Protection et pérennisation du patrimoine scientifique numérique d'un laboratoire de recherche

Cyril Bras

cyril.bras@cermav.cnrs.fr
CNRS/CERMAV
Service système d'information
Domaine universitaire
601, rue de la chimie
38400 St Martin d'Hères

Abstract. La production de données scientifiques numériques a explosé au cours de la décennie passée, à l'instar de ce qui a été observé dans l'usage courant de l'informatique (Demarthon, Delbecq, & Fléchet, 2011).

Dans le même temps, les menaces et les attaques informatiques sont devenues de plus en plus répandues et en être victime pour une organisation n'est qu'une question de temps (German, 2016).

Dans ce contexte, il est important pour un laboratoire de recherche scientifique de garantir l'intégrité et la pérennité des données. Pour cela, le CERMAV (Centre d'Études et de Recherche sur les MAcromolécules Végétales) et son service système d'information ont mené plusieurs actions visant à tenter de contenir la menace tout en protégeant les données.

Keywords: Cyber attaque, cyber défense, vulnérabilités, menaces, pare-feu, IDS, bigdata, cloud computing, chiffrement, imputabilité, opendata, PSSI, RGS

1 Introduction

Le CERMAV est un laboratoire de recherche fondamentale sur les glycosciences, dont il est le leader européen. Il s'agit d'une unité propre du CNRS (UPR5301), située sur le campus universitaire de Grenoble. Elle comporte un effectif moyen d'environ 120 personnes ; composée de chercheurs et enseignants chercheurs, de personnels techniques et administratifs mais aussi de doctorants, de stagiaires… Chaque année le CERMAV produit en moyenne 80 publications scientifiques et génère environ 40 To de données tant de recherche, qu'administratives qu'il convient de protéger des dommages accidentels ou des actes malveillants. Pour y parvenir, le service système d'information est intervenu sur trois axes. Le premier, concerne la gestion et la protection des flux de données. Le second a pour objectif la pérennisation des données. Enfin, le dernier volet vise à la formation et la sensibilisation des personnels aux règles d'hygiènes informatique et à la structuration des données.

adfa, p. 1, 2011.

2 Contexte

Le CNRS, dont dépend le laboratoire, a été depuis de nombreuses années, précurseur sur les aspects sécurité des SI avec la rédaction d'une première version de Politique de Sécurité des Systèmes d'Information (PSSI) en 2006 (Illand, 2006). Depuis, la prise de conscience d'un besoin fort de sécurisation des SI au niveau étatique, a été observée avec la création de l'Agence Nationale de Sécurité des Systèmes d'Information (ANSSI) en 2009 et la mise en œuvre de la PSSI de l'état en juillet 2014 (ANSSI, 2014). Cette dernière s'appliquant à toutes les entités de la fonction publique.

Le CNRS a alors défini un cadre de référence dans la protection des informations qui constituent son patrimoine universel au travers d'une Politique Générale de Sécurité de l'Information (PGSI-CNRS). Celle-ci étant déclinée dans les structures opérationnelles de recherche et de service par une Politique de Sécurité des Systèmes d'Information (PSSI) opérationnelle (CNRS, 2013). Au moment de la rédaction de cet article, le CERMAV réalise une déclinaison de PSSI spécifique au fonctionnement du laboratoire.

De par ses dimensions, le SI du CERMAV peut s'apparenter à celui d'une PME et bien qu'étant une composante du CNRS, les enjeux et besoins de sécurisation sont proches car chaque laboratoire gère sous la responsabilité de son directeur son propre SI. Toutefois, il reste spécifique en ce qui concerne ses acteurs. En effet, les chercheurs et enseignants chercheurs ont de par la nature même de leurs activités à interagir et échanger des données avec des personnes extérieures au laboratoire dans le cadre de collaborations scientifiques par exemple.

L'augmentation des attaques informatiques dans lesquelles peuvent être impliquées des États et des acteurs non-étatiques (BANNELIER & CHRISTAKIS, 2017) met bien en évidence la nécessité de protéger le SI des atteintes à la disponibilité, l'intégrité ou la confidentialité (CARPENTIER, 2016). Pour cela il est nécessaire de mettre en œuvre des axes de stratégie de sécurisation.

3 La gestion et la protection des flux de données

Comme évoqué dans le paragraphe précédent, le CERMAV doit prendre en compte à la fois des problématiques classiques de PME mais aussi des besoins en échanges de données avec des personnes étrangères au SI. Le simple filtrage de flux basé sur des ouvertures de ports et d'adresses IP n'est aujourd'hui plus suffisant ; d'autant que l'explosion des offres d'hébergement gratuite de données dans le cloud est une réalité quotidienne qu'il est alors quasiment impossible de contrôler.

Une première étape a consisté à l'analyse des flux réseau en utilisant des sondes de détection d'intrusion IDS basées sur le logiciel SURICATA. Ces sondes, dont une présentation a été faite en 2014 au cours des journées C&esar (Feuillet & Diallo, 2014), nous ont permis d'identifier les flux qui transitaient vers l'extérieur mais aussi entre les différents VLAN du laboratoire. Après plusieurs mois d'utilisation sur le

réseau du laboratoire (entre mars et septembre 2015 cela représentait environ 320 000 incidents de niveau haut, 105 000 de niveau moyen et 220 000 de niveau bas) (Bras, 2015), nous avons été en mesure de détecter plusieurs types de problèmes de sécurité et notamment des utilisations de solutions cloud publiques pour le partage de données avec l'extérieur. C'est sur cette découverte que nous allons nous attarder car c'est un exemple assez représentatif de la gestion des flux de données au sens patrimoine scientifique.

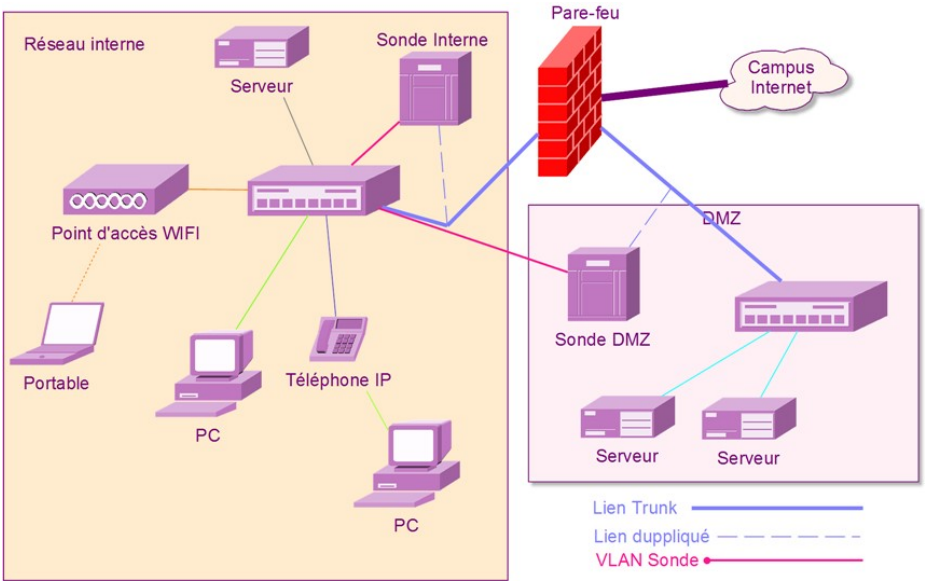

Fig. 1. Schéma de principe du réseau du CERMAV

Les mesures qui ont été prises ont alors été de mettre en service notre propre solution d'hébergement de *cloud computing*. Cette mesure ne répondait que partiellement au problème puisque les utilisateurs pouvaient continuer à utiliser les solutions publiques (ex : dropbox…). Un blocage avec un pare-feu standard (qui équipait alors le CERMAV) étant quasiment impossible du fait de la multitude d'adresses IP à bloquer pour un fournisseur de service.

Cette solution basée sur un contrôle des flux au niveau de la couche transport des modèles OSI ou TCP/IP n'est aujourd'hui plus suffisante. Il est nécessaire de s'élever jusqu'à la couche application afin d'identifier les flux, ce qu'ils contiennent réellement afin de déterminer leur légitimité et ainsi améliorer la cyberdéfense.

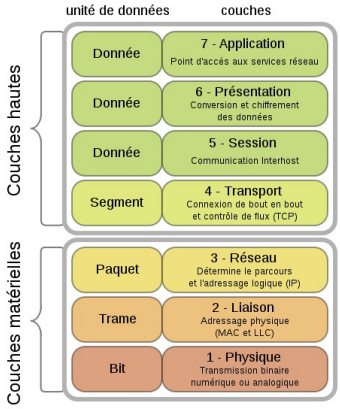

Fig. 2. Diagramme du modèle OSI. Source : Wikipédia

Pour cela en mai 2016, le CERMAV a fait l'acquisition d'un pare-feu applicatif de nouvelle génération de la marque Palo-Alto. Le choix de cette solution commerciale a été motivé par la certification de sécurité obtenue auprès de l'ANSSI sur ce type de produits (Palo Alto Networks, 2014) et la finesse des règles de filtrage. Il est alors devenu aisé d'interdire l'accès aux solutions de cloud publiques ou externes au laboratoire. La définition des règles permet de descendre jusqu'au type d'action autorisé. Par exemple, il est possible d'autoriser la consultation de contenu sur Facebook mais d'interdire l'envoi de documents ou encore d'autoriser non plus simplement une adresse IP mais une adresse IP et un utilisateur à accéder à un contenu identifié.

Un autre choix opéré par le service SI a été de trouver une solution pour garantir la légitimité d'accès d'un équipement informatique au réseau du laboratoire. Pour y parvenir, nous nous sommes appuyés sur le standard de sécurité 802.1X (Loos, 2014) et notre système d'authentification centralisé.

Les sondes, combinées au pare-feu de niveau applicatif et l'utilisation du standard 802.1x ont permis une amélioration significative dans la gestion et l'identification des flux de données mais ne peuvent pas nous prémunir totalement de cyber-attaques. Il convient donc de s'y préparer ce qui passe par la capacité à remettre le système en route ou tout du moins protéger les données de la destruction et ainsi « échapper » à l'attaque (Boyer, 2014).

4 Pérennisation des données

Dans le cas d'attaques informatiques, les données peuvent être endommagées ou détruites ce qui a forcément un impact sur l'activité scientifique d'un laboratoire de recherche (perte de publication en cours de rédaction, de données d'acquisition…). Les dernières « épidémies » de *ransomware* (Wannacrypt, Petya) ont touché de nombreux systèmes, qui n'étaient pas à jour (Mattei, 2017). La seule parade reste alors la restauration de sauvegardes (Spector, 2016).

Au CERMAV, nous avons entrepris un plan de sauvegarde du patrimoine scientifique et technique numérique. La première étape a consisté à acquérir le matériel nécessaire pour sauvegarder sans intervention humaine l'intégralité de nos données (librairie LTO-6). Le pilotage de ces matériels et la réalisation des sauvegardes se fait en utilisant le logiciel libre Bacula[1].

Du point de vue organisation, un plan de sauvegarde a été défini et permet la mise sur bande de l'intégralité des données avec une durée de rétention de 3 mois pour les données courantes et 12 mois pour les archives. Nous avons également défini un plan de test de restauration de données régulier et consigné au niveau du service SI.

Du côté poste de travail, nous avons mis en place un système de réplication automatique du dossier « documents » des postes de travail vers les serveurs en s'appuyant sur le logiciel SyncToy de Microsoft. Ce dossier « documents » a été structuré de façon automatique sur tous les postes avec un sous dossier PERSO contenant les données personnelles (non répliquées) et un sous dossier CERMAV contenant les données scientifiques et/ou techniques de chaque agent. Il a été demandé à chaque personne de structurer les données suivant un mode opératoire interne permettant un classement suivant qu'il s'agit par exemple de données d'acquisition brutes ou traitées, de communications scientifiques… Ce qui nous conduit à notre partie sur l'implication des utilisateurs dans la sécurité numérique.

5 Formation et sensibilisation à la sécurité numérique

Ce n'est pas nouveau (C, 1994), la technologie seule ne peut garantir une sécurité numérique. Il est donc essentiel d'obtenir l'adhésion des utilisateurs pour qu'ils deviennent acteurs de la sécurité informatique. L'ANSSI le rappelle d'ailleurs dans son guide d'hygiène informatique où elle indique que « Chaque utilisateur est un maillon à part entière de la chaîne des systèmes d'information. À ce titre et dès son arrivée dans l'entité, il doit être informé des enjeux de sécurité, des règles à respecter et des bons comportements à adopter en matière de sécurité des systèmes d'information à travers des actions de sensibilisation et de formation » (ANSSI, 2017).

Au CERMAV, nous effectuons des actions de formation à l'hygiène informatique à destination de tous les nouveaux entrants au même titre que la formation à la sécurité générale du laboratoire. Chaque année nous effectuons environ 10 sessions de ce genre. Pour le reste du personnel, nous communiquons lors des assemblées générales (2 fois par an) sur les nouvelles dispositions prises ou sur les nouvelles menaces informatiques avec pour objectif de transmettre une information compréhensible de tous. Le personnel est également averti lors d'alertes nationales de sécurité informatique transmises par la chaine fonctionnelle de sécurité du CNRS.

Le plus efficace reste encore la sensibilisation par l'action. Les campagnes de *phishing* et de *spear phishing* sont de plus en plus élaborées et réalistes. Nous avons décidé de mener un exercice d'attaque en *spear phishing* sur les personnels permanents du laboratoire. Le scénario était le suivant :

1. Campagne nationale au niveau du CNRS sur les dossiers de carrière des agents

[1] http://blog.bacula.org/

2. Envoi d'un courriel reprenant l'aspect habituel des courriels adressés dans le cadre des dossiers de carrière et contenant un lien vers une adresse en crns.fr. La fausse adresse étant déclarée au niveau de nos serveurs DNS pour pointer sur une machine de test. Le courriel est envoyé à 13h39, tous les agents ne sont pas encore revenus de leur pause.

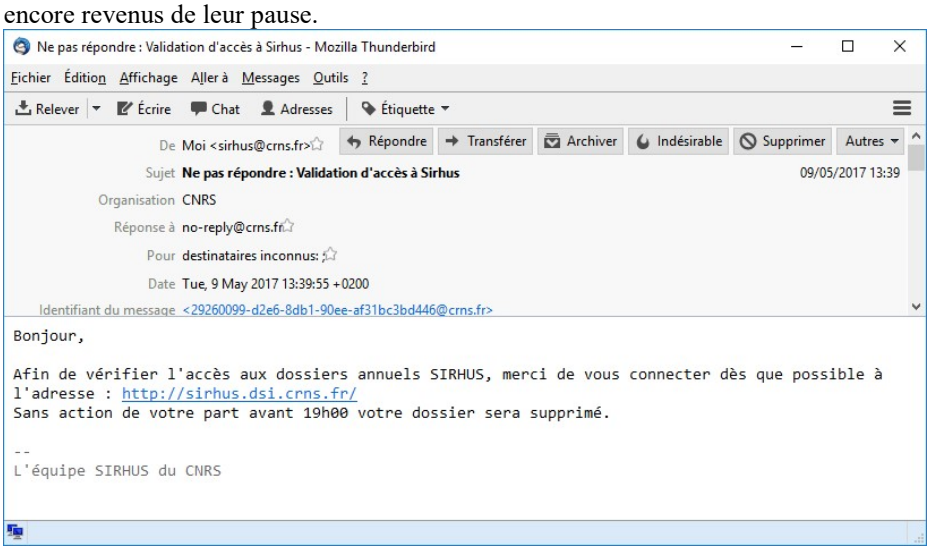

Fig. 3. Copie d'écran du courriel de *spearphishing* envoyé au cours de l'exercice

3. Quelques secondes après l'envoi 7 clics se produisent mais dans le même temps 5 autres personnes contactent le service SI pour signaler la réception de ce courriel suspect.
4. Retour sur l'exercice au cours d'une assemblée générale où tout le personnel était présent.

Cet exercice a été l'occasion d'appuyer de façon concrète la nécessité d'être vigilant lors de la réception de courriels y compris lorsque ces derniers semblent légitimes. Bien que les personnels soient sensibilisés régulièrement, on constate que si les éléments choisis par l'attaquant « parlent » à la cible, la victime baisse facilement sa garde. La pratique étant un des vecteurs à utiliser pour l'amélioration des connaissances des utilisateurs.

6 Perspectives

« Connais ton ennemi et connais-toi toi-même ; eussiez-vous cent guerres à soutenir, cent fois vous serez victorieux » (Tzu), cette recommandation ancienne s'applique sans conteste à la sécurité des SI, pourtant au moment de la rédaction de ce document, des actions d'entraînement à grande échelle impliquant le CERMAV n'ont pas été conduites. Le fournisseur d'accès au réseau Internet de l'enseignement supérieur et de la recherche (GIP RENATER) propose pourtant aux RSSI des établissements de la communauté un outil de tests de vulnérabilités mais là encore, ces outils ne sont pas

appliqués à notre laboratoire. Ceci pouvant s'expliquer par le fait que notre laboratoire n'est pas classé en Zone à Régime Restrictif (ANSSI). Cependant comme toute structure reliée au réseau mondial, nous subissons des attaques allant du simple scan réseau, aux tentatives intrusions et les classiques campagnes de courriels malveillants affectant nos utilisateurs.

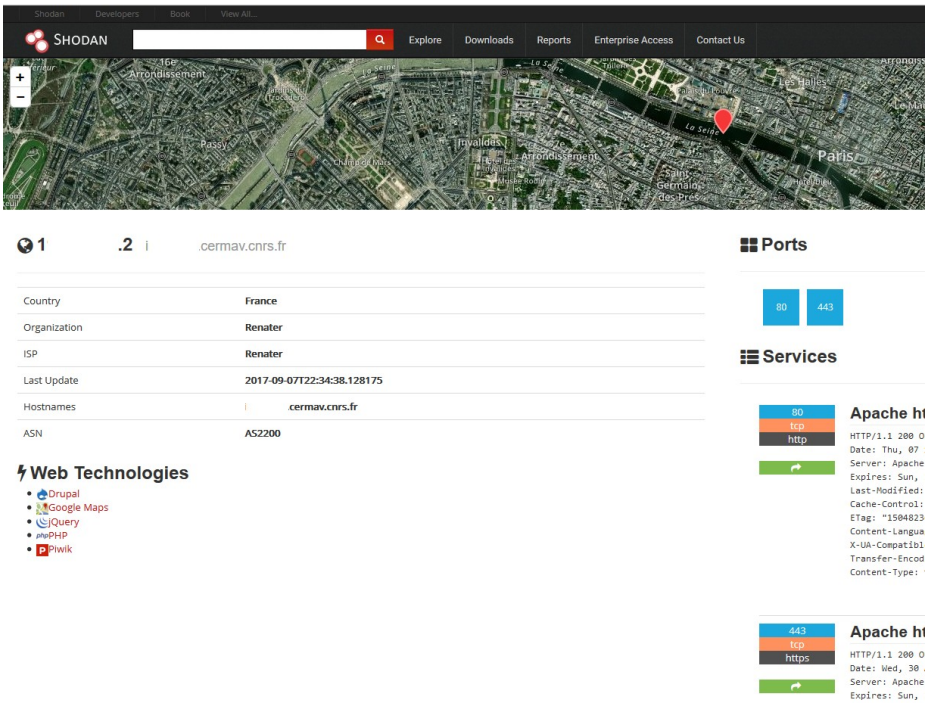

Fig. 4. Résultat d'un test effectué sur le moteur de recherche Shodan

A l'image des exercices d'évacuation incendie, il est essentiel de se préparer en confrontant son système d'information à des tests réguliers tout en essayant d'imaginer ce qu'un attaquant pourrait y faire. Au niveau laboratoire, plusieurs solutions restent utilisables pour tester le système d'information. L'appel à des outils en ligne tel le moteur de recherche Shodan (Shodan), permet d'avoir une « vue » externe de son SI et vérifier la conformité des services exposés sur Internet. La consultation des bases en ligne de comptes compromis du site « ';--have i been pwned? » (Hunt) constitue une bonne source de renseignement sur l'état de compromission éventuels des comptes utilisateurs.

Afin d'affiner la détection des anomalies une piste d'amélioration pourrait reposer sur l'utilisation d'un SIEM (Security Information and Event Management) dans le but de regrouper toutes les sources d'information de sécurité utilisables. Ceci constituera un des chantiers à mettre en œuvre au niveau du laboratoire.

Du point de vue utilisateurs, les exercices et actions de sensibilisation ont un effet bénéfique qui permet, outre l'actualité riche en incidents informatiques médiatisés (ZDNet, 2017), une prise de conscience collective sur l'importance de la sécurité informatique. Lorsque nous notifions les utilisateurs de problèmes détectés sur leur poste (comportement anormal par exemple), ces derniers se trouvent rassurés et appliquent les mesures demandées.

7 Conclusion

La maitrise des flux de données ne se limite pas à l'exploitation de solutions techniques mais nécessite d'intégrer tous les acteurs du SI, de la production jusqu'au stockage. L'organisation de la donnée ne peut se faire efficacement sans s'appuyer sur l'humain qu'il convient d'intégrer pleinement dans toutes les solutions de sécurisation pour en faire un maillon essentiel de la protection et maintenir ainsi la confiance dans l'usage des technologies numériques (Ashenden, 2016).

Bibliographie

ANSSI. (2014, Juillet 17). *La Politique de Sécurité des Systèmes d'Information de l'État (PSSIE).* Consulté le Juin 30, 2017, sur ANSSI: https://www.ssi.gouv.fr/entreprise/reglementation/protection-des-systemes-dinformations/la-politique-de-securite-des-systemes-dinformation-de-letat-pssie/

ANSSI. (2017, Janvier 23). *Guide d'hygiène informatique.* Consulté le Juillet 04, 2017, sur ANSSI: https://www.ssi.gouv.fr/guide/guide-dhygiene-informatique/

ANSSI. (s.d.). *Protection des systèmes d'informations*. Consulté le Septembre 08, 2017, sur ANSSI: https://www.ssi.gouv.fr/administration/reglementation/protection-des-systemes-informations/

Ashenden, D. (2016). The human shield. *The chemical engineer*, 22-25.

BANNELIER, K., & CHRISTAKIS, T. (2017). Cyberattaques - Prévention-réactions : rôle des Etats et des acteurs privés. *Revue Défense Nationale*.

Boyer, B. (2014). *Cybertactique : Conduire la guerre numérique.* Nuvis.

Bras, C. (2015). Surveiller un réseau de laboratoire avec Suricata. *JRES 2015.* Montpellier.

C, S. (1994, Septembre 14). *Roussel Uclaf : « La sécurité repose surtout sur les utilisateurs ».* Consulté le Juillet 04, 2017, sur Les Echos: https://www.lesechos.fr/14/09/1994/LesEchos/16729-116-ECH_roussel-uclaf-----la-securite-repose-surtout-sur-les-utilisateurs--.htm

CARPENTIER, J.-F. (2016). *La sécurité informatique - dans la petite entreprise (3ième édition).* ENI.

CNRS. (2013, Novembre 6). *Politique Sécurité des Systèmes d'Information opérationnelle applicable aux laboratoires du CNRS.* Consulté le Juin 30, 2017, sur CNRS Délégation Centre Limousin Poitou Charentes: http://www.dr8.cnrs.fr/dr_a_votre_service/sti/securite_systeme/documents/PSSIO_Laboratoires_6_novembre_2013.pdf

Demarthon, F., Delbecq, D., & Fléchet, G. (2011). La déferlante des Octets,. *CNRS le Journal n°269*.

Feuillet, M., & Diallo, D. (2014). Détection d'intrusion dans les systèmes industriels. Suricata et le cas de Modbus. *C&esar 2014.* Rennes.

German, P. (2016). Face the facts - your organization will be breached, . *Network Security*.

Hunt, T. (s.d.). *';--have i been pwned?* Consulté le Septembre 08, 2017, sur https://haveibeenpwned.com/

Illand, J. (2006). *Politique de Sécurité des Systèmes d'Information (PSSI).* Consulté le Juin 30, 2017, sur CNRS: https://aresu.dsi.cnrs.fr/IMG/pdf/PSSI-V1.pdf

Loos, J. (2014). *Implementing IEEE 802.1x for Wired Networks.* The Pennsylvania State University CiteSeerX.

Mattei, T. A. (2017). Privacy, Confidentiality, and Security of Health Care Information: Lessons from the. *Wolrd Neurosurgery News*.

Palo Alto Networks. (2014, Octobre 07). *Cyber défense Palo Alto Networks, la première solution de sécurité réseau certifiée par l'ANSSI en France.* Consulté le Juin 30, 2017, sur Palo Alto Networks: https://www.paloaltonetworks.com/company/press/2014/cyber-defense-palo-alto-networks-la-premiere-solution-de-securite-reseau-certifiee-par-lanssi-en-france

Shodan. (n.d.). *Shodan*. Retrieved Septembre 08, 2017, from The search engine for security: https://www.shodan.io/

Spector, L. (2016, Mai 06). *How to stop ransomware: Backup can protect you, but only if you do it right.* (P. World, Ed.) Retrieved Juillet 04, 2017, from PC World from IDG: http://www.pcworld.com/article/3056907/security/how-to-stop-ransomware-backup-can-protect-you-but-only-if-you-do-it-right.html

Tzu, S. (s.d.). *L'art de la guerre.*

ZDNet, L. r. (2017, Août 02). *Ransomwares : la prise de conscience des PME françaises.* Consulté le Septembre 08, 2017, sur ZDNet: http://www.zdnet.fr/actualites/ransomwares-la-prise-de-conscience-des-pme-francaises-39855710.htm