UNIVERSITY
OF TRENTO

DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

AUTOMATED ACCESS CONTROL RULE GENERATION
VIA SEMANTIC MATCHING

Rui Zhang

December 2008

Technical Report # DISI-08-082

Also: accepted as a poster for the WMM, AISB2009 convention

Automated Access Control Rule Generation via Semantic Matching[1]

Rui Zhang[2]

Abstract. Semantic Web techniques bring us help in many fields. In this paper, we propose a way to use Semantic Matching on access control. We illustrate the motivation with an eBusiness access control schema based on $RelBAC$ (for Relation Based Access Control). Semantic Matching techniques are applied on the lightweight ontologies of the subjects and the objects to find the semantic similarities that can be used to suggest new rules, to reuse the existing rules or to separate the duties of semantically disjoint user sets.

1 Introduction

Information Era releases people and data from centralized local environment to dynamic evolving communities and distributed information resources of various scales and types. $RelBAC$ [8] has been proposed as a new model for the dynamic evolving community access control scenario such as eBusiness. One important feature of $RelBAC$ is that hierarchies are naturally represented with a partial order '$\geq$' which formalized as subsumption in the logical framework of $RelBAC$, an access control domain specific description logic. OWL-DL can be used to represent the knowledge as an ontology. This brings us not only the expressiveness, but also the possibility of applying semantic web techniques on the model. Meanwhile, dynamic community access control requires powerful management and administration on various scaled information. To generate new rules on the fly for this vast amount of changes will be time-consuming and error-prone. Thus suggestions to create new rules or to reuse existing rules arouse the interests of research.

We present in this paper a new way of applying Semantic Matching techniques on access control. With a running example of an eBusiness schema, we show how to use these matching results to find semantically related subjects and object in order to suggest possible permission assignment; and to find the similar subject/object sets that can reuse existing rules. The mismatch between subject/object ontologies are also useful such as for separation of duties.

The paper is organized as follows: Sec.2 describes an eBusiness access control schema based on $RelBAC$ as the motivation; Sec.3 shows how to apply Semantic Matching on access control; Sec.4 lists the state of the art and we conclude in Sec.5.

2 Motivation: eBusiness via $RelBAC$

Nowadays, eBusiness becomes so popular that the person sitting beside you on a trolley bus might be an eBusiness vendor of several online shops. Here we suppose an example in an eBusiness solution.
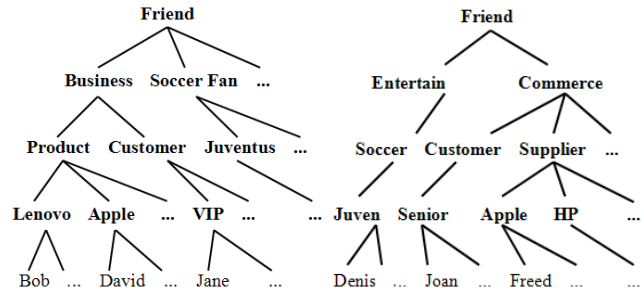

Figure 1. Alice's Social Ontology   **Figure 2.** Bob's Social Ontology

An online vendor, Alice, has a shop on eBay selling digital devices. Her social network consists of many persons, e.g., Bob and David have business relations with her and Chris and George are just common friends, etc. With the continuous growth of this network, Alice wants to manage these friends in her own way, so that she can easily find the 'proper' profile of a friend whenever necessary. For instance, David is a business friend who works as the sales of Apple company, and he will inform Alice the news of Apple products and special offers such that Alice can put it on her website in time. Jane is a representative of the best customer because she visits Alice's online shop frequently and comments on the deals she completed such that potential customers will get an impression on the quality of service and goods. Of course, Alice is happy to give Jane VIP prices as rewards. In general, Alice has a social network with various people and different social interactions.

As an eBusiness runner, Alice likes and has to control the access to the data she puts online. A natural and flexible access control model is $RelBAC$. As described in [8] $RelBAC$ is a model for community access control. It has common components such as SUBJECTs and OBJECTs, and a special part PERMISSIONs as binary relations. A PERMISSION is a named pair $P(s, o)$ where s is a SUBJECT, o is an OBJECT and P is the PERMISSION describing the action that u intends to perform over o such as $Read$ and $Write$. $RelBAC$ defines a common relation with partial order '$\geq$' such that all these three components can be organized in hierarchies as a tree (or DAG). An access control domain-specific Description Logic is used to formalize the $RelBAC$ model. SUBJECTs and OBJECTs are formalized as concepts, and PERMISSIONs as roles. Hierarchies in the model can be formalized as subsumption axioms. All the system states and access control policies are formalized as logical formulas on which automated reasoning can be performed.

Alice may build a tree-like structure as Figure 1 to classify people
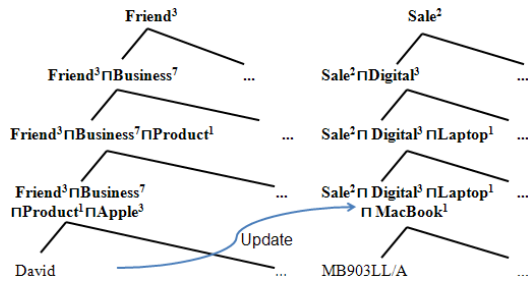

Figure 3. Permission Assignment in $RelBAC$

in her complex social network according to the social relations. The access control is simplified as managing the links between the subject and object ontologies. By exploiting the theory of Lightweight Ontology as described in [5, 7], the arbitrarily manual structure is transformed into a lightweight ontology where implicit semantics on the tree edges are unified into explicit 'IS-A' relations, and the natural language labels of nodes are disambiguated with natural language processing [12] into logical formulas. For instance, Figure 3 shows parts of the lightweight ontologies built by Alice and the assignment of 'Update' to user 'David' on the set of objects 'MacBook'. In the left lightweight ontology, David is classified as an instance of the set '$Friend^3 \sqcap Business^7 \sqcap Product^1 \sqcap Apple^3$' according to his social position that he has a $Business^7$ relation with Alice and he works for $Apple^3$ (the superscript depicts the 3rd sense in the knowledge base, i.e., an IT company rather than a fruit). Symmetrically, in the right ontology (of the goods on sale), there is a class of objects '$Sale^2 \sqcap Digital^3 \sqcap Laptop^1 \sqcap MacBook^1$' where $Sale^2$ is a branch of $Business^7$, $MacBook^1$ is a $Laptop^1$ as a $Product^1$ of $Apple^3$. Apparently the two concepts are syntactically different, but semantically overlapping.

Things become more complicated when new ontologies arrive, e.g, if Alice likes to collaborate with some other eBusiness vendor who has her own user community and product category ontologies heterogeneously. The traditional way to solve the heterogeneity is to merge the database and create new rules for the 'new' knowledge base. Figure 1 shows part of Alice's social ontology and Bob, another eBusiness vendor has his own social ontology as Figure 2. The collaboration of Alice and Bob might lead to integration of these social 'resource's, such as product supplier, transporter, customer, etc. in addition to the integration of the physical resources such as goods.

So the motivation lies in at least two aspects:

- Semantic similarities disclose the latent relationships between subjects and objects although they are syntactically different. These latent relationships might suggest rules to be created for these semantically relevant subjects and objects such as to permit David to update the web categories about Apple products.
- Semantic similarities between ontologies of a type, such as between two subject ontologies or two object ontologies or even permission ontologies, provide a way to reuse (i.e., propagate) the permissions assigned by existing rules such as to reuse the rules for 'VIP' users of Alice onto Bob's 'Senior' customers.

3 Semantic Matching for Access Control

$RelBAC$ provides automated reasoning about the knowledge base such as consistency checking and query answering. Thus, *member-*

ship checking, security property enforcement are used at design time to reason about *hierarchy management, permission propagation, separation of duties*, etc. and *query answering* can be used at run time for *access control decision*. However, that is not enough as an access control system for the larger and more complex eBusiness solutions crucially needs help to manage access control rules such as addressed in Section 2 by providing suggestions about candidate rules when the user is not an expert in access control (as it is often the case with social networks), or to provide semantic heterogeneity resolution for relatively large and complex ontologies or highly dynamic policies.

The fact that we handle subject, object and permission hierarchies as lightweight ontologies allows us to deal with the problem of semantic heterogeneity, namely with the fact that in general we will have multiple subject and/or object and/or permission hierarchies which express semantically related notions in many different forms. We can find with Semantic Matching tools that there exists similarity between the subject and object lightweight ontologies although they are heterogeneous and built independently. This will help to generate candidate permissions to be submitted to the user for approval, or generate semantically motivated constraints between subject and object categories, and so on.

To detect these semantic relations between classifications we use S-Match, a Semantic Matching tool described in [6]. The original idea of Semantic Matching is to calculate the semantic similarity such as *equal, overlapping*, etc. between the categories of the two given classifications. The core of a S-Match procedure consists two rounds of matching. The first round match is performed on the *concept at label* which are logical formulas formed with word senses such as the column names of Table 1. WordNet [9] is used as a knowledge base in which possible relations between senses (meanings of word) are provided. Semantic similarities are defined with sense relations. *Equal* $\equiv$: one concept is equal to another if there is at least one sense of the first concept, which is a synonym of the second. *Overlapping* $\sqcap$: one concept is overlapped with the other if there are some senses in common. *Mismatch* $\perp$: two concepts are mismatched if they have no sense in common. *More general / specific* $\sqsupseteq, \sqsubseteq$: One concept is more general than the other iff there exists at least one sense of the first concept that has a sense of the other as a hyponym or as a meronym. These direct results from the knowledge base can be regarded as a preparation for the second round of matching as they discover the relations between senses of single nodes. Afterwards, matching is performed on the *concept at node* which is a *conjunction* of all the *concepts at label* of nodes from the root to current, e.g., DL formulas in Figure 4. The results of the second round match is calculated by checking subsumption with a reasoner.

Let us see how to use these matching results in turn.

3.1 Suggestions for Rule Creation

For any access control systems, the stage of rules creation is very important because a cute rule set will simplify later work as enforcement and management. Semantic Matching between the subject and the object ontologies will find out potential semantic relations between categories of the two ontologies. For example, given the background knowledge about the relations $MacBook^1$ is a $Laptop^1$ as a $Product^1$ of $Apple^3$ etc., we can find the semantic similarities as listed in Table 1. As WordNet does not 'know' the word such as 'MacBook', which is common under the enormous emergences of new words in this Information Era, we should enrich the knowledge base with the facts such as '$Apple^3$ *is a IT company selling digital products such as MacBook and IPod.*'. This is a non-trivial task

Table 1. Semantic Matching on Labels

S-Match	$Friend^3$	$Business^7$	$Product^1$	$Apple^3$	$Lenovo^1$	$Soccer^1 \sqcap Fan^2$
$Sale^2$	$\perp$	$\sqsubseteq$				$\perp$
$Digital^3$						
$Laptop^1$			$\sqsubseteq$			
$MacBook^1$				$\sqcap$	$\perp$	
$Thinkpad^1$				$\perp$	$\sqcap$	

and many domain experts together with volunteers like common web users are contributing in this direction, at least to our own knowledge bases.

From Table 1, we can see the semantic similarities such as $Sale^2 \sqsubseteq Business^7$, etc. These relations provide the following suggestions to create new rules.

Semantically Related The cells marked with '$\sqsubseteq, \sqsupseteq, \equiv, \sqcap$' represent the semantic similarity of the corresponding concepts. It is meaningful to assign corresponding users some access to the objects. For example, the relation $Sale^2 \sqsubseteq Business^7$ suggests that some access, let us say $Read$, should be assigned to the $Business^7$ $Friend^3$ to some $Sale^2$ categories. It is obvious here in the small toy user and object ontologies, but facing a large eBusiness such as Amazon.com, these similarities will be very useful for the administrators in creating new rules. We may also place degrees on similarities. '$\equiv$' weighs more than '$\sqsubseteq$' and '$\sqsupseteq$', which in turn more than '$\sqcap$'. Therefore, it is more likely to assign access between '$\equiv$' related subjects and objects than the others.

Explicit Unrelated The cells marked with '$\perp$' represent that the corresponding concepts are found 'unrelated' in the knowledge base. Here we shorten the axiom '$C_1 \sqcap C_2 \sqsubseteq \perp$' as '$C_1 \perp C_2$'. We have to differentiate the real world semantics of these '$\perp$'s.

- $Sale^2 \perp Friend^3$ is a mismatch because they are referring to object and subject, i.e. an activity and a person respectively. This mismatch comes from the disjointness between person and activity as different subjects but does not prevent that a person can have some relation with an activity such as $Friend^3$ may have access to $Sale^2$.

- $MacBook^1 \perp Lenovo^1$ comes from that '*MacBook is a product of Apple company but not Lenovo.*' This kind of mismatch suggests exactly no access should be assigned.

- $Sale^2 \perp (Soccer^1 \sqcap Fan^2)$ covers both upper cases so it does prevent the access assignment from $Soccer^1$ Fan^2 to $Sale^2$.

The second case is a *strict mismatch* which means 'irrelevant' in common sense. It is important to detect this kind of mismatches because they can suggest for constraints such as *separation of duties* that we will discuss when matching two subject ontologies in the next subsection.

Implicit Unrelated The blank cells of the table mean that the knowledge base doesn't know any existing relation between the corresponding concepts. In this case, no semantical similarities are provided. From Table 1 we can see that this kind of cells are the majority in this example, only because the knowledge base we use is not designed for eBusiness domain. If it is specially enriched with more background knowledge, we believe more semantic relations can be found and more suggestions will be provided.

The interesting thing here is that the relation between $Friend^3$ and $Sale^2$ is *mismatch*. It is weird but true as $Friend^3$ means '*a person with whom you are acquainted*' and $Sale^2$ is '*the general ac-*

tivity of selling'. This is common when we match a subject ontology with an object ontology. If we went on with the second round of S-Match on the *concepts at node* which includes all the semantics from the root to the current node, this mismatch between $Friend^3$ and $Sale^2$ would propagate to all the results and Table 1 would be full of '$\perp$s' simply because of the similarity of the two roots is *mismatch*. We may get nothing from such a table, therefore in this phase, we use only the first round of S-Match on *concepts at label*.

3.2 Automated Rule Reuse

One important evolution of subject and object ontologies is to integrate other similar ontologies. For example, an eBusiness vendor will enlarge her social network to involve more customers and very likely she would integrate the customer ontology of another vendor, or symmetrically integrate the goods ontology. The traditional access control solutions ask an administrator to create new rules for these evolving parts. Even for the similar ontologies, all assignments have to be made once again. For example, the vendor in the scenario of Section 2 would like to merge another ontology of subjects as Bob's Social Ontology as Figure 2. In this case for instance, a customer set called 'Senior' has the similar intuition to the 'VIP' set in previous ontology.

The resulting semantic relations can be used along the lines of what described in the previous sections either to drive the merging of the two ontologies or to create mappings which allow for the propagation of permissions from one ontology to the other. Thus for instance the system administrator might enforce that the equivalence mapping between the two root nodes in Figure 4 means that a Read permission on the left root node propagates to the right root node. These kinds of mappings are very similar to the C-OWL mappings introduced in [1] and should be used whenever a full merge of the two ontologies is not advisable or there are good reasons to keep the two ontologies distinct.

We show in Figure 4 the results of S-Match on two branches of the 'friend' lightweight ontologies generated from the hierarchies in Figures 1 and 2. The semantic similarity axioms can be added to the knowledge base of access control and the rule reuse is done without further efforts. For example,

$$\{(Friend^3 \sqcap Commerce^1) \sqsubseteq (Friend^3 \sqcap Business^7),$$
$$Business^7 \sqsubseteq \alpha\} \models Friend^3 \sqcap Commerce^1 \sqsubseteq \alpha$$

With the help of these semantic similarities found by S-Match, any *subject-centric* rules with permissions assigned to $Business^7$ will also propagate to $Friend^3 \sqcap Commerce^1$ just as a reasoning result without creating new rules for the new subject sets. Similar reuse applies on objects as well when S-Match is used to find the semantic similarities between object ontologies.

Even though indicating 'explicit unrelated', '$\perp$' is an important semantic similarity for rule reuse. Here we refer to the *strict mismatch* discussed in Section 3.1. It means that the two nodes in
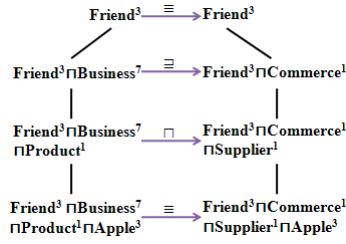
Figure 4. Ontology Matching for Rule Reuse

the two ontologies matched are semantically mutual exclusive, e.g., $HP^2 \perp Lenovo^1$ between the two ontologies in Figures 1 and 2. HP^2 and $Lenovo^1$ represent the set of users belonging to different IT companies, and it is rational to separate the duties from the two sets, i.e. users from company HP^2 should not have the same access as those from $Lenovo^1$. When the two ontologies are both considered as subject knowledge, the matching results suggest a new policy as $HP^2 \sqcap Lenovo^1 \sqsubseteq \perp$ which ensures that users cannot be members of both sets.

4 Related Work

With the arrival of Web2.0 and now coming even Web 3.0, access control over the resources online throughout the evolving social networks demands more automated tools for administration.

Classic access control techniques, e.g., cryptography have been proposed for community access control such as [2]. However, this kind of access control systems focus on protection from security threats rather than taking use of the rich information from the web. The authentication procedure is done once for all which is not enough for fine-grained access control.

Lockr [11] was proposed to fit the situation that the large number of content sharing systems and sites use different access control methods un-reusable for each other. It separates social networking information from the content sharing mechanisms, so that end users do not have to maintain several site-specific copies of their social networks. It also provides a way to use social relationships as an important attribute, *relationship type*, to define access control rules. However, Lockr still uses a public/private key communication and does not consider the semantic similarities.

Another thread similar to our solution is Semantic Based Access Control. Yague et al. discussed the Semantic Access Control model in [3] with a XML based language SPL (Semantic Policy Language). The model is based on the semantic properties of the resources, clients (users), contexts and attribute certificates and relies on the rich expressiveness of the attributes to create and validate access control policies. It is flexible to define access control over attributes but faces the complexity problem of the system. In contrast, our model covers the expressiveness of attributes and takes use of the structure at the same time so that the permission propagation will greatly reduce the number of rules. Pan et al. present a novel middle-ware based system [10] to use semantics in access control. It is based on $RBAC$ model [4] with a mediator to translate the access request between organizations by replacing roles and objects with matched roles and matched objects. For interoperation, they use semantic mapping on roles in order to find the similarity or separation of duties between roles in two ontologies. This is similar to our approach, but we do much further

as the S-Match tools are not domain specific so that we can match a subject ontology with an object ontology for new rule suggestions.

5 Conclusion

Based on the $RelBAC$ formalization of the access control problem in social networks, we can organize users, objects and permissions as (lightweight) ontologies. This allows to represent access control rules and policies as DL formulas and to reason about them using state of the art off-the-shelf reasoners. However, when the knowledge base is more and more complex, the rule management task explodes. Thus it requires automated or semi-automated tools to help creating and reusing rules. In this paper, we have shown how it is possible to use Semantic Matching technology to discover and exploit the underlying semantic relations between subject and object ontologies and between two user or object ontologies belonging to different policies. The resulting automated reasoning capabilities can be exploited to support the user or system administrator in the policy management, an activity which is time expensive and error-prone.

REFERENCES

[1] Paolo Bouquet, Fausto Giunchiglia, Frank Van Harmelen, Luciano Serafini, and Heiner Stuckenschmidt, 'C-owl: Contextualizing ontologies', in *Journal Of Web Semantics*, pp. 164–179. Springer Verlag, (2003).

[2] Barbara Carminati and Elena Ferrari, 'Privacy-aware collaborative access control in web-based social networks.', in *DBSec*, ed., Vijay Atluri, volume 5094 of *Lecture Notes in Computer Science*, pp. 81–96. Springer, (2008).

[3] Mariemma Inmaculada Yage del Valle, Mara del Mar Gallardo, and Antonio Mana, 'Semantic access control model: A formal specification', in *ESORICS*, eds., Sabrina De Capitani di Vimercati, Paul F. Syverson, and Dieter Gollmann, volume 3679 of *Lecture Notes in Computer Science*, pp. 24–43. Springer, (2005).

[4] David F. Ferraiolo, Ravi S. Sandhu, Serban I. Gavrila, D. Richard Kuhn, and Ramaswamy Chandramouli, 'Proposed NIST standard for role-based access control', *Information and System Security*, **4**(3), 224–274, (2001).

[5] Fausto Giunchiglia, Maurizio Marchese, and Ilya Zaihrayeu, 'Encoding classifications into lightweight ontologies.', *J. Data Semantics*, **8**, 57–81, (2007).

[6] Fausto Giunchiglia, Mikalai Yatskevich, and Pavel Shvaiko, 'Semantic matching: Algorithms and implementation.', *J. Data Semantics*, **9**, 1–38, (2007).

[7] Fausto Giunchiglia and Ilya Zaihrayeu, *Encyclopedia of Database Systems*, chapter Lightweight Ontologies, number 978-0-387-35544-3, Verlag, Springer, June 2009.

[8] Fausto Giunchiglia, Rui Zhang, and Bruno Crispo, 'Relbac: Relation based access control', in *International Conference on Semantics, Knowledge and Grid, SKG 2008*, ed., IEEE Computer Society, (2008).

[9] George A. Miller, 'Wordnet: A lexical database for english', *Communications of the ACM*, **38**, 39–41, (1995).

[10] Chi-Chun Pan, Prasenjit Mitra, and Peng Liu, 'Semantic access control for information interoperation', in *SACMAT '06: Proceedings of the eleventh ACM symposium on Access control models and technologies*, pp. 237–246, New York, NY, USA, (2006). ACM.

[11] Amin Tootoonchian, Kiran Kumar Gollu, Stefan Saroiu, Yashar Ganjali, and Alec Wolman, 'Lockr: social access control for web 2.0', in *WOSP '08: Proceedings of the first workshop on Online social networks*, pp. 43–48, New York, NY, USA, (2008). ACM.

[12] I. Zaihrayeu, L. Sun, F. Giunchiglia, W. Pan, Q. Ju, M. Chi, and X. Huang, 'From web directories to ontologies: Natural language processing challenges', in *Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea*, ed., Karl Aberer et al., volume 4825 of *LNCS*, pp. 617–630, Berlin, Heidelberg, (November 2007). Springer Verlag.