

## SECURE GROUP COLLABORATION IN AN OPEN HEALTHCARE ENVIRONMENT

Zhengyi Le and James Ford  
The Dartmouth Experimental Visualization Laboratory,  
The Institute for Security Technology Studies, and  
Computer Science Department  
Dartmouth College  
Hanover, New Hampshire, USA  
E-mail: Zhengyi.Le@Dartmouth.edu

Vangelis Karkaletsis and Vassilios Spiliopoulos  
Software & Knowledge Engineering Laboratory  
Institute of Informatics & Telecommunications  
National Center for Scientific Research (NCSR) "Demokritos"  
Athens, Greece

Sokratis Katsikas  
Department of Information and Communication Systems Engineering  
The University of the Aegean  
Greece

Fillia Makedon  
Office of Cyberinfrastructure  
National Science Foundation  
Arlington, Virginia, USA

### **Abstract**

We introduce the concept of an "open healthcare environment", which is an electronic domain in which multiple healthcare entities need to interact but do not necessarily have complete knowledge of each other. In this setting, we show that a tool like OC (Open Collaboration), a tool being developed to support a variety of electronic collaboration needs, may be useful. OC is built on the open-source JXTA and MyJXTA toolkits. Group and role information is propagated in a peer-to-peer fashion, and peers can share files and send instant messages to any peer who is a member of an appropriate group or role.

### **Introduction and Motivation**

Healthcare is increasingly becoming a distributed service involving stakeholders and resources who may be physically far from each other. An Open Healthcare Environment (OHE) implies that a stranger may need to join a group collaboration where the entities are diverse and autonomous. It is important that OHE provides secure data handling, access to data (or other resources), storage of data and transmission of data. Thus, an effective OHE will enable secure collaboration mechanisms that permit (a) on demand formation of collaboration groups, (b) the ability for qualified strangers to join a collaboration group, (c) the ability to operate in a totally distributed setting without a central administration, and (d) guarantees of privacy and security control by the users of the collaboration system.

In this paper, we describe a system that we are building at Dartmouth jointly with colleagues at the National Center for Scientific Research (Demokritos) and the University of the Aegean in Greece. OC, or Open Collaboration (Le et al., 2005), is an open source resource sharing and collaboration system. Our approach is based on existing Automated Trust Negotiation (ATN) methods using peer-to-peer (P2P) solutions and uses configurable profiles for groups and

individuals to enable privacy and security control. ATN provides the tools need to help a stranger join existing collaborations without human intervention. The fact that OC runs on P2P protocols means this system removes the need for a centralized server, and that any node can be both a resource consumer and a resource provider. OC also applies Role-Based Access Control (RBAC) on shared files, which allows for a more flexible and scalable access authorization solution than traditional Access Control List (ACL) mechanisms.

### Related Work

*Collaboration Systems.* The work most related to OC is by Ellison and Dohrmann (Dohrmann & Ellison, 2002; Ellison & Dohrmann, 2003) and Nita-Rotaru and Li (2004). Ellison and Dohrmann presented a security architecture called NGC (Next Generation Collaboration). They gave a process requiring human and machine interaction for binding a name to a public key and used SDSI names for both groups and individuals. However, NGC's flexibility was limited by the fact that new members could only be added by invitation, and that this could only be done by a subset of the current members pre-authorized to issue invitations. This means that in general only parties known to the core users could participate. Most recently, Nita-Rotaru and Li (2004) presented a framework for role-based access control in group communication systems. They identified the set of all possible group operations that can be controlled and defined the group policy as a mapping between roles and operations using context as constraints. However, theirs is a centralized system, and it does not provide any mechanisms for allowing a stranger to join a collaboration.

There are several complete systems that provide functionality similar to that of OC. Recently the commercial product Groove ([www.groove.net](http://www.groove.net)) has provided a collaborative community for users. Users can create “workspaces” and invite people to join, after which members of the workspace can share files and manage projects together. Groove supports only three fixed roles — Manager, Participant, and Guest — and new members can join only by being invited. Groove is not a pure peer-to-peer system: it needs the help of a reply server. Open-Xchange (OX) and eGroupWare are open source collaboration packages. Both are web-based applications, and a centralized server is thus required. They do not support roles, and users can be added only by a system administrator.

*Role-based Trust Management.* Li and Mitchell (2003) presented a role-based trust-management framework called **RT**. **RT** provides policy language, semantics, a deduction engine, and pragmatic features to address large-scale and decentralized access control and authorization problems.

*Automated Trust Negotiation.* Winsborough *et al.* first introduced the notion of ATN and an architecture for managing the exchange of credentials between two strangers for the purpose of determining bilateral trustworthiness (2000). Researchers have designed ATN systems (Yu *et al.*, 2001; Seamons *et al.*, 2002a; Yu & Winslett, 2003; Yu *et al.*, 2003; Bertino *et al.*, 2004b) and addressed related privacy and security issues (Seamons *et al.*, 2002b; Winsborough & Li, 2002a, 2002b; Yu & Winslett, 2003b; Bertino *et al.*, 2004a). In addition, Li *et al.* (2003) proposed an RSA-based protocol and an ID-cryptography-based protocol to address the cyclic interdependency problem in automated trust negotiation.

*Why knowledge management?* As a result of the continuous evolution in both industrial and academic applications, in conjunction with the distributed settings and varied needs in organizations, knowledge management is required now more than ever. The major approach for dealing with these challenges has been to use schemata and ontologies for resource description, manipulation, and reuse. Several applications in various domains, such as

Database Schema Integration (Shvaiko, 2004), Data Warehouses (Kalfoglou & Schorlemmer, 2002), Web-based Systems Interoperability (Kalfoglou et al., 2005a), and Semantic Web Agents' Interoperability (Baousis et al., 2006) utilize these techniques.

*Overlapping Knowledge and schema/ontology mapping.* In a distributed application, even one where resources can be described unambiguously, the problem of overlapping knowledge exists when ontologies or schemata are employed: users are able to describe the same concept using different formal constructs. Various methods for dealing with this problem by creating mappings between ontology concepts or schemata elements that refer to the same entity have been proposed. These methods come from a variety of fields, including Information Retrieval, Machine Learning, Databases, and Mathematics/Statistics. Depending on the features that the specific ontology/schema languages involved offer (e.g. concepts labels, attribute labels, concept hierarchies) there are different methods that can utilize these features and produce possible mappings. In the case of ontologies in particular, some methods take this process a step further and produce a merged ontology, which incorporates all the available knowledge of the ontologies it came from. Frequently referenced methods include linguistic similarity, similarity flooding, graph matching, coefficient computation, formal concept analysis, information flow, logic satisfiability (SAT), and Description Logics (DL) based inference; for a review, see Kalfoglou et al. (2005b).

COMA/COMA++ (COMbination of MAtching algorithms) is a modular tool that combines several schema matching techniques. As shown in (Do & Rahm, 2001) and (Aumüller et al., 2005), it provides an expandable library of mapping techniques, currently including six simple matchers and five hybrid ones. Furthermore, it can perform matching compositions based on the transitive nature of the relations between schema elements and also has ability to reuse previously located mappings. The former feature is motivated by the fact that many schemata to be matched are similar to or even identical with ones already processed. The majority of the matchers are string-based in various forms (n-gram, affix, edit distance, phonetic similarity), but COMA also utilizes datatype constraints and synonymy/hypernymy relations based on external dictionaries. An extensive evaluation of COMA shows that it outperforms similar tools like Autoplex, Automatch, LSD, GLUE, SF and SemInt (Do, et al., 2002; Shvaiko & Euzenat, 2005).

### **Open Healthcare Benefits and Applications**

We define an Open Healthcare Environment as one in which multiple healthcare-related entities (such as patients, doctors, hospitals, and insurers) need to interact but do not necessarily have prior experience with each other. An open environment is “open” in the sense that there are no absolute barriers to entry: no central authority governs who can interact with whom, and environment participants are free to create, modify, and disband interactions and groups as the need arises. In this situation, entities must still typically make use of some official registrations and endorsements — as represented, for example, by government-issued credentials establishing identity or employer-created credentials attesting to a particular status at their organization — but the exact policies regarding what credentials to require for what purpose are left to the participants themselves.

Consider this scenario: a group of AIDS patients intends to share information and resources within their community, so that each member can search and obtain useful content maintained by other members. Due to the sensitivity of the shared content, patients wish to authenticate each other to ensure that only designated community members will have access to this information. Moreover, AIDS patients are concerned about their privacy; they want their personal information and credentials to be revealed as minimally as possible during the authentication process. In order to achieve these objectives, OC uses automated trust

negotiation to build a secure P2P content search framework for such a situation. Similar sharing-based needs might involve exchanges of data between healthcare institutions and government agencies, or between clinicians, pharmacies, and pharmaceutical manufacturers.

### **Group Collaboration Mechanisms and Applications**

In an open health environment, any interested user may ask to join a collaboration. The traditional approach to joining a collaboration is to let a system administrator review a registration form and all qualification credentials of that user and then make an account for that user (or reject their application). This human-interactive one-way authentication is not suitable for dynamic and large-scale applications. If the applicants have questions about a group, more human intervention and delay will be introduced.

### **Automated Trust Negotiation in OC**

OC uses the concept of automated trust negotiation to avoid this administrative overhead. ATN works as follows: an applicant sends a request to join a group to a group recruiter (which may actually be an agent, i.e. an a computer program running autonomously and without human input), and the group recruiter sends join requirements back to an applicant. The join requirements may include some attribute-based credentials (e.g. Age>18) and possibly other credentials, such as the electronic equivalent of identity or membership cards. After the applicant receives the requirements, they check their local policy pool to see where any required credentials are considered sensitive. If sensitive, they will have a release policy that protects this credential. In this case, the applicant sends back a counter-request indicating the requirements for releasing this credential. If the recruiter can satisfy the counter-request, the applicant will send the requested credentials. Once the recruiter receives and verifies those credentials, the applicant is issued a credential indicating that they are a member of the group. The entire procedure can be executed by the system automatically.

In a specific OHE application, a collaborative group may require that some services, e.g. public file sharing, should be open to everybody while other services, e.g. sensitive file sharing, should be open only to a qualified subset of users. In our approaches, we use role based trust management to control data access. When a user joins a group, they are automatically assigned a role, typically as a guest or junior member. If they want to obtain additional roles, they must repeat the application procedure again specifically for a desired role to get a role certificate. Different services may be protected by different policies, some of which ask the requester to present specific role certificates. When required, a user uses their role certificate to request these services, e.g. the downloading of certain files.

Using a peer-to-peer approach removes the need for a centralized server. A "pure" P2P network does not have the notion of clients or servers, but only equal *peer* nodes that simultaneously function as both "clients" and "servers" with respect to the other nodes on the network. This model of network arrangement differs from the client-server model, where client communication is usually to and from a central server. Since in many situations in this domain each participant could be both a data provider and a data consumer (i.e., be both a server and a client), P2P meets the need of open health environment groups very well. P2P networks are also more efficient for data sharing and avoid the single point of failure problem since data are distributed among the peer nodes (if desired, with some level of redundancy).

### **Resource Description in OC**

In an open source resource sharing and collaboration system like OC, resources should be described by proper schemata. It should be possible, and even encouraged, that users should be able to define their own conceptualization of the knowledge they offer to the network and

also of the knowledge that they consume by participating in it. In other words, different schemata depicting the same resources are expected to be shared over the network. As a consequence, schema mapping using COMA/COMA++ will be a useful future capability for the OC system. As an alternative approach, users could also define their own ontologies in order to describe their resources in a more descriptive manner. In such a case, more advanced techniques could be used, such as ontology merging (Kotis et al., 2006). Specifically, all the ontologies defined over a group could be merged into a central one, which would be usable by all group members as a shared reference ontology.

### Concluding Remarks

Some features of OC are still under development. We are currently implementing a policy module that allows one to describe the requirements for accessing shared resources and assigning roles, and an ATN module that reads policies and enforces them. We also plan to incorporate COMA/COMA++ or a similar approach for schema matching as described above.

This project was supported under Award number 2000-DT-CX-K001 from the U.S. Department of Homeland Security, Science and Technology Directorate. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security or the Science and Technology Directorate.

### References

- Aumüller, D., Do, H.H., Massmann, S., & Rahm, E. (2005). Schema and Ontology Matching with COMA++. In *Proceedings of the International Conference on Management of Data*.
- Baousis, V., Zavitsanos, E., Spiliopoulos, V., Hadjiefthymiades, S., & Merakos, L. (2006). Wireless Web Services Using Agents and Ontologies. In *Proceedings of ICPS'06: The IEEE International Conference on Pervasive Services*. To appear.
- Bertino, E., Ferrari, E., & Squicciarini, A.C. (2004a). Privacy-Preserving Trust Negotiation. In *Proceedings of the 4<sup>th</sup> Workshop on Privacy Enhancing Technologies*.
- Bertino, E., Ferrari, E., & Squicciarini, A.C. (2004b). Trust-X: A Peer-to-Peer Framework for Trust Establishment. *IEEE Trans. Knowl. Data Eng.*, 16, 827–842.
- Do, H.H. & Rahm, E. (2001). COMA — A System for Flexible Combination of Schema Matching Approaches. In *Proceedings of the Very Large Data Bases Conference* (pp. 610–621).
- Do, H.H., Melnik, S., & Rahm, E. (2002). Comparison of Schema Matching Evaluations. In *Proceedings of the Workshop on Web and Databases*.
- Dohrmann, S. & Ellison, C. (2002). Public-key Support for Collaborative Groups. In *Proceedings of the 1st Annual PKI Research Workshop* (pp. 139–148).
- Ellison, C. & Dohrmann, S. (2003). Public-key Support for Group Collaboration. *ACM Trans. Inf. Syst. Secur.*, 6, 547–565.
- Kalfoglou, Y. & Schorlemmer, M. (2002). Information-flow Based Ontology Mapping (Informatics report no.135). University of Edinburgh Division of Informatics.
- Kalfoglou, Y., Hu, B., & Reynolds, D. (2005a). On interoperability of Ontologies for Web-based Educational Systems. In *Proceedings of the Workshop on Interoperability of Web-based Educational Systems at the 14th International World Wide Web Conference*.
- Kalfoglou, Y., Hu, B., Reynolds, D., & Shadbolt, N. (2005b). Capturing Representing and Operationalising Semantic Integration (ECS ePrints report #10842). University of Southampton and HP Labs.
- Kotis, K., Vouros, G.A., & Stergiou, K. (2006). Towards Automatic Merging of Domain

Ontologies: The HCONE-merge Approach. *Journal of Web Semantics*.

- Le, Z., Ouyang, Y., Ford, J., & Makedon, F. (2005). OC: A system for Open Collaborations. In *Proceedings of the First International Workshop on Security, Privacy and Trust in Pervasive and Ubiquitous Computing* (pp. 93–100). IEEE Press.
- Li, N., Du, W., & Boneh, D. (2003). Oblivious Signature-based Envelope. In *Proceedings of PODC '03: The 22<sup>nd</sup> Annual Symposium on Principles of Distributed Computing* (pp. 182–189). ACM Press.
- Li, N. & Mitchell, J.C. (2003). RT: A Role-based Trust-management Framework. In *Proceedings of the Third DARPA Information Survivability Conference and Exposition* (pp. 201–212). IEEE Computer Society Press.
- Nita-Rotaru, C., & Li, N. (2004). A Framework for Role-Based Access Control in Group Communication Systems. In *Proceedings of 2004 International Workshop on Security in Parallel and Distributed Systems*.
- Seamons, K.E., Winslett, M., Yu, T., Smith, B., Child, E., Jacobson, J., Mills, H., & Yu, L. (2002a). Requirements for Policy Languages for Trust Negotiation. In *Proc. of the Third International Workshop on Policies for Distributed Systems and Networks* (pp. 68–79).
- Seamons, K.E., Winslett, M., Yu, T., Yu, L., & Jarvis, R. (2002b). Protecting Privacy during On-Line Trust Negotiation. In *Proceedings of the 2nd Workshop on Privacy Enhancing Technologies* (pp. 129–143).
- Shvaiko, P. (2004). A Classification of Schema-based Matching Approaches. In *Proceedings of the Meaning Coordination and Negotiation Workshop at the 3<sup>rd</sup> International Semantic Web Conference*.
- Shvaiko, P. & Euzenat, J. (2005). A Survey of Schema-based Matching Approaches. *Journal on Data Semantics, IV*, 146–171.
- Winsborough, W.H., Seamons, K., & Jones, V. (2000). Automated Trust Negotiation. In *Proceedings of the DARPA Information Survivability Conference and Exposition* (pp. 88–102). IEEE Press.
- Winsborough, W.H. & Li, N. (2002a). Protecting Sensitive Attributes in Automated Trust Negotiation. In *Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society* (pp. 41–51).
- Winsborough, W.H. & Li, N. (2002b). Towards Practical Automated Trust Negotiation. In *Proceedings of the Third International Workshop on Policies for Distributed Systems and Networks* (pp. 92–103).
- Yu, T., Winslett, M., & Seamons, K.E. (2001). Interoperable Strategies in Automated Trust Negotiation. In *Proceedings of CCS '01: The 8th ACM Conference on Computer and Communications Security* (pp. 146–155). ACM Press.
- Yu, T. & Winslett, M. (2003a). A Unified Scheme for Resource Protection in Automated Trust Negotiation. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 110–122).
- Yu, T. & Winslett, M. (2003b). Policy Migration for Sensitive Credentials in Trust Negotiation. In *Proceedings of WPES '03: The 2003 ACM Workshop on Privacy in the Electronic Society* (pp. 9–20). ACM Press.
- Yu, T., Winslett, M., Seamons, K.E. (2003). Supporting Structured Credentials and Sensitive Policies through Interoperable Strategies for Automated Trust Negotiation. *ACM Trans. Inf. Syst. Secur.*, 6, 1–42.