Pour une gouvernance des données dans les environnements connectés

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
LIUPPA, E2S UPPA, Mont-de-Marsan, France

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

Thèmes – *Informatique - Droit*

Résumé – *Le présent article décrit notre approche pour la gestion du partage des données dans les environnements connectés. Ainsi, nous utilisons la traçabilité des données comme outil de vérification du bon respect de la licence, établie au préalable entre le fournisseur de donnée et son consommateur. Cette licence contient la liste des conditions mises en place par l'utilisateur, qui sont à respecter lors de l'utilisation de ses données. Nous souhaitons permettre d'une part aux utilisateurs d'être certains que leurs ressources sont convenablement utilisées, mais également permettre aux consommateurs de données de s'assurer qu'ils respectent bien la loi. La mise en place de notre proposition permettra de créer un monde numérique sécurisé pour tout le monde.*

Mots-clés – *IoT, provenance des donnés, sécurité de l'information, web sémantique.*

1 Introduction

L'internet des objets (IoT) est une composante majeure du monde connecté. Avec les milliards d'appareils connectés à Internet et l'interconnexion de nombreuses sources de données dans divers secteurs, le partage et la collecte des données constituent aujourd'hui une dynamique propre à promouvoir l'intérêt général et à favoriser le progrès technologique. Dans ce contexte, il existe un enjeu commun à inciter les utilisateurs des environnements connectés à partager leurs données, afin d'améliorer la qualité de vie et de bénéficier des différents services qui émergent dans différents domaines, tels que les transports publics, l'écologie ou encore, l'agriculture connectée. Or, en raison d'un manque de contrôle sur le partage des données et du manque de confiance entre les différentes parties impliquées dans les environnements connectés, les utilisateurs sont de plus en plus réticents à partager leurs ressources avec des systèmes d'informations extérieurs. Par ailleurs, afin de prendre les mesures appropriées, il est important que les consommateurs de ces ressources puissent également avoir confiance dans la fiabilité et l'exactitude des informations communiquées par le système. Ainsi, la sécurité des données est devenue un défi important de nos jours et une préoccupation de taille, notamment avec la pression croissante de la législation.
La provenance des données consiste à sauvegarder, gérer et récupérer des informations sur l'origine et l'historique d'une donnée [2]. Notre proposition est d'introduire les concepts de provenance des données dans le domaine de l'IoT afin de créer des systèmes fiables et dignes de confiance, en enregistrant les informations les plus basiques du système IoT aux plus complexes. La provenance des données est une métadonnée et, en tant que telle, elle nécessite un modèle de données qui décrit les informations spécifiques collectées. Dans cette contribution, nous présentons une approche qui fusionne à la fois la représentation sémantique d'un environnement IoT, et les solutions de traçabilité.

2 Sécurité et gouvernance des données

Des techniques de sécurité et de confidentialité connues, telles que le cryptage, le contrôle d'accès et l'anonymisation, sont appliquées pour garantir la préservation des données brutes capturées au cours de la première couche de l'architecture de l'IoT [3] et éviter les attaques de sécurité. À chaque étape du cycle de vie des données IoT au sein de ces couches, des mécanismes de sécurité classiques sont mis en œuvre pour préserver l'efficacité et l'intégrité du service. Cependant, beaucoup moins de travaux sont ciblés sur les aspects concernant la qualité de l'information capturée et sa gouvernance lors de son partage entre les systèmes d'information. Or, dans les environnements connectés actuels, ce sont là de nouvelles vulnérabilités qu'il est indispensable de prendre en compte.
La stratégie de gouvernance des données est un autre ingrédient clé qui définit la manière dont les informations extraites des données sont partagées, impose une culture d'utilisation des données et révèle les inconvénients auxquels la gouvernance des données peut être confrontée ainsi que le budget nécessaire [1]. Plus précisément, elle définit qui est responsable, redevable et informé des données et comment les décisions seront prises à partir de celles-ci. Elle fait également partie intégrante du processus visant à surmonter les limites de la gestion des risques liés à la sécurité de l'information et aide à atteindre les objectifs et les valeurs attendus par les entreprises. Ainsi, la gouvernance des données fournit la base des processus de gestion des données à suivre par l'ensemble du système.

3 Approche

3.1 Gestionnaire sémantique

Les menaces visant la protection de la confidentialité et les contraintes législatives obligent le propriétaire des données ainsi que les demandeurs à être hésitant vis-à-vis de la notion du partage de données. Dans cette optique, et selon une approche similaire à celle d'une licence, la première partie de notre proposition consiste à donner aux responsable de données le droit de fixer les exigences que les autres entités doivent respecter pour pouvoir utiliser leurs ressources. Cependant, la correspondance entre les préférences du propriétaire des données et la demande d'accès du demandeur de la donnée implique l'utilisation du même vocabulaire que celui qui décrit les exigences d'utilisation. Cette correspondance permet la création d'une politique commune qui peut être appliquée pour protéger la souveraineté du propriétaire des données dans l'environnement IoT. Ces défis ne sont pas limités aux données à caractère personnel, mais portent sur un champ plus large et concernent tout type de données, comme les données industrielles. Afin d'aligner la communication entre les entités IoT concernées, il est nécessaire de définir une description sémantique formelle en utilisant la modélisation ontologique pour décrire les propriétés des données échangées, afin d'appliquer un langage de règles d'expression comme les modèles de contrôle d'usage, les modèles de contrôle d'accès ou les accords de niveau de service basés sur des règles [4] .

3.2 Modèle de provenance de données

La provenance des données offre une solution potentielle pour répondre aux problèmes mentionnés ci-dessus, en stockant des informations sur l'origine des données, les transactions effectuées sur les données et l'historique du traitement depuis sa source initiale jusqu'à son l'état actuel. Ainsi, nous pouvons distinguer deux types de traçabilité :

— Traçabilité ascendante : les consommateurs de don-

nées peuvent retracer l'origine d'un élément de données et déterminer si celle-ci répond aux exigences techniques et juridiques d'une activité donnée et d'en évaluer la qualité.

— Traçabilité descendante : les fournisseurs de données doivent être informés de l'endroit où leurs données sont distribuées et des processus dans lesquels elles sont impliquées, afin de ne pas perdre complètement leur contrôle.

En assurant la traçabilité de l'origine des données, et en combinaison avec des méthodes cryptographiques pour vérifier l'intégrité des métadonnées de provenance, nous visons à fournir un mécanisme générique pour assurer l'exactitude et l'intégrité des applications IoT et ainsi renforcer leur fiabilité pour les cas d'utilisation sensible.

4 Cas pratique : agriculture connectée

Comme cas d'application, nous travaillons sur l'agriculture connectée. Les agriculteurs disposent de plusieurs données relatives à leurs terrains particuliers ainsi qu'à leurs exploitations : dose d'irrigation, structure des vignes, nutriments agricoles/céréales utilisés pour la production des grains, superficie agricole toujours en herbe, etc. Le partage anonyme de ces données avec la communauté permettrait aux agriculteurs de comparer leur production avec celle de leurs voisins, tout en conservant leurs propres objectifs. En parallèle, d'autres organismes pourraient également accéder à ces informations afin de proposer à leur tour des services comme, entre autres, estimer les aides financières collectives et individuelles, proposer une politique d'aménagement de territoire spécifique, ou encore de réfléchir à des procédures anti-gaspillage et anti-pollution.

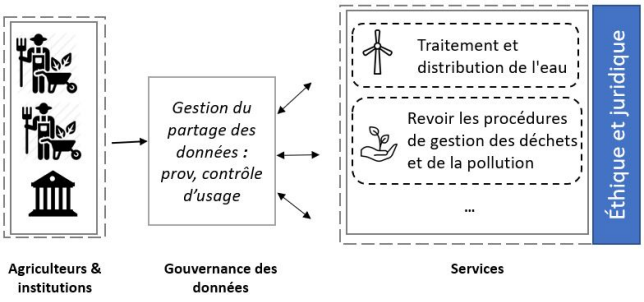

FIGURE 1 – Exemple de l'agriculture connectée

Les systèmes IoT sont des communautés évolutives sans véritable encadrement conventionnel. Dans la figure 1, le partage des données collectées permet d'améliorer la fourniture de services avancés dans un large champ de domaines d'application. De multiples acteurs sont impliqués et chacun cherche à utiliser les mêmes données pour atteindre ses objectifs propres. Cependant, le manque de gouvernance des données IoT fait que ces acteurs n'ont pas la certitude qu'ils remplissent les conditions requises pour exploiter les ressources fournies. Par ailleurs, le propriétaire de l'entité n'a guère de contrôle sur ses données une fois qu'elles sont partagées, et perd donc rapidement sa légitimité. Tout cela crée un environnement IoT potentiellement vulnérable qui manque de transparence entre ses acteurs et peut perturber les services envisagés.

5 Conclusion

Les modèles dirigés par les données connaissent une croissance considérable et deviennent un atout essentiel dans tous les secteurs. Dans ce contexte, la gouvernance des données a évolué pour passer d'une problématique de confidentialité à une véritable préoccupation multidimensionnelle ayant des implications dans divers secteurs, notamment l'économie, le maintien de l'ordre, et même la géopolitique. La mauvaise gestion des données des usagers en raison d'une juridiction limitée, le manque de confiance des utilisateurs et l'accès limité aux données sont des problèmes qui peuvent être résolus par la mise en place d'une solution de gouvernance des données solide et performante.

Notons également que ces travaux sont en parfaite adéquations avec ceux actuellement menés par la Commission européenne sur sa "stratégie européenne pour les données" et le "*Data Governance Act*" [5] .

6 Remerciements

Les auteurs sont reconnaissants pour le soutien du Conseil Départemental des Landes qui finance l'allocation doctorale de Nouha Laamech.

Références

[1] A. Berson, L. Dubov, B.K. Plagman, P Raskas, *Master Data Management and Data Governance.*.McGraw-Hill (2011)

[2] Buneman, P., Khanna, S., Wang-Chiew, T., *Why and where : A characterization of data provenance.*, In : Intl. Conf. on Database Theory. pp. 316,330. Springer (2001)

[3] M. Wu, T.-J. Lu, F.-Y. Ling, J. Sun, and H.-Y. Du, *Research on the architecture of internet of things*, in 2010 3rd international conference on advanced computer

[4] A. Paschke, *Rbsla : Rule-based service level agreements,*, Ph.D. dissertation,Technische Universit¨at M¨unchen, 2007

[5] European Commission *Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)* , Brussels, Belgium : European Commission.