Le code a-t-il fuité ?

Les fuites de code source sont en hausse et sont parfois mal comprises. Maddyness a souhaité donner la parole à Thomas Segura, expert en cybersécurité pour GitGuardian, afin de dresser un panorama sur les fuites de données et les enjeux derrière ce phénomène.

Temps de lecture : minute

---

15 juillet 2022

Il y a quelques semaines à peine, un groupe de hackers connu sous le nom de Lapsus$ était sur toutes les lèvres des analystes de sécurité : il avait volé et divulgué le code source de certaines des plus grandes entreprises technologiques du monde, dont près de 200 Go de code source de Samsung et des parties des projets Bing, Bing Maps et Cortana de Microsoft. Cette affaire a eu lieu quelques mois seulement après que la populaire plateforme de streaming Twitch ait subi le même sort.

Ces dernières années, nous avons assisté à une recrudescence des fuites de dépôts de code d'entreprise en ligne. Bien que toutes les fuites mentionnées aient été motivées par des acteurs malveillants - le motif d'un pirate informatique est de faire pression sur les entreprises en montrant qu'il les a piratées et qu'il a eu accès à des données - n'oublions pas que des erreurs se produisent plus souvent qu'on ne veut bien le croire et sans faire de bruit : un développeur qui pousse par inadvertance vers un dépôt public, alors qu'il pense travailler sur un environnement privé suffit.

Malheureusement, au-delà de l'évident préjudice de réputation, le problème des fuites de code source est parfois mal compris.

Code source et propriété intellectuelle

Dans un monde où 95 % du secteur informatique s'appuie d'une manière ou d'une autre sur des logiciels libres (source : Gartner), il est facile d'oublier que le code source est considéré comme une propriété intellectuelle dès la création de la première ligne de code. Il peut être protégé par des droits d'auteur et, en fait, il est assez courant de voir les fichiers de code source inclure un en-tête indiquant le propriétaire des droits d'auteur et les autorisations.

Numérique par essence, le code source est un actif qui fuit. Les systèmes de contrôle de version distribués et les plateformes de partage de code ont grandement facilité la prolifération et le clonage de bases de code entières sur Internet. Cette situation a alimenté la communauté des logiciels libres et conduit à de grandes réalisations, mais reste problématique pour un nombre croissant d'entreprises pour lesquelles le code source est un actif de valeur fondamentale. Assurer le respect des droits d'auteur du code source à l'échelle mondiale est devenu une tâche presque impossible.

Le code source peut également contenir des secrets commerciaux, qui ne peuvent être protégés par des droits d'auteur, des brevets ou des marques. Dans un secteur où l'innovation et la rapidité sont si cruciales, ne pas réussir à garder ce type d'informations secrètes pourrait signifier la fin du jeu pour certaines jeunes entreprises. Pour les entreprises plus importantes, éviter les frais de justice est également un motif pour mieux surveiller et contrôler quelle partie du code source a été partagée, et avec qui.

Sur un autre plan, le code source publié donne aux pirates une vue de l'intérieur des systèmes déployés en production au service de millions

d'utilisateurs parfois. Cela peut également poser problème.

Les menaces qui se cachent dans le code source

Le code est comparable aux plans des produits physiques. Pour les personnes ayant des connaissances techniques, le fait d'avoir un accès complet aux rouages internes de grands systèmes ou de logiciels signifie qu'elles peuvent facilement repérer les vulnérabilités telles que les informations d'identification codées en dur. Sur les plus de 6 600 clés trouvées dans le code source de Samsung, environ 90 % sont destinées aux services et à l'infrastructure internes de Samsung, tandis que les 10 % restants, critiques, pourraient donner accès aux services ou outils externes de Samsung, tels que AWS, GitHub, Artifactory et Google*.

Bien que cela ne signifie pas nécessairement que les pirates pourraient exploiter des secrets, des failles dans les contrôles d'accès et d'autres vulnérabilités de sécurité juste après qu'une fuite se soit produite, cela reste un atout supplémentaire dans leur boîte à outils. Certains signaux faibles peuvent même être plus précieux pour un attaquant que la logique métier elle-même, comme les informations personnelles identifiables (IPI) des développeurs, leurs habitudes, leur langue et, plus globalement, la culture de l'entreprise. Des milliers de lignes de code peuvent révéler beaucoup plus d'informations qu'on ne le pense.

Comment savoir si un code source a fuité ?

Peu importe que l’entreprise ait subi une violation de données dans le passé ou non, le code propriétaire pourrait être présent publiquement sur

GitHub à l’insu de l’entreprise. Même si cette dernière n'a pas de présence officielle sur la plateforme, il est fort probable que les développeurs contribuent à des projets open-source ou utilisent des dépôts personnels pour partager du code.

Par conséquent, la cartographie de l'empreinte de l’organisation sur GitHub doit être prise au sérieux. L’analyse des menaces doit pouvoir répondre à tout moment aux questions suivantes : quels sont les actifs disponibles publiquement sur Internet ? Comprennent-ils des données utilisateur ? Fournissent-ils des informations critiques et encore valides qui pourraient être exploitées par un attaquant ? Est-il encore possible de contenir les actifs divulgués (en lançant une procédure de retrait DMCA par exemple), ou non ?

Techniquement, la seule solution pour identifier une fuite de code source est de prendre l'empreinte du code propriétaire et de la comparer à une base de données de fichiers publics.

Les fuites de code source sont à prendre au sérieux. Non seulement elles nuisent à la réputation de la marque, mais elles peuvent aussi potentiellement mettre en danger certains des actifs les plus précieux que possède une entreprise. Si les demandes de retrait DMCA permettent de s'assurer que le contenu protégé par le droit d'auteur est retiré des plateformes publiques, les secrets commerciaux peuvent également être révélés et mettre en péril les aspects les plus innovants de l’entreprise. Du point de vue de la sécurité, ils offrent ce que les attaquants apprécient le plus : une image complète du fonctionnement interne d'un logiciel, y compris ses failles, les processus et les personnes impliquées.

Avec l'explosion de la plateforme de partage de code GitHub, la surveillance de l'empreinte publique d'une entreprise est devenue l'une des tâches les plus difficiles pour les analystes sécurité. Mais des outils existent pour aider les entreprises à déterminer si sa propriété

intellectuelle a fuité.

*Chiffres Gitguardian

Article écrit par Thomas Segura, GitGuardian