĩa có thể sử dụng τ = 3 cho miền >= 128 bit với 64 bit dùng cho việc thống kê bảo mật.

2.2.2.4) Lộ trình, mối quan hệ phụ thuộc giữa các chức năng

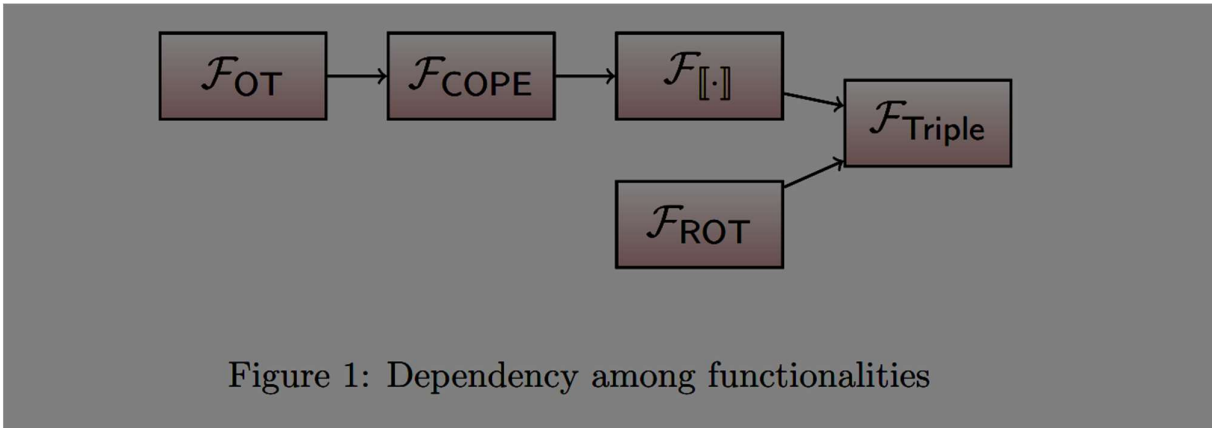


Figure 1: Dependency among functionalities

F_{OT}: mô hình hóa của OT(oblivious transfer)

F_{ROT}: mô hình hóa của OT ngẫu nhiên (random oblivious transfer)

Phần giới thiệu chỉ mô tả sơ lược về Mascot,những thông tin chi tiết sẽ được đề cập trong tài liệu: https://eprint.iacr.org/2016/505.pdf

Tài liệu về SPDZ: https://eprint.iacr.org/2017/1230.pdf