Secure Semantic Web Service Using SAML
JOO-YOUNG LEE and KI-YOUNG MOON
Information Security Department
Electronics and Telecommunications Research Institute
161 Gajeong-dong, Yuseong-gu, Daejeon
KOREA

Abstract: - The Semantic Web’s success will depend on the implementation and use of Web service, becoming preeminent in E-commerce, which will likely be agent-based in the future. Agents use the large amount of information available over the web, but it will allow agents more subtle attacks. Agents and supporting technologies need to be secure and reliable for safe Web service as they adopt more mission critical roles within Web services. In this paper, we’ll discuss agent security on the Semantic Web. On the Semantic Web, agents play different roles in different platforms, dynamically change their access requirements, and act on behalf of users with different access privileges. In order to make agents secure, we’ll propose a method that agents dynamically exchange their authentication and authorization information using SAML. And then, a few threats that may be happened in transit and countermeasures will be presented. Finally, we’ll conclude our talks by presenting further works.

Key-Words: - Semantic Web, SAML, Security, Agent, Web Service

1 Introduction
People currently share their knowledge on the Web in the language intended for other people. This knowledge is potentially of great value for commercial as well as private users. But unfortunately, relevant information is often hidden amongst vast amount of irrelevant information that is also available on the Web. In recent years, researchers have begun to explore the potential of associating web content with explicit meaning so that the web content becomes more machine-readable and intelligent agents can retrieve and manipulate pertinent information readily. Semantic Web proposed by W3C is one of the most promising and accepted approaches. On the Semantic Web, we are able to express ourselves in terms that our computers can interpret and exchange. By doing so, we will enable them to solve problems that we find tedious and to help us find quickly what we’re looking for, such as a movie review, a book purchase order, medical information, etc.

The Semantic Web’s success will depend on the implementation and use of Web service, becoming preeminent in E-commerce, which will likely be agent-based in the future. Using intelligent collaborations, agents can achieved global optimization while adapting to local requirements. This approach will let agents use the large amount of information available over the web, but it will allow agents more subtle attacks, for example, malicious users and their agents can disclose sensitive information or sabotage the information of others. Therefore, agents and supporting technologies need to be secure and reliable for safe Web service as agents adopt more mission critical roles within web services.

In this paper, firstly, we’ll discuss the security of agent, which is an important component of the Semantic Web technologies. On the Semantic Web, agents play different roles in different platforms, dynamically change their access requirements, and act on behalf of users with different access privileges, so they bring some security concerns. Next, in order to make agents secure on the Semantic Web, we’ll propose a method that agents dynamically exchange their authentication and authorization information using Security Assertion Markup Language (SAML), ratified by the OASIS standards. And then, a few threats that may be happened in transit and countermeasures will be presented. Finally, we’ll conclude our talks by presenting further works.

2 Overview of the Semantic Web
The Semantic Web, proposed by W3C as the next generation web, is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in
cooperation [1]. A key element to realizing the Semantic Web is developing a suitably rich language for encoding and describing Web content. Such a language must have a well-defined semantics, be sufficiently expressive to describe the complex interrelationships and constraints between Web objects, and be amenable to automated manipulation and reasoning with acceptable limits on time and resource requirements. Several languages build on XML. These include the Resource Description Framework (RDF), RDF Schema, DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer), and most recently, the Web Ontology Language (WOL).

RDF integrates a variety of applications using XML for syntax and URIs for naming and is a foundation for processing metadata. It provides interoperability between applications that exchange machine-understandable information on the Web. RDF uses XML to exchange descriptions of Web resources and emphasizes facility to enable automated processing. DAML+OIL and OWL are Web ontology languages based on artificial intelligence knowledge representation work in description logics. They provide a natural way to describe class and sub-class relationships between Web vocabulary, as well as constraints on the relationships between classes and between class instances.

2.1 Ontology and Agent
The ontology is a logical theory accounting for the intended meaning of a formal vocabulary. Research on ontology is becoming increasingly widespread in the computer science community. In the semantic web, the ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has taxonomy and a set of inference rules. Ontologies can enhance the functioning of the Web in many ways, and RDF Schema, DAML+OIL and OWL supply the language to define the ontologies.

Agent-based computing can potentially help us recognize complex patterns in the widely distributed, heterogeneous, uncertain information environment. To make agent understanding and interacting with data, we have to provide marked annotations attached to data sources. It makes information available to the agents in a new and exciting way. These kinds of solutions usually involve ontologies. The most typical type of ontology used in building agents involves a structural component. Essentially a taxonomy of class and subclass relations coupled with definitions of the relationships between these things. In addition, the ontologies contain some kind of inference rule. These can be explicit rules about an item, or they can be structural inferences provided by a system. If an ontology can be made machine readable, it allows a computer to manipulate the terms used in the ontology. Figure 1 shows the framework for agents on the Semantic Web.

![Fig. 1. Framework for agents on the Semantic Web](image)

One of the most powerful uses of the Web ontologies and a key enabler for agents on the Web is in the area of Web Services [2]. We can create machine-readable ontologies used by capable agents to find web services and automate their use. Using a combination of web pointers, web markup, and ontology languages, we can do even better than just putting service advertisements into ontologies. Rather, using these techniques we can also include a machine-readable description of a service and some explicit logic describing the consequences of using the service. In fact a service description and a service logic would lead us to the integration of agents and ontologies in some exciting ways. By using these two properties, agents in the Web Service environment can mediate among various heterogeneous Web sites, monitor contents and notify customers and perform precision information filtering and comparison. Additionally, they can provide tailored services according to the specialized needs of customers, assisting customers in making decisions, and act on behalf of customers in matchmaking, server monitoring, negotiation, transfer of goods, and follow-up support.

3 SAML
The OASIS Security Assertion Markup Language (SAML) specification is an industry standard for the extensible Markup Language (XML) for exchanging
authentication, attribute, and authorization information [3]. Such exchanges occur, for example, between interacting applications that do not share the same underlying authentication and authorization infrastructure. The differences may be organizational or platform-based. In any case, SAML enables secure Web single sign on (SSO) to applications between different systems and platforms. SAML statements are called assertions. They are represented as XML constructions and have a nested structure, whereby a single assertion might contain several different information items referring to authentication, authorization decisions, and attributes such as credentials or group membership designator. It is becoming an important standard for federating user identities and it is applicable to both enterprise and value chain environments that need to leverage a user identity that can span traditional security domain boundaries.

4 Security on the Semantic Web

To provide security for Semantic Web services, security models and tools for the underlying technologies, such as XML, RDF, DAML-OIL, need to be developed [4]. Especially, because intelligent agents, enabled to access and process large amounts of Web data, will be able to discover information that might be confidential, the features of interoperation and machine-enabled information processing ability may increase the danger.

Guaranteeing information security means protecting information against unauthorized disclosure, transfer, modification or destruction, whether accidental or intentional. To realize it, any operation should only be accessible to properly authorized agents, proper identity of the agent must be reliably established by employing known authentication techniques, and sensitive data must be encrypted for network communication and persistent storage.

Various security technologies for the web exist at present. These technologies have to be evaluated for the semantic web. Additionally it needs to incorporate security semantics into semantic interoperability. The various logics being developed for the semantic web need to be examined and security properties have to be incorporated. In this section, we will elaborate on the security for agent, which is necessarily required for secure Semantic Web.

4.1 Agent Security

Agent technologies were designed with a focus on interoperability, distributed problem solving, and cooperation. Multiagent systems were intended to be responsive to open environments, such as the Internet, to capitalize on cooperative interaction. But this readiness to interact leaves them vulnerable to agents with harmful intentions. Protecting agent systems requires secure communication protocols and authentication and authorization mechanisms for proper access control, and distributed trust management [5].

4.2 Traditional Agent Security Systems

There are a few applications deploy security measurement. Among of them, we’ll elaborate on the approach of JADE [6], which is principal-based, meaning that it relies on Java security policies and mechanisms for human users. Because a JADE agent platform can be distributed across several hosts and can support the interactions of autonomous entities, each component must belong to some user who is responsible for its actions. Hence, a JADE platform must know all users and be convinced of their authentication. The user in this system must be authenticated on the JADE platform by providing a valid username and password. A JADE policy, along with the permissions of the owners, determines which actions the agent can perform.

Such traditional user and principal based security systems are likely to be insufficient to secure agent-based systems. Agents can play different roles in different platforms, dynamically change their access requirements, and act on behalf of users with different access privileges. Without relying on a centralized trust management system, agents must be able to decide whether or not to trust other agents or platforms.

5 SAML for Agent Security on the Semantic Web

The Organization for Advanced of Structured Information Standards (OASIS) has defined the XML-based Security Assertion Markup Language. It adopted as a standard the exchange of authentication and authorization information. One of the major design goals with SAML is single sign on, which allows a user to authenticate a domain and use
resources in other domains without reauthenticating. This standard is a significant protocol for agents as these agents engage and interact with resources and other agents in various domains, because the SAML support dynamical authentication and authorization mechanism. In this section, we’ll present an agent security scenario based on the SAML and discuss several threats that can be happened during communication and countermeasures for them.

5.1 Scenario
This scenario is the elaboration of a kind of communications between agents or the agent and platform through attribute exchange with authentication reference called an artifact. An artifact is “small” bounded size of byte string. When artifact is conveyed to the source site the artifact unambiguously references an assertion. The artifact is conveyed via redirection to the destination site, which then acquires the referenced assertion by some further step.

An agent that wants to access to information or a resource on a site gets an artifact that asserts it is reliable from a trusted party (Step 1). It searches and requests what it wants, providing the artifact to a destination (Step 2). Then the destination requests the trusted party an authentication document called “assertion”, passing the artifact. In this step, the site also authenticates itself with its certificate that is digitally signed by a certification authority (Step 3). If the artifact is valid, the trusted party returns the authentication document of agent A. This document includes authentication event descriptions and authorization attributes about the agent (Step 4). Finally the site allows the agent to access on information or resource (Step 5). Figure 2 shows a secure communication scenario between an agent and a site using SAML.

This method basically exchanges authentication and authorization information. It allows protecting not only agent execution environment from a malicious agent, but also an agent from a malicious platform by bilateral authentication. In addition, it ensures that the agent is authorized to be access on limited information or resources by sending an attribute assertion. On the Semantic Web, agents can play different roles in different platforms, dynamically change their access requirements, and act on behalf of users with different access privileges. Our proposal enables agents to be able to decide dynamically whether or not to trust other agents or platforms without relying on a centralized trust management system.

5.2 Threat model and countermeasures
In the upper scenario, we can infer a few threat models. In the first place, we assume that a valid agent requests a malicious site some information. After the site allows access to a resource, it stores the artifact sent by the agent A for other purpose. It may impersonate the agent with the artifact on some new site or to agent. To alleviate this concern, the artifact needs to have shortest possible validity period consistent with successful communication of the artifact from source to destination site. This is typically in the order of a few minutes. This ensures that a stolen artifact can only be used successfully within a small time window. The second case is that a malicious agent asks for some information of a site with other’s artifact. A countermeasure is that it makes the trusted party determine if an artifact is being requested by more than one site or agent or not. In such a case, the trusted party must not provide the assertions to the site, then the site will not permit the access.

Finally, the use of SAML assumes and requires trust between participants, but the SAML protocol does not include provisions to establish or guarantee this trust. SAML is not concerned with guaranteeing confidentiality, integrity, or nonrepudiability of the assertions in transit. For these purpose, it needs XML Enc and XML Dsig or other mechanism provided by the underlying communication protocol and platform.
6 Conclusion and Further Works
The Semantic Web will make the Web more accessible to agents by making use of semantic constructs, such as ontologies represented in well-established languages, so that agents can understand what is on a page. And as they are highly desirable for interoperability, this feature is scary from the security perspective. Illegal interfaces, supported by Semantic Web technologies and ontologies, might enable user to access unauthorized information.

In this paper, we have proposed a method that agents exchange their authentication and authorization information using SAML so as to make them secure on the Semantic Web. Because SAML allows exchanging authentication and authorization information between interacting applications that do not share the same underlying security infrastructure, it is suitable to protect agents, which has various role and responsibility in different system with their dynamic access requirements. Our method not only protects the agent against the some sites and other agent, but also guards the host against the malicious agents.

This paper provides only preliminary method for making agents secure on the Semantic Web using SAML. Therefore, we have something to do for increasing security. Most of all the technologies that make up the Semantic Web have to be secure. These include not only agents, but also XML, RDF, the infrastructures, information and data management technologies. And the security is also preserved when integrating data and the technologies. Next, we need to develop the semantic web profile and binding protocol of SAML. Currently, SAML has provided Web browser profile and SOAP binding protocol. Well-defined profile and binding protocol will provide agents on the Semantic Web interoperability as well as security.

References: