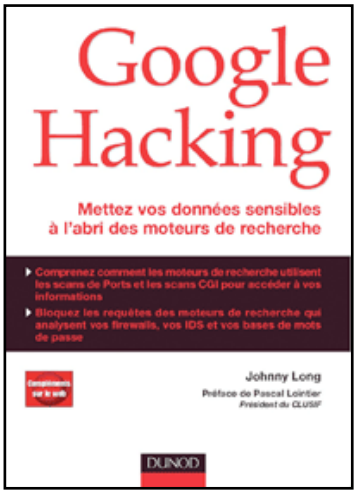
Le Google Hacking, ou comment pirater en utilisant Google

Retour au sommaire de la lettre

Un nouveau livre paraît ce mois-ci chez l'éditeur Dunod, ayant pour titre "Google Hacking". Traduction d'un livre anglais, introduit par Pascal Lointier, président du CLUSIF (Club de la sécurité des systèmes d'information français), il explique comment certains pirates utilisent Google pour obtenir des numéros de cartes bancaires des listes de mots de passe ou autres informations sensibles sur le Web. Nous vous proposons dans cet article une interview de Pascal Lointier, qui présente le livre, ainsi que des extraits de l'ouvrage.



Formidable outil de recherche sur l'Internet, Google serait-il victime de son succès au point de connaître un détournement de ses fonctionnalités aux fins de piratage et d'espionnage ? C'est là tout l'enjeu des techniques du Google Hacking ! En effet, les fantastiques capacités de stockage des pages web et d'indexation de leurs contenus de Google sont aussi une porte ouverte aux hackers qui peuvent accéder, à partir de requêtes judicieusement construites, à quantité d'informations sensibles.

Interview de Pascal Lointier (président du CLUSIF)

Pascal Lointier, président du CLUSIF (Club de la Sécurité des Systèmes d'information français) et préfacier du livre de Johnny Long, Google Hacking (adaptation française, Dunod, 2005), (http://www.dunod.com/pages/ouvrages/ficheouvrage.asp?id=49421) nous dévoile les risques encourus par les organisations et les personnes, pour leur permettre de mettre en place une stratégie de "contre-intelligence économique"…

Pouvez-vous nous présenter l'auteur de Google Hacking, Johnny Long, et votre rôle dans l'adaptation française de ce livre ?

Johnny Long est un spécialiste de la sécurité des réseaux. Il s'occupe, en particulier, de la mise en place des infrastructures de sécurité pour des organisations gouvernementales ou de grandes entreprises. Il a créé un site Internet (http://johnny.ihackstuff.com/) dédié au "Google Hacking" et à ses utilisations. Il a une grande renommée dans le monde "underground" – ce qui n'est jamais évident – et ses analyses font autorité.
Il intervient dans de nombreuses conférences au niveau international.
Pour ma part, je suis le président du CLUSIF (http://www.clusif.asso.fr/), association spécialisée dans la sécurité de l'information, et je suis ingénieur en sécurité chez ACE Europe. Les Éditions Dunod ont fait appel à moi pour rédiger un avant-propos sur le cadre légal dans lequel s'inscrit cet ouvrage, pour le préfacer, et aussi pour en assurer une relecture technique.

Pourquoi Google est-il devenu l'un des outils privilégiés des hackers ?

Il faut en fait se rendre compte que Google donne aujourd'hui accès à d'innombrables informations et documents. Des agents automatiques de Google circulent sur l'ensemble du web et en aspirent le contenu, s'il n'y pas de tag (indicateur présent dans le code source de la page web) spécifique contraignant ou dissuasif, car il est souvent possible de passer outre un tel tag.
Google suit ainsi les ramifications des sites et des pages qu'il rencontre pour indexer les contenus. Ce faisant, il cartographie l'ensemble du réseau et donc aussi les réseaux Intranet d'entreprises ou d'administrations lorsque la sécurité est trop permissive.

Cela signifie que Google permet d'obtenir des informations sur des "entités". Il est en effet possible, via les opérateurs de recherche avancée, d'effectuer des requêtes spécifiques pour retrouver une liste de mots de passe ou un fichier de noms d'utilisateurs, par exemple. On comprend dès lors aisément l'intérêt qu'il y a à détourner un outil qui donne un accès rapide et simple à de telles données…

En quoi le Google Hacking se distingue-t-il des techniques classiques de piratage ?

Traditionnellement, on distingue les tests de vulnérabilité et les tests d'intrusion.
Les tests de vulnérabilité permettent, après avoir cartographié un réseau, de connaître les machines et logiciels utilisés, les numéros de version, etc. – toutes informations qui permettent ensuite de déceler les failles liées à ces machines ou à ces versions de logiciels.
Quant aux tests d'intrusion, ils servent à tester l'accessibilité au réseau même de l'entreprise, en utilisant par exemple les failles de sécurité des logiciels ou des matériels utilisés par l'entreprise, ou encore les mots de passe triviaux de ses utilisateurs.
Le Google Hacking va plus loin que cela, en ce sens qu'il peut s'inscrire dans une démarche d'intelligence économique ou d'espionnage économique. Le but est, par exemple, de récupérer le business-plan d'un concurrent, sans pour autant accéder à son site. En utilisant Google et son cache, on peut ainsi accéder à des informations sensibles, et ce dans un anonymat presque total. Je dis "presque", car il reste tout de même possible de remonter jusqu'à l'émetteur de la requête. Dès lors, pour se protéger contre une telle démarche de renseignement ou d'espionnage, il est nécessaire de connaître les techniques du Google Hacking pour pouvoir effectuer ses propres recherches sur Google. Ce faisant, on apprend quelles informations sont effectivement accessibles sur le réseau et celles qu'il convient donc de mieux protéger – tout en sachant que subsistera toujours le problème de la conservation des données dans la mémoire cache de Google. Il y a donc aussi un travail d'information et de responsabilisation des responsables des systèmes d'information et des utilisateurs à effectuer en matière de sécurité de l'information.

Quelles sont les ressources disponibles dans la mémoire cache de Google ?

La mémoire cache de Google est un outil particulièrement "performant" en matière de conservation de données sensibles. En effet, même si l'on retire une page de son site Internet, celle-ci est conservée dans le cache de Google, sans limitation de durée : il est donc possible de récupérer, par ce biais, des informations qui auraient été supprimées et ce, sans même accéder au site web concerné.
En cela, le livre va bien au-delà de la simple analyse des techniques de recherche et de l'utilisation des opérateurs booléens, car il montre comment un moteur comme Google – mais d'autres outils de recherche sont également concernés – permet de pister et de récupérer des données sensibles, qu'il s'agisse d'informations personnelles ou professionnelles.

Entre hacker et hacké, à qui profitera le plus la lecture de Google Hacking ?

Un détournement de l'objectif du livre est bien sûr toujours possible : lorsque l'on a connaissance de failles, on peut le crier partout et en faire un emploi malveillant. Mais dans le cas présent, il s'agit plus d'un avertissement sur le potentiel d'exposition des informations, à charge pour l'entreprise de prendre des dispositions de sécurité.
L'objectif prioritaire de ce livre est en réalité d'informer sur les manières de se protéger, car ce n'est pas en gardant le silence que l'on fait prendre conscience des risques encourus. Il est donc important de montrer qu'aujourd'hui, chacun peut avoir accès à des informations que l'on croyait inaccessibles et ce, du fait de la démarche globalisante des moteurs de recherche dont le but est bien de vouloir tout indexer.

Les risques engendrés par le Google Hacking sont-ils, selon vous, plus importants en Europe qu'aux États-Unis ?

Contrairement à ce que l'on pense souvent, les entreprises, les administrations et les citoyens aux États-Unis sont loin de bénéficier de la meilleure protection en matière de sécurité de l'information – en témoigne l'usage américain de cartes bancaires à pistes magnétiques, alors que l'on utilise

des cartes à puce en France et qu'elles commencent à se développer en Europe. Il devient alors très facile – et l'auteur le montre dans le livre – d'accéder à des listes de numéros de sécurité sociale américains (numéros SSN) via Google et d'usurper ou de contrefaire de tels numéros. Or, à partir d'un numéro SSN, vous pouvez quasiment tout faire : ouvrir un compte bancaire, commander des billets d'avion, etc. Cela est bien plus complexe en France : les systèmes y sont plus sécurisés, moins perméables et les informations d'identité mieux protégées en raison d'une d'un dispositif légal plus contraignant.

Face à la puissance de Google, comment protéger ses données sensibles ?

D'abord, la notion de "sensibilité" de l'information est quelque chose d'incertain, d'autant que cette sensibilité est très variable dans le temps. Il n'y a donc pas de recette miracle et les moyens et procédures seront adaptés au contexte.
Il faut éviter de mettre en ligne ou sur l'Intranet des informations très sensibles, cela va de soi. Pour les autres données, il faut apprécier le périmètre de diffusion ou le niveau d'accessibilité acceptable.
Il convient aussi d'adapter les requêtes présentées dans l'ouvrage à son propre cas et vérifier si l'on peut, ou non, accéder à ses propres informations, à ses propres documents.
Noyer l'information parmi de nombreuses autres, ce qui aurait pu être une solution il y a quelques années, n'est aujourd'hui pas suffisant : le traitement automatisé des données permet en effet de trouver rapidement le bon document, la bonne information, et de prendre l'avantage sur un concurrent…

A que(s) public(s) recommanderiez-vous plus spécifiquement la lecture de Google Hacking ?

Plus que de sécurité du système d'information, ce livre traite en fait de la veille et de la défense de l'information – et donc de contre-intelligence économique. En cela, il intéresse non seulement les publics informaticiens (Administrateurs de réseaux, DSI, Responsables de la Sécurité), mais aussi les managers soucieux de protéger leur stratégie de veille économique et les intérêts de leurs entreprises.

Merci Pascal Lointier.

Extraits du livre

Nous avons choisi de vous présenter ici les paragraphes "Droit à l'essentiel" qui résument le contenu de certains chapitres du livre, afin que vous puissiez avoir une vision plus concrète du contenu de l'ouvrage, sachant que les chapitres en question sont bien plus détaillés que ce que vous pourrez trouver ci-dessous...

Extrait du chapitre 3 ("Les bases du hacking sur Google")

3.6.1 Conserver l'anonymat avec le cache
- Cliquer sur le lien "En cache" ne charge pas seulement la page à partir de la base de données de Google ; cela déclenche aussi une connexion vers le vrai serveur pour récupérer les images et tout le contenu non HTML.
- En ajoutant &strip=1 à la fin d'une URL du cache, on ne visualisera que le HTML de la page en cache. Lorsqu'on accède de cette façon à une page en cache, on ne se connecte pas au véritable serveur sur le Web; ceci vous permet de protéger votre anonymat, en utilisant la méthode de copier/coller décrite dans ce chapitre.

3.6.2 Utiliser Google comme proxy
- Grâce au service de traduction, on peut utiliser Google comme un serveur proxy transparent.
- Pour cela, vous devrez modifier l'URL, en donnant au paramètre langpair deux valeurs correspondant à la même langue, comme langpair=en%7Cen.

3.6.3 Repérer des contenus de répertoires

- Les pages listant le contenu d'un répertoire apportent de précieux renseignements.
- Un bon moyen de trouver des contenus de répertoires consiste à lancer une requête du genre intitle:index.of "parent directory" ou intitle:index.of name size.

3.6.4 Repérer des répertoires spécifiques dans un contenu de répertoire
- Vous pouvez repérer facilement des répertoires spécifiques dans un contenu de répertoire en ajoutant un nom de répertoire à une recherche index.of.
Par exemple, on peut utiliser intitle:index.of inurl:backup pour rechercher des contenus de répertoires ayant backup dans l'URL. Si le mot backup figure dans l'URL, il y a de grandes chances que ce soit un nom de répertoire.

3.6.5 Repérer des fichiers spécifiques dans un contenu de répertoire
- Vous pouvez repérer facilement des fichiers spécifiques dans un contenu de répertoire en ajoutant le nom de fichier à une recherche index.of, comme par exemple dans intitle:index.of ws_ftp.log.

3.6.6 Repérer la version du serveur dans un contenu de répertoire
- Certains serveurs, dont Apache et ses dérivés, ajoutent un message identifiant le serveur en bas de chaque page listant le contenu d'un répertoire. On peut repérer ces messages en étendant une recherche index.of, pour s'intéresser par exemple à la phrase server at, comme dans intitle:index.of server.at.
- Pour repérer un serveur Web spécifique ou même une version spécifique, vous pouvez étendre votre recherche avec des expressions extraites du message associé à ce serveur ou cette version. Par exemple, la requête intitle:index.of server.at "Apache Tomcat/" repérera toutes les machines utilisant le serveur Apache Tomcat.

3.6.7 Traversée de répertoire
- Une fois que vous avez repéré un répertoire spécifique sur le serveur Web ciblé, vous pouvez utiliser cette technique pour repérer d'autres répertoires ou sous-répertoires.
- La méthode la plus simple consiste à utiliser les contenus de répertoires. Cliquez simplement sur le lien parent directory, qui vous conduit vers le répertoire au-dessus du répertoire courant. Si ce répertoire provoque aussi l'affichage d'un contenu de répertoire, vous pouvez cliquer sur les liens figurant dans cette page afin d'explorer d'autres répertoires. Si le répertoire parent n'affiche pas de contenu de répertoire, vous devrez recourir à une méthode plus complexe, en essayant d'imaginer des noms de répertoires à ajouter au bout de l'URL du répertoire parent. Une autre solution consiste à utiliser les opérateurs site et inurl dans une recherche Google.

3.6.8 Substitution incrémentale
- Effectuer une substitution incrémentale consiste à remplacer un nombre par l'entier inférieur ou supérieur.
- Cette technique peut être utilisée lorsqu'on explore un site utilisant des nombres dans ses noms de répertoires ou de fichiers. Remplacez le nombre par l'entier inférieur ou supérieur, en prenant garde de ne pas modifier le reste du nom (attention aux zéros !). Une autre solution consiste à utiliser l'opérateur site avec les opérateurs inurl ou filetype pour effectuer une recherche Google créative.

3.6.9 Parcours d'extensions
- Cette technique permet de repérer des fichiers (par exemple, des fichiers de sauvegarde) qui ont le même nom de fichier mais une extension différente.
- La technique la plus simple consiste à remplacer une extension par une autre dans l'URL, par exemple pour remplacer html par bak.
- Les contenus de répertoires, et tout particulièrement les contenus de répertoires en cache, permettent de repérer facilement s'il existe des fichiers de sauvegarde, ainsi que les types d'extensions de fichiers que l'on peut rencontrer sur le reste du site.

Extrait chapitre 7 ("Dix recherches simples qui marchent")

Il n'existe pas de liste parfaite, mais ces 10 recherches devraient vous rendre bien des services en attendant que vous compiliez votre propre liste de requêtes gagnantes. Il faut bien comprendre qu'une requête qui fonctionne face à une cible peut fort bien ne pas bien fonctionner face à d'autres cibles. Notez bien les requêtes qui fonctionnent pour vous, et tâchez d'en tirer quelques conclusions sur ce qui marche et ce qui ne marche pas. On peut utiliser des outils automatisés,

comme ceux présentés aux chapitres 11 et 12, pour pouvoir enchaîner un grand nombre de requêtes Google, telles que celles figurant dans la base de données Google Hacking ; mais dans certains cas, le plus simple est aussi le mieux. Si vous avez du mal à déterminer quelles sont les requêtes qui vous conviennent, n'hésitez pas à les conserver dans une liste que vous utiliserez avec l'un des outils automatisés dont nous parlerons plus tard.

7.3.1 site :
• L'opérateur site est pratique pour parcourir tout le contenu que Google a rassemblé à partir d'une cible.
• Cet opérateur est en général combiné à d'autres, afin de restreindre le champ de la recherche à une seule cible.

7.3.2 intitle:index.of
• C'est la requête universelle pour rechercher des listes de répertoires générées par Apache.
• Les listes de répertoires fournissent une foule d'informations aux attaquants.

7.3.3 error | warning
• Les messages d'erreur sont très révélateurs, quel que soit le contexte.
• Dans certains cas, les messages d'avertissement donnent des renseignements importants sur les logiciels mis en œuvre sur la cible.

7.3.4 login | logon
• Cette requête repère des portails de connexion avec beaucoup d'efficacité.
• On peut aussi l'utiliser pour rechercher des noms d'utilisateur ou des procédures d'aide à la connexion.

7.3.5 username | userid | employee.ID | "your username is"
• C'est une des requêtes les plus générales pour rechercher des noms d'utilisateurs.
• Si cette requête ne révèle pas de noms d'utilisateurs, le contexte qui entoure ces mots peut cependant décrire certaines procédures et aider ainsi l'attaquant dans une offensive ultérieure.

7.3.6 password | passcode | "your password is"
• Cette requête reflète l'utilisation usuelle du mot password en anglais. Vous devrez l'adapter pour un site français.
• Cette requête peut dévoiler des documents décrivant les procédures de connexion, de changement de mot de passe, ou donner des indications sur les règles de changement des mots de passe en usage sur la cible.

7.3.7 admin | administrator
• Cette requête, qui utilise les deux termes anglais les plus communs pour désigner le propriétaire ou l'administrateur d'un site et que vous pourrez compléter en français par le mot administrateur, permet aussi de repérer des détails sur les procédures (comment contacter l'administrateur), et même sur les portails d'administration.

7.3.8 –ext:html –ext:htm –ext:shtml –ext:asp –ext:php
• Cette requête, qui doit être combinée avec l'opérateur site, permet d'écarter les types de fichiers les plus communs de façon à révéler des documents plus intéressants.
• On peut adapter cette requête pour exclure également d'autres types de fichiers communs, ceci en fonction du site ciblé.

7.3.9 inurl:temp | inurl:tmp | inurl:backup | inurl:bak
• Cette requête permet de repérer des fichiers et des répertoires temporaires ou de sauvegarde.

7.3.10 intranet | help.desk
• Cette requête récupère des sites intranet (qui sont d'ordinaire censés être protégés du grand public) ainsi que les moyens d'accès au service d'assistance aux utilisateurs.

Extraits du chapitre 9 ("Noms d'utilisateurs, mots de passe et autres secrets...)

Ne nous y trompons pas, il y a des données sensibles sur le Web, et Google sait les trouver. Il n'y a virtuellement pas de limites à ce que l'on peut trouver, dès lors qu'on sait quelle requête formuler. Des noms d'utilisateurs jusqu'aux mots de passe, numéros de carte de crédit et de sécurité sociale,

informations financières, tout est là. En tant que pirate, vous pouvez compter sur la naïveté des autres, mais en tant que responsable de la sécurité chargé de mettre un site à l'abri de cette forme dangereuse de fuite d'informations, vous pouvez être écrasé par l'immensité de votre responsabilité de défenseur.

Aussi bizarre que cela puisse paraître, une politique de sécurité sérieuse et contrôlée est le meilleur moyen d'empêcher la fuite de données sensibles sur le Web. Si les utilisateurs comprennent les risques liés à la fuite d'informations et s'ils comprennent à quelles sanctions ils s'exposent en enfreignant les règles, ils se révéleront plus coopératifs dans ce qui devrait être un partenariat sur la sécurité. Entre-temps, cela ne fait certainement pas de mal d'essayer de comprendre la tactique qu'emploierait un adversaire pour attaquer un serveur Web. Il est clair pour vous maintenant qu'un attaquant dispose d'un nombre extraordinaire de fichiers auxquels il peut s'intéresser. L'un des moyens d'éviter la fuite d'informations sur le Web consiste à refuser l'accès à des types de fichiers non connus. Que votre serveur Web délivre des fichiers CFM, ASP, PHP, ou HTML, il est infiniment plus simple de gérer ce que doit délivrer le serveur Web plutôt que de se focaliser sur ce qu'il ne doit pas délivrer. Configurez vos serveurs et vos équipements périphériques de protection pour ne laisser passer que des contenus ou des types de fichiers spécifiques.

9.6.1 Rechercher des noms d'utilisateurs
• On trouve des noms d'utilisateurs en beaucoup d'endroits.
• Dans certains cas, il faudra peut-être fouiller dans des documents ou dans des dossiers de messagerie.
• Une simple requête telle que "votre nom d'utilisateur est" peut-être très efficace pour repérer des noms d'utilisateurs.

9.6.2 Rechercher des mots de passe
• On trouve des mots de passe en beaucoup d'endroits.
• Une requête du genre "oublié votre mot de passe" permettent de repérer des pages indiquant une procédure à suivre en cas d'oubli du mot de passe.
• intext:(password | passcode | pass) intext:(username | userid | user) est une autre requête qui permet de repérer des données d'identification.

9.6.3 Rechercher des numéros de cartes de crédit, des SSN et plus
• Il existe des documents contenant des numéros de carte de crédit ou de SSN; il y en a même un certain nombre.
• Certains sites d'information particulièrement irresponsables ont révélé comment repérer pratiquement ces renseignements.
• Il y a relativement peu d'exemples de données financières en ligne, mais elles sont variées.
• Dans bien des cas, on peut les rechercher dans des fichiers ayant une extension bien spécifique.

9.6.4 Rechercher d'autres informations fructueuses
• Des carnets d'adresses et journaux de chat jusqu'aux rapports de vulnérabilité en ligne, les informations sensibles mises en ligne ne manquent pas.

Présentation du livre Google Hacking (Johnny Long, adaptation française, Dunod, 2005) :

http://www.dunod.com/pages/ouvrages/ficheouvrage.asp?id=49421

Possibilitté de télécharger la préface, l'avant propos et le chapitre 1 sur le site web de Dunod.