PHIẾU GIAO NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP

1. Mục đích nội dung của ĐATN

Tìm hiểu mô hình xuất bản trực tuyến xác thực và các cấu trúc dữ liệu xác thực. Thiết kế và cài đặt thử nghiệm mô hình trên hệ thống xuất bản dữ liệu hành chính địa lý.

2. Các nhiệm vụ cụ thể của ĐATN

- Tìm hiểu về các mô hình xuất bản trực tuyến.
- Tìm hiểu về các cấu trúc dữ liệu xác thực.
- Tìm hiểu về sự ứng dụng của các cấu trúc dữ liệu xác thực vào trong xuất bản trực tuyến.
- Thiết kế, cài đặt và đánh giá mô hình xuất bản xác thực.

3. Lời cam đoan của sinh viên:

Tôi *Nguyễn Văn Việt* cam kết ĐATN là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS Nguyễn Khanh Văn*.

Các kết quả nêu trong ĐATN là trung thực, không phải là sao chép toàn văn của bất kỳ công trình nào khác.

Hà Nội, ngày 25 tháng 05 năm 2009
Tác giả ĐATN

Họ và tên sinh viên
Nguyễn Văn Việt

4. Xác nhận của giáo viên hướng dẫn về mức độ hoàn thành của ĐATN và cho phép bảo vệ:

Hà Nội, ngày 25 tháng 05 năm 2009
Giáo viên hướng dẫn

TS. Nguyễn Khanh Văn

LỜI CẢM ƠN

Trước hết, em xin được chân thành gửi lời cảm ơn sâu sắc tới **các thầy cô giáo trong trường Đại học Bách Khoa Hà Nội** nói chung và **các thầy cô trong khoa Công nghệ Thông tin, bộ môn Công nghệ phần mềm** nói riêng đã tận tình giảng dạy, truyền đạt cho em những kiến thức, những kinh nghiệm quý báu trong suốt 5 năm học tập và rèn luyện tại trường Đại học Bách Khoa Hà Nội.

Em xin được gửi lời cảm ơn đến **thầy Nguyễn Khanh Văn – trưởng bộ môn Công nghệ phần mềm, khoa Công nghệ Thông tin, trường Đại học Bách Khoa Hà Nội** đã hết lòng giúp đỡ, hướng dẫn và chỉ dạy tận tình trong quá trình em làm đồ án tốt nghiệp.

Cuối cùng, em xin được gửi lời cảm ơn chân thành tới **gia đình, bạn bè** đã động viên, chăm sóc, đóng góp ý kiến và giúp đỡ trong quá trình học tập, nghiên cứu và hoàn thành đồ án tốt nghiệp.

Hà Nội, ngày 25 tháng 05 năm 2009

Nguyễn Văn Việt

Sinh viên lớp **Công nghệ phần mềm B – K49**

Khoa **Công nghệ Thông tin - Đại học Bách Khoa Hà Nội**

TÓM TẮT NỘI DUNG ĐỒ ÁN TỐT NGHIỆP

Trong bối cảnh Internet phổ biến trên toàn thế giới, lĩnh vực xuất bản trực tuyến đang ngày càng trở nên hấp dẫn. So với xuất bản trên giấy, xuất bản trực tuyến (online publishing) có ưu điểm rõ rệt về tốc độ cập nhật thông tin, dễ dàng tìm kiếm qua các công cụ tìm kiếm như yahoo, google, livesearch…. Bên cạnh đó người sử dụng không phải mất công giữ gìn những thông tin họ cần mà khi cần họ lập tức có thể tìm đến nhà xuất bản trực tuyến để lấy thông tin.

Tuy vậy, lĩnh vực xuất bản trực tuyến cũng gặp rất nhiều khó khăn, thách thức để phát triển và đang chỉ ở mức tiềm năng. Những khó khăn lớn nhất có thể nói tới là khó khăn về bảo mật, và chi phí đầu tư cơ sở hạ tầng xuất bản trực tuyến. Đồ án sẽ trình bày một hướng tiếp cận giúp hạn chế những khó khăn này.

Trước tiên, Người viết luận văn (NVLV) sẽ trình bày một hướng tiếp cận hay đối với mô hình xuất bản trực tuyến. Trong mô hình xuất bản được trình bày trong đồ án, Chủ sở hữu dữ liệu không trực tiếp xuất bản dữ liệu của họ cho Người sử dụng dịch vụ xuất bản mà công việc xuất bản dữ liệu của Chủ sở hữu dữ liệu được thực hiện bởi các Nhà xuất bản trực tuyến. Điều đặc biệt là các Nhà xuất bản này không nhất thiết phải là đáng tin cậy. Bởi vì mô hình cung cấp khả năng cho phép Người sử dụng dịch vụ xuất bản xác minh được kết quả truy vấn họ nhận được là chính xác hay không. Để thực hiện được điều này, mô hình sử dụng các cấu trúc dữ liệu xác thực trong việc xuất bản trực tuyến. Mô hình giúp giảm gánh nặng trong việc cung cấp dịch vụ xuất bản cho Chủ sở hữu dữ liệu. Bên cạnh đó, chất lượng của dịch vụ xuất bản tăng và giá cả của dịch vụ xuất bản giảm do bất kỳ một Nhà xuất bản nào cũng có thể tham gia vào việc xuất bản dữ liệu. Về phía Người sử dụng dịch vụ xuất bản, họ cũng sẽ an tâm hơn khi sử dụng dịch vụ xuất bản trực tuyến. Với những ưu điểm lớn đó, mô hình sẽ góp phần thúc đẩy lĩnh vực xuất bản trực tuyến còn rất tiềm năng phát triển.

Tiếp đó, NVLV sẽ tập trung trình bày các vấn đề về mô hình xuất bản xác thực và các cấu trúc dữ liệu xác thực. Dựa vào những lý thuyết đã tìm hiểu NVLV sẽ cài đặt thử nghiệm ứng dụng xuất bản dựa vào mô hình được trình bày trong đồ án và sử dụng một cấu trúc dữ liệu xác thực điển hình. Qua đó, chúng ta sẽ đưa ra một vài đánh giá về mô hình và hướng phát triển trong tương lai.

Đồ án được chia thành các phần chính như sau:

Chương 1 trình bày tổng quan về bài toán xuất bản trực tuyến.

Chương 2 trình bày về cấu trúc dữ liệu Merkle Hash Tree và phương thức xác thực theo hướng từ dưới lên.

Chương 3 trình bày ứng dụng của cấu trúc dữ liệu xác thực Merkle Hash Tree trong việc xuất bản xác thực dữ liệu được quản lý bởi hệ quản trị cơ sở dữ liệu quan hệ.

Chương 4 trình bày một mô hình chung của các cấu trúc dữ liệu xác thực và phương thức xác thực theo hướng từ trên xuống. Chương 4 cũng trình bày một số cấu trúc dữ liệu xác thực điển hình và phương thức xác thực từ trên xuống trên các cấu trúc dữ liệu đó.

Chương 5 trình bày về thiết kế và cài đặt thử nghiệm mô hình xuất bản xác thực thông tin hành chính.

Phần cuối của đồ án sẽ trình bày một số đánh giá và hướng phát triển trong tương lai cho đề tài.

ABSTRACT OF THESIS

In the context of widespread Internet over the world, online data publication is getting more and more attractive. In comparison with paper based data publication, online data publication has more advantages in data updating rate, easier search for data through search engines, such as yahoo, google, livesearch, etc. Moreover, Clients don't have to store and maintain necessary data, they only have to ask online publisher for those data when they need.

On the other hand, online data publication has quite a few difficulties and challenges in development, especially the difficulties in building and running secure system and investing infrastructure of online publication. An approach to improve these difficulties will be presented in the graduation project.

Firstly, we will present an approach to online data publication scheme in the project. In the scheme, Data Owner doesn't directly publish their data to Clients, Publishers are employed to do this publication. A special thing is these publishers are not required to be trusted, because the scheme can allow the Clients to identify whether the answers for queries are true or not. To implement this, the scheme uses authentic data structure in online data publication. The scheme can relieve Data Owner's burden of providing publishing service to Clients. In addition, every publishers can join in this service, which leads to the increase in compatitiveness and as a result, the quality of service is not only improved but the cost of service is also reduced . Moreover, Clients also feel safer when using this service. Thanks to these advantages, the scheme can help to develop this potential online data publication.

Secondly, we will present the application of authentic data structure in online publication, in which the data of Data Owner and Publishers are controlled by relational database management system (DBMS). Besides, we will also give a general model for authentic data structure as well as a generalized model for query verification.

Lastly, we will do experiment in setting up publishing application based on the scheme presented in the project and a typical authentic data structure. Then, we will give some evaluations about the scheme and the plan of development in the future.

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG

DANH MỤC CÁC THUẬT NGỮ, TỪ VIẾT TẮT

Thuật ngữ	Ý nghĩa
Nhà xuất bản	Nhà cung cấp dịch vụ xuất bản trực tuyến
Khách hàng	Người sử dụng dịch vụ xuất bản trực tuyến
MHT	Merkle Hash Tree
VO	Verification Object – Đối tượng xác minh
Search DAG	Search Directed Acyclic Graph – Đồ thị có hướng tìm kiếm

CHƯƠNG 1. TỔNG QUAN VỀ BÀI TOÁN XUẤT BẢN TRỰC TUYẾN VÀ CẤU TRÚC DỮ LIỆU XÁC THỰC

1.1 Đặt vấn đề

Trước kia, nói đến xuất bản là nói đến việc xuất bản các cuốn sách, các tạp chí, các bài báo… trên giấy. Nhược điểm của loại hình xuất bản này là thông tin chậm, khó tìm kiếm khi cần, chi phí sản xuất tốn kém và rất khó bảo quản. Trong bối cảnh Internet phổ biến trên toàn thế giới, lĩnh vực xuất bản trực tuyến đang ngày càng trở lên hấp dẫn. So với xuất bản trên giấy, xuất bản trực tuyến (online publishing) có ưu điểm rõ rệt về tốc độ cập nhật thông tin, dễ dàng tìm kiếm qua các công cụ tìm kiếm như yahoo, google, livesearch…. Bên cạnh đó người sử dụng không phải mất công giữ gìn những thông tin họ cần mà khi cần họ lập tức có thể tìm đến nhà xuất bản trực tuyến để lấy thông tin. Vì đồ án chỉ đề cập đến các vấn đề trong xuất bản trực tuyến nên từ đây Người viết luận văn (NVLV) sẽ sử dụng từ xuất bản để chỉ xuất bản trực tuyến cho ngắn gọn.

Trong xuất bản trực tuyến, Nhà xuất bản (publisher) phải đáp ứng các yêu cầu về bảo mật của Người sử dụng dịch vụ xuất bản (client) như tính toàn vẹn (integrity), tính xác thực (authenticity), và tính không thể chối bỏ (non-repudation) của các thông tin xuất bản. Các yêu cầu trên đặc biệt quan trọng khi dữ liệu xuất bản là các dữ liệu quan trọng, ví dụ như các dữ liệu về tình hình tài chính, đầu tư, cổ phiếu…, các dữ liệu y tế như thuốc, phương pháp điều trị, các loại bệnh…, hay các dữ liệu của chính phủ như các quyết định, các nghị quyết, các luật, các thông tin hành chính… được Người sử dụng dịch vụ xuất bản (từ đây NVLV sẽ sử dụng từ Khách hàng cho ngắn gọn) để đưa ra các quyết định quan trọng mà chỉ cần một chút sai lệch trong thông tin sẽ gây ra hiệu quả hết sức nghiêm trọng. Bên cạnh đó, dịch vụ xuất bản cũng phải đáp ứng các yêu cầu về chất lượng, đặc biệt là tính mở rộng (scalability) và tính sẵn dùng (availability).

Trong mô hình xuất bản mô hình xuất bản truyền thống bao gồm hai bên: Chủ sở hữu dữ liệu (Data Owner) trực tiếp cung cấp dịch vụ xuất bản cho Khách hàng. Do các yêu cầu về bảo mật và chất lượng dịch vụ, Chủ sở hữu dữ liệu phải đầu tư rất tốn kém cho dịch vụ xuất bản trực tuyến. Điều này khiến Chủ sở hữu dữ liệu không muốn hoặc không đủ khả năng cung cấp một dịch vụ xuất bản trực tuyến như thế.

Giải pháp để giải quyết vấn đề của mô hình xuất hai bên là sử dụng thêm bên thứ ba là các Nhà xuất bản đáng tin cậy (trusted publisher). Nhược điểm của mô

hình sử dụng các Nhà xuất bản thông tin đáng tin cậy là yêu cầu tính đáng tin cậy của các nhà xuất bản.

Mô hình mà NVLV trình bày trong đồ án là một mô hình xuất bản trong đó không đòi hỏi tính đáng tin cậy của của các Nhà xuất bản (untrusted Publishers) bằng cách cho phép Khách hàng kiểm tra tính chính xác của kết quả truy vấn mà họ nhận được từ Nhà xuất bản mà họ sử dụng. Đây cũng là lý do mô hình này được gọi là mô hình xuất bản xác thực.

1.2 Các yêu cầu trong xuất bản trực tuyến

Phần trên đã nêu tầm quan trọng của các yêu cầu về bảo mật (tính toàn vẹn, tính xác thực, tính không thể chối bỏ) cũng như các yêu cầu về chất lượng dịch vụ xuất bản (tính mở rộng, tính sẵn dùng) trong xuất bản trực tuyến. Sau đây, NVLV sẽ giải thích chi tiết từng yêu cầu.

Tính toàn vẹn. Nội dung thông tin mà bên nhận nhận được phải không bị thay đổi so với nội dung thông tin mà bên gửi đã gửi. Yêu cầu này hết sức quan trọng, giả sử một nhà đầu tư cổ phiếu A muốn biết thông tin về tình hình cổ phiếu và các tình hình kinh tế hiện tại để quyết định mua cổ phiếu. Nhà đầu tư này lấy thông tin từ các công ty chứng khoán B qua internet. Các Hacker có ý đồ xấu can thiệp thành công làm thay đổi nội dung thông tin được gửi từ B đến A. A quyết định đầu tư dựa trên những thông tin sai lệch về tình hình cổ phiếu và tình hình kinh tế. Do đó, A có thể bị những tổn thất nặng nề.

Tính xác thực. Thông tin chỉ được gửi từ nơi mà người sử dụng thông tin yêu cầu. Giả sử một nhà đầu tư chứng khoán A yêu cầu công ty chứng khoán B cung cấp thông tin về tình hình cổ phiếu qua Internet. Nhưng một hacker C lại can thiệp vào quá trình trao đổi thông tin giữ A và B. C giả danh B gửi cho A những thông tin sai lệch về tình hình cổ phiếu. A tưởng những thông tin đó là do B (một công ty chứng khoán uy tín) cung cấp và dựa vào đó để quyết định mua bán cổ phiếu. Sau đó, A có thể chịu những thiệt hại nặng nề.

Tính không thể chối bỏ. Đảm bảo rằng người đã gửi thông tin M không thể chối bỏ rằng mình đã gửi thông tin M và người đã nhận thông tin M không thể chối bỏ rằng mình đã nhận thông tin M. Giả sử một nhà đầu tư chứng khoán A bị công ty X cung cấp (qua Internet) những thông tin sai lệch M về tình hình kinh doanh của X. A dựa vào những thông tin đó (M) để quyết định mua bán cổ phiếu của X và bị thiệt hại do những thông tin X đã cung cấp là không đúng với thực tế. Sau đó, A có thể kiện X vì đã đưa thông tin sai lệch gây thiệt hại cho A. Vấn đề là làm sao chứng minh được X đã cung cấp thông tin M cho A?

Tính mở rộng. Khả năng hệ thống có thể mở rộng để đáp ứng số lượng truy vấn ngày càng tăng. Giả sử một công ty chứng khoán B cung cấp dịch vụ tư vấn đầu tư chứng khoán. Dựa vào tình hình thị trường chứng khoán lúc đó, B thiết kế một

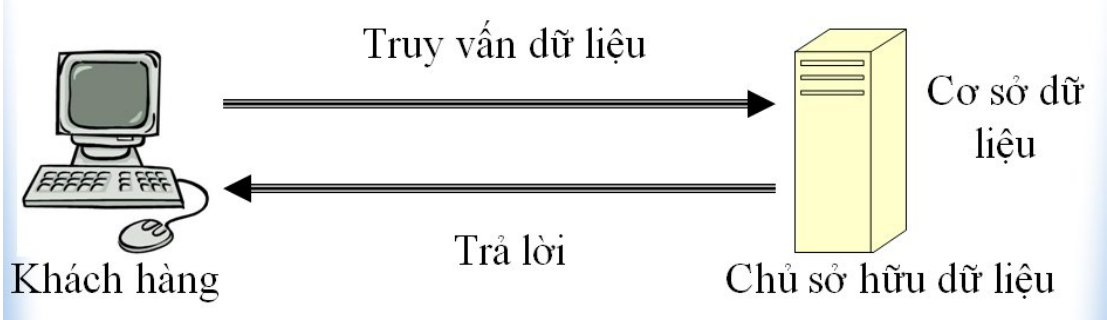
dịch vụ có khả năng đáp ứng khoảng 1000 lượt truy cập một ngày. Nhưng hiện nay tình hình thị trường chứng khoán tăng trưởng nhanh. Số lượng truy cập mỗi ngày tăng nhanh. Do đó để đáp ứng nhu cầu của các nhà đầu tư chứng khoán quan tâm đến dịch vụ của công ty, B phải xây dựng một dịch vụ mới mạnh hơn hoặc phải nâng cấp dịch vụ của B. Trong đó, việc nâng cấp dịch vụ sẽ ít tốn kém hơn so với việc phải xây dựng một dịch vụ mới toàn bộ.

Tính sẵn dùng. Hệ thống luôn luôn có khả năng cung cấp dịch vụ khi cần. Một dịch vụ dù tốt đến đâu, được đầu tư nhiều đến đâu cũng sẽ là vô nghĩa nếu người sử dụng không thể truy cập dịch vụ đó họ khi cần.

1.3 Các mô hình xuất bản trực tuyến

Trong phần này, NVLV sẽ mô tả chi tiết các mô hình xuất bản trực tuyến đã được nêu trong phần 1.1.

1.3.1 Mô hình xuất bản hai bên



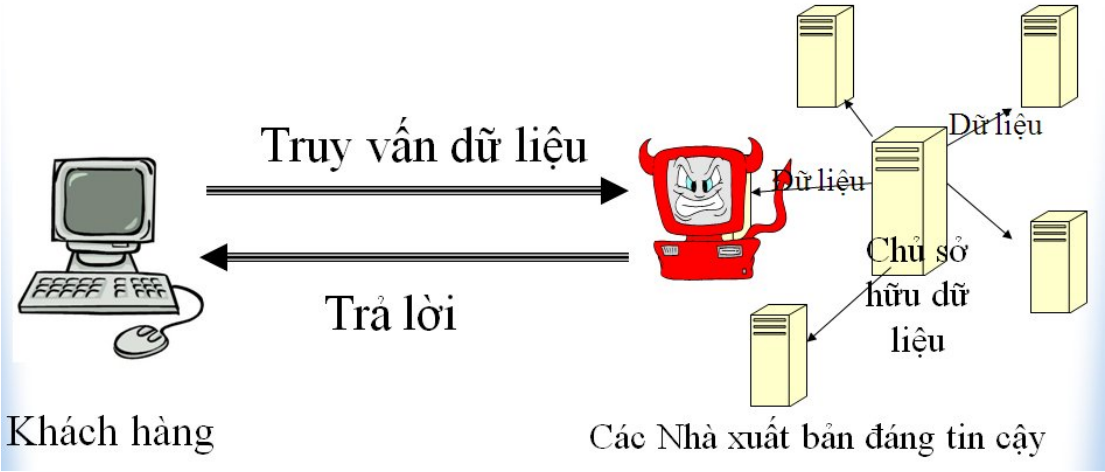
Hình 1. Mô hình xuất bản dữ liệu hai bên

Trong mô hình xuất bản hai bên, Khách hàng gửi trực tiếp yêu cầu truy vấn dữ liệu đến Chủ sở hữu dữ liệu. Chủ sở hữu dữ liệu xử lý truy vấn và trả lại kết quả cho Khách hàng. Kết quả truy vấn được đòi hỏi phải được trả về cho khách hàng một cách nhanh chóng, tin cậy và chính xác. Để đảm bảo các yêu cầu về bảo mật và các yêu cầu chất lượng dịch vụ trong xuất bản trực tuyến, Chủ sở hữu dữ liệu sẽ phải đầu tư một cơ sở hạ tầng về cả phần cứng lẫn phần mềm rất tốn kém cho dịch vụ xuất bản trực tuyến. Bên cạnh việc đảm bảo tính sẵn dùng cho dữ liệu và dịch vụ truy vấn, Chủ sở hữu phải đảm bảo tính không thể chối bỏ bằng cách ký vào câu trả lời cho câu truy vấn của Khách hàng. Do đó, Chủ sở hữu dữ liệu còn phải bảo mật chữ ký điện tử của mình một cách trực tuyến. Để cung cấp một dịch vụ thỏa mãn các yêu cầu trên, Chủ sở hữu phải đầu tư rất tốn kém. Do đó, mô hình này phát sinh vấn đề: Chủ sở hữu dữ liệu có thể không muốn hoặc không đủ khả năng cung cấp dịch vụ xuất bản có độ tin cậy cao, tốc độ nhanh và có khả năng đáp ứng số lượng truy vấn rất lớn mỗi ngày. Điều này hoàn toàn dễ hiểu nhất là khi Chủ sở hữu dữ

liệu không phải là một nhà xuất bản chuyên nghiệp, hoặc các thông tin mà Chủ sở hữu dữ liệu muốn cung cấp rất quan trọng nhưng không nhằm mục đích lợi nhuận cho Chủ sở hữu dữ liệu, ví dụ như một công ty dược tất nhiên không muốn đầu tư quá nhiều cho dịch vụ cung cấp trực tuyến các thông tin về thuốc của họ nhưng những thông tin này nếu bị sai lệch (do bị hacker tấn công) có thể gây nguy hiểm đến tính mạng con người; hay bất kỳ một công ty nào muốn công bố tình hình kinh doanh, cổ phiếu… của họ qua Internet tất nhiên sẽ không muốn đầu tư quá tốn kém cho công việc này nhưng những thông tin đó nếu bị sai lệch (do bị hacker tấn công) sẽ gây ra hậu quả cho các nhà đầu tư chứng khoán đã sử dụng những thông tin đó. Để khắc phục vấn đề này, giải pháp là sử dụng thêm bên thứ ba là các Nhà xuất bản đáng tin cậy.

1.3.2 Mô hình xuất bản sử dụng các Nhà xuất bản đáng tin cậy

Chủ sở hữu dữ liệu sẽ cung cấp dữ liệu cần xuất bản cho các Nhà xuất bản đáng tin cậy và chỉ phải cung cấp lại khi dữ liệu của Chủ sở hữu dữ liệu thay đổi. Chủ sở hữu dữ liệu cũng có thể ký lên dữ liệu trước khi gửi cho các Nhà xuất bản đáng tin cậy để đảm bảo tính không thể chối bỏ giữa Chủ sở hữu dữ liệu và các Nhà xuất bản này. Trong mô hình này, Chủ sở hữu dữ liệu không phải đầu tư vào dịch vụ xuất bản mà dịch vụ này sẽ được các Nhà xuất bản đáng tin cậy cung cấp. Các Khách hàng sẽ không lấy thông tin trực tiếp từ Chủ sở hữu dữ liệu mà sẽ lấy thông tin qua dịch vụ xuất bản của các Nhà xuất bản đáng tin cậy.



Hình 2. Mô hình xuất bản sử dụng các Nhà xuất bản đáng tin cậy

Ưu điểm của mô hình. Thứ nhất, Các nhà xuất bản chuyên nghiệp thường có hoặc sẵn sàng đầu tư một cơ sở hạ tầng xuất bản trực tuyến tin cậy, hiệu quả, bảo mật. Do đó, chất lượng của dịch vụ xuất bản được cung cấp bởi các nhà xuất bản chuyên nghiệp sẽ tốt hơn so với dịch vụ được cung cấp bởi một Chủ sở hữu dữ liệu

không phải là một nhà xuất bản chuyên nghiệp. Thứ hai, tính mở rộng và sẵn dùng của dịch vụ được tăng cường. Khi cần dữ liệu của Chủ sở hữu dữ liệu, Khách có thể truy cập đến bất kỳ Nhà xuất bản đáng tin cậy nào tham gia xuất bản dữ liệu của Chủ sở hữu dữ liệu. Do đó, chúng ta có thể coi tập hợp của tất cả dịch vụ xuất bản trực tuyến với cùng một dữ liệu của Chủ sở hữu dữ liệu của tất cả Nhà xuất bản đáng tin cậy là một hệ thống xuất bản trực tuyến. Chính vì vậy, một hệ thống xuất bản trực tuyến càng có nhiều Nhà xuất bản đáng tin cậy tham gia vào xuất bản dữ liệu của Chủ sở hữu dữ liệu thì tính mở rộng và tính sẵn dùng của toàn hệ thống đó càng tăng. Mặc nhiên càng có nhiều Nhà xuất bản đáng tin cậy tham gia vào một hệ thống xuất bản trực tuyến thì khả năng đáp ứng số lượng lớn truy vấn càng tăng. Khi số lượng truy vấn tăng, khả năng đáp ứng truy vấn của hệ thống có thể tăng bằng cách sử dụng thêm các Nhà xuất bản đáng tin cậy. Do vậy, tính mở rộng của hệ thống là cao. Ngoài tính mở rộng của toàn hệ thống có thể được tăng cường bằng cách tăng cường tính mở rộng của dịch vụ của mỗi Nhà xuất bản đáng tin cậy. Mặt khác, khi dịch vụ của một Nhà xuất bản đáng tin cậy trong hệ thống mà Khách hàng truy cập đến gặp sự cố, Khách hàng có thể ngay lập tức chuyển sang một Nhà xuất bản đáng tin cậy khác cùng xuất bản dữ liệu mà Khách hàng cần. Đây chính là lý do tính sẵn dùng của toàn hệ thống tăng.

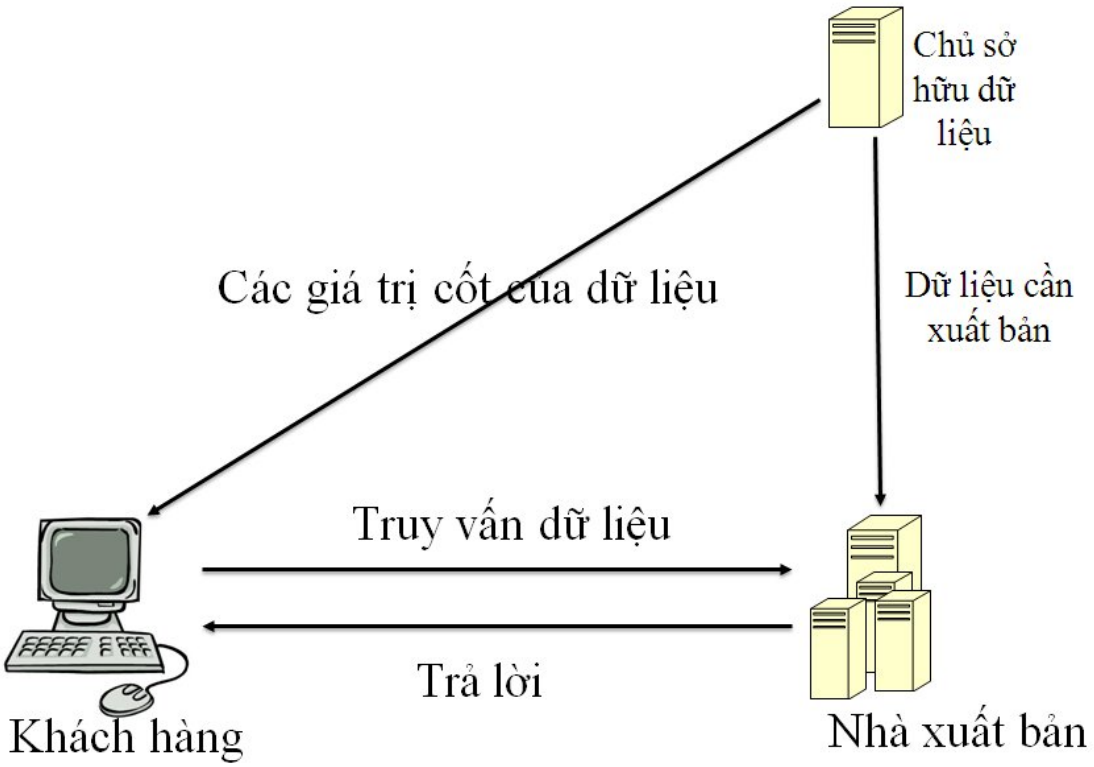
Nhược điểm của mô hình. Mô hình đòi hỏi các Nhà xuất bản tham gia xuất bản dữ liệu phải là đáng tin cậy. Sự đáng tin cậy của một Nhà xuất bản có thể do danh tiếng của chính Nhà xuất bản đó. Ví dụ, chúng ta "ngầm" hiểu dịch vụ xuất bản được cung cấp bởi các công ty có danh tiếng như FPT, VTC,…;hoặc các cơ quan tổ chức đáng tin cậy như Thông tấn xã Việt Nam, Nhà xuất bản khoa học kỹ thuật … là đáng tin cậy. Ngoài ra, sự tin cậy của một Nhà xuất bản cũng có thể xác minh bởi Chủ sở hữu dữ liệu. Trong trường hợp này, Chủ sở hữu dữ liệu vẫn phải bảo mật hệ thống xác minh sự đáng tin cậy của các Nhà xuất bản. Hơn nữa, yêu cầu sự đáng tin cậy của Nhà xuất bản sẽ hạn chế sự tham gia vào dịch vụ xuất bản trực tuyến của các Nhà xuất bản. Do đó, mặc nhiên tính mở rộng và tính sẵn dùng của toàn hệ thống cũng giảm. Mặt khác, yêu cầu này cũng làm giảm tính cạnh tranh về chất lượng cũng như giá thành trong lĩnh vực cung cấp dịch vụ xuất bản trực tuyến.

1.3.3 Mô hình xuất bản xác thực

1.3.3.1Giới thiệu

Sau đây, NVLV sẽ trình bày một mô hình sẽ khắc phục các nhược điểm trên bằng cách sử dụng bên thứ ba là các nhà xuất bản bất kỳ. Điều đặc biệt trong mô hình này là các nhà xuất bản thông tin không nhất thiết phải được xác nhận là đáng tin cậy (untrusted publishers) mà Khách hàng vẫn có thể biết được kết quả truy vấn họ nhận được từ các Nhà xuất bản là chính xác hay không. Chính điều này đã giải quyết tính mở rộng và tính sẵn dùng một cách hiệu quả và ít tốn kém. Việc bất kỳ

Nhà xuất bản nào cũng có thể tham gia vào xuất bản thông tin cũng làm tăng tính cạnh tranh trong xuất bản trực tuyến, góp phần làm tăng chất lượng và giảm giá thành của dịch vụ xuất bản trực tuyến. Trong mô hình này, Chủ sở hữu dữ liệu cũng không cần phải quan tâm đến vấn đề bảo mật chữ ký trực tuyến vì các thông tin mà chủ sở hữu dữ liệu cần gửi cho client hoàn toàn không bí mật và Chủ sở hữu dữ liệu chỉ cần trực tuyến (online) khi dữ liệu thay đổi của họ đòi hỏi phải gửi lại chữ ký cho Khách hàng và dữ liệu cần xuất bản cho các Nhà xuất bản sau đó có thể ngoại tuyến (offline). Việc bảo mật hệ thống của các Nhà xuất bản cũng được giảm bớt rất nhiều. Bởi vì, một kẻ xấu tấn công vào Nhà xuất bản sẽ không thể lừa gạt được Khách hàng sử dụng dịch vụ xuất bản của Nhà xuất bản đó. Đối với Khách hàng, họ có nhiều lựa chọn trong việc lấy thông tin từ nhà xuất bản nào. Khi Nhà xuất bản mà Khách hàng truy cập có vấn đề, họ có thể dễ dàng chuyển sang một Nhà xuất bản bất kỳ khác cũng xuất bản thông tin đó. Sự phát triển của các công cụ tìm kiếm



Hình 3. Mô hình xuất bản xác thực

trên Internet giúp cho việc tìm kiếm nhà xuất bản xuất bản thông tin mà Khách hàng cần trở nên dễ dàng.

1.3.3.2Các bước cơ bản trong mô hình xuất bản xác thực

Mô hình xuất bản trực tuyến sử dụng các Nhà xuất bản bất kỳ được mô tả khái quát như sau (Hình 3):

1) Chủ sở hữu dữ liệu xây dựng các cấu trúc dữ liệu xác thực cho các dữ liệu cần xuất bản và tính toán các cốt (digest, ∑s) của các cấu trúc dữ liệu này sử dụng một hàm băm phi đụng độ (One-Way hash function). Giá trị cốt ∑ này liên quan tới toàn bộ dữ liệu trong cấu trúc dữ liệu do đó khi có bất kỳ dữ liệu nào trong cấu trúc dữ liệu thay đổi thì giá trị cốt của cấu trúc dữ liệu đó cũng thay đổi.
2) Chủ sở hữu dữ liệu gửi các cốt ∑s đã tính đến Khách hàng theo một giao thức bảo mật và gửi dữ liệu cần xuất bản cho các Nhà xuất bản.
3) Khi Khách hàng cần sử dụng dữ liệu của Chủ sở hữu dữ liệu, Khách hàng sẽ gửi truy vấn q đến một Nhà xuất bản bất kỳ xuất bản dữ liệu của Chủ sở hữu dữ liệu. Khách hàng sẽ tính toán câu trả lời A cho câu truy vấn q. Bên cạnh đó, Nhà xuất bản cũng tính toán một đối tượng xác minh (Verification Object - VO) cho q. Sau đó, Nhà xuất bản sẽ gửi cả A và VO cho Khách hàng.
4) Khách hàng sử dụng đối tượng xác minh mà Nhà xuất bản cung cấp để tính lại giá trị cốt và so sánh với cốt ∑ đã được cung cấp bởi Chủ sở hữu dữ liệu để xác minh câu trả lời A được cung cấp bởi Nhà xuất bản có đúng với dữ liệu của Chủ sở hữu dữ liệu hay không. Trường hợp hai giá trị này khác nhau, Khách hàng ngay lập tức có thể kết luận kết quả truy vấn A là giả mạo.

Đặc trưng của mô hình. Mô hình dựa vào khả năng xác thực dữ liệu của các cấu trúc dữ liệu xác thực. Độ an toàn của mô hình xuất bản xác thực phụ thuộc vào độ an toàn của hàm băm được sử dụng để tính cốt của các cấu trúc dữ liệu xác thực trong mô hình. Do đó, yêu cầu hàm băm phải an toàn, tức là tính phi đụng độ cao. Ngoài ra, Đối tượng xác minh phải có kích thước nhỏ. Đối tượng xác minh đảm bảo rằng câu trả lời là đúng: câu trả lời bao gồm tất cả các bản ghi đã yêu cầu và không có bản ghi nào dư thừa hay bị bỏ sót. Chi phí để Nhà xuất bản tính toán Đối tượng xác minh và chi phí Khách hàng xác minh kết quả truy vấn phải thấp. Một câu trả lời hay một Đối tượng xác minh sai sẽ bị Khách hàng phát hiện. Nhà xuất bản không thể giả mạo một Đối tượng xác minh hợp lệ cho một câu trả lời truy vấn sai. Bên cạnh đó, chúng ta có thể thấy rằng mô hình này áp dụng đặc biệt hiệu quả với những hệ thống trong đó ít có sự thay đổi dữ liệu của Chủ sở hữu dữ liệu, tức là hiệu quả đối với hệ thống có dữ liệu tương đối tĩnh. Nói cách khác, dữ liệu càng ít thay đổi Chủ sở hữu dữ liệu càng ít phải tính toán lại giá trị cốt của dữ liệu, Nhà xuất bản càng ít phải xây dựng lại các cấu trúc dữ liệu xác thực, và sự truyền dữ liệu của Chủ sở hữu dữ liệu cho Khách hàng và Chủ sở hữu dữ liệu cho Nhà xuất bản càng giảm.

1.4 Xác định nội dung cụ thể của đồ án

Trong đồ án, NVLV tập trung nghiên cứu, tổng hợp các các cấu trúc dữ liệu xác thực và các phương pháp xác thực câu trả lời cho truy vấn. Đây là những vấn đề

côt lõi trong mô hình xuất bản xác thực. Việc lựa chọn cấu trúc dữ liệu cùng với phương pháp xác thực đi kèm cấu trúc dữ liệu xác thực đó có ảnh hưởng quyết định đến hiểu quả của mô hình xuất bản xác thực. Do đó, NVLV sẽ dựa vào những nghiên cứu về mặt lý thuyết để lựa chọn một cấu trúc dữ liệu xác thực điển hình đảm bảo tất cả các yêu cầu của mô hình xuất bản xác thực, kèm theo nó là một phương pháp xác thực kết quả truy vấn đảm bảo chính xác. Cấu trúc dữ liệu này cũng đảm bảo hiệu quả trong thực nghiệm. Trong quá trình nghiên cứu, NVLV thấy cấu trúc dữ liệu Merkle Hash Tree[1] thỏa mãn tất cả những yêu cầu của xuất bản xác thực. Do phần lớn dữ liệu hiện tại được lưu trữ dưới dạng mô hình quan hệ và được quản lý bởi các hệ cơ sở dữ liệu quan hệ như MS SQL Server, PostgreSQL, MySQL …nên trong đồ án, NVLV sẽ đề cập chi tiết đến vấn đề ứng dụng của cấu trúc dữ liệu xác thực Merkle Hash Tree vào việc xuất bản xác thực cơ sở dữ liệu quan hệ. Cụ thể, NVLV sẽ đề cập việc ứng dụng Merkle Hash Tree để xử lý các dạng truy vấn phổ biến, được công thức hóa bằng các phép toán quan hệ trong đại số quan hệ như: phép chiếu, phép chọn, phép kết nối …, và phương thức xác minh câu trả lời cho các dạng truy vấn đó.

Cuối cùng, NVLV sẽ áp dụng những nghiên cứu về mặt lý thuyết để cài đặt hệ thống xuất bản xác thực dữ liệu hành chính địa lý của các đơn vị hành chính Nhà nước. Lý do NVLV chọn hệ thống xuất bản thông tin hành chính địa lý là vì đây là hệ thống mang đầy đủ những thể hiện về mặt lý thuyết của mô hình xuất bản xác thực và là một hệ thống xuất bản dữ liệu điển hình. Tức là, đây là một hệ thống khá phổ biến với Chủ sở hữu dữ liệu là Chính phủ do đó đảm bảo dữ liệu cung cấp cho Khách hàng là đáng tin cậy. Bên cạnh đó, những dữ liệu về hành chính địa lý thường khá tĩnh, phù hợp với mô hình xuất bản xác thực. Dịch vụ xuất bản dữ liệu hành chính địa lý cũng có đầy đủ các dạng truy vấn mà bất kỳ dịch vụ xuất bản dữ liệu nào cũng phải sử dụng. Do đó, những thiết kế, cài đặt cho hệ thống xuất bản dữ liệu hành chính địa lý có thể sử dụng cho hầu hết các hệ thống xuất bản thông tin khác.

1.5 Bố cục của đồ án.

Với những nội dung đã nêu trong mục 1.4, phần còn lại của đồ án sẽ được cấu trúc như sau:

Chương 2 sẽ trình bày về cấu trúc dữ liệu Merkle Hash Tree và phương thức xác thực theo hướng từ dưới lên (bottom-up).

Chương 3 sẽ trình bày ứng dụng của cấu trúc dữ liệu xác thực Merkle Hash Tree trong việc xuất bản xác thực dữ liệu từ CSDL quan hệ.

Chương 4 sẽ trình bày một mô hình chung của các cấu trúc dữ liệu xác thực và phương thức xác thực theo hướng từ trên xuống (top-down). Chương 3 cũng

trình bày một số cấu trúc dữ liệu xác thực điển hình và phương thức xác thực từ trên xuống trên các cấu trúc dữ liệu đó.

Chương 5 sẽ trình bày về thiết kế và cài đặt thử nghiệm mô hình xuất bản xác thực dữ liệu hành chính địa lý của đang đơn vị hành chính Nhà nước.

Phần cuối của đồ án sẽ trình bày một số đánh giá và hướng phát triển trong tương lai cho đề tài.