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

Thierry Berthier[1], Olivier Kempf, Eric Hazane[2], Thomas Anglade[3]

[1] Université limoges - IUT dpt informatique - thier.berthier@orange.fr
[2] HUB IA France - eric.hazane@protonmail.com
[3] ITrust - tanglade@itrust.fr

Résumé

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'honeypot dédié au leurrage (expérimentation Cybereason, 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).

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