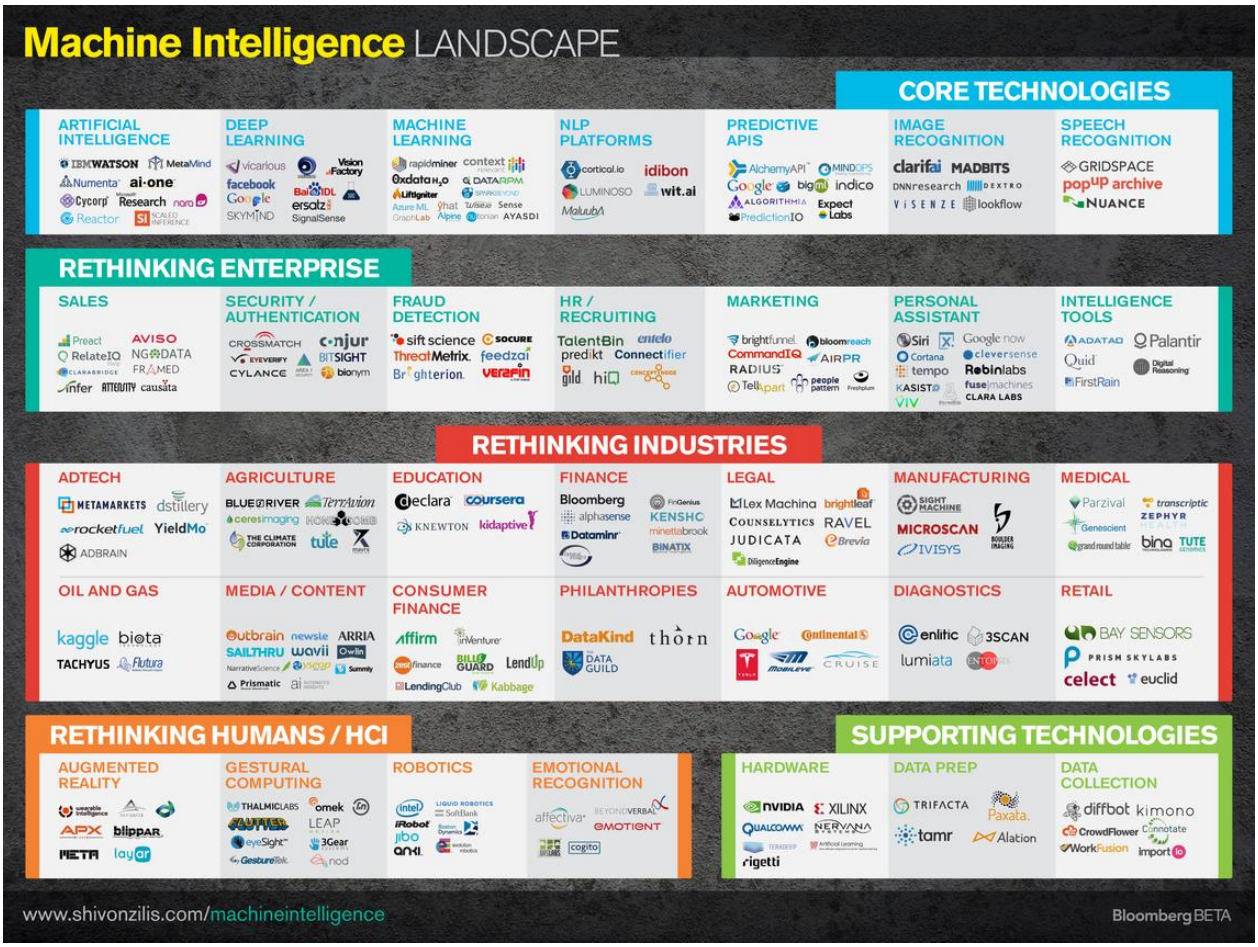
François Bourrée H5

Sécurité informatique et intelligence artificielle

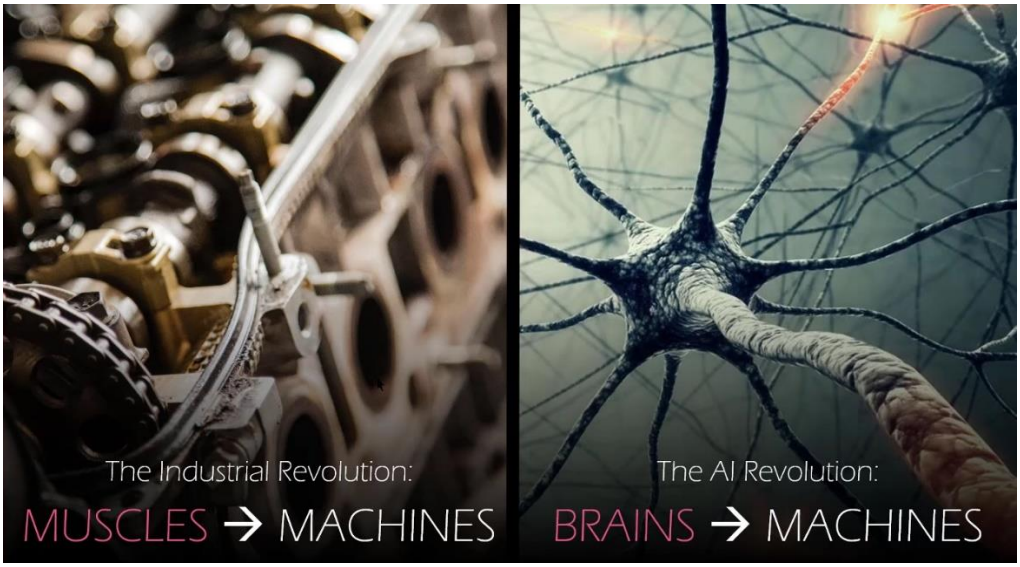
Il est tout d'abord nécessaire de définir ce qu'est l'« Intelligence Artificielle » ou « IA ». L'IA est l'ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence.

Actuellement, des sommes colossales sont investies dans le domaine de l'intelligence artificielle. Tous les pans de la société sont ou seront affectés dans l'avenir par cette technologie. Il est tout à fait légitime de penser que l'IA sera le moteur de la prochaine révolution industrielle.



Les domaines où se développe l'IA

Si l'on regarde la précédente révolution industrielle, il était question de remplacer la force physique humaine par celles de machines, plus puissantes, plus précises et ne se fatiguant pas. Désormais, on peut penser que la prochaine révolution industrielle visera à remplacer la force de réflexion humaine par celles de l'IA, plus intelligentes sur certains aspects, plus rapides, ne se fatiguant pas et capable de résonner à des échelles nous échappant.



On pourrait être tenter de penser que cela est loin mais en réalité, l'IA est déjà présente dans notre environnement, : prédiction d'achat, traitement d'images, reconnaissance vocale, voiture autonome, calcul d'assurance risque régissant le secteur financier, diagnostiques médicaux…

Les solutions technologies se reposant sur l'IA nécessitent 2 prérequis importants. Premièrement, une très grande quantité de données, au point de couvrir tous les champs possibles d'un problème. Secondement, il faut à la fois une expertise pour l'IA, des gens comprenant le fonctionnement des mathématiques et sachant paramétrer les algorithmes, et une expertise du domaine concerné.

En effet, avec la technologie actuelle, nous sommes encore loin de développer une IA capable d'apprendre entièrement d'elle-même pour résoudre un problème. Il est encore nécessaire d'avoir quelqu'un connaissant le domaine. Par exemple pour faire de la reconnaissance de voix, il fat quelqu'un capable de comprendre la structure du langage humain, de la sémantique. Il en est de même pour la reconnaissance d'image où il faut quelqu'un capable de comprendre la photographie numérique.

Et si l'on veut appliquer l'IA à la sécurité informatique, il faut quelqu'un maitrisant l'univers de la sécurité informatique.

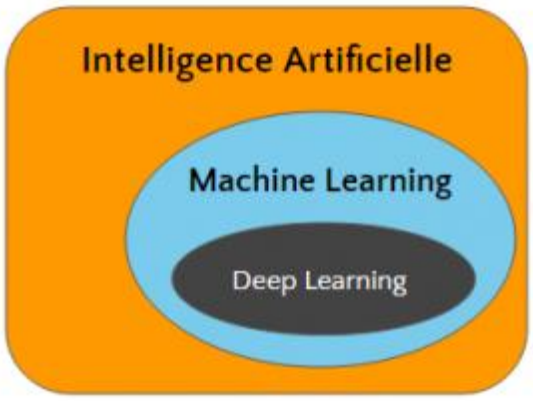
Les problèmes majeurs du développement de l'IA au sein de la sécurité informatique sont d'un coté le manque de données et de l'autre le manque d'expertise.

L'accès aux données de sécurité est restreint, surtout pour les petites entreprises, et il n'y a pas de données pertinentes et disponibles dans le domaine publique. Ainsi il est difficile de pouvoir entrainer les algorithmes par manque de données à leurs injecter. Même les grosses entreprises ont des difficultés à récolter assez de données car les clients sont souvent frileux de partager leurs données.

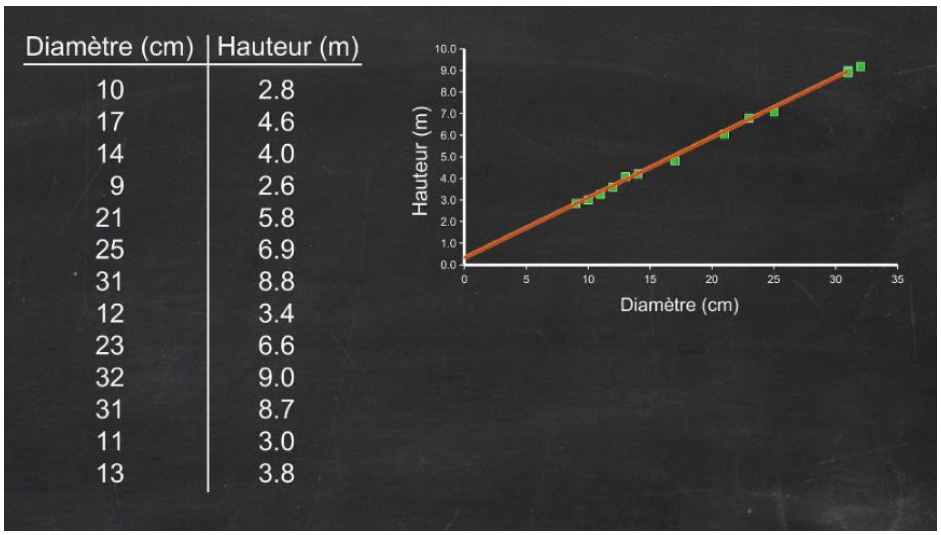
L'autre problème intrinsèque des systèmes se reposant sur l'IA est que le résultat proposé est souvent obscur pour l'utilisateur. On appelle cela « l'effet boite noire » : on ne sait pas comment l'algorithme arrive à ce verdict. Ce qui fait qu'on est obligé de faire confiance « à l'aveugle » à l'IA que le résultat est le bon. Cela aurait pu être acceptable, mais les systèmes fonctionnant sous IA sont connus pour avoir un taux de fausse détection assez élevé. Or dans le domaine de la sécurité informatique, une fausse détection ou une non-détection peut avoir de lourdes conséquences.

Et pourtant, malgré toutes ses difficultés l'IA est en train de bouleverser le domaine de la sécurité informatique. Pourquoi ? Parce que l'IA permet de traiter les problèmes à une échelle sans précédent. Elle permet désormais d'automatiser des tâches réalisables auparavant uniquement par des professionnels talentueux. Ces derniers étant en nombre peu élevé, cela réduit de facto l'ampleur des opérations réalisables. Il est maintenant possible d'aligner la quantité de données récoltées avec une puissance d'analyse de grande échelle.

On peut noter que contrairement à l'apprentissage supervisé, le Deep Learning « Apprentissage Profond » permet de se passer de l'expertise du domaine. Le « Machine Learning » ou « Apprentissage Automatique » est un champ d'étude de l'IA qui étudie comment des algorithmes peuvent apprendre en étudiant des exemples. Enfin le Deep Learning est une manière spécifique de faire du Machine Learning dont on va expliquer le fonctionnement.



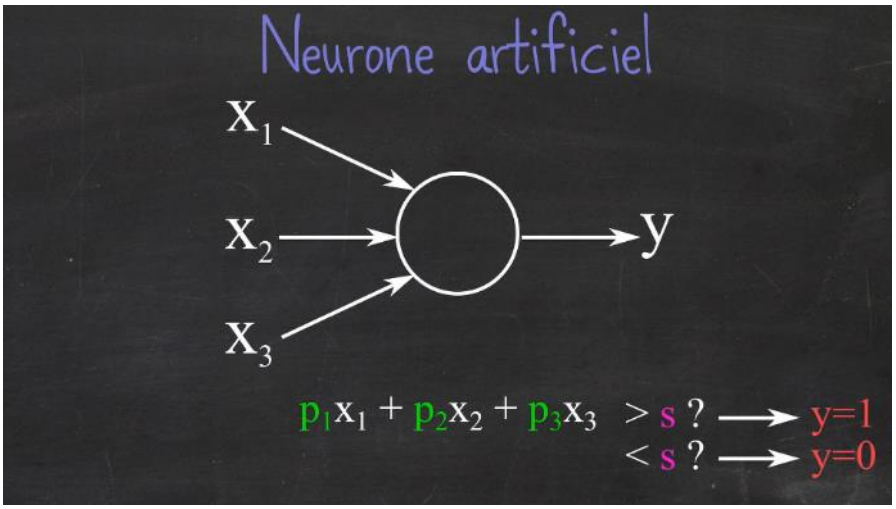
Pour permettre à un algorithme d'apprendre, on va lui fournir des données, découvrir un lien entre ces données et en généralisant, faire des prédictions sur les futures données.



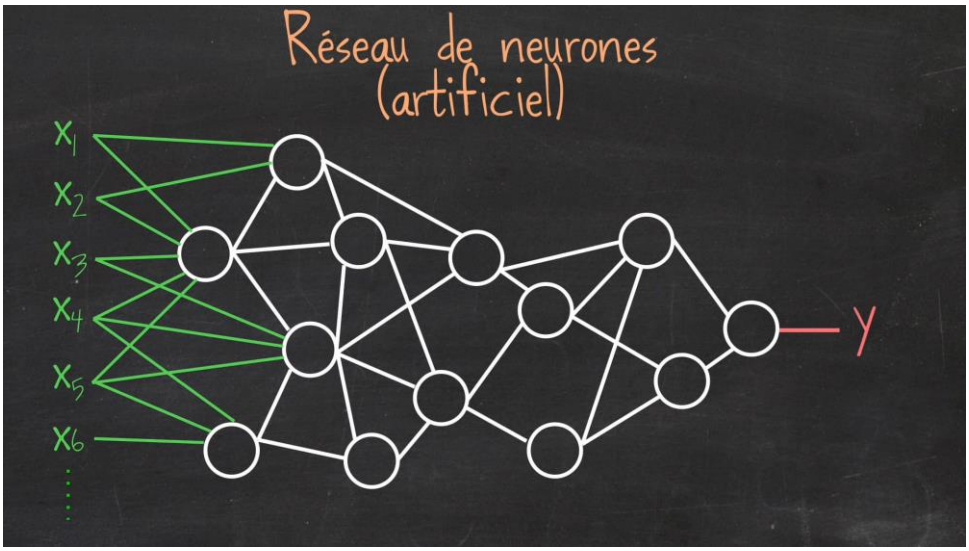
Par exemple, les étapes ici sont :

- De relever les données
- Etablir un lien (ici une simple droite)
- Prévoir la valeur attendue pour une donnée

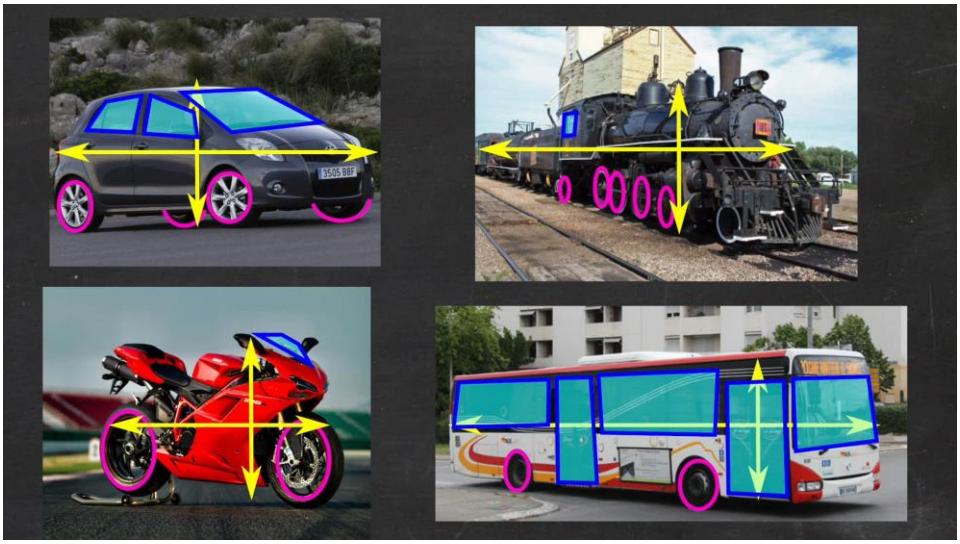
Ici le système est très simple car on possède une unique valeur en entrée et une unique valeur en sortie. Or en réalité, les entrées et sorties sont bien plus nombreuses et surtout les relations sont bien plus complexes. C'est ici qu'interviennent les neurones. Un « neurone » est une fonction mathématique, comme cette droite, qui met en relation des entrées et des sorties. Par exemple, si on choisit plusieurs paramètres d'entrée (comme x_1, x_2 et x_3) et un seul paramètre y en sortie, le neurone représente une somme dont le résultat donnera 1 ou 0 en sortie. Cependant, toutes les entrées n'ont pas forcément la même importance. Ici les coefficients p_1, p_2 et p_3 sont appelés des « poids » car ils n'ont pas le même poids dans l'équation.



Bien que cette relation soit déjà un peu plus complexe, elle reste très basique vis-à-vis des relations que l'on souhaite simuler. Sauf qu'il est possible d'en associer un grand nombre ensemble pour créer des fonctions plus compliquées. On appelle cela des « réseaux de neurones ».

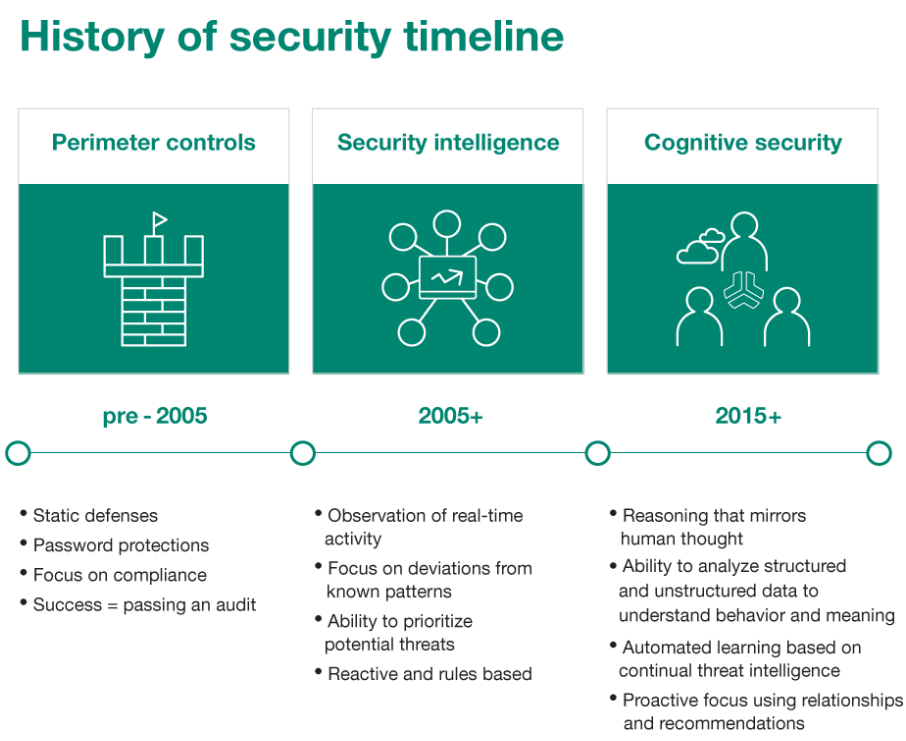


La phase d'apprentissage, celle où l'algorithme détermine un lien entre les données, devient plus longue avec un nombre de couches de neurones grandissant. Il est donc nécessaire de déterminer des caractéristiques intermédiaires qui serviront d'entrées, comme les humains le font. Pour reconnaitre les différents véhicules, on regarde le nombre de roues, la taille ou les vitres.



Maintenant que nous avons brièvement vulgarisé le fonctionnement du Machine Learning, nous allons voir de quelle manière il s'est introduit dans le monde de la cybersécurité.

Histoire de la sécurité informatique



1) Avant 2005 : Les périmètres de contrôles

La sécurité informatique a commencé avec des défenses statiques pour bloquer ou limiter les flux de données, tels que les pares-feux ou les antivirus. Le but étant de bloquer ou restreindre l'accès aux information sensibles grâce a des mots de passe ou autres contrôles d'accès. Même si ces stratégies sont encore utilisées elles ne sont plus désormais suffisantes.

2) 2005 et après : Le renseignement de sécurité

Avec le temps, on a progressé vers des systèmes plus sophistiqués et riches en données, ce qui permet de découvrir les vulnérabilités et de prioriser les attaques potentielles.

Cette transition a conduit à mettre l'accent sur l'information en temps réel pour détecter les activités suspectes. Aujourd'hui, le renseignement de sécurité est la collecte, la normalisation et l'analyse en temps réel de données structurées, produites par les utilisateurs, les applications et l'infrastructure.

Le renseignement de sécurité fait appel à l'analyse pour déceler les écarts par rapport aux tendances régulières, découvrir les changements dans le trafic réseau et trouver les activités qui dépassent les niveaux définis. En déterminant quels écarts sont significatifs, le renseignement de sécurité peut non seulement aider à détecter plus rapidement les compromis, mais aussi à réduire les faux positifs pour économiser du temps et des ressources.

3) 2015 et après : La sécurité cognitive

Les systèmes cognitifs sont des systèmes d'auto-apprentissage qui utilisent l'exploration de données, l'apprentissage automatique, le traitement du langage naturel et l'interaction homme-ordinateur pour imiter le fonctionnement du cerveau humain.

Fondée sur le renseignement de sécurité, la sécurité cognitive se caractérise par une technologie qui est capable de comprendre, de raisonner et d'apprendre. Les systèmes cognitifs capables de traiter et d'interpréter 80 % des données non structurées d'aujourd'hui, comme le langage écrit et parlé, permettent maintenant d'accéder à une échelle beaucoup plus grande de données de sécurité pertinentes.

Après avoir ingéré un corpus de connaissances, élaboré par des experts sur un sujet donné, un système de sécurité cognitive est formé par une série de questions-réponses. Les techniques de défense peuvent maintenant être entraînées à analyser des milliers de rapports de recherche, de documents de conférence, d'articles universitaires, d'articles de presse, de blogues et d'alertes de l'industrie, tous les jours.

Au fur et à mesure que les systèmes cognitifs continuent d'observer les événements et les comportements, la capacité d'utiliser des défenses intégrées pour bloquer de nouvelles menaces devient de plus en plus forte.

Un des pionniers dans ce domaine est Watson d'IBM.

L'équipe d'IBM s'appuie de plus en plus sur sa plate-forme d'apprentissage cognitif Watson pour les tâches de "consolidation des connaissances" et de détection des menaces basées le Machine Learning.

« Beaucoup de travail dans un centre d'opérations de sécurité aujourd'hui est routinier ou répétitif, et si nous pouvons en automatiser une partie en utilisant l'apprentissage automatique ? »

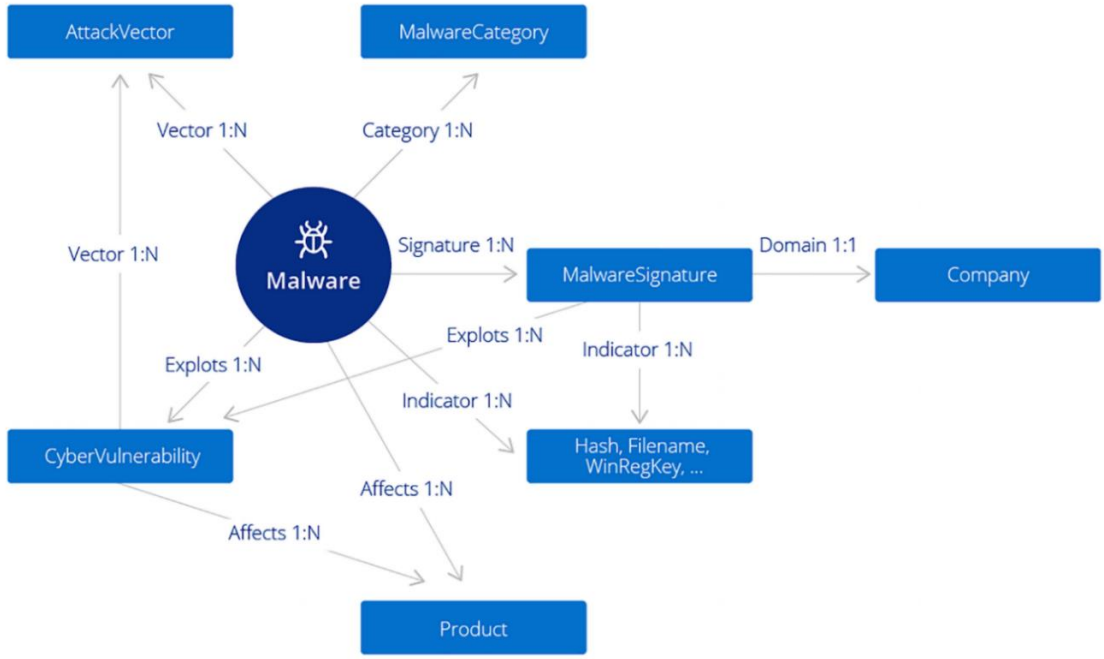
Koos Lodewijkx, vice-président et directeur des opérations techniques de sécurité chez IBM Security

Le fonctionnement

Pour comprendre comment l'intelligence artificielle et l'apprentissage machine peuvent être bénéfiques pour la cybersécurité, il est utile de comprendre comment les machines donnent un sens à de grandes quantités de données. Généralement, ils le font à l'aide de représentations de la connaissance telles que les ontologies.

En termes simples, les ontologies sont des systèmes composés de " choses " distinctes, appelées entités, et leurs relations les unes avec les autres.

Dans ce cas, un simple diagramme de Venn :

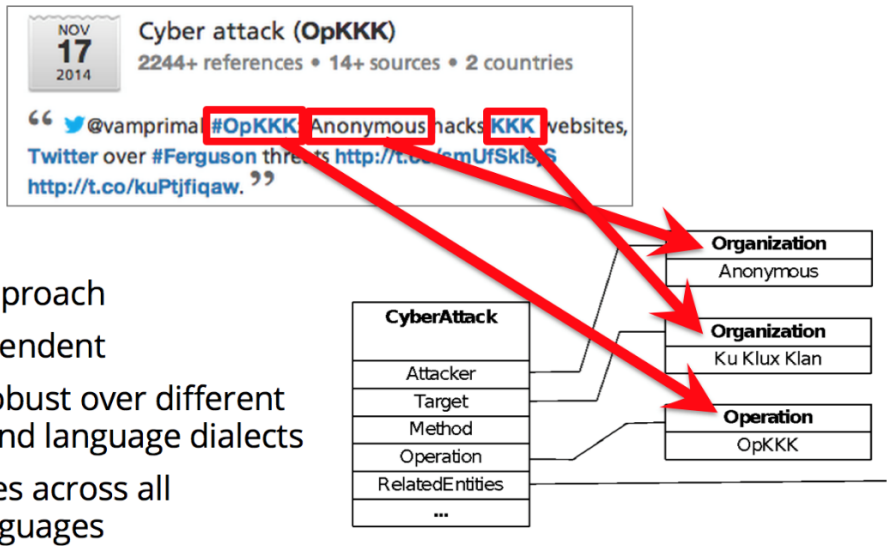


Dans cette ontologie, vous pouvez voir des logiciels malveillants se trouvant au centre, entourés de diverses autres entités qui pourraient être liées à ce logiciel malveillant. Par exemple, l'entité 'MalwareCategory' pourrait être un cheval de Troie bancaire ou un ver, et l'entité 'AttackVector' pourrait indiquer soit un spam, soit une injection SQL, soit une vulnérabilité particulière que le malware exploite.

En utilisant ce type d'ontologie, une machine peut commencer à "comprendre" le monde réel - dans ce cas, les menaces auxquelles un réseau est confronté.

Bien sûr, pour qu'une machine puisse dresser un tableau complet, elle doit consommer des données et classer tous les points de données qu'elle reconnaît qui se rapportent à ces entités particulières. A partir de là, la prochaine étape est de convertir ces entités en événements, qui à leur tour auront leurs propres classifications, par exemple, l'attaquant, ou la cible, ou la méthode.

Dans l'image ci-dessous, vous pouvez voir un exemple de la façon dont cela pourrait fonctionner dans la pratique. Le point de données, en l'occurrence un message sur Twitter, est analysé à l'aide d'une approche simple basée sur des règles.



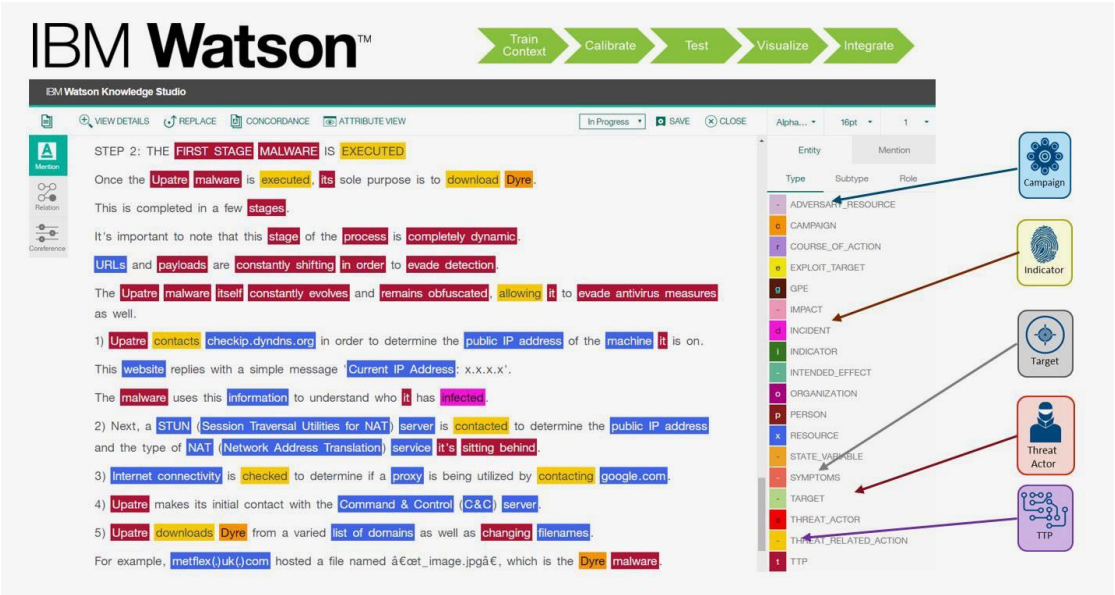
- Rule-based approach
- Language dependent
- Surprisingly robust over different media types and language dialects
- 25+ event types across all supported languages

De manière générale, on utilise des datasets, c'est-à-dire un ensemble de données structurées, pour fournir les nombreuses données nécessaires à l'entrainement.

Origin	Destination	Packet length	Payload length	TPC Port	Source IP	Destination IP	Label
Germany	Portugal	420	220	25	143.7.13.20	145.17.14.90	Normal
Germany	France	512	254	25	143.7.13.20	143.7.13.20	Normal
Portugal	Portugal	512	220	80	152.7.13.24	152.7.13.24	Normal
Russia	USA	1024	400	80	129.7.13.23	129.7.13.23	Intrusion
USA	Spain	820	416	465	97.7.13.21	97.7.13.21	Normal
France	France	718	512	465	13.77.13.22	13.77.13.22	Normal

Ici les données sont structurées.

Voici un exemple pour Watson où les données ne sont pas structurées :



La vaste bibliothèque de recherche d'IBM facilite la tâche. Mais ce n'est pas aussi simple que de montrer à Watson un tas d'articles et de documents de recherche. Il faut lui apprendre ce que tout signifie, avant qu'il ne puisse apprendre lui-même comment ils interagissent.

Pour aider Watson à démarrer, les chercheurs d'IBM annotent manuellement les documents qui entrent dans son système (pour l'instant, ils sélectionnent manuellement les documents et les sources). Au fur et à mesure que Watson commencera à maîtriser certains concepts et démontrera qu'il est capable d'annoter par lui-même, ils accéléreront le processus, avec l'aide d'étudiants de huit universités américaines. Au cours de cette première phase de formation, Watson ingérera jusqu'à 15 000 documents de sécurité par mois, en se connectant à diverses bibliothèques et flux d'informations pour s'assurer qu'ils restent à jour. Si un superordinateur peut le faire, Watson peut le faire.

Entrainement

Un exemple connu de ML et d'apprentissage supervisé est une solution de sandboxing pour identifier les logiciels malveillants ou les domaines malveillants. L'algorithme extrait les caractéristiques pertinentes, attribue de l'importance à chacune d'elles et peut ensuite prédire si le fichier d'entrée inconnu est malveillant ou bénin sur la base d'une valeur en pourcentage comme indiqué dans le tableau.

	Document valide	Document malveillant
Nom du document	0,9%	84,4%
Sans titre	6,6 %	50,2 %
Code impénétrable	0,1%	39,6%
Accède au fichier hôte	21,8%	49,5%
Résolution DNS	27,4 %	50,4 %

Il existe beaucoup d'autres fonctionnalités que l'on peut examiner comme :

- API accédées

- Champs accédés sur le disque

- Accès aux périphériques (caméra, clavier, etc.)

- Puissance consommée du processeur

- Consommation de bande passante

- Quantité de données transmises sur Internet

- Réutilisation de codes provenant de virus déjà détectés

Ce mode d'entrainement peut cependant entrainer un nombre élevé de fausses alertes (faux positifs). Des comportements jamais vus avant mais pourtant légitimes peuvent être reconnues comme des

anomalies. D'un autre côté, il est possible de rater des attaques (faux négatifs). Il est donc nécessaire de trouver un juste équilibre.

Voici un tableau récapitulatif :

Detection Rate vs False Alarm Rate

Standard metrics for evaluations of intrusions (attacks)

Standard metrics		Predicted connection label	
		Normal	Intrusions (Attacks)
Actual connection label	Normal	True Negative (TN)	False Alarm (FP)
	Intrusions (Attacks)	False Negative (FN)	Correctly detected intrusions - Detection rate (TP)

Un autre exemple est d'analyser les caractéristiques des en-têtes HTTP pour identifier des modèles uniques de comportement de commande et de contrôle qui n'existent généralement pas dans le trafic de données normal. Comme condition préalable, les spécialistes des données et les ingénieurs de sécurité doivent analyser un large éventail de trafic de commande et de contrôle et se concentrer sur les caractéristiques communes à de nombreux types de logiciels malveillants.

Ces informations sont introduites dans l'algorithme d'apprentissage et un modèle est généré qui peut ensuite prédire si une communication de commande et de contrôle basée sur HTTP se produit. Au lieu d'essayer de suivre les attaquants lorsqu'ils changent de domaine et d'adresse IP, ce modèle détecte rapidement les communications de commande et de contrôle sans utiliser de signatures.

Les cas d'utilisation classiques de ML avec apprentissage non supervisé sont basés sur le principe de la recherche de groupes logiques pour identifier les valeurs aberrantes des normes locales. Après une période de référence, un trafic réseau anormal provenant d'un hôte peut être un indicateur d'activité malveillante. Deuxièmement, l'identification de l'accès aux ressources auxquelles un utilisateur ou un hôte n'a généralement pas accès pourrait également présenter des valeurs aberrantes par rapport aux normes locales. Un troisième exemple est le modèle de comportement qui est trop régulier pour un humain. Il est essentiel que les valeurs aberrantes ne signifient pas nécessairement des incidents de sécurité. Il dit plutôt qu'il faut faire enquête et mettre à jour les données de référence.

Utilisation

Par exemple, l'IA est couramment utilisée dans la cybersécurité aux fins suivantes :

Reconnaissance de patterns - Identifier les courriels de phishing en fonction du contenu ou de l'expéditeur (ou en utilisant des modèles probabilistes de réputation, d'attaques), identifier les logiciels malveillants, filtrer les spams, les pages web ou emails au contenu inapproprié

Détection d'anomalies - Détection d'activités, de données ou de processus inhabituels (détection de fraude pour les services bancaires en ligne ou le jeu, détection de réseau zombies)

Détection d'intrusions – fichier Javascript ou autre script malveillant, fichier exécutable (ou non). Contrairement aux documents qui doivent respecter un certain nombre de règles, sont limités par le système d'exploitation, les exécutables sont autorisés à faire bien plus de choses. Cependant, un domaine d'expertise existe et est capable de comprendre comment les hackers agissent. Par exemple, ils n'écrivent rarement voire jamais le programme du début à la fin. Ils réutilisent plutôt des morceaux de codes déjà existant ou une logique similaire dans le but de faire la même action. On utilise alors ces similarités pour identifier si le programme est malveillant.

Traitement du langage naturel - Conversion de texte non structuré tel qu'une page Web en intelligence structurée (comme le fait Watson)

Analyse prédictive - Traitement des données et identification des patterns afin de faire des prédictions et d'identifier les valeurs aberrantes. Par exemple, si l'on détecte un IOC (Indicateur de compromission : ensemble de données sur la menace), comme une url malveillante, un humain voudra chercher d'autres IOC similaires. Cela peut être des URL enregistrées par la même personne, dans la même fenêtre de temps, utilisant une structure lexicale similaire. L'intuition nous guidera à nous dire que ces url sont semblables à la première. Il est désormais possible d'automatiser cette réflexion et l'appliquer non pas sur quelques IOC mais jusqu'à des millions de IOC. Ce changement d'échelle permet d'attribuer une attaque isolée à une campagne d'attaque globale. Toutes ces informations permettent d'entrainer des IA sur la prédiction non pas d'attaques mais de campagnes d'attaques.

Un dernier exemple serait la possibilité d'analyser non plus la menace en elle-même mais d'y ajouter aussi le contexte l'accompagnant. On peut par exemple prendre en compte si l'élément vient d'une pièce jointe d'email ou un téléchargement depuis le navigateur. Si c'est un téléchargement, comment l'utilisateur est arrivé à ce lien. Peut-être que le lien vient d'un sms qu'il a reçu. Tous ces paramètres sont injectés dans un programme basé sur une IA pour aider à la détermination s'il y a menace.

Conclusion

On peut ainsi penser que le Machine Learning va permettre non plus d'identifier les menaces existantes mais aussi de prévoir les menaces futures. Il faut cependant garder à l'esprit que les outils développés servent à la fois la sécurité et les pirates informatiques. Avec le développement croissant de la puissance des logiciels, porté par l'I.A, il devient plus que jamais nécessaire d'innover sous peine de ne plus pouvoir garantir la sécurité informatique, et ce dans un monde qui s'éloigne peu à peu de l'échelle de l'Homme.