Pirater un PC via un code malveillant écrit sur un brin d'ADN

Les pirates ont souvent un coup d'avance, mais de temps en temps les chercheurs en sécurité imaginent les évolutions des menaces pour mieux les contrer. A cette fin, des chercheurs de plusieurs disciplines (biologie, sécurité informatique, etc.) de l'Université de Washington travaillent depuis quelques années sur la sécurité des systèmes d'analyse et de séquençage de l'ADN. Ils ont pu trouver pas mal de vulnérabilités dans les solutions Open Source utilisées. Dans leurs derniers travaux, ils ont réussi à intégrer du code malveillant dans un brin d'ADN et celui-ci a été lu par un PC qui s'est retrouvé infecté.

Un encodage épineux

Pour réaliser cela, ils se sont servis des derniers travaux sur l'encodage d'information au sein de l'ADN. Microsoft est très en pointe sur ce sujet avec la capacité d'écrire 200 Mo de données sur des brins d'ADN. La firme de Redmond envisage même de créer des systèmes de stockage à base d'ADN dans un proche avenir. Dans la démonstration des chercheurs, ils ont encodé une attaque simple de débordement de mémoire (buffer overflow). Mais l'encodage a été compliqué. Les séquenceurs d'ADN fonctionnent en mélangeant de l'ADN avec des produits chimiques capables de distinguer les bases (comprenant des chaines de nucléotides A, C, G, T (adénine, cytosine, guanine et thymine)) en fonction des couleurs qu'elles émettent. Pour accélérer le processus d'analyse, ces millions de bases sont divisées elles-mêmes en millions de morceaux pour être analysées en parallèle.

Les données de l'attaque ont dû être adaptées à quelques centaines de bases pour rester intactes à la suite du traitement en parallèle du séquenceur. « *L'exploit a été intégré dans 176 bases longues* », explique Karl Koscher, un des chercheurs ayant participé à l'aventure et co-auteur du rapport présenté à la conférence Usenix. Il ajoute, « *un programme de compression transforme chaque base en deux bits, qui sont regroupés, ce qui entraîne un exploit de 44 octets lors de l'analyse* ». Dans le détail, il précise « *la plupart de ces octets sont utilisés pour coder une commande shell ASCII. 4 octets servent à faire retourner la fonction de conversion à la fonction système () via une bibliothèque C standard, qui exécute les commandes shell, et 4 octets supplémentaires ont été utilisés pour indiquer au système () où la commande est en mémoire* ».

Une question d'équilibre des paires A-T et G-C

Une fois les bases encodées sous la forme As, Ts, Gs et Cs et prêtes pour l'analyse, les chercheurs ont constaté que l'ADN répondait à certaines contraintes physiques. Ainsi, pour assurer une stabilité à leur échantillon, les chercheurs doivent maintenir un ratio entre entre Gs et Cs et Ts et As. En effet, la stabilité naturelle de l'ADN impose une proportion régulière de paires A-T et G-C. Conséquence de cette spécificité, les spécialistes ont dû réécrire plusieurs fois le code de l'attaque pour trouver une forme viable et exploitable par les analyseurs d'ADN. Enfin pour mener à bien l'expérience, ils ont modifié le programme Open Source de compression, fqzcomp, pour lire

correctement l'ADN et diffuser son contenu malveillant.

Les différents spécialistes soulignent qu'il faut beaucoup d'efforts et d'argent pour mener à bien ce type d'attaque. Mais elle a le mérite d'exister et de montrer sa faisabilité. Les cybercriminels disposent souvent du temps et de l'argent, une attaque par ADN corrompu peut leur ouvrir un champ des possibles. Et ce n'est pas de la science-fiction.

A lire aussi :

Microsoft en route pour une baie de stockage à base d'ADN

Un algorithme de streaming vidéo aide au stockage sur ADN