Les données sont-elles devenues le premier enjeu de la cybercriminalité ?

Par Éric FREYSSINET
Officier général de gendarmerie

Les données sont au cœur des enjeux de la cybersécurité, en tant que cible principale des cybercriminels et comme outil non seulement pour ces mêmes délinquants mais aussi pour les défenseurs des systèmes d'information. Cette préoccupation concerne tant les données à caractère personnel que l'ensemble des données sensibles des organisations. Et les risques ne sont pas liés qu'aux célèbres rançongiciels, même s'ils constituent la menace la plus dynamique.

Chaque semaine, plusieurs alertes passent dans les flux d'actualités informant de détournements de données importantes : ainsi, au moment même de la rédaction de cet article, en étaient victimes un groupe de gestion de patrimoine français, une ONG ou encore un casino. Aux États-Unis, un rapport de l'ITRC[1] relevait en 2021 une augmentation de 68 % du nombre de compromissions de données. Au titre de la même année, la CNIL rapportait[2] 2 150 violations de données personnelles.

Derrière ces chiffres se cachent des réalités très variables. Un cas emblématique en est, par exemple, le vol et la diffusion en 2021 des données de 491 840 patients de laboratoires de biologie médicale de l'ouest de la France avec, dans certains cas, non seulement la captation de leurs coordonnées mais également d'informations à caractère médical, telles que les maladies pour lesquelles ils sont suivis ou leurs traitements.

Force est de constater que les données sont devenues le carburant ou, en tous cas, la face visible d'une grande partie des attaques cybercriminelles modernes. Nous nous proposons ici d'explorer les différentes facettes de cette réalité : le cadre légal, l'évolution des pratiques des criminels et les stratégies de ceux qui protègent les systèmes d'information.

La protection des données, un enjeu traduit depuis longtemps dans le droit

Le législateur français ne s'y est pas trompé en faisant de la donnée le cœur de la protection des systèmes d'information. Ainsi, dès 1978, avec la loi Informatique et Libertés[3], le Parlement consacrait l'importance de la protection contre les abus dans l'utilisation des données personnelles : « Article 25 – La collecte de données opérée par tout moyen frauduleux, déloyal ou illicite est interdite. »

Et, dix années plus tard, la loi Godfrain[4] est venue définir les infractions commises à l'encontre des systèmes informatiques en consacrant le concept de « système de traitement automatisé de données », qui se révèle à la fois très général – puisqu'il s'applique non seulement à des ordinateurs, mais également dans les faits à tous les systèmes automatisés qui traitent des données, et donc les réseaux informatiques, les terminaux ou les supports de stockage – et très spécifique, en ce qu'il définit très clairement l'objectif de la loi : protéger les données.

On retrouvera ensuite ce concept dans les traités internationaux qui visent le même objectif, avec, par exemple, la convention du Conseil de l'Europe sur la cybercriminalité qui, adoptée en 2001, définit ainsi un système informatique (lequel doit être protégé par les lois et les enquêtes judiciaires) : « L'expression "système informatique" désigne tout dispositif isolé ou ensemble de dispositifs interconnectés ou apparentés, qui assurent ou dont un ou plusieurs éléments assurent, en exécution d'un programme, un traitement automatisé de données. »

D'autres formes de données bénéficient de protections spécifiques en droit. Ainsi, l'on notera que le Code de la propriété industrielle protège notamment les logiciels informatiques ou les bases de données contre les différentes formes de contrefaçon et d'usage détourné.

Cette approche correspond donc à la fois à une réalité technique – celle des systèmes d'information dont l'objet et les principes de fonctionnement tournent autour de la donnée – et à une réalité opérationnelle – la nécessité de protéger les données et les systèmes qui les traitent. Nous verrons d'ailleurs que les traitements de données sont non seulement une finalité mais aussi un outil pour les attaquants eux-mêmes.

(1) Identity theft resource centre, https://www.idtheftcenter.org/post/identity-theft-resource-center-2021-annual-data-breach-report-sets-new-record-for-number-of-compromises/

(2) Rapport annuel 2021 de la Commission nationale informatique et libertés, https://www.cnil.fr/sites/default/files/atoms/files/cnil_-_42e_rapport_annuel_-_2021.pdf

(3) Loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés.

(4) Loi n°88-19 du 5 janvier 1988 relative à la fraude informatique (dite loi Godfrain).

Le rôle des données dans les pratiques des cybercriminels

Les lois sur la cybercriminalité visent donc à protéger les données des atteintes illégales ; des atteintes qui sont en pleine croissance, notamment dans leur forme la plus visible, à savoir les menaces de divulgation de données que profèrent les groupes criminels qui se cachent derrière les rançongiciels.

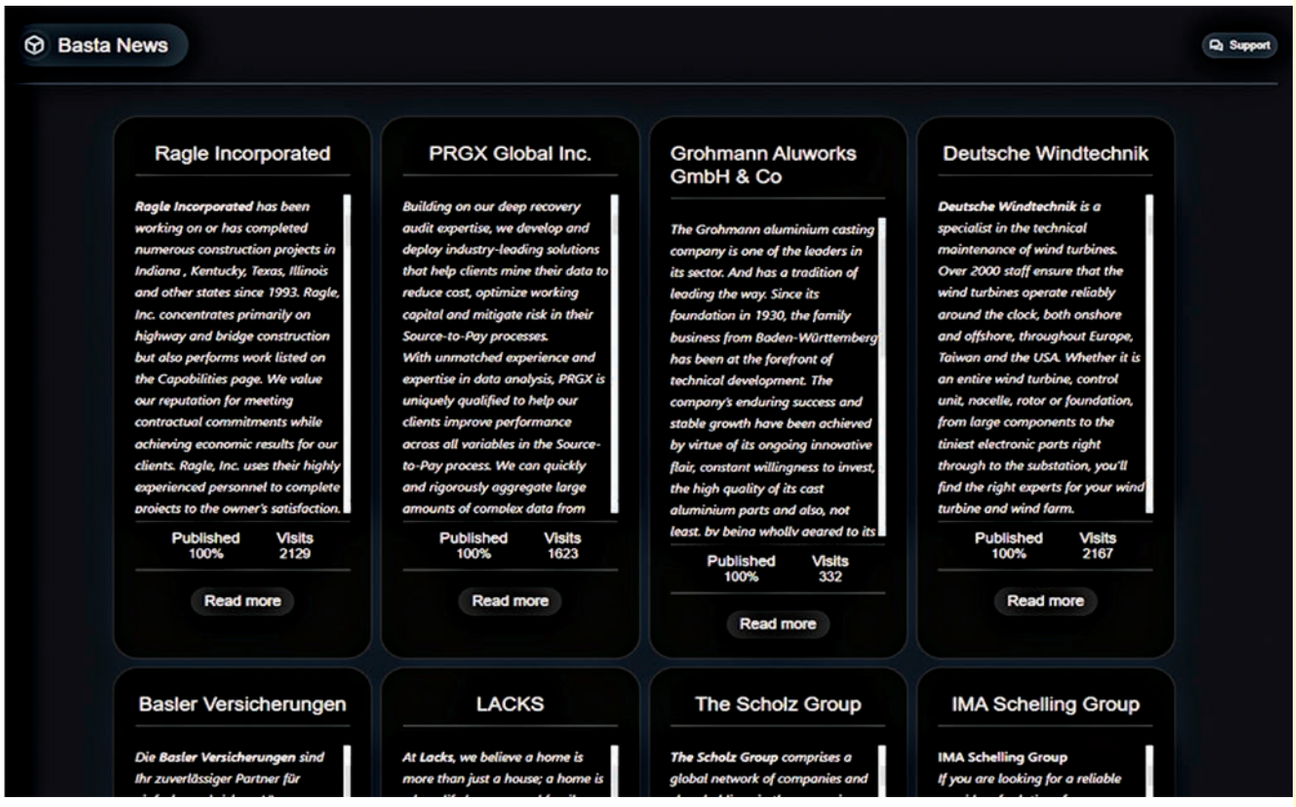
Au nombre d'environ 150, ces groupes de rançonneurs aux organisations variables – certains regroupant quelque dizaines d'affiliés utilisant leur plateforme pour attaquer des victimes spécifiques – utilisent tous aujourd'hui la stratégie dite de la double extorsion, à savoir la menace de ne plus permettre aux utilisateurs légitimes d'accéder à leurs données (le rançongiciel chiffrant les données sur les systèmes informatiques victimes de l'attaque), mais aussi la menace de rendre publiques les données confidentielles de l'organisation ciblée.

La copie d'écran ci-après est celle de la page de l'un des plus récents de ces groupes cybercriminels (surnommé BlackBasta), avec la liste de ses victimes et indiquant le nombre des visiteurs des sites piratés et le volume de données effectivement publiées.

Mais essayons d'avoir une vision plus systématique des différentes stratégies de détournement des données que l'on a pu déjà constater :

- la collecte détournée ou l'*hameçonnage* qui consiste à créer un faux site Web qui reproduit le visuel et le fonctionnement d'un site légitime et dont l'adresse est diffusée par un procédé de masse (typiquement le *spam* diffusé par courrier électronique, SMS ou messagerie instantanée) ;
- une variante de cette technique est, quant à elle, rencontrée dans le monde plus physique des cartes bancaires par le biais d'appareils permettant la capture de copies des pistes magnétiques de ces cartes ou des codes frappés sur les claviers des distributeurs automatiques ou des terminaux ;
- le vol de bases de données à travers l'exploitation d'une faiblesse dans un logiciel d'administration desdites bases (par exemple, grâce à l'injection de commandes dans un formulaire d'un site Web qui normalement contrôle l'accès aux données) ;
- la capture exhaustive d'une base de données en libre accès, dans une logique commerciale en dépit de l'intention de son propriétaire (technique souvent appelée en anglais *Web scraping*), par exemple pour copier un annuaire ou le catalogue d'un concurrent ;
- l'intrusion dans le système d'information de la victime, puis l'extraction de données confidentielles depuis les systèmes attaqués – c'est souvent la première étape de l'attaque mise en œuvre par les groupes criminels de rançonneurs ;
- enfin, une autre variante consiste à utiliser un logiciel malveillant qui va de façon automatique collecter et exfiltrer des données depuis les ordinateurs d'un réseau informatique attaqué.

Mais outre ces actions criminelles, il ne faut pas oublier les divulgations involontaires : certaines organisations ayant mal sécurisé l'accès à leurs données voient celles-ci temporairement ou parfois sur de longues durées librement accessibles à tous les utilisateurs d'Internet. On parlera alors de fuites de données. C'est assez souvent le cas, notamment lorsqu'un système de stockage dans le *Cloud* n'est pas correctement sécurisé.



Les données ciblées sont de différentes natures ; elles peuvent être regroupées dans les catégories suivantes :

- les identifiants et les mots de passe ;
- les données bancaires (y compris les identifiants d'accès aux comptes bancaires) ;
- les bases de données ;
- les fichiers stockés.

Peuvent aussi être ciblées, de façon plus spécifique, des copies d'écran, des images ou des enregistrements sonores détournés à partir des caméras et microphones des ordinateurs ou téléphones de leurs propriétaires.

Toutes ces données ont différents usages et peuvent donc présenter de la valeur pour les attaquants, et ce à différents niveaux :

- les données elles-mêmes peuvent être revendues : chaque catégorie de données a une valeur unitaire qui varie de plusieurs euros à plusieurs centaines d'euros (c'est le cas, en particulier, des identifiants et mots de passe ou des données bancaires) ;
- parfois, la revente concerne plus spécialement un accès détourné à des systèmes d'information (ainsi certaines personnes fournissent à la demande de leurs clients des données qu'ils vont, grâce à un accès légitime, piocher dans un système exploité par la police ou un opérateur de communications) ;
- les données peuvent aussi être utilisées pour démultiplier le potentiel d'une autre attaque : c'est notamment le cas des identifiants de messagerie ou de réseaux sociaux qui peuvent être exploités pour contacter les amis et contacts professionnels d'une première victime dans but d'escroquer de nouvelles victimes mises en confiance par un échange présumé être avec une personne qu'elles connaissent ;
- enfin, le rançonnement à un ou deux niveaux, comme évoqué précédemment.

On observe dans ce champ particulier du marché des données volées une des caractéristiques importantes des activités cybercriminelles modernes, à savoir l'existence d'un véritable écosystème, où des acteurs cybercriminels commercialisent entre eux différents services : il s'agit de développeurs de solutions logicielles (les logiciels malveillants, les plateformes permettant de les piloter), des administrateurs de ces systèmes, de ceux qui les louent pour les mettre en œuvre (souvent appelés affiliés), de spécialistes de l'intrusion, de revendeurs de données et, le cas échéant, de spécialistes du blanchiment des bénéfices retirés de ces différentes opérations.

On le voit donc très clairement, l'exploitation commerciale de la donnée est la motivation principale d'une grande partie des activités cybercriminelles, vraisemblablement sa principale source de financement – à travers la revente ou en favorisant l'extorsion de fonds dans le cadre de chantages.

La donnée peut aussi être utilisée comme un outil pour commettre de nouvelles attaques par rebond vers de nouvelles victimes. D'ailleurs, c'est peut-être dans le cadre de cette dernière dimension que l'on verra s'accroître le nombre des escroqueries dans les années à venir. Car de la même façon que la donnée est indispensable pour protéger les systèmes d'information comme nous l'évoquerons *infra*, elle l'est tout autant pour les attaquants dans leur construction de modèles et de stratégies d'attaque (par une meilleure connaissance de leurs cibles potentielles et de leurs vulnérabilités). Tout comme le seront demain les techniques de l'intelligence artificielle qui nécessiteront pour les cybercriminels l'analyse de gros volumes de données d'apprentissage, mais leur permettront aussi d'étendre et d'accélérer leurs activités. En effet, on peut parfaitement imaginer les cybercriminels exploiter, par exemple, des *chatbots* pour démultiplier le nombre des victimes avec lesquelles ils interagissent.

Les données au cœur des stratégies de cybersécurité

On le voit la première préoccupation d'un responsable d'un système d'information et de ses équipes en charge de la sécurité de celui-ci est donc la protection des données de ce système. Il faut donc être en mesure de maîtriser quels types de données on possède, quelles sont les plus sensibles, quels systèmes les stockent et les traitent, qui y a accès ? Et, au bout du compte, d'être capable de détecter les tentatives d'accès à ces données ou, le cas échéant, les données qui ont pu fuiter.

Cette approche « données » est tout aussi valide dans le cas d'un système informatique servant à réaliser des opérations de production industrielle, car c'est à tout le moins l'intégrité et la disponibilité des données permettant au système industriel de fonctionner que l'on cherchera à préserver et, dans beaucoup de cas, leur confidentialité.

Mais les architectures de protection des systèmes d'information génèrent une couche supplémentaire de données qui servent à protéger lesdits systèmes. Ainsi, on collectera des traces d'activité liées à différentes applications (notamment les traces d'accès aux données), des traces d'interactions sur les réseaux et, de plus en plus souvent, des données liées au fonctionnement des systèmes informatiques eux-mêmes (avec un modèle de sécurité informatique évoluant de l'antivirus vers la notion d'EDR (ou *endpoint detection and response*)).

Au passage, ces données de sécurité sont tout aussi importantes que celles qui étaient initialement protégées ; et la sécurité de leur collecte, de leur stockage et de leur traitement doit elle aussi être préservée.

L'utilisation de ces données doit permettre idéalement de détecter les opérations suspectes et les intrusions en cours. Mais elles peuvent aussi avoir un usage rétrospectif : par exemple, pour comprendre le fonctionnement et l'impact d'une attaque une fois qu'elle a été révélée. C'est l'une des premières opérations que l'on réalisera lorsque l'on découvrira une attaque par un rançongiciel afin, par exemple, de distinguer les systèmes qui sont touchés de ceux qui ne le sont pas.

Les investigations rétrospectives se révèlent également très utiles pour la révélation d'une faille de sécurité inconnue jusque-là (souvent appelées vulnérabilités 0-day). Il s'agit ici de vérifier, grâce à des traces caractéristiques (ou indices de compromission, les IoC), que le système d'information que l'on gère n'a pas été l'objet d'une attaque permettant de l'exploiter. Les mêmes données recueillies dans le cadre de ces investigations pourront bien évidemment être aussi exploitées par les enquêteurs judiciaires.

De même, ces données sont aussi exploitées en masse par les analystes en géostratégie cyberriminelle ou encore par des chercheurs qui s'intéressent au fonctionnement des groupes cybercriminels… Ils vont même jusqu'à exploiter des conversations collectées (ou même qui ont fuité) sur des forums sur lesquels échangent des cybercriminels[5].

Les données sont donc l'objet sur lequel travaillent les responsables de la sécurité des systèmes d'information et sont, dans le même temps, un outil pour eux. Plus que les systèmes eux-mêmes, c'est l'ensemble du cycle de vie des données qu'ils doivent donc maîtriser et savoir exploiter.

[5] C'est un des sujets qui ont plus particulièrement émergé lors de la conférence Botconf, qui s'est tenue à Nantes du 26 au 29 avril 2022 (https://www.botconf.eu/botconf-2021/botconf-2021-22-final-schedule/), comme le relève Louis Adam dans son article *Botconf* : *la nécessité de surveiller les écosystèmes cybercriminels* (https://www.zdnet.fr/actualites/botconf-la-necessite-de-surveiller-les-ecosystemes-cybercriminels-39941429.htm).

Conclusion

On l'a vu, l'enjeu principal de la cybercriminalité peut aujourd'hui se décrire au travers des données qui sont protégées par leurs propriétaires et utilisateurs légitimes, mais qui sont aussi la cible et l'objet de détournements par les délinquants. Ces données sont aussi un outil pour ceux qui sont chargés de les préserver. Demain, la collecte de données permettra tout autant aux défenseurs de connaître les stratégies des attaquants potentiels (*threat intelligence*), qu'aux criminels de développer de nouvelles techniques d'attaque.

Une des conclusions que l'on peut tirer des développements qui précèdent, est qu'il est indispensable aujourd'hui, pour sécuriser les systèmes d'information, de se doter d'une stratégie tournant autour de la donnée et de disposer de bons outils, voire des bons spécialistes du traitement de ces données. Cette vision doit être partagée et transverse (données personnelles et confidentielles, données de production, données de sécurité). C'est à cette fin que de nouveaux métiers ou, en tous cas, de nouvelles compétences se développent dans les équipes de sécurité informatique.