Rapport du webinaire sur la génomique et la cybersécurité avec Renaud Lifchitz

Par GUARDIA Quentin, quentin.guardia@etu.u-paris.fr, M1 Cybersécurité FI

Renaud Lifchitz, ingénieur senior en sécurité informatique, a commencé le webinaire avec pour problématique : comment améliorer l'ingénierie sociale en utilisant la génomique ?

La génomique a évolué rapidement. On est passé de plusieurs années à deux semaines pour séquencer un ADN humain en entier, depuis 2003. Le séquençage s'utilise beaucoup à but récréatif aux USA, pour déterminer l'origine ethnique. Cela est interdit en France. En revanche, il est utilisé pour analyser une scène de crime, diagnostiquer une maladie ou reconnaître des membres de la même famille.

Mais quel est le rapport avec la cybersécurité ? Et bien la génomique peut révéler les faiblesses humaines. Elles peuvent faire office de failles pour des personnes malveillantes. Avant de développer cet aspect, petit rappel sur la génétique.

L'ADN contenu dans les chromosomes est composé de nucléotides, dont les expressions sont : A, C, G ou T. Cet ensemble forme le génotype. Les humains partagent 99,9 % du même ADN. Par contre, les SNP, polymorphisme nucléotidique, regroupent 90 % de la partie de l'ADN qui diffère d'un individu à l'autre. Ces différences se manifestent sur le phénotype. À noter que le phénotype varie aussi sous l'effet de l'environnement, d'où le non-déterminisme génétique.

Les cybercriminels et grandes organisations commencent à s'intéresser à l'ADN. En juin 2019, un groupe Iranien a piraté une machine de séquençage connectée au web. Fin 2019, la société de séquençage Veritas s'est sûrement fait voler les données. En même temps en Floride, toute séquence ADN est rendue accessible à la justice. Les données ne sont plus protégées. En chine, l'ADN de chaque nouveau-né est séquencé. En parallèle, les séquençages se popularisent: plus de 30 millions d'individus sont volontairement « analysés » aux USA. Il suffit d'envoyer un prélèvement salivaire par courrier et attendre que les données soient disponibles numériquement. Les tests grand public sont souvent partiels car centrés sur certains SNP, ce qui vaut un fichier de 100 Mo en moyenne. Les analyses complètes sont à peine plus chères mais accessibles.

Des génotypes se trouvent sur Google par indexage inopiné (Google Dorks). On peut même cibler les SNP en cherchant bien. Sauf que le moteur de recherche dédoublonne les SNP car ils sont similaires. À cela s'ajoute des bases de données génétiques open source. On peut voir à quoi correspond chaque SNP. Entre autres, grâce à dbSNP ou SNPedia.

Bien que l'origine ethnique soit difficilement prédictible, le lien familial l'est. Ainsi, en séquençant 1 % d'une population (taux dépassé aux USA), on peut désanonymiser 100 % de celle-ci. De plus, le SNP détermine : les traits physiques, allergies, … Et même les préférences, capacités physiques, logiques, la personnalité, faiblesses, etc. Qui sont influencées par le génome ! L'usage pour le social engineering est donc très pertinent. On peut se faire une idée sur Genomilink.

L'usage en cybercriminalité est alors clairement établi. Par exemple, quelqu'un qui est faible en reconnaissance faciale ou peu consciencieux, peut être facilement victime d'intrusion physique. Quelqu'un qui est dépendant à la récompense rapide peut aisément être victime de phishing. Une personne anxieuse est une bonne victime pour la fraude au président. Les exemples d'attaques personnalisées sont nombreuses. D'où l'intérêt de voler ces données et les revendre sur le marché noir d'internet. Ou même de les récupérer à l'insu des personnes, sur leurs couverts par exemple.

Il faut donc toujours relire les conditions d'utilisation des sites de séquençage pour éviter la revente d'informations, et s'assurer du respect des normes ISO 27001, HIPAA et RGPD de la société.