Mathématiques contre logiciels malveillants

Existe-t-il une solution plus efficace ? Ou comment les mathématiques peuvent vaincre les logiciels malveillants.

« Les mathématiques sont préférables à toute autre connaissance que nous aient enseignée les hommes », René Descartes, philosophe, mathématicien et scientifique français

Le problème, bien que peu soient prêts à l'admettre, c'est que les salariés chargés de la sécurité de l'entreprise défendent un château truffé de trous, de passages secrets et protégé par des barrières inefficaces. Ces points faibles sont la conséquence de logiciels de sécurité de qualité médiocre, d'équipement ancien et, dans certains cas, de portes dérobées créées par des collaborateurs malveillants. Résultat : une acceptation à contrecœur selon laquelle les cybercriminels gagnent la guerre.

Les attaques sont motivées par de multiples raisons, proviennent de différentes zones géographiques et leur complexité évolue à mesure que les technologies progressent. Cette évolution implique que les menaces modernes emploient couramment des techniques d'évasion conçues pour contourner les mesures de sécurité existantes. Le simple fait de détecter ces menaces avancées après leur exécution est déjà compliqué, sans parler de protéger l'ensemble d'une entreprise avant leur exécution.

Et s'il existait une solution plus efficace ?
Et s'il était possible de défendre le château ?
Et s'il était possible d'arrêter une menace avant qu'il ne soit trop tard ?

Le facteur humain

Afin de se prémunir contre les cybercriminels modernes, les technologies de sécurité doivent évoluer en conséquence, sans se fier à l'intervention humaine. C'est là où les mathématiques et l'apprentissage automatique entrent en jeu. S'il est possible de classer objectivement des fichiers comme « valides » ou « malveillants » en se basant sur des facteurs de risques mathématiques, il est donc possible de former une machine à prendre les bonnes décisions sur ces fichiers en temps réel.

Dans les prochaines pages, nous revendiquons qu'une approche de la sécurité informatique basée sur les mathématiques et l'apprentissage automatique changera fondamentalement notre manière d'évaluer, de classifier et de contrôler l'exécution des fichiers. Nous expliquerons également comment les produits Dell tirent profit de cette approche et démontreront leur différence par rapport à toutes les autres offres de sécurité du marché.

Pendant des années, les secteurs de la santé, des assurances et des transactions à haute fréquence ont appliqué les principes de l'apprentissage automatique afin d'analyser des quantités considérables de données d'entreprise pour prendre des

décisions de manière autonome. Au cœur de chaque implémentation se trouve un « cerveau » de traitement de données hautement extensible, capable d'appliquer des modèles mathématiques de haute précision à des quantités astronomiques de données quasiment en temps réel.

Appliquer l'apprentissage automatique à la classification des fichiers

Définition de l'apprentissage automatique

« L'apprentissage automatique, champ d'étude de l'intelligence artificielle, traite de la conception et de l'étude de systèmes capables d'apprendre à partir de données… La base de l'apprentissage automatique repose sur la représentation et la généralisation. La représentation d'instances de données fait partie de tout système d'apprentissage automatique. La généralisation est la propriété selon laquelle le système fonctionnera correctement sur des instances de données inconnues. Les conditions de garantie de la présente définition sont le principal objet d'étude du champ d'application de la théorie d'apprentissage informatisée » - Wikipédia

Au cours des dernières décennies, des milliards de fichiers ont été créés, certains malveillants, d'autres non. Pendant cette période, des schémas ont émergé et influencé la manière de créer des types spécifiques de fichiers. Ces schémas peuvent subir des variations ainsi que des anomalies, mais en règle générale, le processus informatique reste raisonnablement cohérent.

Les schémas deviennent encore plus cohérents, si l'on regarde différentes sociétés de développement telles que Microsoft®, Adobe® et d'autres grands éditeurs de logiciels. La cohérence est également présente dans les processus de développement utilisés par des développeurs spécialisés autant que par les cybercriminels. Le défi est le suivant : identifier les schémas, comprendre la manière dont ils se manifestent sur des millions d'attributs et de fichiers, et reconnaître les schémas cohérents qui nous informent sur la nature de ces fichiers.

En raison du volume de données impliquées, la tendance à la partialité et le nombre de calculs nécessaires, les humains sont incapables de tirer profit de ces données pour déterminer si un fichier est malveillant ou non. Malheureusement, la majorité des entreprises utilisent toujours le facteur humain pour ces déterminations. Elles embauchent de nombreux salariés qui vont parcourir des millions de fichiers pour déterminer les fichiers valides et les fichiers malveillants.

Les êtres humains n'ont ni les ressources intellectuelles ni l'endurance physique pour faire face au volume colossal et à la sophistication des menaces modernes. Des avancées ont été recensées sur l'analyse des vulnérabilités et des comportements, ainsi que sur l'identification des indicateurs de compromission, mais ces « avancées » présentent toutes le même inconvénient majeur. Elles sont toutes basées sur la perspective et l'analyse humaine d'un problème, et les humains se trompent en cherchant toujours à tout simplifier.

Les machines, en revanche, sont impartiales.

Fonctionnement

L'apprentissage automatique et le data mining vont de pair. L'apprentissage automatique se concentre sur la prédiction basée sur des propriétés apprises à partir de données antérieures. C'est ainsi que Dell fait la différence entre les fichiers malveillants et les fichiers sûrs ou légitimes. Le data mining se concentre sur la recherche de propriétés de données inconnues auparavant, afin que ces propriétés puissent être utilisées dans de futures décisions d'apprentissage automatique.

L'apprentissage automatique utilise un processus en quatre phases : collecte, extraction, apprentissage et classification.

Collecte

Tout comme une analyse ADN ou une étude actuarielle, l'analyse de fichiers commence par la collecte d'énormes volumes de données, et dans ce cas, des fichiers de types spécifiques (exécutables, .pdf, .doc. Java, flash, etc.). Des centaines de millions de fichiers sont collectés via les flux issus des sources du secteur, des référentiels structurels propriétaires et des informations en direct tirées d'ordinateurs actifs équipés d'agents Dell1

L'objectif de la collecte est de s'assurer de :

- Disposer d'une taille de fichiers statistiquement significative.
- Disposer de fichiers qui couvrent la plus large variété possible de types et d'auteurs de fichiers (ou groupes d'auteurs tels que Microsoft, Adobe, etc.)
- NE PAS avoir de collecte biaisée due à une collecte surabondante de types de fichiers spécifiques

Une fois ces fichiers collectés, ils sont étudiés et classés en trois catégories : « connu et valide », « connu » et « malveillant et inconnu ». Il est essentiel de garantir la fiabilité de ces catégories : l'ajout de fichiers malveillants dans la catégorie valide ou de fichiers valides dans la catégorie malveillant introduirait de la partialité.

Extraction

L'étape suivante du processus d'apprentissage automatique est l'extraction d'attributs. Ce processus est fondamentalement différent du processus d'identification des comportements ou d'analyse des programmes malveillants actuellement utilisé par les personnes qui cherchent les menaces.

Plutôt que de chercher des éléments évocateurs de malveillance, Dell utilise la capacité de calcul des machines et les techniques de data mining pour identifier le plus de caractéristiques possibles d'un fichier. Ces caractéristiques peuvent être basiques, comme la taille d'un fichier PE ou l'outil de compilation utilisé, ou complexes, comme l'étude du premier niveau de l'arborescence logique du système binaire. Dell extrait les caractéristiques atomiques uniques du fichier selon son type (.exe, .dll, .com, .pdf, .java, .doc, .xls,.ppt, etc.).

En identifiant un ensemble d'attributs le plus large possible, Dell supprime la partialité introduite par la classification manuelle des fichiers. Avec l'utilisation de centaines de milliers d'attributs, les cybercriminels doivent faire face à des coûts bien plus élevés pour créer un logiciel malveillant non détecté par Dell.

Le résultat de ce processus d'identification et d'extraction des attributs est la création d'un génome de fichiers très similaire à celui utilisé par les biologistes pour créer un génome humain. Ce génome devient ensuite la base à partir de laquelle des modèles mathématiques peuvent être créés pour déterminer les caractéristiques attendues des fichiers, tout comme l'analyse d'ADN humain est utilisée pour déterminer les caractéristiques et les comportements des cellules.

Apprentissage et formation

Une fois les attributs collectés, le résultat est normalisé et converti en valeurs numériques qui peuvent être utilisées dans les modèles statistiques. C'est à ce moment que la vectorisation et l'apprentissage automatique sont appliqués pour supprimer les erreurs humaines et accélérer le processus analytique. En exploitant les millions d'attributs de fichiers identifiés au cours de l'extraction, les mathématiciens Dell développent ensuite des modèles statistiques qui prédisent avec précision quels sont les fichiers valides et quels sont les fichiers malveillants.

Des dizaines de modèles sont créés avec des mesures clés afin de garantir la précision prédictive des modèles finaux utilisés. Les modèles inefficaces sont supprimés et les modèles efficaces passent plusieurs niveaux de tests. Le premier niveau commence par quelques millions de fichiers connus et les dernières étapes incluent l'ensemble du corpus de fichiers (des dizaines de millions de fichiers). Les modèles finaux sont ensuite extraits du corpus de test et chargés dans l'environnement de production pour servir à la classification des fichiers.

Il ne faut pas oublier que pour chacun des fichiers, des milliers d'attributs sont analysés pour différencier les fichiers légitimes des programmes malveillants. C'est ainsi que le moteur Dell identifie les logiciels malveillants (compressés ou non, connus ou non) et atteint un niveau inégalé de précision. Il divise un seul fichier en un nombre astronomique de caractéristiques et analyse chacune d'entre elles en tenant compte de centaines de millions d'autres fichiers pour prendre une décision concernant la normalité de chaque caractéristique.

Classification

Une fois les modèles statistiques créés, le moteur peut être utilisé pour classifier les fichiers inconnus (par exemple, des fichiers qui n'ont encore jamais été vus ou analysés par d'autres listes blanches ou noires). Cette analyse ne prend que quelques millisecondes et est extrêmement précise, en raison de l'ampleur des caractéristiques de fichiers analysées.

L'analyse se faisant à l'aide de modèles statistiques, la classification n'est pas exécutée dans une boîte noire. Dans le cadre du processus de classification, Dell fournit à l'utilisateur une « note de confiance ». Cette note offre à l'utilisateur des informations incrémentielles qu'il pourra utiliser pour réfléchir aux mesures à prendre pour un fichier spécifique : bloquer, mettre en quarantaine, surveiller ou analyser plus en détail.

La distinction entre l'approche de l'apprentissage automatique et l'approche traditionnelle de recherche de menaces est substantielle. Avec l'approche mathématique, nous créons des modèles qui déterminent spécifiquement si un fichier est valide ou malveillant. La réponse pourra également être « suspect » si la confiance vis-à-vis de ses intentions malveillantes est inférieure à 20 % et s'il n'y a pas d'autre indication d'intention malveillante. Ainsi, l'entreprise bénéficie d'une perspective holistique des fichiers exécutés dans son environnement. Ce fonctionnement permet également de supprimer la partialité existant actuellement dans le secteur : en effet, ceux qui recherchent les menaces ne font que déterminer la malveillance d'un fichier et les fournisseurs de liste blanche ne déterminent que la validité d'un fichier.

Outre les avantages évidents liés à la détection de nombreuses menaces, d'autres avantages plus subtils sont à l'œuvre avec cette approche : chaque fichier analysé est évalué à l'aide d'algorithmes de classification. Même si cela peut paraître simple, les fournisseurs d'antivirus traditionnels évaluent seulement un fichier spécifique par rapport à une liste limitée de signatures conçue pour détecter les programmes malveillants sur la base d'une analyse humaine. Bien qu'ils utilisent certaines techniques automatisées, ils sont limités en termes de création de signatures, car certains éléments spécifiques de fichiers ont déjà été identifiés par quelqu'un d'autre comme un logiciel malveillant connu. Non seulement ces techniques ne laissent que peu ou pas de place à la proactivité, mais elles classent tout simplement des objets dans la catégorie valide à partir du moment où ils ne correspondent à aucune signature particulière. À l'inverse, le moteur Dell analyse chaque fichier et donne une classification définitive pour chacun d'entre eux : malveillant, suspect ou validé. Ainsi, l'équipe chargée de la sécurité aura une idée très claire des fichiers exécutés dans leur environnement.

Sécurité sur le long terme

En appliquant les modèles mathématiques au point de terminaison, le moteur Dell de prévention des menaces avancées surpasse largement l'ensemble des méthodes traditionnelles de détection et de prévention contre les logiciels malveillants. L'objectif est d'arrêter l'exécution des fichiers malveillants avant qu'ils provoquent des dégâts. Avec cette approche, le point de terminaison reste sécurisé et inviolé, même si le fichier est stocké sur le disque.

Prévention des menaces avancées Dell

La solution Dell Data Protection | Endpoint Security Suite Enterprise est le produit d'entreprise phare de la société et elle exploite toute la puissance du moteur de prévention des menaces avancées pour empêcher l'exécution des menaces avancées en temps réel sur chaque point de terminaison de l'entreprise.

Principales fonctionnalités :

- Protection et détection de menaces avancées auparavant indétectables
- Non tributaire du Cloud pour les environnements sensibles
- Pas de mises à jour .DAT quotidiennes, ce qui supprime le besoin de connexion permanente
- Impact sur les performances particulièrement faible au lancement de l'exécution pour réduire considérablement la surcharge
- Facile à gérer avec une console Web simple

Dell propose une détection et une prévention des programmes malveillants en temps réel. Cette solution analyse les fichiers potentiellement dangereux et susceptibles d'exécuter des logiciels malveillants sur le système d'exploitation (SE) et dans les couches de mémoire afin d'éviter des charges malveillantes. La protection de la mémoire est conçue pour ne nécessiter que très peu d'actions afin de ne pas impliquer de surcharge au niveau des performances. Au contraire, la protection de la mémoire renforce les fonctions de protection de base du système d'exploitation, telles que DEP, ASLR et EMET en offrant une couche supplémentaire pour détecter ou refuser certains comportements très souvent utilisés par les exploits (fichiers de vulnérabilité).
Ces deux principales fonctions sont prises en charge par un éventail de fonctions secondaires nécessaires au bon fonctionnement de l'entreprise, notamment :

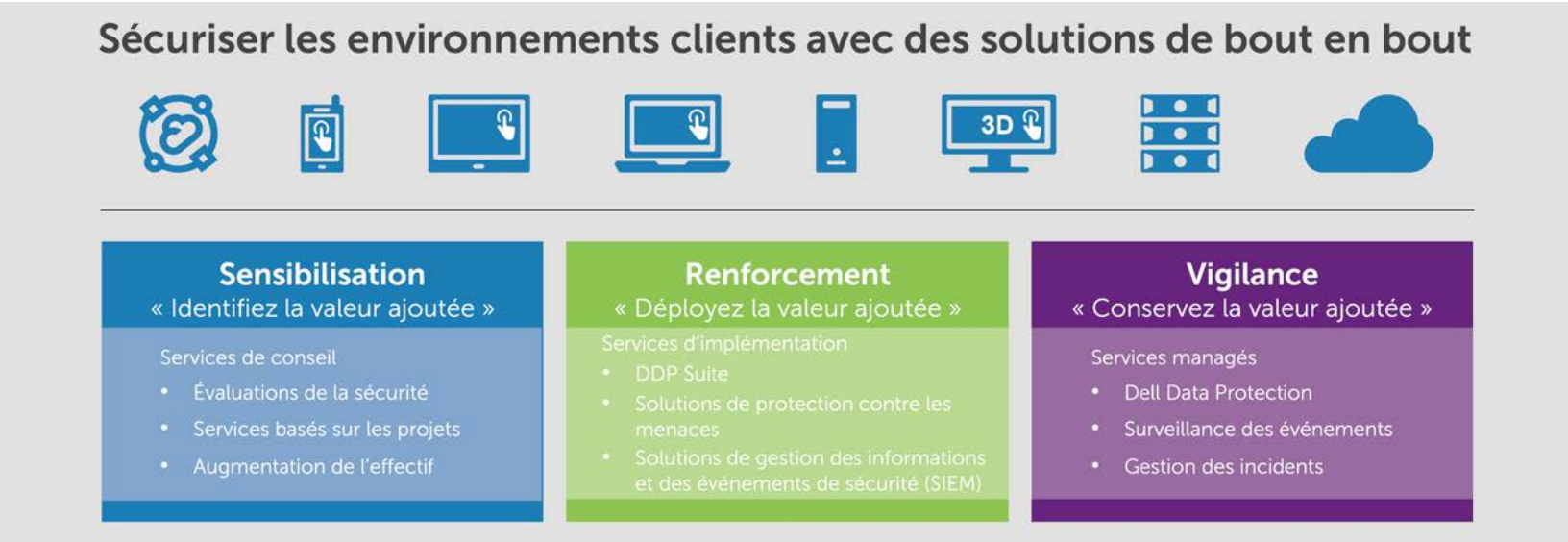
- Prise en charge des listes blanches et noires pour une granularité administrative
- Mode de détection uniquement (mode audit)
- Auto-protection (prévention contre la falsification d'utilisateur)
- Rapports complets sur le contrôle, la souplesse de configuration et la conformité depuis la console de gestion

Services professionnels de cybersécurité Dell

Dell vous propose des solutions de bout en bout pour identifier et résoudre des risques de sécurité dans votre environnement. L'équipe Dell de professionnels de la sécurité vous aidera à détecter les risques de sécurité, à implémenter des solutions et à sécuriser votre entreprise. La suite de services Dell a été conçue en tirant parti d'une expertise reconnue, des efforts de collaboration et des situations d'urgence.

Les principaux services offerts sont les suivants :

- Évaluations
- Services d'implémentation
- Services managés



Synthèse

Dell est persuadé que la modélisation mathématique et l'apprentissage automatique sont les clefs d'un avenir sécurisé. Chaque produit et service proposé par Dell est étroitement intégré à son moteur de prévention des menaces avancées, en offrant une précision et des informations inégalées sur le paysage de menaces moderne. Mieux encore, avec l'apprentissage et la formation continus basés sur les nouvelles données, le moteur Dell fonctionne vraiment sur le long terme et ne perdra pas en efficacité dans le temps, même si les cybercriminels changent leurs stratégies.

Pour plus d'informations sur les solutions de points de terminaison Dell qui vous aident à protéger vos données et à éviter les menaces, consultez le site Dell.com/DataSecurity.

[1] Les fichiers sont uniquement téléchargés à partir d'ordinateurs actifs si le client décide d'activer cette option.

Advanced Threat Prevention
POWERED BY
CYLANCE