Proceedings of I-KNOW '05
Graz, Austria, June 29 - July 1, 2005

Security Enhanced Distributed Knowledge Management Architecture

Petros Belsis
(Department of Information and Communication Systems Engineering
University of the Aegean, Karlovasi, Samos, Greece
pbelsis@aegean.gr)

Stefanos Gritzalis
(Department of Information and Communication Systems Engineering
University of the Aegean, Karlovasi, Samos, Greece
sgritz@aegean.gr)

Christos Skourlas
(Department of Informatics, Technological Education Institute of Athens, Greece
cskourlas@teiath.gr)

Abstract: From the emergence of Knowledge Management (KM) systems until today, the vast majority of the implemented systems have concentrated on a centralised architecture, which utilizes knowledge within a single organizational domain. Lately, there is much focus on alternative, distributed approaches, which attempt to overcome the single organization's KM paradigm, and develop inter-organizational knowledge exchange infrastructures. Effective management from a security perspective is always a serious and difficult to achieve challenge, especially when it comes to managing resources from cooperating autonomous domains. Security issues in Information Systems coalition enabling environments are treated in this paper and a secure distributed KM architecture is being presented.

Keywords: Distributed Knowledge Management, Security policies
Categories: H.4

1 Introduction

Traditional Knowledge Management (KM) systems fail to utilize knowledge that resides in other organizations. Their centralized orientation and their closed architecture, prevent from exploiting adequately organization's intellectual capital, to its full extent. Apparently, the limitations of centralized approaches are not technological, but mainly organizational. They create a mismatch between social and technological architectures [Bonifacio, 2002], which limit knowledge acquisition and dissemination possibilities. The advent of networked infrastructures on the other hand, arises as a challenge for cooperation between Information Systems, in terms of knowledge exchange and diffusion. Even though distributed architectures in KM emerge as a promising solution, they pose a number of serious problems, concerning the efficient and secure management of jointly owned resources. The complexity of the problem resides mainly in the contradictory requirements of such a task: on one hand there is a necessity for sharing and collaborative contribution to the knowledge assets, on the other there is a necessity to protect knowledge assets so that they will be accessed only by those who retain legal access rights. To make things harder, there is often a tendency of partner organizations to minimize their willingness for sharing resources, which complicates the management of knowledge assets, especially when it comes to automating the negotiation procedures.

In this paper, we present a distributed, resilient KM architecture characterized by its scalability potential and concurrently the capability to handle knowledge assets exchange between autonomous domains. The rest of the paper is organized as follows: Section 2 presents related work, section 3 presents the basic notions of policy based security frameworks and presents the basic security concepts adopted by our approach and gives an overview of our architecture, while section 4 concludes the paper.

2 Related work

The benefits and the obvious superiority of distributed Knowledge Management (DKM) systems have attracted considerable interest to both the academic community as well as to the industry. A large number of systems have emerged, others focusing on the expansion of knowledge sharing capabilities, others emphasizing on authentication schemes.

XAROP [Tempich, 2004] is a peer-to-peer system, which manages heterogeneous knowledge sources through the usage of ontologies. Due to the technical challenges when handling multiple heterogeneous resources, the notion of security has not attracted the main focus of research within the XAROP framework. Classification of documents requires from the user to manually classify users and groups. This approach raises difficulties to the user, while it poses concerns relative to retaining the system's scalability potential. Moreover, storing access permission details for each document within such a distributed system becomes an untreatable problem.

ADAM [Seleznyov 2004] is a distributed system, which utilizes trust based negotiation procedures for the establishment of transactions between users. Its architecture is agent-based, where one agent is responsible for gathering knowledge from distributed nodes and a second agent is responsible to handle authorization procedures on behalf of the user. ADAM manages mainly knowledge about its users and bases the authorization process on grounds of reputation collected for a user from other nodes. Even though it handles scalability issues very efficiently, this system gives the chance to somebody to create a new identity or retain multiple identities concurrently and attempt to enter into relations with the system. The application of these principles on systems such as Internet transactions (e-commerce) - where a security failure could direct to short-scale financial loss - is not doubted relative to its validity. ADAM authorizes transactions and not users. Furthermore, it functions on total absence of explicitly stated organizational policy.

SemanticLIFE [Weippl, 2004] is a project that stores an individual's entire digital life and makes it available to co-workers. The security scheme effectuation is based mainly on implementing role-based access controls through the usage of database systems. Security policies, as utilized in our architecture, offer a more flexible and much robust way for security administration, while they can provide the system with greater scalability potential without lacking in attack resistant metrics.

Our approach emphasizes on providing a resilient and concurrently robust system, which is characterized by a high scalability potential, while it provides a flexible means for managing distributed knowledge based infrastructures. In order to handle security issues effectively, security policies have been conceptualized and a mapping between organizational structure and security related duties has been applied, while by making use of the XACML [XACML] policy language, user credentials and her permissions together with context related details can be encoded in XML and transferred from one node to another, minimizing for the user the necessary authentication procedures, and providing to her transparent access to knowledge assets that her security clearance level allows, no matter if they belong to the same organization or they have been retrieved from a cooperating domain.

3 Security policy based framework for managing distributed infrastructures

Various approaches emerged lately in the area of distributed computing, relative to access control enforcement. Among other models, the Role Based Access Control Model (RBAC) [Sandhu 2000] seems to be the more prominent. Key RBAC advantages are: simplification of security-related user management; scalability; correspondence of the security mechanisms with organizational structure (organizational roles map easily to security roles and to the establishment of security related role attributes).

Implementation of security management relies often on security policies, which have concentrated much focus on the area of distributed computing. By adopting a policy-based approach, we facilitate the management of resources, which span over a number of different organizational domains. Furthermore, policies are interpreted by automated agents and their behaviour can be modified dynamically. The usage of security policies is a necessity in distributed systems, since access control mechanisms have to be applied across many heterogeneous components, which enforce authorization mechanisms for a variety of target objects. When we consider several autonomously managed domains which cooperate on grounds of knowledge sharing, security management becomes more complicated, expensive and error prone; different domains have different demands relative to the levels of security restrictions. Therefore, through a policy based approach, we simplify security management and we provide flexible intra-organizational knowledge assets exchange.

In our system we handle several KM related activities:

- Knowledge assets discovery, which is handled by security agents and is based on the use of an appropriate ontology,
- Cross Lingual Retrieval capabilities
- Authorization process, which identifies according to the user that requests access to an asset if he has the appropriate level of classification
- Negotiation, when it comes to inter-organizational knowledge transfer
- Semantic based Retrieval using Meta data

3.1 System architecture

Our architecture is materializing a distributed organizational memory. An Organizational Memory (OM) comprises a variety of information sources where information elements of all kinds, structures, contents and media types are available [Abecker 2003]. In addition, a distributed OM utilizes knowledge from interconnected domains, representing knowledge assets in a location independent form. Several instances of an organizational memory are established in different organizational domains and are stored on local nodes.

The organizational memory module supports storage and retrieval for semi-structured documents with multilingual support [Belsis 2004b]. Organizational experience is being codified by subject in semi-structured documents, which consist of raw text (in two languages), brief abstract description and keywords in order to facilitate retrieval. The system also attempts to provide support for tacit knowledge exploitation through its capability to interconnect users with experts based on multimedia modules, and it provides inter-organizational cooperation for geographically dispersed organizations or cooperating organizations through a Geographical Information System (GIS) interface, that provides users with facilities to seek experts in different locations and establish direct contact with them (Fig. 1). Lately an image repository has been integrated in order to provide support for heterogeneous types of files. The image repository is implemented in Oracle 9i database environment, while access to the database is provided through a Java based module which provides the possibility for storing and retrieving images based on the usage of metadata for each item inserted in the database.

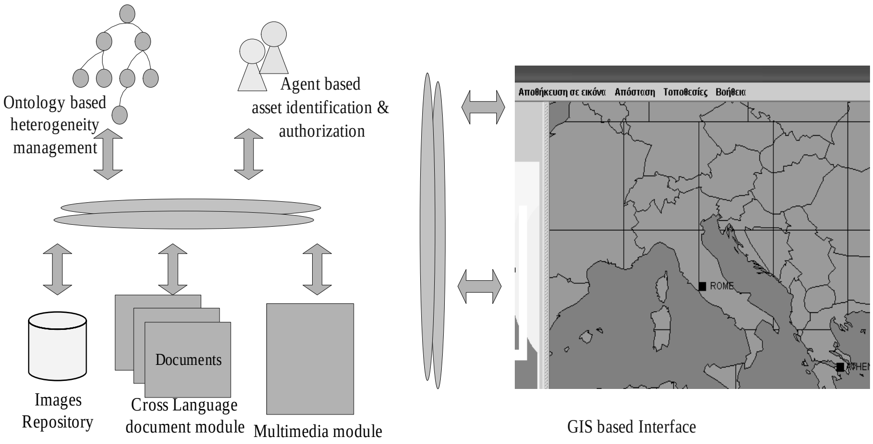
Ontology based heterogeneity management
Agent based asset identification & authorization
Images Repository
Documents
Cross Language document module
Multimedia module
Αποθήκευση σε εικόνα
Απόσταση
Τοποθεσίες
Βοήθεια
ROME
ATHEN
GIS based Interface

Figure 1: System Components

This implementation scheme is replicated on different nodes and supposing that different domains would like to contribute their knowledge potential, our aim is to ensure first that only authorized personnel between the two domains will have access to the knowledge sources. This situation is typical in medical environments, where due to the mobility of patients, a lot of patient related data are distributed in interconnected hospitals, or in e-Government environments, where all cooperating agencies need to share access to each other's data for a common purpose. Relative to these issues, in our research, we consider two main problems:

- First, the problem of knowledge discovery upon different domains
- Second, how a user from one domain can be authorized to access resources from another domain and how this procedure can happen transparently and securely, which means by minimizing the effort on the user's behalf and at the same time without exposing the knowledge assets to unauthorized disclosure or modification.

Relative to the first problem, the role of ontology is crucial. Each domain maintains its own domain ontology. We also introduce a central ontology repository, accessible from all the domains for retrieving domain ontologies. Ontologies define the concepts for each domain and their properties and enable semantically enabled description and querying over the knowledge assets. In order to enable transparent knowledge assets identification we utilize software agents, which act on the user's behalf and query the distributed domains. Ontologies play also a crucial role in facilitating communication between software agents as they enable standardization of terminology in agent communication messages.

Relative to the second, we adopt a security policy based approach and we introduce a security-agent that handles the necessary negotiations in order to authorize a user to gain access over a specific asset. The security policy defines the roles that deserve access to a specific asset, and the agent carries the user credentials which enable user identification, and accordingly by a policy interpretation the user is authorized or not to retrieve the remote asset.

3.2 Implementation details

We implemented a prototype by utilizing the following technologies:

1. The organizational memory module, which stores and retrieves semi-structured multilingual documents and interconnects users with experts, the Java and Oracle 9i technology have been used.
2. The GIS interface and the asynchronous and synchronous communication tools [Belsis 2004a] have been implemented on Java (JDK and JMF) and XML (Apache Xerces parser), with the usage of the OpenMap Java Bean component [OpenMap].
3. For distributed authentication, we have used the XACML [XACML] policy language, which maps users to roles and related with these roles credentials and accordingly records these details together with context parameters to XML files. This choice enables operational interoperability, because the authentication scheme can be independent of the nature of the applications among the several participating domains.
4. For the agent-based tasks, such as authentication and cross-organizational knowledge discovery, the JADE [JADE] agent platform has been utilized.

3.3 Application scenario

The goal of the system is to engage users in a knowledge acquisition and dissemination procedure that enables both the utilization of tacit and explicit knowledge, merging knowledge from different organizations in a transparent to the user process. The system, upon a user's request is looking in the codified semi-structured documents repository, for documents that match the user's criteria. Additively, experts relevant to the thematic areas of the retrieved documents are presented, in case the documents are not efficiently related with the user's problem. It is upon the user's choice, if the retrieved documents fail to provide efficient assistance to the problem, either to utilize the asynchronous application and send targeted messages to the most specialized experts, or in the case some of them are available, to connect through the GIS interface and the multimedia module with other domains (in the case of geographically dispersed organizations) and acquire online assistance. Except from these options, the system has the ability to merge different domain organizational memories, without demanding additive effort from the user. The knowledge discovery agent handles the communication and query details with other domains, and identifies the most relevant sources with the posed query. Accordingly, the authorization process needs to identify which of the knowledge assets are allowed for the user's security clearance level.

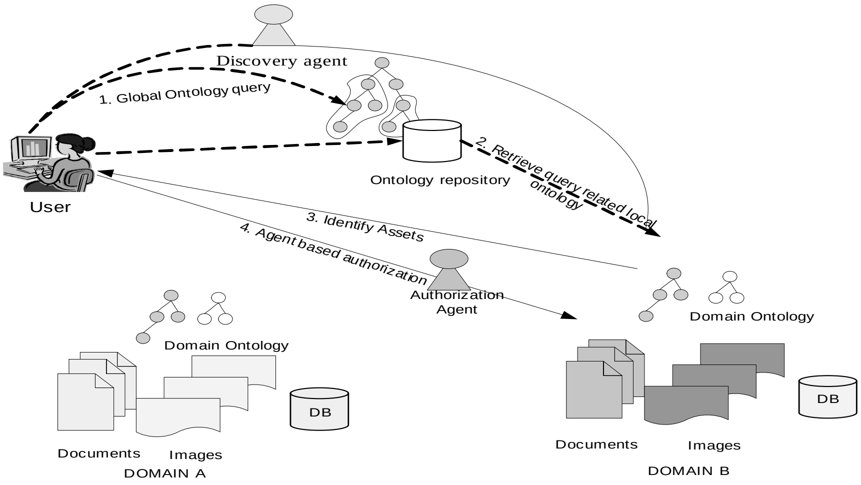

Figure 2: Transparent, agent-based identification and authorization sequence

Figure 2 shows the sequence of the identification and authorization related tasks, relative to the organizational memory utilization. The user poses a question, attempting to seek the most relevant sources to her query. At a first step, the most relevant domain ontologies are retrieved from the central ontology repository. The identification agent queries the domains that are more likely to maintain knowledge assets relative to the desired topic of interest. A list with the most relevant files is identified. Accordingly the authorization agent carries the credentials, fetches and presents to the user the files that correspond to the user's authorization level. In order to make the policy based authorization process more clear, we will explain accordingly in brief its main concepts.

On each knowledge node, we implement a Policy Enforcement Point (PEP). The PEP is responsible for allowing access to a resource. The authorization decision is made by another module, the Policy Decision Point (PDP). The PDP can collect also context related attributes for a role from another module, the context handler, enabling thus context-based authorization. The XACML policy framework now operates as follows (Fig. 3): The policy administrator, who is responsible for determining access rights, edits the security policy and makes it available to the PEP. When a request is made, it is directed to the PEP.

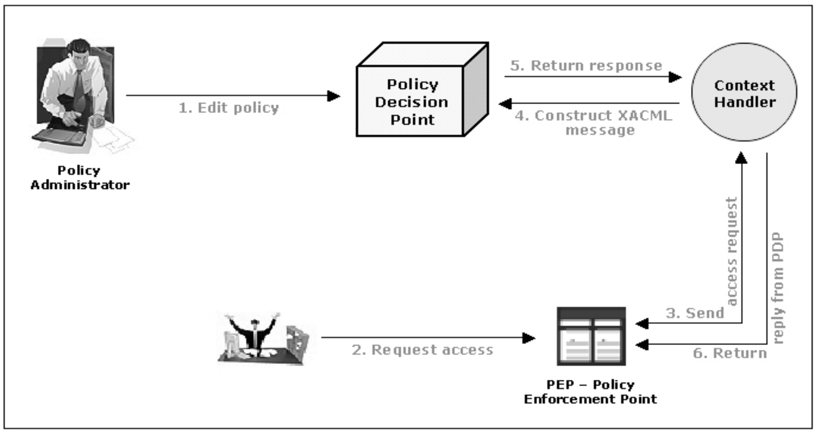

Figure 3: XACML based authorization

The PEP is requesting additive context related information, through another module, the context handler, responsible for constructing the messages in XACML format and collecting additive information, such as subject, action, resource and environment related attributes. Then, this XACML message is transmitted to the PDP, which decides upon providing authorization. Accordingly, the PDP returns the response to the context handler in XACML native format and at the end the message is directed to the PEP, which fulfils its obligations, authorizing or not the requester to perform the requested action over the resource. For transparency reasons, both the identification of knowledge assets as well as the authorization related tasks, through the implemented policy based modules, are performed by user agents, which encapsulate in the exchanged messages the necessary query related information or authorization specific information.

4 Conclusions – further directions

Knowledge Management has almost become a necessity for the modern organization, in order to become competitive and to respond quickly to technological challenges and changes. Usually, different organizations need to collaborate on the grounds of a common target; still their knowledge potential cannot be utilized uniformly due to the fear for unauthorized disclosure of knowledge assets to the other partner. This situation leads to limited performance while the collaboration is active. We present a solution to the aforementioned problem, where all the necessary resources can be shared and with clearly efficient and robust from security perspective manner. Our architecture is resilient and scalable. It enables the joint collaboration of several organizations and demands little extra cost in order to administer the system. We have presented the architecture of our prototype, the functionalities of which are continuously extending, covering so far support for different types of files, and enabling transparent to the user knowledge assets identification and access control enforcement. We are attempting to expand the policy-based framework, so as to include negotiation support for different policy models and policy languages.

Acknowledgements

This work was co-funded by 75% from E.E. and 25% from the Greek Government under the framework of the Education and Initial Vocational Training Program – Archimedes.

References

[Abecker 2003]. Abecker A., Bernardi A., Elst L., "Agent technology for distributed organizational memories" The Frodo project, ICEIIS 2003, pp.2-10

[Belsis 2004a] Belsis P., Gritzalis S., Malatras A., Skourlas C., Chalaris I., "Enhancing Knowledge through the use of GIS and Multimedia", in *Proceedings of the PAKM 2004 5th International Conference on Practical Aspects of Knowledge Management*, U. Reimer, D. Karagiannis (Eds.), December 2004, Vienna, Austria, Lecture Notes in Computer Science LNCS, Springer

[Belsis 2004b] Belsis P., Gritzalis S., ''Distributed autonomous Knowledge Acquisition and Dissemination ontology based framework'', Workshop on Enterprise Modeling and Ontology: Ingredients for Interoperability H. Kuhn (ed.) Dec. 2004 Vienna Austria, Univ. of Vienna.

[Bonifacio, 2002]. M. Bonifacio, P. Bouquet, and P. Traverso. Enabling distributed knowledge management. Managerial and technological implications. *Informatik – Informatique*, 1/2002.

[JADE] http://jade.tilab.com/

[OpenMap] Open Systems Mapping Technology, OpenMap TM, http://openmap.bbn.com

[Sandhu 2000]. Ravi Sandhu, David Ferraiolo, and Richard Kuhn. The NIST model for role based access control: towards a unified standard. In Proceedings of the Fifth ACM Workshop on Role-Based Access Control (RBAC'00), pages 47–63, 2000.

[Tempich, 2004] Tempich C., Ehrig M., Fluit C., Haase P., Marti E.L., Plechawski M., Staab S. "XAROP: A Midterm Report on Introducing a Decentralized Semantics based Application, Proceedings of Practical Aspects of Knowledge Management (PAKM) 2004, Vienna Austria, D. Karagiannis, U. Reimer (eds) LNAI 3336 Kluwer Academic publishers, pp. 259-270.

[Weippl 2004] Weippl E., Schatten A., Karim S.,, Tjoa A. SemanticLIFE Collaboration: Security Requirements and solutions – security aspects of semantic knowledge management. Proceedings of Practical Aspects of Knowledge Management (PAKM) 2004, Vienna Austria, D. Karagiannis, U. Reimer (eds) LNAI 3336 Kluwer Academic publishers, pp. 365-377.

[XACML] ''Extensible access control markup language specification 2.0'', OASIS Standard, 2004 (available at http://www.oasis-open.org).