International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 5 Issue 6, June-2016, Impact Factor: 1.544

A Review on Google Hacking Database

Vaibhav Sharma [1], Dr. Yashpal Singh[2]

[1] Research Scholar, Computer Science & Engineering Deptt., GITAM, Kablana Jhajjar (Haryana)
[2] Associate Professor, Computer Science & Engineering Deptt., GITAM, Kablana Jhajjar (Haryana)

ABSTRACT

Basically Dorks are used to incase specific information about a topic excluding the needless content. A Dork is a set of command and the keyword about which you urge information. You all must have seen whenever you search something about a special topic, Google gives millions of results and from those results you have to do Endeavour to find your interested information. To avoid this, Google dorks were introduced to comfort out the search and it's being used in Google search area thereby it is named as Google Dorks. Google hacking can be used to situate vulnerable web servers and websites which are catalogued in the Google search engine database. In other terms, hackers can locate several thousands of vulnerable websites, web servers and online devices all around the world and choose their targets randomly. It is professed that the Google hacking procedure is based on certain keywords, which could be used successfully if they are used by some internal commands of the Google search engine. These commands can be used to help hackers slim cut their search to locate sensitive data or vulnerable devices. Nevertheless, the success of Google hacking techniques depends on the survival of vulnerable sites, servers and devices. However, we should not disregard the power of the search engines in providing information about the targets to the hackers in the reconnaissance phase.

I. INTRODUCTION

GOOGLE Dorks

Almost every website you visit has a private ‘effective notebook’, also known as a database. A database records everything you do on their website. If you give the website your phone number, credit card number or social security number, it stores your information in the ‘effective notebook’ of the website. When you put down the website you’d think that you are the only one that can see your information but, unfortunately, the whole world can obtain any information you have entered on almost every website, but only if your website doesn’t cover private information from Google search. The search indexes of Google make everything public, including those ‘effective notebooks’ and everything stored in those notebooks. If you gave your credit card number on a site, then there is a high possibility that the site isn’t secured and has put all the information public, Google will then automatically add the information to its search.

This information is very simple to find for anyone and especially for cyber-criminals because Google has made it so that anyone can do a google search with the word filetype: and then have access to the ‘effective notebook’. A Google dork is an worker who unknowingly exposes sensitive corporate information on the Internet. The word dork is slang for a slow-witted person .Google dorks put business information at risk because they unwittingly create back doors that allow an attacker to penetrate a network without permission and/or gain access to unauthorized information.

Related Work

The idea of "Google Hacking" dates back to 2002, when Johnny Long began to collect interesting Google search queries that exposed vulnerable systems and/or sensitive information disclosures - labeling them **google Dorks**. The directory of google Dorks grow into large dictionary of queries, which were finally organized into the original Google Hacking Database (GHDB) in 2004. These Google hacking techniques were the heart of a book released by Johnny Long in 2005, called **Google Hacking for Penetration Testers, Volume 1**.Since , the concepts explored in Google Hacking have been extended to other search engines, such as Bing . Automated attack tools use custom search dictionaries to discover vulnerable systems and sensitive information disclosures in public systems that have been indexed by search engines.

II. ADVANCED OPERATORS

2.1 Basics

All of us have used Google search for searching answers for our queries. What most of don't recognize is the advantage of forming the search queries in Google to expose sensitive information that we require to perform a successful attack. This can be able by using the advanced operator features of Google. The basic syntax for using advanced operator in Google is as follows.

Operator_name: keyword

The syntax as shown above is a Google advanced operator followed by a colon, which is once more followed by the keyword without any space in the query. This put together becomes an advanced query string. The practice of advanced operators in Google is termed as dorking. Dorks are of two forms Simple dorks and complex dorks. Using a single advanced operator as your search query is called as simple dork whereas multiple advanced operators put together in a single search query is called as advanced dork. Each keyword/advance operator has a unique meaning to the Google engine. It helps you sort out the unwanted results and narrow down your searches by a great margin when these dorks are used.

2.2 Types of Advanced Operators

OPERATOR	PURPOSE
Intitle	Search page title
allintitle	Searches for occurrences of keywords all at a time
inurl	Search URL
allinurl	Searches for a URL matching all the keywords in the query
site	Specifically searches that particular site and lists all the results for that site
filetype	Searches for a particular filetype mentioned in the query
Link	Searches for external links to pages
Allintext	Searches for occurrences of all the keywords given
Intext	Searches for the occurrences of keywords all at once or one at a time
Numrange	Used to locate specific numbers in your searches
Daterange	Used to search within a particular date range
Inanchor	Search link anchor text
author	Group author search
group	Group name search

2.3 Mixes With Other Operators?

OPERATOR	YES / NO
Intitle	Yes
allintitle	No
inurl	Yes
allinurl	No
site	Yes
filetype	Yes

Link	No
Allintext	Not really
Intext	No
Numrange	Yes
Daterange	Yes
Inanchor	Yes
Author	Yes
Group	Not really

2.4 Can Be Used Alone?

OPERATOR	YES / NO
Intitle	Yes
Allintitle	Yes
Inurl	Yes
Allinurl	Yes
Site	Yes
Filetype	No
Link	Yes
Allintext	Yes
Intext	Yes
Numrange	Yes
Daterange	No
Inanchor	Yes
Author	Yes
Group	Yes

III. GOOGLE DORK QUERY

A Google dork query, is just referred to as a dork, is a search query that uses advanced search operators to find information that is not willingly available on a website. Google dorking, also known as Google hacking, can return information that is not easy to locate through simple search queries. That explanation includes information that is not intended for public viewing but that has not been sufficiently protected.

As a passive attack method, Google dorking can return usernames and passwords, email lists, sensitive documents and website vulnerabilities. That information can be used for so many number of illegal activities, including cyber bullying, industrial espionage, identity theft and cyber stalking.

A search parameter is a restriction applied to a search. Here are a few examples of advanced search parameters:

- **site:** returns files situated on a particular website or domain.
- **filetype:** followed by a file extension returns files of the specified type, such as PPT, DOC, PDF, XLS. Multiple file types can be searched simultaneously by separating extensions with "|".
- **inurl:** followed by a particular string returns results with that sequence of characters in the URL.
- **intext:** followed by the word or phrase returns files with the string anywhere in the text.

Multiple parameters can be used, for example, to search for files of a certain type on a firm website or domain. The Public Intelligence website provides this example:

"website services" filetype:pdf site:uplca.com

Those search parameters return PDF documents on that website's servers with the string "website services" anywhere in the document text.

Access to internal documents can give way further sensitive information. For example, document metadata frequently contains more information than the author is aware of, such as history, deletions, dates and author names. The practice of document sanitization is planned to make sure that only the projected information can be accessed.

IV. THE WAY HACKERS EXTRACT THE RESULTS THAT NORMAL USERS CAN'T

So now I am going to clarify you how hacker use these google dorks to form complex query to google search engine to extract the results that normal user can't. Let's suppose hacker wants to find admin login page of all the site, so we have dork for this inurl because hacker wants to search admin login page and normally admin login page's name look like adminlogin, admin, login etc.

So we write dork: inurl:adminlogin

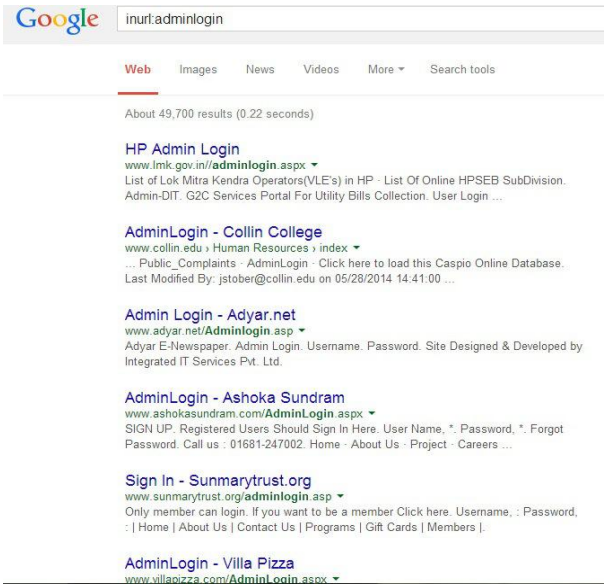

Fig 1: searching admin login page with the help of inurl dork.

You can see in the screenshot that we have admin login pages for plenty of sites. Similarly, now we create this dork for any particular website, for a particular site we have google dork **site** so now we construct dork: site:twitter.com inurl:login in the above dork we unite two dorks **site** and **inurl** it will give us twitter login page not admin login page because twitter restrict that page from google.

V. WE CAN DEFEND OUR WEB SERVER

The public server is always used for storing data which are frequently being accessed by the public and if you are really worried of keeping the data private, then the easiest and the best way is to maintain it away from the public server. Though such documents are kept isolated, it is easy to obtain access to such pages. All know the risk related with directory listings, which can allow user to see most of the files stored inside the directory, the sub directories, etc. Sometimes even the system access file is being listed which actually is used to protects the directory contents from illegal access but a simple miscon figuration allows this file to be listed and also read. Since many have the tendency of uploading important data on their servers to enable access from anywhere and they are listed by the web search crawlers. One of the straightforward rules is that Web site administrators can create a robots.txt file that specifies particular locations, so that the search engine should not discover and store in its cache. To care for yourself, use robots.txt file to avoid listing of such documents or folders. E.g. User-agent: *Disallow: /documents Also to block individual pages or if you don't want the page to be listed by any search engine, we can use something like meta tag "meta name='sipder_name' content='NOarchive'

CONCLUSION

Google hacking is a hacking technique that uses Google Search and other Google applications to find security flaws in the configuration and computer code that websites use Google is a very enormous web search engine and is capable of doing many things which are very positive for a hacker. Using simple Google dorks, people are gifted to hack a website and many web developers are not or unable to protect themselves or their customers data from such attacks. For example, using Google dorks, the attacker can mine various information like the database configuration details,

username, passwords, directory listings, error messages, etc.

REFERENCES

[1]. http://whatis.techtarget.com/definition/Google-dork-query
[2]. 'Google as an Advance E-Learner' Mrunal Shidurkar and Binal Savla (Jawaharlal Darda Institute of Engineering and Technology)
[3]. http://resources.infosecinstitute.com/defending-from-google-hackers/
[4]. http://whatis.techtarget.com/definition/Google-dork
[5]. Badri Nath (Security Analyst) at Secure Vision Labs.
[6]. 'Characterizing Google Hacking: A First Large-Scale Quantitative Study' Jialong Zhang, Jayant Notani, and Guofei Gu(SUCCESS Lab, Texas A&M University)
[7]. Nadeem ansari Student Of Secure Vision Labs Pvt. Ltd
[8]. 'Google hacking: A tool for the Security Professional' Brian Dzik GIAC Security Essentials Certification
[9]. http://securityaffairs.co/wordpress/19570/hacking/search-engines-reconnaissance-magic-weapons.html
[10]. 'The Dark Side of Google' T.Ly and M. Papadaki (Network Research Group, University of Plymouth, Plymouth, United Kingdom).