Données de santé, le nouvel El Dorado

David SYGULA
Analyste Senior en cybersécurité
CybelAngel

En décembre 2020, j'ai vécu une expérience saisissante : je suis allé chez le dentiste. Cela faisait plusieurs années - ne me jugez pas s'il vous plaît, mes dents vont bien - et je n'avais pas imaginé une telle numérisation des outils. Dès mon arrivée j'ai dû remplir un formulaire sur une tablette, qui s'est verrouillée le temps que je m'installe, et dont les questions étaient pour certaines assez personnelles. Inutile de retourner demander le code : ma tentative de bruteforce a réussi au bout d'un essai (123456). Je l'ai ensuite rendue au secrétariat et, en feignant de ne pas apercevoir le mot de passe du poste sur un petit bout de papier collé à l'écran (admin123), je demande, un peu inquiet : "Savez-vous ce qu'il advient des données de votre questionnaire, qui y a accès ? Sont-elles chiffrées ?" Regard coi.

L'ironie, c'est que quelques jours plus tôt, nous venions de publier une étude sur l'exposition mondiale des données médicales, notamment issues de cabinets de dentistes[1]. Les mois qui ont suivi, les données de santé, particulièrement françaises, ont par ailleurs fait couler beaucoup d'encre : attaques par rançongiciel de centres de soins[2], vente d'accès vers des applications d'hôpitaux[3], fuite de centaines de milliers de données de patients[4], vol d'1,4 million de résultats de tests Covid[5], etc.

Comment trouve-t-on ces données ?

Il serait facile de porter le blâme sur de mauvaises pratiques de sécurité d'utilisateurs (quels qu'ils soient), mais ne tirons pas sur l'ambulance, la réalité est beaucoup plus complexe. Comme souvent en cybersécurité, c'est l'enchaînement de plusieurs actions (ou inactions) qui permettent à une attaque d'aboutir, quand il ne s'agit pas de l'emploi d'une faille 0-day[6]. On peut par exemple y voir l'ouverture d'une pièce jointe malveillante comme un antispam ou un antivirus défaillant en premier lieu.

Mais allons encore plus loin : sans même attaquer qui que ce soit, Internet offre un choix insensé de données sur des espaces non protégés qui sont faciles à identifier pour qui sait où chercher. Pour donner un ordre d'idée, chaque jour 3 milliards de documents et lignes de bases de données passent par les moteurs de CybelAngel.

[1]https://cybelangel.com/blog/medical-data-leaks/
[2]https://www.01net.com/actualites/ransomware-les-attaques-sur-les-hopitaux-francais-se-multiplient-2035000.html
[3]https://cybelangel.com/blog/heathcare-data-targeted/
[4]Ibid
[5]https://www.numerama.com/tech/740608-apres-la-fuite-des-resultats-de-14-million-de-tests-covid-lap-hp-a-bien-ecrit-a-ses-patients.html
[6]https://www.lemagit.fr/actualites/252507013/Une-0-day-au-cur-du-vol-de-donnees-de-14-million-de-Franciliens-testes-Covid-19

Ces éléments sont trouvés via des protocoles non sécurisés (utilisés par des NAS par exemple), mais également des applications Cloud, type Google Drive, Slideshare et autres Dropbox.

C'est également sans compter les téraoctets de documents exfiltrés et mis en libre téléchargement par des groupes de rançongiciels. Bien que ces derniers déclarent ne pas s'en prendre aux hôpitaux ni aux centres de soins, ils ne se privent pas pour attaquer cliniques, laboratoires, prestataires du domaine médical et puis bon, de temps en temps, un hôpital par-ci par-là malgré tout.

Régulièrement, sur les forums underground, quelques mois plus tard, ces mêmes données sont compilées et vendues, alors que l'acteur fait ressortir leur valeur - "dossiers de patients US", "10 millions de numéros de sécurité sociale", "personnel hospitalier allemand", etc.

Bref, l'exposition est énorme, et nous n'avons même pas parlé des données patients qui se retrouvent indexées dans les moteurs de recherche à cause d'un serveur Web non protégé, ni de tous ces objets que nous nous évertuons à connecter à Internet, pour le meilleur et pour le pire - et comme chacun sait, tout ce qui est connecté est vulnérable.

Pourquoi un tel acharnement sur les données de santé ?

Précisons tout de suite que de leur côté, les acteurs cybercriminels n'ont pas attendu la crise sanitaire mondiale pour s'intéresser aux données médicales. Si des cas sont de plus en plus rapportés dans la presse en 2021, notamment via les attaques de rançongiciel contre les hôpitaux, sur les forums cybercriminels elles ont toujours eu la côte, où elles sont vendues près de dix fois plus cher que des données de type carte bancaire, email, numéros de téléphone, etc.

Les données de santé sont une mine inépuisable d'informations pour un large panel d'acteurs. D'abord, elles font fi des différences entre chaque individu (sexe, âge, classe sociale, nationalité, etc.), tout le monde est logé à la même enseigne, et elles sont internationales. Un groupe sanguin, une maladie, un indice de masse corporelle n'a pas de frontière. Ensuite, elles sont propres à chaque individu, étant majoritairement rattachées à un numéro de sécurité sociale ou en tout cas un numéro d'identification citoyenne, peu importe le pays. Ce sont également des données que nous ne contrôlons pas, nous les "subissons", dans le sens où elles nous collent à la peau, mais nous ne pouvons pas les modifier, contrairement à toute l'empreinte numérique que nous créons quotidiennement - publications sur les réseaux sociaux, achats en ligne, commentaires sur des articles de journaux, etc. Mais surtout, elles contiennent des informations intimes qui ne regardent que le patient et son praticien, informations qui dans certains contextes peuvent mettre un individu à l'écart voire en danger.

Le premier risque auquel on pense est donc naturellement le chantage, à juste titre. Il y a aussi de quoi alimenter la fraude : aux États-Unis, le phénomène des "ghosts clinics" coûte des centaines de millions de dollars par an aux assurances - il s'agit de créer de faux rendez-vous patients, voire de faux établissements de santé, avec de vraies données.

Pour les groupes de rançongiciel, il y a une autre raison, plus pragmatique : en dehors de toute considération d'assurance cyber, en attaquant un centre de soins il y a plus de chances de recevoir un paiement qu'en attaquant la PME qui fabrique des lunettes. L'arrêt de la PME entraîne une perte

de revenus pour la société, voire un arrêt définitif, mais il n'y a pas mort d'homme. Dans le cas du centre de soins, le scénario est tout à fait envisageable, en particulier en ces temps de pandémie.

Enfin, sans tomber dans la paranoïa, il y a clairement un intérêt qui dépasse la fraude, l'usurpation d'identité et le rançonnage. Qui achète ces données lorsqu'elles transitent sur les places de marché noir ? Mystère. Et qu'en font-elles ? Re-mystère.

Et pourtant ces datasets, d'une tranche de la population ou d'un pays entier, ont une valeur inestimable, au moins marchande, sinon les géants de l'Internet ne s'y intéresseraient pas[7].

La mine d'or n'est pas près de se tarir

Les données de santé sont au centre de tous les paradoxes :
- Elles sont de plus en plus numérisées donc nous élargissons la surface d'attaque, or nous avons de plus en plus à cœur le droit au respect de notre vie privée, notre anonymat ;
- Elles sont d'une grande valeur et criticité, or elles ne sont pas toujours bien protégées - on pourrait citer le manque de régulations (ou en tout cas leur application), le manque de sensibilisation et le manque de ressources comme explications principales, mais chaque raison vaudrait une tribune en soi ;
- Elles doivent cependant être protégées, or elles doivent être facilement accessibles afin d'être utilisables en urgence.

Paradoxe ultime, les smartphones proposent de plus en plus des fonctions natives pour indiquer ses informations de santé de base, qu'un secouriste pourrait consulter librement en prenant en charge le patient dans l'incapacité de répondre aux questions.
À l'ère du big data, nous ne devons pas oublier que nos données de santé ne sont pas des données classiques, même si tout ce qui transite ne semble être que des suites infinies de 0 et de 1. Elles doivent être traitées avec toutes les précautions qui s'imposent, et la première de toutes les étapes est la sensibilisation des personnes qui doivent les manipuler.

Je fus ainsi ravi lorsqu'en partant du cabinet de mon dentiste le secrétariat m'a interpellé : oui les données étaient stockées dans un endroit sûr, chez un prestataire certifié HDS et utilisées uniquement à des fins médicales. Le post-it était toujours là, mais petit à petit, les consciences évoluent.

[7] https://www.lopinion.fr/edition/wsj/enquete-comment-google-collecte-donnees-medicales-americains-208360