

Sur la croyance, la plausibilité et l’immersivité associées à un réseau de profils fictifs utilisé comme un dispositif de sonde

Thierry Berthier, Olivier Kempf, Eric Hazane, Thomas Anglade

Abstract.

Keywords: cybersécurité, leurre, ADFI, Sentinelles 2022, fonctions de croyance, plausibilité, immersivité, Dempster-Shafer

1 - Introduction

Les Architectures de Données Fictives Immersives (ADFI) ont été décrites dans [1] à l’aide du formalisme des projections algorithmiques. Elles peuvent être utilisées par un attaquant durant une phase initiale d’ingénierie sociale et de collecte de données ou dans le cadre d’une installation d’honeytrap dédié au leurrage (expérimentation Cyberreason, juin 2020 [2]) ou encore lors du déploiement de sondes statistiques sur certains réseaux sociaux (projet Sentinelles 2022). Nous proposons dans cet article d’explorer les trois paramètres déterminant l’efficacité d’une ADFI : la croyance, la plausibilité et l’immersivité de l’architecture déployée. La croyance et la plausibilité peuvent être quantifiées par les fonctions de croyance (croyance que la vérité est dans une hypothèse A) et par les fonctions de plausibilité définies dans la théorie de Dempster-Shafer. Plus complexe à définir, l’immersivité tient compte des qualités intrinsèques de l’ADFI. En particulier de sa capacité à bloquer l’arrivée d’informations extérieures (auprès de l’observateur) susceptibles de susciter son doute sur la véracité de l’architecture et de remettre en question sa légitimité. Une ADFI cumulant les trois qualités (croyance, plausibilité, immersivité) offre des garanties fortes d’efficacité à celui qui la déploie, dans la durée et dans la largeur du spectre des observateurs leurrés.

Nous proposons d’appliquer la théorie de Dempster-Shafer sur un réseau de profils fictifs (ou réseau de comptes tests) déployé sur certains réseaux sociaux dans l’objectif de construction d’une architecture fonctionnant en mode « sonar passif ». Typiquement, cette architecture de comptes tests peut servir à mesurer le niveau d’activité et de diffusion de messages à caractères politiques produits par des robots conversationnels (bots russes, chinois, ...) en période préélectorale (projet initial canadien, repris et adapté dans le projet Sentinelles 2022 porté par le groupe « sécurité-IA » du Hub France IA).

2 - Modèle de réseau de profils fictifs

Nous appliquons la théorie des fonctions de croyance au contexte de supervision d'une architecture constituée d'un ensemble de profils fictifs interagissant avec un ensemble de profils extérieurs.

Il est important de préciser le sens du mot « fictif ». Un profil fictif est un ensemble de données synthétiques formant un profil plus ou moins réaliste créé sur un réseau social ou sur une plateforme en ligne et destiné à tromper ceux avec qui il interagit. Ces données synthétiques peuvent ne correspondre avec aucun individu existant (on parlera dans ce cas de profil pleinement synthétique) ou au contraire, elles peuvent usurper l'identité d'un utilisateur bien réel.

Notations et modèle

Un réseau fictif $\mathcal{R}_{\text{fict}}$ est la donnée du triplet $(\mathcal{P}_{\text{fict}}, \mathcal{S}_{\text{fict}}, \mathcal{A}_{\text{fict}})$ formé d'un ensemble de profils fictifs opérés par un superviseur et évalués en aveugle par un analyste.

Profils fictifs et superviseur

$\mathcal{P}_{\text{fict}}$ désigne un ensemble de n profils fictifs $\{ PF_1, \dots, PF_n \}$ opérés par un superviseur $\mathcal{S}_{\text{fict}}$ qui a accès à la totalité de l'information transitant par les profils fictifs qu'il commande et contrôle. Son objectif est de maximiser la durée d'action de $\mathcal{P}_{\text{fict}}$ et de non-identification de son architecture par les profils extérieurs.

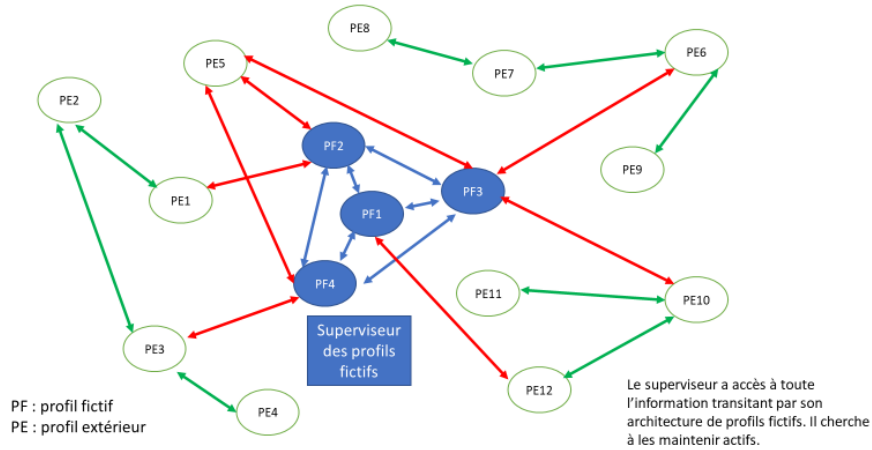


Fig.1 – le réseau de profils vu du Superviseur $\mathcal{S}_{\text{fict}}$

Profils extérieurs

\mathcal{P}_{ext} désigne l'ensemble (évolutif) des profils extérieurs avec lesquels les profils fictifs communiquent. A l'instant t , $\mathcal{P}_{\text{ext}} = \{ PE_1, \dots, PE_m \}$. Initialement ($t=0$), les profils extérieurs n'ont aucune information sur le caractère fictif du sous-réseau $\mathcal{P}_{\text{fict}}$. Au fil du temps, des échanges d'information s'établissent entre \mathcal{P}_{ext} et $\mathcal{P}_{\text{fict}}$. De nouveaux profils extérieurs se connectent aux profils fictifs alors que d'autres profils extérieurs ne communiquent plus ou se déconnectent de leurs interlocuteurs fictifs. L'indicateur principal est celui de la « confiance – défiance » qu'attribue un profil extérieur à un profil fictif avec lequel il interagit. Lorsque le niveau de confiance diminue, le profil extérieur peut, dans le doute, restreindre ou suspendre ses communications avec le profil fictif concerné et/ou décider de rompre définitivement sa connexion. Dans ce cas, l'intégrité fonctionnelle (attractivité des profils) du réseau fictif peut être compromise. Lorsqu'un profil extérieur découvre et déclare le caractère fictif de l'ensemble $\mathcal{P}_{\text{fict}}$, le superviseur $\mathcal{S}_{\text{fict}}$ met fin à l'activation de l'architecture fictive.

Analyseur

L'architecture fictive est équipée d'un analyseur $\mathcal{A}_{\text{fict}}$ qui ne dispose pas d'information a priori (au temps $t=0$) sur l'existence ou non de profils fictifs dans le réseau analysé mais qui cherche à les détecter en fonction des comportements et des échanges auquel il a accès. L'analyseur apporte ainsi un regard extérieur sur l'architecture de profils fictifs. Son rôle est de détecter « en aveugle » d'éventuelles contradictions ou anomalies susceptibles de révéler le caractère fictif d'un profil.

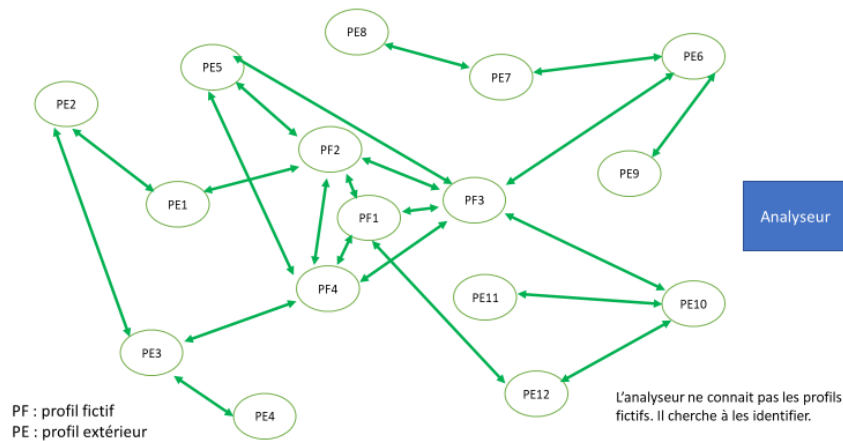


Fig.2 – le réseau de profils vu de l'analyseur $\mathcal{A}_{\text{fict}}$

On cherche à automatiser le plus de processus au sein du réseau fictif $\mathcal{R}_{\text{fict}} = (\mathcal{P}_{\text{fict}}, \mathcal{S}_{\text{fict}}, \mathcal{A}_{\text{fict}})$, en particulier, les fonctionnalités de l'analyseur $\mathcal{A}_{\text{fict}}$.

3 - Théorie de Dempster-Shafer et fonctions de croyance

Historique

La théorie de l'évidence ou théorie des fonctions de croyance a été introduite par Dempster en 1967 [3] sur les bornes inférieures et supérieures d'une famille de distributions de probabilités et puis complétée par Shafer en 1976. Shafer a construit un modèle mathématique de croyance ou théorie des fonctions de croyance pour la modélisation de connaissances incertaines. Ce modèle permet de manipuler des degrés de confiance ou masse de croyance associés à la validité d'une information (Tacnet et al., 2010; Pearl, 1990). Cette théorie est basée sur un modèle d'inférence statistique qui généralise l'inférence Bayésienne.

Plusieurs approches distinctes de la théorie de croyance ont été proposées : (Simon et al., 2007, 2008) utilisent un réseau de fonctions de croyance pour l'étude de la fiabilité des systèmes s'appuyant sur les algorithmes d'inférence des réseaux Bayésiens. (Denoeux et Govaert, 1996) construisent une approche basée sur les procédures supervisées et non supervisées en s'appuyant sur la théorie de Dempster-Shafer pour le diagnostic des systèmes. (Smets et Kennes, 1994) définissent un modèle des croyances transférables. (Sallak, 2007) a montré une évaluation de paramètres de sûreté de fonctionnement en présence d'incertitudes et d'aide à la conception.

Eléments de la théorie des fonctions de croyances

La théorie des fonctions de croyances de Dempster – Shafer fournit une approche intéressante pour modéliser l'incertitude et l'arbitraire dans le cadre du modèle des croyances transférables.

L'analyseur $\mathcal{A}_{\text{fict}}$ fonctionne sans information préalable sur l'existence ou non de profils fictifs dans le réseau qu'il observe. Il construit ses propres croyances au fil des échanges de données entre $\mathcal{P}_{\text{fict}}$ et \mathcal{P}_{ext} .

L'analyseur $\mathcal{A}_{\text{fict}}$ cherche à répondre à la question centrale :

$$Q = \ll \text{Qui sont les profils fictifs ?} \gg.$$

Cadre de discernement

Soit $\Omega = \{ \omega_1, \dots, \omega_K \}$ un ensemble fini de réponses à la question Q.

L'ensemble Ω est le cadre de discernement de l'analyseur.

Masse de croyance

Une fonction de masse de croyance sur Ω est une application m définie sur l'ensemble des parties de Ω telle que $m : 2^\Omega \rightarrow [0,1]$ vérifie $\sum_{A \subseteq \Omega} m(A) = 1$

Toute partie $A \subseteq \Omega$ telle que $m(A) > 0$ est appelée élément focal de m

La fonction de masse m représente l'état de la connaissance d'un agent rationnel à un instant t , relativement à la question Q . C'est une opinion pondérée. La masse $m(A)$ est la part de croyance allouée à A (et à aucun autre sous-ensemble strict). La masse $m(\Omega)$ correspond au degré d'ignorance totale.

Cas particuliers de fonctions de masse	Distribution de la croyance
Fonction de masse catégorique sur $B \subseteq \Omega$	$m_B(B) = 1$
Fonction de masse vide (ignorance totale)	$m_\Omega(\Omega) = 1$
Fonction de masse Bayésienne	Les éléments focaux sont des singletons
Fonction de masse consonante	Les éléments focaux sont emboîtés
Fonction de masse simple	Il existe, au plus, deux éléments focaux dont Ω

Fonction de croyance bel(.)

Pour tout $A \subseteq \Omega$, $bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B)$

La croyance $bel(A)$ représente la part totale de croyance soutenant A

Fonction de plausibilité $pl(\cdot)$

Pour tout $A \subseteq \Omega$, $pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$

La plausibilité $pl(A)$ représente la part maximale de croyance qui pourrait soutenir A .

Fonction d'implicabilité

Pour tout $A \subseteq \Omega$, $b(A) = \sum_{B \subseteq A} m(B)$

$b(A)$ représente la somme des masses allouées aux sous-ensembles de A , $b(\Omega) = 1$

Fonction de communalité

Pour tout $A \subseteq \Omega$, $q(A) = \sum_{B \supseteq A} m(B)$

$q(A)$ représente la somme des masses allouées aux sur-ensembles de A , $q(\emptyset) = 0$

Les fonctions $bel(\cdot)$ et $pl(\cdot)$ sont des mesures non additives :

$$bel(A \cup B) \geq bel(A) + bel(B) - bel(A \cap B)$$

$$pl(A \cup B) \leq pl(A) + pl(B) - pl(A \cap B)$$

Relations entre ces fonctions

Pour tout $A \subseteq \Omega$, on a : $m(A) = \sum_{B \subseteq A} bel(B)(-1)^{|A| - |B|}$ si $A \neq \emptyset$

et $m(A) = 1 - bel(\Omega)$ si $A = \emptyset$

Pour tout $A \subseteq \Omega$, $bel(A) = pl(\Omega) - pl(A^c)$

Pour tout $A \subseteq \Omega$, $m(A) = \sum_{B \subseteq A} b(B)(-1)^{|A| - |B|}$

Pour tout $A \subseteq \Omega$, $m(A) = \sum_{A \subseteq B} q(B)(-1)^{|B| - |A|}$

Changement de cadre de discernement et transfert d'information

Soit ρ une application de θ dans Ω , soit m_θ une fonction de masse sur θ traduisant un certain état de connaissance. On exprime cet état de connaissance dans un référentiel plus fin en transportant m_θ dans Ω :

pour tout $A \subseteq \theta$, $m_\Omega(\rho(A)) = m_\theta(A)$

Ω est un raffinement de θ alors que θ est un grossissement de Ω

Fusion d'informations

Le contexte est celui où différentes sources d'information s'expriment sur le même cadre de discernement et donnent des informations sur le même objet. Il est alors possible de combiner ces informations via une seule masse de croyance. Il existe plusieurs types de combinaison de masses de croyances : la combinaison conjonctive, la combinaison conjonctive normalisée, la combinaison disjonctive.

Combinaison conjonctive

On suppose que m_1 et m_2 sont deux masses issues de deux sources d'informations fiables et distinctes. La combinaison conjonctive $m_1 [\cap] m_2$ est définie par :

$$\text{Pour tout } A \subseteq \Omega, \quad m_{1[\cap]2}(A) = m_1 [\cap] m_2(A) = \sum_{B \cap C = A} m_1(B)m_2(C)$$

La combinaison conjonctive est associative, commutative, de neutre m_Ω . Si m_A et m_B sont deux fonctions de masse catégoriques, alors $m_A [\cap] m_B = m_{A \cap B}$

Le conditionnement de m par A est défini par $m [\cap] m_A$

Combinaison conjonctive normalisée

La forme normalisée est donnée par la règle de Dempster-Shafer

$m_{1\oplus 2}(A) = m_1 \oplus m_2(A) = \frac{m_{1[\cap]2}(A)}{1 - K_{12}}$ si $A \neq \emptyset$ avec $K_{12} = m_{1[\cap]2}(\emptyset)$ le degré de conflit

$$m_{1\oplus 2}(A) = 0 \text{ si } A = \emptyset$$

Combinaison disjonctive

On suppose que m_1 et m_2 sont deux masses telles que au moins une des deux sources est fiable (sans savoir laquelle des deux est fiable).

$$\text{Pour tout } A \subseteq \Omega, \quad m_{1[\cup]2}(A) = m_1 [\cup] m_2(A) = \sum_{B \cup C = A} m_1(B)m_2(C)$$

Affaiblissement et doute sur la fiabilité d'une information

Lorsqu'il existe un doute sur la fiabilité d'une des informations m fournie par une source S , on lui applique un taux d'affaiblissement $\alpha \in [0,1]$. Le degré de fiabilité de la source S est $1-\alpha$

$$\text{Pour tout } A \subseteq \Omega, \quad {}^\alpha m(A) = (1-\alpha)m(A) \text{ et } {}^\alpha m(\Omega) = (1-\alpha)m(\Omega) + \alpha$$

$${}^\alpha m = (1-\alpha)m + \alpha m_\Omega$$

Prise de décision

Formalisation de la prise de décision dans un cadre Bayésien classique

Dans un cadre bayésien classique, il faut fixer :

L'ensemble des décisions pouvant être prises $\mathcal{D} = \{d_1, \dots, d_n\}$,

L'ensemble des états ou cadre de pari Γ

Une fonction de coût : $C : \mathcal{D} \times \Gamma \rightarrow \mathbb{R}$ telle que $C(d, \gamma)$ représente le coût de décider d alors que la vérité est γ

L'incertitude sur Γ étant décrite par une mesure de probabilité P , on choisit la décision d minimisant le risque espéré $R_P(d) = E_P[C(d, \cdot)] = \sum_{\gamma \in \Gamma} C(d, \gamma)P(\{\gamma\})$

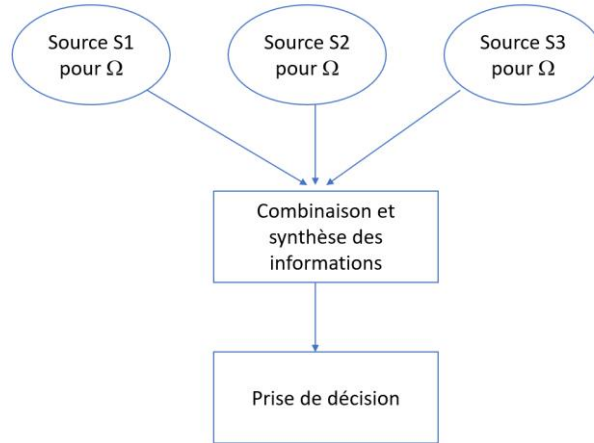


Fig.3 – Processus de prise de décision

Formalisation de la prise de décision étendue au cadre des fonctions de croyance

On intègre les fonctions de croyance au processus de prise de décision en distinguant ce qui relève de la modélisation et manipulation d'informations d'un côté (niveau crédal) et de l'autre (niveau pignistique), ce qui relève de la prise de décision. Une mesure de probabilité est nécessaire pour éviter une séquence de paris perdants (Dutch book problem). On utilise pour cela la transformation pignistique définie par :

Γ une composition de grossissements/raffinements de Ω
 $m^\Omega \rightarrow m^\Gamma \rightarrow$ probabilité BetP

$$\text{BetP}(\{\gamma\}) = \sum_{A \subseteq \Gamma, \gamma \in A} \frac{m(A)}{|A|} \times \frac{1}{1 - m(\emptyset)}$$

On choisit la décision minimisant R_{BetP}

Fonction de coût 0-1

$\Gamma = \{\gamma_1, \dots, \gamma_n\}$ et $\mathcal{D} = \{d_1, \dots, d_n\}$ où la décision d_i signifie choisir γ_i

La fonction de coût C est définie par :

$C(d_i, \gamma_j) = 0$ si $i=j$ (bonne réponse) et $C(d_i, \gamma_j) = 1$ si $i \neq j$ (mauvaise réponse)

Dans ce cas, minimiser le risque R_{BetP} revient à choisir la décision maximisant BetP

Risques espérés inférieurs et supérieurs

Le risque espéré inférieur est défini par : $R_{\text{inf}}(d) = \sum_{A \subseteq \Gamma} m(\{A\}) \min_{\gamma \in A} C(d, \gamma)$

Le risque espéré supérieur est défini par : $R_{\text{Sup}}(d) = \sum_{A \subseteq \Gamma} m(\{A\}) \max_{\gamma \in A} C(d, \gamma)$

La stratégie de décision optimiste consiste à choisir la décision d qui minimise le risque inférieur R_{inf}

La stratégie de décision pessimiste consiste à choisir la décision d qui minimise le risque supérieur R_{sup}

Cas particulier du coût 0-1 :

$\Gamma = \{\gamma_1, \dots, \gamma_n\}$ et $\mathcal{D} = \{d_1, \dots, d_n\}$ où la décision d_i signifie choisir γ_i

La fonction de coût C est définie par $C(d_i, \gamma_j) = 0$ si $i=j$ (bonne réponse) et $C(d_i, \gamma_j) = 1$ si $i \neq j$ (mauvaise réponse)

Dans ce cas, le risque inférieur est $R_{\text{inf}}(d_i) = 1 - pl(\{\gamma_j\})$ où pl est la fonction de plausibilité.

Le risque supérieur est $R_{\text{sup}}(d_i) = 1 - bel(\{\gamma_j\})$ où bel est la fonction de croyance.

La stratégie de décision optimiste consiste à choisir la décision d qui maximise la plausibilité $pl(\cdot)$

La stratégie de décision pessimiste consiste à choisir la décision d qui maximise la fonction de croyance $bel(\cdot)$

Méthode d'analyse multicritères AHP (Analytic Hierarchy Process)

La méthode d'analyse hiérarchique multicritères AHP repose sur la décomposition hiérarchique du problème en sous-problèmes. Les données pertinentes sont ensuite extraites à l'aide d'un ensemble de comparaisons par paire. Ces comparaisons sont utilisées pour obtenir les poids d'importance des critères mais également les performances des alternatives pour chaque critère.

Le premier principe de la méthode conduit à décomposer le problème en ses composantes principales et à réitérer cette décomposition sur les composantes ainsi obtenues autant de fois que nécessaire. Ainsi, la construction de la hiérarchie s'opère selon une démarche descendante. Le nombre de niveaux de cette hiérarchie dépend de la complexité du problème et de son degré de détail. Une hiérarchie AHP possède au moins trois niveaux. L'objectif principal du problème est représenté au plus haut niveau de la hiérarchie. Ensuite, chaque niveau de la hiérarchie contient des critères ou sous-critères qui influencent la décision. Le dernier niveau de la structure contient les alternatives.

L'ensemble d'alternatives $\Omega = \{\omega_1, \dots, \omega_K\}$ correspond à l'ensemble des profils analysés par l'analyste $\mathcal{A}_{\text{fict}}$

L'ensemble des critères $\Theta = \{c_1, \dots, c_n\}$ permettent à l'analyste de décider si un profil lui semble fictif.

Critères associés à la cohérence d'un profil

Cri- tères	Descriptif	Commentaire	Score de signifiante (de 1 à 10)
Critères associés à la photo de profil			
C ₁	Absence de photo de profil		5
C ₂	Photo de profil déjà employée dans un autre profil sous un nom diffé- rent	Détection par une requête sur un mo- teur de recherche d'image inversée	10
C ₃	Photo de profil retou- chée, modifiée, ou syn- thétique (GAN)	Photo récupérée sur « this person does not exist » https://thispersondoes- notexist.com/	9
C ₄	Cohérence « photo - titres et profession »	Incohérence entre l'âge et le niveau de responsabilité ou d'activité du profil	4
Critères associés aux données et métadonnées du profil			
C ₅	Date de création du profil (compte débutant son activité)	Utilisation brusque et intensive d'un compte inactif	2
Critères associés à l'attractivité du profil			
C ₆	Faiblesse du nombre d'amis ou de relations du profil par rapport la date de création du profil		2

C ₇	Ratio entre le nombre de demande de connexions reçues et le nombre de demande de connexions envoyées par le profil		3
C ₈	Forte dispersion géographique des amis ou des relations du profil		4
Critères associés au niveau d'activité de publication du profil			
C ₉	Niveau d'activité et de publications ouvertes du profil sur une période donnée	Volume de messages publiés Cohérence avec le volume moyen d'activité sur le réseau social concerné	3
C ₁₀	Ratio de republication de messages	Nombre de republication divisé par le nombre total de messages publiés sur une période donnée	8
C ₁₁	Date, horaires et fréquences de publication	Compatibilité avec le fuseau horaire de la position	3
Critères associés au contenu publié ou échangé par le profil			
C ₁₂	Ratio image et vidéo	Nombre d'images et vidéos publiées divisé par le nombre total de publications sur une période donnée	3

C ₁₃	Taux de messages polarisants ou clivants	Nombre de messages clivants divisé par le nombre total de messages	7
C ₁₄	Cohérence sémantique des messages publiés ou échangés	Cohérence des messages avec les titres, fonctions et activités du profil	3

Les critères sont hiérarchisés par paire selon les règles du tableau suivant.

Valeur numérique	Définition	Description
1	Importance égale des deux critères.	Les deux critères contribuent autant au critère père.
3	Faible importance d'un critère par rapport à un autre.	L'expérience et l'appréciation personnelles favorisent légèrement un élément par rapport à un autre.
5	Importance forte ou déterminante d'un critère par rapport à un autre.	L'expérience et l'appréciation personnelles favorisent fortement un critère par rapport à un autre.
7	Importance attestée d'un critère par rapport à un autre.	Un critère est fortement favorisé et sa dominance est attestée dans la pratique.
9	Importance absolue d'un critère par rapport à un autre.	Les preuves favorisant un critère par rapport à un autre sont aussi convaincantes que possible.

2,4,6,8	Valeurs intermédiaires entre deux appréciations voisines.	Un compromis est nécessaire entre deux appréciations.
---------	---	---

Echelle de mesure de la méthode AHP.

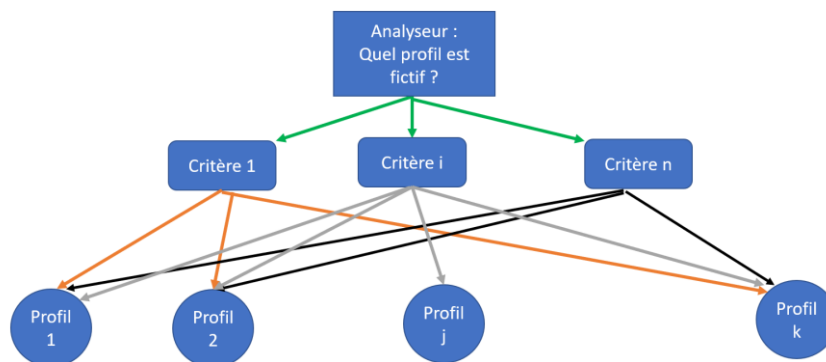


Fig. 4 – Hiérarchie pour la sélection des profils fictifs dans la cadre de la méthode AHP

Comparaison par paire

Une fois la hiérarchie construite, la seconde phase de la méthode consiste à réaliser les différentes comparaisons afin de définir les priorités. Ces comparaisons sont réalisées à chaque niveau de la hiérarchie. On commence donc à comparer les critères deux à deux par l'intermédiaire d'une question de type : "Quelle est l'importance du critère c_i par rapport au critère c_j ?". Ces questions permettent d'établir les poids des critères. Des questions similaires sont posées au niveau hiérarchique inférieur (alternatives) mais en conservant le lien avec le nœud supérieur. A ce niveau, les questions sont du type : "En considérant le critère c_i , quelle est l'importance de l'alternative ω_i par rapport à l'alternative ω_j ?". Les comparaisons par paire sont évaluées par l'intermédiaire d'une échelle numérique définie par Saaty composée de 9 niveaux. Le choix de cette dernière repose sur la capacité naturelle de l'être humain à "établir des distinctions dans la force des relations qui existent entre les éléments".

Les comparaisons par paire sont présentées dans une matrice carrée, notée $M = (v_{ij})$ pour les alternatives $(i,j) \in \{1, \dots, k\}$. La valeur v_{ij} quantifie l'importance de la proposition i sur la proposition j . Lorsqu'une proposition est comparée à elle-même $V_{ii} = 1$ et afin de respecter la transitivité $v_{ij} = 1/v_{ji}$

Indice de cohérence

Afin de vérifier que les comparaisons sont cohérentes entre elles, on peut calculer un indice de cohérence CI défini par : $CI = \frac{\lambda_{\max} - n}{n-1}$ avec λ_{\max} la valeur propre maximale de la matrice étudiée et n son ordre. Le taux de cohérence $CR = CI / RI$ où RI est un indice dépendant de l'ordre de la matrice. Si CR est inférieur à 0,1 alors la matrice est considérée comme cohérente. Si ce n'est pas le cas, il faut changer de matrice.

Vecteur de Priorité

La synthèse des comparaisons précédentes est réalisée en appliquant la moyenne arithmétique sur chaque ligne des matrices. On obtient dans le cas de la matrice d'importance des critères :

$$v_i = \left(\prod_{j=1}^m v_{ij} \right)^{1/n}$$

Le vecteur de poids des critères $T = \{ \tau_1, \dots, \tau_m \}$ est alors défini par :

$$\tau_i = v_i / \sum_{k=1}^m v_k$$

τ_i représente le poids du critère c_i . Ce poids est noté par la suite p_i . On peut appliquer la même démarche pour les matrices concernant les alternatives. Dans ce cas, on note $v_{ij} = \tau_i$ l'importance de l'alternative i selon le critère j .

Agrégation

La dernière phase de la méthode AHP est la synthèse des appréciations sous la forme d'une valeur unique. Les priorités pour chaque alternative ω_k désignées par A_k^{AHP} en fonction de tous les critères combinés sont déterminés selon la formule suivante :

$$A_k^{AHP} = \sum_{j=1}^m v_{kj} p_j \text{ pour tout } k = 1, \dots, K$$

Où v_{kj} représente l'importance de l'alternative ω_k sachant le critère C_j
 p_j indique le poids du critère C_j .

Les limites de l'analyse multicritères AHP

Il est souvent difficile d'évaluer avec précision l'importance entre les alternatives, pour un critère donné. De la même façon, il est difficile d'évaluer l'importance des

critères. L'échelle d'évaluation fixée n'offre pas une précision optimale. Elle ne tient pas compte de l'ignorance partielle de l'expert qui fournit des évaluations et qui ne connaît pas toujours l'importance de toutes les alternatives pour tous les critères. Si la théorie des ensembles flous apporte des solutions partielles à ces limitations, elle ne tient pas compte des relations conditionnelles entre les alternatives et les critères. Dans l'approche AHP, on demande à l'expert d'évaluer l'importance des alternatives par des comparaisons deux à deux selon un critère. Toutefois, la fonction de croyance obtenue n'est pas conditionnée à ce critère ce qui est en contradiction avec le raisonnement demandé à l'expert.

En 2012, Éric Lefèvre [4] a proposé une approche associant les fonctions de croyance à l'analyse AHP.

Analyse multicritères AHP et fonctions de croyance [4]

Dans une première phase, il est nécessaire d'identifier les critères et les alternatives afin de procéder à la comparaison par paire. Souvent l'expert n'est pas en mesure de discerner l'importance entre certains critères. La méthode développée par E. Lefèvre consiste à proposer à l'expert d'exprimer ses opinions sur des groupes de critères au lieu de le faire sur un seul critère. L'expert rassemble les critères dans le même sous-ensemble lorsqu'il n'y a pas suffisamment d'informations pour les comparer.

Soit $\Theta = \{ c_1, \dots, c_m \}$ l'ensemble des critères. On note \mathcal{C}_k un sous-ensemble de Θ appartenant à 2^Θ

Les groupes de critères peuvent être définis par $\mathcal{C}_k > \mathcal{C}_j \quad \forall k, j$ tels que $\mathcal{C}_k, \mathcal{C}_j \in 2^\Theta$ $\mathcal{C}_k \cap \mathcal{C}_j = \emptyset$ où $>$ traduit la préférence. On procède de la même manière pour les alternatives : le décideur peut comparer des alternatives seules mais également les regrouper dans des sous-ensembles pour comparer ces sous-ensembles. Soit $\Omega = \{ \omega_1, \dots, \omega_K \}$ l'ensemble des alternatives, soit $A_k \in 2^\Omega$

Construction de la fonction de croyance associée aux critères

Une fois les critères et les alternatives définis, l'expert précise ses préférences afin d'obtenir les poids des critères et les préférences des alternatives selon chaque critère. On utilise l'échelle de Saaty afin d'évaluer l'importance des paires d'éléments. Les fonctions de croyance sont ensuite générées en utilisant la méthode décrite dans la partie précédente. Le calcul de l'indice de consistance permet de vérifier que les matrices de comparaison par paire sont cohérentes. Pour les critères, une fois la comparaison par paire réalisée, on obtient des poids qui peuvent être assimilés à des croyances reflétant l'importance des critères. La fonction de croyance ainsi obtenue est notée m^Θ

Construction de la fonction de croyance associée aux alternatives

De la même façon, l'expert exprime ses préférences sur les ensembles d'alternatives selon chaque critère et non quels que soient les critères. En conséquence, afin de marquer l'influence des critères, on utilise une fonction de croyance conditionnelle.

Étant donné une matrice de comparaison par paire qui compare les ensembles d'alternatives selon un critère spécifique, on obtient les masses de croyance suivantes :

$$m^{\Omega}[C_j](A_k) = v_{kj}, \quad \forall A_k \in 2^{\Omega} \text{ et } \forall C_j \in \Theta$$

où v_{kj} est l'importance du k -ème ensemble d'alternatives selon le critère C_j et $m^{\Omega}[C_j](A_k)$ est la croyance en l'importance de A_k selon C_j

On combine les fonctions de croyances conditionnelles obtenues à la fonction de croyance issue des critères (qui est définie sur le cadre de discernement Θ) alors que les fonctions de croyance pour les alternatives sont définies sur le cadre de discernement Ω). Il faut donc procéder à une extension du cadre de discernement de Θ à $\Omega \times \Theta$:

$$m^{\Theta \uparrow \Omega \times \Theta}(B) = m^{\Theta}(C_i) \text{ avec } B = \Omega \times C_i, \quad C_i \subseteq \Theta$$

Au niveau des alternatives, E. Lefèvre utilise le déconditionnement sur un espace produit afin de transformer la fonction de croyance conditionnelle $m^{\Omega}(C_j)$ en une nouvelle fonction de croyance

$m^{\Theta}[C_j]^{\uparrow \Theta \times \Omega}$. Cette étape de déconditionnement de $m^{\Omega}(C_j)$ sur l'espace produit $\Omega \times \Theta$ est obtenue par :

$$m^{\Omega}[C_j]^{\uparrow \Omega \times \Theta}(A_k \times C_j \cup \Omega \times C_j \bar{c}_j) = m^{\Omega}[C_j](A_k), \quad \forall A_k \subseteq \Omega$$

Une fois cette opération d'unification réalisée, il devient possible de combiner l'ensemble des fonctions de croyance via la combinaison conjonctive :

$$m^{\Omega \times \Theta} = [[\cap]_{j=1, \dots, m} m^{\Omega}[C_j]^{\uparrow \Omega \times \Theta}] [\cap] m^{\Theta \uparrow \Omega \times \Theta}$$

où $m^{\Omega \times \Theta}$ représente l'importance des alternatives selon le poids des critères.

La dernière étape consiste à prendre une décision concernant la meilleure alternative en utilisant la probabilité pignistique. Pour cela, on redescend l'espace produit $\Omega \times \Theta$ sur Ω en posant :

$$m^{\Omega \times \Theta \downarrow \Omega}(A_j) = \sum_{\{B \subseteq \Omega \times \Theta \text{ tel que } Proj(B \downarrow \Omega) = A_j\}} m^{\Omega \times \Theta}(B), \quad \forall A_j \subseteq \Omega$$

On calcule alors la probabilité pignistique :

$$\text{BetP}(\omega_j) = \sum_{A_i \subseteq \Omega} \frac{|\omega_j \cap A_i|}{|A_i|} m^{\Omega \times \Theta \downarrow \Omega}(A_i) / (1 - m^{\Omega \times \Theta \downarrow \Omega}(\emptyset)), \quad \forall \omega_j \in \Omega$$

Application de la méthode AHP-classique dans un cadre simple

Nous considérons un ensemble d'alternatives formé par 4 profils qui sont pris en compte par l'analyseur $\mathcal{A}_{\text{fict}}$ selon 3 critères.

$\Omega = \{ \omega_1, \omega_2, \omega_3, \omega_4 \}$ et $\Theta = \{ C_1, C_2, C_3 \}$ avec :

C_1 = (Photo de profil anormale)

C_2 = (Taux de republication élevé, dépassant un seuil)

C_3 = (Dispersion géographique des relations élevée, dépassant un seuil)

Question Q : « Lequel des 4 profils semble, pour l'analyseur, le plus proche d'un profil fictif ? »

Tableau 1 – Importance relative entre les critères

	C_1	C_2	C_3	Moyenne géométrique	Poids
C_1	1	3	4	2,289	0,625
C_2	1/3	1	2	0,873	0,238
C_3	1/4	1/2	1	0,5	0,137

Le coefficient v_{ij} quantifie l'importance de la proposition i sur la proposition j avec $v_{ii} = 1$ et $v_{ij} = 1/v_{ji}$

Valeur propre maximale : $\lambda_{\text{max}} = 3,018$

L'indice de Saaty permet de vérifier la cohérence des coefficients de la matrice.

$CI = (\lambda_{\text{max}} - n) / (n - 1)$ où n est l'ordre de la matrice. Si $CR = CI/RI < 0,1$ la matrice est cohérente.

Calcul de l'indice de cohérence de Saaty CR : $n=3, CI = 0,009 \quad RI = 0,58 \quad CR = CI/RI = 0,015 < 0,1$

Matrice cohérente

Tableau 2 – Importance relative entre les profils en considérant le premier critère (Photo)

Critère 1 (photo)	ω_1	ω_2	ω_3	ω_4	Moyenne géométrique	Poids
ω_1	1	1/4	1/2	1/3	0,451	0,096
ω_2	4	1	4	1/2	1,681	0,358
ω_3	2	1/4	1	1/2	0,707	0,150
ω_4	3	2	2	1	1,861	0,396

$$\mu_1 = (1 \times 1/4 \times 1/2 \times 1/3)^{1/4} = 0,45 \text{ et } p_1 = \mu_1 / \sum \mu_i$$

Valeur propre maximale : $\lambda_{\max} = 4,221$

Calcul de l'Indice de cohérence de Saaty CR : $n=4$, $CI = 0,073$ $RI = 0,90$ $CR = CI/RI = 0,081 < 0,1$ Matrice cohérente.

Tableau 3 – Importance relative entre les profils en considérant le deuxième critère (republication)

Critère 2 (republication)	ω_1	ω_2	ω_3	ω_4	Moyenne géométrique	Poids
ω_1	1	1/3	1/2	1/4	0,451	0,096
ω_2	3	1	2	1/3	1,189	0,253
ω_3	2	1/2	1	1/2	0,840	0,179
ω_4	4	3	2	1	2,213	0,472

Valeur propre maximale $\lambda_{\max} = 4,124$

Calcul de l'Indice de cohérence de Saaty CR : $n=4$, $CI = 0,041$ $RI = 0,90$ $CR = CI/RI = 0,045 < 0,1$ Matrice cohérente.

Tableau 4 – Importance relative entre les profils en considérant le troisième critère (dispersion)

Critère 3 (dispersion)	ω_1	ω_2	ω_3	ω_4	Moyenne géométrique	Poids
ω_1	1	1/3	1/2	1/2	0,5372	0,118
ω_2	3	1	1/3	2	1,1892	0,262
ω_3	2	3	1	3	2,0597	0,453
ω_4	2	1/2	1/3	1	0,7598	0,167

Valeur propre maximale : $\lambda_{\max} = 4,261$

Calcul de l'Indice de cohérence de Saaty CR : $n=4$, $CI = 0,087$ $RI = 0,90$ $CR = CI/RI = 0,096 < 0,1$ Matrice cohérente.

Agrégation et calcul des priorités

La dernière phase de la méthode AHP synthétise les appréciations sous la forme d'une valeur unique.

Les priorités pour chaque profil ω_k , désignées par A_k^{AHP} , en fonction de tous les critères combinés sont déterminés selon la formule :

$$A_k^{AHP} = \sum_{j=1}^m v_{kj} p_j$$

Où les coefficients v_{kj} figure,t dans les matrices précédentes.

On obtient finalement le tableau des priorités :

Critères	C ₁ Photo	C ₂ Republica- tion	C ₃ Dispersion	Priorité
	0,625	0,238	0,137	
ω_1	0,096	0,096	0,118	0,099
ω_2	0,358	0,253	0,262	0,319
ω_3	0,150	0,179	0,453	0,198
ω_4	0,396	0,472	0,167	0,382

L'analyseur retient l'alternative ayant la plus grande priorité, c'est-à-dire ω_4 qui lui semble le plus proche d'un profil fictif.

Durant la présentation orale, nous proposerons un second exemple numérique basé cette fois sur la méthode AHP intégrant les fonctions de croyances.

4 Application à l'ADFI de comptes tests « Sentinelles 2022 »

Nous appliquons la théorie de Dempster-Shafer à une architecture composée de 500 comptes fictifs créés sur les réseaux sociaux (Facebook, LinkedIn, Twitter). L'objectif est de pouvoir ajuster certains paramètres de l'architecture fictive afin de maximiser sa durée d'activation, son attractivité et sa crédibilité. Les qualités de l'ADFI s'expriment à l'aide des fonctions de croyance et de plausibilité de chaque profil fictif puis par application de la règle de combinaison de Dempster-Shafer à la structure complète. Enfin, nous définissons de nouvelles fonctions d'évaluation du caractère immersif de l'ADFI.

Le projet Sentinelles 2022

L'enjeu principal du projet Sentinelles 2022 est d'offrir un outil d'alerte en temps réel sur l'activité de bots étrangers producteurs de messages politique sur les réseaux sociaux destinés à perturber et à influencer l'élection présidentielle de 2022. Imitant un projet canadien similaire, SENTINELLES 2022 est une architecture de 500 comptes tests créés sur les réseaux sociaux afin de mesurer la variation de production de messages à caractère politique en phase pré-électorale. Le dispositif agit comme un sonar passif détectant les éventuelles campagnes de bots étrangers en amont de l'élection présidentielle. C'est un outil d'alerte.

Le dispositif SENTINELLES 2022 est adaptable à d'autres cas d'usages que l'élection présidentielle de 2022. Les fonctionnalités de sonar passif peuvent s'appliquer à de la détection de montée en puissance de risques sociaux.

Références

1. From digital traces to algorithmic projections, 1st Edition, Thierry Berthier, Bruno Teboul, ISTE Wiley, Elsevier, September 2018.
2. Cybereason – Honeypot & ransomware, juin 2020
<https://www.cybereason.com/blog/cybereason-honeypot-multistage-ransomware>
3. Un état de l'art sur les fonctions de croyance appliquées au traitement de l'information ; Patrick Vannoorenbergh, 2002
https://www.irit.fr/journal-i3/volume03/numero02/article03_02_01.pdf
4. Eric Lefèvre, « Fonctions de croyance : de la théorie à la pratique », rapport de synthèse HDR, 2012, https://www.lgi2a.univ-artois.fr/spip/IMG/pdf/hdr_eric_lefevre.pdf