HAL open science

Détection des attaques de confiance dans l'Internet des Objets Social

Wafa Abdelghani, Florence Sèdes, Corinne Amel Zayani, Ikram Amous

To cite this version:

Wafa Abdelghani, Florence Sèdes, Corinne Amel Zayani, Ikram Amous. Détection des attaques de confiance dans l'Internet des Objets Social. 38éme Congrès INFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID 2020), 2020, Dijon, France. pp.155-170. hal-03184015

HAL Id: hal-03184015

https://hal.science/hal-03184015v1

Submitted on 29 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection des attaques de confiance dans l'Internet des Objets Social

Wafa Abdelghani [1], **Florence Sèdes** [1], **Corinne Amel Zayani** [2], **Ikram Amous** [2]

1. IRIT, Université Paul-Sabatier, Toulouse, France
2. Miracl, Université de Sfax, Sfax, Tunisia

RÉSUMÉ. *L'Internet des Objets Social (SIoT) est un paradigme dans lequel l'Internet des Objets (IoT) est fusionné avec les réseaux sociaux. Dans ce type d'environnement, les participants sont en compétition afin d'offrir une variété de services attrayants. Néanmoins, certains d'entre eux ont recours à des comportements malveillants afin de propager des services de mauvaise qualité. Ils lancent ce qu'on appelle des attaques de confiance et brisent les fonctionnalités de base du système. Plusieurs travaux de la littérature ont traité ce problème et ont proposé différents modèles de confiance. Néanmoins, ces derniers proposent de classer les meilleurs noeuds du réseau SIoT. Ils ne permettent pas de détecter les noeuds malveillants. Pour remédier à ce problème, nous proposons un nouveau modèle de gestion de la confiance, capable de détecter et bloquer les noeuds malveillants afin d'obtenir un système fiable et résilient.*

ABSTRACT. *The Internet of Things Social (SIoT) is a paradigm where the Internet of Things (IoT) is merged with social networks. In this type of environment, participants compete to offer a variety of attractive services. Nevertheless, some of them resort to malicious behavior in order to spread poor quality services. They commit so-called trust-related attacks and break the basic functionality of the system. Several works in the literature have addressed this problem and have proposed different trust-models. Nevertheless, they propose to classify the best nodes of the SIoT network. They do not detect different types of trust attacks or malicious nodes. To address this problem, we propose a new trust evaluation model able to detect and block malicious nodes in order to ensure a reliable and resilient system.*

MOTS-CLÉS : *Internet des Objets Social, Réseaux sociaux, Gestion de la confiance, Attaques de confiance.*

KEYWORDS: *Social Internet of Things, Social Networks, Trust Management, Trust attacks.*

DOI:10.3166/HSP.x.1-16 © 2014 Lavoisier

1. Introduction

L'Internet des Objets (IoT) est dominé par un grand nombre d'interactions entre des milliards d'objets intelligents. L'intégration de la composante sociale dans l'Internet des Objets a donné naissance à l'Internet des Objets Social (SIoT). Le SIoT

est apparu suite à un processus évolutif qui a transformé les objets du quotidien en objets pseudo-sociaux capables d'interagir avec leur environnement, puis en objets sociaux, ayant la possibilité d'établir des relations avec d'autres objets, d'une manière autonome ((Atzori *et al*., 2014)). Cette nouvelle vision a permis de simplifier la navigabilité et la découverte des ressources ((Ali, 2015)), de garantir la scalabilité comme dans les réseaux sociaux classiques ((Atzori *et al*., 2012)) offrant une source de données plus riche et plus variée ((Geetha, 2016)). Dans ce type d'environnement, les participants sont en compétition afin d'offrir une variété de services attrayants. Néanmoins, certains d'entre eux ont recours à des comportements malveillants afin de propager des services de mauvaise qualité. Ils lancent ce qu'on appelle des attaques de confiance et compromettent les fonctionnalités de base du système.

Dans la littérature, la gestion de la confiance a été largement étudiée dans divers domaines. Plusieurs travaux se sont intéressés à ce problème et ont proposé différents modèles, basés sur différents facteurs et mesures. Notre contribution se résume comme suit: Contrairement à la majorité des systèmes de gestion de la confiance existants qui se limitent à classer les meilleurs noeuds du réseau, notre objectif est de détecter les noeuds malveillants. Cela permet de les isoler et d'obtenir un système de confiance. Pour ce, nous proposons un modèle de confiance basé sur de nouveaux facteurs qui sont dérivés de la description de chaque type d'attaque. Nous proposons, également, via l'apprentissage supervisé, de combiner les différents facteurs proposés afin de distinguer les comportements malveillants de ceux qui sont légitimes.

Le papier est organisé comme suit. Dans la section 2, nous présentons une étude des travaux de la littérature qui s'intéressent à la gestion de la confiance. Dans la section 3, nous détaillons le modèle de confiance proposé. Dans la section 4, nous présentons les expérimentations qui nous ont permis de prouver la résilience du modèle d'évaluation de la confiance. Enfin, nous concluons en section 5 et indiquons nos perspectives.

2. Concepts de base

L'internet des Objets Social permet aux personnes et aux objets d'interagir dans un cadre social pour soutenir un nouveau type de navigation. La structure du réseau SIoT peut être façonnée selon les besoins afin de faciliter la navigabilité, permettre la découverte d'objets et de services et garantir la scalabilité comme dans les réseaux sociaux humains. Toutefois, la confiance doit être assurée pour tirer parti des avantages multiples de ce paradigme.

La confiance est un concept complexe utilisé dans divers contextes et influencé par de nombreuses propriétés mesurables et non mesurables telles que la croyance, la fiabilité, l'intégrité ou encore l'aptitude. Il n'y a pas de définition consensuelle de ce concept. En effet, bien que son importance soit largement reconnue, les multiples approches pour la définition de la confiance ne se prêtent pas à l'établissement de mesures et de méthodologies d'évaluation.

La confiance peut être définie comme la croyance d'une entité en une autre pour accomplir un objectif selon ses attentes. Dans l'environnement SIoT, les entités peuvent être des êtres humains, des dispositifs, des systèmes, des applications ou encore des services. La mesure de la confiance peut être absolue (par exemple, la probabilité) ou relative (par exemple, un degré de confiance). L'objectif de la confiance peut être une action ou une information.

Différents modèles d'évaluation de la confiance sont proposés pour garantir la confiance dans différents types de systèmes. Leur rôle consiste à fournir (calculer) un score de confiance, qui aidera les acteurs à prendre la décision d'invoquer ou non les services fournis par d'autres participants. Il existe plusieurs attaques qui sont conçues pour briser spécifiquement cette fonctionnalité. Nous présentons dans cette section les principales attaques de confiance citées dans la littérature (Bao *et al*., 2013 ; R. Chen *et al*., 2016 ; Abdelghani *et al*., 2016).

2.1. Les attaques de confiance dans les réseaux SIOT

Une attaque est un comportement malveillant lancé sciemment par un noeud pour détruire, bloquer ou dégrader les fonctionnalités de base d'un système. Les attaques de confiance représentent un sous-ensemble des attaques possibles dans les environnements IoT et SIoT. Dans ce type d'attaques, un noeud malveillant peut promouvoir sa propre réputation pour accéder à des fonctions supérieures ou perturber le système de manière à réduire son efficacité. Ainsi, un dispositif IoT malveillant (sous contrôle d'un propriétaire malveillant) peut effectuer les attaques suivantes.

– **Bad Mouthing Attacks (BMA)** : est une attaque dans laquelle des noeuds malveillants tentent de détruire la réputation des noeuds bienveillants (en leur donnant de mauvais votes) afin de diminuer leurs chances d'être sélectionnés comme fournisseurs de services.

– **Ballot Stuffing Attacks (BSA)** : est une attaque dans laquelle des noeuds malveillants tentent de promouvoir la réputation d'autres noeuds malveillants afin d'augmenter leurs chances d'être sélectionnés comme fournisseurs de services.

– **Self Promoting Attacks (SPA)** : est une attaque dans laquelle des noeuds malveillants, fournissant des services de mauvaise qualité, tentent de renforcer leur réputation (en s'octroyant des votes élevés) afin d'être sélectionnés comme fournisseurs de services.

– **Discriminatory Attacks (DA)** : est une attaque dans laquelle des noeuds malveillants s'attaquent à d'autres noeuds qui ne présentent pas de relation sociale forte avec eux.

Dans le tableau 1, nous proposons une spécification informelle du comportement malveillant pour chaque type d'attaque de confiance.

Dans les différents types d'attaques, c'est le noeud qui invoque un service et l'évalue ensuite qui est malicieux. En effet, tous les types d'attaques de confiance opèrent par le biais de votes erronés et non représentatifs. Ce noeud malicieux (dit invoca-

Tableau 1. Spécification informelle du comportement malveillant pour chaque type d'attaque de confiance.

	Invocateur(u_i)	**Fournisseur**(u_j)	**Interaction**(u_i, u_j)
BMA	Noeud Malicieux: - Mauvaise réputation - Services de mauvaise qualité	Noeud bénin: - Bonne réputation - Service de bonne qualité	- Grand nombre d'interactions - Majorité de votes négatifs
BSA		Noeud Malicieux: - Mauvaise réputation - Services de mauvaise qualité	- Grand nombre d'interactions - Majorité de votes positifs
SPA		Noeud Malicieux: - Mauvaise réputation - Services de mauvaise qualité	- Grand nombre d'interactions - Majorité de votes positifs - Similarité
DA	Noeud Malicieux: - Mauvaise réputation	Noeud Malicieux/ Noeud bénin	Majorité de votes négatifs

teur) est caractérisé par une mauvaise réputation dans le réseau et par des services de mauvaise qualité. En effet, ce sont ces deux caractéristiques qui font qu'il ne parvient pas à propager ses services d'une manière légitime et a recours aux attaques de confiance pour le faire. Néanmoins, dans l'attaque BMA, l'invocateur va cibler un autre utilisateur (fournisseur de service) légitime, bien réputé et offrant des services qualifiés. Dans les attaques BSA et SPA, par contre, le noeud malicieux va cibler un autre noeud malicieux (lui-même dans le cas de SPA), dans l'objectif de s'entraider. Par contre, dans l'attaque DA, le noeud malicieux choisit ses cibles de manière aléatoire, sans se soucier du fait qu'elles soient légitimes ou malicieux. De ce fait, dans cette attaque, nous ne trouverons pas un grand nombre d'interactions avec un noeud donné. Or, dans les attaques BMA, BSA et SPA, le noeud malicieux va s'acharner sur une cible donnée, ce qui se reflète par un grand nombre d'interactions avec cette dernière. Enfin, dans les attaques BMA et DA, nous retrouvons une majorité de votes négatifs, car les noeuds invocateurs ont pour objectif de ruiner la réputation d'autre noeuds. Or, dans l'attaque BSA et SPA, les noeuds malicieux cherchent à promouvoir la réputation d'autres noeuds malicieux, engendrant une majorité de votes positifs.

2.2. Evaluation et gestion de la confiance

Les mécanismes de gestion de la confiance (MGC) permettent d'assurer le processus d'établissement, de propagation et de mise à jour de la confiance (Guo *et al.*, 2017). La figure 1 montre les différentes étapes d'un MGC.

L'étape d'établissement de la confiance se base sur "un modèle d'évaluation de la confiance" qui est construit en deux étapes. **(i) L'étape de composition** consiste à

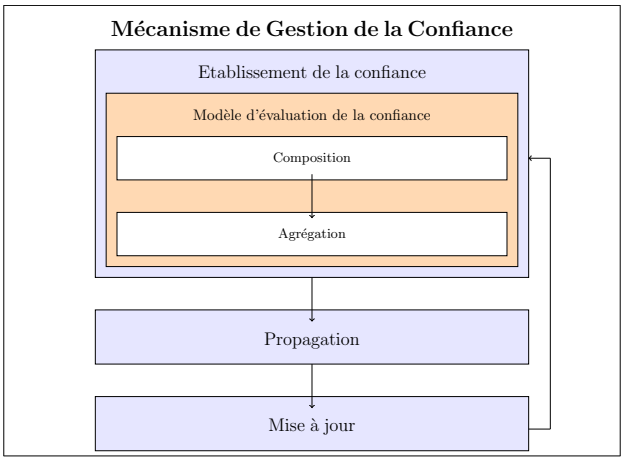

Figure 1. Architecture d'un mécanisme de gestion de la confiance

sélectionner les facteurs à prendre en compte dans le calcul des valeurs de confiance. Plusieurs facteurs ont été proposés dans la littérature, tels que l'honnêteté, la coopération, la similarité des profils, la réputation,... Ces derniers peuvent être classés selon différentes dimensions : (i) globaux ou locaux ; (ii) implicites ou explicites ; (iii) symétriques ou asymétriques. Pour les mesurer, les auteurs utilisent des informations relatives aux noeuds, telles que leur localisation ou leur historique d'interaction. **(ii) L'étape d'agrégation** consiste à choisir une méthode pour agréger les valeurs des différents facteurs afin d'obtenir la valeur de confiance finale. A cette fin, les auteurs de la littérature utilisent la moyenne pondérée, la logique floue, les modèles probabilistes, etc.

L'étape de propagation consiste à choisir une méthode pour propager dans le réseau les valeurs de confiance obtenues après l'étape d'agrégation. Deux méthodes sont utilisées. Dans la méthode dite centralisée une entité centrale fait les différents calculs pour tous les noeuds du réseau. Dans la méthode dite décentralisée, chaque noeud fait ses propre calculs. Certains travaux de la littérature utilisent la méthode de propagation centralisée, arguant que les noeuds impliqués dans les réseaux SIoT ont une capacité limitée (en termes de de calcul, de stockage, etc. ...). D'autres optent pour une approche décentralisée afin d'améliorer la scalabilité du système face à la montée en échelle (grand nombre de noeuds impliqués).

L'étape de mise à jour consiste à choisir une méthode pour mettre à jour les valeurs de confiance. Deux méthodes sont utilisées. Dans la méthode dirigée par le temps, les mises à jour sont faites d'une manière périodique. Dans la méthode dirigée par les évènements, les mises à jour se font à chaque fois qu'un nouvel évènement se produit.

Les étape de propagation et de mise à jour n'affectent pas la pertinence du MGC en termes de résilience face aux attaques. Cependant, elles ont un impact direct sur la performance du système. Nous nous concentrons dans ce travail sur l'étape principale

qui est celle de l'établissement de la confiance. En effet, la performance du système de gestion de la confiance dépend essentiellement du modèle mis en place pour évaluer le degré de confiance qui peut être accordé aux différentes entités impliquées dans le système.

3. Scénario de motivation

Prenons l'exemple d'un scénario dans le domaine des réseaux véhiculaires. Dans ce dernier, les conducteurs collaboreront pour connaître l'état de la route ou ils circulent. Les informations sur l'état de la route (accident, embouteillage, impasse, travaux, inondations, secousses, route étroite, etc.) peuvent être détectées automatiquement par différents types d'objets intelligents (véhicules intelligents, téléphones intelligents, capteurs, ...), ou signalées manuellement par différents conducteurs. Dans ce type de scénario, les services fournis sont les informations sur l'état d'un itinéraire donné à un moment donné. Une requête dans ce scénario se réfèrent à la position actuelle du conducteur en termes de longitude et de latitude et à un Δ_t se réfère à l'intervalle de temps actuel. Un conducteur qui emprunte un itinéraire donné lancera donc sa requête. Le système fonctionnera de manière à lui donner une réponse fiable. Dans un tel scénario, des attaques peuvent être effectuées pour différentes raisons. Certains conducteurs peuvent s'amuser à signaler des incidents (attaque discriminatoire). D'autres conducteurs peuvent s'entraider pour signaler un incident juste pour libérer le trafic sur leur trajet en effectuant des atttaques de type BSA. Mais d'autres raisons plus graves peuvent être à l'origine de ces les attaques telles que un vol ou un enlèvement à l'aide des attaques de type BMA.

4. Travaux connexes

Les modèles d'évaluation de la confiance se composent de deux étapes, à savoir (i) **l'étape de composition** et (ii) **l'étape d'agrégation de la confiance**. Le tableau 2 présente les facteurs proposés dans la littérature pour l'étape de composition. Ces facteurs représentent des concepts abstraits visant à quantifier le niveau de confiance des noeuds et sont calculés par différentes mesures en fonction de l'objectif et du contexte de l'auteur. Par exemple, dans (Jayasinghe *et al*., 2016), le facteur *recommandation* est mesuré comme le nombre de noeuds directement connectés à un noeud donné u_i. Toutefois, dans (Truong *et al*., 2016), le facteur *recommandation* est mesuré comme la moyenne totale des votes donnés à un noeud u_i. Cette même mesure (moyenne des votes) est appelée *réputation* dans certains autres ouvrages. Le facteur *coopération* est considéré comme un indicateur pour mesurer la connaissance d'un noeud dans (Truong *et al*., 2016) et est calculé comme la fréquence des interactions sociales entre deux noeuds. Toutefois, dans (R. Chen *et al*., 2016), le facteur *coopération* est calculé comme le nombre d'amis communs entre deux noeuds.
Etant donné qu'il n'existe pas de consensus sur la définition du concept de confiance, et compte tenu de la divergence des facteurs proposés, ainsi que des mesures proposées pour chaque facteur, nous avons choisi dans ce travail de partir de la définition

de chaque type d'attaque. En effet, nous estimons qu'un modèle d'évaluation de la confiance doit avant tout remplir le rôle de garant de la fiabilité du système dans lequel il est impliqué. Cette fiabilité est compromise par les différents types d'attaques de confiance.
Nous pensons que certains facteurs et mesures proposés dans la littérature, tels que le nombre d'amis communs ou le nombre de relations dans le réseau, n'ont aucun rapport avec les attaques de confiance citées. Il est, effectivement, courant (comme dans les réseaux sociaux classiques) qu'un noeud malveillant augmente le nombre de ses relations avant de procéder à des attaques. D'autres mesures, telles que la moyenne des votes reçus, pourraient donner une idée de l'historique d'un noeud et pourraient donc permettre de détecter certains types d'attaques. Les facteurs proposés dans la littérature restent insuffisants pour détecter tous les types d'attaques. En effet, aucun facteur ne permet, par exemple, de détecter l'attaque SPA dans laquelle un noeud est caché sous une fausse identité.

Pour conclure, la performance d'un modèle d'évaluation de la confiance dépend principalement des facteurs et des mesures choisies dans la phase de composition. Néanmoins, elle dépend également de la méthode choisie dans la phase d'agrégation. Le tableau 2 montre que la moyenne pondérée est la méthode d'agrégation la plus utilisée. Cependant, les comportements réalisés pour chaque type d'attaque de confiance ne sont pas similaires. Une moyenne pondérée ne peut pas détecter tous les types d'attaques car les facteurs considérés et les poids attribués à chaque facteur peuvent différer d'un type d'attaque à l'autre. En effet, prenons le cas de l'attaque SPA, le facteur similarité qui permet de détecter que c'est le même utilisateur sous une fausse identité est primordial. Alors qu'il n'a aucune importance dans le cas des attaques BMA, BSA et DA.
Le deuxième critère de comparaison concerne la résilience aux attaques de confiance. Certains des travaux cités s'intéressent à la détection des attaques (Z. Chen *et al.*, 2016 ; R. Chen *et al.*, 2016). Cependant, ils ne permettent pas de détecter tous les types d'attaques. Le tableau 2 montre que la majorité des travaux connexes proposent des modèles permettant d'attribuer un degré de confiance à chaque noeud du réseau (Truong *et al.*, 2017 ; Huang *et al.*, 2016 ; Militano *et al.*, 2016). Ces modèles proposent de classer les meilleurs noeuds du réseau en fonction de leurs valeurs de confiance. Leurs objectif est de recommander les meilleurs noeuds du réseau. Cependant, ce type de modèle ne permet pas de détecter et d'isoler les noeuds malveillants. Ceci leur donne libre accès pour établir différents types d'attaques dans le réseau. Le but de notre travail est d'isoler les noeuds malveillants afin d'obtenir un système fiable. Les noeuds jugés comme malveillants ne sont pas bloqués, mais seront naturellement moins sollicités, en raison.

5. Etape de composition: Sélection des facteurs

Dans cette section, nous présentons l'étape de composition de notre modèle d'évaluation de la confiance. Nous proposons de nouveaux facteurs permettant de décrire et de quantifier les différents comportements opérant dans les systèmes SIoT. Nos fac-

Tableau 2. Comparaison des travaux connexes

	Composition	Agrégation	Objectif
(Truong *et al*., 2017)	Connaissance Réputation Experience	LF	C
(Huang *et al*., 2016)	Consistence Intention Capacité	MP	C
(Truong *et al*., 2016)	Recommandation Réputation Experience	LF	C
(R. Chen *et al*., 2016)	Honnêteté Coopération Intêrets-communs	LC	DA
(Militano *et al*., 2016)	Fiablité Réputation	MP	C
(Z. Chen *et al*., 2016)	Réputation Relation Sociale Niveau d'énergie	MP	DA

teurs sont dérivés de la description informelle de chaque type d'attaque de confiance et permettent de distinguer les comportements malveillants des comportements bénins.

5.1. Réputation

Ce facteur représente la réputation globale d'un utilisateur u_i dans le réseau et est désigné par $Rep(u_i)$. Il est calculé comme le quotient entre le nombre d'interactions positives de u_i et le nombre total d'interactions (eq.1). Les interactions positives sont des interactions ayant reçu des valeurs de vote élevée. Les noeuds ayant une valeur de réputation élevée sont plus susceptibles d'être attaqués par d'autres noeuds. Les noeuds ayant une valeur de réputation faible sont plus susceptibles de lancer des attaques de confiance. Le facteur réputation, combiné à d'autres facteurs, permet de révéler des attaques de type BMA, BSA, SPA et DA.

$$Rep(u_i) = \frac{1}{N^i} \sum_{k=0}^{N^i} r_k \tag{1}$$

(Avec N^i est le nombre de votes attribués à l'utilisateur u_i et $r_k \in [0, 5]$ la valeur du vote .)

5.2. Honnêteté

Ce facteur permet d'estimer si un utilisateur est honnête et est désigné par Hon(u_i). Un utilisateur est considéré honnête si ses votes reflètent son opinion réelle, ce qui signifie qu'il n'essaie pas de donner des votes erronés dans l'objectif de promouvoir ou ruiner la réputation des autres utilisateurs. En effet, dans les attaques BMA, BSA et SPA, le noeud malveillant présente un comportement malhonnête. Dans l'attaque BMA, le noeud malveillant donne des valeurs de votes basses à un noeud qui fournit des services de bonne qualité, afin de ruiner sa réputation. Dans l'attaque BSA, le noeud malveillant donne des votes élevés à un autre noeud malveillant qui fournit des services de mauvaise qualité, dans le but de l'aider à promouvoir sa réputation. Dans l'attaque SPA, le noeud malveillant tente de promouvoir sa propre réputation en s'accordant de bons votes alors que ses services sont de mauvaise qualité.
La figure 2 montre comment ce facteur est calculé. Nous générons tous d'abord le vecteur de votes moyen $\bar{r}$ qui représente la moyenne des votes de tous les utilisateurs du système. Nous le comparons ensuite au vecteur de votes r_i de l'utilisateur u_i à l'aide de la similarité Cosinus.

$$\begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \begin{array}{c} \begin{array}{cccccccccc} u_1 & u_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & u_9 & u_{10} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \\ 4 & 5 & 5 & 4 & 4 & 5 & 5 & 4 & 5 & 4 \end{bmatrix} \\ \textit{Matrice de votes} \\ \textit{utilisateurs/ services} \end{array} \xrightarrow{Moyenne} \begin{array}{c} \bar{r} \\ \begin{bmatrix} 0 \\ 0.5 \\ 1 \\ 5 \\ 4.5 \end{bmatrix} \end{array} \xleftrightarrow[Cosinus]{Similarité} \begin{array}{c} r_i \\ \begin{bmatrix} 5 \\ 5 \\ 5 \\ 0 \\ 1 \end{bmatrix} \end{array}$$

Figure 2. Mesure du facteur "Honnêteté"

Le facteur Honnêteté peut simplement indiquer qu'un utilisateur a des opinions différentes, mais, associé à d'autres facteurs, il peut révéler différents types d'attaques. Pour mesurer et quantifier ce facteur, nous comparons le vecteur de votes de l'utilisateur $Rvec(u_i)$ avec la matrice de votes du réseau en utilisant la similarité cosinus (eq.2).

$$Hon(u_i) = MAX_r - \frac{1}{S}\sum_{j=0}^{S}\sqrt{(r_{i,j} - \bar{r}_j)^2} \tag{2}$$

(Avec $r_{i,j}$ la valeur du vote attribuée par l'utilisateur u_i au service s_j, $\bar{r}_j$ est la moyenne des votes attribués par tous les utilisateurs du réseau au service s_j, S est le nombre total de services et MAX_r est une variable statique indiquant la valeur maximale de vote.)

5.3. Qualité du fournisseur

La facteur Qualité du fournisseur permet de juger de la qualité des services fournis par un utilisateur donné. Il est désigné par $QoP(u_i)$. En effet, le noeud malveillant vise

à propager des services de mauvaise qualité. Les services de bonne qualité acquièrent naturellement une bonne réputation dans le réseau. Le noeud malveillant doit recourir à un comportement malveillant pour propager des services de mauvaise qualité et, par conséquent, il lance des attaques de type BMA, BSA, SPA et DA pour atteindre cet objectif. Le facteur QoP est donc essentiel pour distinguer les noeuds susceptibles d'opérer des comportements malveillants, des autres noeuds qui fournissent des services de bonne qualité et qui n'ont pas besoin d'avoir recours à des attaques pour les propager.

$$QoP(u_i) = \sum_{s_k \in S(u_i))} \alpha * QoS(s_k) + (1-\alpha) * mr(s_k) \quad (3)$$

Avec S_{u_i} l'ensemble des services fournis par un utilisateur u_i, $QoS(s_k)$ est la valeur de QoS du service s_k, $mr(s_k)$ la moyenne des votes attribués à s_k et α est un poids.

5.4. Similarité

La similarité fait référence à la similitude entre l'utilisateur u_i et l'utilisateur u_j et est désignée par $SimU(u_i, u_j)$. Ce facteur est calculé sur la base de différentes caractéristiques telles que les profils, les intérêts, les services fournis, les dispositifs utilisés et la fréquence de proximité entre un couple d'utilisateurs. Elle vise à détecter les affinités entre les utilisateurs mais peut également révéler une attaque SPA dans laquelle le même utilisateur tente de promouvoir sa propre réputation sous une fausse identité.

5.5. Fréquence des votes

Ce facteur désigne la fréquence de votes attribués par un utilisateur u_i à un utilisateur u_j, et est noté par $RateF(u_i, u_j)$. Il est calculé comme le nombre de votes attribués par un utilisateur u_i à un utilisateur u_j divisé par le nombre total de votes donnés par l'utilisateur u_i. En effet, si un utilisateur u_i effectue une attaque contre un utilisateur u_j, nous trouverons probablement un grand nombre de votes attribués par l'utilisateur u_i à l'utilisateur u_j. Selon que ces votes soient positifs ou négatifs et en fonction de certains autres facteurs tels que la réputation et la qualité du fournisseur de l'utilisateur cible u_j, nous pouvons détecter une attaque de type BMA ou BSA.

5.6. Experience Directe

L'Expérience directe fait référence à l'opinion d'un noeud u_i sur ses interactions passées avec un noeud u_j, désignée par $ExpD(u_i, u_j)$. Elle est calculée comme le quotient des interactions réussies entre le noeud u_i et le noeud u_j, divisé par le nombre total d'interactions entre eux. Le facteur expérience directe ne peut donc pas révéler directement une attaque. Mais, combiné à d'autres facteurs, il permet de repérer le type de l'attaque. En effet, prenons l'exemple d'un noeud u_i qui attaque un noeud u_j. Ceci se traduit par des valeurs de réputation $Rep(u_i)$ et de Qualité de fournisseur $QoP(u_i)$

faibles pour u_i, ainsi que par une valeur de fréquence de votes $RateF(u_i, u_j)$ élevée qui reflète que u_i s'acharne à donner des votes au noeud u_j. Une valeur d'honnêteté de $Hon(u_i)$, vient confirmer l'hypothèse qu'il s'agit d'une attaque. Néanmoins, ces 4 facteurs combinés ensembles ne permettent pas de distinguer s'il s'agit d'une attaque BMA ou BSA. Le facteur Expérience directe permet de faire cette distinction. En effet, dans l'attaque BMA, le noeud u_i vise à ruiner la réputation de u_j et fournira donc des votes négatifs qui se traduiront par une valeur de $ExpD(u_i, u_j)$ faible, alors que, dans l'attaque BSA, le noeud u_i vise à promouvoir la réputation de u_j, ce qui donnera une valeur élevée de $ExpD(u_i, u_j)$.

5.7. Tendance des votes

Le facteur Tendance des votes est mesuré par le nombre de votes positifs divisé par le nombre total de votes fournis par un utilisateur. Elle vise à révéler si un utilisateur est plutôt optimiste ou pessimiste. Elle permet de détecter l'attaque discriminatoire (DA) dans laquelle l'utilisateur fournit des votes négatifs de manière aléatoire.

6. Etape d'agrégation: Conception de la fonction de classification

Une fois que nous avons choisi les différents facteurs qui permettent de décrire le comportement des utilisateurs du réseau, l'étape suivante consiste à choisir une méthode pour les agréger, afin d'obtenir la valeur de confiance finale. Dans la littérature, la méthode la plus courante est la moyenne pondérée. Cependant, nous estimons que la performance du système dépend, dans ce cas, principalement des poids attribués à chaque facteur. Ces derniers sont généralement fixés de manière empirique, or, l'importance des facteurs est subjective et dépend clairement des priorités de l'utilisateur. En outre, le comportement effectué pour chaque type d'attaque de confiance est différent. Une moyenne pondérée ne peut pas détecter tous les types d'attaques car les facteurs considérés et les poids attribués à chaque facteur différent d'un type d'attaque à un autre. En effet, si nous prenons l'exemple du facteur Similarité, ce dernier est essentiel pour détecter une attaque de type SPA, car il révélera qu'il s'agit du même utilisateur sous une fausse identité. Cependant, ce facteur n'a aucune importance dans le cas des attaques de type BMA, BSA ou DA.

La détection des noeuds malveillants étant considérée comme un problème complexe nécessitant une analyse approfondie du comportement des noeuds, nous proposons d'utiliser les techniques d'apprentissage automatique. Ainsi, nous considérons notre système comme un problème de classification. En effet, notre objectif est de détecter si un utilisateur est malveillant ou bénin. Un utilisateur est considéré comme malveillant s'il tente d'effectuer une attaque BMA, BSA, SPA ou DA. Si l'utilisateur n'a effectué aucune des attaques citées, il est considéré comme bénin. Ainsi, pour chaque couple d'utilisateurs (u_i, u_j), nous récupérons toutes les interactions passées. Nous calculons sur la base de ces interactions la valeur des différents facteurs liés à u_i, u_j et (u_i, u_j) (voir tableau 3). L'entrée de l'algorithme est l'ensemble de ces valeurs.

L'analyse de ces valeurs permettra de détecter si une attaque a eu lieu. En fonction de cela, l'utilisateur u_i sera jugé comme malveillant / bénin.

Tableau 3. Entrée de l'algorithme d'apprentissage automatique

$$\begin{array}{ccc} u_i & (u_i, u_j) & u_j \\ \begin{bmatrix} Rep(u_i) \\ Hon(u_i) \\ QoP(u_i) \\ RateT(u_i, u_j) \end{bmatrix} & \begin{bmatrix} SimU(u_i, u_j) \\ RateF(u_i, u_j) \\ ExpD(u_i, u_j) \end{bmatrix} & \begin{bmatrix} Rep(u_j) \\ Hon(u_j) \\ QoP(u_j) \\ RateT(u_i, u_j) \end{bmatrix} \end{array}$$

7. Expérimentations et évaluations

7.1. Description du jeu de données et méthodologie

En raison du manque de données réelles, la majorité des travaux proposent des expérimentations basées sur des simulations. Dans notre travail, nous avons évalué les performances de notre modèle en nous basant sur des simulations appliquées à un jeu de données réel intitulé Sigcomm [1]. Ce dernier contient des utilisateurs, leurs profils, leurs listes d'intérêts, leurs relations sociales, leurs interactions et leurs localisations. Nous avons généré pour chaque utilisateur un ou plusieurs dispositifs et nous avons réparti ses interactions sur l'ensemble des dispositifs qui lui sont attribués. Nous avons considéré que 50% des utilisateurs de notre réseau sont malicieux et nous avons simulé pour chaque utilisateur malicieux l'une ou plusieurs des 4 types d'attaques de confiance décrits précedemment. Le tableau 4 présente les statistiques du data-set.

Tableau 4. Statistiques des données de test.

Contenu		Nombre
Utilisateurs	Malicieux	38
	Bénin	38
Profils	Institut Ville Pays	76
Intérêts		711
Relations sociales		531
Interactions		32000
Dispositifs		300
Services		364
Proximité		285788

Pour prouver la performance des facteurs proposés, nous avons mesuré le gain d'information pour chaque facteur séparément. Nous avons, ensuite, testé les diffé-

1. http://crawdad.org/thlab/sigcomm2009/20120715/

rents algorithmes d'apprentissage mis en oeuvre dans l'outil WEKA (Hall *et al.*, 2009) pour construire notre modèle d'apprentissage. Ceci nous a permis de choisir la méthode d'apprentissage la plus adaptée à notre problématique. Enfin, afin de valider la méthode d'agrégation proposée (Apprentissage automatique supervisé), par rapport à la méthode d'agrégation la plus utilisée dans la littérature (Moyenne pondérée), nous avons comparé les résultats obtenus par (i) les autres travaux agrégés avec la moyenne pondérée, (ii) les facteurs que nous proposons, agrégés avec la moyenne pondérée et (iii) les facteurs que nous proposons, agrégés avec l'apprentissage automatique.

7.2. Sélection des facteurs et de l'algorithme d'apprentissage

Le gain d'information est une mesure d'évaluation utilisée pour sélectionner les attributs discriminiatifs et éliminer les attributs redondants, présentant des correlations ou inutiles. Elle se consiste à mesurer la variation de l'entropie en présence/absence d'un attribut (Azhagusundari, Thanamani, 2013 ; Lee, Lee, 2006 ; Yang, Pedersen, 1997). La figure 3 montre le gain d'information pour chaque facteur séparément. Le facteur *Similarité* (SimU) a la plus grande valeur de gain d'information. Cela s'explique par le fait qu'il est le seul facteur à permettre la détection les attaques de type SPA. Les facteurs *Fréquence de votes (RateF)*, *Qualité du fournisseur (QoP)*, *Tendance des votes (RateT)*, *Honnêteté (Hon)* et *Réputation (Rep)* présentent des valeurs de gain d'information presque égales. En effet, ils sont discriminatifs d'une manière égale pour la détection des attaques de type BMA, BSA et DA. Le facteur expérience directe a la plus faible valeur de gain d'information. En effet, ce facteur ne permet pas de détecter des attaques, mais, permet de faire la différence entre une attaque de type BMA et une attaque de type BSA. Nous avons ensuite

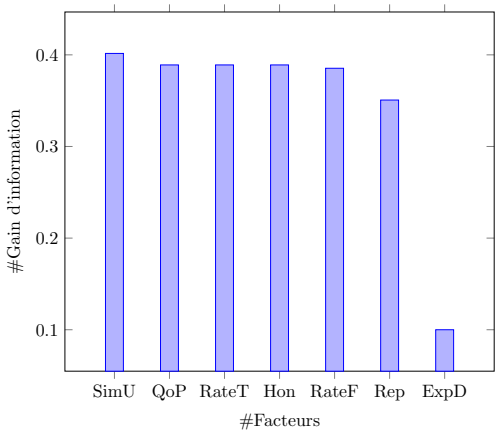

Figure 3. Gain d'information.

testé les différents algorithmes d'apprentissage mis en oeuvre dans l'outil WEKA ((Hall *et al.*, 2009)) pour construire notre modèle d'apprentissage. Nous rapportons, dans Figure 4, les résultats obtenus pour les algorithmes : Naive Bayes, Multi-Layer Perceptron et Random Tree. Nous avons finalement opté pour le Multi-Layer Perceptron, car il a donné les meilleurs résultats en termes de F-Mesure.

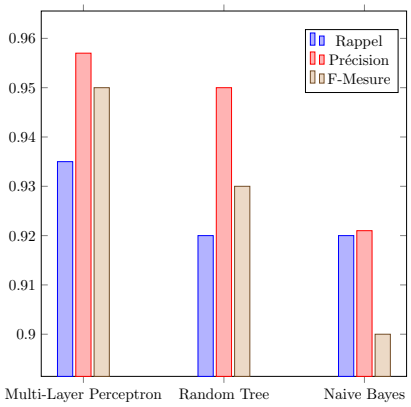

FIGURE *4. Comparaison des techniques d'apprentissage supervisé.*

7.3. Comparaison aux travaux connexes

Nous comparons les sept facteurs que nous proposons aux dix facteurs les plus utilisés dans la littérature. Nous avons implémenté ces dix facteurs pour les expérimenter sur notre jeu de données. Etant donné que les travaux de la littérature utilisent la moyenne pondérée pour agréger leurs facteurs, nous avons effectués différents essais pour fixer les poids et les seuils pour chaque travail. Nous avons ensuite utilisé la moyenne pondérée pour agréger les facteurs que nous proposons dans ce travail (F+MP). Ceci nous a permis de comparer et de valider la pertinence des facteurs proposés par rapport à ceux de l'état de l'art. Autrement dit, de valider notre proposition pour l'étape de composition indépendamment de la méthode d'agrégation que nous proposons ensuite. Enfin, nous avons appliqué notre méthode d'agrégation, notamment l'apprentissage automatique supervisé (F+AA), pour prouver sa pertinence par rapport à la méthode d'agrégation la plus utilisée dans la littérature (la moyenne pondérée). La figure 5 montre les résultats obtenus. Les facteurs proposés donnent de meilleurs résultats en termes de rappel, de précision et de F-mesure par rapport aux autres travaux, même dans le cas de l'agrégation avec la moyenne pondérée. Les résultats sont encore meilleurs lorsque nous appliquons la technique d'apprentissage automatique.

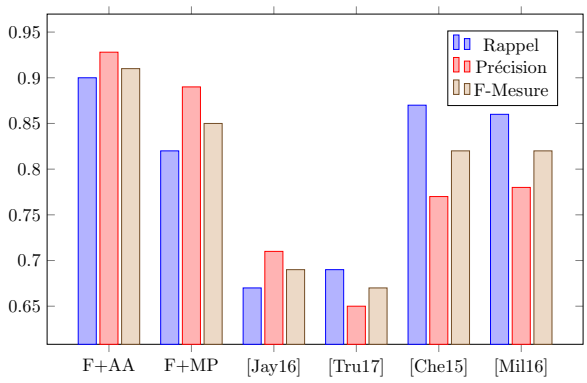

Figure 5. Comparaison avec les travaux connexes

8. Conclusion et perspectives

Nous avons proposé un modèle d'évaluation de confiance, capable de détecter les noeuds malveillants dans les environnements SIoT. Ce modèle repose sur de nouveaux facteurs, permettant de quantifier le comportement des utilisateurs, ainsi qu'une nouvelle méthode d'agrégation permettant d'analyser ces comportements. Des expérimentations basées sur des données réelles nous ont permis de prouver la pertinence du modèle proposé. Dans la suite de ce travail, nous nous intérressons aux étapes de propagation, de stockage et de mise à jour des valeurs de confiance qui sont primordiales pour la mise en oeuvre d'un mécanisme de gestion de la confiance adapté aux contraintes des environnements SIoT (scalabilité, dynamisme, minimisation de la consommation des ressources...).

Bibliographie

Abdelghani W., Zayani C. A., Amous I., Sèdes F. (2016). Trust management in social internet of things: A survey. In *Social media: The good, the bad, and the ugly*, p. 430–441. Swansea, Springer.

Ali D. H. (2015). *A social internet of things application architecture: applying semantic web technologies for achieving interoperability and automation between the cyber, physical and social worlds*. Thèse de doctorat non publiée, Institut National des Télécommunications.

Atzori L., Iera A., Morabito G. (2014). From" smart objects" to" social objects": The next evolutionary step of the internet of things. *IEEE Communications Magazine*, vol. 52, n° 1, p. 97–105.

Atzori L., Iera A., Morabito G., Nitti M. (2012). The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization. *Computer networks*, vol. 56, n° 16, p. 3594–3608.

Azhagusundari B., Thanamani A. S. (2013). Feature selection based on information gain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 2, nº 2, p. 18–21.

Bao F., Chen I., Guo J. (2013). Scalable, adaptive and survivable trust management for community of interest based internet of things systems. In *11th international symposium on autonomous decentralized systems*, p. 1–7. Mexico City, IEEE.

Chen R., Bao F., Guo J. (2016). Trust-based service management for social internet of things systems. *IEEE transactions on dependable and secure computing*, vol. 13, nº 6, p. 684–696.

Chen Z., Ling R., Huang C., Zhu X. (2016). A scheme of access service recommendation for the social internet of things. *Int. J. Communication Systems*, vol. 29, nº 4, p. 694–706.

Geetha S. (2016). Social internet of things. *World Scientific News*, vol. 41, p. 76.

Guo J., Chen R., Tsai J. J. (2017). A survey of trust computation models for service management in internet of things systems. *Computer Communications*, vol. 97, p. 1–14.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, vol. 11, nº 1, p. 10–18.

Huang J., Seck M. D., Gheorghe A. (2016). Towards trustworthy smart cyber-physical-social systems in the era of internet of things. In *System of systems engineering conference (sose), 2016 11th*, p. 1–6. Kongsberg, Norway, IEEE.

Jayasinghe U., Truong N. B., Lee G. M., Um T.-W. (2016). Rpr: A trust computation model for social internet of things. In *Ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress*, p. 930–937. Toulouse, France, IEEE.

Lee C., Lee G. G. (2006). Information gain and divergence-based feature selection for machine learning-based text categorization. *Information processing & management*, vol. 42, nº 1, p. 155–165.

Militano L., Orsino A., Araniti G., Nitti M., Atzori L., Iera A. (2016). Trusted d2d-based data uploading in in-band narrowband-iot with social awareness. In *Personal, indoor, and mobile radio communications (pimrc), 2016 ieee 27th annual international symposium on*, p. 1–6. Valencia, Spain, IEEE.

Truong N. B., Um T.-W., Lee G. M. (2016). A reputation and knowledge based trust service platform for trustworthy social internet of things. *Innovations in Clouds, Internet and Networks (ICIN), Paris, France*.

Truong N. B., Um T.-W., Zhou B., Lee G. M. (2017). From personal experience to global reputation for trust evaluation in the social internet of things. In *Globecom 2017-2017 ieee global communications conference*, p. 1–7. Singapore, IEEE.

Yang Y., Pedersen J. O. (1997). A comparative study on feature selection in text categorization. In *Icml*, vol. 97, p. 35.