Modèle et mesures de confiance pour la sécurité des systèmes d'information

Cyril Ray, Gouenou Coatrieux, Benjamin Coste

Modèle et mesures de confiance pour la sécurité des systèmes d'information

Cyril RAY [2], Gouenou COATRIEUX [3], Benjamin COSTÉ [1]

1. *Chaire de Cyber Défense des Systèmes Navals*
   Ecole navale - CC 600 29240 Brest Cedex 9, FRANCE
   benjamin.coste@ecole-navale.fr
2. *Ecole navale - CC 600 29240 Brest Cedex 9,* FRANCE
   cyril.ray@ecole-navale.fr
3. *IMT Atlantique*
   Technopole Brest-Iroise, CS 83818 29238 Brest Cedex 3, FRANCE
   gouenou.coatrieux@imt-atlantique.fr

RÉSUMÉ. *La multiplication des capteurs et des objets communicants de tous types a significativement enrichi le contenu des systèmes d'information (SI). Cependant, ces sources, souvent non maîtrisées, peuvent être leurrées ou corrompues par un tiers qui falsifie les informations produites. Cela soulève des questions relatives à la confiance accordée tant aux informations qu'aux sources et systèmes impactés. Cet article aborde la sécurité des SI sous l'angle de la confiance dans les sources d'information. La définition puis l'évaluation de la confiance dans un SI sont introduits avant de proposer une modélisation des sources d'information. La confiance dans ces dernières est abordée au travers de deux caractéristiques (la compétence et la sincérité) dont la mesure permet d'évaluer la confiance. Une expérimentation basée sur plusieurs sources simulées à partir d'un jeu de données réelles montre la pertinence de l'approche, transposable à d'autres SI. Cette étude est appliquée à l'analyse des données de navigation d'un navire.*

ABSTRACT. *The proliferation of sensors and communicating devices has significantly enhanced information systems (IS). However, these sources can be deceived or under a third party's control who can forge pieces of information produced by them. This raises new issues concerning trust one has in information, sources and IS affected. This article studies security of IS thanks to trust in sources. Definition then evaluation of trust in this context are introduced. Thus, a model of sources taking into account competence and sincerity is proposed. Those two features are studied and measured to allow trust's evaluation. Simulations based on a real dataset shows relevance of our approach. This study is applied to ship's navigation data analysis.*

MOTS-CLÉS : *confiance, sécurité des systèmes d'information.*

KEYWORDS: *trust, security of information systems.*

1. Introduction

Avec l'essor des technologies mobiles (e.g. smartphones, tablettes, objets connectés) embarquées ou distribuées (par ex. véhicules intelligents, automates industriels), les systèmes d'information s'adaptent et évoluent. S'ils doivent rendre différents services et assurer des tâches, ils gèrent et traitent par ailleurs de multiples informations qui les renseignent tout aussi bien sur leur état interne (par ex. alimentation, température, orientation) que sur leur environnement (informations géographiques, météorologiques, etc.). Sans être exhaustif, ces données sont issues de capteurs (par ex. GPS, gyroscope), d'équipements industriels (automates, actionneurs), de logiciels (IHM, microcodes, noyau) ou même d'opérateurs humains (e.g. administrateur). Sur cette base, chaque élément ou composant d'un système d'information peut être vu comme une source émettant des informations de différentes natures et formes qui contribuent au bon fonctionnement de l'ensemble.

Néanmoins, ces dernières sont susceptibles d'être altérées à chaque étape de la collecte et des traitements dont elles font l'objet, de manière accidentelle ou malveillante. Dans le dernier cas, des cyberattaques peuvent mettre en danger l'ensemble du système avec des conséquences plus ou moins graves en fonction du système visé. Dans ce travail, nous considérons plus particulièrement les systèmes d'information navals, qui contrôlent la totalité des navires et de leur survie, pour lesquels des travaux récents ont mis en évidence des vulnérabilités du système de navigation (Balduzzi *et al.*, 2014 ; Bhatti, Humphreys, 2015). Il a été montré qu'un tiers peut prendre le contrôle d'un navire et le détourner en falsifiant les informations de position qui lui sont transmises, avec des finalités que l'on peut imaginer (e.g. actes de terrorisme, banditisme).

Divers moyens peuvent être mis en place pour protéger un système d'information (SI), dans sa globalité ou au niveau de chacun de ses constituants (authentification, contrôle d'accès, confidentialité des échanges, etc.), ou même pour garantir l'intégrité de l'information elle-même (codes correcteurs d'erreurs, fonctions de hachage, codes d'authentification de message, tatouage de données, etc.). Ces outils ont pour objet d'assurer différents objectifs de sécurité (confidentialité, intégrité, disponibilité, traçabilité ...) et garantir un niveau maximal de sécurité. La confiance de l'utilisateur dans un SI s'appuie aujourd'hui sur la capacité de ces outils à contrer des menaces très variées. Cependant, elles restent limitées car elles ne permettent pas de savoir qu'un élément est leurré ou sous l'emprise d'un tiers malveillant (qui peut affirmer aujourd'hui que son ordinateur est exsangue de tout virus ou logiciel malveillant ?). Se pose alors la question de savoir quelle confiance accorder au SI, et plus particulièrement quelle confiance le SI lui-même peut accorder aux sources et aux informations qui l'alimentent.

Dans cet article, nous abordons la sécurité des systèmes d'information sur la base de la confiance qu'ils peuvent avoir de leur environnement et de leur état interne. La confiance est une notion complexe qui permet de raisonner en présence d'incertitudes (Abdul-Rahman, Hailes, 2000). Comme nous le verrons, la confiance a été abordée

dans différents domaines clés de la société avec pour objectif de mieux comprendre les relations entre différents acteurs (e.g. économie, sociologie, réseaux sociaux). À notre connaissance, la prise en compte d'une mesure de confiance à des fins de détection de cyberattaques n'a pas encore été étudiée. Pourtant, dans ce contexte, pouvoir mesurer la confiance au sein d'un SI nous semble adapté pour gérer l'absence de preuve formelle de compromission du SI et ainsi d'être capable de faire face à des attaques inconnues a priori. Dans ce travail, nous proposons une définition de la confiance et une mesure de celle-ci réalisée à partir de l'analyse des multiples informations reçues, collectées et manipulées par le système. Cette mesure est dépendante des caractéristiques intrinsèques d'une source mais également de l'évolution des informations qu'elle transmet, au regard des données transmises par les autres sources du système d'information.

Le reste de cet article est organisé de la manière suivante. La section 2 aborde la question du comment définir la confiance dans des sources qui composent et alimentent un SI sur la base de différentes définitions et mesures de la littérature. En section 3, nous proposons et présentons une modélisation possible des sources d'un SI et des informations qu'elles produisent. Ce modèle nous servira par la suite en section 4 pour établir un ensemble de mesures de la confiance dans et au sein d'un système d'information. Ces mesures sont par la suite expérimentées en section 5 dans le cadre d'un simulateur d'un système de navigation ; un SI naval clé à bord des navires. La dernière section conclu cet article.

2. Modèles de confiance

Divers modèles de confiance existent. Tandis que certains cherchent à définir cette notion complexe (Demolombe, 2004) d'autres tentent de la mesurer (Capra, Musolesi, 2006). Cette section présente une synthèse des contributions que nous avons pu trouver en lien avec ces thèmes.

2.1. Définir la confiance

De nombreux travaux ont cherché à définir voire à modéliser la confiance dans divers domaines (Blomqvist, 1997 ; McKnight, Chervany, 2000). En économie, les modèles proposés privilégient généralement la *coopération* entre des agents (e.g. entreprises, banques, consultants) qui « travaillent et agissent ensemble sur une tâche, et partagent les bénéfices qui en découlent » (Marsh, 1994).

En approfondissant le lien entre les notions de coopération et de confiance, l'article (Demolombe, 2001) a étendu ces modèles avec la *sincérité*. En effet, l'auteur fait remarquer que la confiance d'un individu A envers un individu B n'est possible que si B est sincère aux yeux de A. Plus clairement, B ne doit pas dissimuler d'informations ayant un intérêt pour A (Lorini, Demolombe, 2008). Cependant, B n'a peut-être aucun intérêt à se montrer sincère. La sincérité permet donc de prendre en compte la dépendance de la confiance vis-à-vis des objectifs ou des intérêts de chacun. Si A

et B ont des intérêts contraires, alors il est de bon sens que chacun se méfie de l'autre avec pour conséquence une coopération quasi impossible.

En sciences humaines, et avant les années 2000, la confiance avait pour but d'appréhender les rapports entre individus. Elle considérait les *émotions* d'un sujet (Lewis, Weigert, 1985) et son état psychologique au moment de prendre une décision (Deutsch, 1958). L'ajout d'une dimension émotionnelle à la confiance peut compenser un manque de *connaissances* lors de la prise de décision dans des situations nouvelles (Luhmann, 1979). Ces caractéristiques (émotions, connaissance) contribuent à qualifier la confiance.

Cependant, dans le contexte considéré à l'époque, un même individu entretenait un nombre faible de relations de confiance ; lesquelles s'établissaient sur le long terme. Avec l'émergence d'Internet, l'entourage d'une personne s'est considérablement élargi. Un individu n'est plus seulement en relation avec d'autres personnes de son entourage mais est connecté à une multitude d'entités ; entités qui peuvent être tout aussi bien des personnes que des services. Cette nouvelle masse de relations diminue considérablement le nombre d'interactions entre chaque entité et limite du même coup l'accroissement du niveau de connaissance d'un individu sur ses relations. Ainsi un entourage réduit mais bien connu a laissé place à de nombreuses relations méconnues (Grandison, Sloman, 2000). Dans le même temps, les nouveaux moyens de communication empêchent parfois d'identifier la nature du correspondant (un être humain, un automate ou une intelligence artificielle ?). Le peu d'interactions couplé à l'anonymat relatif d'Internet rend cette connaissance difficile à acquérir. Pour pallier les faiblesses des définitions précédentes, Grandison et Sloman (2000) ont défini la confiance comme « la *croyance* ferme en la *compétence* d'une entité à agir de manière fiable au travers d'un contexte spécifique ». Une définition dans laquelle la notion de compétence s'exprime comme « la capacité d'une entité à assurer les fonctions qui lui sont attribuées ». En fondant la confiance sur la compétence, la définition s'adapte au mélange hommes-machines des systèmes d'information.

Plus récemment encore, l'émergence et la prolifération rapide des capteurs de tous types alimentant les systèmes d'information (notamment mobiles) suscitent de nouvelles questions quant à l'expression de la confiance vis-à-vis des informations collectées et transmises. En effet, les sources sont liées aux informations qu'elles émettent (Jousselme *et al*., 2014), ainsi certaines de leurs caractéristiques (e.g. *fiabilité*, *confiance*, *compétence*) influencent les informations produites. Il est possible, par exemple, en mesurant certaines propriétés des informations (e.g. *qualité*, *confiance*, *crédibilité*), d'estimer celles de la source. L'inverse est également vrai : la confiance dans une information dépend de celle accordée à la source. Paglieri *et al*. (2014) analysent ce lien et établissent la confiance dans une source comme la qualité attendue d'une information. Sur la base de leur analyse, plus la confiance dans la source est élevée et plus la qualité de l'information l'est aussi. En mesurant la qualité de l'information, la confiance dans la source s'adapte en impactant deux caractéristiques de cette dernière : sa compétence et sa sincérité. Ces caractéristiques sont choisies d'après le modèle de Demolombe (2001). Elles ne sont pas les seules à caractériser le compor-

tement d'une source mais Liu et Williams (2002) ont montré qu'il est possible de ramener les autres critères (vigilance et coopération) à ces deux là.

Afin de participer au renforcement de la sécurité d'un système d'information, la confiance doit prendre en compte l'éventuelle malveillance de la source qui se traduit par une falsification volontaire de l'information. Cependant, une source peut également commettre des erreurs (e.g. en donnant accidentellement une fausse information). Ainsi, pour modéliser séparément les erreurs accidentelles d'une source, de ses falsifications intentionnelles, un modèle pertinent de confiance devrait s'appuyer a minima sur les notions de compétence et de sincérité (Costé *et al.*, 2016).

2.2. Mesures de confiance

Plusieurs travaux ont cherché à mesurer la confiance. La plupart de ces contributions sont basées sur un modèle de réseau dans lequel les divers nœuds interagissent. Les interactions sont alors sources de recommandations faites par les divers membres du réseau pour calculer leurs indices de confiance (Yu, Singh, 2002 ; Yan *et al.*, 2003 ; Teacy *et al.*, 2006 ; Das, Islam, 2012 ; Josang *et al.*, 2015). La recommandation est le processus par lequel un nœud i va communiquer sa confiance $C_{i,j}$ dans le nœud j. Très utilisée, cette mesure suppose cependant que les nœuds ont conscience les uns des autres. Si chaque nœud est isolé des autres et n'a pas conscience du réseau alors la recommandation est impossible. Bien que cette hypothèse soit vérifiée dans les réseaux sociaux ou sur le web, elle ne l'est cependant pas en général. Par exemple, la recommandation est difficile dans les réseaux dits centralisés où un serveur communique avec plusieurs clients qui ne se connaissent pas entre eux.

Lorsqu'il n'est pas possible d'obtenir les diverses appréciations des sources entre elles, il est encore possible de mesurer la confiance à partir de l'analyse des informations transmises par la source (Matt *et al.*, 2010). Beaucoup de ces mesures sont construites sur la base de la théorie de l'argumentation (Dung, 1993) qui modélise un ensemble de propositions appelées *arguments* et d'*attaques* entre ces arguments. Les arguments sont assimilés aux nœuds d'un réseau et les attaques à des arêtes unidirectionnelles. La théorie de l'argumentation cherche à établir quels arguments sont rationnellement acceptables. Plus clairement, et comme illustré en figure 1, l'argument d est acceptable puisqu'il n'est attaqué par aucun autre. Il en est de même pour l'argument f. En revanche, l'argument e est contesté à la fois par b et f.

Sur la base de cette théorie, divers modèles de confiance qui utilisent des sources d'informations ont été suggérés (Stranders *et al.*, 2008 ; Parsons *et al.*, 2011 ; Villata *et al.*, 2013 ; Paglieri *et al.*, 2014). Ces modèles reposent sur deux hypothèses : l'ensemble des arguments utilisables ainsi que leurs liens (i.e. les attaques) sont connus et sont en nombre fini. Pour juger de la recevabilité d'un argument, il est donc nécessaire de pouvoir comparer l'ensemble de ceux à disposition et donc de pouvoir clairement identifier les attaques. Cela n'est pas toujours possible, notamment en présence d'incertitude. En effet, les arguments peuvent ne pas formellement s'opposer. Par exemple, les deux assertions « il fait chaud » et « il fait froid » ne s'opposent pas

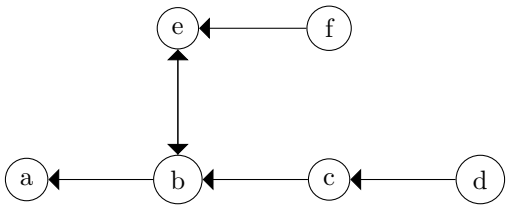

Figure 1. Exemple d'un modèle utilisant 6 arguments et 6 attaques entre ces arguments

nécessairement : elles peuvent indiquer une température modérée, intermédiaire. La théorie de l'argumentation n'est donc pas adaptée lorsqu'un conflit entre informations n'est pas clairement identifié et est donc incertain.

Néanmoins, plusieurs travaux ont cherché à pallier cette faiblesse. Parmi ceux-ci, Da Costa Pereira *et al*. (2011) proposent un modèle dans lequel l'acceptabilité des arguments (c.-à-d. le degré de croyance qu'ils sont vrais) est évaluée selon la confiance attribuée à la source. Contrairement aux modélisations de Dung (1993) et Villata *et al*. (2013) où un argument est soit accepté soit rejeté, l'acceptabilité d'un argument est ici continue. Cependant, la confiance est considérée comme un concept unidimensionnel alors qu'elle est multidimensionnelle pour Villata *et al*. (2013) qui la modélisent à partir de la compétence et de la sincérité de la source.

D'autres théories plus adaptées à la gestion de l'incertitude ont été utilisées (Capra, Musolesi, 2006 ; Sun *et al*., 2006 ; Wang, Singh, 2007). En particulier, Sun *et al*. (2006) arguent que la confiance est une mesure de l'incertitude et définissent ainsi leur mesure de confiance à partir de la probabilité qu'une entité effectue une certaine action. De même, Wang et Singh (2007) considèrent l'importance de la prise en compte de la certitude comme critère pour mesurer la confiance. Malgré une gestion efficiente des grandeurs réelles, ces travaux se basent exclusivement sur une confiance monodimensionnelle.

Nous souhaitons donc étendre ces modèles en proposant une mesure de la confiance qui soit multidimensionnelle, fondée sur la compétence et la sincérité. Cette mesure ne repose pas, voire peu, sur une connaissance *a priori* et ne nécessite pas d'interactions entre les sources. Elle doit, de plus, être adaptée à des informations continues telles que des valeurs réelles.

3. Modélisation des producteurs et des sources d'informations

Un système d'information est composé de multiples blocs fonctionnels interconnectés qui mesurent, analysent, traitent voire prennent des décisions et émettent de l'information. Ces blocs, quelle que soit leur fonction, peuvent être vus comme des *producteurs d'informations*. Ces producteurs (capteurs, automates, humains, etc.) peuvent aussi être perçus comme mono ou multisources.

Au contraire d'un producteur, une source est à l'origine d'une information d'une nature ou d'un type particulier. Il peut s'agir d'une entité physique comme un capteur. Cette section présente notre modélisation des sources et des producteurs d'information. Cette modélisation est nécessaire à l'établissement d'une mesure de confiance.

3.1. Modélisation des producteurs d'informations

Les producteurs d'informations sont de différentes natures. Ils produisent diverses informations (e.g. position, vitesse) sous différentes formes (nombre, texte, image, son, vidéo, etc.). Plusieurs types de producteurs peuvent être distingué : mono-source mono-information, multi-source mono-information (cas d'un système constitué de plusieurs capteurs de même type) ou multi-source multi-information.

La figure 2 illustre la modélisation, sous forme de diagramme UML, du producteur « GPS », et la figure 3, les sources qui le constituent.

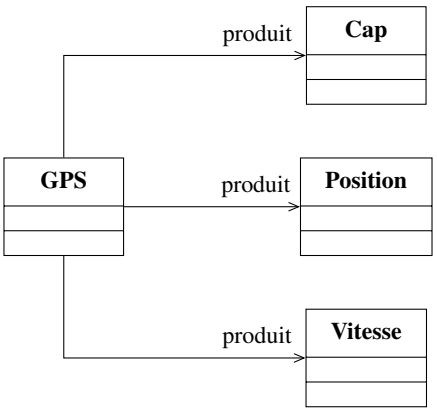

Figure 2. Modélisation du GPS sous la forme d'un producteur multi-informations

Cette modélisation permet de simplifier l'ajout ou la suppression d'un producteur au niveau du système (e.g. nouveau capteur installé, capteur en panne). La notion de producteur permet également de prendre en compte le fait que des sources et leurs informations sont liées à un même composant du système. Cette représentation est utile notamment en cas d'attaque par leurre du producteur de données. Dans ce cas, toutes les sources constituant le producteur sont leurrées également.

Un producteur est cependant plus complexe à manipuler du fait du nombre d'informations transmises à un instant t qui n'est pas forcément fixe. Les informations ne sont pas émises à la même période, certaines pouvant être envoyées occasionnellement (e.g.

les alertes SAR [1] ou CPA [2] du système AIS [3]). D'où l'intérêt de pouvoir modéliser un producteur comme constitué de sources d'informations. Ce dernier modèle allie simplicité (une source est spécialisée, c'est-à-dire qu'elle n'envoie qu'un seul type d'information et sert une unique fonctionnalité) et souplesse (il est facile d'ajouter ou d'enlever des sources d'un producteur, en cas de défaillance par exemple).

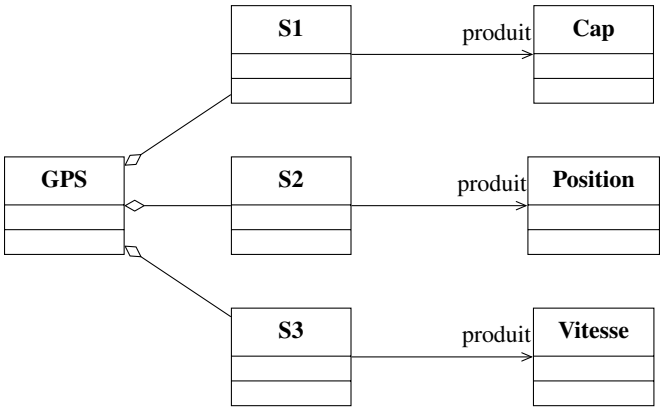

Figure 3. Modélisation multi-sources mono-information du GPS

Si l'on revient à l'exemple de la figure 3, le producteur GPS est constitué de trois sources distinctes émettant respectivement les informations de cap, position et vitesse. Si le GPS est éteint alors trois sources distinctes n'émettront plus d'information. Ce modèle prend en compte les liens qui existent entre les différentes sources, en particulier le fait qu'elles font partie d'un même producteur.

Enfin, pour aller plus loin, il est possible de rassembler plusieurs producteurs en un sous-système, comme illustré en figure 4, où un sous-système regroupe un ensemble de producteurs qui n'interagissent qu'entre eux.

Pour mesurer la confiance, avec cette modélisation, nous pourrons l'évaluer au niveau des sources, des producteurs et des sous-systèmes.

3.2. Modélisation des sources

Une source d'information est une entité qui observe un phénomène et le restitue au système sous différentes formes (valeur numérique, texte, image, vidéo, etc.). Les concepts développés sont d'ordre général. Toutefois, l'approche présentée est centrée sur les valeurs numériques, lesquelles sont des mesures fournies par des capteurs

1. *Search and Rescue*, alerte pour le sauvetage en mer.

2. *Closest Point of Approach*, alerte anti-collision avertissant d'un obstacle sur la route poursuivie par le navire.

3. *Automatic Identification System*, système standardisé par l'Organisation maritime internationale pour la diffusion en temps réel d'informations de navigation par VHF

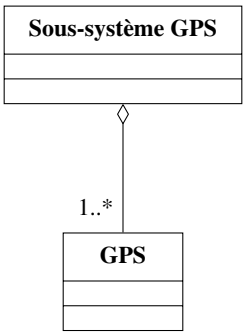

Figure 4. Modélisation d'un sous-système GPS regroupant l'ensemble des producteurs GPS du navire

embarqués. Dans le cas où la source est un capteur, elle mesure une grandeur physique (vitesse, température, etc.).

Cette mesure est imparfaite et entachée d'erreurs. En effet, la mesure est dépendante des caractéristiques du capteur (précision, sensibilité, usure, etc.). Deux sources mesurant le même phénomène et ayant les mêmes caractéristiques ne rendront pas forcément compte de la réalité de la même manière en raison d'un bruit dans la mesure. Cependant, ces mesures ne seront pas très éloignées et en tous cas seront proches de la réalité à moins de la défaillance du capteur ou d'une attaque. Suivant la complexité des capteurs, des phénomènes physiques observés et des composants électroniques utilisés, il est plus ou moins difficile de quantifier cette erreur dans la mesure. Néanmoins, une solution simple consiste à résumer l'ensemble des bruits de la chaîne d'acquisition à un bruit blanc additif ; c'est-à-dire également distribué sur toutes les fréquences du signal. Il s'agit d'un modèle de bruit très largement utilisé en traitement du signal (Papoulis, Pillai, 1986).

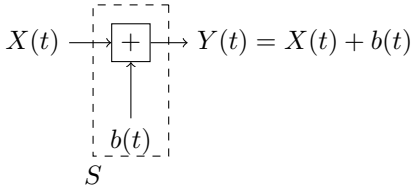

Figure 5. Modélisation d'une source par un canal gaussien

Comme illustré en figure 5, une source S observe le phénomène réel $X(t)$, une fonction dépendante du temps (température, vitesse, position), transmet sa mesure $Y(t) = X(t) + b(t)$ où b est une variable aléatoire de loi normale $\mathcal{N}(\mu(t), \sigma(t))$ de moyenne $\mu(t)$ et d'écart-type $\sigma(t)$, également dépendants du temps.

Les moments statistiques du bruit b (μ et σ) peuvent varier. En effet, les erreurs de mesures sont sujettes à l'usure des sources. Celles-ci ont une durée de vie limitée pendant laquelle elles se détériorent au fil de leurs sollicitations. Cette usure impacte

négativement la précision de la mesure. Une source est donc de moins en moins précise au fur et à mesure de son fonctionnement.

Toutefois, pour certaines applications (e.g. industrielles), les sources ont une durée de vie très longue. Elles sont alors très fiables, c'est-à-dire résistantes aux pannes ou à l'usure. C'est par exemple le cas dans les systèmes industriels composés de multiples capteurs, automates et actionneurs soumis à des conditions parfois extrêmes. Afin de prévenir toute atteinte à la sûreté de fonctionnement de l'installation, ceux-ci ont donc une durée de vie pouvant atteindre plusieurs dizaines d'années (ANSSI, 2015). Dans ces cas-là, nous pouvons donc supposer que les moments statistiques de b sont stables, c'est-à-dire μ et σ sont constantes.

Les erreurs peuvent provenir d'un bruit dans la mesure mais également d'un biais dû à un mauvais calibrage ou à l'environnement (par ex. des vibrations). Cette erreur, qualifiée d e s ystématique, s e m odélise p ar l 'ajout d 'un b iais d éterministe d ans la mesure (c.-à-d. $\mu \neq 0$). Ce type d'erreur est cependant identifiable l ors d e tests préalables du capteur. Sans perte de généralité, nous supposerons donc que b est centrée (*i.e.* $\mu = 0$).

Par la suite, une source sera dite *idéale* ou *parfaite* si celle-ci renvoie telle quelle l'information observée, c'est-à-dire $X(t) = Y(t)$ pour tout t (*i.e.* $\mu = 0$, $\sigma = 0$).

Dans cette section, nous avons modélisé des sources d'information. À partir de ce modèle, nous présentons dans la section suivante les mesures de compétence, sincérité et confiance dans une source d'information.

4. Mesurer la confiance

Comme décrit dans la section précédente, nous définissons la confiance accordée à une source comme une fonction de la compétence et de la sincérité de cette dernière. Nous précisons ci-après les concepts et comment les mesurer.

4.1. Mesure de compétence

Grandison et Sloman (2000) ont défini la compétence d'une source comme « sa capacité à remplir les fonctions qui lui sont attribuées ». Cette capacité est dépendante des caractéristiques intrinsèques de la source.

D'après cette définition et le modèle de source proposé, la compétence $Comp$ d'une source est donc dépendante de l'imprécision de sa mesure. Ainsi, une source *idéale* est jugée compétente car elle remplit sa fonction en fournissant exactement la mesure réelle. Nous avons

$$Comp = f(b) \overset{b \text{ est centrée}}{=} f(\sigma)$$

où f est une fonction de la mesure de la compétence à définir (telle que $Comp = f(\sigma) \in [0; 1]$). Considérant que si la source est parfaite alors la compétence est

maximale, i.e. $f(\sigma = 0) = 1$ et qu'a contrario, si la source est très imprécise (i.e. $\sigma \to +\infty$) alors elle est incompétente, i.e. sa compétence tend vers 0 ($\lim_{\sigma \to +\infty} f(\sigma) = 0$), nous proposons de définir la fonction f telle que

$$Comp = \frac{1}{1+\sigma}$$

Malgré sa simplicité, la section 5.2 montrera que cette fonction répond au besoin. Par exemple, dans le cas d'un GPS qui mesure une latitude avec une précision de l'ordre de 10^{-5}, la compétence de la source associée à cette mesure est de l'ordre de 0.99999.

4.2. Mesure de sincérité

La sincérité d'une source est par nature difficile à évaluer. Liu et Williams (2002) proposent de la mesurer à partir de la croyance que les sources ont dans l'information qu'elles envoient. Ils l'identifient à la différence entre ce que la source « dit » et ce que la source, un humain dans le contexte, « pense », ou « sait ». Dans le contexte de cet article, sur la base du modèle de source évoqué en Section 3.2, ce concept de « pensée » d'une source n'est pas valide. Nous proposons donc de comparer les informations des différentes sources entre elles à l'instar de Paglieri *et al*. (2014). Plus clairement, la sincérité d'une source est évaluée à partir des informations émises par les autres sources.

Il est également important de souligner qu'il existe un phénomène de dépendance entre la compétence et la sincérité. En effet, lorsqu'une source est incompétente, elle émet une information très imprécise qui complexifie sa comparaison avec des informations fournies par des sources compétentes. Plus une information est imprécise et plus celle-ci sera éloignée de la réalité et, par voie de conséquence, des autres informations plus précises et de même nature. Dès lors, dans le cas où la compétence de la source est faible, sa sincérité doit l'être également. Par contraposition, lorsque la compétence de la source est élevée (*i.e.* proche de 1), aucune conclusion ne peut être induite sur sa sincérité. Nous proposons alors de borner la mesure de sincérité d'une source par sa compétence :

$$\forall i \geq 1 \quad Sinc_i(t) = min(p_i(t), Comp_i(t))$$

où $p_i \in [0;1]$ représente le degré d'accord de la source i avec les autres à l'instant t. Le degré d'accord d'une source avec les autres se mesure en comparant les informations que celle-ci fournit avec celles émises par les autres sources. Il sera élevé si l'information émise par la source est en accord avec celle des autres.

Ainsi, considérant un ensemble de sources compétentes, une source émettant une information similaire à la majorité sera jugée plus sincère qu'une source contestée (*i.e.* en accord avec une minorité). Tel que défini, le degré d'accord est une mesure de consensus, c'est-à-dire à quel point la source est supportée par les autres sources. Elle peut être vu comme le ratio entre le nombre de sources en accord avec la source i à l'instant t, et le nombre total de sources. Pour mesurer l'accord entre deux sources,

une solution possible est de passer par un consensus binaire, comme proposé dans (Paglieri *et al.*, 2014) : deux sources sont complètement d'accord ou en complet désaccord. Dans le contexte de cet article, et avec le modèle de source vu en section 3.2, cette approche n'est pas la plus judicieuse, car l'information correspond à des nombres réels. Nous proposons plutôt d'utiliser une fonction de similarité, notée Sim, continue pour mesurer le consensus prenant en compte les informations émises aux instants précédents, c'est-à-dire :

$$p_i(t) = \begin{cases} 1 & n = 1 \\ \dfrac{1}{n-1} \sum\limits_{\substack{j=1 \\ j \neq i}}^{n} Sim(\{Y_i(t)\}_{t>0}, \{Y_j(t)\}_{t>0}) & n \geq 2 \end{cases}$$

où n est le nombre de sources et $\{Y_i(t)\}_{t>0}$ l'ensemble des informations émises par la source i jusqu'à l'instant t. De manière à garantir $p_i = 1$ lorsque toutes les sources sont en accord et inversement si $p_i = 0$ lorsque la source i s'oppose à toutes les autres, la fonction de similarité utilisée ci-après correspond à une mesure de corrélation entre les informations des différentes sources. Un autre intérêt de cette mesure est que la valeur de p_i est relativement stable lorsque le nombre n de sources est « suffisamment » grand.

Au contraire, dans le cas particulier d'une unique source, le consensus ne peut être mesuré à cause du manque d'informations supplémentaires. Par convention, nous proposons alors de poser $p_1(t) = 1$ ce qui symbolise l'accord de la source avec elle-même. Il en résulte alors une égalité directe entre la sincérité d'une source unique et sa compétence (*i.e.* $Sinc_i(t) = Comp_i(t)$ pour tout t).

4.3. De la compétence et de la sincérité à la confiance

Pour obtenir une mesure de confiance $Conf(S_i)$ à partir des mesures de compétence et de sincérité (*i.e.* $Conf(S_i) = Conf(Comp(S_i), Sinc(S_i))$), plusieurs solutions ont été définies dans (Liu, Williams, 2002). Ces mesures respectent toutes les contraintes suivantes :

- $Conf(1,1) = 1$
- $Conf(0,0) = 0$
- $Conf(Comp, 1) = Comp,\ Comp \in [0;1]$
- $Conf(1, Sinc) = Sinc,\ Sinc \in [0;1]$

Les auteurs proposent ainsi plusieurs mesures en adéquation avec ces contraintes :

$$Conf_1(Comp, Sinc) = Comp * Sinc \quad (1)$$

$$Conf_2(Comp, Sinc) = min(Comp, Sinc) \quad (2)$$

$$Conf_3(Comp, Sinc) = 1 - (1 - Comp)(1 - Sinc) \quad (3)$$

La mesure $Conf_3$ ne traduit pas nécessairement l'absence de confiance en une source incompétente ou non sincère. En particulier, $Conf_3$ est non nulle lorsque la compétence ou la sincérité de la source est nulle, propriété cependant souhaitée dans notre contexte. Cela revient à ajouter au jeu de contraintes précédent, les règles supplémentaires suivantes :

- $Conf(0, Sinc) = 0,\ Sinc \in [0; 1]$
- $Conf(Comp, 0) = 0,\ Comp \in [0; 1]$

Le tableau 1 résume les différentes propriétés des mesures évoquées dans cette section. Ce tableau expose également les monotonies des différentes mesures selon les variables dont elles dépendent. Par exemple, la compétence est décroissante selon l'imprécision de la source : plus une source est imprécise et moins elle est compétente. *A contrario*, la sincérité est croissante selon la compétence et le consensus. De deux sources de compétence identique, la plus sincère est celle en accord avec la majorité. Inversement, si deux sources s'accordent autant l'une que l'autre avec les sources restantes alors la plus compétente est jugée plus sincère. Les mêmes propriétés s'appliquent à la mesure de confiance : entre deux sources de compétence (resp. de sincérité) identiques, celle qui est la plus de confiance est celle qui est la plus sincère (resp. la plus compétente).

Dans la section suivante, nous expérimentons ces différentes mesures sur des données réelles.

5. Expérimentations

5.1. Scénario

Les mesures de confiance, de compétence et de sincérité définies précédemment ont été testées sur des données provenant de l'*Automatic Identification System* d'un bateau de type cargo, à proximité de Brest. L'AIS fournit différentes informations de navigation : position, vitesse, identifiant du bateau, etc. Ces informations, qui servent à empêcher ou limiter les collisions entre navires, sont vulnérables à la falsification (Ray *et al.*, 2015). Nous les avons utilisées ici comme base de simulation.

À partir de ces données, nous avons simulé 3 producteurs d'information (cf. section 3.2) : deux GPS et un Loch Doppler. Ces producteurs d'informations peuvent se retrouver embarqués sur des navires tels que des navires de croisière par exemple. Dans notre contexte expérimental, les deux GPS sont situés respectivement à l'avant et à l'arrière du navire et le Loch Doppler en son milieu. Si un GPS est constitué de trois sources donnant trois types distincts d'informations : la position, la vitesse et le cap, un Loch Doppler ne comporte qu'une seule source : la vitesse. En effet, un Loch Doppler mesure la vitesse du navire par rapport au fond en utilisant un signal ultrasonore.

Pour les besoins de l'expérience, nous avons choisi de fixer le nombre de sources à trois. En effet, pour être mesurée, la sincérité nécessite de la redondance, c'est-à-dire

Tableau 1. Propriétés des mesures servant à l'évaluation de la confiance

Mesures	Propriétés
Compétence	• Fonction de l'imprécision de la mesure : $Comp_i(t) = f_c(\sigma_i(t))$ • $f_c(0) = 1$ • $\lim_{\sigma_i(t) \to +\infty} f_c(\sigma_i(t)) = 0$ • Monotonie : $\sigma_1(t) < \sigma_2(t) \Rightarrow Comp_1(t) \geq Comp_2(t)$
Sincérité	• Compétence **faible** $\Rightarrow$ On considère une sincérité **faible** • Compétence **forte** $\not\Rightarrow$ sincérité **forte** • $Sinc_i(t) = f_s(p_i(t), Comp_i(t))$, $p_i(t)$ une mesure du consensus entre les sources à l'instant t • Monotonie par rapport au consensus : $p_1(t) < p_2(t)$ et $Comp_1(t) = Comp_2(t)$ $\Rightarrow Sinc_1(t) \leq Sinc_2(t)$ • Monotonie par rapport à la compétence : $p_1(t) = p_2(t)$ et $Comp_1(t) < Comp_2(t)$ $\Rightarrow Sinc_1(t) \leq Sinc_2(t)$
Confiance	• $Conf = f(Comp, Sinc)$ • $f(0,0) = 0$ • $f(1,1) = 1$ • $f(Comp, 1) = Comp,\ Comp \in [0;1]$ • $f(1, Sinc) = Sinc,\ Sinc \in [0;1]$ • $f(0, Sinc) = 0$ • $f(Comp, 0) = 0$ • Monotonie par rapport à la compétence : $Comp_1(t) < Comp_2(t)$ et $Sinc_1(t) = Sinc_2(t)$ $\Rightarrow Conf_1(t) \leq Sinc_2(t)$ • Monotonie par rapport à la sincérité : $Comp_1(t) = Comp_2(t)$ et $Sinc_1(t) < Sinc_2(t)$ $\Rightarrow Conf_1(t) \leq Conf_2(t)$

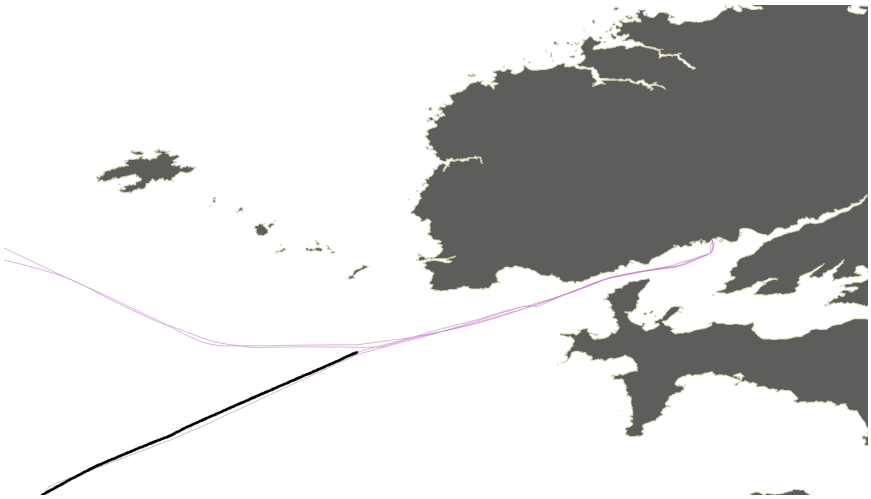

Figure 6. Trajectoire et positions du navire étudié

de disposer de plusieurs sources émettant une même information. De plus, les navires ne disposent pas souvent, en pratique, de beaucoup de capteurs émettant une même information. Cependant, sur des bateaux d'une certaine taille (tels que des navires de croisière), cela est plus fréquent. Simuler 3 sources est donc un bon compromis : un nombre de capteurs qui est réaliste et qui permet de tester les mesures élaborés dans cet article en exploitant la redondance des informations. Pour simuler les trois sources, un bruit gaussien centré a été ajouté suivant le modèle de source décrit en section 3.2. Les trois bruits gaussiens sont de variance identique, tirée des spécifications des constructeurs. Le bruit d'une source $i \in \{1, 2, 3\}$ est de variance $\sigma_i = 0,1$ ce qui donne une mesure de compétence $Comp_i(t) = \frac{1}{1+\sigma_i} \approx 0,91$.

Dans notre scénario, nous souhaitons simuler une attaque sur l'un des capteurs. En effet, un attaquant peut vouloir envoyer de fausses informations de vitesse pour ralentir le navire (e.g. pour faciliter son interception par des pirates) ou bien le faire accélérer (e.g. surconsommation, usure prématurée du moteur ou de la ligne d'arbre). Cette attaque, bien que peu subtile (un regard à l'historique suffit à la détecter), peut se révéler gênante voire dangereuse sur le long terme. Nous avons également testé une version plus graduelle de cette attaque : l'augmentation de vitesse n'est plus soudaine mais progressive, via des falsifications successives.

Pour ce faire, nous avons extrait une suite de 1 000 données de vitesse parmi les 9 693 fournies par l'AIS. Parmi les informations dont nous disposons figurent également un identifiant du navire, sa longitude, sa latitude, son cap et son statut (e.g. « En mer », « Au mouillage »). Nous avons implémenté (en Python) le bruitage des informations pour simuler les sources ainsi que les diverses mesures décrites dans les sections précédentes. Avant d'être bruitées, les informations de vitesse sont modifiées. Ceci simule la falsification, par un attaquant, des informations destinées

au Loch Doppler. Le fait que les informations soient altérées avant d'être bruitées est caractéristique d'une attaque par « leurrage ».
La section suivante présente différents résultats obtenus dans le cadre du scénario évoqué.

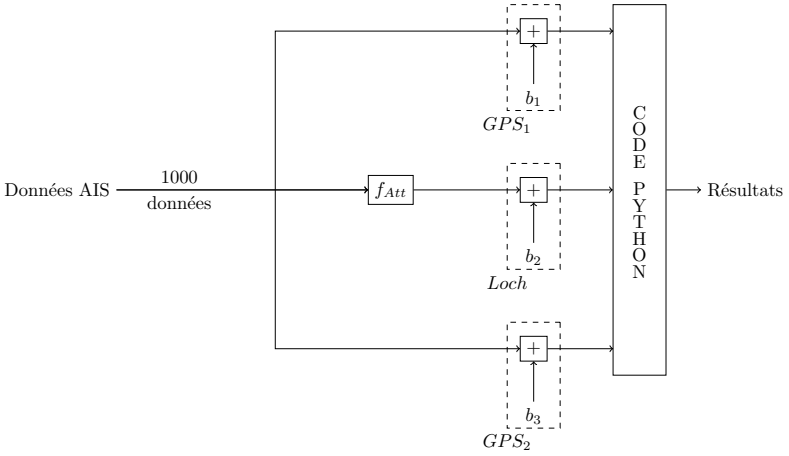

Figure 7. Protocole de génération des résultats

5.2. Résultats

La figure 8 montre le comportement des différentes mesures de compétence, de sincérité et de confiance en simulant les trois sources de vitesse des producteurs à partir des données de vitesse produites par l'AIS. Pour simuler une attaque de « leurrage », le Loch Doppler émet de fausses informations à partir de l'instant $t = 500$; instant à partir duquel la vitesse transmise est supérieure de 1 nœud à la vitesse réelle.

La première ligne de courbes en figure 8 montre la vitesse telle que perçue par chaque source avec une précision d'environ $0,1$ nœud (spécifications constructeur). En particulier, la deuxième courbe met en évidence la falsification des informations transmises par le Doppler. La dernière ligne montre l'évolution de la compétence, de la sincérité et de la confiance de chacune des sources au fil du temps. Comme le montre la figure 8, la compétence de chaque source est identique, du fait qu'un bruit de même variance a été ajouté aux données réelles. Nous constatons l'impact de l'attaque du Loch Doppler au niveau de la mesure de sincérité des trois sources. Dès que ce dernier indique une vitesse différente de celle mesurée par les GPS, la mesure de sa sincérité tend à la baisse et de manière plus forte que les mesures de sincérité des deux GPS.

Comme Bhatti et Humphreys (2015) l'ont démontré, il est possible d'avoir une falsification incrémentale de l'information afin de masquer l'attaque. Sur la figure 9, nous avons rejoué le scénario en modifiant le comportement de l'attaquant. Le

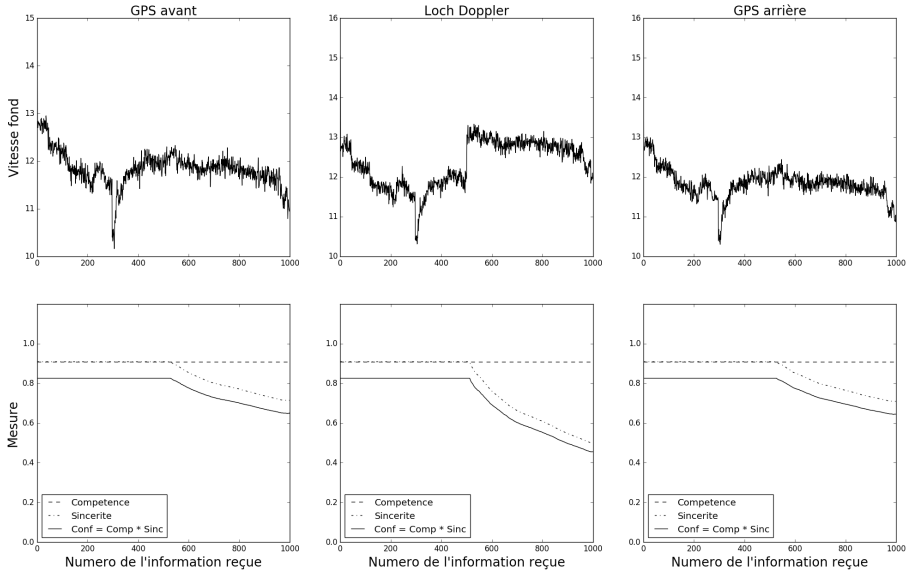

Figure 8. Évolution des mesures de compétence, sincérité et confiance en présence de trois sources de vitesse sur un navire

but étant de tester le comportement des mesures de compétence, de sincérité et de confiance, selon différents niveaux de complexité d'attaques. Une nouvelle attaque de « leurrage » est simulée, le Loch Doppler émet de fausses informations à partir de l'instant $t = 500$. À partir de cet instant, la vitesse transmise par le Loch augmente progressivement jusqu'à être supérieure de 1 nœud à la vitesse réelle.

La sincérité (et par conséquent la confiance) du Loch est la plus faible. L'attaque n'est cependant pas détectée dès le moment où elle survient. En effet, les informations sont falsifiées à partir de l'instant $t = 500$ mais la sincérité commence à varier à partir de l'instant $t = 743$. En falsifiant les informations progressivement, sans changement important, il est donc possible de modifier et de contrôler la vitesse du navire.

D'autres types d'attaques utilisant la manipulation d'informations peuvent affecter un système d'information. Par exemple, les *attaques par rejeu* consistent pour un attaquant à répéter une information déjà émise par une source. Ce type d'attaque a généralement pour but de corrompre le SI, sans nécessiter de lever la confidentialité des échanges. Les informations peuvent en effet être chiffrées et donc incompréhensibles pour un attaquant. Cependant, celui-ci peut les enregistrer (sous forme chiffrée) et les réémettre ultérieurement dans le but de perturber le bon fonctionnement du système.

Dans le cas d'un navire, une telle attaque peut répéter une séquence ayant pour but de stopper le navire ou bien le faire accélérer lors d'un accostage. Dans ce dernier cas, le navire nécessite de manœuvrer à faible vitesse. Rejouer une séquence d'informa-

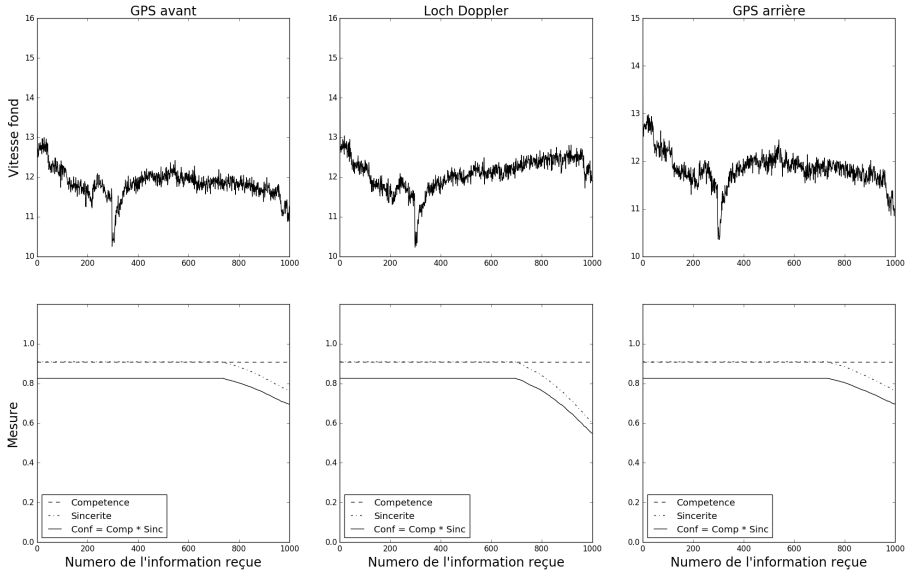

Figure 9. Évolution des mesures de compétence, sincérité et confiance en présence d'une falsification incrémentale des informations d'une source parmi trois

tions, enregistrées lors d'une phase d'accélération (e.g. au moment de quitter un port), conduit donc à un comportement qui est dangereux dans une telle situation.

Sur la figure 10, le Loch Doppler ré-émet une séquence d'informations (entre $t = 250$ et $t = 350$) à partir de l'instant $t = 500$. Cette séquence comprend un ralentissement brutal du navire. Comme pour la première attaque présentée, la sincérité ainsi que la confiance de la source attaquée diminuent au moment de la malversation.

Les attaques présentées ci-dessus ont été expérimentées dans des scénarios à 3 sources d'information. Comme expliqué en section 5.1, ceci est un bon compromis entre la réalité et nos objectifs expérimentaux. Toutefois, afin de confirmer que les mesures élaborées sont toujours adaptées en présence d'un nombre de sources plus important, nous avons également expérimenté plusieurs attaques en augmentant le nombre de sources.

Les résultats présentés en figure 11 illustrent l'attaque incrémentale avec 5 puis 10 sources. Les résultats montrent que le nombre de sources influence la confiance dans celles qui ne sont pas attaquées. En effet, une seule source étant ciblée, le nombre de sources « saines » augmente. Ces sources sont toutes en accord ce qui augmente de ce fait leurs sincérités respectives. En présence d'un nombre important de sources « saines », une source attaquée est donc la seule à subir un impact sur sa confiance.

La figure 12 a été obtenue après plusieurs rejeux du scénario (cf section 5.1) en faisant varier, à chaque itération, la compétence d'une des sources (le GPS avant). Pour chaque scénario joué, la sincérité de la source est calculée.

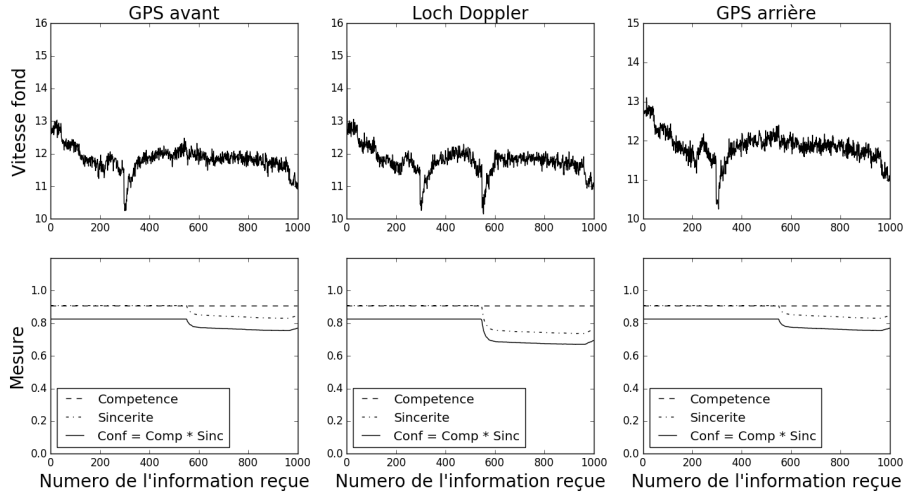

Figure 10. Évolution des mesures de compétence, sincérité et confiance en présence d'un rejeu des informations d'une source parmi trois

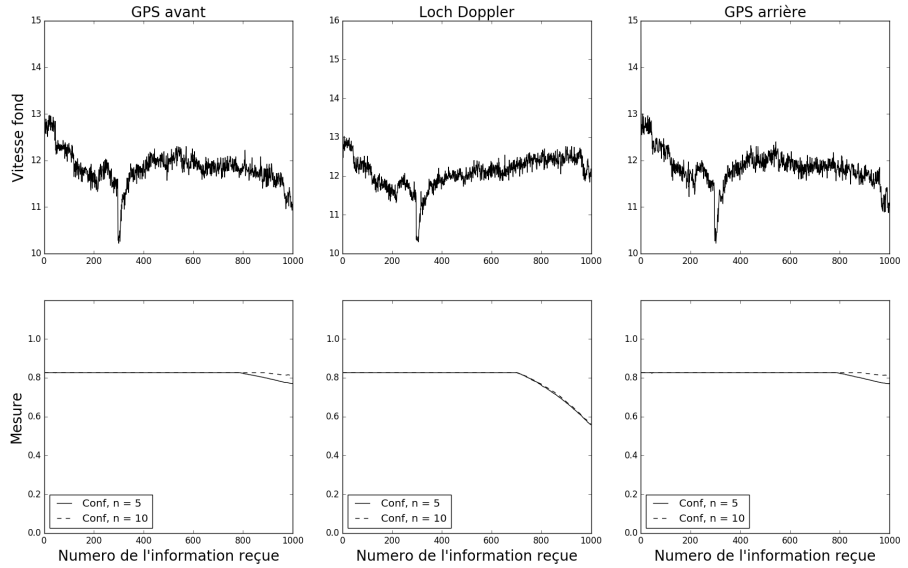

Figure 11. Évolution de la confiance en présence d'une falsification incrémentale des informations d'une source parmi cinq et dix

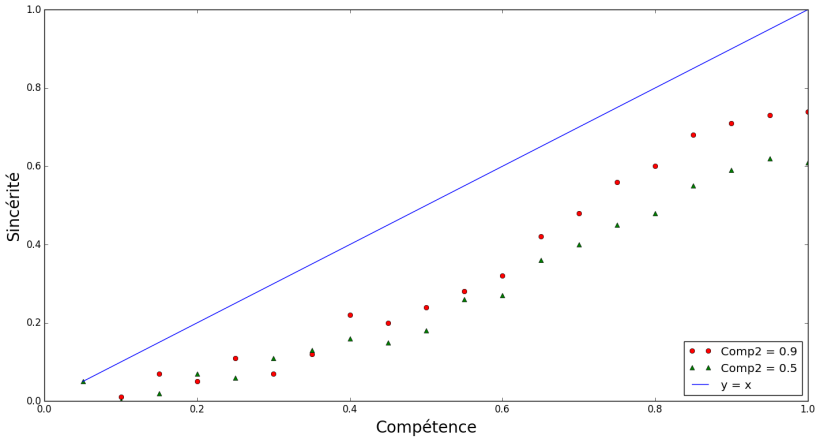

Figure 12. Évolution de la sincérité en fonction de la compétence d'une source parmi trois

La courbe en bleu est la droite d'équation $y = x$. Les points rouges représentent la sincérité de la source à la fin de la simulation, lorsque les compétences des deux autres sources sont égales à $0,9$. De même, les triangles verts représentent la sincérité de la source lorsque les compétences des deux autres sont égales respectivement à $0,5$ et $0,9$. Pour une compétence supérieure à $0,35$, les points rouges sont au-dessus des triangles verts. Ceci montre l'influence de la compétence des sources sur la sincérité des autres. Moins les sources sont compétentes et moins celles-ci sont sincères.

Conformément à ce qui a été énoncé en section 4.2, nous pouvons donc observer que la mesure de la sincérité est dépendante de la compétence.

L'ensemble des expérimentations présentées dans cette section montrent que des variabilités de la confiance peuvent servir à des fins de détection de cyberattaques. Malgré quelques limites en cas de falsification incrémentale, la confiance dans la source attaquée est la plus faible pour chacun des scénarios. De plus, en présence d'un nombre important de sources, celle qui est attaquée est la seule à subir une baisse de son niveau de confiance. Il est donc possible de détecter la source à l'origine de la malversation (*i.e.* celle qui diffuse les informations falsifiées).

6. Conclusion

Les systèmes d'information (SI) modernes s'appuient sur les nouvelles technologies mobiles, embarquées ou distribuées et leurs utilisateurs. Ils sont informés en continu de leur environnement et de son évolution par de multiples sources d'informations. Vulnérables, celles-ci peuvent être leurrées ou malveillantes, mettant en danger la sécurité des utilisateurs ou de l'environnement du système d'information.

Cet article aborde la notion de confiance dans un système d'information et plus particulièrement envers les sources qui le constituent et les informations qu'il manipule. Sur la base d'un modèle de sources et de producteurs d'informations le constituant, une mesure de confiance a été élaborée. Celle-ci repose sur un modèle fondé sur la compétence et la sincérité des sources. Des test ont été menés pour comparer les mesures dans une situation de falsification d'informations. Les résultats obtenus valident la pertinence de l'utilisation de la confiance dans un processus de détection de cyberattaque ; cet aspect n'ayant pas encore été abordé à notre connaissance.

Les perspectives de ce travail sont doubles. Tout d'abord, les sources sont jugées indépendamment les unes des autres. La prise en compte des relations qui les unissent, notamment de dépendance (e.g. proximité géographique, même type, causalité), pourrait permettre de limiter la portée des collusions éventuelles. Plus largement, les phénomènes de propagation de confiance (Esfandiari, Chandrasekharan, 2001 ; Guha *et al.*, 2004 ; De Cock, Da Silva, 2006 ; Josang *et al.*, 2006 ; Wang, Singh, 2006) et de rétroaction (*i.e.* comment la confiance de la source, basée sur l'analyse des informations passées, ainsi que les nouvelles informations émises, influencent la confiance de la source à l'instant présent) (Villata *et al.*, 2013 ; Paglieri *et al.*, 2014) restent à explorer.

Remerciements

Cette recherche est co-financée par la « Chaire de Cyber Défense des Systèmes Navals » et la région Bretagne. Les auteurs les remercient pour leur soutien.

Bibliographie

Abdul-Rahman A., Hailes S. (2000). Supporting trust in virtual communities. In *Proceedings of the 33rd annual hawaii international conference on system sciences*, p. 9–19.

ANSSI. (2015). *Cybersecurity for industrial control systems*. Rapport technique. Agence Nationale pour la Sécurité des Systèmes d'information.

Balduzzi M., Pasta A., Wilhoit K. (2014, dec). A security evaluation of automated identification system. In *Proceedings of the 30th annual computer security applications conference*, p. 436–445.

Bhatti J., Humphreys T. (2015). Hostile control of ships via false gps signals: Demonstration and detection. *submitted to Navigation, in review*.

Blomqvist K. (1997). The many faces of trust. *Scandinavian journal of management*, vol. 13, n° 3, p. 271–286.

Capra L., Musolesi M. (2006). Autonomic trust prediction for pervasive systems. In *20th international conference on advanced information networking and applications*, vol. 2, p. 48–59.

Costé B., Ray C., Coatrieux G. (2016). Évaluation de la confiance dans un environnement multisources. In *Informatique des organisations et systèmes d'information et de décision (inforsid), atelier sécurité des systèmes d'information : technologies et personnes.*

Da Costa Pereira C., TeTettamanzi A. B., Villata S. (2011). Changing one's mind: Erase or rewind? possibilistic belief revision with fuzzy argumentation based on trust. In *Proceedings of the twenty-second international joint conference on artificial intelligence*, vol. 1, p. 164–171.

Das A., Islam M. M. (2012). Securedtrust: a dynamic trust computation model for secured communication in multiagent systems. *IEEE Transactions on Dependable and Secure Computing*, vol. 9, n° 2, p. 261–274.

De Cock M., Da Silva P. P. (2006). A many valued representation and propagation of trust and distrust. In *Fuzzy logic and applications*, p. 114–120. Springer.

Demolombe R. (2001). To trust information sources : a proposal for a modal logical framework. In *Trust and deception in virtual societies*, p. 111–124. Springer.

Demolombe R. (2004). Reasoning about trust : A formal logical framework. In *Trust management*, p. 291–303. Springer.

Deutsch M. (1958). Trust and suspicion. *Journal of conflict resolution*, p. 265–279.

Dung P. M. (1993). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. In *International joint conferences on artificial intelligence*, p. 852–857.

Esfandiari B., Chandrasekharan S. (2001, may). On how agents make friends : Mechanisms for trust acquisition. In *4th workshop on deception, fraud and trust in societies*, vol. 222.

Grandison T., Sloman M. (2000). A survey of trust in internet applications. *Communications Surveys & Tutorials, IEEE*, vol. 3, n° 4, p. 2–16.

Guha R., Kumar R., Raghavan P., Tomkins A. (2004). Propagation of trust and distrust. In *Proceedings of the 13th international conference on world wide web*, p. 403–412.

Josang A., Hayward R., Pope S. (2006). Trust network analysis with subjective logic. In *Proceedings of the 29th australasian computer science conference*, vol. 48, p. 85–94.

Josang A., Ivanovska M., Muller T. (2015, jul). Trust revision for conflicting sources. In *Proceedings of the 18th international conference on information fusion (fusion 2015)*, p. 550–557.

Jousselme A.-L., Boury-Brisset A.-C., Debaque B., Prévost D. (2014). Characterization of hard and soft sources of information : A practical illustration. In *17th international conference on information fusion*, p. 1–8.

Lewis J. D., Weigert A. (1985). Trust as a social reality. *Social Forces*, vol. 63, n° 4, p. 967–985.

Liu W., Williams M.-A. (2002). Trustworthiness of information sources and information pedigree. In *Intelligent agents viii*, p. 290–306. Springer.

Lorini E., Demolombe R. (2008). From binary trust to graded trust in information sources: A logical perspective. *LNAI 5396*, p. 205–225.

Luhmann N. (1979). *Trust and power*. U.M.I.

Marsh S. P. (1994). Formalising trust as a computational concept (Thèse de doctorat, Department of Computer Science and Mathematics, University of Stirling). *Computing Science and Mathematics eTheses*, p. 184.

Matt P.-A., Morge M., Toni F. (2010). Combining statistics and arguments to compute trust. In *Proceedings of 9th international conference on autonomous agents and multiagent systems*, p. 209–216.

McKnight D. H., Chervany N. L. (2000). What is trust? a conceptual analysis and an interdisciplinary model. *Americas Conference on Information Systems*, p. 827–833.

Paglieri F., Castelfranchi C., Costa Pereira C. da, Falcone R., Tettamanzi A., Villata S. (2014). Trusting the messenger because of the message: feedback dynamics from information quality to source evaluation. *Computational and Mathematical Organization Theory*, vol. 20, n° 2, p. 176–194.

Papoulis A., Pillai S. U. (1986). *Probability, random variables, and stochastic processes*. McGraw Hill, New York.

Parsons S., Tang Y., Sklar E., McBurney P., Cai K. (2011). Argumentation-based reasoning in agents with varying degrees of trust. In *Proceedings of the international conference on autonomous agents and multiagent systems.*

Ray C., Gallen R., Iphar C., Napoli A., Bouju A. (2015, May). Deais project: Detection of ais spoofing and resulting risks. In *Oceans 2015 - genova*, p. 1–6.

Stranders R., Weerdt M. de, Witteveen C. (2008). Fuzzy argumentation for trust. In *Computational logic in multi-agent systems*, p. 214–230. Springer.

Sun Y. L., Han Z., Yu W., Liu K. R. (2006). A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks. In *Infocom*, p. 1–13.

Teacy W. T. L., Patel J., Jennings N. R., Luck M. (2006). TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, vol. 12, n° 2, p. 183–198.

Villata S., Boella G., Gabbay D. M., Torre L. van der. (2013). A socio-cognitive model of trust using argumentation theory. *International Journal of Approximate Reasoning*, vol. 54, n° 4, p. 541–559.

Wang Y., Singh M. P. (2006). Trust representation and aggregation in a distributed agent system. In *Aaai*, vol. 6, p. 1425–1430.

Wang Y., Singh M. P. (2007). Formal trust model for multiagent systems. In *International joint conference on artificial intelligence*, p. 1551–1556.

Yan Z., Zhang P., Virtanen T. (2003). Trust evaluation based security solution in ad hoc networks. In *Proceedings of the seventh nordic workshop on secure it systems*, vol. 14.

Yu B., Singh M. P. (2002). An evidential model of distributed reputation management. In *Proceedings of the first international joint conference on autonomous agents and multiagent systems: part 1*, p. 294–301.