

Tận dụng Blockchain cho Lưu trữ và Chia sẻ Dữ liệu Địa lý

Đào Văn Tùng^{1,*}, Trần Trung Chuyên¹, Nguyễn Thị Phương Bắc¹

¹ Trường Đại học Mở - Địa chất

TÓM TẮT

Trong thời đại số hóa ngày nay, việc quản lý, lưu trữ và chia sẻ dữ liệu không gian địa lý trở nên ngày càng phức tạp và quan trọng hơn bao giờ hết. Công nghệ blockchain đã nổi lên như một công cụ quan trọng có khả năng cải thiện tính an toàn, minh bạch và hiệu quả của quá trình này. Bài báo này tập trung vào việc khám phá cách mà blockchain có thể được áp dụng để lưu trữ và chia sẻ dữ liệu địa lý. Ban đầu, chúng tôi sẽ xem xét cách mà blockchain cung cấp một hệ thống bảo mật mạnh mẽ thông qua mã hóa và tính chống thay đổi của nó. Tiếp theo, chúng tôi sẽ nghiên cứu cách mà blockchain hỗ trợ quản lý quyền truy cập và xác thực dữ liệu, đảm bảo rằng chỉ những người được ủy quyền mới có thể truy cập vào thông tin địa lý cụ thể. Cuối cùng, bài báo sẽ đi sâu vào việc khám phá các cơ hội và thách thức khi triển khai công nghệ blockchain cho mục đích lưu trữ và chia sẻ dữ liệu địa lý, đồng thời đánh giá khả năng áp dụng của công nghệ này trong lĩnh vực quản lý dữ liệu không gian địa lý. Việc đánh giá này sẽ giúp ta hiểu rõ hơn cách sử dụng công nghệ chuỗi khối để tích hợp các công nghệ không gian địa lý như GIS và viễn thám.

Từ khóa: Blockchain; Geospatial data; Data storage; Data sharing; Blockchain applications

1. Đặt vấn đề

Với sự gia tăng nhanh chóng của các loại cảm biến và công nghệ định vị di động, lượng dữ liệu không gian đang phát triển mạnh mẽ. Dữ liệu không gian đã trở thành một tài sản quý giá trong nền kinh tế hiện đại; việc khai thác dữ liệu này có thể mang lại lợi ích kinh tế đáng kể và giúp người dùng đưa ra những quyết định thông minh hơn. Để đảm bảo sự lưu thông và sử dụng dữ liệu hiệu quả, nhiều tổ chức mới về chia sẻ và giao dịch dữ liệu không gian đã ra đời trong những năm gần đây. Bên cạnh phương thức lưu thông dữ liệu truyền thống, thị trường giao dịch dữ liệu lớn cũng đã xuất hiện, tạo điều kiện thuận lợi cho việc kết nối nhu cầu dữ liệu với các nguồn dữ liệu.

Trong một thị trường dữ liệu truyền thống, người bán dữ liệu thường gửi dữ liệu đến một nền tảng giao dịch tập trung. Dù nền tảng này giúp đẩy nhanh sự lưu thông của dữ liệu, nhưng vẫn còn nhiều vấn đề tồn tại:

- Người mua dữ liệu có thể có ý định xấu, lưu trữ và bán lại dữ liệu gốc với giá thấp hơn sau khi đã mua được, gây thiệt hại cho người bán dữ liệu.
- Nền tảng giao dịch tập trung có thể không đáng tin cậy, có thể lưu trữ và bán lại dữ liệu gốc mà không có sự cho phép của người bán, gây tổn hại đến lợi ích của cả hai bên.
- Nền tảng giao dịch tập trung với độ chịu lỗi thấp có thể dễ dàng bị tấn công, làm gián đoạn giao dịch dữ liệu giữa người mua và người bán, thậm chí gây mất mát dữ liệu quan trọng. Hơn nữa, nền tảng giao dịch tập trung thường thiếu kênh thông tin hiệu quả giữa người mua và người bán, dẫn đến hiệu quả giao dịch thấp.

Để khắc phục những nhược điểm của thị trường dữ liệu tập trung, như vấn đề bảo mật và quyền riêng tư, bảo vệ bản quyền dữ liệu và hiệu suất chia sẻ dữ liệu, thị trường dữ liệu phi tập trung đã được phát triển. Kiến trúc của thị trường dữ liệu phi tập trung có khả năng nâng cao tính minh bạch và độ tin cậy của các giao dịch dữ liệu bằng cách loại bỏ các điểm lỗi duy nhất và tắc nghẽn hiệu suất. Tuy nhiên, việc thiết kế hệ thống và đảm bảo an ninh cho nó sẽ phức tạp hơn so với thị trường dữ liệu tập trung do không có quản lý tập trung. Ví dụ, vấn đề thanh toán kép vẫn là một thách thức lớn trong hệ thống phân tán.

Việc áp dụng công nghệ blockchain vào hệ thống thị trường dữ liệu cho phép người bán dữ liệu giao dịch trực tiếp với người mua mà không cần phụ thuộc vào bên thứ ba, từ đó củng cố quyền sở hữu dữ liệu

* Tác giả liên hệ

Email: mail.daovantung@gmail.com

và đảm bảo tính công khai, minh bạch trong quá trình giao dịch. Điều này có thể giải quyết một phần các nhược điểm của thị trường dữ liệu tập trung. Tuy nhiên, nó cũng đặt ra các vấn đề về lưu trữ dữ liệu và bảo vệ quyền riêng tư, như sau:

- Lưu trữ dữ liệu: Hệ thống blockchain chỉ phù hợp để lưu trữ một lượng nhỏ dữ liệu giao dịch, không thể xử lý khối lượng lớn tệp dữ liệu. Do đó, các tệp dữ liệu không nên được lưu trữ trực tiếp trên blockchain. Hơn nữa, mô-đun lưu trữ tệp dữ liệu phải đảm bảo không tiết lộ thông tin danh tính, đồng thời người mua và người bán không thể biết được danh tính thực sự của đối phương.
- Bảo vệ quyền riêng tư: Cần phải thiết kế cơ chế bảo vệ quyền riêng tư trong quá trình giao dịch dữ liệu để bảo vệ danh tính của người dùng khỏi bị lộ và đảm bảo an ninh cho các giao dịch.

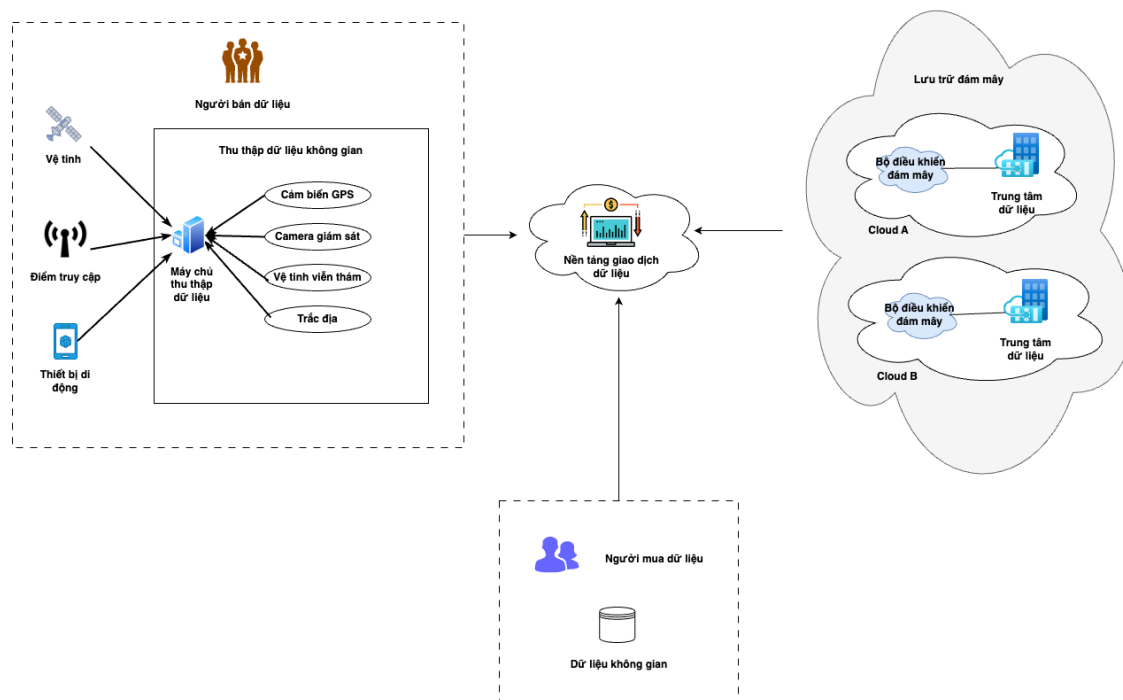
Để giải quyết các vấn đề này, chúng tôi đề xuất một khung giao dịch dữ liệu không gian dựa trên blockchain. Khung này tận dụng các lợi thế của blockchain để xây dựng một nền tảng phi tập trung cho giao dịch dữ liệu, bao gồm lưu trữ dữ liệu, truy vấn dữ liệu, định giá dữ liệu, uy tín dữ liệu và tính toán an ninh.

Đối với thách thức đầu tiên, chúng tôi sử dụng IPFS (Hệ thống tệp liên hành tinh) để lưu trữ các tệp dữ liệu. IPFS là một hệ thống lưu trữ tệp phân tán, khi một tệp được tải lên IPFS, nó trở nên khả dụng cho tất cả các nút trong mạng IPFS. Người dùng sẽ nhận được một chỉ số băm, cho phép truy xuất tệp sau này. Chỉ số này sẽ thay thế dữ liệu lưu trữ trong hợp đồng thông minh, giảm bớt gánh nặng cho toàn bộ hệ thống.

Đối với thách thức thứ hai, chúng tôi áp dụng hợp đồng thông minh để thực hiện các giao dịch dữ liệu. Hơn nữa, việc kết hợp công nghệ mã hóa và cơ chế tài khoản Ethereum sẽ giải quyết các vấn đề liên quan đến bảo vệ quyền riêng tư và an ninh dữ liệu trong giao dịch.

2. Cơ sở lý thuyết và phương pháp nghiên cứu

Nhờ các đặc điểm phi tập trung, không thể thay đổi và có thể truy xuất, công nghệ blockchain đã trở thành tâm điểm trong thị trường giao dịch dữ liệu trong những năm gần đây, thu hút sự quan tâm lớn từ ngành công nghiệp. Dữ liệu không gian trong bài báo này bao gồm: viễn thám, lập bản đồ và các dữ liệu thô khác, chẳng hạn như hình ảnh vệ tinh độ phân giải thấp, trung bình, cao, siêu cao dưới mét, dữ liệu chụp ảnh hàng không, dữ liệu vector theo tỷ lệ, dữ liệu địa hình và các loại dữ liệu không gian gốc khác.

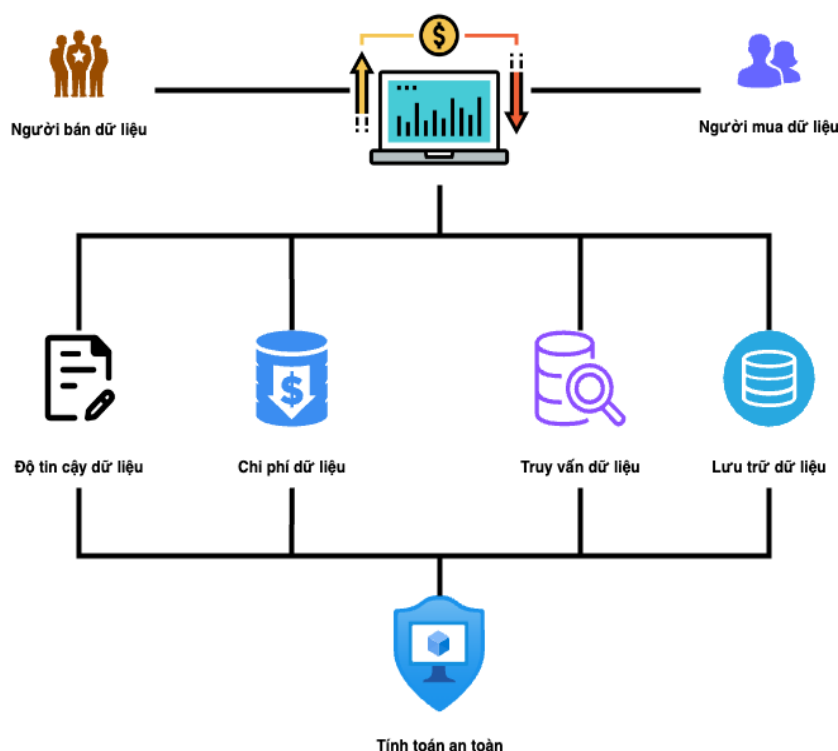


Hình 1. Hệ thống giao dịch dữ liệu không gian

Như minh họa trong Hình 1, sau khi thu thập dữ liệu không gian thông qua các thiết bị như vệ tinh,

điểm truy cập và điện thoại di động, chủ sở hữu dữ liệu tải chúng lên nền tảng lưu trữ đám mây qua mạng. Sau đó, họ bán dữ liệu này cho người mua thông qua nền tảng giao dịch dữ liệu. Người mua có thể sử dụng dữ liệu này để đạt được định vị mục tiêu chính xác cao, giám sát điều kiện đường xá hoặc thu thập thông tin về thảm họa, mang lại nhiều lợi ích thiết thực trong các lĩnh vực khác nhau.

Hệ thống giao dịch dữ liệu không gian dựa trên blockchain được thiết kế trong bài báo này bao gồm ba thành phần chính: hợp đồng thông minh trong blockchain Ethereum, hệ thống ứng dụng giao dịch cho người dùng và mạng truyền dữ liệu điểm-điểm. Như minh họa trong Hình 2, sau khi thu thập một số dữ liệu để bán, người bán đăng ký và thêm thông tin tiêu đề dữ liệu vào hợp đồng thông minh trong khi dữ liệu gốc được lưu trữ trong IPFS thông qua mô-đun lưu trữ dữ liệu. Khi cần dữ liệu để thực hiện một nhiệm vụ, người mua sử dụng mô-đun truy vấn dữ liệu và phát hành một đơn hàng chứa yêu cầu dữ liệu thông qua hợp đồng thông minh. Hệ thống sau đó sẽ truy vấn dữ liệu đủ điều kiện bằng phương thức tính toán bảo mật. Sử dụng phương pháp đấu giá, mô-đun định giá sẽ xác định giá trị của dữ liệu. Sau khi thanh toán hoàn tất, hệ thống sẽ chuyển dữ liệu đến người mua một cách an toàn. Cuối cùng, khi dữ liệu đã được chuyển giao và sử dụng, mô-đun uy tín sẽ đánh giá chất lượng dữ liệu và uy tín của người bán, đảm bảo sự minh bạch và công bằng trong hệ thống giao dịch.



Hình 2. Nền tảng giao dịch dữ liệu

Chúng tôi đã triển khai một hệ thống giao dịch dữ liệu không gian dựa trên chuỗi Ethereum riêng tư, bao gồm hệ thống ứng dụng giao dịch cho người dùng, hợp đồng thông minh trên mạng Ethereum và mạng truyền dữ liệu. Cụ thể, hệ thống ứng dụng được viết bằng JavaScript, hợp đồng thông minh Ethereum được viết bằng Solidity, và giao diện IPFS được sử dụng cho việc truyền dữ liệu.

Chúng tôi đơn giản hóa hệ thống thành một người mua và nhiều người bán để giao dịch. Cơ chế định giá áp dụng phương pháp đấu giá giá thứ hai, trong đó dữ liệu giao dịch sẽ chọn giá thầu thấp nhất của người bán, trong khi việc thanh toán dữ liệu sẽ theo giá thầu thấp thứ hai.

Đầu tiên, người bán dữ liệu đăng ký và thêm thông tin dữ liệu mới vào hợp đồng thông minh, trong khi bản thân dữ liệu được lưu trữ trên IPFS, được triển khai trong môi trường thực thi tin cậy (TEE), và người mua dữ liệu phát hành một đơn hàng chứa các yêu cầu dữ liệu. Sau khi điều kiện yêu cầu dữ liệu của người mua trong đơn hàng được khớp, người bán sẽ được phép tham gia đấu giá trong giao dịch. Sau đó, hệ thống sẽ thực hiện cơ chế định giá để hoàn tất giao dịch giữa người mua và người bán, và chuyển dữ liệu kết quả cho người mua thông qua phương pháp tính toán an toàn. Sau khi dữ liệu đã được chuyển giao và sử dụng, người mua sẽ chấm điểm cho người bán và dữ liệu của họ dựa trên chất lượng dữ liệu. Dựa trên kết quả đánh giá, giá trị uy tín của người bán sẽ được điều chỉnh để đảm bảo độ tin cậy của giao

dịch dữ liệu.

Để mô tả cách hoạt động của khung, chúng tôi giới thiệu quy trình chính của hệ thống:

- Bước 1: Đăng ký tài khoản. Để thực hiện giao dịch trong hệ thống giao dịch dữ liệu, người bán cần đăng ký tài khoản Ethereum trong mô-đun quản lý tài khoản Ethereum tương ứng. Bên cạnh đó, người bán cũng cần đăng ký tài khoản trong hợp đồng thông minh.
- Bước 2: Đăng ký dữ liệu. Sau khi thu thập một số dữ liệu để bán, người bán có thể mã hóa và tải dữ liệu lên IPFS, sau đó đăng ký dữ liệu vào tài khoản hợp đồng thông minh, bao gồm địa chỉ lưu trữ, giá trị băm và thời gian đăng ký của dữ liệu, tóm tắt dữ liệu liên quan đến ảnh thu nhỏ và siêu dữ liệu, v.v.
- Bước 3: Phát hành đơn hàng. Nếu một người mua dữ liệu muốn mua dữ liệu cụ thể trong hệ thống giao dịch, họ sẽ thêm một đơn hàng vào hợp đồng thông minh. Đơn hàng chứa yêu cầu của người mua về dữ liệu, bao gồm loại dữ liệu, điều kiện truy vấn dữ liệu, thời gian đăng ký dữ liệu, giá cả, số lượng dữ liệu, giá trị uy tín của người bán, v.v. Người mua có thể thiết lập yêu cầu về thời gian đăng ký dữ liệu và giá trị uy tín của người bán khi thêm đơn hàng. Sau đó, những người bán dữ liệu đáp ứng yêu cầu về thời gian đăng ký và uy tín sẽ đủ điều kiện để đấu giá đơn hàng, nhằm cải thiện tính kịp thời và độ tin cậy của giao dịch dữ liệu.
- Bước 4: Cung cấp dữ liệu. Người bán dữ liệu chọn những dữ liệu đáp ứng nhu cầu của người mua và đưa ra mức giá dự kiến đã được niêm phong trong một giá trị băm làm giá thầu của họ.
- Bước 5: Thông báo đấu giá. Người mua dữ liệu thông báo hợp đồng thông minh chuẩn bị cho cuộc đấu giá của người bán. Đối với những người bán đã được chấp nhận tham gia giao dịch, họ sẽ được thông báo tiếp tục đấu giá; trong khi những người bán mới sẽ không được phép tham gia giao dịch.
- Bước 6: Gửi giá thầu. Người bán dữ liệu gửi giá thầu của mình, và giá thầu này phải khớp với giá trị băm trước đó.
- Bước 7: Cơ chế định giá. Chúng tôi sử dụng cơ chế đấu giá ngược Vickrey cho việc định giá dữ liệu, đảm bảo tính xác thực, nên mỗi người bán đều có động lực đưa ra đánh giá trung thực về dữ liệu của mình.
- Bước 8: Thanh toán dữ liệu. Giao dịch được thực hiện ở mức giá thầu thấp nhất của người bán, và người bán được thanh toán theo mức giá thầu thấp thứ hai.
- Bước 9: Giao hàng dữ liệu. Dữ liệu được mã hóa của người bán thắng thầu sẽ được chuyển đến người mua để hoàn tất đơn hàng.
- Bước 10: Đánh giá uy tín. Sau khi dữ liệu đã được giao và sử dụng, người mua sẽ chấm điểm cho người bán và dữ liệu dựa trên chất lượng dữ liệu. Dựa trên kết quả đánh giá, giá trị uy tín của người bán sẽ được đánh giá lại dưới sự bảo đảm của cơ chế đánh giá uy tín.

Hợp đồng thông minh là một tập hợp các cam kết kỹ thuật số được triển khai tại một địa chỉ cụ thể trên blockchain. Nền tảng giao dịch dữ liệu tạo ra một hợp đồng thông minh cho các giao dịch dữ liệu. Tất cả các thành viên tham gia giao dịch dữ liệu đều có thể đăng ký sau khi đồng ý với các điều khoản và điều kiện của hợp đồng. Điều kiện kích hoạt của hợp đồng được quy định trong hợp đồng thông minh. Tại một thời điểm nhất định, khi các điều kiện đáp ứng yêu cầu kích hoạt, hợp đồng sẽ tự động được thực thi.

2.1 Truy vấn dữ liệu

Người mua dữ liệu có thể tìm kiếm các nguồn dữ liệu không gian thông qua nhiều phương pháp truy vấn khác nhau như truy vấn từ khóa, truy vấn theo khu vực, truy vấn lân cận gần nhất, truy vấn lân cận xa nhất, và nhiều phương pháp khác. Họ có thể xem chi tiết dữ liệu không gian qua tiêu đề dữ liệu, hình thu nhỏ và siêu dữ liệu để xác định yêu cầu của mình. Hơn nữa, người mua có thể chuyển các yêu cầu dữ liệu cụ thể thành biểu thức logic hoặc hàm toán học và lưu trữ chúng trên blockchain để người bán có thể truy vấn. Dựa trên yêu cầu dữ liệu của người mua, người bán cũng có thể tiến hành truy vấn để xác nhận xem dữ liệu có đáp ứng yêu cầu của người mua hay không.

Dù các điều kiện truy vấn dữ liệu không gian có vẻ đơn giản, việc tải trực tiếp các yêu cầu truy vấn của người mua lên blockchain có thể không an toàn, vì kẻ tấn công có thể suy luận yêu cầu dữ liệu của người mua và xâm phạm quyền riêng tư của cả hai bên, làm tăng nguy cơ rò rỉ thông tin cá nhân. Để bảo vệ quyền riêng tư của người dùng, chúng ta cần thực hiện truy vấn dữ liệu trong khi ẩn đi các điều kiện truy vấn. Một giải pháp phổ biến là mã hóa hàm, trong đó người mua dữ liệu với khóa giải mã có thể nhận giá trị truy vấn từ dữ liệu đã mã hóa mà không tiết lộ thông tin về dữ liệu gốc.

2.2 Lưu trữ dữ liệu

Hầu hết các hệ thống blockchain công cộng hiện nay gặp phải hạn chế về số lượng và không gian của các giao dịch trong khối do kích thước khối có hạn (như hệ thống Bitcoin) hoặc lượng "Gas" tiêu tốn trong khối (như hệ thống Ethereum). Vì vậy, việc lưu trữ khối lượng lớn dữ liệu trực tiếp trên blockchain trong hệ thống giao dịch dữ liệu không gian là không khả thi. Gần đây, hệ thống lưu trữ dữ liệu phân tán InterPlanetary File System (IPFS) đã được áp dụng rộng rãi để lưu trữ dữ liệu chia sẻ trong các lĩnh vực như chăm sóc sức khỏe, điện toán đám mây, IoT, và nhiều lĩnh vực khác.

IPFS là một hệ thống lưu trữ tệp phân tán, peer-to-peer, và có khả năng xác định nội dung, sử dụng một mạng lưới máy tính kết nối. Khi một tệp được tải lên IPFS, nó sẽ có sẵn cho tất cả các nút trong mạng IPFS. Tệp được tải lên sẽ được chia thành các phần nhỏ, mỗi phần được gán một mã băm mã hóa duy nhất. Nhờ vào mã băm này, dữ liệu trên IPFS có thể được xác định theo nội dung, tạo điều kiện cho việc truy xuất nhanh chóng và hiệu quả. IPFS sử dụng bảng băm phân tán (DHTs) để định vị các tệp. Trong hệ thống giao dịch dữ liệu, các tệp dữ liệu được lưu trữ trên IPFS, trong khi các mã băm liên kết với IPFS được lưu trong các khối của blockchain, giúp giảm thiểu chi phí lưu trữ lớn. Tóm lại, IPFS cung cấp một mô hình lưu trữ an toàn với hiệu suất cao, hỗ trợ truy cập đồng thời và dung lượng lưu trữ lớn.

Dịch vụ lưu trữ đám mây của nền tảng giao dịch dữ liệu không gian lớn chủ yếu nhằm thiết lập một trạm lưu trữ hiệu quả, sử dụng công nghệ máy tính, Internet, và Internet of Things để quản lý hàng ngày các sản phẩm và dịch vụ được giao dịch trên nền tảng, cũng như các thông tin phát sinh trong quá trình giao dịch dữ liệu. Dịch vụ này có khả năng nhanh chóng và chính xác tổng hợp thông tin giao dịch sản phẩm và dịch vụ, nhờ vào các ưu điểm như truy xuất nhanh, độ tin cậy cao, dung lượng lưu trữ lớn và bảo mật tốt, cho phép lưu trữ một lượng lớn dữ liệu một cách an toàn.

Mặc dù việc lưu trữ trực tiếp dữ liệu gốc trên blockchain không thực tế, mã băm liên kết với dữ liệu cụ thể có thể được lưu trữ trên blockchain, trong khi dữ liệu gốc được lưu trữ trên IPFS. Khi dữ liệu được lưu trữ thành công trên IPFS, người dùng sẽ nhận được một chỉ số băm, giúp truy xuất tệp sau này. Chỉ số này sẽ thay thế dữ liệu lưu trữ trong hợp đồng thông minh, làm giảm tình trạng tắc nghẽn lưu trữ trong toàn bộ hệ thống.

Chúng tôi đại diện dữ liệu không gian dưới dạng $T_1 = \{\text{hashid, time, space, other attributes}\}$, trong đó hashid là giá trị băm được tạo ra từ việc băm dữ liệu không gian, tức là $\text{hashid} = \text{hash}(\text{time, space, other attributes})$; time là thuộc tính thời gian cho biết thời điểm dữ liệu được tạo ra; space là thuộc tính không gian, thường được biểu thị dưới dạng tọa độ kinh độ và vĩ độ. Để thuận tiện cho việc truy vấn và lưu trữ dữ liệu không gian, thuộc tính không gian được chuyển đổi bằng thuật toán GeoHash (Geographical Hash), giúp băm thông tin kinh độ và vĩ độ. Các thuộc tính khác biểu thị các đặc điểm khác của dữ liệu không gian ngoài thuộc tính thời gian và không gian, bao gồm thông tin chữ ký.

2.3 Chi phí dữ liệu

Dữ liệu, như một hàng hóa đặc biệt, mang những đặc điểm riêng biệt, làm cho việc định giá trở nên phức tạp hơn. Đầu tiên, chi phí cận biên của dữ liệu là rất thấp (chi phí cận biên ở đây là chi phí sao chép dữ liệu), vì vậy một khi người mua đã nhận được dữ liệu từ người bán, họ có thể dễ dàng bán lại dữ liệu đó. Thứ hai, giá trị của dữ liệu không chỉ phụ thuộc vào khối lượng mà còn vào nội dung và chất lượng của nó. Ví dụ, một tập hợp các hình ảnh khuôn mặt sẽ không có giá trị nhiều đối với người cần hình ảnh viễn thám. Thứ ba, việc ước lượng giá trị dữ liệu rất khó khăn. Hơn nữa, sự đánh giá giá trị có thể rất khác biệt giữa các người dùng khác nhau.

Nghiên cứu cho thấy hầu hết người mua dữ liệu chỉ cần các kết quả thống kê như giá trị trung bình của tập dữ liệu hoặc các đặc trưng trích xuất từ dữ liệu cho mục đích học máy, thay vì chính dữ liệu. Do đó, người mua chủ yếu mua quyền sử dụng dữ liệu thay vì quyền sở hữu dữ liệu. Theo cách này, dữ liệu thực sự không nằm trong quyền sở hữu của người tiêu dùng.

Mặc dù giá trị của dữ liệu rất khó định lượng, việc đánh giá giá trị dữ liệu là cần thiết vì người mua thường cần dữ liệu từ nhiều người bán khác nhau, và sự đánh giá giá trị của dữ liệu có thể rất khác nhau giữa các người bán. Để tính toán giá trị dữ liệu, giá trị Shapley (một phương pháp phân chia lợi ích và chi phí công bằng cho nhiều bên tham gia) có thể được sử dụng để tính toán sự đóng góp của từng mảnh dữ liệu. Do độ phức tạp tính toán của giá trị Shapley gia tăng theo cấp số nhân với số lượng dữ liệu, các thuật toán xấp xỉ hoặc phân tán thường được áp dụng trong thực tế.

Trong nghiên cứu này, chúng tôi áp dụng lý thuyết đấu giá để thiết kế cơ chế định giá trong hệ thống giao dịch dữ liệu không gian. Mỗi người bán dữ liệu có một giá trị riêng, bao gồm đánh giá rủi ro liên quan đến quyền riêng tư của dữ liệu; ngược lại, mỗi người mua dữ liệu cũng có một mức giá trị cho dữ liệu mà họ muốn mua. Với thông tin không đồng đều về giá trị, cả người mua và người bán có thể báo cáo không chính xác giá trị dữ liệu của mình. Để khắc phục vấn đề này, chúng tôi sử dụng lý thuyết đấu giá để thiết kế một cơ chế khuyến khích phù hợp, giúp hai bên đạt được lợi ích tối đa khi báo cáo giá trị thực của mình, bổ sung thêm bằng thuật toán tính toán giá trị Shapley, qua đó làm cho việc thiết kế cơ chế

định giá trở nên dễ dàng hơn.

2.4 Độ tin cậy dữ liệu

Người mua dữ liệu thường gặp khó khăn trong việc đánh giá chất lượng dữ liệu của người bán ngay lập tức trước khi sử dụng. Trong môi trường giao dịch dữ liệu trực tuyến, người bán thường sở hữu lợi thế thông tin vượt trội hơn so với người mua. Khi thông tin không đồng đều, người bán dữ liệu không đáng tin cậy có thể lừa đảo bằng cách đưa ra giá thấp để bán dữ liệu kém chất lượng và thu lợi bất chính.

Để khắc phục vấn đề này, bài báo đề xuất một cơ chế đánh giá uy tín nhằm đảm bảo tính bền vững lâu dài của giao dịch dữ liệu. Uy tín, như một chỉ báo quan trọng về chất lượng dữ liệu, trở thành tài sản vô hình có giá trị đối với người bán dữ liệu. Khi chất lượng dữ liệu không thể quan sát trực tiếp, uy tín đóng vai trò then chốt, giúp giảm thiểu sự không chắc chắn trong giao dịch. Điểm uy tín của người bán dữ liệu được tính toán dựa trên phản hồi từ người mua sau khi họ tiếp nhận và sử dụng dữ liệu, phản ánh chính xác chất lượng dữ liệu mà người bán cung cấp.

2.5 Độ an toàn

Trong hệ thống blockchain, thông tin thường được công khai cho tất cả người dùng, và mọi giao dịch hoặc kịch bản đều minh bạch nhờ tính mở và minh bạch của công nghệ blockchain. Tuy nhiên, do hạn chế kích thước khối, khả năng tính toán của hợp đồng thông minh trên blockchain thường khá yếu, khiến việc thực hiện trực tiếp các giao dịch từ người mua trong hợp đồng thông minh trở nên không an toàn và không khả thi do chi phí tính toán quá cao. Vấn đề này cũng xuất hiện trong các quá trình truy vấn dữ liệu và định giá dữ liệu. Thêm vào đó, các nút trên chuỗi công khai như Bitcoin, Ethereum không hoàn toàn tin cậy lẫn nhau, làm cho việc bảo vệ quyền riêng tư trở nên khó khăn hơn vì chúng hoàn toàn mở và minh bạch.

Để giải quyết vấn đề này, phần cứng tin cậy như môi trường thực thi tin cậy (TEE) là một phương pháp phổ biến để thực hiện tính toán an toàn. TEE đảm bảo rằng mã và dữ liệu được tải vào nó được bảo vệ về mặt bảo mật và toàn vẹn. Người bán dữ liệu sẽ gửi dữ liệu của mình đến thiết bị TEE của người mua an toàn, người sở hữu phần cứng TEE, từ đó nhiệm vụ tính toán sẽ được thực hiện trong TEE và kết quả sẽ được trả lại cho người mua một cách an toàn. Cách ly phần cứng trong TEE bảo vệ dữ liệu và dịch vụ tính toán khỏi các ứng dụng chạy trên hệ điều hành, trong khi các ứng dụng tin cậy chạy trong TEE có thể truy cập đầy đủ chức năng của bộ xử lý chính và bộ nhớ của thiết bị. Các công nghệ phần cứng điển hình hỗ trợ TEE bao gồm ARM TrustZone và Intel SGX (Software Guard Extensions).

Intel Software Guard Extensions (SGX) cung cấp một triển khai TEE phổ biến cho tính toán mục đích chung, với các đơn vị được gọi là enclaves trong SGX. Mã chạy bên trong một enclave có không gian địa chỉ được bảo vệ. Khi dữ liệu trong enclave di chuyển ra khỏi bộ xử lý đến bộ nhớ, nó được mã hóa một cách minh bạch bằng các khóa chỉ có sẵn cho bộ xử lý. Do đó, hệ điều hành, hypervisor và các người dùng khác không thể truy cập bộ nhớ của enclave. Trong enclave, mã và dữ liệu được đo lường tại giai đoạn khởi động và phép đo này được ký vào báo cáo chứng thực dựa trên gốc tin cậy phần cứng. Báo cáo này có thể được xác minh để chứng minh mã enclave không bị sửa đổi, từ đó người dùng có thể xác nhận tính bảo mật của enclave.

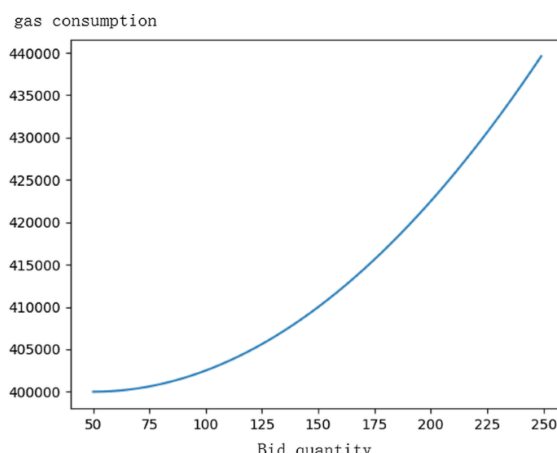
Trong bài báo này, chúng tôi sử dụng Intel SGX làm ví dụ điển hình để xây dựng một hệ thống trao đổi tin cậy bằng cách sử dụng môi trường thực thi tin cậy, hỗ trợ trong việc thanh toán công bằng cho các giao dịch.

3. Kết quả và thảo luận

Trong hệ thống, quyền riêng tư về danh tính và dữ liệu của những người tham gia được bảo vệ. Cả người mua và người bán dữ liệu đều sử dụng tài khoản riêng cho mỗi giao dịch, điều này giúp giảm thiểu thông tin liên quan giữa các giao dịch khác nhau. Trong quá trình thanh toán, cả hai bên đều phải cung cấp địa chỉ tài khoản và chữ ký khóa liên quan để hoàn tất giao dịch. Sau khi thanh toán, người mua sẽ nhận được dữ liệu từ hệ thống tệp IPFS dựa trên địa chỉ lưu trữ tệp dữ liệu và giá trị băm tương ứng, mà không cần liên hệ trực tiếp với người bán dữ liệu. Tất cả điều này cho phép người mua và người bán dữ liệu ẩn danh tính của mình trong quá trình giao dịch dữ liệu, loại bỏ nguy cơ tiết lộ quyền riêng tư về danh tính. Do đó, quyền riêng tư về danh tính dựa trên tính ẩn danh của blockchain, trong khi quyền riêng tư về dữ liệu được bảo vệ hoàn toàn thông qua cơ chế mã hóa dữ liệu và tính toán bảo mật, ngăn chặn kẻ tấn công thu thập thêm thông tin để tấn công trong quá trình giao dịch.

Vì việc định giá dữ liệu được thực hiện trên chuỗi khối Ethereum thông qua các hợp đồng thông minh, chi phí tính toán cần được nhấn mạnh do khả năng tính toán hạn chế của hợp đồng thông minh. Chi phí định giá dữ liệu dựa trên đấu giá ngược Vickrey được đánh giá dựa trên mức tiêu thụ gas trên chuỗi khối Ethereum. Hình sau cho thấy mối quan hệ giữa mức tiêu thụ gas của việc định giá dữ liệu và số lượng giá

thầu cho các đơn hàng dữ liệu. Chúng tôi nhận thấy rằng khi có một số lượng lớn người bán dữ liệu đầu giá cho cùng một đơn hàng dữ liệu, chi phí định giá dữ liệu sẽ tăng cao rõ rệt, điều này trở thành một hạn chế của hệ thống giao dịch dựa trên chuỗi khối Ethereum. Do đó, trong các nghiên cứu tiếp theo, việc giảm chi phí định giá dữ liệu sẽ là một trong những hướng nghiên cứu trong tương lai.



Chúng tôi đã triển khai một hệ thống giao dịch dữ liệu không gian dựa trên blockchain. Kết quả thử nghiệm cho thấy hệ thống này vượt trội về khả năng thông lượng, yêu cầu dung lượng lưu trữ thấp hơn và có khả năng chịu lỗi tốt hơn, đáp ứng hiệu quả các yêu cầu giao dịch dữ liệu không gian với tần suất cao và thời gian thực. Tính bền vững của hệ thống giao dịch là yếu tố quan trọng để đảm bảo hoạt động lâu dài và hiệu quả của hệ thống. Việc tích hợp cơ chế uy tín vào hệ thống giao dịch đã tạo thêm một lớp bảo vệ hữu ích cho cơ chế đấu giá, nâng cao độ tin cậy của các giao dịch. Tuy nhiên, một vấn đề có thể xảy ra là những người tham gia đấu giá có chất lượng cao có thể không được chọn nhiều lần, điều này có thể làm giảm động lực tham gia của họ do cơ hội chiến thắng thấp. Để khắc phục điều này, chúng tôi dự định kết hợp cơ chế uy tín với thuật toán “tín dụng tham gia ảo” nhằm nâng cao cơ hội cho những người thua cuộc được chọn, từ đó xây dựng một hệ thống giao dịch lâu dài và bền vững.

Ngoài ra, sự mâu thuẫn không thể tránh khỏi giữa khả năng mở rộng, phân quyền và bảo mật là một vấn đề lớn trong các hệ thống blockchain. Vì hệ thống hoàn toàn phân quyền và không dựa vào bất kỳ bên thứ ba đáng tin cậy nào, việc đạt được sự cân bằng giữa khả năng mở rộng và bảo mật trở nên đặc biệt thách thức. Để đảm bảo bảo mật, hệ thống mà chúng tôi đề xuất đã phải hy sinh một phần khả năng mở rộng. Mặc dù các công nghệ phân vùng đồng thời và tối ưu hóa song song đã cải thiện hiệu quả giao dịch, điều này đã được xác nhận trong và các thí nghiệm của chúng tôi, nhưng việc đạt được mức bảo mật rất cao vẫn là một thách thức lớn. Trong các triển khai tương lai của hệ thống giao dịch dữ liệu, chúng tôi sẽ tiếp tục nghiên cứu sự đánh đổi giữa khả năng mở rộng, phân quyền và bảo mật.

Chúng tôi đã triển khai một hệ thống giao dịch dữ liệu không gian dựa trên chuỗi khối Ethereum riêng tư, hợp đồng thông minh trên mạng Ethereum và mạng truyền tải dữ liệu. Tất cả các bài kiểm tra của hệ thống này được hoàn thành trên máy tính PC với 18 lõi ảo (Intel Xeon Platinum 8124M 3 GHz). Mô-đun giao dịch trực tuyến của hệ thống giao dịch dữ liệu chủ yếu được triển khai bằng ngôn ngữ Python.

Bài báo này chủ yếu kiểm tra thông lượng của hệ thống, dung lượng lưu trữ và khả năng chịu lỗi. Đối với thông lượng hệ thống, chúng tôi kiểm tra lượng dữ liệu có thể xử lý mỗi đơn vị thời gian dưới điều kiện số lượng nút và dữ liệu khác nhau thông qua thử nghiệm đồng thời. Đối với dung lượng lưu trữ, chúng tôi kiểm tra dung lượng lưu trữ cần thiết cho dữ liệu với các số lượng nút khác nhau. Đối với khả năng chịu lỗi, chúng tôi so sánh tỷ lệ lỗi của nút khi bị tấn công độc hại trong các chế độ khác nhau. Để cải thiện hiệu suất giao dịch của hệ thống, công nghệ tối ưu hóa phân chia đồng thời và song song đã được áp dụng trong thí nghiệm.

Trong các hệ thống chuỗi khối liên minh như Fabric, thông lượng của nút bị hạn chế bởi độ trễ truyền thông, thời gian xác minh giao dịch và thời gian thực hiện hàm băm. Trong một hệ thống với N nút và trung bình mỗi khối có T giao dịch, giá sử độ trễ truyền thông của nút là t_c , thời gian xác minh giao dịch là t_v , và thời gian thực hiện một hàm băm là t_o , thì thông lượng của toàn hệ thống được biểu diễn bằng phương trình:

$$TPS = T / (N * t_c + T * t_v + T * t_o)$$

Để kiểm tra yêu cầu lưu trữ của các giao dịch trong hệ thống, bài báo này so sánh sự thay đổi dung lượng lưu trữ với số lượng nút. Trong một hệ thống chuỗi khối với N nút, số lượng giao dịch là NT, giá

sử dụng lượng lưu trữ trung bình của mỗi giao dịch là St , và dung lượng lưu trữ của một kết quả hàm băm là Sh . Khi đó, chúng ta có thể ước tính dung lượng lưu trữ là:

$$S = NT * N * (St + Sh)$$

4. Kết luận

Trong bài báo này, chúng tôi đề xuất khung giao dịch dữ liệu không gian dựa trên blockchain, trong đó chúng tôi thiết kế một nền tảng phân cấp với môi trường thực thi đáng tin cậy (TEE) bằng cách sử dụng khung.

Nền tảng này có đặc điểm là khả năng lưu trữ và xử lý dữ liệu viễn thám khổng lồ, được thiết kế để xử lý không chỉ giao dịch dữ liệu không gian mà còn cả các dịch vụ liên quan trong tương lai. Nó sẽ cung cấp giao diện dịch vụ, trên đó người dùng, không cần cài đặt trước, có thể nhận được các dịch vụ tích hợp của dữ liệu viễn thám, phần mềm và tài nguyên máy tính thông qua thiết bị đầu cuối dịch vụ đám mây, có thể sử dụng thông tin viễn thám có độ phân giải cao một cách thuận tiện và tiết kiệm, giảm thiểu hiệu quả ngưỡng của viễn thám về chi phí và bảo trì.

5. Tài liệu tham khảo

1. D. Zhao, C. Ye, B. Zhang, Data marketplace and its values in data trading. *Libr. Inf. Serv.* 61(13), 5–12 (2017)
2. G.S. Ramachandran, R. Radhakrishnan, B. Krishnamachari, Towards a decentralized data marketplace for smart cities, in 2018 IEEE International Smart Cities Conference (ISC2) (2019)
3. M. Li, J. Weng, A. Yang, W. Lu, Y. Zhang, L. Hou, L. Jia-Nan, Y. Xiang, R. Deng, Crowdbc: a blockchain-based decentralized framework for crowdsourcing. *IEEE Trans. Parallel Distrib. Syst.* 1 (2018)
4. W. Shi, C. Wu, Z. Li, A Shapley-value mechanism for bandwidth on demand between datacenters. *IEEE Trans. Cloud Comput.* 6(1), 19–32 (2018)
5. J. Zhang, S. Zhong, J. Wang, X. Yu, O. Alfarraj, A storage optimization scheme for blockchain transaction databases. *Comput. Syst. Sci. Eng.* 36(3), 521–535 (2021)
6. F. Baig, F. Wang, Blockchain enabled distributed data management—a vision, in 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW) (2019)
7. S. Dziembowski, L. Ekey, S. Faust, Fairswap: How to fairly exchange digital goods, in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15–19, 2018, pp. 967–984 (2018)
8. J.Z.F. Pang, H. Fu, W.I. Lee, A. Wierman, The efficiency of open access in platforms for networked cournot markets, in IEEE Infocom-IEEE Conference on Computer Communications (2017)
9. L. Ruinian, S. Tianyi, M. Bo, L. Hong, C. Xiuzhen, S. Limin, Blockchain for large-scale internet of things data storage and protection. *IEEE Trans. Serv. Comput.* 12, 762–771 (2018)
10. Y. Chen, H. Li, K. Li, J. Zhang, An improved p2p file system scheme based on ipfs and blockchain, in 2017 IEEE International Conference on Big Data (Big Data) (2017)
11. D.D. Nguyen, M.I. Ali, Enabling on-demand decentralized iot collectability marketplace using blockchain and crowdsensing, in Global IoT Summit (2019)
12. N. Eltayieb, R. Elhabob, A. Hassan, F. Li, A blockchain-based attribute-based signcryption scheme to secure data sharing in the cloud. *J. Syst. Archit.* 102, 101653 (2019)
13. G. Su, W. Yang, Z. Luo, Y. Zhang, Y. Zhu, Bdtf: a blockchain-based data trading framework with trusted execution environment (2020)
14. G. Zyskind, D.M.S. Zekrifa, P. Alex, O. Nathan, Decentralizing privacy: using blockchain to protect personal data, in IEEE Security and Privacy Workshops (2015)
15. L.D. Nguyen, S.R. Pandey, B. Soret, A. Broering, P. Popovski, A marketplace for trading AI models based on blockchain and incentives for iot data. *CoRR arXiv:2112.02870* (2021)