01

Février 2020

CYBELIUS REVIEW

INDUSTRIAL CYBER SOLUTIONS

La Cybelius Review : votre rendez-vous en cybersécurité industrielle

Changement d'approche sur les données pour la cybersécurité industrielle

Cybelius Review

Une étude menée par nos experts en cybersécurité industrielle.

Ce premier numéro a été réalisé par *Frédéric Planchon,* fondateur et CEO de Cybelius.

Prochaines publications

- **Numéro 1 :** Les données
- **Numéro 2 :** L'humain dans la cybersécurité industrielle
- **Numéro 3 :** Les promesses mesurées de l'AI en cybersécurité industrielle
- **Numéro 4 :** La fascination de la menace

Synthèse

La focalisation sur les données pour la cybersécurité de l'IT se heurte à la culture du fonctionnement pour les systèmes industriels. Cette focalisation est vécue comme abstraite et ne permet pas une prise en compte de la sécurité cohérente et complète. Pour remédier à cela, l'approche proposée consiste à partir des fonctions puis à aller aux données. De cette manière, les outils de sécurité des données peuvent être compris et dimensionnés convenablement, adossés à une présentation des informations explicite pour les exploitants et les automaticiens.

Cybelius utilise cette approche dans son analyse de risque **APERO** et sa sonde de détection d'intrusion **CyPRES**.

Les données

Rôle en IT

La pierre angulaire de la protection cyber « classique » consiste à identifier des données sensibles et à les protéger. Il y a plusieurs raisons à cela :

- Ce que produit un Système d'Information, c'est essentiellement de la donnée dans des fichiers ;
- La corruption ou l'effacement des données peut avoir un impact très fort sur l'activité de l'entreprise, des ventes aux finances, en passant par les communications ou les documents contractuels et réglementaires ;
- Le vol des données est également critique, pour l'évasion de secrets d'entreprise ou pour respect de la vie privée du personnel ;
- Enfin, l'IA génère aujourd'hui une abondante littérature sur les données, avec le Data Mining, et toutes les approches d'apprentissage qui se fondent sur des données en masse, ce qui contribue à une focalisation (excessive ?) sur les données.

Pour cela, les critères de protection DICT (Disponibilité, Intégrité, Confidentialité, Traçabilité / Imputabilité) sont adaptés. Plus généralement, les données d'un SI sont l'élément à protéger, l'actif essentiel au sens EBIOS, la finalité même de toutes actions de sécurité.

Perspective OT : la sécurité du fonctionnement

Lorsqu'on regarde du côté des systèmes industriels, il se trouve que les données n'ont pas cette prérogative. Un système industriel a en effet par lui-même une finalité concrète, physique : le process. Depuis qu'il existe des processus industriels, la finalité de toutes les sécurités en est le bon fonctionnement, et cela indépendamment des technologies utilisées.

Une petite anecdote : ce qui a justifié le passage de la technologie électromécanique (des armoires de relais réalisant les fonctions d'automatismes) à la technologie électronique, la justification a pour partie été le coût, mais surtout la capacité de l'électronique d'effectuer un autocontrôle et de signaler une défaillance avant qu'elle n'ait un impact. Alors qu'un relais collé, ce n'est qu'à la sollicitation du côté opposé que l'on constate le défaut.

Quant au passage au numérique, le monde industriel l'a adopté pour accepter la poussée des fabricants, qui y ont vu un nouveau marché, et pas vraiment pour les gains en prix, qui n'étaient pas démontré à l'orée des années 90 – 2000. De fait, des rangées d'armoires à relais posaient peut-être des problèmes de maintenance, mais la durée de vie se comptait en décennies, et beaucoup de personnes avaient la compétence pour intervenir sur du relayage. Le numérique apportait une flexibilité évolutive un peu meilleure, mais un gain en fiabilité espéré assez faible. Les clients ont surtout dû acheter ce que les constructeurs proposaient…. Une illustration en est le retour à l'électromécanique pour des éléments de sécurité dans plusieurs métiers (oil & gas, transport notamment), là où la défiabilisation liée à l'informatique n'est ni compensable ni tolérable.

Ce très bref historique permet aussi de comprendre pourquoi un industriel, qui n'était pas demandeur de technologie numérique, sera encore moins demandeur d'investir et de protéger le système de contrôle-commande qu'on lui a vanté, et qui présente aujourd'hui des failles de sécurité informatique alarmantes.

Au-delà de cette évolution subie, il faut comprendre qu'un système informatique n'est pas une fin en soir pour un industriel, et que les données de ce système ne sont pas non plus une finalité intéressante. C'est même plutôt un passage obligé et une contrainte peu appréciée. Mais la finalité, comme avant, c'est le bon fonctionnement de l'installation.

Ainsi, les données ne sont intéressantes, pour un industriel, que subordonnées au bon fonctionnement. Cela pose de nombreux problèmes de transposition de la sécurité IT vers la sécurité OT.

Premier point : comment relier les données au bon fonctionnement ?

Un bon fonctionnement est physique, c'est celui d'un process qui effectue ce que l'on attend de lui, soit automatiquement, soit manuellement. La fonction « lavage d'un filtre », par exemple, repose sur l'automate qui l'effectue, le programme qui a été mis en œuvre, les données d'exploitation et de maintenance qui conditionnent l'enclenchement du lavage, sa terminaison, les éléments qualitatifs et quantitatifs associés à la fonction de laver, les éléments de contexte qui génèrent des cas d'exception tels que démarrage, arrêt, panne etc.

Cette description est assez encourageante car on voit se dessiner une organisation apparemment solide pour relier les données aux fonctions : le bon fonctionnement d'une installation correspond au bon fonctionnement de chacune de ses fonctions, et pour chaque fonction on a un ensemble de données que l'on peut cerner, décrire, et qui se placent dans un petit nombre de catégories :

- Le programme,
- Les données d'exploitation,
- Les données de maintenance,
- Les éléments quantitatifs et qualitatifs,
- Les données de contexte.

Cependant, toutes les fonctions ne sont pas d'égale importance pour le bon fonctionnement global, et il n'existe pas de modèle générique qui donne une hiérarchie. Une fonction de secours, par exemple, qui permet un arrêt propre d'une machine ou d'un atelier sur incident, ne contribue pas franchement au « bon fonctionnement » à moins de l'étendre à la reprise après incident.

Ensuite, les données précitées sont à considérer dans leur existence, leur nom, leur prise en compte, leur diffusion, leur complétude. Une perte de charge trop importante pour un flux d'eau traversant un filtre déclenche le lavage du filtre : il y a donc la perte de charge, le niveau déclenchant le lavage du filtre, le programme de lavage du filtre. Il y a tous les paramètres temporels associés : la perte de charge est mesurée de manière continue, sa valeur est changeante dans le temps – et sa valeur n'a de sens qu'associée au temps, comme à peu près toutes les données d'exploitation et maintenance sur un process industriel.

Toutes ces données ne sont pas ensemble dans une base, bien rangées et prêtes à être scrutées par un algorithme. Elles sont réparties. Elles sont organisées par rapport au process, à l'architecture du contrôle-commande (est-ce le même automate qui détecte la perte de charge trop élevée, et celui qui effectue le lavage du filtre ?), à la manière dont le programme a été réalisé, etc.

Finalement, l'organisation que nous pensions solide pour relier les données au fonctionnement est un peu difficile à mettre en œuvre concrètement, à moins d'y apporter des simplifications.

Approche Cybelius

Partir des fonctions puis aller aux données

L'approche Cybelius consiste à partir des fonctions et pas des données. A partir des fonctions, on regarde quelles sont les données concernées. En effet, une fonction se modélise (et se code) avec des variables internes, mais surtout des données entrantes, des données sortantes et des paramètres. Cette approche simple, tirées de la méthodologie IDEF0 et popularisée avec SADT, permet de passer des fonctions aux données de manière très claire et parlante pour l'exploitation.

Si maintenant les données ne sont plus la finalité, et que d'autre part on part des fonctions pour aller aux données, doit-on encore utiliser le modèle DICT ? Quelles données doivent être protégées pour rester disponibles, intègres, confidentielles, imputables ?

Une réponse consiste à adapter le modèle DICT. De fait, un système de contrôle-commande a essentiellement besoin de disponibilité et d'intégrité, moyennement d'imputabilité, rarement de confidentialité. Mais surtout, pourquoi ne pas appliquer ces critères aux fonctions ? Une fonction disponible et intègre, c'est finalement une fonction « qui fonctionne bien », et nous nous rapprochons de la demande des exploitants. De plus, une fonction « qui fonctionne bien » c'est aussi l'objectif de la sûreté de fonctionnement. Et cela est très fructueux :

- Toutes les études de sûreté de fonctionnement disponibles vont permettre de mettre l'accent sur les fonctions critiques identifiées
- Les données entrantes de ces fonctions devront par nature être protégées contre une non-disponibilité ou une non-intégrité, sans quoi la fonction sera impactée en disponibilité et en intégrité ;
- Les exploitants, intégrateurs, automaticiens peuvent être impliqués de manière très naturelle dans la cybersécurité, qui s'appuie sur les études de fiabilité qui leur sont familières.

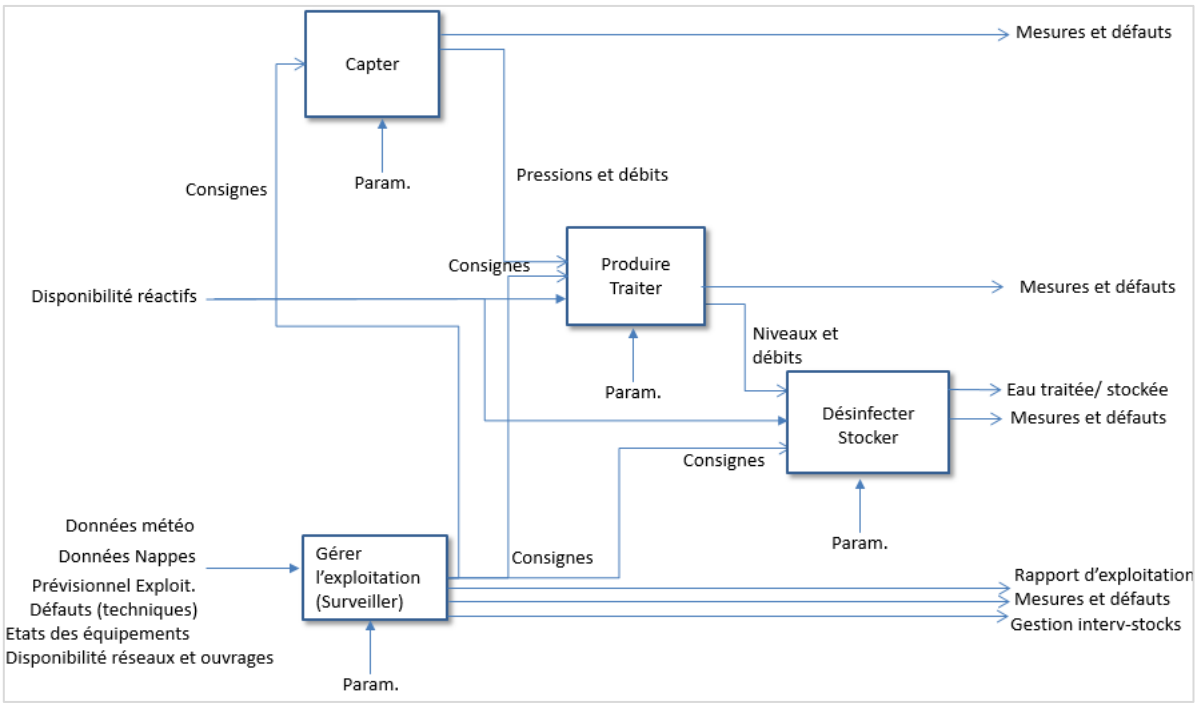
Modèle statique et dynamique fonctionnelle

Il reste cependant des obstacles. La non-intégrité des fonctions correspond-elle aux modes dégradés de la fonction tels qu'ils sont appréhendés usuellement ? L'intégrité d'une fonction, que signifie-t-elle exactement ? Ne comprend-elle pas, en plus des données d'entrée, son exécution, c'est-à-dire son programme et tous les services externes nécessaires à sa bonne exécution (firmware, OS, services réseau…) ?

C'est d'autant plus compliqué qu'une fonction, au sens de l'analyse fonctionnelle, n'est pas une fonction au sens informatique.

La réponse apportée par Cybelius se trouve à plusieurs endroits.

Pour les analyses de risque, la méthode **APERO** que nous avons développé se base sur une décomposition fonctionnelle, en reprenant le formalisme IDEF0 / SADT. Cela permet d'avoir un modèle statique associant des fonctions au sens process, à des typologies de données (mesures, commandes, etc.). Ce cadre permet de placer les besoins en sécurité sur les données entrantes, les paramètres, et la réalisation des fonctions. Voici un exemple pour le traitement de l'eau :



Pour la surveillance du bon fonctionnement, notre produit **CyPRES** intègre une approche assez pragmatique. Les fonctions sont classées en 3 catégories :

- Les fonctions principales, qui sont celles qui sont les plus courantes et qui sont vraiment les fonctions de conduite « directe » du process industriel ;
- Les fonctions secondaires : ce sont des fonctions de conduite qui sont plus rares, soit parce que leurs actions sont occasionnelles (liées au dépotage de produit chimique par exemple), soit parce qu'elles s'activent dans des conditions particulières (orage, pollution, réduction de régime, forçage…)
- Les fonctions de secours : pour des raisons de sûreté de fonctionnement, les automates intègrent tous des fonctions de secours qui ont pour but de réagir au mieux en cas d'incident.

CyPRES présente le déroulement de ces fonctions de manière temporelle avec la possibilité, pour chaque trait indiquant une fonction, de lui donner un nom parlant pour l'exploitant (par défaut, celui qui a été donné dans l'analyse fonctionnelle).

Bien que sommaire, cette présentation présente plusieurs avantages :

1- Etant graphique, elle permet de former le regard de l'exploitant à des patterns, car tous les process sont assez répétitifs
2- Elle complète assez bien une IHM de SCADA qui présente des valeurs, des synoptiques et des courbes mais qui est essentiellement sur du temps présent et assez peu sur un déroulé temporel
3- Elle présente un contexte de fonctionnement que l'on relie à une détection d'anormalité (pour **CyPRES**, cela s'appelle un symptôme) par l'instant de la détection. De fait, si quelque chose d'anormal est détecté par **CyPRES**, et qu'à ce moment on a des fonctions de secours ou des fonctions secondaires rares qui sont actives, il est probable que c'est un contexte exceptionnel que la sonde rencontre pour la première fois. L'investigation immédiate sera plutôt vers un phénomène non-cyber. Et cette participation à la levée de doute est faite par l'exploitant, qui est la personne qui connait en général le mieux son usine et les activités en cours.

Conclusion

Pour le monde industriel, il est préférable de subordonner l'analyse des données au fonctionnement du système. Cela rapproche la cybersécurité de la sécurité fonctionnelle, connue des utilisateurs OT. L'approche consistant à partir du fonctionnement du système, de manière statique et dynamique, pour aller aux données, permet d'appliquer de manière profitable les critères usuels de Disponibilité, Intégrité, Confidentialité et Traçabilité. L'analyse de risque **APERO** et la sonde de détection d'intrusion **CyPRES** de Cybelius tirent pleinement parti de ces approches.