http://www.information-security.fr/decouverte-google-dorks/

MAJ Ultérieur à faire.

III. Des exemples de Google Dorks

Cet article n'a pas pour but de vous former à l'utilisation des Google Dorks, dont l'utilisation est rarement à des fins louables, hormis quand il s'agit d'un pentest ou un audit de sécurité. Néanmoins, nous allons voir quelques exemples pour comprendre leur utilisation et ainsi voir concrètement de quoi il s'agit.

Voici d'abord quelques-uns des opérateurs Google qui sont utilisés pour forger des requêtes très précises, notons que ces opérateurs n'ont pas du tout été créés à la base pour des besoins de sécurité ou de hacking bien entendu. C'est l'utilisation qu'en font les attaquants qui relève de la sécurité informatique :

- **intext : « chaine de caractère »** – permet de rechercher une chaine de caractère spécifique dans le code source ou le texte d'une page/d'un fichier comme vu plu haut
- **ext : « extension »** – permet de ne ressortir que des extensions de fichier spécifique, par exemple « TXT » ou les fichiers PDF (Exemple – ext: »docx »)
- **inurl : « fichier/répertoire »** – permet de retrouver des chaines de caractères dans un URL (Exemple – inurlext: »docx »)
- **site : « nomdedomaine » –** Cet opérateur est particulièrement utile lorsque l'on recherche une information sur un domaine/site précis (Exemple – site: »information-security.fr »)

Bien entendu, on peut combiner, ajouter et exclure l'ensemble de ces opérateurs (donc la liste n'est pas exhaustive, retrouvez-les tous ici : Liste des opérateurs Google).

Voici maintenant certaines requêtes Google Dorks que j'ai récupérées sur le site exploit-db.com qui contient une section « GHDB » (**Google Hacking Database),** j'ai parlé de ce site dans un billet antérieur : Découvrez le site exploit-db.com

inurl:robots.txt intext:CHANGELOG.txt intext:disallow ext:txt -site:github.com

Cette requête par exemple utilise :

- **inurl :** Pour retrouver « robots.txt » dans l'URL
- **intext:** Pour retrouver la chaine « CHANGELOG.txt » et ainsi faire apparaitre tous les fichiers Changelog, un second intext est ajouté pour retrouver la chaine « disallow »
- **ext :** Pour retrouver les fichiers aux formats « txt », fichiers texte
- **-site :** on retrouve ici l'opérateur « site » avec un « – » devant, ce qui permet d'exclure les résultats venant du site « github.com »

Concrètement, lorsque l'on souhaite sécuriser un site web, il est courant de demander aux moteurs de recherches de ne pas indexer les fichiers Changelog.txt justement dans le but d'éviter que des versions obsolètes soient retrouvées via des Google Dorks ciblant les fichiers Changelog ou leur contenu. Pour ne pas que les spiders des moteurs de recherches n'indexent un fichier, il est courant de les indiquer dans un fichier nommé « robots.txt », c'est justement cette configuration spécifique qui est visée via cette requête. Pour faire simple, on vise l'interdiction d'indexation.

Voici un second exemple :

```
intext:"root:x:0:0:root:/root:/bin/bash"
inurl:*=/etc/passwd
```

Ici, on va recherche une chaine de caractère qui est forcément présente dans les fichiers« **/etc/passwd** » des OS Linux, il s'agit de la déclaration de l'utilisateur « root » dans le fichier des utilisateurs « passwd ». Voici les opérateurs utilisés :

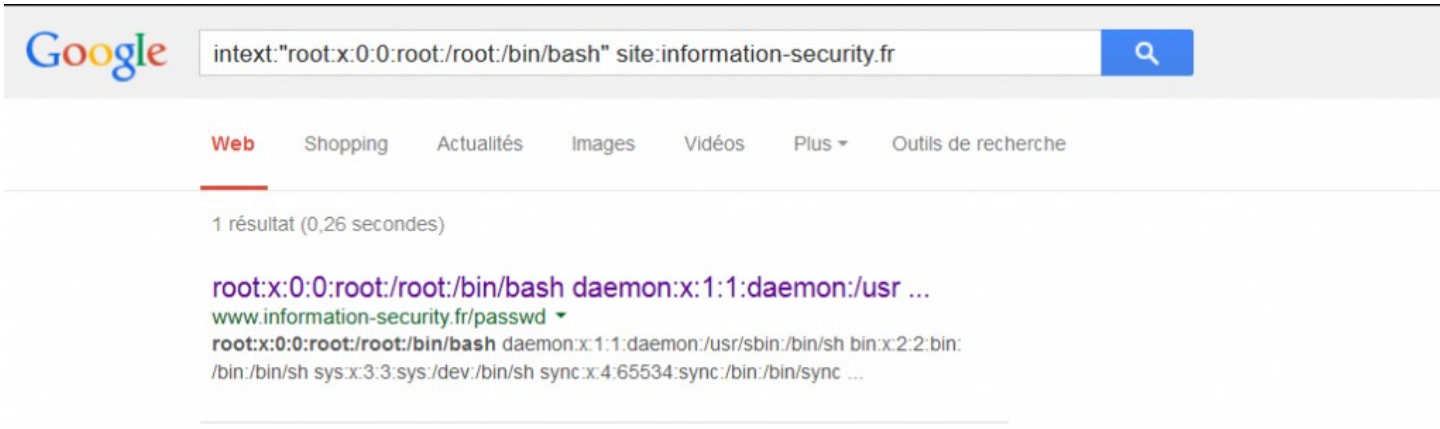
- **intext :** on recherche ici la chaine de caractère spécifique à la déclaration de l'utilisateur « root »
- **inurl :** on recherche ici la présence de « /etc/passwd » dans l'URL

Cela est plutôt rare, mais un fichier « **/etc/passwd** » indexé même s'il ne s'agit pas de sa version la plus à jour permet de retrouver des informations comme des utilisateurs présents sur les machines du parc informatique. Cela permettra de compléter la liste d'information que l'attaquant cherche à récolter. De même, si votre DSI s'appelle « Albert Dupont » et que le fichier indexé comporte un « a.dupont », on pourra directement deviner le login de tous les utilisateurs du SI.

Afin de faire un test dans le cadre de mon article, j'ai volontairement positionné un fichier « passwd » contenant cette chaine de caractère sur information-security.fr. Supposons que Information-Security soit le nom d'une société qu'un pirate cherche à attaquer, il va donc exécuter des requêtes Google Dorks uniquement sur le nom de domaine de cette société et ajouter « **site:information-security.fr**« . Le lendemain de l'ajout de ce fichier, j'ai exécuté la requête suivante :

> **intext:"root:x:0:0:root:/root:/bin/bash" site:information-security**

La requête est alors tellement précise que Google ne me ressort qu'un résultat :



Google Dorks sur un fichier passwd d'un entreprise précise

En enlevant l'opérateur « **site:** »**,** je retrouve le-dis fichier « **passwd** » en 4e page de Google.