CS437 XML Webdata Management

Semester Project: Secure XML Publishing

Phase 1 Report

Aaron Powers. Tom Guenther. Matt Nuckolls

Problem Statement

XML provides a versatile way to store data. Data stored in an XML format can be published on the web for distribution purposes. This allows for a convenient method of distributing data to other individuals interested in the data. Naturally, some of this data may be sensitive or may only be intended for a certain individual to read. A secure method of publishing XML data is needed to prevent unauthorized access. The method of secure publishing should support the ability to verify a receiving individual’s allowance access as well as preventing the individual from seeing portions of the document they are not intended to see. Additional features that would be beneficial would be for the publisher to have the ability to publish only changes to an existing document instead of republishing the entire document as well as the receiving individual having the ability to only receive changes to the document relevant to the portions they are allowed to access instead of re-downloading the entirety of the document.

Analysis of the Problem

XML is a flexible markup language. It is intended to be both machine redable as well as human readable. Though it has a specific structure defined by tags, elements, attributes, etc. it has no specific format nor requirement for tag sets. This flexibility makes it versatile but also vulnerable.

Since XML simply a markup language, it has no intrinsic security mechanisms. Therefore, XML data storage and transmission security must be handled externally to the XML content. XML document encryption is necessary for secure storage and transmission. Furthermore, XML document access must also be handled externally. Again, XML has no intrinsic access controls.

Since access and transmission control must be handled externally to the XML document (implicitly: the XML parser), a “3rd party” application may be required to authorize access over a secure transmission protocol such as SSL (secure socket layer).

Possible Solutions

One possible solution is to implement a publish / subscriber middle layer that handles all transactions between the publisher (entity containing the data) and the subscriber. The middle layer would serve as a gateway to provide secure access to the stored XML data. At some level it will be assumed that the content provider who originates the data to be published will take reasonable precautions to ensure that the data is presented to the publisher in a secure way. The content provider will ensure that the middle layer is informed of any document access restrictions and constraints under which the publisher must act. At this point we don't have a clear picture of what this middle layer will look like. More effort will be spent on fleshing out the details. At the present time this is the solution we are exploring.

Review of Related Papers

Currently, there is a large amount of work being done in this area. Mohamed Nabeel and Elisa Bertino published a paper titled “Secure Delta-Publishing of XML Content”. In this paper they propose a method of secure XML publishing that supports small incremental changes without the publisher needing to republish the entire paper. Their method also allows that the third party publishers do not need to be trusted due encryption used by the original publisher.

Erwin Leonardi, Sourav S. Bhowmick, and Mizuho Iwaihara published a paper titled “Efficient Database-Driven Evaluation of Security Clearance for Federated Access Control of Dynamic XML Documents”. The authors of this paper propose a method of Secure XML distribution that utilizes a “policy enforcer” and “data provider”. Essentially, the data provider stores the XML documents as well as the access policy of those xml documents. A user can query the policy enforcer for data from the XML documents which the policy enforcer passes on to the data provider. The data provider runs the query then returns to the policy enforcer the documents found in which the user is allowed to access.

A sophisticated approach to the problem of secure xml publishing was presented in a paper by Mohammad Ashiqur Rahaman, Yves Roudier, and Andreas Schaad titled “A Publish/Subscribe Model for Secure Content Driven XML Dissemination”. This paper presents a very detailed example of a publish/subscribe model and how it could be used in a real world scenario between different organizations. This method ensures that the XML schema is confidential as well as the data itself, ensures the integrity of transmitted data is and removes the coupling of publishers with subscribers by inducing a dissemination layer between the publishers and subscribers. The method also ensures the xml data cannot be read by the dissemination layer. In this model a publisher first establishes authorization policies that are enforced by the dissemination network then a user sends a subscription request along with his credentials to the dissemination network. When a publisher publishes his encrypted XML document he annotates it with information regarding the “concept” of the document and the dissemination layer forwards the document to the subscribed users. In this model, users only get portions of the document they are authorized for, they can get them automatically by distribution from the dissemination layer, there is confidence in the integrity of the data, and the data is kept hidden from anyone that is not authorized to see it. This model in and of itself seems to solve many of the problems surrounding secure XML publishing as well as providing efficiency within the dissemination network layer. It does not appear that this model supports incremental updates from publishers.

Another method has been proposed by B. Carminati, E. Ferrari, and E. Bertino in their paper “Secure Third Party Distribution of XML Data”. This is yet another method that provides confidentiality, integrity, and authenticity in secure XML data publishing but it also provides completeness. Completeness is provide by ensuring the user that they are receiving every portion of the document that they are allowed to receive based on their access control policy. Like the method proposed by Rahaman, Roudier, and Schaad, this method does not require that the third party publishers be trusted, however data still remains confidential. This is achieved by the user receiving keys directly from the data publisher.