Attaques par HoaxCrash et par Faux Ordres de Virement : la puissance du leurre cognitif

Thierry Berthier[1]
[1] Chaire de cyberdéfense & cybersécurité Saint-Cyr, France

Abstract. L'hyperconnexion des systèmes et l'accélération des échanges d'information réalisés en haute fréquence induisent de nouvelles vulnérabilités liées à l'utilisation de structures de données fictives ou fausses données comme vecteur d'entrée d'une attaque. Ces fausses données sont de plus en plus souvent employées pour tromper un opérateur humain et l'amener à déclencher un protocole de virement sur un compte frauduleux. Les attaques de type FOVI (Faux ordre de virement), fraude au Président et changement de RIB se sont multipliées depuis 2010, impactant de nombreuses entreprises. Le préjudice global en France dépasse les 485 millions d'euros avec des centaines de PME, PMI piégées et plus de 2300 plaintes déposées durant ces cinq dernières années. La détection automatisée des structures de données fictives constitue alors un enjeu majeur de la cybersécurité. Après avoir décrit les mécanismes de ces attaques construites sur de fausses données, nous explorons quelques pistes pour la détection des cyberattaques de type Fovi et HoaxCrash, en nous appuyant sur des techniques d'analyse sémantique et de cartographie des scénarii de hoax.

Introduction - Facteur humain et leurre cognitif : les clés des attaques par HoaxCrash et FOVI

Le facteur humain est souvent évoqué pour désigner le maillon faible de la chaîne de sécurité d'un système d'information. L'exploitation des vulnérabilités humaines installe et entretient la menace. Les biais cognitifs, la crédulité, parfois la naïveté et toutes les petites négligences dont chacun fait preuve à un instant ou à un autre de sa pratique numérique, sont autant de vulnérabilités biologiques que l'attaquant sait exploiter. Lorsque l'on parle d'ingénierie sociale, on sous-entend que cet attaquant met en œuvre une stratégie dont l'objectif est d'obtenir des informations sur le système qu'il souhaite cibler. Le facteur humain constitue alors souvent le premier mécanisme qu'il doit actionner pour obtenir les données clés d'une future intrusion. Parfois au contraire, l'attaquant n'attend pas de retour d'information pour mener son opération mais se limite à diffuser un ensemble plus ou moins sophistiqué des fausses données crédibles et cohérentes. Leur prise en compte comme corpus d'information valide engendre ensuite une série d'actions souhaitées par l'attaquant qui saura en tirer profit. Ce mécanisme de leurre cognitif peut provoquer de fortes turbulences sur un environnement hyperconnecté et saturé d'informations. On le retrouve dans les attaques de type FOVI (Faux Ordres de Virement), arnaques au Président, changement de RIB et dans les attaques par HoaxCrash dont l'impact immédiat et scalable peut être mondial.

L'accélération de la diffusion de l'information concerne aujourd'hui l'ensemble des secteurs d'activités humaines, en particulier l'économie et la finance. En apportant de la fluidité et de la réactivité dans les échanges, cette accélération, de nature systémique, a ouvert de nouveaux espaces d'interaction en mode "haute fréquence" et, corrélativement, a créé de nouvelles vulnérabilités. Si les opérations d'influence et de désinformation ont toujours accompagné l'histoire de la communication depuis l'Antiquité, elles prospèrent désormais sur les infrastructures numériques et se nourrissent du déluge de données. Le canular, pratiqué par les grecs et les romains, a traversé deux millénaires pour devenir un puissant outil d'influence et de manipulation. Lorsqu'il est utilisé pour déstabiliser le cours d'une action en créant une volatilité artificielle sur le titre, on parle alors de "HoaxCrash" (Hoax pour canular et Crash pour le Flash Crash boursier qui en résulte). De telles cyber-opérations méritent une attention particulière car les turbulences qu'elles engendrent sont souvent très violentes et coûteuses pour les victimes de l'attaque. A partir du HoaxCrash qui a ciblé le groupe Vinci le 22 novembre 2016, notre étude passe en revue les cas antérieurs puis réalise une analyse de ce type de cyberattaque. Elle montre en particulier qu'un HoaxCrash peut se définir formellement par la donnée de trois paramètres : sa durée de validité, son efficacité et sa puissance.

Les fausses données sont également de plus en plus souvent employées pour tromper un opérateur humain et l'amener à effectuer un virement sur un compte frauduleux. Les attaques de type FOVI (Faux ordre de virement), fraude au Président et changement de RIB se sont multipliées depuis 2010, impactant de nombreuses entreprises. Le préjudice global en France dépasse les 485 millions d'euros avec des centaines de PME, PMI piégées et plus de 2300 plaintes déposées durant ces cinq dernières années. La détection automatisée des structures de données fictives constitue un enjeu majeur de la cybersécurité. L'article explore plusieurs pistes pour lutter algorithmiquement contre les HoaxCrash et les attaques FOVI.

1 Les attaques par HoaxCrash

1.1 Le cas du HoaxCrash Vinci

Le mardi 22 novembre 2016 vers 16 h 05, le groupe Vinci est la cible d'une opération visant à déstabiliser le cours de son action en s'appuyant sur la publication d'un faux communiqué de presse "officiel" transmis aux opérateurs boursiers. Ce faux message diffusé par l'agence de presse spécialisée Bloomberg fait état d'une alerte de révision des comptes consolidés 2015 et du premier semestre 2016 ainsi que du renvoi du directeur financier de Vinci, après la découverte d'erreurs comptables portant sur plus de 3,5 milliards d'euros. Le titre Vinci chute alors presque instantanément de 18,28 % (Fig.1) , variant de 61,81 euros à 49,93 euros en quelques minutes. La valorisation du groupe passe d'environ 36 milliards d'euros le mardi 22 novembre au matin à 29 milliards au plus fort du dévissage, ce qui représente une perte momentanée de plus de 7 milliards d'euros.

Le premier démenti de Vinci intervient (par téléphone) à 16h10, soit cinq minutes seulement après la première publication du Hoax par Bloomberg. Le second démenti officiel est publié par Vinci à 16 h 49 sur son site internet :

«Un faux communiqué de presse Vinci a été publié par Bloomberg le 22 novembre à 16 h 05. Vinci dément formellement l'ensemble des «informations» figurant dans ce faux communiqué et étudie toutes les actions judiciaires à donner suite à cette publication».

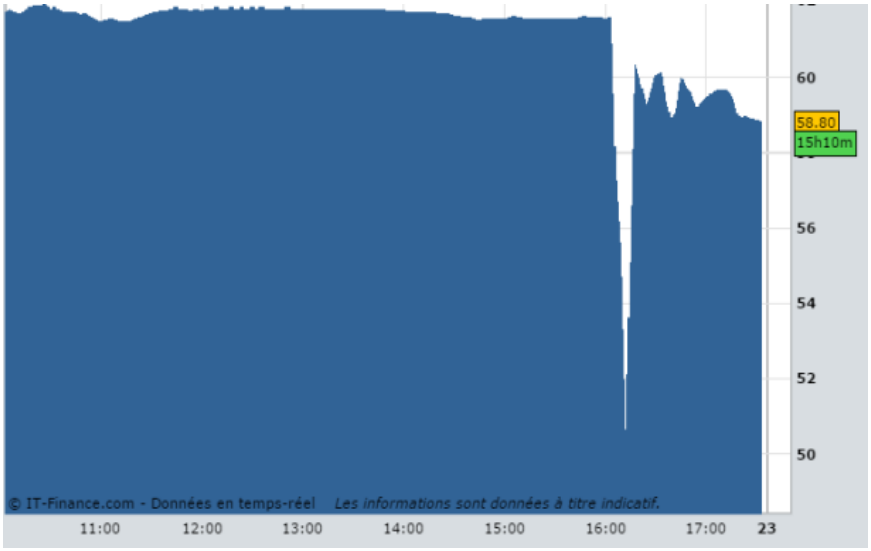

Fig.1 - Profil du Flash Crash Vinci du 22 novembre 2016

L'attaquant publie son propre "faux démenti" à 16h27 dans le but de créer et d'exploiter une volatilité à la hausse sur l'action. Le titre Vinci termine la séance de cotation le mardi soir à 58,80 euros, en repli de 3,76 %, pour un volume total d'échanges atteignant cinq fois celui réalisé en moyenne dans un contexte "normal". Vinci dépose plainte contre X le jour même de l'attaque et saisit l'Autorité des Marchés Financiers (AMF). Celle-ci devra identifier précisément l'identité des opérateurs (humains ou robots HFT) qui ont exploité cette déstabilisation, c'est-à-dire ceux qui ont été présents et actifs au bon moment, au bon endroit, durant l'opération. On notera enfin qu'un message de revendication de l'opération de déstabilisation du titre Vinci a été envoyé durant l'attaque le 22/11 et diffusé à la presse le 23 novembre : *«Action mediatico-boursière contre Vinci : revendication - la forêt de notre dame des landes a elle même sentie le béton reculer et ces occupants ont fêté se nouveau coup porté directement dans la bourse de ce monstre de béton».* Il convient de considérer ce message avec toutes les précautions d'usage, compte tenu notamment des capacités opérationnelles des Zadistes, opposants au futur aéroport de Notre-Dame-des-Landes. Il est fort possible et probable que cette revendication relève elle

aussi du canular opportuniste. L'opération semble avoir été construite pour créer une forte volatilité de manière artificielle sur le titre Vinci puis exploiter ses variations pour en tirer bénéfice. La chronologie du HoaxCrash montre que la période utile de l'attaque ne dépasse pas les 8 minutes après publication du faux message.

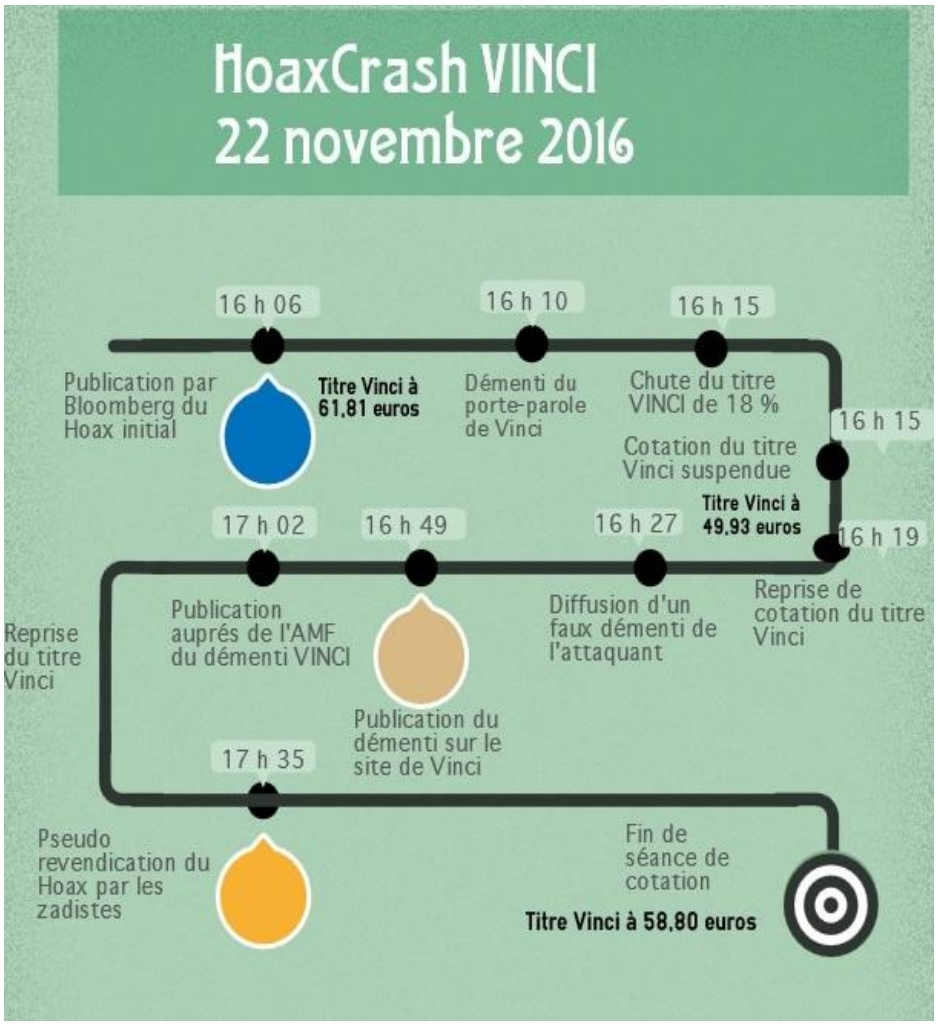

Fig.2 - Chronologie de l'attaque par HoaxCrash VINCI

On notera que le flash crash provoqué sur le titre Vinci n'a pas eu de répercussions sur d'autres valeurs boursières. Cette absence de contagion au marché est essentiellement due à la nature très ciblée du Hoax et au démenti rapidement publié par l'attaquant, puis par Vinci. La séquence de publication du Hoax puis, quelques minutes plus tard, du démenti démontre que l'attaquant ne recherchait que la volatilité sur le titre ciblé et le flash crash avec un unique objectif de profit spéculatif à court

terme. Cette séquence "Hoax-démenti" écarte également l'objectif d'un simple effondrement (sans remontée) de l'action Vinci qui aurait été celui d'un groupe d'activistes affichant des motivations écologistes ou politiques.

Le service de communication du groupe Vinci a réagi très rapidement, dès la quatrième minute de l'attaque, en produisant des démentis par téléphone auprès des agences de presse puis en publiant un communiqué officiel sur son site, une quarantaine de minutes plus tard. Au regard de cette chronologie de gestion de crise, il semble difficile de réduire encore le temps de réaction humain face à une tentative de HoaxCrash. Seule une réponse algorithmique, en haute fréquence et à large spectre de diffusion, permettrait de stopper le processus dans sa propre temporalité, avant qu'il ne produise son effet sur les marchés.

Le HoaxCrash Vinci doit plus généralement nous interroger sur la responsabilité des médias, de la presse spécialisée, des portails d'information financière et des réseaux sociaux dans la validation et la diffusion de l'information. La sensibilité et la réactivité des audiences face à une donnée numérique, vérifiée ou non, semblent intimement liées à la vulnérabilité du système quand celui-ci est confronté à des opérations de désinformation et d'influence parfois très rudimentaires dans leur conception. Comment définir une chaîne de responsabilité qui tienne compte à la fois des temporalités "haute fréquence" et de l'impact d'une fausse information sur des systèmes interconnectés hypersensibles ? C'est finalement la question de la résilience ou de l'antifragilité [1] définie par Nassim Nicholas Taleb qui se pose dans les mécanismes du HoaxCrash.

1.2 Les mécanismes du HoaxCrash

Un HoaxCrash débute toujours par la publication d'un "Hoax" ou canular réalisé sous usurpation d'identité d'une autorité. Cette fausse information, considérée comme légitime par les opérateurs boursiers provoque alors un flash crash de l'action de l'entreprise ciblée. Le mode opératoire du HoaxCrash s'articule selon la séquence suivante : l'attaquant commence par usurper l'identité d'un membre de la direction d'un groupe industriel X coté en bourse. Il rédige un message d'alerte "crédible" imitant le mieux possible le canal de communication de l'entreprise. Cette alerte, signée du nom de l'un de ses dirigeants, révèle des difficultés financières ou des malversations dans l'activité du groupe. Elle est immédiatement diffusée auprès d'agences de presse spécialisées comme Bloomberg et des principaux opérateurs boursiers. Si le message ne suscite pas de doute, l'effet est immédiat : le cours de l'action X dévisse brusquement. Quelques minutes plus tard, un second communiqué émis par l'attaquant vient démentir la première alerte, ce qui provoque la remontée immédiate du titre. Cette volatilité contrôlée dans le temps est alors exploitée par un deuxième acteur, complice de l'attaquant ou commanditaire du HoaxCrash, qui profite des variations artificielles engendrées par la prise en compte par le marché de la fausse alerte. L'avantage informationnel et temporel, créé de toute pièce par l'attaquant, lui permet de réaliser de très gros profits sur les fluctuations du titre, qu'il connait à l'avance. On évoque enfin souvent le trading haute fréquence (HFT) comme troisième acteur de l'opération puisque capable d'amplifier et de tirer bénéfice du flash crash né de la fausse information. Ce type d'opération de déstabilisation d'un titre

boursier ne commence pas avec l'affaire Vinci. En avril 2013, l'Armée Electronique Syrienne (SEA) prenait le contrôle du compte Twitter officiel de l'agence Associated Press. Un faux message était alors publié sur le compte piraté indiquant qu'une explosion avait eu lieu à la Maison Blanche et que Barack Obama était blessé. Le faux tweet provoquait le dévissage immédiat de toutes les bourses occidentales sous la forme d'un flash crash à Wall Street à hauteur de 136 milliards de dollars et la chute de l'indice boursier Standard&Poor's 500 de 145 points en trois minutes [2]. Certains analystes ont rapidement attribué le violence de ce flash crash au trading haute fréquence mais une enquête approfondie a montré que les ordres ont surtout été passés par des opérateurs humains pris de panique [3]...

Fig.3 - Faux tweet de la SEA usurpant l'identité de l'agence Associated Press

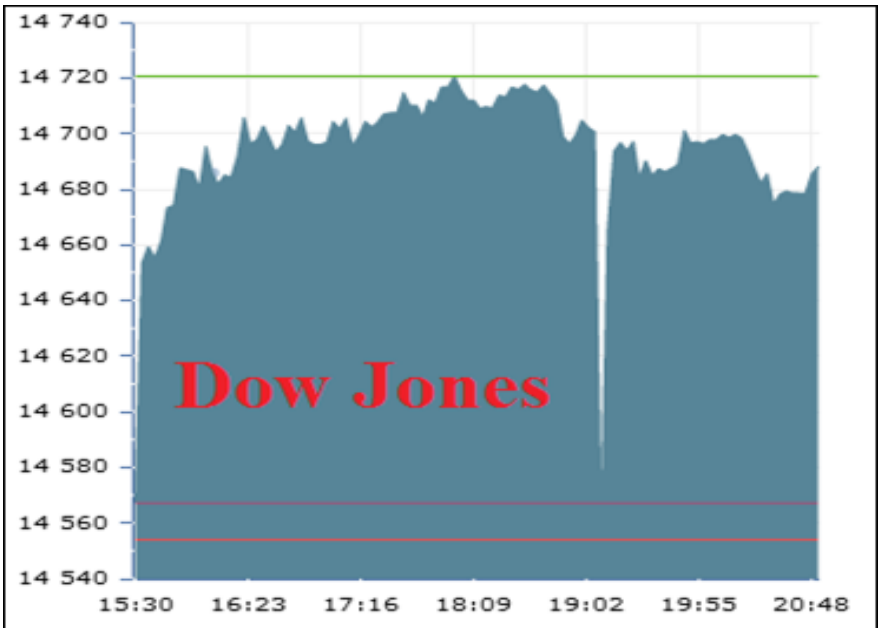

Fig 4 - Flash Crash du Dow Jones provoqué par le faux tweet de la SEA

En 2013, c'est le groupe australien Whitehaven Coal qui subit une attaque provenant d'un groupe de militants écologistes. Ces derniers usurpent l'identité d'un des responsables du groupe et envoient un faux communiqué de presse qui annonce la non rentabilité d'un des sites d'exploitation minière de Whitehaven Coal. Le cours de bourse de l'action s'effondre alors temporairement.

En 2014, la société de sécurité G4S assurant la surveillance des camps de réfugiés et migrants subit le même type d'attaque. Un faux communiqué de presse établi par usurpation d'identité annonce que la société rencontre des difficultés, qu'elle procède à une révision des comptes depuis 2013 et que son directeur financier a été démis de ses fonctions. Le cours de l'action G4S dévisse dès la publication de cette fausse alerte.

En 2015, le groupe américain Avon (Fig 7) est à son tour ciblé par la publication d'un faux message annonçant une offre publique d'achat (OPA) qui déstabilise le cours de son titre et oblige Avon à démentir l'OPA dans l'urgence.

Fig. 5 - Perturbation du titre Avon en 2015

En novembre 2016, la société américaine Fitbit voit le cours de son action fortement perturbé par la publication d'un faux communiqué annonçant une offre de rachat par le fond d'investissement chinois ABM Capital.

Fig. 6 - Perturbation du cours de l'action Fitbit

Les motivations des auteurs de HoaxCrash sont variées, comme le montre le tableau suivant.

HoaxCrash	**Motivation(s) de l'attaquant**
SEA - AP (2013)	Politique - Hacktivisme (conflit syrien)
Whitehaven Coal (2013)	Politique - activisme d'un groupe d'écologistes
G4S (2014)	Politique - activisme
AVON (2015)	Economique - (dégradation d'image - spéculation)
FITBIT (2016)	Economique et activisme
VINCI (2016)	Economique (volatilité - spéculation)

Le cours de la cryptomonnaie Bitcoin a été ciblé lors de plusieurs attaques par HoaxCrash. La dernière date du 11 septembre 2017. Une fausse information affirmant

que les autorités chinoises s'apprêtaient à interdire les plateformes d'échanges Bitcoin sur le territoire, a été relayée par la presse américaine. La fausse nouvelle a provoqué une chute immédiate de la valeur de la cryptomonnaie puis une remontée après démenti.

1.3 Efficacité et puissance d'un Hoaxcrash

Il est possible de définir l'efficacité et la puissance d'un HoaxCrash construit sur la publication d'un message "m" contenant une fausse information relative à l'activité d'un groupe industriel coté en bourse.

Efficacité d'un HoaxCrash :

L'efficacité d'un HoaxCrash est obtenue en divisant le gain net G(m) obtenu par l'attaquant par la complexité algorithmique du message "m" et du corpus informationnel S(m) qu'il a mis en place pour mener son attaque.

$$E(m) = G(m, S(m)) / K(m, S(m))$$

G(m, S(m)) désigne le gain (net) obtenu par l'attaquant après déroulement de la séquence suivante :

1 - Création éventuelle d'un support de publication S(m) imitant le support de communication officiel de la cible ou prise de contrôle (hacking) d'un support de communication légitime.

2 - Publication du message m (le Hoax) sur le support S(m).

3 - Exploitation des variations du cours de l'action ciblée et passages d'ordres pendant la durée effective du HoaxCrash.

4 - Réalisation du gain G(m, S(m)) par l'attaquant.

K(m,S(m)) désigne la complexité algorithmique (complexité de Kolmogorov) du message publié et de son support de publication créé spécifiquement pour l'opération.

Puissance d'un HoaxCrash sur une action :

La puissance d'un HoaxCrash est obtenue en divisant la valeur totale des variations du cours de l'action A ciblée (évaluée une fois le flash crash terminé après publication du démenti officiel) par la complexité déjà citée.

$$P(m) = V(A, T(m)) / K(m, S(m))$$

A désigne l'action ciblée, T(m) la durée de validité du Hoax avant publication du démenti officiel et V(A, T(m)) la valeur totale des variations du cours de l'action A évaluée après publication du démenti officiel, une fois le flash crash terminé.

Un HoaxCrash H(m) est alors quantitativement et qualitativement déterminé par la donnée du triplet $H(m) = \{ T(m), P(m), E(m) \}$.

La morphologie des attaques de type HoaxCrash risque d'évoluer en se complexifiant fortement. Les techniques d'imitation permettent déjà de reproduire certains sites web à l'identique en conservant une adresse apparente semblable à l'adresse officielle ciblée. C'est l'infrastructure informationnelle utilisée en amont pour rendre crédible le faux message qui va nécessiter les efforts les plus importants de la part de l'attaquant. Les plateformes de validation et de détection des HoaxCrash seront capables, grâce à l'intelligence artificielle, de détecter en temps réel les faux messages les plus rudimentaires. Elles produiront alors une alerte qui évitera le flash crash.

2 Les attaques par FOVI (Faux ordres de virements) et arnaques au Président

2.1 FOVI, arnaques au Président, des attaques ciblées et lucratives

La "**Fraude au président**" consiste pour des escrocs à convaincre le collaborateur d'une entreprise d'effectuer en urgence un virement important à un tiers pour obéir à un prétendu ordre du dirigeant, sous prétexte d'une dette à régler, de provision de contrat ou autre.

Le "**Changement de RIB**" consiste pour les fraudeurs à envoyer un mail à un salarié du service de comptabilité ou trésorerie de l'entreprise en se faisant passer pour un fournisseur, et lui demander de diriger ses versements vers un autre compte bancaire appartenant aux escrocs.

Réalisée par téléphone ou par mail, l'escroquerie aux Faux Ordres de Virement (FOVI) concerne les entreprises de toute taille et de tous les secteurs. Souvent situés à l'étranger, les escrocs collectent en amont un maximum de renseignements sur l'entreprise. Cette connaissance de l'entreprise associée à un ton persuasif et convaincant est la clé de réussite de l'arnaque. L'opération est alors lancée sur les personnes capables d'opérer les virements (services comptables, trésorerie, secrétariat...).

Apparues en France dès 2010, les attaques FOVI (Faux Ordres de Virement) ou "arnaques au Président" font partie des attaques ciblées nécessitant une bonne connaissance de l'entreprise visée, de son personnel et de sa direction. La phase initiale d'ingénierie sociale est ainsi primordiale dans le bon déroulement de cette escroquerie numérique. Elle permet de recueillir des informations sur l'activité commerciale de l'entreprise, sur sa production, sur ses fournisseurs, sur ses clients et sur son organigramme. Une fois cette collecte réalisée, l'attaquant envoie des messages électroniques et des appels téléphoniques en demandant un virement en urgence d'une importante somme d'argent sur un compte international. En général, il dénonce le retard ou le non paiement d'une facture ou d'une prestation en usurpant l'identité d'une autorité ou d'un tiers de confiance. Le message est presque toujours "confidentiel et urgent". Les principales cibles visées par ces envois sont les assistantes de direction, les secrétaires comptables et les services de comptabilité des entreprises, PME et PMI. Un important travail de sensibilisation aux tentatives de FOVI s'avère nécessaire auprès des personnels exposés. Cela dit, cette sensibilisation ne peut être suffisante pour réduire le risque et doit nécessairement être couplée avec des solutions de détection automatisée.

2.2 Les chiffres des arnaques au Président

Depuis 2010, les escroqueries de "fraude au président" ou de "changement de RIB" ont fait de nombreuses victimes parmi les entreprises françaises. Plusieurs centaines de faits ou de tentatives de FOVI ont été recensées pour un préjudice global de 485 millions d'euros. L'Office central de répression de la grande délinquance financière (OCRGDF), appelle les sociétés à la vigilance : **"***C'est un véritable fléau économique. Il faut être vigilant, la trêve des confiseurs est souvent synonyme de*

relâchement dans les sociétés et les escrocs en profitent." Durant les cinq dernières années, 2.300 plaintes ont été déposées. Cela dit, de nombreuses entreprises n'osent pas déposer plainte par crainte de mauvaise publicité et d'atteinte à leur image.

L'Association Nationale des Directeurs Financiers et de Contrôle de Gestion (DFCG) a produit un baromètre de la fraude en entreprise en collaboration avec Euler Hermes [4], leader de l'assurance-crédit. Publiée fin 2016, cette enquête a révélé que 93% des entreprises françaises ont déclaré avoir été victimes d'au moins une tentative de fraude en 2015 et qu'une fois sur deux, elles ont été confrontées à une arnaque au Président. Ces attaques ont eu lieu le plus souvent durant les périodes de fêtes, lorsque les effectifs sont réduits et lorsque la vigilance se relâche. L'augmentation de leur fréquence est sensible puisqu'en 2014, seules 77 % des entreprises ont été ciblées... Plus préoccupant en terme de préjudices, une fois sur trois, la tentative d'attaque n'est pas déjouée, la victime n'est pas en mesure de détecter la fraude et le virement est effectué. Le dénominateur commun de ces attaques reste l'usurpation d'identité initiale qui a pour objectif d'installer la confiance chez l'employé de la société ciblée. Chaque type d'autorité peut être invoquée par l'attaquant dans le message frauduleux afin de créer un climat de confiance et d'urgence : faux avocat, faux banquier, fournisseurs, commissaire aux comptes et faux dirigeant de l'entreprise dans le cadre d'une arnaque au président.

L'attaquant doit convaincre sa cible, par téléphone ou par courriel, d'effectuer un virement sur un compte donné ou de modifier les coordonnées bancaires du destinataire du règlement (attaque par changement de RIB). Sa capacité de persuasion est donc déterminante dans le bon déroulement de l'opération. Pour cette raison, il effectue toujours une recherche préliminaire sur l'entreprise, sur son activité et ses clients. Il cible ensuite les personnels qui ont la capacité d'effectuer des paiements et des opérations bancaires (salarié d'un service financier, agent comptable, assistante ou assistant du dirigeant pour les petites sociétés).

Le DFCG a produit une liste de recommandations de bon sens qui permettent de limiter le risque. Certaines d'entre elles sont "automatisables" :

- Sensibiliser et inciter les salariés à ne pas communiquer sur les réseaux sociaux des informations relatives au fonctionnement de leur entreprise et de ses clients.
- Sensibiliser les personnels en contact téléphonique avec l'extérieur (standard téléphonique, responsable des achats, comptabilité).
- Mettre en place des vérifications systématiques et des signatures multiples pour les paiements internationaux.
- Imposer une demande de confirmation avant toute opération d'envoi de fonds non planifiée.
- En cas de requête effectuée par courriel, envoyer un mail à l'adresse habituelle du donneur d'ordre au lieu de répondre simplement à la requête initiale.
- Renforcer la vigilance en période de fêtes et de vacances lorsque l'effectif est réduit.

Enfin, lorsque l'employé ciblé effectue le virement sans avoir détecté l'arnaque, il faut alors alerter la banque le plus rapidement possible et déposer plainte auprès d'un service de police.

3 Détection automatisée des HoaxCrash et des FOVI

3.1 Les approches possibles dans la lutte contre les HoaxCrash et FOVI

La lutte contre la diffusion de fausses informations sur les réseaux sociaux rejoint les priorités de cybersécurité des grands acteurs du numérique. Ainsi, Google et Facebook viennent d'annoncer le développement d'outils spécifiques pour détecter le faux et certifier le vrai. Facebook souhaite par ailleurs interdire les publicités sur les pages colportant de fausses informations. Google déréférencera les sites contenant des Hoax destinés à tromper ou orienter le public. La prise en compte de la fausse donnée semble se généraliser notamment à la suite de l'élection de Donald Trump. Cela dit, la détection en amont d'une fausse information reste techniquement complexe.

Lutter contre le HoaxCrash consiste à mettre en place des mesures de détection des faux messages selon une temporalité "haute fréquence". Même lorsque le démenti est diffusé seulement dix minutes après la publication initiale du hoax, le "mal est fait" puisque ces dix minutes suffisent largement à perturber et à influencer les cours boursiers. Il est donc nécessaire d'aligner la fréquence de lutte contre le HoaxCrash sur celle de l'opération elle-même, en développant des agents logiciels actifs en temps réel, capables de mesurer la véracité instantanée d'un message et d'évaluer son impact potentiel sur le marché en cas de diffusion frauduleuse. A l'image des détecteurs de spam, l'agent évalue la véracité d'un message en lui attribuant une probabilité de véracité instantanée en fonction de sa forme, de son contenu et du contexte extérieur. Cette mesure de véracité peut alors se construire selon deux approches.

La première approche repose sur la mise en place d'un réseau "d'agents correspondants" couvrant l'ensemble des acteurs du marché boursier et des entreprises cotées. L'agent évaluateur d'un message doit pouvoir interroger un agent "correspondant" au sein de l'entreprise ou de l'administration figurant dans le message, afin d'obtenir la validation ou la répudiation du message. La mise en place d'un tel réseau d'agents communiquant spécifiquement sur la véracité des messages à impact boursier, permettrait d'optimiser le temps de réaction des systèmes et de produire une publication automatisée de démenti en cas de HoaxCrash. La construction d'un réseau d'agents "correspondants" pourrait également s'appuyer sur une architecture de type Blockchain afin d'éviter une supervision centralisée, souvent coûteuse et potentiellement vulnérable.

La seconde approche intervient indépendamment ou en l'absence de toute confirmation ou réfutation de l'agent correspondant "légitime". L'agent évaluateur se contente du message, de son contenu, de ses métadonnées et du contexte extérieur pour lui attribuer une valeur de véracité qui, en dessous d'un certain seuil, déclenche l'alerte et la réfutation du message sur l'ensemble du réseau. Dans le cadre du HoaxCrash Vinci, le message initial présentait des particularités qui auraient été facilement détectables par un agent logiciel : le numéro de téléphone de l'attaché de presse de Vinci figurant en fin de message était erroné et l'adresse du site publiant le message apparaissait en "vinci.group" alors que l'adresse légitime du groupe Vinci est "vinci.com". Parfois, ce sont les fautes d'orthographe ou de style grammatical du texte qui doivent alerter. Dans le cas du HoaxCrash Vinci, ces éléments suffisaient à mettre

en doute la véracité du message mais la précipitation qui règne au sein des agences de presse spécialisées en finance (Bloomberg) et le facteur humain ont eu raison du principe de précaution minimal (c'est ce qui rend nécessaire l'évaluation des messages par un réseau d'agents logiciels). L'utilisation de solutions à base d'apprentissage automatisé peut également renforcer la détection de faux communiqués en se référant cette fois à un historique des communiqués légitimes des groupes concernés. C'est en croisant et en combinant plusieurs méthodes algorithmiques qu'il devient possible d'évaluer la véracité du message avec précision.

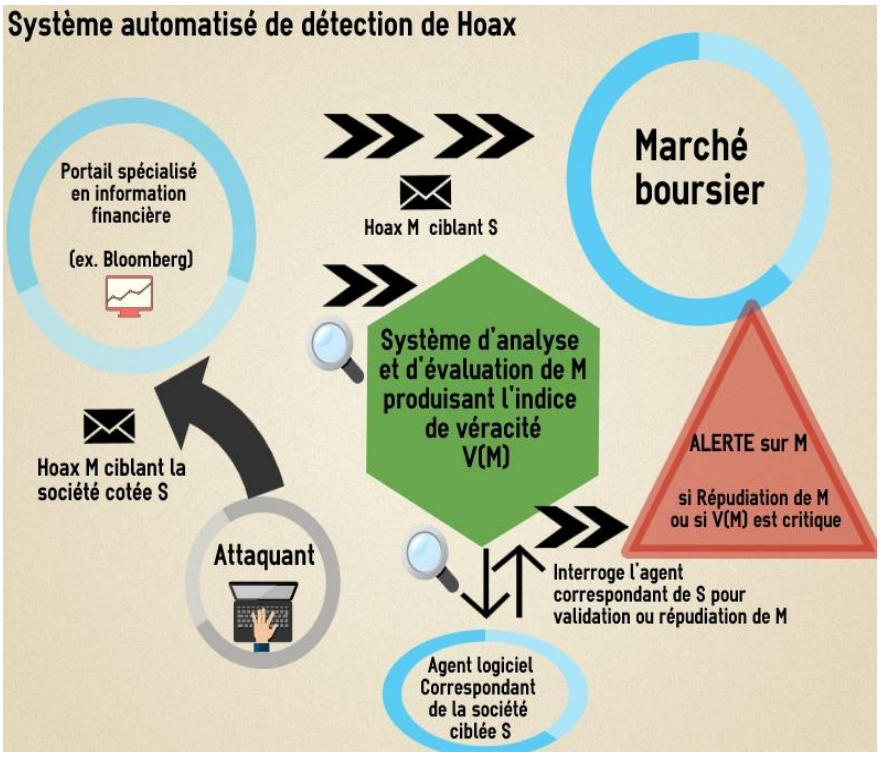

Fig.7 - Schémas de détection automatisée des attaques par HoaxCrash

3.2 Développement de la plateforme ALETHEIA

La plateforme ALETHEIA contient deux solutions de lutte contre les attaques de type HoaxCrash :

- HoaxDetect dédiée à la détection des messages de Hoax potentiellement reçus par les agences de presse (Bloomberg, Associated Press, Reuters, AFP,...).
- HoaxBlock, architecture blockchain sécurisant la diffusion des communiqués de presse par les services de communication des groupes cotés auprès des agences

spécialisées. Dérivant de la technologie HoaxDetect, la solution FOVIDetect est destinée à la détection des arnaques au président réalisées par envoi de courriels malveillants.

Ces trois solutions sont en cours de développement dans le cadre de la création de la startup ALETHEION. La description des trois produits reste générale en raison de la clause de confidentialité associée au développement.

HoaxDetect - Compte-tenu de la faible quantité de données d'apprentissage (messages utilisés lors des HaoxCrash) , il n'a pas été possible de créer un système d'apprentissage automatique efficace. Seul un moteur de règles, à large spectre d'application, pouvait répondre au problème de détection sans remonter trop de "faux positifs". HoaxDetect est donc construit sur un premier moteur de règles scalable, capable de s'adapter aux volumétries des serveurs de messagerie d'une agence de presse internationale et d'agir en moins de 4 minutes en cas d'alerte (on considère que l'attaque par HoaxCrash se déroule sur une durée moyenne de 7 minutes). Le détecteur produit en sortie des alertes en cas de forte probabilité de tentative de HoaxCrash sous la forme d'une hauteur de Hoax. Il produit également une hauteur d'impact qui renseigne sur l'impact potentiel de l'information en cas de prise en compte du message par le destinataire. Ces deux hauteurs sont calculées en tenant compte à la fois des métadonnées du message et de son contenu, via une analyse sémantique classique (NLP, dictionnaires [5]).

HoaxBlock - Il s'agit d'une architecture Blockchain dédiée à la diffusion des communiqués de presse (mail, fichiers word, pdf,...) par les services de communication des sociétés cotées en bourse auprès des agences de presse. La répartition des nœuds en trois tiers renforce la robustesse de la chaine tout en incluant une possibilité d'authentification forte du diffuseur du communiqué au moment de la connexion. HoaxBlock répond ,entre autres, aux demandes de sécurisation des communiqués de presse, formulées par l'AMF (Autorité des Marchés Financiers) en février 2017, après la seconde attaque par HoaxCrash ciblant le groupe VINCI.

FOVIDetect - Cette solution est dédiée à la détection automatique des faux ordres de virement, et d'autres arnaques au Président Le développement de FOVIDetect s'effectue comme sous-produit du détecteur HoaxDetect. On notera que le moteur de règles construit sur de l'analyse sémantique "à l'ancienne" produit peu de faux positifs contrairement aux systèmes d'apprentissage automatisés. Pour autant, il est prévu de compléter l'architecture FOVIDetect par un second système fonctionnant cette fois par apprentissage machine. L'hybridation des deux techniques peut apporter en adaptativité et en facilité de maintenance de l'ensemble.

A noter : Le projet ALETHEION est actuellement inscrit sur deux Topics "Cybersécurité" du programme européen Horizon 2020 (H2020). Il est également référencé par le nouveau Hub National FranceIA.

Pour conclure...

Les attaques par diffusion de leurres cognitifs montent en puissance. elle coûtent cher à l'économie française en produisant des préjudices qui peuvent mettre en danger la survie même d'une PME. Dans un avenir proche, il faut s'attendre à une complexification des structures de données fictives sur lesquelles l'attaquant va s'appuyer pour mener son opération. La détection automatique du faux, en haute fréquence sera la seule réponse pertinente, accompagnée d'une sensibilisation des personnels aux biais et aux leurres cognitifs.

References

1. Taleb, NN. "*Antifragile - Les bienfaits du désordre*" , 2013, Ed. Les Belles Lettres
2. Teboul, B., Berthier, T.,- "*Valeur et Véracité de la donnée: enjeux pour l'entreprise et défis pour le data scientist*", Hal, 19 mai 2015, Actes du colloque "La donnée n'est pas donnée" Ecole Militaire, 23 mars 2015.
3. Berthier, T., - "*Sur la valeur d'une donnée*", Mai 204, Art. IV.3, Publications de la Chaire de Cyberdéfense & Cybersécurité Saint-Cyr.
4. Etude Euler Hermes / DFCG 2017, 3eme édition, "De la cybercriminalité à la fraude : une menace en pleine mutation", mai 2017
5. Bird, S., Klein, E., Loper, E,. Natural Language Processing with Python, O'Reilly, 2009