Gouvernance des données basée sur le web sémantique

Nouha Laamech[1] Manuel Munier[2] Congduc Pham[3]

laamech.nouha@univ-pau.fr manuel.munier@univ-pau.fr congduc.pham@univ-pau.fr

[1,2] IUT des Pays de l'Adour
Universite de Pau et des Pays de l'Adour
E2S UPPA, LIUPPA, Mont-de-Marsan, France

[3] Universite de Pau et des Pays de l'Adour
E2S UPPA, LIUPPA, Pau, France

THÈMES – *Informatique - Droit*

RÉSUMÉ –

MOTS-CLÉS – *Le présent article décrit notre approche pour la gestion du partage des données dans les environnements connectés. L'exploitation des données générées par les ressources de l'internet des objets ou IoT soulève des risques de sécurité en raison du manque de transparence entre les différents acteurs de l'environnement. Ainsi, nous proposons une nouvelle solution de gestion du partage des données appelée IdSM (IoT data sharing management) basée sur le web sémantique, le modèle de contrôle d'usage OrBAC (Organization-Based Access Control) et les techniques de provenance, afin de répondre aux préoccupations de sécurité de l'information dans les environnements IoT.*

1 Introduction

Nos travaux de recherche se focalisent principalement sur trois questions complémentaires (i) la définition d'une couche sémantique qui décrit les exigences en matière de sécurité de l'information pour la gestion de partage des données (ii) l'application d'une politique de sécurité qui prend en compte à la fois les préférences d'utilisation du propriétaire des données, mais également les conditions de service du consommateur de données (iii) la construction d'une solution de provenance de données IoT dans une architecture décentralisée en éliminant le besoin de faire confiance à des tierces parties.

Afin de répondre à ces problématiques nous avons proposé dans un premier lieu une ontologie appelée IdSM-o [4] qui modélise les métadonnées ainsi que les contraintes d'utilisation des données dans les environnements IoT. Ensuite, afin d'aligner sémantiquement les exigences en matière de gestion du partage entre les producteurs des données et les consommateurs des données, nous avons étendu l'ontologie par des règles sémantiques d'inférence qui génèrent un package contenant une observation, ses métadonnées et sa licence. Cette licence décrit comment les données doivent être manipulées une fois partagées avec un consommateur donné. Afin de garantir le respect de cette politique, nous introduisons l'utilisation des techniques de provenance de données afin de prendre en compte d'où vient la donnée, dans quelle contexte est-elle utilisée, ainsi que sa qualité.

Dans cet article, nous présentons une approche qui fusionne à la fois la représentation sémantique d'un environnement IoT, les politiques de contrôle d'usage et les solutions de traçabilité.

2 Exigences de sécurité pour le partage de l'information

Sur la base des préoccupations actuelles en matière de gestion des données, nous pouvons identifier les concepts clé suivants de notre système de sécurité de l'information :

- Génération de la licence : une licence est générée sur la base des exigences des producteurs des données, établies en suivant un modèle OrBAC [1] et est attribuée aux données IoT, de sorte que la licence restera attachée aux données, peu importe où elles seront diffusées.
- Gouvernance des données : notre travail rejoint le " Data Governance Act" (DGA) [2] de la Commission européenne, qui vise à promouvoir le partage des données en créant un environnement numérique de confiance entre les différents acteurs. Cette vision peut s'appliquer à différents domaines : agriculture, énergie, villes intelligentes, etc.
- Bien commun : notre solution favorise un environnement IoT où différentes ressources peuvent être partagées au sein de la communauté au profit d'un intérêt général.

3 Solution proposée : système de gestion du partage des données

3.1 Scénario motivant

Nous illustrons nos idées dans le contexte de l'agriculture connectée, mais notre solution est agnostique en termes d'applications et peut être appliquée dans d'autres contextes IoT. Nous définissons Sam comme un individu de la classe "Producteur de données" possédant un capteur de météo pour son champ de blé, qui est une instance de la classe sosa :Sensor. Ce capteur a deux sorties : le taux d'humidité et la température. Une règle de préférence d'utilisation est définie par Sam pour chacune des deux observations du dispositif. En effet, Sam accorde aux agriculteurs la permission d'utiliser l'observation "température" comme ils le désirent. Pour l'observation "humidité", il souhaiterait cette fois-ci que seules les personnes ayant le rôle "chercheur" puissent avoir le droit de la "consulter".

Alice, Aline, Bob et Ben sont des "consommateurs de données" qui souhaiteraient utiliser les observations partagées. Aline a le rôle "Farmer". Alice a le rôle de "Nursery_wroker", qui est un sous-rôle du rôle "Farmer". Bob à le rôle "Professor" et Ben a le rôle "Researcher", les deux sont des sous-rôles de "Academy". Nous définissons quatre types de requêtes qui impliquent la création d'instances de la classe Operation.

- Opération 1 : Bob demande l'accès à l'observation "humidité"
- Opération 2 : Ben souhaiterait apporter des modifications à l'observation "Humidité"
- Opération 3 : Aline souhaiterait lire et modifier l'observation "Temperature"
- Opération 4 : Alice souhaiterait rediffuser l'observation "Temperature"

3.2 Politique de contrôle d'usage et provenance de données

Notre approche est divisée en deux parties :
(1) D'abord, nous utilisons OrBAC comme outil de modélisation du contrôle d'usage pour aider les producteurs de données à définir des règles d'utilisation. En utilisant ce modèle, chaque utilisateur peut définir des règles de sécurité pour spécifier si certains rôles sont autorisés, interdits ou requis pour effectuer certaines activités sur certaines vues. L'avantage d'OrBAC est que ses règles ne sont pas statiques mais dépendent également de conditions contextuelles permettant d'activer ou de désactiver certaines règles. Une fois ce modèle défini, en utilisant l'ontologie IdSM-o, nous pouvons traduire ces exigences d'usage en matière de sécurité de l'information énoncées par les producteurs de données en règles SWRL (Semantic

Web Rule Language) [3], et ainsi exploiter toute la puissance du web sémantique.
(2) Une fois les exigences de l'utilisateur définies, nous utilisons la provenance des données pour vérifier que ces conditions sont effectivement bien respectées. Plus simplement, la provenance des données permet de répondre à des questions telles que "comment ont-elles été produites", "où sont elles distribuées", et "par qui ont-elles été produites". La provenance des données est une métadonnée associée aux obsevations, qui détaille l'origine, les modifications et les détails permettant de confirmer la crédibilité ou la validité des données. Cette démarche est importante pour repérer les utilisations inhabituelles et les attribuer aux sources. En outre, la provenance des données peut être utile pour la création de rapports de recherche ou des processus d'audit. Exemple de figure

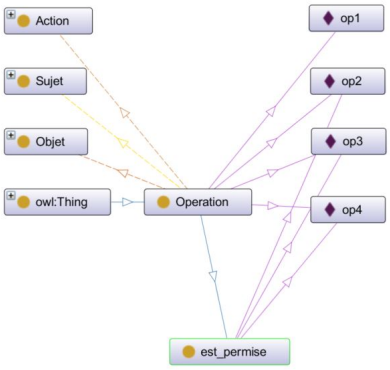

FIGURE 1 – Résultat après le lancement du moteur d'inférence

Après avoir lancé le moteur d'inference, les politiques d'utilisation appropriées sont dérivées et l'une des dérivations obtenues est illustrée à la figure 1. Le système a déduit, à partir des règles de Sam, que opération 2, opération 3 et opération 4 sont des opérations permises. Ce résultat prouve que notre processus de raisonnement réussit à déduire une politique de sécurité en cas de correspondance entre les règles de préférences et les requêtes des consommateurs de données.

4 Architecture de notre solution

La figure ci-dessous montre l'architecture que nous proposons et qui permet de gérer le partage des données en utilisant l'approche définie précédemment. L'architecture implique deux acteurs principaux, à savoir le consommateur de données et le producteur de données. D'abord, le producteur de données précise les règles de partage qui concernent chacune de ses observations. Sur la base de ces exigences, une licence est générée et restera attachée à l'observation où qu'elle soit partagée.

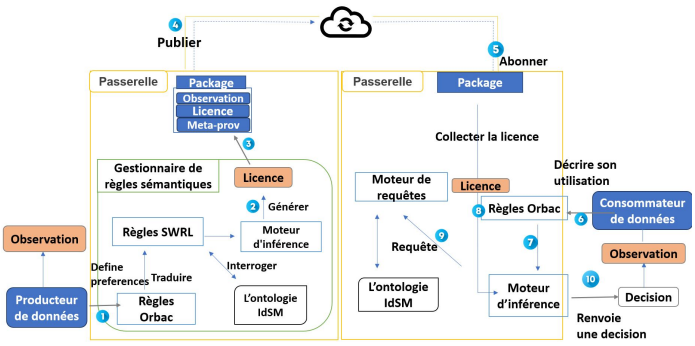

FIGURE 2 – Architecture de notre solution

Une fois que le consommateur de données a défini à son tour son usage, il reçoit le package verrouillé qui contient l'observation, sa licence, ainsi qu'un certain nombre de meta données ajoutées au fur et à mesure de l'acheminement de l'information. Le moteur d'inférence décide à partir de la licence et du descriptif d'utilisation du consomateur de données d'accorder ou non l'autorisation d'accès à l'observation contenue dans le package.

5 Conclusions et perspectives

Nos travaux de recherches sont en cours de développement, et nous sommes actuellement engagés dans la réalisation d'un système de provenance des données comme mécanisme de vérification de la sécurité. Ainsi, lorsqu'un producteur de données partage une observation avec sa licence d'utilisation, il peut contrôler depuis sa propre passerelle si les conditions d'utilisation qu'il a définies sont effectivement respectées par les consommateurs de données.

6 Remerciements

Les auteurs sont reconnaissants pour le soutien financier du Conseil Départemental des Landes.

Références

[1] Cuppens Frédéric and Miège Alexandre, *OrBAC, Organization Based Access Control*, The Review of Politics, Journées Druide (2004)

[2] European Commission, *Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)*, 2020.Vol. 55. COM/2020/767 final. Brussels, Belgium :.

[3] I. Horrocks, P.F. Patel-Schneider, H. Boley, Said Tabet, Benjamin Grosof, and Mike Dean, *SWRL : A semantic web rule language combining oWL and ruleML.* , W3C Member submission 21 (01 2007).