DC Proposal: Knowledge Based Access Control Policy Specification and Enforcement

Sabrina Kirrane

Digital Enterprise Research Institute,
National University of Ireland, Galway
sabrina.kirrane@deri.org
http://www.deri.ie

Abstract. The explosion of digital content and the heterogeneity of enterprise content sources have resulted in a pressing need for advanced tools and technologies, to support enterprise content search and analysis. Semantic technology and linked data may be the long term solution to this growing problem. Our research explores the application of access control to a knowledge discovery platform. In order to ensure integrated information is only accessible to authorised individuals, existing access control policies need to be associated with the data. Through in-depth analysis we aim to propose an access control model and enforcement framework which can be used to represent and enforce various access models both inside and outside the enterprise. Furthermore, through experimentation we plan to develop a methodology which can be used as a guideline for the lifting of distributed access control policies from the individual data sources to a linked data network.

Keywords: Policy, Access Control, Reasoning, Information Analysis, Knowledge Discovery, Linked Data.

1 Introduction

The Internet is growing exponentially, fuelling research into the next generation of Internet technologies. Over the past two decades much research has gone into the use of semantic technology and linked data for data integration, search and information analysis. This exponential expansion of data and many of the challenges that come with it are mirrored within the enterprise. Our use case examines how such techniques can be used to build an integrated data network for enterprise content analysis and knowledge discovery. Furthermore, we investigate how subsets of interconnected enterprise data can be shared with partner companies or exposed on the Internet.

When we integrate data from multiple line of business (LOB) applications and document repositories and share a subset of this information externally we need to ensure that we don't expose sensitive data to unauthorised individuals. For example, an application could be developed to extract and link data from both a document management system (DMS) and a human resource (HR) application. If several documents were written by the management team over the past

L. Aroyo et al. (Eds.): ISWC 2011, Part II, LNCS 7032, pp. 293–300, 2011.

month, discussing future plans for the company, terms such as "takeover", "acquisition", "redundancy", "office closure", "pay cuts" and "lay-offs" may feature prominently in a list of last month's frequent phrases. Worse still, by integrating the data from both systems it might be possible to determine the offices that are scheduled to be closed and the individual employees that will be laid off. Much of the information within the enterprise is highly sensitive and as such the access policies of the underlying data source applications need to be extracted, represented and enforced in the linked data network.

In recent years, there has been a great deal of research into the use of policies for access control [1], [2] and [3]. The term "policy", in this context, is used to refer to the access control model which describes the blueprint, the policy language which defines both the syntax and the semantics of the rules and the framework which is a combination of the access control model, the language and the enforcement mechanism. Natural language, programming languages, XML based languages and ontologies can all be used to express policies. XML and ontologies are two popular choices for representing policy languages as they benefit from flexibility, extensibility and runtime adaptability. However, ontologies are better suited to modelling the semantic relationship between entities. In addition, the common framework and vocabulary used by ontologies, to represent data structures and schemas, provides greater interpretability and interoperability. Regardless of the language chosen, a logic based underlying formalisation is crucial for automatic reasoning in a linked data network. To date, very little research has been done on the elevation and the representation of policies in the linked data network. Many of the proposed access control models lack a formal enforcement framework.

In this paper, we identify the need for a policy language which can represent multiple access control models and a framework which will allow enterprises to abstract information from existing content sources and publish subsets of this data to the linked open data (LOD) cloud. With a view to addressing the aforementioned issues our research focuses on the:

- Examination of existing enterprise content and access control models, with a view to proposing an integrated model
- Analysis of the expressivity of existing policy languages and their ability to represent multiple access control models
- Identification of a reasoning and enforcement framework taking into consideration flexibility, interoperability, scalability and usability
- Development of a methodology for the automatic lifting of policies from the individual data source to the linked data network
- Implementation of a prototype in order to evaluate the overall effectiveness of the access control framework

The remainder of the paper is structured as follows: In section 2, we detail our approach, the chosen research methodology and our evaluation strategy. In section 3, we describe the shape of the literature with respect to related access control models, languages and enforcement frameworks. Finally, we conclude and give the direction of our future work in section 4.

2 Research Objectives and Plan

2.1 Approach

In this section we detail our strategy for the conceptualisation and realisation of a linked data access control model, policy language and enforcement framework.

Access Control Model. Our initial objective is to examine existing access control models and to generate an integrated model which can be used to represent the access control policies of a linked data network either inside or outside the enterprise. Our analysis will consider heterogeneous access policies from existing enterprise content sources. It will be designed to be flexible enough to cater for both authenticated and non authenticated users. In addition, the model will take into account existing standards such as the initiative by the OASIS Integrated Collaboration Object Model (ICOM)[1] and the Distributed Management Task Force (DMTF) Common Information Model (CIM)[2].

Policy Language. Once we have produced the integrated access control model we will analyse existing policy languages to determine their suitability for representing the authorisations, obligations and rules in a linked data network and their degree of support for our integrated access control features. In order to facilitate reasoning over policies in the linked data network, the policy language will need to have underlying formal semantics. The language must also provide support for propagation based on the semantic relationship between all entities and both positive and negative authorisations and obligations. Our analysis will also take into account non functional requirements such as flexibility, extensibility, runtime adaptability and interoperability. We will examine the level of support without need for change, the adaptability of the policy language and the scope of the changes required. Depending on the results of our analysis we will choose the policy language that most closely meets our needs or, if appropriate, we will integrate and extend components from one or more languages.

Enforcement Framework. We will examine each of the reasoning and enforcement frameworks based on functional requirements identified in the literature as important for the representation and enforcement of access control policies. These are namely: propagation and inference [1] [2] [3] [4], conflict management [1] [2] [3] [4], distributivity [2] [3], exception handling [1] [2] and support for standards [5]. Given our use case investigates the lifting and representation of data from existing distributed content sources we will add provenance, policy lifting and policy merging to the list of criteria. In addition, we will consider non functional requirements such as flexibility, interoperability, scalability and usability. Our objective is to propose a new framework which builds on existing work and addresses any shortcomings.

[1] http://www.oasis-open.org/committees/icom/
[2] http://www.dmtf.org/standards/cim

In order to examine the effectiveness of the model, language, lifting methodology and framework we plan to implement a prototype. Figure 1, provides a high level overview of the components that make up the proposed information analysis and knowledge discovery platform. Domain data and relationships will be modelled as ontologies in the abstraction layer. The LOD cloud will be used to enrich the semantics of both the enterprise and access control models. Instance data will be abstracted from enterprise content repositories and LOB applications, stored in the enterprise data integration layer and appropriate indexes will be generated. The enforcement framework in turn will be composed of several interfaces, a rules engine, a query engine and local storage for both ontologies and access control rules. The policy integration interface will facilitate the retrieval of existing data access policies. Data from existing content sources will be obtained through the data integration interface. The policy specification interface will provide authenticated access to the policies and the data governed by the relevant policies. All data access queries will be intercepted by the query reasoning engine which will grant or deny access to data based on a combination of rules and inference.

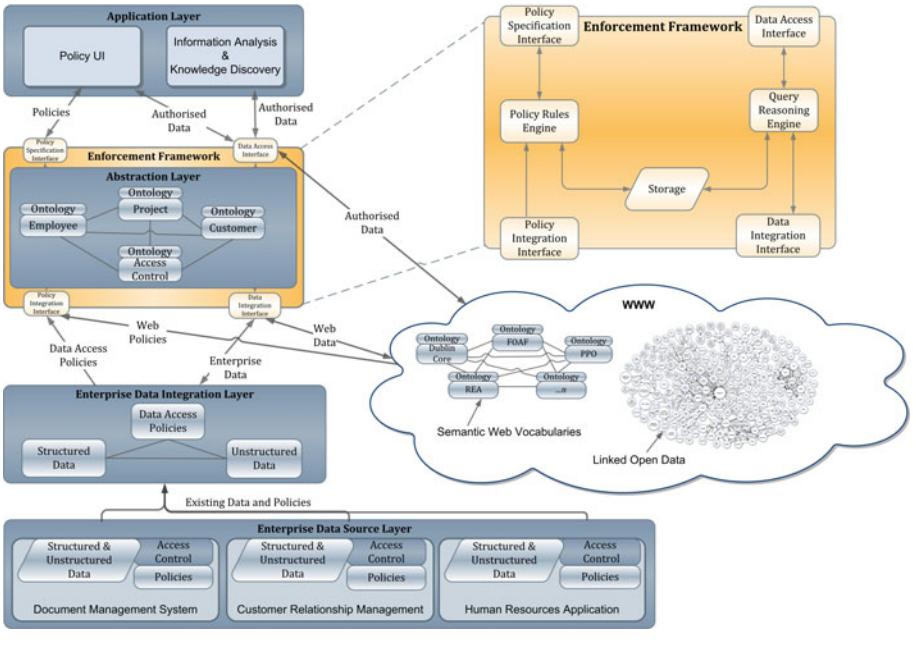

Fig. 1. Information analysis and knowledge discovery platform

We will investigate the applicability of existing standard technologies for conceptualization as they offer many well known benefits such as interoperability, extensibility, scalability and re-usability. We will utilise a number of methods in order to determine the best way to extract access constraints from the LOB

applications. Through experimentation we aim to development a methodology which can be used to as a guideline for the lifting of access control policies from the individual data sources to the abstraction layer. Many researchers are currently investigating the representation and enforcement of privacy on the web in general and in semantically aware environments in particular. Although we do not plan to specifically address this problem our prototype will take into consideration current work in this area.

2.2 Methodology

We have chosen two complimentary methodologies, the Design Science Research (DSR) methodology proposed by Peffers *et al* [6] and Action Research (AR) as defined in [7], to guide our research. DSR involves the identification of the research problem and the motivation behind the research, the specification of the objectives and the development of a research artefact to meet these objectives. The artefact is subsequently evaluated and the problem, process, outcomes etc are communicated to academics and practitioners alike. AR is an iterative process which involves the identification of a problem, the taking of action and the examination of the outcomes of the action. If the examination determines that the problem has not been solved, the researcher takes what they have learned and further refines the problem and takes further action in an attempt to resolve the problem. This process continues until the researcher is satisfied that they have solved the problem or concludes that the problem can't be solved. The relationship between AR and DSR is multifaceted e.g. AR can be wrapped around DSR, DSR can be wrapped around AR or a number of AR iterations could form part of one or more individual DSR steps. In our research project we have adopted the DSR methodology to guide the overall research from conception through to conclusion. The AR methodology will in turn be used in the design and development step of the DSR methodology to enable the refinement of both the access control framework and the lifting methodology through iterative reflection.

2.3 Evaluation

A number of evaluation methods will be used throughout the research project, both artificial and naturalistic. The access control model, language, enforcement framework and the lifting methodology will be constructed and evaluated in an iterative manner as part of the AR methodology. The model and the policy language will be evaluated using qualitative means, such as interviews and desk studies, to confirm that they are flexible enough to support both authenticated and non authenticated users and a variety of access control models. The lifting methodology will be evaluated using precision and recall metrics to determine the effectiveness of the methodology. Finally the enforcement framework will be evaluated in our real world use case scenario through the implementation of a prototype to ensure that it addresses both the functional and non functional requirements highlighted in our approach as important for a knowledge discovery platform. A combination of experiments and interviews will be used to evaluate

the expressivity of the policy language and scalability, performance and usability of our prototype.

3 Related Work

Access Control Models. Role-based access control (RBAC), task-based authorisation control (TBAC) and attribute-based access control (ABAC) are three of the most common approaches proposed for restricting access to information systems. In RBAC a number of roles are defined, users are assigned to appropriate roles and access to resources is granted to one or more roles [8]. RBAC policies have been the de facto authorisation mechanism used in enterprises for many years. The TBAC model proposed by [9] facilitates task oriented access controls required for workflow and agent based distributed systems. Essentially, TBAC models access controls from a task-oriented as opposed to a subject-orientated perspective. ABAC [10], which was also designed for distributed systems, provides an alternative means of access control where the requester is unknown prior to the submission of the request. ABAC grants or denies access to resources, based on attributes of the requester and/or the resource. Early work on semantic access control policies by [11] was based on RBAC. However, most of the recent work in this area is based on ABAC [4], [1], [2]. We propose an integrated approach which will include a number of access control models and will be flexible enough to represent access control both inside the enterprise and on the web.

Policy Language. Policy languages can be categorised as general or specific. In the former the syntax caters for a diverse range of functional requirements (e.g. access control, query answering, service discovery, negotiation etc...), whereas in contrast the latter focuses on just one functional requirement. Two of the most well-known access control languages, KAos [12] and Rei [13], are in fact general policy languages. A major drawback of both KAos [12] and Rei [13] is they have little or no support for policy propagation. Although Kolovski *et al* [14] propose a description logics formalisation for XACML it also does not consider policy propagation based on the semantic relationship between entities. Concept-level access control (CLAC) [4], semantic-based access control (SBAC) [1] and the semantic network access control model proposed by [2] all focus on the policy specification and reasoning in the conceptual layer of the semantic aware environment. Qin *et al* [4] allow propagation of access controls based on the semantic relationships among concepts. [1], [2] and [3] enhance the semantic relations by allowing propagation based on the semantic relationships between the subjects, objects, and permissions. The framework proposed in Amini *et al* [3] enables the specification of policy rules in both conceptual and ground levels. However [3] specifically states that access controls should be based on attributes as opposed to identity. In open distributed environments such as the web ABAC is the ideal choice however, in an enterprise setting we can't ignore the RBAC that already exist in the underlying data sources. None of the authors consider either TBAC or a combination of access control models.

Enforcement Framework. A framework is made up of the syntax of the policy language, the semantics of the language and the execution model. In addition to providing a mechanism to represent and enforce the criteria specified by the policy constructs, features such as distributivity, propagation and inference, policy merging, conflict management, exception handling and provenance need to be catered for by the framework. At first glance it may appear as if well known access control frameworks such as KAos [12], Rein [13] and PROTUNE [5] support many of the aforementioned features. However, as their underlying policy languages do not consider semantic relations between all of the policy entities the frameworks themselves can not fully exploit propagation and inference based on subject, objects and actions. The primary issues with (CLAC) [4] are the lack of both an underlying formalism and an enforcement mechanism. The policy language proposed by [1] is not based on well defined semantics and their solution does not scale well due to their centralised approach. Although [3] caters for many of the features listed above it does not support exception policies and more importantly their policy language is not based on standards and as such raises questions with respect to usability and interoperability. The framework detailed in [2] is the closest match for our requirements. However, they do not take existing policy representation and provenance into consideration and it is not clear the extent of their use of existing standard based technologies. [1] and [3] propose access control models and enforcement frameworks for linked data networks. However they both dismissed RBAC stating that is it not suitable for semantically aware environments. We argue that linked data network access control framework must cater for various access models e.g. RBAC, TBAC and ABAC. In addition, in an enterprise setting it makes sense to populate the linked data network based on data stored in content repositories and LOB applications. None of the authors have provided any guidance with respect to the automatic population of the linked data network in general and the access controls in particular, both of which we plan to address in our research.

4 Conclusions and Future Work

Our research explores the application of access control to a linked data network. We identify the need for a policy language which can represent multiple access control models and highlight the importance of representing existing access control policies in the linked data network. The research problem, motivation, objectives and the approach have been described in this paper. The next step is to analyse existing access control models, languages, frameworks and standards in detail and to generate a detailed design based on both the requirements and the results of our in depth analysis. The prototype will subsequently be developed based on the design. Once we have completed the development of the prototype we will investigate if the artefact meets the objectives or if further iterations of design, development and reflection are required.

Acknowledgements. This work is supported in part by the Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2), the Irish Research Council for Science, Engineering and Technology Enterprise Partnership Scheme and Storm Technology Ltd.

References

1. Javanmardi, S., Amini, M., Jalili, R., GanjiSaffar, Y.: SBAC: A Semantic Based Access Control Model. In: 11th Nordic Workshop on Secure IT-systems (NordSec 2006), Linkping, Sweden (2006)
2. Ryutov, T., Kichkaylo, T., Neches, R.: Access Control Policies for Semantic Networks. In: 2009 IEEE International Symposium on Policies for Distributed Systems and Networks, pp. 150–157. IEEE (July 2009)
3. Amini, M., Jalili, R.: Multi-level authorisation model and framework for distributed semantic-aware environments. IET Information Security 4(4), 301 (2010)
4. Qin, L., Atluri, V.: Concept-level access control for the Semantic Web. In: Proceedings of the 2003 ACM Workshop on XML Security - XMLSEC 2003, Number Cimic, p. 94. ACM Press (2003)
5. Bonatti, P.A., De Coi, J.L., Olmedilla, D., Sauro, L.: Rule-Based Policy Representations and Reasoning. In: Bry, F., Małuszyński, J. (eds.) Semantic Techniques for the Web. LNCS, vol. 5500, pp. 201–232. Springer, Heidelberg (2009)
6. Peffers, K., Tuunanen, T., Rothenberger, M.: A design science research methodology for information systems research. Management Information Systems 24, 45–77 (2007)
7. Checkland, P., Holwell, S.: Action Research: Its Nature and Validity. Systemic Practice and Action Research 11(1), 9–21 (1998)
8. Sandhu, R., Coyne, E., Feinstein, H., Youman, C.: Role-based access control: a multi-dimensional view. In: Tenth Annual Computer Security Applications Conference, pp. 54–62 (1994)
9. Thomas, R., Sandhu, R.: Task-based authorization controls (TBAC): A family of models for active and enterprise-oriented authorization management. Database Security, 166–181 (1998)
10. McCollum, C., Messing, J., Notargiacomo, L.: Beyond the pale of MAC and DAC-defining new forms of access control. In: Proceedings of 1990 IEEE Computer Society Symposium on Research in Security and Privacy, 1990, pp. 190–200. IEEE (1990)
11. Yague, M., Maña, A., López, J., Troya, J.: Applying the semantic web layers to access control. In: Proceedings of 14th International Workshop on Database and Expert Systems Applications, 2003, pp. 622–626. IEEE (2003)
12. Bradshaw, J., Dutfield, S., Benoit, P., Woolley, J.: KAoS: Toward an industrial-strength open agent architecture. In: Software Agents, pp. 375–418 (1997)
13. Kagal, L., Finin, T.: A policy language for a pervasive computing environment. In: Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks, pp. 63–74. IEEE Comput. Soc. (2003)
14. Kolovski, V., Hendler, J., Parsia, B.: Analyzing web access control policies. In: Proceedings of the 16th International Conference on World Wide Web WWW 2007, p. 677 (2007)