

Using Google to Find Vulnerabilities In Your IT Environment

Attackers are increasingly using a simple method for finding flaws in websites and applications: they Google them. Using Google code search, hackers can identify crucial vulnerabilities in application code strings, providing the entry point they need to break through application security. Sound scary? It is, but there is good news: You can use these same methods to find flaws before the bad guys do. In this special report, we outline methods for using search engines such as Google and Bing to identify vulnerabilities in your applications, systems and services—and to fix them before they can be exploited.

By Francis Brown

Presented in conjunction with

SECURITY
dark READING
Protect The Business  Enable Access



CONTENTS

TABLE OF

- 3 Author's Bio
- 4 Executive Summary
- 5 The Search is on for Vulnerabilities
- 5 Figure 1: Google Results for Common Docs
- 7 New Weapons of Choice
- 7 GoogleDiggity
- 7 BingDiggity
- 8 DLPDiggity
- 8 FOCA
- 9 The Harvester
- 9 Maltego
- 10 Deep Magic
- 10 BingLinkFromDomainDiggity—
Footprint Tool
- 10 Shodan
- 11 FlashDiggity
- 11 PasteBin Leaks
- 11 Conclusion
- 12 Figure 2: Google Hacking: A History
of Key Events
- 13 Defensive Alerts Provide New Protections
- 13 Diggity Alert Feeds
- 14 Diggity Alert Clients
- 15 More Like This



ABOUT US

InformationWeek Reports' analysts arm business technology decision-makers with real-world perspective based on qualitative and quantitative research, business and technology assessment and planning tools, and adoption best practices gleaned from experience. To contact us, write to managing director **Art Wittmann** at awittmann@techweb.com, content director **Lorna Garey** at lgarey@techweb.com, editor-at-large **Andrew Conry-Murray** at acmurray@techweb.com, and research managing editor **Heather Vallis** at hvallis@techweb.com. Find all of our reports at reports.informationweek.com





Francis Brown
Stach & Liu

Francis Brown, CISA, CISSP, MCSE, is a managing partner at Stach & Liu, a security consulting firm that provides IT security services to *Fortune* 500 and global financial institutions, as well as U.S. and foreign governments. Before joining Stach & Liu, he served as an IT security specialist with the Global Risk Assessment team of Honeywell International. There, Francis performed network and application penetration testing, product security evaluations, incident response and risk assessments of critical infrastructure. Prior to that, Francis was a consultant with the Ernst & Young Advanced Security Centers. Francis has presented his research at leading conferences, including Black Hat USA, DEF CON, InfoSec World Conference and Expo, ToorCon and HackCon. He has been cited in numerous industry and academic publications. Francis holds a Bachelor of Science and Engineering degree from the University of Pennsylvania, with a major in computer science and engineering and a minor in psychology. While at Penn, Francis taught operating system implementation and C programming, and participated in DARPA-funded research into advanced intrusion prevention system techniques.

SUMMARY

EXECUTIVE

Google, Bing and other major search engines, have made it easy to find all manner of information—including everything from exposed password files to SQL injection points. This led to the emergence of Google hacking, a technique used to identify and then exploit system and data vulnerabilities. Google hacking's popularity waned in the last few years, due in large part to Google shutting down the Google SOAP API. However, with aggressive R&D efforts fueled by innovative thinking, as well as significantly more data available on the Web and stored in the cloud, Google hacking is on the rise again. While this gives IT security professionals yet another battle to fight, the good news is that they can leverage the very tools and techniques hackers use to identify and fix any vulnerabilities their companies may have. In other words, they can Google themselves to find security problems before the bad guys do. In this report we will examine a slew of new tools and techniques that will allow security professionals to leverage Google, Bing, Baidu and other open search interfaces to proactively track down and eliminate sensitive information disclosures and vulnerabilities in public systems and also take a look at defensive tools designed to pull thousands of real-time RSS updates from search engines to provide users with alerts—a sort of intrusion detection system (IDS) for Google hacking. Malicious hackers have already embraced search engine hacking as an effective way to target and exploit vulnerabilities on a massive scale. It is imperative that security professionals learn to take equal advantage of these techniques to help safeguard their organizations.

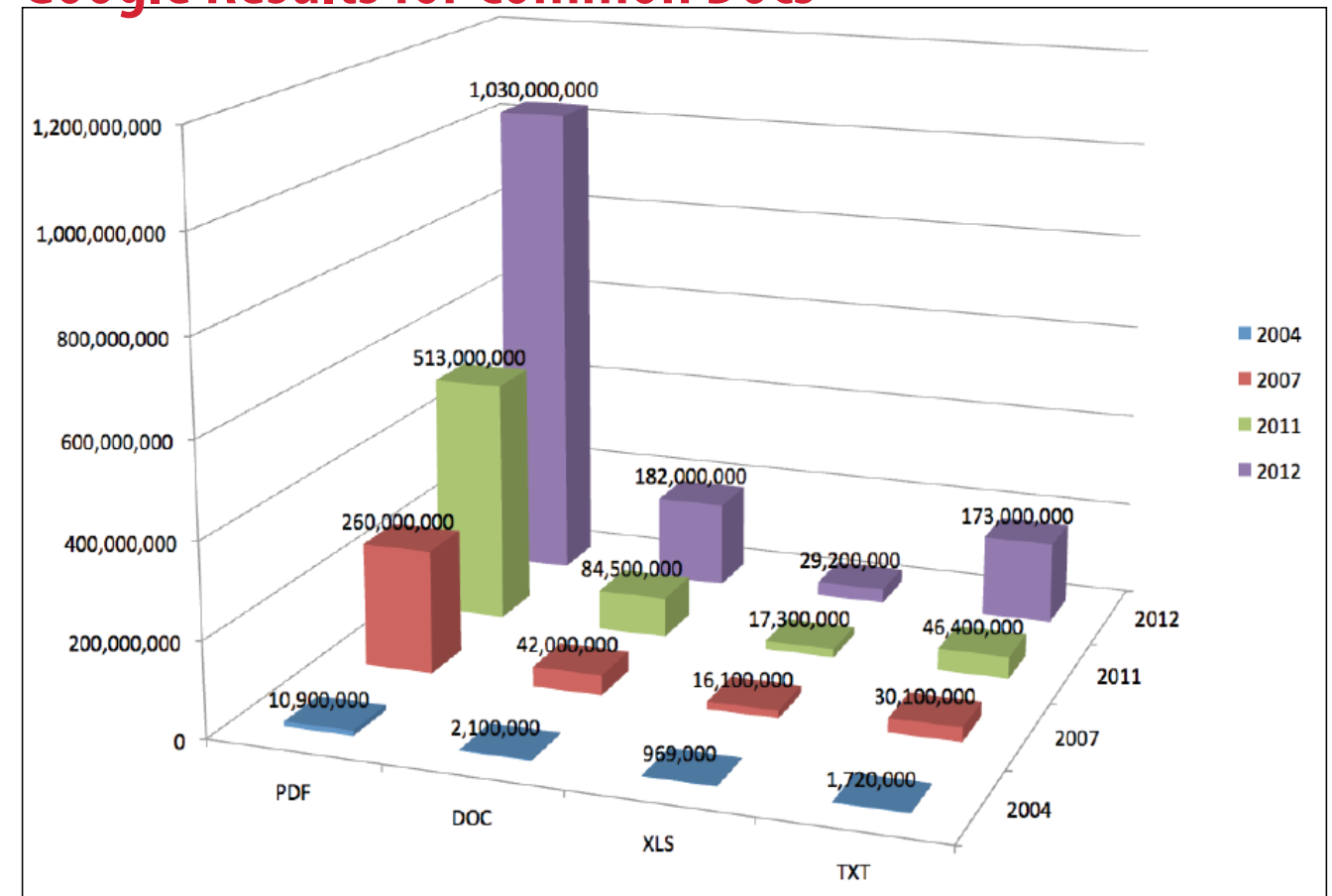
The Search is on for Vulnerabilities

The vast volumes of information available on the Internet are of great value to businesses—and to hackers. For years, hackers have been using Google and other search engines to identify vulnerable systems and sensitive data on publicly exposed networks. The practice, known as Google hacking, has seen a resurgence of late, providing new challenges for IT professionals striving to protect their companies from threats growing in number and sophistication.

Google hacking—a term used for penetration testing using any search engine—surged in popularity around 2004, when computer security expert Johnny Long first released his book *Google Hacking for Penetration Testers* and the Google Hacking Database (GHDB). The database was designed to serve as a repository for search terms, called Google-Dorks, that exposed sensitive information, vulnerabilities, passwords and much more. The beginning of the end for the Google hacking tools available at the time came when Google stopped issuing new Google Simple Object Access Protocol

Figure 1

Google Results for Common Docs



The rate of growth of the documents available for data mining via Google has risen exponentially in the last few years. Source: Stach & Liu

(SOAP) API keys in. The Google SOAP API was the interface that hacking tools used to make Google queries. Google retired the SOAP API to make room for its Google Asynchronous JavaScript and XML (AJAX) API. Google hacking tools hobbled along until September 2009, when Google closed down the Google SOAP API entirely.

There recently has been an upswing in Google hacking, with a few factors playing a role in the practice's growth. For one thing, the amount of data indexed and searchable by Google and other search engines has skyrocketed in the last few years. Simply put, this has given hackers much more to work with. The rate of growth of the documents available for data mining via Google is illustrated in Figure 1, page 5. There has also been a significant increase in new search engine interfaces to various types of data. Examples include Google Health and Google Code Search (recently [shut down](#)), which have allowed users to search through patient health records and open source code projects, respectively. We've also seen the emergence of security-specific search



As the trend of moving data storage to the cloud continues, so does the risk of data loss. Most cloud-based storage offerings are not immune to Google hacking. As shown in the screen, Google can be used to find documents stored in major cloud providers such as Dropbox.

engines such as [Shodan](#) come into play. These search engines were created specifically for use by penetration testers to identify vulnerable Web applications on the Internet.

We saw a frightening example last year of just how effective Google hacking can be

when the group LulzSec used Google hacking techniques to go on an epic spree that left in its wake a number of victims, including Sony, PBS, Arizona's Department of Public Safety, FBI affiliate InfraGard and the CIA. In other news, the recent LizaMoon mass SQL injection

attack was reported to have affected as many as 4 million websites, a large step up in scale from the few hundred thousand sites affected by the mass SQL injection attack from the previous June that most notably compromised *The Wall Street Journal* and *The Jerusalem Post* websites.

New Weapons of Choice

With the retirement of Google's SOAP API in 2009, most of the security utilities available for

Google hacking ceased to function, leaving the security industry with a need for new and innovative tools. It did not take long for the vacuum to be filled. For example, in November 2010, the Exploit

Database (exploit-db.com) took over maintenance of the Google Hacking Database and began adding new Google Dorks, the search terms that reveal vulnerabilities and sensitive information disclosures.

However, what's good for hackers is also

good for corporate security professionals, who can make use of Google hacking tools and other resources to identify—and then eliminate—vulnerabilities in their data systems. The majority of these tools are free and easy to use, although experience and skills associated with Web application security are helpful in validating scan results and identifying which results are real security issues and which are not.

There is no one tool that will serve as a silver bullet in eliminating search engine exposures. We encourage security professionals to try out and regularly use as many as possible to gain as much security coverage as possible over their network perimeter.

Here is an overview of some of the tools currently available.

>> GoogleDiggity

GoogleDiggity is a utility from Stach & Liu's [Google Hacking Diggity Project](#) that leverages the Google JSON/ATOM Custom Search API, which means users will not be blocked by

Google bot detection while scanning. Also, unlike other Google hacking tools available, GoogleDiggity allows users to specify a Google Custom Search Engine (CSE) ID. This allows Google hacking vulnerability checks to be run against a customized version of Google that will return results tailored to specific organizations.

>> BingDiggity

Also from the Google Hacking Diggity Project, BingDiggity uses the Bing 2.0 API and Stach & Liu's newly developed Bing Hacking Database (BHDB) to find vulnerabilities and sensitive information disclosures that are exposed via Microsoft's Bing search engine. This utility also provides footprinting capabilities that allow users to enumerate URLs, hosts, domains and IP-to-virtual host mappings for target companies. In the past, MSN/Bing hacking utilities were limited to footprinting techniques because Microsoft took steps to prevent traditional hacking techniques via Bing back in 2007. The BHDB

Security professionals can expect to see a continued increase in development of new tools being released that tap into data indexed by search engines such as Google and Bing.



Follow



Like

was created so that the Bing search engine could be used to discover vulnerabilities within target applications and infrastructure. It provides more than 1,300 Bing searches that return vulnerability data.

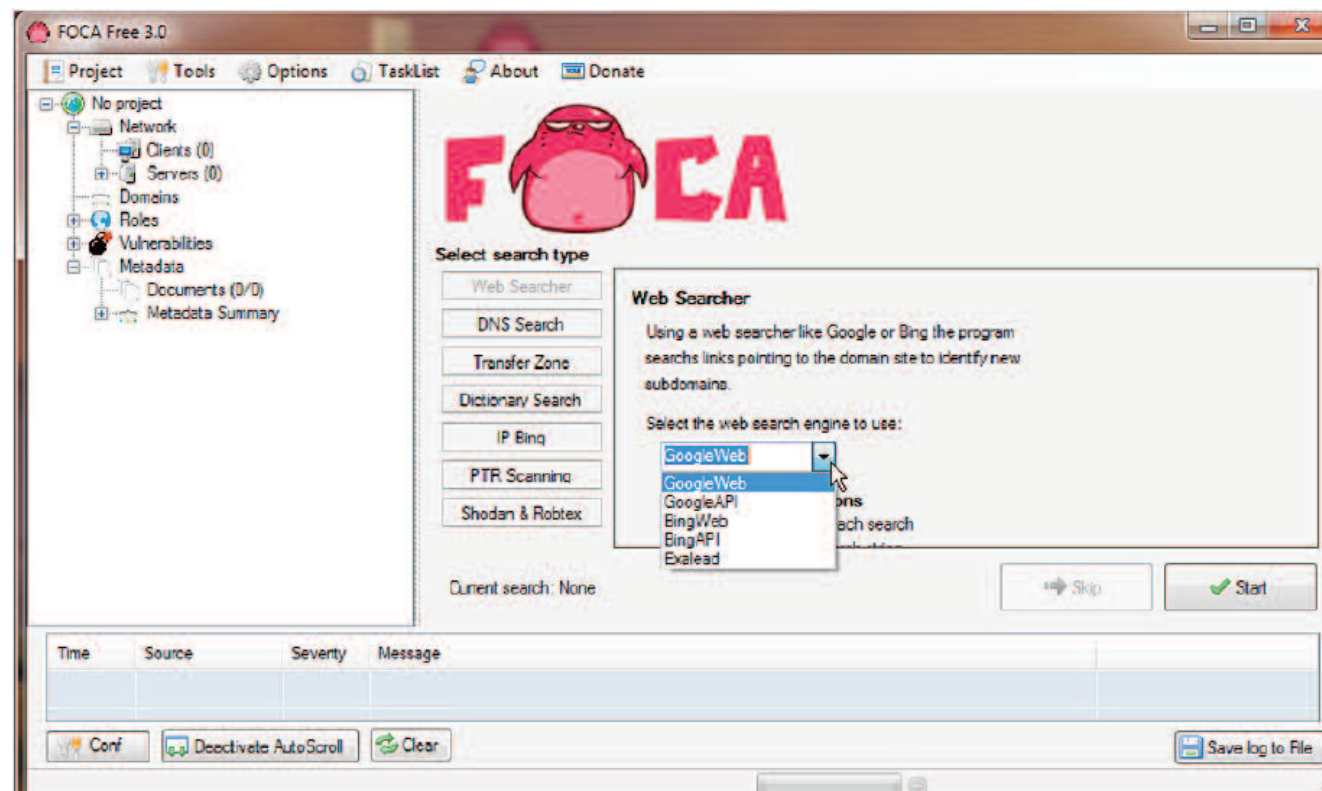
>> DLPDiggity

DLPDiggity is a data loss prevention (DLP) tool that leverages Google/Bing to identify exposures of sensitive information, such as Social Security and credit card numbers, via common document formats including .doc, .xls and .pdf. GoogleDiggity or BingDiggity are first used to locate and download files belonging to target domains/sites on the Internet, then DLPDiggity is used to analyze those downloaded files for sensitive information disclosures. DLPDiggity uses iFilters to search through the actual contents of files, as opposed to just the metadata. Using .NET regular expressions, DLPDiggity can find almost any type of sensitive data within common document file formats.

Like This Report? Rate It!

Something we could do better? Let us know.

Rate



FOCA is a free fingerprinting tool that can be used for information gathering on the Web.

>> FOCA

FOCA FREE 3.0.2 is a tool used to carry out information gathering fingerprinting in Web audit work. The free version of the tool en-

ables users to find servers, domains, URLs and documents published, as well as versions of software on servers and clients. FOCA became famous for extracting meta-

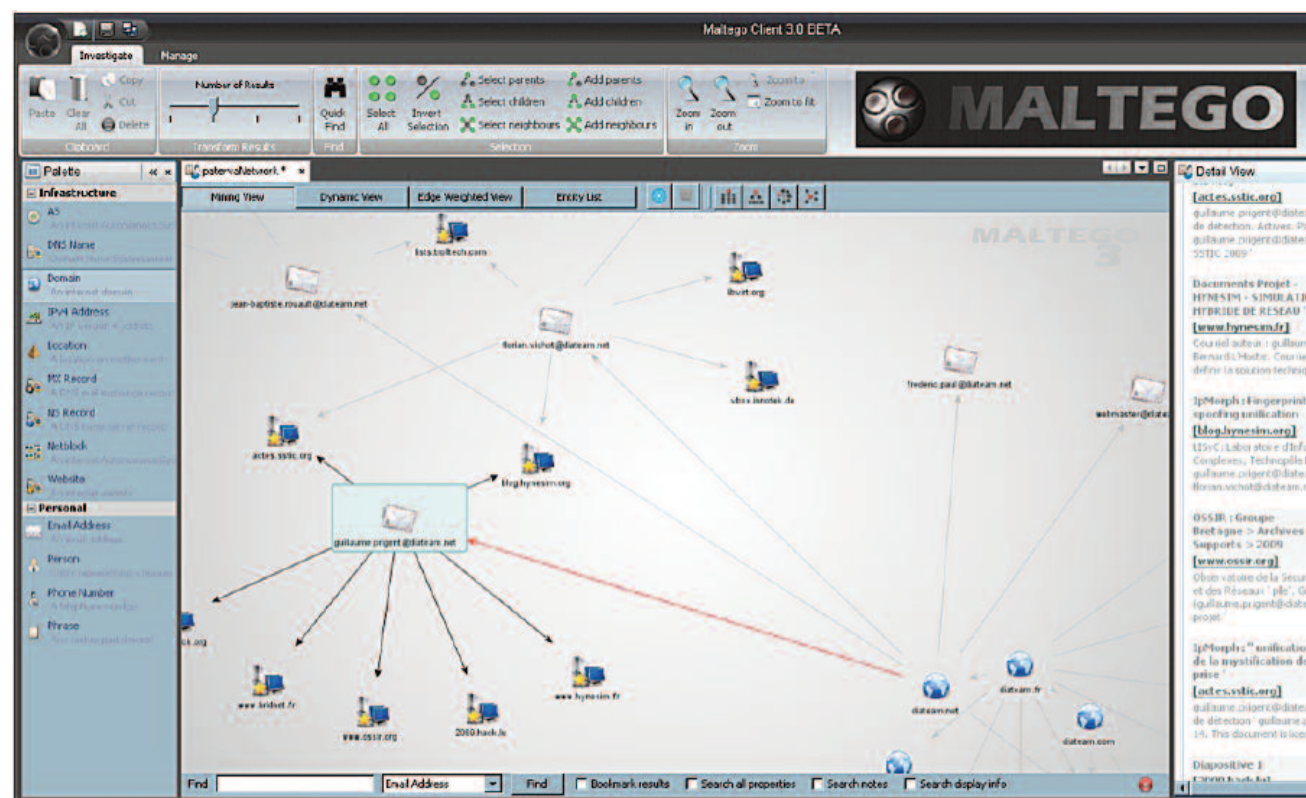
data of public documents, but today the tool is capable of much more than that.

>> TheHarvester

Users have always been, and will always be, the weakest link in the security chain. In the past few years, there has been an explosion of personal information made readily available on the Web. That information is easily searchable on sites such as LinkedIn and Facebook. Getting a list of executives, network administrators, developers or even IT security staff associated with a target company is now a pretty trivial thing to do. TheHarvester gathers emails, subdomains, hosts, employee names, open ports and banners from different public sources, including search engines, PGP key servers and the SHODAN computer database.

>> Maltego

We've seen the emergence of several new toolkits, such as the Social Engineering Toolkit (SET), that have helped automate (and dumb down) the process of mounting what were once highly sophisticated phishing and social



Tools like Maltego reduce the complexities involved with phishing and social engineering attacks.

engineering attacks. Maltego is an open source intelligence and forensics application that enables the mining and gathering of information, as well as the representation of this information in a meaningful way. The unique

perspective that Maltego offers to both network- and resource-based entities is the aggregation of information posted all over the Internet. Whether it's the current configuration of a router poised on the edge of your

network or the current whereabouts of a company vice president on an international visit, Maltego can locate, aggregate and visualize this information.

>> DeepMagic

DeepMagic DNS is a security-specific search engine that allows attackers to quickly build a footprint of a target organization by allowing them to easily develop a list of odds names and live hosts.

>> Bing LinkFromDomainDiggity—Footprinting Tool

Bing LinkFromDomainDiggity is an attack footprinting tool that leverages Bing's [linkfromdomain: directive](#) to find off-site links. From those results, the tool then enumerates lists of applications, host names, domains and subdomains related to an attack target. We used Bing LinkFromDomainDiggity to build out a list of targets on the Chinese government's domains by analyzing the offsite links of the Chinese government's main website. (see screenshot at right)

>> Shodan

Shodan is a search engine that lets users find specific computers (such as routers or servers) using a variety of filters. Some have also described Shodan as a public port scan

directory or a search engine of banners. Shodan can be used, for example to target supervisory control and data acquisition (SCADA) systems that control the critical infrastructure of a country.

The screenshot shows the Bing LinkFromDomainDiggity tool interface. At the top, there are several tabs: GoogleDiggity, CodeSearchDiggity, BingDiggity, LinkFromDomainDiggity (highlighted), DLPDiggity, FlashDiggity, and MalwareDiggity. Below the tabs is a 'Query Appender' field containing 'site:gov.cn'. To the right of the field are 'SCAN' and 'Cancel' buttons. Further right is a 'Sites/Domains' field containing 'www.gov.cn [Remove]' and an 'Add' button. Below the 'Query Appender' field is a 'Hosts' tab with a list of results: 2010.visithainan.gov.cn, app.mps.gov.cn, bg.mofcom.gov.cn, bjsat.gov.cn, bjyouth.gov.cn, catf.agri.gov.cn, and cc.fjkd.gov.cn. To the right of the 'Hosts' list is a 'Domains' tab. Below the 'Hosts' list is an 'Output' section with the following text: 'Using [redacted] F9367F8FD32. Advanced Scan started [9/10/2011 2:16:54 PM] Found 445 result(s) for query: 'linkfromdomain:www.gov.cn site:gov.cn'. Total Results: 445. Scan Complete. [9/10/2011 2:17:26 PM]'. Three red callout boxes provide annotations: 1. 'Running Bing's linkfromdomain:www.gov.cn to get list of off-site links from China's government main website' (pointing to the 'Sites/Domains' field). 2. 'Also filtering results to just those also part of the gov.cn domain' (pointing to the 'Query Appender' field). 3. 'Results in large list of other valid Chinese government hostnames on the gov.cn domain' (pointing to the 'Hosts' list).

We used Bing LinkFromDomainDiggity to build out a list of targets on the Chinese government's domains.



Ensuring Secure Database Access

Role-based access control based on least user privilege is one of the most effective ways to prevent the compromise of corporate data. But proper provisioning is a growing challenge, due to the proliferation of big data, NoSQL databases and cloud-based data storage.

Download

>> FlashDiggity

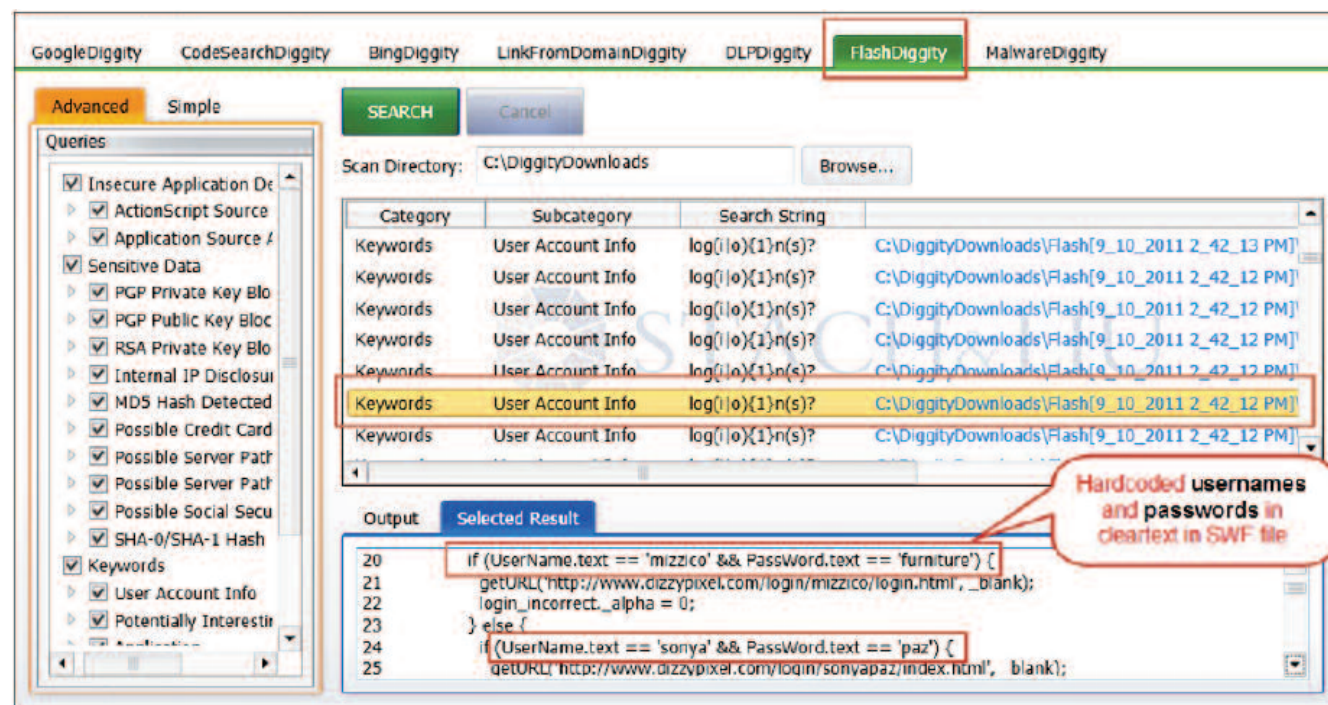
FlashDiggity automates Google searching/downloading/decompiling/analysis of **SWF files** to identify Flash vulnerabilities and information disclosures. FlashDiggity first leverages the GoogleDiggity tool to identify **Adobe Flash SWF** applications for target domains via Google searches, such as `ext:swf`. Next, the tool is used to download all of the SWF files in bulk for analysis. The SWF files are disassembled back to their original **ActionScript** source code, and then analyzed for code-based vulnerabilities.

>> PasteBin Leaks

Information disclosures for your organization may occur on sites unrelated to your organization. For example, **PasteBin leaks** is a Twitter feed that tracks the disclosure of sensitive information via the Pastebin.com website.

Conclusion

Security professionals can expect to see a

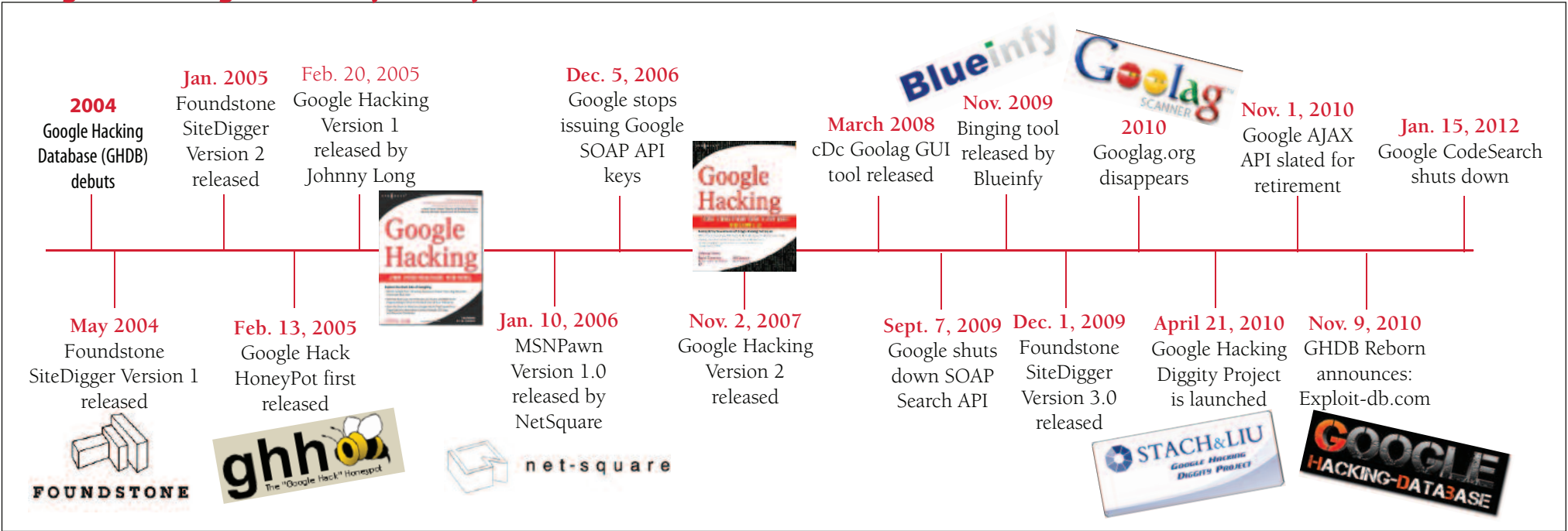


FlashDiggity can be used to identify vulnerabilities in Flash applications.

continued increase in development of both innovative methods and new tools being publically released that tap into the vast amounts of data indexed by search engines such as Google and Bing. Opportunistic hackers will undoubtedly be employing these new

tools to target the low-hanging (and not-so-low-hanging) fruit of vulnerabilities exposed on the Web. It is therefore imperative that security professionals learn to take equal advantage of these techniques to help safeguard their organizations.

Figure 2
Google Hacking: A History of Key Events



Defensive Alerts Provide New Protections

Defensive strategies for protecting your organization from Google hacking attacks traditionally have been limited, mostly falling back on the approach of “Google hack yourself.” This approach has several shortcomings. While a few free tools exist that allow security professionals to Google hack their organization, they are typically inconvenient, utilizing only one search engine and providing only a snapshot in time of your organization’s exposure. Security consultancy Stach & Liu, which works with *Fortune* 500 and government agencies, has created free defensive tools to help protect your organization from exposing vulnerabilities via Google, Bing and other popular search engines. The tools fall into two major categories—alert RSS feeds and alert RSS monitoring tools. Together, they form a type of intrusion detection system (IDS) for Google hacking.

Diggity Alert Feeds

Google Hacking Alerts provide real-time vulnerability updates via convenient RSS feeds.



Bing Hacking Alerts employs a similar approach, but instead leverage Stach & Liu’s Bing Hacking Database (BHDB) in conjunction with Microsoft Bing’s &format=rss directive to turn Bing searches into RSS feeds.

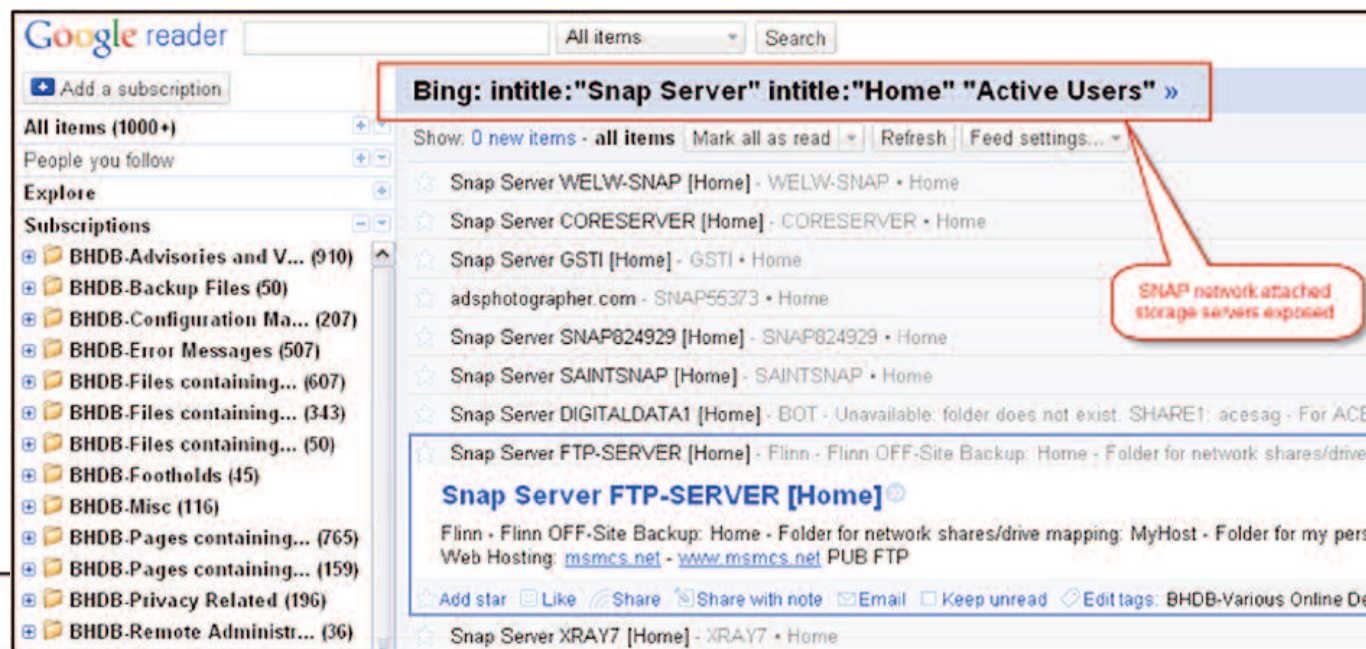


These feeds in turn produce RSS entries that identify vulnerable Web applications on the Internet.

Google Alerts have been created for all 1,623 GHDB/FSDB search strings, generating a new alert each time newly indexed pages match regular expressions.

Diggity Alert Clients

If you would like to filter these alerts on domains and other criteria, making them more relevant to sites you care about, you can use either the Microsoft Windows desktop client (AlertDiggity) or iPhone client (iDiggity Alerts) to filter these alerts. —Francis Brown



These alerts have accumulated several thousand hits per day, and now form what is likely the largest repository of live vulnerabilities on the Internet.

MORE
LIKE THIS
MORE

Want More Like This?

InformationWeek creates more than 150 reports like this each year, and they're all free to registered users. We'll help you sort through vendor claims, justify IT projects and implement new systems by providing analysis and advice from IT professionals. Right now on our site you'll find:

Strategy: Database Defense: The biggest threat to your company's most sensitive data may be the employee who has legitimate access to corporate databases but less-than-legitimate intentions. And while the incidence of insider data breaches has decreased, external attacks often imitate them—and do serious damage. Follow our advice to mitigate the risk.

Strategy: Understanding Software Vulnerabilities: To protect company and customer data, we need to determine what makes it so vulnerable and appealing. We also need to understand how hackers operate, and what tools and processes they rely on. In this report we explain how to ensure the best defense by thinking like an attacker and identifying the weakest link in your own corporate data chain.

Private Cloud Blueprint: We outline plans to help you move beyond server virtualization to build a more flexible and efficient data center. From network and storage virtualization to automation and orchestration to promoting your private cloud services, our guide takes you through the essentials needed to build a private cloud.

PLUS: Find signature reports, such as the *InformationWeek Salary Survey*, *InformationWeek 500* and the annual State of Security report; full issues; and much more.

Newsletter

Want to stay current on all new *InformationWeek* Reports? Subscribe to our weekly newsletter and never miss a beat.

[Subscribe](#)