A simple search can expose security secrets.

Dorks love Google. Being a Googledork, however, is nothing to aspire to.

Yes, that's an actual term. Haven't heard of it? Well, it's about time you did. Googledork refers to anyone who has unknowingly exposed sensitive information on the Web, enabling search engines to index material that wasn't intended for public consumption.

Are you a Googledork? Are your coworkers? Think long and hard. Then do a pen test to make sure you and your company are in the clear. You might be surprised at what you find.

Google hacking—the practice of using specially crafted search engine queries to cull information about a target—is now a feather in virtually every black hat's cap. They're pulling off real intrusions, using real information

Hacking

Black hats are aware of it. Are you?

By Michael S. Mimoso

gleaned from the simplest of queries on Google and other search engines to attack unsuspecting companies. And, they do it without leaving a trace.

What should worry CISOs most? There's the troubling fact that you may be unaware that your company's sensitive information is a sitting duck on the Internet. Just as disturbing, though, is the alarming simplicity with which these hacks can be executed. All the tools to pull off a successful Google hack are readily available on the Internet.

Security practitioners are starting to learn that this is more than silly Web gibberish. If it's your job to secure information, being a Googledork could land you in the unemployment line. You can't be in the dark about this phenomenon…or how to keep from falling prey to it.

Illustration by JEFF MANGIAT

Beating Google Hackers

Don't be a Googledork.
Follow these tips to stop suspicious searches:

1. Restrict open directories on Web servers and ensure you have an index file defined.
2. Use a robots.txt file to block Web crawlers.
3. Employ NOARCHIVE and NOSNIPPET meta tags to limit caching and snippets.
4. Use password protection. Google can't traverse protected Web sites.
5. Assess yourself. Regularly run Google queries against your organization to see what is available.
6. Keep Web servers patched.
7. If you don't want data on the Internet, keep it off your Web servers.
8. Use Google's online "remove" form to delete search results from its cache.

Sources: Dave Shackleford, Vigilar; John Penrod, The Weather Channel; *Google Hacking for Penetration Testers* by Johnny Long

So Simple, It's Scary

Johnny Long literally wrote the book on the subject: *Google Hacking for Penetration Testers*. A white hat, Long created a site (http://johnny.ihackstuff.com/) that hosts the Google Hacking Database (GHDB), a trove of queries that admittedly have value to hackers and pen-testers alike. He says hackers of some repute—like Mark "Simple Nomad" Loveless and Ryan Russell—were among the first to tap search engines. Long's book and dynamic presentations at industry conferences have made Google hacking part of the security lexicon.

"The simple fact is that, if you put a searchable interface on any pile of data, I think bad guys will eventually figure out you can do not-so-nice things with that," Long says.

In part, Google hacking is a misnomer. A large part of it is information gathering, turning Google's extensive search powers loose on an enterprise's vulnerable servers and files, password logs, open directories, Web-based device-management panels, remote desktop protocol clients, and administration interfaces for routers and switches. Intent separates pen-testers from black hats.

The hacks don't always require a lot of sophistication. The right combination of advanced operators—special terms that enable more sophisticated queries—and search terms can open your eyes to enterprise security secrets you'd never believe were readily available on the Internet. It's up to the security manager to make Google hacking part of any penetration test, and to design and implement security policies and procedures that review what data and infrastructure controls are exposed to the Internet.

"If the purpose of your [search] is to gain access to a network and hack into something, security knowledge is going to make or break that. It's not going to be the sort of thing where you stumble through somebody's firewall by using Google," Long says. "If you come in with some knowledge of security, Google is a great tool and will facilitate—for good guys and bad—getting what [you're] after. That's what made this so universal. Techies understand how far reaching this is. Non-techies realize it's something simple."

Long's site contributes to that simplicity. The GHDB is made up of 14 categories of queries and more than 1,200 entries, submitted from a community of hundreds of contributors. The queries run the gamut—they might find error messages that reveal too much about a failed login, or uncover information about online devices like printers and Webcams. Google can also generate much more dangerous results, such as vulnerability data from IDS and firewall logs, or vulnerable Web server versions.

One security expert shed some light on the simplicity of Google hacks: During a short phone call, he showed us how to search Google for remote desktop protocol extensions. Using a particular advanced operator-search term combination, we got 193 results. Clicking on a random return produced a dialogue box asking us if we wanted to open or save the remote desktop. The expert cautioned us not to go further. Had we done so, he said, we likely could have watched someone as they navigated through their desktop.

Weathering the Storm

Type "weather" into a Google search and The Weather Channel's site is the first returned result. From a marketing perspective, director of network architecture John Penrod loves the top ranking. From a security perspective, he realizes the depth of danger that a malicious query against the site, www.weather.com, could bring.

Penrod says his group has a clean track record fending off hackers, due in large part to an efficient QA process, and stringent security and code reviews applied against The Weather Channel's Web development. Experts say those are an enterprise's best defenses against Google hacking.

> Google is a great tool and will facilitate—

"The big thing for us is brand name protection," Penrod says. "The worst thing that could happen is for our site to be attacked or brought down."

Part of any risk assessment is examining liability and risk of information exposure, what your company is willing to share with the rest of the world and what it's willing to lose. It's imperative that companies understand which assets are Web-facing and if they're secure. This is a difficult issue

Photograph by KATHERINE LAMBERT

for good guys and bad—getting what [you're] after.

—Johnny Long, author of *Google Hacking for Penetration Testers*

> "The big thing for us is brand-name protection."
>
> —John Penrod, The Weather Channel's director of network architecture

to contend with for enterprises whose Web presence grows quickly. Strict corporate policies must address what it takes to put an application on the Web, and those edicts must be signed by managers, administrators and developers alike. The process must be policed regularly.

When a new version of The Weather Channel site goes into QA, it's reviewed internally by a team that verifies that established security processes were followed before the site is launched into production. On the back end, the administrative side is kept off the Internet. That's solid strategy. Search engines are indiscriminate and are likely to find a file on the Web that is unprotected or reveals too much.

"[QA teams] look at every file, every directory accessible by the public. Don't assume because you don't see it in browsing that it can't be found," Penrod says. "It's about ensuring the OS is secure, ensuring the Web server application is secure, and the dynamic-content-building process is secure from the ground up."

Protection mechanisms include simple things like dropping into a root directory a text file like robot.txt, a standard file for robot exclusion, which keeps Google's spiders from caching directories. Using meta tags like NOARCHIVE and NOSNIPPET prevents Google from caching specific pages. Applying password protection to applications also keeps Google away.

"Everyone we show Google hacking results to can't believe it," says Dave Shackleford, solution engineering manager with consultancy Vigilar. "They're used to using Google every day; it's an objective tool you can use for innocuous queries. But it's used to find information you'd never want to see as well. That's what takes them aback."

Companies should also dedicate staff time to Google hacking. Run queries against your company and seek out dangerous nuggets before a hacker beats you to it. If sensitive pages are found online, companies should remove the pages from directories or password-protect them. It's also imperative to request via an online form that Google take down such pages from its search results and its cache.

The Catch-22 for sites like The Weather Channel that want strong Google returns is that using exclusionary techniques will damage a site's search engine ranking.

"We definitely want Google to look at our site—it's good for business," Penrod says. "Not looking at our site is better for security. We just have to put everything on a scale and weigh the risks."

Photograph by KEN HAWKINS

Google Gobbledygook

oogle hackers speak a searcher's slang. The following words will help you gab with the Googlers.

Googleturd A search that shouldn't return any results because of a syntax error; or an incorrect query that returns legitimate results.

Googledork An inept person or company whose sensitive information has been revealed by Google.

Advanced operator Special searching techniques offered by Google that enable advanced queries. The syntax of a Google advanced operator is operator:search_term.

NOARCHIVE meta tag A command that prevents Google from including cached links in search results

NOSNIPPET meta tag A command that prevents Google from returning summary information with search results; also prevents Google from caching page

Source: *Google Hacking for Penetration Testers* by Johnny Long

A Cache Cow

The risks are substantial if you fall victim to a Google hack. While it's impossible to estimate how many businesses have fallen prey, the potential figure is staggering.

Hackers troll search engines armed with queries that enable them to do everything from network mapping to carrying out the final phases of an actual attack. In recent months, Long says, newly submitted queries to the GHDB have found Web interfaces for VoIP equipment without login or password protection. Another uncovered an interface that would enable you to turn off a business' lights. It's not unusual to find an exposed Linksys router or Cisco VPN Concentrator management interface. Google hacks aren't parlor tricks.

Hackers love Google because it's anonymous; they can do target reconnaissance without anyone knowing. Google caches every page it crawls, ensuring that a copy is stored somewhere, even if the original has long been pulled from your site. The rub is that while the hacker scans a cached page looking at the leftover, forgotten goodies, there isn't a trace of his steps on your server logs. You'll never know your sensitive data wound up in the wrong hands.

Long cautions that making sure a cached page and the original link to a page are no longer referenced is not enough to keep your data from being accessible via a search engine. Security managers need to ensure that the page summary that appears with each result on the main search page is taken away as well. Hackers can use that snippet to reconstruct portions of a Web page you may not want them to see.

"There's a lot of technology around [caching], but it boils down to the same thing. You need to know what you want to get rid of and be proactive about getting it removed," Long says. "It's not just firing off the remove form to Google, but following it up and using the same techniques bad guys use to make sure it's actually gone."

Defending against Google hacks requires not only a process change, but also shifts in cultural attitudes toward security. Sensitive information often falls through the cracks because Web apps are rushed to market without code reviews or pen tests against a Web infrastructure.

The Ethics of Sharing

Long, a professional pen-tester with Computer Sciences Corp., concedes to a moral dilemma over hosting this type of information on his site. In the end, he says full disclosure wins out.

"People may get affected in a negative way, but open communication fosters more education on all parts," Long says. "Yeah, it helps the bad guys, but after sitting back and watching the discussion unfold about vulnerabilities and whether they should be open, it would be silly to think I'm protecting anyone by sitting on the information."

The GHDB is rolled into a short list of tools that can be modified to automatically run queries against your company's domain. Long has written an open-source tool called Gooscan, which conducts bulk Google searches. Athena is a similar tool that, like Gooscan, is not based on the Google API and is a violation of Google's terms of service. Google has the option of banning a violator's IP range from using its search engine. Other tools like Witko and Foundstone's SiteDigger are based on the Google API and require a license key from Google.

"One of the things we're struggling with is figuring out how public and accessible we make [the GHDB]," Long says. "We're at the point now that we realize there's enough awareness around it. It's high time we start releasing it and making it as open as possible. That was our goal from the beginning—publicize this and raise awareness."

Then there's the question of whether Google has any responsibility not to disclose information that could imperil businesses—beyond honoring remove requests. A Google representative said the company's job is to bring the Internet to users. He declined further comment.

Long agrees that, while Google may have an opportunity to make a business of alerting companies that are being scanned, it doesn't have a responsibility to do so.

"It's not their data; Google doesn't own the data. It's the responsibility of the [business'] security people to keep their own space in order," Long says.

Michael S. Mimoso is senior editor of Information Security. *Send your thoughts on this article to feedback@infosecuritymag.com.*

Find tools to run Google scans against your company's domain at www.searchsecurity.com/ismag