Cyberconflictualité, hacking d'influence et prévisibilité des attaques

Thierry Berthier

Chaire de cybersécurité & cyberdéfense Saint-Cyr / Sogeti / Thales

Abstract. Les guerres, les révoltes et d'une façon générale les conflits humains se projettent systématiquement sur le cyberespace. Le hacking d'influence occupe désormais une position prépondérante dans l'arsenal des cellules activistes ou des structures étatiques. L'asymétrie militaire qui hiérarchise certains conflits physiques n'a plus de sens sur le cyberespace, elle fait place à des concurrences algorithmiques dont la complexité évolue rapidement. Pourtant, certaines cyberagressions obéissent à des mécanismes qui deviennent prévisibles, moyennant une veille exhaustive et une analyse adaptée. Le développement d'infrastructures de prévisions des attaques peut alors contribuer à renforcer les protections algorithmiques et physiques conventionnelles.

Keywords: cyberconflictualité ; cyberdéfense ; hacking d'influence ; prévisions ; concurrences ; duels algorithmiques.

1 Cyberconflictualité

Les problématiques liées à la cyberconflictualité occupent un espace de réflexion considéré aujourd'hui comme hautement stratégique pour des nations connectées à toute échelle. En France, la cyberdéfense, la cybersécurité et la lutte contre la cybercriminalité constituent l'une des priorités définies par le dernier livre blanc de la Défense Nationale 2013 [1]. Les grands acteurs étatiques, économiques et industriels s'accordent désormais sur la nécessité de doter la nation d'infrastructures de sécurité actives, résilientes, et lorsque la technologie le permet, antifragiles [2,3,4]. L'un des enjeux sous-jacents est de faire émerger une véritable culture populaire des turbulences numériques. Celle-ci sous-entend une prise de conscience rationnelle des cybermenaces, une meilleure évaluation des risques individuels liés aux pratiques numériques et une information objective sur le devenir des données personnelles créées par l'utilisateur afin de dissiper une illusion d'anonymat qui leurre souvent le cybercitoyen. La complexité des interactions opérant sur les espaces numériques multiplie les angles d'études. Qu'elle provienne de l'expert en stratégie, du géopolitologue, du militaire, de l'analyste Data centré, du développeur codeur, de l'anthropologue, du juriste ou du psychanalyste, chacune de ces contributions apporte un éclairage complémentaire dans la compréhension d'événements numériques souvent rapides, furtifs et violents. Aujourd'hui, tout conflit armé généralisé ou localisé, symétrique ou asymétrique, toute guerre civile, toute insurrection intègre systématiquement une compo-

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

sante « cyber » dans les luttes de rationalités qui l'animent. Une guerre moderne se projette naturellement sur le cyberespace, depuis son origine, puis tout au long de son développement et bien après sa conclusion. L'information qui guide les choix stratégiques des donneurs d'ordres oscille constamment entre les couches physiques, logiques et sémantiques du cyberespace. Elle est tantôt projetée sur la sphère numérique, tantôt relevée sur l'espace réel. C'est elle qui déclenchera l'action militaire, c'est elle qui l'orientera, la limitera ou l'éteindra. La « guerre de l'information » , présente dès le début des conflits humains s'intensifie à mesure que les technologies assurant sa collecte et son traitement se développent.

La contre-information, la cybermanipulation, la cybersubversion viennent renforcer l'arsenal du combattant qui doit agir conjointement sur les espaces physiques et numériques en anticipant l'effet des interactions qu'il suscite.

L'asymétrie de certains conflits, que l'on mesure classiquement en terme de puissance des armées mises en confrontation, doit être totalement repensée sur le cyberespace : un attaquant disposant d'infrastructures de calculs standards au sein d'un groupe rebelle irrégulier peut mener des offensives destructrices sur des composantes vitales d'un état doté pourtant d'une armée moderne, régulière et technologique [5,6,7]. L'interconnexion des structures vitales (centres de production d'énergie, centres de traitement de l'eau, réseaux de communications, de transports, ou de commandement) apporte à un pays une grande modernité mais aussi une grande vulnérabilité face à des attaques de plus en plus sophistiquées et planifiées. La production exponentielle de données au niveau mondial constitue aujourd'hui la plus grande source d'information pour les services de renseignements qui, pour exploiter ces données souvent ouvertes, développent des centres de collecte, de stockage, de traitement et d'analyse toujours plus puissants. L'information, considérée par de nombreux opérateurs comme le nouveau pétrole, joue un rôle fondamental dans les mécanismes de cyberconflictualités et dans la cristallisation des concurrences et des duels algorithmiques. Les notions de traces ou d'empreintes numériques enregistrées par les systèmes, les métadonnées associées à nos interactions quotidiennes participent alors très directement à la constitution d'un corpus informationnel sensé représenter l'utilisateur des moyens numériques. Généralisant les traces et empreintes numériques, le concept de projections algorithmiques d'un individu développé en 2013 [8] permet de structurer la donnée selon l'algorithme qui est à l'origine de sa création. Le formalisme projectif facilite l'approche systémique de certaines situations en maintenant les liaisons entre contenu informationnel, algorithme générateur et système sur lequel il est exécuté.

2 Concurrences et duels algorithmiques

Les traces numériques ou projections algorithmiques d'un opérateur interviennent dans les contextes de concurrences et de duels algorithmiques. Une agression informatique induit presque sûrement des projections caractéristiques de cette attaque. Le conflit de rationalité sous-jacent « s'exporte » naturellement dans les projections de ses acteurs. Celles-ci fournissent alors des informations utiles à l'anticipation et la prévision des réactions de l'adversaire. La collecte et l'analyse (sémantique) des pro-

jections algorithmiques de l'ennemi participent directement à la maîtrise stratégique de la confrontation. Ceci sous-entend d'être en mesure d'évaluer à tout moment la valeur d'une donnée collectée : sa valeur stratégique, sa valeur d'impact sur un contexte, sa valeur instantanée relativement à l'algorithme qui l'interprète sur ce contexte [9]. Là encore, une approche systémique associant projection algorithmique et valeur de son contenu facilite l'analyse des interactions qui opèrent en situation de conflit. Les exemples et cas d'étude concrets se multiplient à grande vitesse. Le conflit syrien apparaît comme très emblématique dans le sens où il réunit une confrontation « cinétique » (bombardements, opérations militaires et civiles, renseignement, influence) et une cyberconfrontation (guerre de l'information, attaques informatiques, renseignement, subversion, ingénierie sociale et cyberagressions multiples). Dès le début de cette guerre civile, en 2011, les belligérants (le régime de Bachar el-Assad et la rébellion syrienne) investissent le cyberespace et projettent le duel cinétique sur l'espace numérique [7]. En mai 2011, la SEA (Syrian Electronic Army) est créée en soutien du régime syrien. Elle conjugue les opérations de contre-information et les attaques ciblées sur l'ensemble des soutiens occidentaux et arabes de la rébellion (plus de 200 attaques répertoriées). La SEA développe une infrastructure de propagande et de revendication particulièrement efficace, organisée militairement sur les réseaux sociaux (Twitter, Facebook). Les projections algorithmiques des membres actifs de la cellule s'étendent bien au-delà de la simple publication de textes, d'images et vidéos d'exécutions sommaires ou de massacres attribués à la rébellion. Elles témoignent d'activités de renseignement opérées contre les structures numériques organisées soutenant la rébellion. L'utilisation de malwares (Trojan, Keylogger) destinés à la collecte furtive d'information a été régulièrement dénoncée par les membres de la rébellion qui, de leur côté, symétrisent le procédé contre les membre de la SEA. Des boucles rétroactives du type :

« projections algorithmiques offensives – actions cinétiques – projections défensives et riposte» s'installent en marge du conflit armé. Elles ne se limitent pas aux seuls intérêts syriens puisqu'elles impactent de nombreuses cibles occidentales. Ainsi, le 24 avril 2013, la SEA attaque le compte Twitter de l'agence Associated Press (AP) [9]. Elle en prend momentanément le contrôle et publie à 13h07 le message suivant : « Breaking : Two Explosions in the White House and Barack Obama is injured » soit : « Deux explosions à la Maison Blanche, Barack Obama est blessé »

Les 1.9 millions d'abonnés au compte Twitter d'Associated Press reçoivent le faux message posté par la SEA en le considérant comme authentique. La réaction des marchés financiers est presque immédiate : entre 13h08 et 13h10, l'indice principal de Wall Street, le Dow Jones (DJIA) perd 145 points soit l'équivalent de 136 milliards de dollars (105 milliards d'euros) en raison notamment du trading haute fréquence (HFT) qui a interprété et « réagi » au faux tweet. Les actions Microsoft, Apple, Mobil perdent plus de 1% presque instantanément. Quelques minutes plus tard, Associated Press reprend le contrôle de son compte et publie immédiatement un tweet annonçant que le message précédent était un faux et qu'il résultait du piratage de son compte. Aussitôt, l'indice Dow Jones remonte avec l'ensemble des valeurs qui venaient de chuter et reprend rapidement son cours normal. La courte durée de prise en compte du faux message publié par la SEA a suffi à modifier un indice boursier stratégique. L'ac-

tivité des systèmes automatisés de trading haute fréquence, capables de passer des ordres en quelques millisecondes, a modifié les lignes de prise de décision repoussant le contrôle humain en fin d'opération. La validation automatique et la prise en compte d'une information fausse (mais d'une projection algorithmique authentique) peuvent donc avoir un impact considérable sur un environnement interconnecté. Les projections algorithmiques engendrées par les membres de la SEA forment une séquence cohérente en offensif comme en défensif. Elles possèdent des spécificités prononcées qui facilitent l'attribution (ou la non attribution) des attaques revendiquées par cette cellule.

3 Hacking d'influence et prévisibilité des attaques

3.1 Typologie du hacking

Lorsque l'on cherche à construire une typologie du hacking, il faut distinguer trois grandes familles de cyberagressions : le hacking ludique ou opportuniste, le hacking d'acquisition-exploitation et le hacking d'influence.

Le hacking ludique rassemble les cyberagressions conduites par un attaquant isolé ou une cellule de hackers qui sélectionnent une cible selon l'unique critère de présence et d'identification d'une faille de sécurité exploitable. Le contenu informationnel de la cible n'intervient pas dans sa sélection. L'objectif du hacker se confond avec sa volonté de réussir l'attaque, de l'enregistrer et de faire valider son attaque sur un registre de trophées numériques.

L'ensemble des cibles potentielles d'un hacking ludique est particulièrement vaste puisqu'il n'est contraint que par le niveau de sécurité protégeant la cible. La découverte d'une vulnérabilité résulte de recherches exhaustives, parfois automatisées ou intervient de façon accidentelle selon le principe de sérendipité (ou sagacité accidentelle).

Le hacking d'acquisition-exploitation repose sur un objectif de collecte de données puis d'exploitation afin d'en tirer un avantage au niveau économique ou informationnel. C'est typiquement le cas du vol de bases de données bancaires permettant d'effectuer des paiements en ligne ou se négociant sur les marchés spécialisés du hacking.

Le hacking d'acquisition, qui intervient également dans les opérations de renseignement ou d'intelligence économique, mobilise souvent des techniques algorithmiques sophistiquées pour réaliser l'intrusion puis procéder à l'exfiltration des données captées. La furtivité de l'attaque participe directement à sa réussite. Il n'existe, en général, aucune revendication une fois l'attaque effectuée. Les cellules assurant ce type de hacking peuvent agir pour leur propre compte ou sur commande d'un état ou d'une organisation privée. Les composantes « cyber » des armées nationales pratiquent très régulièrement ce type d'opérations.

Le hacking d'influence est construit sur un objectif de diffusion d'un message militant en liaison avec un conflit présent sur l'espace physique ou une lutte de rationalités qui s'exporte sur le cyberespace. Il agit parfois comme le bras armé d'une cyberdiplo-

matie dont il renforce la parole. En ciblant les opinions publiques, il installe un contexte de duel algorithmique et accompagne l'opération d'influence qui opère de plus en plus souvent sur les réseaux sociaux et les espaces de communication partagés. La SEA a construit son activité sur le format du hacking d'influence. Il en est de même pour les armées électroniques iraniennes et turques.

3.2 Défacements

L'anglicisme « défacement » désigne une cyberattaque qui modifie intentionnellement le contenu d'un site web par l'ajout ou la suppression d'information (on parle parfois de défiguration). La pratique du défacement se décompose chronologiquement en trois phases. La première repose sur une campagne d'hameçonnage et d'ingénierie sociale destinée à collecter des identifiants et mot de passe d'opérateurs disposant de droits privilégiés sur le site ciblé. Cette opération peut également s'appliquer à un registraire qui gère un ensemble de noms de domaines. La seconde phase du défacement consiste à prendre le contrôle du site ciblé, lorsque l'attaquant dispose des droits d'administrateur, et à le transformer en lui ajoutant un message de revendication ou de justification de la cyberagression. Si la prise de contrôle n'est pas possible, l'attaquant procède à une redirection de l'adresse légitime vers un faux site qui contiendra le message justifiant l'opération. La troisième et dernière phase consiste, pour la cellule de hacking, à faire valider et enregistrer son attaque sur un site officiel d'archivage des défacements. Typiquement, Zone-H offre ce service à la communauté du hacking depuis une dizaine d'années et archive ainsi plus de dix millions d'attaques par défacement.

3.3 L'étude d'une communauté de hacking d'influence

Nous avons mené une étude détaillée sur un groupe de 320 cellules de hacking, actif sur Zone-H durant la période avril-mai 2014 et responsable à lui seul de 1 091 996 défacements réalisés depuis l'origine de l'enregistrement des cellules du groupe. Cette étude a d'abord mis en lumière le phénomène assez classique de distribution de Pareto des cellules les plus efficaces dans le groupe. Les 25 cellules les plus actives du groupe produisent à elles seules plus d'un million d'attaques, les 295 suivantes assurant le reliquat. Ainsi, 90 % des cyberagressions est attribuable à moins de 10 % des cellules. La seconde propriété déduite de l'étude concerne l'existence d'un « liant » assurant la cohésion et facilitant la coopération entre les sous-groupes de cellules pour organiser une opération de grande envergure. Le liant prédominant sur la famille étudiée relève clairement de l'appartenance religieuse musulmane. Il permet de fédérer de petites structures de hacking, réparties dans l'ensemble des pays musulmans, et de les faire coopérer sur des projets collectifs d'attaques d'intérêts numériques américain, israéliens, européens ou russes.

Enfin, l'étude a démontré qu'une veille exhaustive des projections algorithmiques des cellules de hacking les plus actives permettait, dans certaines circonstances, de construire des prévisions de futures attaques incluant un ensemble de cibles potentielles et une temporalité. Cette méthodologie de veille a permis de prévoir une OPI-

sraël II, sa date de réalisation et un ensemble de cibles possibles qui ont effectivement fait l'objet d'une cyberattaque à cette date...

Pour conclure...

La projection des confrontations humaines sur l'espace numérique influence le déroulement des guerres contemporaines. L'information, la contre-information, la subversion, le renseignement obtenu à partir de sources ouvertes ou celui obtenu par des attaques informatiques, complètent aujourd'hui l'arsenal conventionnel du combattant. Les conflits modernes ont définitivement investi le cyberespace obligeant les responsables en charge de la sécurité des systèmes à réinventer constamment leur défenses. L'anticipation et la prévision des cyberagressions vont contribuer à renforcer ces défenses.

Bibliographie

[1] Livre blanc de la Défense Nationale 2013
http://fr.calameo.com/read/000331627d6f04ea4fe0e

[2] N.N. TALEB, Antifragile, Les bienfaits du désordre, 2013, Édition Les belles lettres.

[3] BERTHIER T., Concurrences et duels algorithmiques, Revue de Défense Nationale, juin 2013

[4] BERTHIER T., Créons l'observatoire des évolutions algorithmiques, Défense et Sécurité Internationale, mai 2013.

[5] E. FILIOL, Les virus informatiques, théorie, pratique et applications, deuxième édition, Collection IRIS, 2009, Springer - Verlag.

[6] BERTHIER T., Cyberchronique – Décomposition systémique d'une cyberattaque, dissymétries et antifragilité, Publications de la chaire de cyberstratégie CASTEX, janvier 2014

[7] KEMPF O, BERTHIER T., L'armée électronique syrienne : entre cyberagression et guerre de l'information - Revue de Défense Nationale, - Volume Guerre de l'information - mai 2014

[8] BERTHIER T., Projections algorithmiques et cyberespace, Vol 5/2 2013, pp179-195, R2IE, Revue internationale d'intelligence économique.

[9] BERTHIER T., Sur la valeur d'une donnée – Publications de la Chaire de cyberdéfense Saint-Cyr-Sogeti-Thales – mai 2014.