

Utilizando Google Hacking para encontrar vulnerabilidades em sites

Bruno Rodrigo Barbosa Cortes⁽¹⁾; Carlos Eduardo de Carvalho Dantas⁽²⁾

⁽¹⁾ Estudante, Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro, Uberlândia, Minas Gerais, brudrigo@gmail.com; ⁽²⁾ Professor; Instituto Federal do Triangulo Mineiro; Uberlândia; Minas Gerais; carloseduardodantas@iftm.edu.br.

RESUMO: Este trabalho apresenta uma análise de como a configuração indevida de um servidor pode expor informações sensíveis de uma empresa na base de dados do Google e demonstra as técnicas utilizadas pelos hackers para explorar estas falhas, Google Hacking. Estar bem classificado nas pesquisas do Google é um dos principais objetivos de uma empresa que busca visibilidade na internet, entretanto, o mecanismo de indexação do Google poderá registrar informações sensíveis de sua empresa e abrir uma porta para a ação de hackers maliciosos. O trabalho realiza a exposição do método utilizado por hackers para obtenção de informações sensíveis e a descoberta de possíveis alvos que utilizam softwares vulneráveis.

Palavras-chave: Google Hacking, Segurança da Informação, Ataques Cibernéticos.

INTRODUÇÃO

Com o advento da internet e sua constante expansão, tornou-se fundamental a presença das empresas na rede mundial de computadores, seja para expor trabalhos, atrair clientes, fornecer serviços e atividades de comércio online (*e-commerce*), quanto para outras muitas finalidades que tornam as empresas cada vez mais dependentes das facilidades providas pela conectividade da internet. Entretanto, publicar um site ou serviço web não é o suficiente para atrair visitantes, é necessário ter visibilidade, ou seja, ser visto pelo seu público-alvo. Diante desta demanda, surgiram os sites para busca de conteúdo web, que tem por objetivo retornar o conteúdo relacionado a demanda de um usuário.

Dentre este universo, o *Google* é, atualmente, o buscador mais usado, estando à frente de outros concorrentes como *Ask*, *Yahoo* e *Bing*. O buscador *Google* se destaca dos demais por sua eficiente atualização e classificação de informações. Sua base de informações é diariamente atualizada por meio de seu *crawler*, o *Googlebot*, um “robô” que varre a rede mundial de computadores em busca de informações novas.

Entretanto, as facilidades providas pelos buscadores de conteúdo web também são o pivô para uma série de ataques cibernéticos, pois, assim como são eficientes ferramentas para pesquisa de conteúdo, possibilitam a *hackers* maliciosos encontrar vulnerabilidades conhecidas e realizar ataques a diversos alvos pela rede.

Neste trabalho, será exposto como os criminosos utilizam o *Google* para obter acesso a informações sensíveis e encontrar alvos para vulnerabilidades conhecidas, bem como, será apresentado alternativas para proteção desta exposição indevida.

MATERIAL E MÉTODOS

Conforme PAIVA (2015), a pesquisa no Google não se fundamenta especificamente na busca por informações sensíveis como usuários e senhas, mas se fundamenta no que é procurado, buscando usar essas informações para seus próprios objetivos. Encontrar informações sensíveis faz parte da rotina de um *Google Hacker*, que pode utilizar o *Google* na busca de servidores negligenciados, diretórios expostos, relatórios de segurança expostos e na busca de informações pessoais e documentos compartilhados por engano na Internet como: planilhas, tabelas, vídeos, documentos do *Word*, fotos, bancos de dados e outros arquivos que possuam alguma informação relevante.

Segundo LONG (2004), o *Google* permite o uso de certos operadores para ajudar a refinar as pesquisas. A utilização de técnicas avançadas com operadores é muito simples, desde que seja dada atenção à sintaxe.

Com o emprego de algumas técnicas, é possível otimizar as pesquisas feitas no Google. Os operadores de busca nada mais são que convenções definidas pelo próprio buscador para auxiliar quem procura por resultados avançados. A pesquisa é feita na tradicional caixa de busca do Google, porém, com alguns códigos adicionais inseridos antes dos termos utilizados. Um dos recursos mais poderosos do Google, e ao mesmo tempo desconhecidos pela maioria dos usuários, são os ditos “operadores avançados”. Na confecção deste trabalho foram levantados os principais operadores avançados nas obras de JOHNNY LONG, *The Google Hacker’s Guide. Understanding and Defending Against the Google Hacker*, de 2004, e *Google Hacking for Penetration Testers. Google Hacking: Teste de Invasão*, de 2007, e nos sites: www.johnny.ihackstuff.com/ghdb, www.exploit-db.com/google-dorks e www.googleguide.com/advanced_operators_reference.html. Os principais exemplos de operadores, no contexto deste trabalho, são:

1. Subtrair resultado

Deve-se adicionar um traço (-) antes de uma palavra ou um site para excluir todos os resultados que incluem essa palavra. Isso é útil especialmente para diferenciar palavras com vários significados.

Exemplo: Eleições –governador, Gol –carro

2. Pesquisa exata

Usam-se aspas para pesquisar uma palavra exata ou um conjunto de palavras em uma página da Web. Termos entre aspas filtram a busca somente para resultados exatos, ou seja, exatamente como o pesquisador está procurando. Deve ser usado somente se estiver procurando uma palavra ou frase exata. Caso contrário, a busca excluirá muitos resultados úteis por engano.

Exemplo: “Luiz Fernando da Costa”³⁴

3. Curingas

Usa-se um asterisco em uma pesquisa como um marcador para termos desconhecidos ou caracteres coringa. Aspas podem ser usadas para encontrar variações da frase exata ou para lembrar palavras no meio de uma frase.

Exemplo: “Forças * revolucionárias da **”

4. Busca alternativa

Usa-se “OR” quando se deseja pesquisar páginas que contenham apenas uma entre várias palavras, deve-se incluir “OR” (em maiúsculas) entre as palavras. Sem o “OR”, os resultados normalmente mostram somente páginas correspondentes a ambos os termos.

5. Restringindo pesquisa a site específico

Se o pesquisador incluir o operador “site” em sua consulta, o Google irá restringir os resultados da pesquisa do site ou domínio que o pesquisador especificar. Por exemplo, é possível encontrar todas as referências a “terrorismo” no website da BBC.

Exemplos: terrorismo site:bbc.co.uk/portuguese

6. Buscando por cache

Caso o pesquisador utilize o operador “cache”, será exibida a versão de uma página web em cache do Google correspondente ao termo buscado. Este operador permite visualizar como estava a página na última vez que o Google rastreou o site.

Exemplo: cache:www.mpl.org.br

7. Buscando por tipo de arquivo

Caso o pesquisador utilize o operador “filetype” este se trata de um recurso empregado para selecionar o tipo de arquivo que se deseja em uma pesquisa. Busca apenas em arquivos de um tipo específico. Este operador instrui o Google para pesquisar apenas dentro do texto de um determinado tipo de arquivo. Este operador requer um argumento adicional da busca.

Exemplo: download Constituição Federal filetype:pdf

8. Buscando termos no texto de um documento

Caso o pesquisador utilize o operador “intext” os resultados serão restritos a documentos que contenham o termo no texto. O comando abaixo retornará documentos que mencionam a palavra “terremoto” no texto, e mencione os nomes “Missão”, “Paz” e “Haiti” em qualquer parte do documento (texto ou não).

Exemplo: Missão de Paz Haiti intext:terremoto

9. Buscando termos simultâneos em um texto

Caso o pesquisador utilize o operador “allintext” o Google restringirá os resultados para aqueles que contenham todos os termos da consulta que o pesquisador especificar no texto da página. O comando abaixo retornará somente as páginas em que as palavras “Exército”, “fronteira” e “operação” aparecem no texto da página.

Exemplo: allintext:Exército fronteira operação

10. Buscando termo em um título de documento

Caso o pesquisador utilize o operador “intitle” restringirá os resultados a documentos que contenham o termo no título. Este comando faz com que o sistema de buscas foque somente no título das páginas dos sites indexados para encontrar os resultados relevantes para o pesquisador. O comando abaixo retornará documentos que mencionam a palavra “amazônia brasileira” em seus títulos, e mencione as palavras “garimpo” e “ilegal” em qualquer parte do documento.

Exemplo: garimpo ilegal intitle:amazônia brasileira

11. Buscando termos simultâneos em um título de um documento

Caso o pesquisador utilize o operador “allintitle” o Google restringirá os resultados para aqueles que contenham todos os termos da consulta que o pesquisador especificar no título. O

comando abaixo retorna somente documentos que contenham as palavras "FARC" e "terrorismo" no título. Isso é equivalente a uma série de pesquisas 'intitle' individuais.

Exemplo: `allintitle:FARC terrorismo`

12. Buscando termo em uma URL

Caso o pesquisador utilize o operador "inurl" em sua consulta, o Google irá restringir os resultados a documentos que contenham essa palavra na URL. Este operador instrui o Google a pesquisar somente dentro da URL ou endereço web de um documento.

Exemplo: `inurl:admin senha`

13. Buscando termos simultâneos em uma URL

Caso o pesquisador utilize o operador "allinurl" o Google restringirá os resultados para aqueles que contenham todos os termos da consulta que o pesquisador especificar na URL. O comando abaixo mostrará somente documentos que contenham as palavras "black" e "bloc" na URL.

Exemplo: `allinurl:black bloc`

14. Buscando termo uma localidade específica

Caso o pesquisador utilize o operador "location" em sua consulta no Google, apenas artigos do local que o pesquisador especificar serão devolvidos. O comando abaixo mostrará artigos que correspondam ao termo "eleições" de sites no Brasil.

Exemplo: `eleições location:brasil`

RESULTADOS E DISCUSSÃO

A fim de demonstração da técnica utilizada por *Hackers* na identificação de sistemas vulneráveis utilizando os serviços de busca do *Google*, iremos considerar a situação hipotética de que um *hacker* encontrou uma vulnerabilidade no Portal Padrão adotado pelo Governo Brasileiro e deseja utilizar o *Google* para encontrar outros sites que adotam este sistema e possuem a vulnerabilidade encontrada. A vulnerabilidade fictícia seria a possibilidade de *Sql Injection* (quando o atacante consegue inserir uma série de instruções SQL dentro de uma consulta através da manipulação das entradas de dados de uma aplicação) na página de contatos do Portal Padrão, disponível em: www.portalpadrao.gov.br/contact-info.



Figura 1 – Página de contato do Portal Padrão do Governo Brasileiro.

Uma vez encontrada a vulnerabilidade, o *hacker* deverá construir uma *Dork*, combinação de termos e operadores avançados de pesquisas que retornará os sites vulneráveis a falha pesquisada. Neste exemplo, o *hacker* buscará por características da página que servirão como parâmetros de

pesquisa para os operadores do *Google* a fim de filtrar os resultados encontrados para os sites que adotam o Portal Padrão do Governo Brasileiro e possuem a página de contato vulnerável.

Todas as páginas do Portal Padrão possuem, por padrão, o seguinte texto no rodapé “Desenvolvido com o CMS de código aberto Plone” o que possibilita a utilização do operador *intext* para procurar as páginas que possuem o trecho pesquisado. Entretanto, apenas este critério não será necessário para encontrar as páginas vulneráveis, uma vez que outros sites, que não utilizam o Portal Padrão, também foram retornados na pesquisa.

Para direcionar a pesquisa aos resultados desejados, será adotado um segundo critério: a inclusão do operador *inurl* que irá filtrar os resultados para as páginas que, além do primeiro critério, possuem “/contact-info” em sua url. Desta maneira, o *Dork* utilizado para retornar o conteúdo desejado seria: `intext:"Desenvolvido com o CMS de código aberto Plone" inurl:/contact-info`.

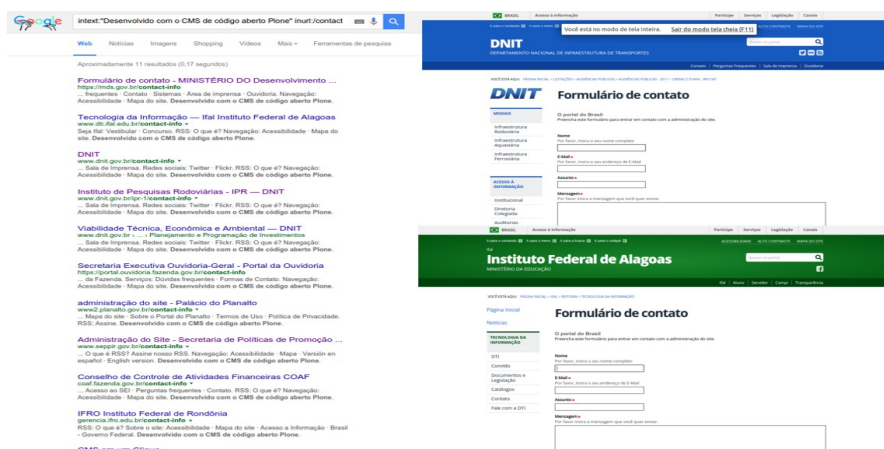


Figura 2 – Exemplo de pesquisa que retornariam páginas, hipoteticamente, vulneráveis.

Na Figura 2 podemos observar o retorno obtido com a utilização do *Dork* construído e exemplos de sites que estariam vulneráveis a falha encontrada.

Na internet estão disponíveis sites como o *Google Hacking Database* (<https://www.exploit-db.com/google-hacking-database/>) que possuem um banco de dados de *Dorks* pré-definidas para encontrar sistemas com vulnerabilidades conhecidas.

Desta maneira, pode-se observar como a ferramenta de pesquisa do *Google* torna-se um eficiente aliado aos *hackers* e criminosos cibernéticos. Para se proteger destas ameaças, torna-se necessário proteger a indexação de conteúdo pelos *Googlerobots*, o que pode ser feito com a correta configuração do arquivo *robots.txt* na raiz da aplicação, documento que orienta o que deve e o que não deve ser indexado pelos sites de busca em seu sistema. Além disto, deve-se buscar disfarçar características das tecnologias utilizadas, a fim de evitar a fácil identificação em buscas realizadas, por exemplo, no caso acima, os sites que utilizam o Portal Padrão e alteraram o texto exibido no rodapé não são exibidos nos resultados de pesquisa, entretanto, estariam igualmente vulneráveis.

CONSIDERAÇÕES FINAIS

Segurança da informação deve ser uma preocupação constante nos dias atuais, visto que estão, cada vez mais constantes, ataques cibernéticos como, por exemplo, ataques de sequestro de dados, onde criminosos criptografam dados do usuário e solicitam um pagamento para liberação da senha de acesso, o que pode resultar em sérios prejuízos a uma empresa ou pessoa física.

Neste trabalho, pode-se constatar a técnica utilizada por *hackers* para combinar buscadores de conteúdo web, como o *Google*, para identificação de sistemas vulneráveis e mal configurados. Conclui-se que é importante manter atualizadas as tecnologias utilizadas nos sistemas, bem como,

as configurações adequadas para se evitar a exposição de conteúdos indevidos na rede mundial de computadores.

AGRADECIMENTOS

Agradeço a Deus, a minha família por todo apoio durante meus estudos e ao meu orientador pelo apoio prestado na elaboração deste trabalho.

REFERÊNCIAS

DHANJANI, Nitesh; RIOS, Billy e HARDIN, Brett. **Hacking: A Próxima Geração**. Editora: Alta Books. Rio de Janeiro, 2011.

LONG, Johnny. 2007. **Google Hacking for Penetration Testers. Google Hacking: Teste de Invasão**. Rockland, Massachusetts, EUA: Syngress.

PAIVA, Newton. **Google Hacking**. Disponível em <<http://blog.newtonpaiva.br/pos/wp-content/uploads/2013/04/PDF-E6-SI491.pdf>>. Acesso em 22 Set 2015.

TOFFLER, Alvin. **The Third Wave (A Terceira Onda)**: tradutor João Tavora, 4a Edição, Rio de Janeiro, RJ, Record, 1980.

TOFFLER, Alvin e TOFFLER, Heidi. **Guerra e antiguerra: sobrevivência na aurora do terceiro milênio**. Vol. 302. Tradução de Luiz Carlos do Nascimento Silva. Rio de Janeiro, RJ, Biblioteca do Exército, 1995.