InData

La lettre connectée à l'actualité digitale

OCTOBRE 2023

L'intelligence artificielle (IA) appliquée à la cybersécurité : Outil offensif ou défensif ?

LENDYS
Pôle Transformation
Finance & Digital

Introduction

Dans une société qui se dématérialise un peu plus chaque jour, l'explosion de l'utilisation de l'IA représente un outil majeur en cybermenaces. Néanmoins, ces technologies peuvent servir d'arme à double tranchant aux mains des cybercriminels. Par conséquent, il nous paraît important d'aborder le rôle de l'IA aussi bien d'un point de vue défensif qu'offensif. Après avoir présenté la notion de cybermenaces et identifié les auteurs et leurs motivations, nous aborderons le rôle de l'IA dans la cybersécurité tant sur le plan préventif que proactif face au risque spécifique de cybercriminalité.

Cybermenaces : panorama des principales cyberattaques

Les cybermenaces sont des tentatives malveillantes destinées à perturber un système informatique ou un réseau en volant des données ou en accédant à des fichiers non autorisés. Par environnement de cybermenace, on entend l'espace virtuel où les auteurs de cybermenace mènent des activités malveillantes. Il comprend les réseaux, les dispositifs et les processus qui sont connectés à Internet et qui peuvent être la cible d'auteurs de cybermenace, ainsi que les méthodes utilisées par ces auteurs pour atteindre les systèmes. Les cybermenaces touchent aussi bien les particuliers que les entreprises. Elles peuvent être une source de chantage (ransomware), de monétisation (revente d'infos personnelles sur le dark web), ou d'usurpation d'identité (phishing, etc.).

PRINCIPALES CYBERMENACES DANS L'UE

Attaques au moyen de rançongiciels
Attaques dans le cadre desquelles des cybercriminels prennent le contrôle d'un système cible et demandent une rançon pour rétablir sa disponibilité.

Menaces de déni de service distribué
Attaques empêchant les utilisateurs d'un réseau ou d'un système d'accéder à des informations, services et autres ressources utiles.

Logiciels malveillants
Programmes mavleillants conçus pour accéder à un appareil sans autorisation, endommager un appareil ou perturber l'accès à un appareil.

→ Rien qu'en juin 2022, des cheveaux de Troie ont été téléchargés environ 10 millions de fois par le biais de logiciels publicitaires.

Menaces d'ingénierie sociale
Menaces visant à exploiter une erreur humaineou un comportement humain pour obtenir l'accès à des informations ou des services.

→ 82% des violations de données reposaient sur un facteur humain.

Menaces sur les données
Attaques visant à accéder sans autorisation à des données et à les manipuler pour interférer avec le comportement des systèmes.

→ Les cibles les plus fréquentes de ces attaques ont été les serveurs (près de 90% des cas).

Menaces pour l'Internet
Attaques ayant des conséquences sur la disponibilité de l'Internet. Piratage du protocle BGP (Border Gateway Protocole), par exemple.

→ En juin 2022, 15% de l'infrastructure Internet de l'Ukraine avaient été détruits par la Russie.

Désinformation - malinformartion
Une attaque intentionnelle consistant à créer ou à partager des informations fausses et trompeuses pour manipuler l'opinion publique.

→ Des campagnes de désinformation massive aviaent déjà visé l'Ukraine avant le début de son invasion par la Russie.

Attaques de la chaîne d'approvisionnement
Une stratégie d'attaque visant une organisation à travers des vulnérabilités dans sa chaîne d'approvisionnement avec le potentiel d'induire des effets en cascade.

→ Les incidents sur la chaîne d'approvisionnement ont représenté 17% des intrusions en 2021, contre moins de 1% en 2020.

Source : Agence de l'UE pour la cybersécurité (Données de juillet 2021 à juillet 2022)
© Union européenne, 2022

AUTEURS DE CYBERMENACES ET LEURS MOTIVATIONS

Les auteurs de cybermenace peuvent être classés en fonction de leur motivation et dans une certaine mesure, de leur niveau de compétence technique. Leur objectif est d'accéder à des dispositifs et des réseaux pour diverses raisons, telles que la manipulation ou le vol d'informations, la perturbation des performances d'un réseau et l'extorsion des propriétaires. Certains auteurs de cybermenace ciblent spécifiquement des individus ou des organisations, tandis que d'autres agissent de manière opportuniste en visant des systèmes vulnérables. Chaque catégorie d'auteurs de cybermenace est motivée par un objectif. Une « taxonomie » que nous pouvons résumer comme suit :

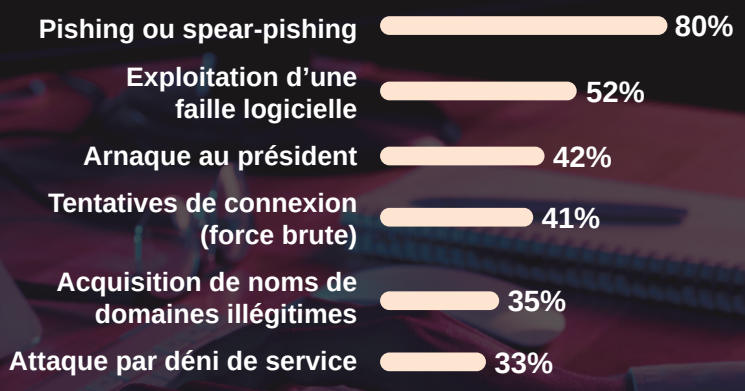
AUTEURS DE CYBERMENACE	MOTIVATION
ÉTATS-NATIONS	GÉOPOLITIQUE
CYBERCRIMINELS	GAINS FINANCIERS
HACKTIVISTES	IDÉOLOGIQUE
GROUPES TERRORISTES	VIOLENCE IDÉOLOGIQUE
AMATEURS DE SENSATIONS FORTES	SATISFACTION
MENACES INTERNES	MÉCONTENTEMENT

LES CYBERCRIMINELS, QUI SONT-ILS ?

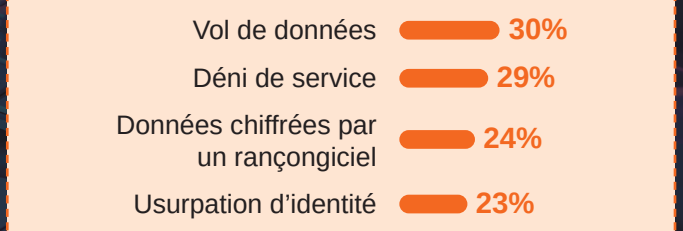
Les cybercriminels sont principalement motivés par le gain financier et le degré de sophistication de leurs méthodes varie beaucoup. Que ce soit par le biais de l'hameçonnage[1], de l'ingénierie sociale ou d'autres tactiques, les cybercriminels exploitent souvent les failles techniques et/ou la naïveté humaine pour pénétrer dans les systèmes. Aussi, la cybercriminalité concerne au premier plan les structures et organisations du quotidien. En effet, les entreprises, les institutions sanitaires et les gouvernements sont les premières victimes de pertes de données sensibles, de charges financières énormes et de dommages à leur image de marque. Ce constat s'est accentué avec la crise sanitaire qui a confronté les entreprises à de nouveaux cyber-risques, en lien notamment avec la généralisation du télétravail et l'usage d'applications et de services Cloud dont la sécurité fait parfois défaut.

[1]L'hameçonnage ou le phishing consiste à obtenir du destinataire d'un courriel, d'apparence légitime, qu'il transmette ses coordonnées bancaires ou ses identifiants de connexion à des services financiers, afin de lui dérober de l'argent.

Les cyberattaques les plus courantes contre les entreprises françaises *(constatées en 2020)*



Principales conséquences des attaques :



L'IA en tant qu'outil défensif

Lors de nos précédentes publications sur l'utilisation du machine learning (ML) dans les secteurs banque et assurance, nous avons présenté les méthodologies d'algorithmes prédictifs de cette technologie. Entre autres, nous avons évoqué les mécanismes de détection de fraude qu'il s'agisse de fraude à la carte bancaire ou à l'assurance. A cet effet, nous avons parlé de stratégies de détection de signaux faibles permettant de recevoir des d'alertes à chaque suspicion de fraude. Dans ce cas, l'approche est dite plus fine dans la détection en comparaison avec des logiciels de surveillance classique. Cette finesse permise par le Machine Learning aiderait à réduire le nombre de fausses notifications et permettrait de focaliser[2] l'effort sur les « vraies menaces ». En somme, ce que nous appelons protection prédictive repose sur une IA entrainée_via des algorithmes de ML_ à discriminer les programmes en logiciels « malveillants » et « non malveillants ». Les cas d'usages possibles de cette approche prédictive sont multiples. Une simple recherche avec le moteur « Google » permet d'analyser les pages web abordant les sujets de l'usage de l'IA en cybersécurité défensive. Le tableau ci-dessous est une synthèse non exhaustive des cas d'usages possibles.

DOMAINE D'APPORT DE L'IA	CAS D'USAGE
RECONNAISSANCE ET INTERPRÉTATION DES DONNÉES	→ Identification des fraudes → Sécurisation des authentifications → Identification des tentatives de phishing → Analyse des comportements malveillants → Détection de l'usurpation d'identité
AIDE À LA DÉCISION	→ Corrélation d'événements → Analyse et évaluation de risques
TRAITEMENT DU LANGAGE NATUREL	→ Détection de spam → Désinformation → Identification
PRÉDICTION	→ Prédiction des risques

Cependant, rappelons que les technologies d'IA/ML reposent sur la collecte de données essentiellement récupérées en open source. Autrement dit, l'utilisation croissante des données par les systèmes d'IA/ML pourrait être propice à un détournement de leur usage initial par des acteurs malveillants. Par conséquent, sans supervision humaine de la matière première, l'IA préventive sera moins probante voire inefficace. Dans ce cadre, L'Agence Nationale de la Sécurité des Systèmes d'Information[3] (ANSSI) a détecté différents types de menaces à l'image de «l'empoisonnement de données» (modification des données durant l'apprentissage), du model évasion (déviation par rapport à ce que le modèle est supposé faire) ou du model inversion (extraction de certaines données secrètes du modèle). Une IA entraînée sur cette base polluée, laissera passer le code des cybercriminels sans rien soupçonner. C'est tout l'enjeu de la protection proactive ! Cette dernière analyse le comportement du logiciel une fois qu'il commence à fonctionner afin de stopper sa progression. On parle dans ce cas d'IA en tant qu'outil offensif.

Des technologies d'intelligence artificielle sont déjà intégrées dans des outils comme les anti-virus, les passerelles mail ou web, les firewalls…soit des équipements qui réagissent automatiquement aux attaques en filtrant le trafic malicieux. La plupart des éditeurs le font y compris en utilisant certaines technologies avancées telles que le deep learning (Google l'a intégré depuis 2015 pour filtrer les spams dans Gmail).

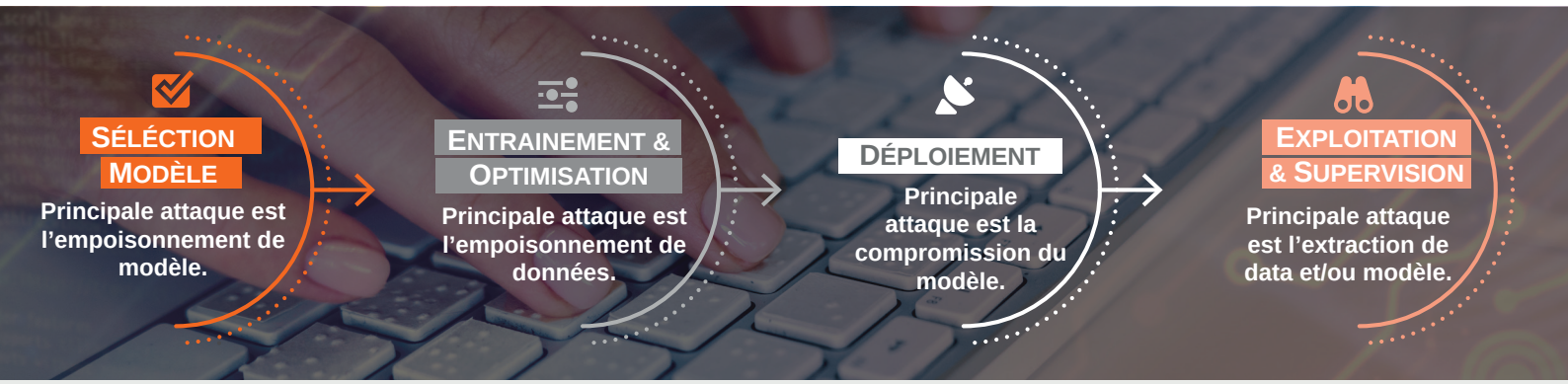
[2]Les tâches les plus chronophages et fastidieuses seraient ainsi gérées par la machine, limitant la fatigue et les pertes de temps des équipes qui pourraient alors se concentrer sur les actions à forte valeur ajoutée.

[3]La revue annuelle de l'ANSSI, édition 2020

L'IA en tant qu'outil offensif

Comme nous l'avons vu, si l'IA/ML est un outil intéressant en cyberdéfense, son usage peut être détourné par les cybercriminels. L'usage de l'IA/ML (en conjonction avec l'automatisation des cyberattaques) permettra d'atteindre des niveaux d'industrialisation des attaques encore inédits selon le Gartner et ce risque est d'autant plus sensible que beaucoup de compagnies adoptant ces technologies ne sont pas systématiquement au courant de ces menaces (Ijlal, 2022). Le schéma ci-dessous illustre les différentes étapes du développement d'un système basé sur l'IA/ML et les types d'attaques auxquels il peut être exposé à chaque étape. Le processus de développement d'un tel système comprend généralement quatre phases :

- Le choix du modèle
- L'entraînement / optimisation
- Le déploiement
- La supervision de l'exploitation



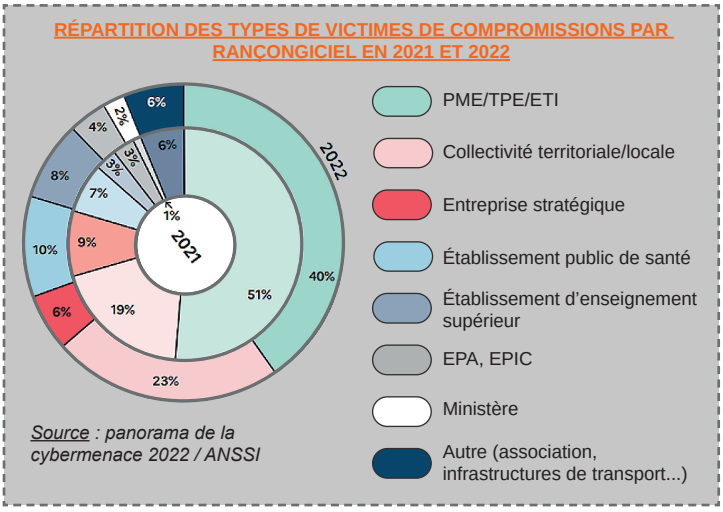
L'empoisonnement se fait via des « malwares » difficilement détectables utilisés souvent dans des modèles existants (voir annexe)

L'empoisonnement des données se produit lorsque des attaquants altèrent les données d'entraînement utilisées pour créer des modèles d'apprentissage.

L'attaque vise généralement l'infrastructure supportant la solution et l'éco-système applicatif.

L'attaquant peut créer une copie via l'API du modèle en requêtant et en observant les réponses du système ce qui lui permet de contourner voire de reconstruire le modèle ou la data.

Dès lors, la sécurisation passe indubitablement par la surveillance renforcée du comportement du système IA/ML de l'entrée à la sortie de la data afin de détecter toutes les dérives non contrôlées par le système. Dans ce contexte, le besoin de nouvelles solutions avancées de cybersécurité offensive basées sur l'IA est plus que jamais d'actualité. Qu'en est-il sur le terrain ? Peut-on parler d'une IA qui va booster l'innovation et assurer la cyber sécurité ?



[4] Gartner : Les principales tendances sécurité 2022

La promesse de l'IA augmentée dans la cybersécurité

Tout au long de nos récentes publications nous avons essayé d'exposer l'idée selon laquelle le succès de l'IA réside essentiellement de la massification des données ainsi que de l'accroissement des puissances de calcul. Finalement, la barrière réside plus dans l'accès aux données et à l'accès à une puissance de calcul suffisante. Du point de vue des pratiques, dans les années à venir, la cybersécurité consistera probablement moins à « défendre » qu'à progresser vers l'acceptation des cyber-risques permanents, en mettant l'accent sur le renforcement de la résilience et de la capacité de récupération des systèmes. Autrement dit, c'est une démarche proactive qui nécessite un investissement continu et adapté aux évolutions des menaces cybernétiques. En témoignent les derniers chiffres d'une étude publiée par le cabinet Grand View Research . On peut y lire que la taille du marché mondial de l'intelligence artificielle appliquée à la cybersécurité est évaluée à 16,48 milliards de dollars en 2022. Elle devrait atteindre les 93,75 milliards de dollars en 2030, se développant à un taux de croissance annuel composé de 24,3 %. Preuve en est que l'expertise en cybersécurité n'est pas innée, elle est acquise !

LA FORMATION CONTINUE COMME MEILLEURE DÉFENSE

Pour maintenir leur supériorité face à des cybercriminels de plus en plus agiles dans la maîtrise des technologies numériques, les entreprises doivent aujourd'hui anticiper les systèmes de défense à base d'IA qui sont eux-mêmes vulnérables aux leurres que les attaquants mettent en place pour les contourner. Par conséquent, les entreprises et les administrations peuvent développer une expertise similaire en investissant dans la formation de leur personnel. Cependant, il est nécessaire de reconnaître que cette formation doit être continue et adaptée aux évolutions constantes du paysage des menaces cybernétiques. Par exemple, le Danemark, qui est considéré comme le pays le plus sûr en matière de cybersécurité selon une étude de Comparitech de septembre 2022, investit fortement dans la formation en cybersécurité et a mis en place des initiatives pour encourager l'apprentissage continu dans ce domaine. En somme, la question n'est pas de savoir si les cybercriminels ont toujours un temps d'avance, mais plutôt de savoir comment rattraper le retard et renforcer la résilience face aux cyberattaques. Ces incidents soulignent l'importance de la cybersécurité et la nécessité d'investir dans la formation, l'éducation au numérique et d'avoir une gestion proactive des vulnérabilités. A titre d'exemple, les experts en sécurité doivent sensibiliser l'ensemble de personnel. Le meilleur moyen d'y parvenir est de simuler des menaces réelles. Pour cela, des exercices devraient être organisés régulièrement dans les entreprises à des fins de formation, pour mesurer comment les équipes réagissent lors d'une violation de la sécurité. L'objectif étant d'utiliser ce retour d'expérience pour améliorer les procédures de réponse aux incidents et renforcer les réflexes à adopter par les équipes lorsque les risques se présentent (simulation d'ingénierie sociale, phishing…) en complément des mesures de sécurité indispensables. Réaliser un test sous forme de simulation une fois par an ne suffit plus, il faut désormais informer, former et sensibiliser les collaborateurs en continu, et entraîner leurs réflexes sur les nouveaux types de menaces intégrant l'IA.

[5] Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution, By Technology (Deep Learning, Machine Learning), By End-use, By Region, And Segment Forecasts, 2023 - 2030

[6] Pour augmenter leur potentiel de nuisance à moindre coût, pour sélectionner leurs cibles, optimiser leurs capacités d'intrusion (ex : phishing intelligent), déterminer les meilleures failles à exploiter (ex : scan de pages web intelligent), passer autant que possible sous le radar des systèmes de détection, etc.

[7] Which countries have the worst (and best) cybersecurity? September 26, 2022

Conclusion

L'IA procure aux organisations un ensemble d'outils supplémentaires pour renforcer leur sécurité et celle de leurs utilisateurs contre les cyberattaques. Le rôle de l'intelligence artificielle dans la cybersécurité continuera de se développer, ce qui aidera les entreprises à améliorer leur profil de sécurité. C'est généralement le rôle de l'IA défensive. Cependant, l'apprentissage de l'IA obéit à des étapes de développement qui l'expose à des risques de cyberattaques multiples qu'elle soit ciblée ou détournée (un hackeur néophyte peut mettre la main sur le code d'une IA défensive et la réutiliser à des fins malveillantes).

Par conséquent, la meilleure chance de contrer ce type de menace est d'associer la formation réactive et continue et les technologies de l'IA préventive. Autrement dit, pour pouvoir faire face à de nouvelles formes de menaces, les entreprises doivent intégrer le changement de paradigme. Il est primordial d'impliquer les équipes, de les sensibiliser à l'importance de la sécurité au sein de l'organisation, de leur démontrer qu'elle est au cœur de l'entreprise et vitale pour sa pérennité. Rappelons que l'accompagnement au changement dans les entreprises est un enjeu crucial pour assurer une transition réussie vers une organisation de travail évoluant dans un environnement saturé par l'IA.

« L'intelligence artificielle est un moyen d'amplifier l'intelligence humaine, de même que les machines sont un moyen d'amplifier la force physique ».

Yann Le Cun (Chief AI Scientist for Meta)

Enfin, le recours à des techniques sophistiquées est tributaire de la résilience des techniques standards et actuelles. En effet, une grande partie des cyberattaques réussies peuvent être attribuées à la vulnérabilité humaine. Par conséquent, dans une société digitalisée, il est intéressant de se demander si les cybercriminels ont réellement toujours un temps d'avance, ou s'il s'agit juste d'une expression de la difficulté que l'on a à identifier et admettre nos propres manquements et expertises ? ■

■ **Mohammed MEHANAOUI**
Responsable d'Études R&D et Innovation | Pôle Transformation, Finance & Digital

RÉFÉRENCES

► Ijlal, T., 2022. Artificial Intelligence (AI) Governance and Cyber-Security: A beginner's handbook on securing and governing AIsystems. Broché – 21 avril 2022.

► Directive (UE) 2016/1148 du Parlement européen et du Conseil du 6 juillet 2016 concernant des mesures destinées à assurer un niveau élevé commun de sécurité des réseaux et des systèmes d'information dans l'Union.

► Proposition de DIRECTIVE DU PARLEMENT EUROPÉEN ET DU CONSEIL concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, abrogeant la directive (UE) 2016/1148.

ANNEXES

► ***Les « Trojans » :*** il s'agit d'exécutables qui apparaissent comme légitimes et sans danger. Ces programmes malveillants se cachent dans des programmes d'apparence inoffensive. Mais une fois que ces derniers sont lancés, ils exécutent des instructions malveillantes sur l'ordinateur hôte et peuvent par exemple ouvrir une porte d'accès dérobée.

► ***Les « Botnets » :*** ce sont des programmes dont l'objectif principal est de se propager sur autant d'hôtes que possible au sein d'un même réseau, dans le but, par exemple est de détourner leur capacité de calcul au service des attaquants

► ***Les virus :*** un virus est un programme qui s'auto-reproduit en infectant d'autres programmes. Très populaires sous MS-DOS et Windows 3.1, ils sont en nette perte de vitesse comparés aux autres formes de malwares.

► ***Les « Downloaders » :*** ce sont des programmes qui une fois installés téléchargent des librairies ou du code au contenu malveillant à partir du réseau. Une fois téléchargé, le code malveillant est exécuté sur l'hôte infecté.

► ***Les « RootKits » :*** ces malwares sont conçus pour permettre aux pirates d'installer une série d'outils permettant d'accéder à distance aux ordinateurs infectés. Ils compromettent le plus souvent les hôtes au niveau de leur système d'exploitation. Ils prennent souvent la forme de « drivers » ce qui rend la lutte contre ces programmes très compliquée et souvent inefficace.

► ***Les « Ransomwares » :*** ce sont des programmes malveillants dont la fonction est de chiffrer les données présentes sur le disque dur des ordinateurs infectés. Les victimes sont alors contactées pour payer une rançon contre les clés de déchiffrement des disques durs infectés.

► ***Les spywares :*** il s'agit de logiciels espions qui permettent de connaître l'activité sur l'ordinateur infecté. On y retrouve deux grandes familles : les Keyloggers (les «Enregistreurs de touches» stockent dans un fichier souvent chiffré toutes les touches frappées sur le clavier) et les Adwares (qui modifient la page de démarrage du navigateur internet, ou installent un plugin de recherche sur internet).

► ***Les Droppers :*** ce sont des fichiers exécutables qui paraissent anodins mais qui installent un ou plusieurs malwares.

► ***Les « Macros » :*** les malwares sous forme de macros s'insèrent dans des documents de type»suites bureautiques».

► ***Les « APTs » :*** il s'agit là de cyberattaques prolongées et ciblées visant des vulnérabilités spécifiques du parc d'ordinateurs installés, par lesquelles un pirate et/ou une organisation non autorisée accède au réseau d'une entreprise et passe inaperçue pendant une longue période. Une attaque APT vise généralement à surveiller l'activité réseau et à voler des données plutôt qu'à porter atteinte au réseau ou à l'organisation.

► ***Les « Zero days » :*** il s'agit de malwares qui exploitent des vulnérabilités qui n'ont pas encore été publiée par la communauté des chercheurs et analystes. Ces malwares ne sont par définition pas détectable et possède une très grande valeur pour les pirates ou certaines organisations gouvernementales.