Vulnérabilités d'un système de fusion crédal à des sources corrompues : analyses et mesures de modèles d'attaque

Quentin Saint-Christophe*, Christophe Osswald**
Cyril Ray***, Abdel-Ouahab Boudraa***

*Chaire de Cyberdéfense des Systèmes Navals [1]
BCRM Brest, CC 600, 29240 Brest Cedex 9, France
q.saint-christophe@ecole-navale.fr
**Lab-STICC (UMR CNRS 6285) ENSTA Bretagne
2 Rue François Verny, 29200 Brest, France
***IRENav (UR 3624), École navale/Arts-et-Métiers Institute of Technology
BCRM Brest, CC 600, 29240 Brest Cedex 9, France

Résumé. La théorie de Dempster-Shafer, ou théorie de l'évidence, est un cadre mathématique permettant de décrire l'état d'un système à partir de connaissances incomplètes et imparfaites. Son usage dans la propagation de croyances et la combinaison d'informations devient pertinent en présence de données imprécises et incertaines. En introduisant la notion de vulnérabilité de l'information, nous nous intéressons au cas où un agent extérieur corrompt intelligemment une minorité de sources. Nous considérons le conflit ainsi généré comme une quantité valorisable pour la détection de corruption. Plusieurs modèles d'attaques sont étudiés au cours de simulations numériques dans un système de gestion de connaissances. Nous analysons la robustesse du système de gestion de connaissances en fonction de la discrétion de l'agent corrupteur. La règle de combinaison usuelle de l'information est comparée à une nouvelle règle permettant une gestion fine du conflit.

1 Introduction

Les entités épistémiques, appelées propositions ou événements, représentent dans ce travail l'état la connaissance. À un instant donné, ces entités sont supposées exister dans un et un seul état. Toutefois, les données recueillies sont trop souvent hétérogènes, imprécises et incertaines. C'est pourquoi les combiner revient en général à raffiner l'information afin de produire de la connaissance de meilleure qualité. Les techniques de fusion d'information sont basées sur des théories de mesure de confiance telles que les probabilités conditionnelles de Bayes, les probabilités supérieures et inférieures de Dempster ou encore la théorie de Dempster-Shafer (DS) Shafer (1976). La théorie de DS est une extension ensembliste de la théorie des probabilités à l'ensemble des partitions de l'univers. Elle permet notamment de conditionner une information incomplète par une autre afin de produire une connaissance combinée quel que soit son degré

1. Founded and supported by École navale, Thales, ENSTA Bretagne, Naval Group, ENSM and IMT Atlantique

de conflit Dempster (1968). La croyance totale est déduite au travers d'une règle de combinaison, à choisir dans une vaste bibliothèque. Ce cadre se retrouve dans nombreux systèmes de fusion pratiques dans divers domaines, comme l'imagerie médicale ou le consensus d'experts, et apparaît ainsi dans des systèmes sensibles pouvant être la cible d'attaques malveillantes. Ce travail explore les possibilités de telles attaques. Il pose le contexte d'un système générique, qui agrège les résultats de sources imprécises mais honnêtes. Il analyse et simule les possibilités d'impact de la corruption d'une des sources, poursuivant le double objectif de faire varier la décision synthétique et de minimiser les écarts par rapport à la source infectée. Ce travail s'articule comme suit. Dans la section 2, nous rappelons les fondamentaux de la théorie de l'évidence. Une présentation du système d'étude et des modèles d'attaques sont détaillés dans la section 3. L'analyse des résultats de simulations numérique mettant en exergue les modèles d'attaque est traitée dans la section 4. Enfin, la section 5 présente les remarques finales et les perspectives de travaux futurs.

2 La théorie de l'évidence

La théorie de l'évidence a été initiée par Dempster Dempster (1968) et formalisée par Shafer Shafer (1976) pour modéliser l'incertitude d'entités épistémique. Une propriété importante de cette théorie est sa capacité à traiter les sous-ensembles indépendamment des éléments qu'ils englobent.

2.1 Fondamentaux

La théorie de l'évidence est utilisée afin de mesurer l'incertitude dans un univers de discours appelé *cadre de discernement* et noté Ω. Ses éléments sont utilisés pour discerner les états réels du système (e.g. hypothèses, événements, propositions). Ainsi, Ω est un ensemble fini non vide, et d'éléments mutuellement exclusifs.

Définition 1 ***(Fonction de croyance).*** Les preuves à chaque proposition sont attribuées par une *fonction de croyance* (FC) appelée m, telle que $\forall X \subseteq \Omega,\ \sum \mathrm{m}(X) \triangleq 1$.

Définition 2 ***(Éléments focaux).*** $\mathcal{F}(\mathrm{m}) = \{X \subseteq \Omega | \, \mathrm{m}(X) > 0\}$ est appelé *ensemble d'éléments focaux* ; les éléments sur lesquels se focalise l'opinion de m sont les *éléments focaux*.

La fonction m mesure la confiance attribuée exactement à un événement précis et non la confiance globale accordée aux ensembles englobés par cet événement.

2.2 Combiner les preuves

Dans le respect des axiomes de Kolmogoroff, les hypothèses de Shafer sont (i) les FC combinées sont indépendantes ; (ii) les éléments de Ω sont mutuellement exclusifs ; (iii) les éléments de Ω sont exhaustifs (hypothèse du monde fermé). La règle de combinaison se base fondamentalement sur la conjonction des éléments focaux des FCs combinées (cf. Définition 3).

Définition 3 ***(Règle de combinaison conjonctive).*** Soient m_p et m_q deux FCs. Leur combinaision conjonctive est la somme orthogonale suivante : $\mathrm{m}^{\cap}_{p,q}(Z) \triangleq \sum_{X \cap Y = Z} \mathrm{m}_p(X)\,\mathrm{m}_q(Y)$ où $\mathrm{m}^{\cap}_{p,q}(\emptyset)$ est le *conflit*.

Dans le respect de l'hypothèse du monde fermé de Shafer, la FC combinée est normalisée sur la projection des éléments non-vides (cf. Définition 4).

Définition 4 ***(Règle de combinaison DS).*** Soient m_p et m_q deux FCs de conflit différent de 1 (cf. Définition 3). $\forall Z \neq \emptyset,\ \mathrm{m}_{p,q}(Z) \triangleq \frac{\mathrm{m}^{\cap}_{p,q}(Z)}{1-\mathrm{m}^{\cap}_{p,q}(\emptyset)}$ est leur *combinaision DS*.

Ce paradigme du monde fermé suppose que ce conflit provient de sources non fiables, hypothèse que questionnent principalement Smets et Kennes Smets et Kennes (1994). Dans le *Transferable belief model* (TBM), la fiabilité de l'information et la remise en cause de l'exhaustivité de Ω provoquent ainsi la perte de l'axiome de Kolmogoroff. Deux avancées majeures dans la théorie de l'évidence s'ensuivent alors. Premièrement la naissance du paradigme du monde ouvert, qui suppose la non-exhaustivité de Ω comme origine du conflit et autorise l'ensemble vide comme élément focal. Et deuxièmement l'utilisation retrospective de la règle conjonctive (cf. Définition 3), autorisant le conditionnement d'une masse par une autre à l'instar de l'inférence bayésienne. Également en monde ouvert, la règle de combinaison cohérente (cf. Définition 5) proposée par les auteurs, permet d'agréger différemment des FCs dont les conflits internes seraient redondants avec le TBM et al. (2022).

Définition 5 ***(Règle de combinaison cohérente).*** Soit $\mathrm{m}_{p,q}$ la combinaison de m_p et m_q par la *règle de combinaison cohérente* (RCC), sa masse est assignée de la façon suivante :

$$\mathrm{m}_{p,q}(Z) \triangleq \begin{cases} \mathrm{m}^{\cap}_{p,q}(Z) + \sum\limits_{\substack{X \cap Y = \emptyset \\ X' \cap Y \cup X \cap Y' = Z}} \mathrm{m}_p(X)\,\mathrm{m}_q(Y)\,\mathrm{m}_p(X')\,\mathrm{m}_q(Y') & \text{si } Z \neq \emptyset \\ \sum\limits_{\substack{X \cap Y = \emptyset \\ X' \cap Y \cup X \cap Y' = \emptyset}} \mathrm{m}_p(X)\,\mathrm{m}_q(Y)\,\mathrm{m}_p(X')\,\mathrm{m}_q(Y') & \text{sinon} \end{cases}$$

2.3 Connaissances pignistique et doxastique

Dans la théorie de l'évidence, on distingue deux niveaux d'étude : doxastique et pignistique. Le niveau doxastique englobe toutes les connaissances collectées et décrit l'état actuel du système sous la forme d'opinion. Les métriques de décision sont quant à elles utilisées afin de parier sur un état du système. Toute décision est naturellement prise sur des éléments de Ω. Smets préconise alors l'emploi de la *probabilité pignistique* Smets (2005a).

Définition 6 ***(Probabilité pignistique).*** Soit m une FC. Sa *probabilité pignistique* p est la transformation pignistique bet de m : $\forall \omega \in \Omega,\ \mathrm{bet} \circ \mathrm{m}(\omega) = \frac{1}{1-\mathrm{m}(\emptyset)} \sum_{Y | \omega \in Y} \frac{\mathrm{m}(Y)}{|Y|}$.

Remarque 1. L'ensemble vide est par nature inutile en tant qu'élément focal au niveau pignistique car il ne permet pas de projeter une quelconque décision. □

Dans ce travail, l'agent corrupteur s'attaque aux connaissances partielles du niveau doxastique afin de modifier la décision au niveau pignistique.

3 Description du système de gestion de connaissances

Dans cette section est présentée l'architecture du système de gestion de connaissances. On y propose des modèles d'attaque autour de deux critères : la modification de la décision et la discrétion.

3.1 Architecture du système

3.1.1 Présentation générale

On considère un système de gestion de connaissances composé d'entités d'acquisition d'information sur lesquelles se base la décision. Les sources sont supposées hétérogènes, imparfaites et véhiculent chacune une connaissance partielle sous forme d'information indépendante vers un centre de collecte. Celui-ci transfère la connaissance combinée à un centre de décision qui raffine cette dernière en connaissance décisionnelle (cf. Définition 6). On suppose que le système comporte plusieurs sources vulnérables. Ces sources sont sujettes à la manipulation de l'information qu'elles véhiculent. Puisque l'infection d'une majorité de sources vulnérables par un agent assurerait trivialement la corruption du système, nous nous penchons dans ce travail au cas intéressant où une minorité de sources est vulnérable.

3.1.2 Cas d'utilisation à trois sources

Dans l'optique d'étudier un cas représentatif, nous nous intéressons au comportement d'un système de gestion de connaissances composé de trois sources (S_{in1}, S_{in2} et S_{vul}). Une source vulnérable S_{vul} est compromise par l'agent de corruption. Celui-ci y a accès en lecture et écriture (cf. Figure 3.1.2). Dans la suite, on note les FCs des deux sources intègres m_{in1} et m_{in2}, la FC vulnérable m_{vul} et la FC corrompue m_{cor}. La FC du centre de collecte m_{fus} est transformée par le centre de décision en probabilité pignistique p_{fus}. Le rôle de l'agent de corruption est de substituer m_{cor} à m_{vul} afin d'injecter sa propre information.

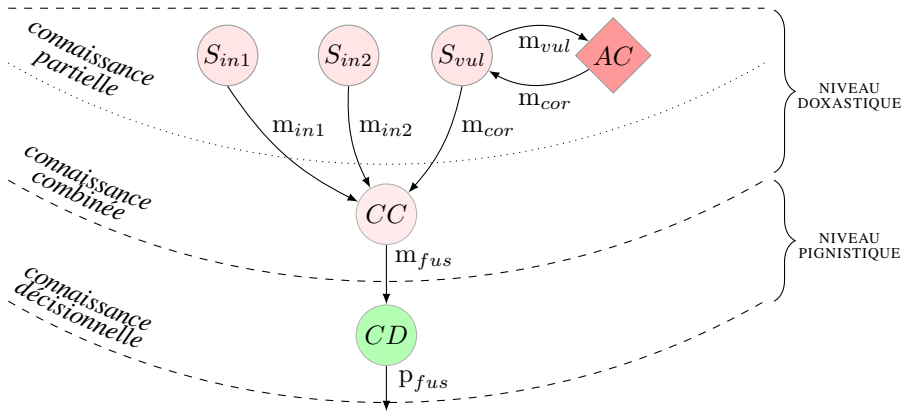


FIG. 1 – *Description du système de gestion de connaissance.*

3.2 Objectif de l'agent de corruption : influencer la décision

On considère que l'agent de corruption a compromis le système. En modifiant l'information de S_{vul}, il tente de dévoyer la décision et ainsi corrompre le système (cf. Définition 7).

Définition 7 ***(Corruption).*** Soit $\mathrm{p}_{fus}[\mathrm{m}]$ la probabilité du centre de décision du système sachant m. Le système est corrompu si : $\arg\max \mathrm{p}_{fus}[\mathrm{m}_{vul}] \neq \arg\max \mathrm{p}_{fus}[\mathrm{m}_{cor}]$.

On présente d'abord un exemple d'attaque où l'agent de corruption a connaissance de la vérité ω_t (cf. Définition 8) et injecte dans le système son complémentaire dans Ω.

Définition 8 ***(Interdiction).*** Soit ω_t la vérité, $\mathrm{m}_{cor} \triangleq \mathrm{m}_{\Omega\setminus\{\omega_t\}}$ est une attaque par interdiction.

Une seconde technique est de choisir la négation de la FC de la source vulnérable (cf. Définition 9) Smets (2005b). En effet, l'ordre de probabilité des éléments de la transformation pignistique est trivialement inversé.

Définition 9 ***(Inversion).*** Soit m_{vul} la FC vulnérable, $\mathrm{m}_{cor} \triangleq \overline{\mathrm{m}}_{vul}$ est une attaque par inversion si $\forall X \subseteq \Omega,\ \overline{\mathrm{m}}_{vul}(X) = \mathrm{m}_{vul}(\Omega \setminus \{X\})$.

3.3 Objectif de l'agent de corruption : faire preuve de discrétion

Le second objectif de l'agent de corruption est de faire preuve de discrétion. Une fois entré dans le système, l'agent de corruption cherche à s'y maintenir. L'agent veut influencer juste assez la décision en modifiant m_{vul} le moins possible. Formellement, on définit ici la discrétion comme une fonction de la distance de Jousselme et al. (2001).

Définition 10 ***(Distance de Jousselme).*** Soient m_p et m_q dans Ω. La distance de Jousselme $\mathrm{d}(\mathrm{m}_p, \mathrm{m}_q)$ vérifie : $\mathrm{d}(\mathrm{m}_p, \mathrm{m}_q) \triangleq \sqrt{\frac{1}{2}(\underline{\mathrm{m}_p} - \underline{\mathrm{m}_q})\underline{\underline{\mathrm{D}}}^{\,t}(\underline{\mathrm{m}_p} - \underline{\mathrm{m}_q})}$; avec $\underline{\underline{\mathrm{D}}}$ une matrice $\left(2^{|\Omega|}\right)^2$ des indices de Jaccard tels que : $\forall (X,Y) \subseteq \Omega^2,\ \underline{\underline{\mathrm{D}}}(X,Y) = \frac{|X\cap Y|}{|X\cup Y|}$ et où $\underline{\mathrm{m}} = (\mathrm{m}(X_i))_{i\leqslant 2^{|\Omega|}}$ est l'écriture vectorielle de m telle que $\sum_i \mathrm{m}(X_i) = 1$.

Sans perte de généralité, on observe dans la suite une notion qualitative de la discrétion.

Définition 11 ***(Discrétion).*** Soient m_{vul} la FC vulnérable et m_{cor} la FC injectée par l'agent de corruption. Pour ϵ un seuil de discrétion, si $\mathrm{d}(\mathrm{m}_{vul}, \mathrm{m}_{cor}) \leqslant \epsilon$, alors l'agent est discret.

En étendant les modèles d'attaque montrés précédemment à la logique de discrétion, la corruption peut être une moyenne arithmétique d'un signal utile m_{vul} et d'une perturbation m_{att} telle que $\mathrm{m}_{cor} = (1-\epsilon)\,\mathrm{m}_{vul} + \epsilon\,\mathrm{m}_{att}$. De plus, afin d'identifier des types d'attaques nouveaux, une méthode consiste à générer aléatoirement ces m_{att} et d'étudier leur impact sur la décision. On choisit de faire un tirage de Monte-Carlo des m_{cor} dans le voisinage de m_{vul} telles que $d(\mathrm{m}_{vul}, \mathrm{m}_{cor}) \leqslant \epsilon$ (cf. Définition 11). On propose ainsi de générer les m_{cor} à partir de la moyenne pondérée de m_{vul} et de FC aléatoires m_{att} (cf .Algorithme 1). La FC aléatoire m_{al}^{opt} qui corrompt le plus discrètement le système donne une idée empirique de la direction dans laquelle chercher des modèles d'attaque nouveaux ($\mathrm{m}_{att} \neq \mathrm{m}_{\Omega\setminus\{\omega_t\}}$ et $\mathrm{m}_{att} \neq \overline{\mathrm{m}}_{vul}$).

4 Résultats numériques

On s'intéresse à la discrétion optimale pour laquelle le système est corrompu. L'information est alors représentée par une FC aléatoire m_{al} dont l'élément le plus probable est la vérité

Entrées : N : nombre de tirages, ϵ : distance maximale, m_{vul} : FC vulnérable, l'élément de vérité : ω_t.
Sorties : m_{att}^{opt} : FC optimale
$\mathrm{m}_{att}^{opt} \leftarrow \mathrm{m}_{vul}, d^{opt} \leftarrow \epsilon, i \leftarrow 0$
tant que $i < N$ **faire**
  $\mathrm{m}_{att} \leftarrow$ FC candidate
  $\mathrm{m}_{cor} \leftarrow (1-\epsilon)\,\mathrm{m}_{vul} + \epsilon\,\mathrm{m}_{att}$
  $\mathrm{m}_f \leftarrow$ fusion de m_{in1},m_{in2} et m_{cor}
  $\mathrm{p}_f \leftarrow \mathrm{bet} \circ \mathrm{m}_f$
  $\omega_{max} \leftarrow \arg\max \mathrm{p}_f$
  $d \leftarrow \mathrm{d}(\mathrm{m}_{vul}, \mathrm{m}_{cor})$
  si $\omega_t \neq \omega_{max}$ *et* $d < d^{opt}$ **alors**
    $\mathrm{m}_{att}^{opt} \leftarrow \mathrm{m}_{att}$
    $d^{opt} \leftarrow d$
    $i \leftarrow i + 1$
retourner m_{att}^{opt}

Algorithme 1 : Recherche par Monte-Carlo de FC curruptrices de discrétion maximale

selon le système $\omega_t = \arg\max \mathrm{p}_{fus}$. Les $\mathrm{m}_{al}(\omega_k)$ sont distribuées uniformément sur le morceau d'hyperplan de $[0;1]^n$ tel que $\sum_{k=1}^{n} \mathrm{m}_{al}(\omega_k) = 1$ avec $\arg\max_{\omega_k} \mathrm{bet} \circ \mathrm{m}_{al}(\omega_k) = \omega_t$.

4.1 Règles de combinaison du centre de collecte

Dans le cadre de la fusion instantanée de multiples FC, la propriété d'associativité de la règle de combinaison permet d'éviter la question de priorité des combinaisons deux-à-deux. De fait, le TBM étant commutatif et associatif, l'ordre de combinaison ds FCs n'importe pas. Ainsi proposons-nous une méthode de fusion multi-source (cf. Définition 12).

Définition 12 ***(Fusion multi-source).*** Soient N FCs m_i et $(\sigma_{i,j})_{i,j\leqslant N,N!}$ les $N!$ permutations de $[\![1,N]\!]$. La FC m_{fus} est la moyenne arithmétique des permutations de combinaison : $\mathrm{m}_{fus} = \frac{1}{N!}\sum_{j=1}^{N!} \bigoplus_{i=1}^{N} \mathrm{m}_{\sigma_{i,j}} = \frac{1}{N!}\sum_{j=1}^{N!} \left(\cdots\left(\left(\mathrm{m}_{\sigma_{1,j}} \oplus \mathrm{m}_{\sigma_{2,j}}\right) \oplus \mathrm{m}_{\sigma_{3,j}}\right)\cdots \oplus \mathrm{m}_{\sigma_{N,j}}\right)$

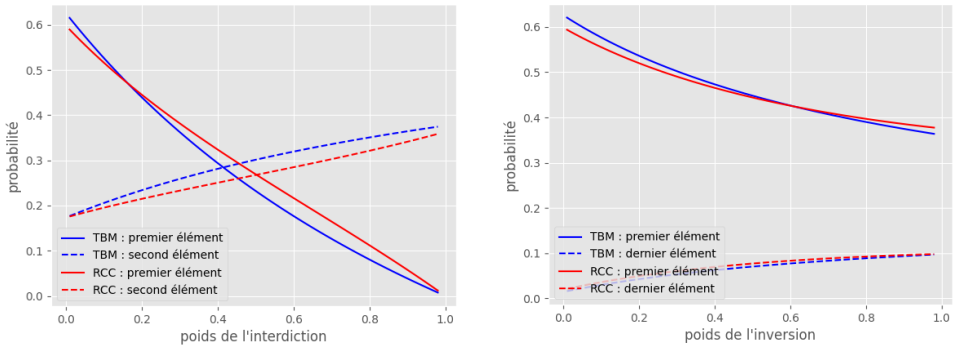
Remarque 2. De par la commutativité, il existe $\frac{N!}{2}$ ordres de combinaison deux-à-deux des N FCs. Pour trois sources : $N = 3 \implies \frac{3!}{2} = 3$ ordres de combinaison différents. □

Afin de comparer les résultats du TBM et de la RCC sur la même méthode, on peut noter que l'associativité du TBM donne : $\forall k \leqslant N!,\ \bigoplus_{i=1}^{N} \mathrm{m}_{\sigma_{i,k}} = \frac{1}{N!}\sum_{j=1}^{N!} \bigoplus_{i=1}^{N} \mathrm{m}_{\sigma_{i,j}}$. La fusion multi-source (cf. Définition 12) dégénère donc en fusion séquentielle pour le TBM.

4.2 Simulations numériques

Dans cette partie, la discrétion maximale de l'agent de corruption est comparée selon la règle de combinaison. La Figure 2 compare l'impact des attaques de l'agent dans le cas d'une interdiction et d'une inversion. Les probabilités des éléments ω_t et $\omega_f = \arg\max_{\Omega\setminus\{\omega_t\}} \mathrm{p}_{fus}$ (second élément le plus probable) sont comparées en fonction du poids ϵ. En effet, on recherche le point de bascule dans la décision entre ces deux éléments. Les courbes de même couleur se

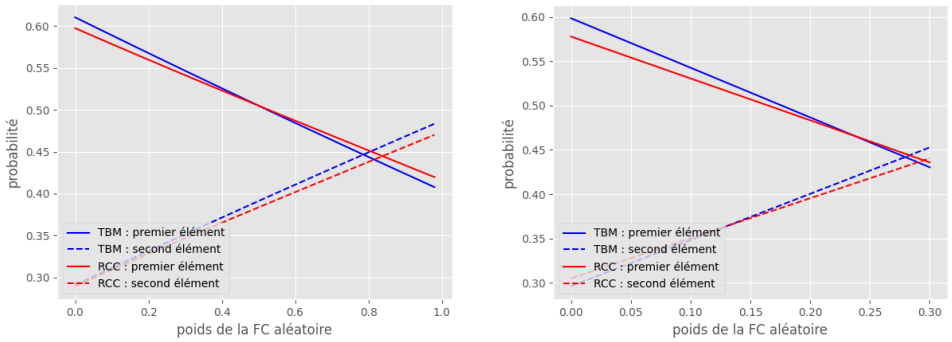
croisent lorsque $\mathrm{p}_{fus}(\omega_t) = \mathrm{p}_{fus}(\omega_f)$ auxquels cas le système est corrompu. Lors d'une attaque par interdiction, on remarque que le système est plus discrètement corrompu avec le TBM qu'avec la RCC. L'attaque par inversion ne permet pas quant à elle de corrompre le système. La Figure 3 du tirage de Monte-Carlo de m_{att}^{opt} permettant de faire basculer la décision



(a) (b)

FIG. 2 – *Robustesse de la RCC et du TBM pour (a)* $\mathrm{m}_{att} = \mathrm{m}_{\Omega\setminus\{\omega_t\}}$ *et (b)* $\mathrm{m}_{att} = \overline{\mathrm{m}}_{vul}$

(cf. Algorithme 1). Une fois m_{att}^{opt} trouvée, on fait varier son poids ϵ afin de faire basculer la décision pour les deux modèles de combinaison. Dans la partie (a) de la figure, l'efficacité de la FC optimale corrompant le système par le TBM est mise en regard du même système quand il fonctionne par la RCC. La partie (b) est l'exact inverse. Ces simulations permettent d'analyser la pertinence des directions choisies par les deux tirages sous TBM et RCC. Dans les deux cas le poids ϵ minimum requis tel que $\mathrm{p}_{fus}(\omega_t) = \mathrm{p}_{fus}(\omega_f)$ est supérieur pour la RCC. Ces résultats montrent donc que l'agent doit produire une attaque moins discrètes pour corrompre un système combinant l'information recueillie avec la RCC.



(a) (b)

FIG. 3 – *Attaques de type* $\mathrm{m}_{cor} = (1-\epsilon)\,\mathrm{m}_{vul} + \epsilon\,\mathrm{m}_{att}^{opt}$ *de distance minimale. En bleu le TBM et en rouge la RCC. Étude de la robustesse respectivement (a) de la RCC et (b) du TBM subissant une attaque optimale contre respectivement (a) le TBM et (b) de la RCC.*

5 Conclusion

Dans ce travail ont été explorés différents modèles d'attaque sur un système de gestion de connaissances dont la faille est une source vulnérable. Ces attaques sont menées par un agent de corruption dont le double objectif a été de modifier la décision du système et de rester discret. En définissant la discrétion selon une distance à la source vulnérable, le seul levier de corruption de l'agent a été l'accès en écriture et en lecture sur cette unique source, minoritaire dans un système à trois sources. Deux types de combinaison de l'information ont été comparés (TBM et RCC) afin d'étudier leur robustesse face à ces attaques. Les simulations numériques montrent que la RCC est plus robuste que le TBM aux attaques les plus discrètes, qui peuvent également être considérées comme les pannes les plus malchanceuses. Une suite complémentaire à ce travail serait d'étendre la simulation à N sources et d'exposer le système à des attaquants communiquant et coopérant. Il serait également intéressant d'étudier le comportement d'un tel système piloté par d'autres types de combinaisons gérant le conflit différemment.

Références

Dempster, A. P. (1968). A generalization of bayesian inference. *J. Royal Statistical Society Series B 30*(2), 205–32.

et al., A.-L. J. (2001). A new distance between two bodies of evidence. *Information fusion 2*(2), 91–101.

et al., Q. P. B. S.-C. (2022). Du concept de conflit cohérent en fusion d'informations vulnérables. In *28e conférence sur le traitement du signal et de l'image*, Volume 001-102, pp. 409–12. GRETSI.

Shafer, G. (1976). *A mathematical theory of evidence*.

Smets, P. (2005a). Decision making in the tbm : the necessity of the pignistic transformation. *Inter. Journal of Approximate Reasoning 38*(2), 133–47.

Smets, P. (2005b). Managing deceitful reports with the transferable belief model. In *7th International Conference on Information Fusion*, Volume 2, pp. 7–pp. IEEE.

Smets, P. et R. Kennes (1994). The transferable belief model. *Art. Intel. 66*, 191–234.

Summary

Dempster-Shafer theory, or evidence theory, is a mathematical framework for describing the state of a system based on incomplete and imperfect knowledge. Its use in the propagation of beliefs and the combination of information becomes relevant in the presence of imprecise and uncertain data. By introducing the notion of information vulnerability, we focus on the case where an external agent intelligently corrupts a minority of sources. We consider the conflict thus generated as a valuable quantity for corruption detection. Several attack models are studied during numerical simulations in a knowledge management system. We analyse the robustness of the knowledge management system as a function of the discretion of the corrupting agent. The usual rule for combining information is compared with a new rule for fine-grained conflict management.