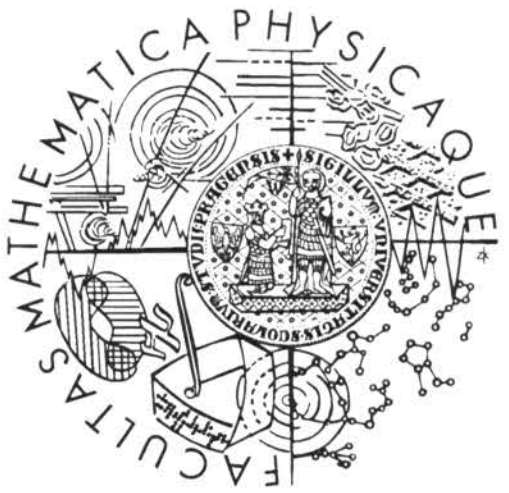
Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Norbert Hanuska

Google hacking

Katedra softwarového inženýrství

Vedúci bakalárskej práce: Doc. RNDr. Pavel Pyrih, CSc.,

Katedra matematické analýzy

Štúdijný program: Informatika, Obecná informatika

2008

Poďakovanie

Moje poďakovanie patrí predovšetkým vedúcemu práce Doc. RNDr. Pavlovi Pyrihovi, CSc. za odborné vedenie a pripomienky k danej téme.

Prehlasujem, že som svoju bakalársku prácu vypracoval samostatne a použil k tomu len literatúru, ktorú uvádzam v zozname priloženom k bakalárskej práci. Súhlasím so zapožičiavaním práce a jej zverejňovaním.

V Prahe dňa 5. 8. 2008 Norbert Hanuska

Norbert Hanuska

OBSAH

ÚVOD **5**

1 VYHĽADÁVANIE NA INTERNETE **6**

1.1 HISTÓRIA A SÚČASNOSŤ GOOGLU 6

1.2 WEBOVÉ ROZHRANIE GOOGLU 8

1.3 VYHĽADÁVACIE TIPY, NEBEZPEČNÉ GOOGLE SLUŽBY 9

1.4 VYTVÁRANIE DOTAZOV 10

1.4.1 Základné pravidlá pri vyhľadávaní pomocou Googlu 10

1.5 BOOLEOVSKÉ OPERÁTORY 12

1.6 ADRESY URL GOOGLU 13

1.7 POKROČILÉ OPERÁTORY 14

2 ZÁKLADY HACKINGU GOOGLOM **20**

2.1 ANONYMITA S GOOGLE ARCHÍVOM 21

2.2 VÝPISY ADRESÁROV A PÁTRANIE PO SÚBOROCH 22

2.3 INFORMÁCIE O SIEŤACH A SYSTÉMOCH 25

2.4 DATABÁZY A GOOGLE 28

2.5 PÁTRANIE PO UŽÍVATEĽSKÝCH MENÁCH, HESLÁCH 30

2.6 OSOBNÉ DÁTA A DÔVERNÉ INFORMÁCIE 33

2.7 SIEŤOVÉ ZARIADENIA 34

2.8 WEBOVÉ UTILITY NEPOCHÁDZAJÚCE Z GOOGLU 36

2.9 EXPLOITY A ICH VYUŽITIE 37

2.10 AJAX SEARCH API - PREKVAPIVÉ MOŽNOSTI GOOGLE SLUŽIEB 40

2.11 ZAUJÍMAVÉ DÁTA 41

3 OBRANA PROTI HACKEROM GOOGLU **43**

3.1 VÝPISY ADRESÁROV 43

3.2 BLOKOVANIE INDEXOVACÍCH ROBOTOV 44

3.3 META TAGY 46

3.4 ŠTANDARDNÉ NASTAVENIA 47

3.5 HĽADANIE BEZPEČNOSTÝCH RIZÍK – HACKNUTIE VLASTNÉHO WEBU 47

3.5.1 Automatizácia vyhľadávania a automatizačné nástroje 48

3.6 OKAMŽITÉ ODSTRÁNENIE Z GOOGLE INDEXU 50

3.7 FILTROVANIE VYHĽADÁVACÍCH FRÁZ 51

4 GOOGLE, HACKING A ČESKÉ PROSTREDIE **52**

5 GOOGLE HACKING A OSTATNÉ VYHĽADÁVAČE **55**

ZÁVER **56**

ZOZNAM POUŽITEJ LITERATÚRY **57**

Názov práce: Google hacking

Autor: Norbert Hanuska

Katedra (ústav): Katedra softwarového inženýrství

Vedúci bakalárskej práce: Doc. RNDr. Pavel Pyrih, CSc., Katedra matematické analýzy

e-mail vedúceho bakalárskej práce: Pavel.Pyrih@mff.cuni.cz

Abstrakt: Cieľom bakalárskej práce je podať prehľad o možnostiach využitia vyhľadávača Google ako hackovacieho nástroja a to ako teoreticky, tak aj demonštratívne pomocou konkrétnych príkladov. Rozoberajú sa možnosti pokročilého vyhľadávania pomocou vyhľadávača, jeho kladné, ale aj zneužiteľné stránky, vysvetľujú sa princípy tvorby efektívnych dotazov. Práca sa zameriava najmä na popis jednotlivých metód Google hackigu, teda získavania citlivých alebo inak zaujímavých dát, venuje sa ochrane proti takýmto typom útokov a na základe dostupných informácií sa snaží o zhodnotenie možných rizík.

Kľúčové slová: hacking, Google, vyhľadávanie

Title: Google hacking

Author: Norbert Hanuska

Department: Department of software engineering

Supervisor: Doc. RNDr. Pavel Pyrih, CSc., Department of mathematical analysis

Supervisor's e-mail adress: Pavel.Pyrih@mff.cuni.cz

Abstract: The aim of the bachelor thesis is to teoretically and practiacally with examples demonstrate possibilities of using Google search as a hacking tool. It analysis possibilities of advanced searching by the Google and it's pros and cons that can be used for malicious puropuses. Next, thesis also explains principles of building effective Google queries. Orientation is mainly to review methods of google hacking, to serve possible solutions of protection and it tries to evaluate and summarize possible risks based on the avilable information.

Key words: hacking, Google, searching

ÚVOD

Internet v poslednom období zaznamenáva obrovský rozmach, každý deň sa k nemu pripájajú milióny užívateľov aby si prečítali maily, oboznámili sa s dianím vo svete, alebo jednoducho našli zdroj relaxácie. Internet sa teda stal súčasťou, ďalšou dimenziou nášho života.

Google ako jeden z najpopulárnejších prostriedkov na vyhľadávanie informácii na Internete sa stal pojmom. Potrebujeme nájsť najbližšiu čistiareň? Chceme zistiť ako uvariť chutný obed? Môžeme na svoju web stránku uverejniť skutočne čokoľvek? Kedykoľvek nás napadne nejaká otázka, obrátime sa s ňou na Google. A keď vieme klásť tie správne otázky, je vysoká pravdepodobnosť, že sa k vytúženej odpovedi skutočne dopátrame.

Nie každý si však uvedomuje, akou silou Google oplýva a že sa dá zneužiť aj k dopátraniu sa k veľmi citlivým informáciám, akými sú mená, heslá, osobné údaje, konfiguračné súbory, alebo informácie týkajúce sa bezpečnosti webov a ich zraniteľnosti.

Vo svojej práci sa pokúšam upozorniť na toto nebezpečenstvo, podať prehľad účinných vyhľadávacích stratégií a samozrejme sa zamerať aj na možnosti obrany pred takýmto typom útokov, nazývaných aj Google hacking. Dobré schopnosti v Googli sú však výhodné aj mimo oblasť informačnej bezpečnosti, preto sa v práci venujem aj základom vyhľadávania pomocou Googlu.

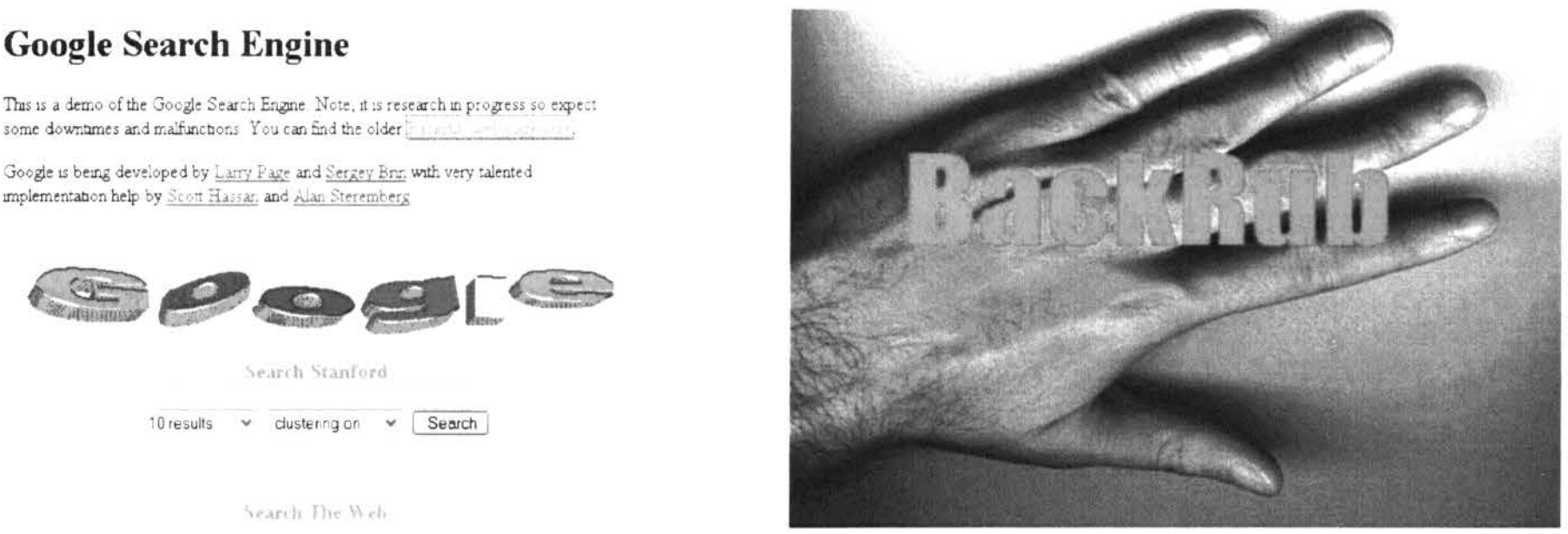
Hranica medzi Google hackingom a bežným vyhľadávaním nie je až taká pevná. V prípade Google hackingu je potrebné zadávať dotazy veľmi konkrétne a premyslene, aby sme maximalizovali efektivitu a relevanciu našich výsledkov, k čomu slúžia najmä pokročilé vyhľadávacie techniky ponúkané Googlom, a tiež znalosť chodu rôznych softwarových produktov. Informácie a dáta získané prostredníctvom Google hackingu môžu byť následne zneužívané a útočníkovi nápomocné v budúcnosti, či už sa jedná o rôzne exploity, štatistiky webových serverov, alebo osobné údaje využiteľné sociotechnikmi.

1 VYHĽADÁVANIE NA INTERNETE

Nikto asi nedokáže presne povedať, aké ohromné množstvo dát sa nachádza na Internete. Odhady počtu stránok sa pohybujú v miliardách až desiatkach miliárd, každý deň vznikajú milióny nových, ďalšie sa menia, iné mažú. Na Internete je teda neobvykle rušno. Aby v takomto extrémnom množstve dát nevládol zmätok a dokázali sme sa v ňom orientovať, musíme informácie nejakým spôsobom získavať, vyhľadávať. Nájdenie konkrétnej informácie v mori stránok zabezpečuje práve proces zvaný vyhľadávanie a stránky, ktoré ho sprostredkúvajú, sa nazývajú vyhľadávače [1].

1.1 História a súčasnosť Googlu

Za vznikom celosvetovo úspešného vyhľadávača s názvom Google stoja dvaja postgraduálny študenti, Larry Page a Sergey Brin. Google vznikal ako ich spoločné dielo pri príprave dizertačnej práce na prestížnej americkej univerzite Stanford. Larry Page dostal zdanlivo šialenú myšlienku, že stiahne do svojho počítača celý web. Svojmu konzultantovi tvrdil, že to zvládne v priebehu jedného týždňa. Za jeden rok sa mu toho naozaj kúsok stiahnuť podarilo. Spolu uviedli do prevádzky prvú verziu vyhľadávača, ešte pod názvom BackRub (obrázok 1) a prevádzkovali ho v rámci univerzitnej siete Stanfordu. Postupne sa ich projekt rozrastal a keď v roku 1997 hľadali nový názov pre svoj vyhľadávač, ujal sa „Google“ odvodený od slova „googol“. Slovo „googol“ bolo vymyslené Miltonom Sirottom, 9-ročným vnukom matematika Edwarda Kasnera a predstavuje číslo 10 umocnené na 100. Použitie tohoto slova reflektuje vôľu zakladateľov indexovať, prehľadávať a radiť pre ľudí nepreberné množstvo dát. Larry Page si k 15. septembru 1997 zaregistroval doménu *google.com* a neskôr, k 27. septembru 1998 aj firmu Google Inc., aby napokon 21.septembra 1999 bol Google oficiálne spustený [2].



Obr. 1. Stránky Google a BackRub z roku 1997

V dnešnej dobe je Google jednotkou medzi vyhľadávačmi s najväčším podielom na trhu (celosvetovo približne 62% všetkých vyhľadávaní [3]), konkurovať sa mu snažia hlavne Yahoo a Microsoft. Za svoj úspech vďačí najmä jednoduchosti, pokročilým možnostiam zadávania dotazov a prepracovaným systémom generovania výsledkov. Práve systém hodnotenia webových stránok, nazývaný aj „page rank", ktorý vracia odkazy na určité stránky častejšie a s vyššou prioritou než na tie ostatné, je jedno z najväčších obchodných tajomstiev Googlu. Pri tejto príležitosti si neodpustím jednu poznámku, a to že „page rank" je vlastne „Page rank", pretože ho zostavil jeden zo zakladateľov spoločnosti, Larry Page [2].

Veľké pozitívum spoločnosti Google je aj to, ako sa píše v knihe Google Story, že „na rozdiel od ostatných spoločností, kde si výkonní a výrobní riaditelia lámu hlavu nad tým, ako zarobiť čo najviac peňazí a až potom premýšľajú, aké výrobky či služby by asi tak mohli predávať, Google na to ide opačne a ako prví sa zamýšľajú technológovia; až potom – ak vôbec niekedy – sa uvažuje, ako výsledky speňažiť" [2].

1.2 Webové rozhranie Googlu

Webové rozhranie Googlu je pre užívateľa veľmi jednoduché a prehľadné, no napriek tomu neobyčajne silné. Jedna historka hovorí o tom, že keď bola stránka Googlu po prvýkrát spustená a začala sa načítavať, po načítaní jej obsahu sa prihliadajúci experti na ňu chvíľu iba dívali a čakali, kedy sa načítavanie skončí. Lenže ona už bola načítaná celá! To prinútilo zakomponovať na úvodnú stránku klasické záhlavie a zápätie s copyrightom, aby bolo jasné: „To je všetko, ráčte vyhľadávať!“. Dnes sa už nad hlavnou stránkou Googlu takmer nikto nepozastaví a nájdeme ju na adrese *www.google.com* (poprípade *www.google.cz*, stránky „českého“ Googlu – s českým jazykovým rozhraním).

Obr. 2. Hlavná stránka Google Česká republika

Aj keď na prvý pohľad skutočne vyzerá trochu chudobne, môžeme odtiaľto uskutočňovať rôznorodé vyhľadávania. Jeho základom je vstupné textové pole, do ktorého napíšeme to, čo hľadáme a stlačíme buď klávesu ENTER, alebo klikneme na tlačítko *Vyhledat Googlem* (*Google Search*), aby sme sa dostali na stránku s výsledkami nášho dotazu. Nad vstupným textovým poľom sa objavujú aj odkazy otvárajúce ďalšie sekcie vyhľadávania. Základná vyhľadávacia funkcionalita je rovnaká, ale v niektorých sekciách sa vyskytujú odlišné schopnosti, napríklad sa príjmajú iné vyhľadávacie operátory. Po spracovaní dotazu Googlom sa zobrazí stránka výsledkov, kde každý z nich má svoj nadpis (titulok stránky), ktorý zároveň slúži aj ako hypertexový odkaz. Výsledky sú zobrazené podľa relevantnosti, teda Google nám ponúkne na prvých miestach tie odkazy, o ktorých predpokladá, že sú pre nás primárne. Pri každej položke na stránke výsledkov vypíše Google názov webu, zhrnutie o webe na niekoľko málo riadkov, URL stránky obsahujúcu zhodu a jej veľkosť, odkaz *Archív (Cache)*, ktorý zobrazí stránku tak, ako vypadala keď ju Google prehľadal naposledy a odkaz na stránky s podobným obsahom.

Samozrejme Google má aj určité možnosti nastavenia, ku ktorým je možné pristúpiť z ktorejkoľvek vyhľadávacej stránky po kliknutí na odkaz *Nastavení (Preferences)*. Primárne sa voľby týkajú jazyka, v ktorom sa nám bude Google prihovárať a jazyka hľadania (štandardne Google prehľadáva stránky vytvorené v akomkoľvek jazyku), ďalej počtu zobrazených výsledkov na stránku a možnosť voľby otvárania odkazov zo stránky s výsledkami hľadania v novom okne prehliadača.

Za zmienku taktiež stojí odkaz *Pokročilé vyhledávání (Advanced Search)*. Z rozhrania stránky, ktorá sa nám zobrazí po kliknutí na odkaz je možné vykonávať pokročilé vyhľadávania s možnosťami, ktoré by nám boli za normálnych okolností prístupné iba prostredníctvom modifikácie URL adresy, alebo niektorých pokročilých operátorov.

1.3 Vyhľadávacie tipy, nebezpečné Google služby

Google odvádza skvelú prácu pri zavádzaní nových projektov, doplnkov a služieb. Ukazuje, že veci sa dajú robiť aj inak, umožnil rozkvet mnohých internetových projektov, vytvoril novátorské postupy, priniesol nový typ reklamy... Dobrú prácu odvádza aj čo sa týka užívateľskej prívetivosti a možností, čo všetko sa dá zistiť priamo z vyhľadávacieho rozhrania. Stačí do vyhľadávacieho poľa napísať `weather Prague`, `time London`, `5*3+(sqrt 10)` a vďaka chytrým nástrojom Googlu sa nám dostane konkrétnej odpovede, ktorú by sme intuitívne očakávali. Medzi takéto skutočne šikovné pomôcky patrí napríklad rozoznanie konverzie meny (`150 GBP in USD`), prevod jednotiek (`10.5 cm in inches`) alebo hľadanie určitej definície (`define Exploit`). Podobné zaujímavosti sú podrobnejšie rozpracované na oficiálnych stránkach Google Search Features (*http://www.google.com/help/features.html*). Možností, ktoré Google ponúka aj na tomto poli, je neúrekom.

Všetko má však svoje pre a proti a aj využívanie niektorých pokročilejších služieb Googlu so sebou prináša tienisté stránky. Študovanie správania svojich užívateľov prinieslo veľké množstvo vylepšení služieb, no zároveň umožňuje Googlu spracovávať a používať naše dáta. Objavili sa už aj rôzne konšpiračné teórie, že Google spolupracuje s americkou vládou a dodáva jej určité dôverné dáta a podobne. Faktom však ostáva, že Google o nás skutočne môže vedieť toho dosť, podrobný zoznam nájdeme na *http://blog.synopsi.com/2008-02-19/co-vsetko-o-vas-google-vie-doplnene-o-video.*
Námatkou vyberiem, že pri používaní Google AdSense môže Google vedieť naše celé

meno, adresu, bankový účet, IP adresu osoby, ktorá navštívi stránky s AdSense, jej polohu a podobne. Google Analytics zase odhaľuje zameranie našich stránok, z akej sme krajiny, mesta, akí užívatelia stránku navštevujú. Takto by sme mohli pokračovať, určitú hrozbu prináša napríklad aj používanie Google Books, Gmail, Picasa, Calendar... Google toho o nás môže vedieť naozaj dosť, treba len dúfať, že s týmito informáciami je nakladané profesionálne a dôverne a tak to ostane aj v budúcnosti.

1.4 Vytváranie dotazov

Ako som už spomínal, keď potrebujeme na Internete nájsť nejakú informáciu, väčšina užívateľov siahne práve po Googli. A keď vieme ako sa máme pýtať, Google je veľmi mocný nástroj. Je ľahké vyrobiť hľadanie neúčinné, ktoré nám neposkytne požadované výsledky, a preto vytvorenie správneho Google dotazu môžeme označiť za určitý proces . Samozrejme v našom záujme je čo najefektívnejšie využitie Googlu a chceme, aby výsledky dotazov boli čo najviac podobné našim predstavám. Preto sa využíva technika zvaná zužovanie (narrowing), čiže redukcia výsledkov hľadania. Základom efektívneho hľadania je zvládnutie syntaxu google dotazov, tým ďalším faktorom je určitá prax, skúsenosť a cit pre redukciu pri hľadaní.

1.4.1 Základné pravidlá pri vyhľadávaní pomocou Googlu

Predtým než sa zmienim o vyhľadávaní v Googli, pokúsim sa predstaviť niekoľko základných pravidiel [4].

V dotazoch Googlu sa nerozlišuje veľkosť písmen. Teda nie je podstatné či napíšeme dotaz spôsobom `gOogLE`, `GooGLe`, alebo `google`, vždy to bude pre Google rovnaké slovo a vráti nám rovnaké výsledky. Jedinou výnimkou je slovo *or*, ktoré keď chceme použiť ako Booleovský operátor, musíme napísať veľkými písmenami (OR).

Zástupné znaky Googlu (wildcardy) sú niečo iné než je väčšina programátorov bežne zvyknutá. Tí obvykle chápu zástupné znaky ako symbolickú reprezentáciu akéhokoľvek jediného znaku (napríklad v Unixe otáznik), alebo sériu znakov, ktorú predstavuje hviezdička. Kdežto v Googli je to inak. Zástupný znak hviezdička (*) reprezentuje v dotaze jedno, alebo niekoľko celých slov [5]. Napríklad hľadanie výrazu `kurzy * varenia` zobrazí výsledky pre frázy typu

```
"kurzy vegetariánskeho varenia"
```

```
"kurzy aranžovania kvetín, varenia"
```

a pod. Rovnako je možné použiť viacero hviezdičiek v jednom vyhľadávacom dotaze, napríklad `vitamin * je zdravy pre *`. Bohužiaľ vyhľadávanie zástupného znaku funguje iba pre celé slová alebo frázy, Google nepodporuje vyhľadávanie, v ktorom hviezdička označuje časť, alebo pokračovanie slova a teda pri dotaze typu

```
brit* awards
```

nedostaneme výsledok vyhľadávania

```
british awards
```

Google disponuje aj vlastnosťou automatického zkracovania slov (anglicky stemming). To znamená, že tam, kde je to vhodné, Google použije aj možný skrátený zápis slova, takže vyhľadáva aj podobné výrazy. Napríklad ak hľadáme výraz `pet lemur dietary needs`, Google hľadá aj `pet lemur diet needs` a podobné variácie výrazov. Nezabudnime však, že táto schopnosť môže mať za následok aj nepredvídateľné výsledky vyhľadávania (pri češtine a slovenčine väčšinou pozná iba skratky najznámejších slov - pri hľadaní `prodej ojetych automobilu` sa vyhľadá tiež `prodej ojetych aut`, ale aktuálne sa táto služba rozširuje a zavádza sa aj práca s gramatickými formami slov).

Google pri hľadaní ignoruje veľmi bežné anglické slová, znaky, číslice. Napríklad *who*, *where*, *what*, *the*, *a* alebo *an*. Zaujímavé však je, že logika pre vylúčenie určitého slova môže byť pri rôznych výrazoch rôzna. Napríklad ak hľadáme výraz `what the car in`, Google ignoruje slová *what*, *the* a *in*. Ak ich však hľadáme jednotlivo, Google ich prijme ako platné termíny a hľadanie termínu `the` vedie na vyše 14 miliárd výsledkov (jún 2008).

Pri vytváraní dotazu, ktorý nám vráti presne tie výsledky, aké požadujeme, začíname od určitého základu, ktorý potom rôzne modifikujeme. V tomto procese zohráva významnú úlohu aj technológia prideľovania poradia Googlu, pretože stránky, ktorým bolo pridelené najvyššie poradie, prídu na prvú stránku výsledkov.

Najjednoduchší dotaz sa skladá z jediného slova alebo kombinácie slov, napr. `google`, `google hacking`, `FBI hacker mitnick`. O niečo zložitejšie je hľadanie fráz. Fráza je slovné spojenie uzatvorené v úvodzovkách, ktorým hovoríme Googlu aby hľadal dané slovné spojenie so všetkými termínami a v presnom poradí, ako je to uvedené

v úvodzovkách. V týchto prípadoch Google nevylučuje bežné slová, ktoré inak ignoruje (and, the, a...). Príklady fráz: `"vacation hawaii"`, `"luxury hotels mauii"`.

Google ignoruje v dotazoch bodku, takže frázy môžeme tiež písať technikou [4]:

```
vacation.hawaii
```

1.5 Booleovské operátory

Frázy sú však stále pomerne elementárna forma vyhľadávania. Pre efektívnejšie vyhľadávanie a konkrétnejšie výsledky vracajúce sa z dotazov potrebujeme poznať a vedieť používať aj booleovské operátory AND, OR a NOT.

Najbežnejším je operátor AND. S jeho pomocou môžeme do dotazu začleniť viacero termínov. Napríklad výraz `johnny AND long` nám vráti výsledky, kde sa vyskytuje súčasne slovo johnny aj long. Treba však poznamenať, že pre vyhľadávací engin Googlu je operátor AND nadbytočný, pretože automaticky hľadá všetky termíny, ktoré sme zaradili do dotazu (až na spomínané výnimky). Google nás na túto skutočnosť aj patrične upozorní, ako môžeme pozorovať na obrázku 3.

Obr. 3. Upozornenie Googlu

Užitočný je symbol plus (+), ktorým si vynútime, aby sa do vyhľadávania zaradilo slovo, ktoré za ním nasleduje. Čiže ak chceme v danom výraze hľadať aj slová ako *what*, *the*, *in* napíšeme ich so znamienkom plus, za ktorým však nesmie byť medzera.

Napríklad:

```
+the car +in
```

Podobne by fungovalo aj spojenie s danými slovami uvedenými v úvodzovkách:

```
"the" car "in"
```

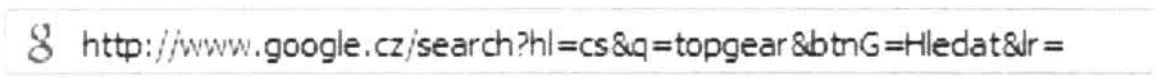
Ďalším bežne používaným operátorom je NOT, ktorý je funkčným opakom operátoru AND, to znamená vylučuje slovo z hľadania. Jeho ekvivalentom je znamienko mínus (-), ktoré uvedieme pred slovo a rovnako ako pri symbole + nesmie byť medzi ním a termínom

medzera. Pri zadaní dotazu `bill -gates` sa teda z výsledkov odstránia stránky, ktoré pojednávajú o Billovi Gatesovi.

Medzi bežné operátory patrí aj Booleovský operátor OR. Môžeme ho reprezentovať aj zvislicou (|) a jeho význam je vyhľadanie buď jedného termínu, alebo druhého. Pri jeho používaní by sme však mali byť opatrný, pretože v niektorých prípadoch môže byť jeho použitie mätúce. Je to z dôvodu rovnakej váhy operátorov OR, AND, ako aj pokročilých operátorov (o nich budem ešte hovoriť). Ak však budeme dotaz chápať ako vetu, ktorú čítame zľava doprava, nejasnostiam sa vyhneme. Na predídenie rôznym zmätkom môžeme použiť aj zátvorky, ktorým Google dobre rozumie a a vo väčšine prípadov je takto zapísaný výraz lepšie čitateľný.

1.6 Adresy URL Googlu

Každý dotaz Googlu sa dá reprezentovať pomocou URL, ktorá ukazuje na stránku výsledkov. Stránka výsledkov nie je statická, ale dynamicky sa meniaca v závislosti na tom, čo hľadáme. Odoslaním vyhľadávacieho dotazu prostredníctvom webového rozhrania sa teda dostaneme na stránku výsledkov, ktorá sa dá reprezentovať URL adresou. Predstavme si dotaz `topgear`. Akonáhle tento dotaz zadáme, Google nás presmeruje na adresu podobnú adrese na obrázku 4.



Obr. 4. URL adresa konkrétneho hľdania vygenerovaná Googlom

Ak by sme túto adresu zadali do poľa *Adresa* nášho prehliadača, Google znovu spracuje hľadanie `topgear` a zobrazí stránku výsledkov. Táto URL adresa sa dá veľmi ľahko modifikovať zmenou reťazca `topgear` na nejaký iný, a teda aj výsledky vyhľadávania pri jej opätovnom spracovaní. Pozrime sa na jej syntax podrobnejšie. Prvá časť URL *www.google.cz/search* je umiestnenie vyhľadávacieho skriptu Googlu [4]. Otáznik ďalej znamená, že sa skriptu budú predávať určité parametre, čiže voľby, aby Google skutočne niečo urobil. Parametre sa oddeľujú znakom ampersand (&) a sú zložené z premenných, za ktorými nasleduje znak rovná sa (=) a hodnota, na ktorú sa má premenná nastaviť. Premenná *q* (query) vyjadruje odoslaný dotaz, *hl* zase jazyk rozhrania. Pozor na to, že premenná *hl* sa ukladá, to znamená, že ak jej hodnotu zmeníme v URL, automaticky

sa použije aj pri ďalšom hľadaní až do doby, než to zmeníme prostredníctvom URL, stránky nastavení, poprípade otvoríme nove okno prehliadača. Ak potrebujeme pridať do vyhľadávania nejaké ďalšie parametre (premenné), jednoducho ich môžeme dopísať priamo do URL v ľubovoľnom poradí, to isté platí aj pre odstraňovanie parametrov. Základná syntax teda vyzerá nasledovne:

```
www.google.com/search?premenna1=hodnota&premenna2=hodnota
```

Najjednoduchši dotaz Googlu by však vzhľadom na to, že vo väčšine prípadov je jediným povinným parametrom parameter query (q), vyzeral asi takto:

www.google.com/search?q=topgear

Charakteristiky jednotlivých premenných som sa už zľahka dotkol. Popri premenných *q* a *hl* patrí medzi frekventovane používané aj premenná *lr,* ktorá obmedzuje výsledky vyhľadávania iba na stránky vytvorené v konkrétnom jazyku (*lr=lang_dk* vráti stránky iba v dánštine). Občas sa pletie s premennou *restrict*, aj keď tá nemá s jazykom veľa spoločného. Dáva možnosť obmedziť výsledky hľadania na jednu zem alebo niekoľko zemí určených názvom domény najvyššej úrovne (napríklad .cz) a/alebo zemepisným umiestnením IP adresy serveru [4]. Jej význam by sa mohol zdať tak trochu nepresný, no napriek tomu premenná funguje prekvapivo dobre. Do našej URL zaradíme premennú *restrict* a jej hodnotu, napríklad countryDK (alebo countryCZ) a Google nám vráti stránky, o ktorých si myslí, že sú fyzicky umiestnené v Dánsku. Tento predpoklad o polohe daného webu v určitom zemepisnom regióne si môžeme ľahko overiť nástrojmi *host* a *whois* [4]. Často veľmi nápomocný je aj parameter *as_qdr*, ktorý umožňuje odfiltrovať výsledky staršie ako určité časové obdobie. Napríklad keď do URL pridáme *as_qdr=m3*, výsledkom nášho dotazu budú iba stránky mladšie ako tri mesiace, ktoré sú pre nás v danom momente relevantné a aktuálne.

Zoznam väčšiny bežne používaných parametrov môžeme bez väčších problémov nájsť na Internete, napríklad na adrese *http://www.joostdevalk.nl/wp-content/uploads/2007/07/google-url-parameters. pdf*.

1.7 Pokročilé operátory

Okrem základných vyhľadávacích techník Google ponúka aj špeciálne prvky nazývané pokročilé operátory, ktoré môžeme chápať ako rozširujúcu syntax slúžiacu na zužovanie

výsledkov nášho hľadania. Bez použitia pokročilých operátorov v našich dotazoch Google hľadá uvedené termíny vo všetkých oblastiach webovej stránky, teda v titulku, texte, URL a podobne. Práve pokročilé operátory dovoľujú užívateľovi zamerať svoje hľadanie len na určité špecifické časti stránok a hľadať špecifický druh informácií.

Pokročilý operátor je vlastne iba časť dotazu, ktorý predávame Googlu bežne. Avšak ich syntax je dosť striktná a musí sa dodržiavať. Základná syntax je teda:

```
operátor:hľadaný_termín
```

Dôležité je poznamenať, že medzi operátorom, dvojbodkou a termínom nemôžu byť žiadne medzery. V opačnom prípade by sme dostali nežiadúce výsledky, pretože Google by pochopil syntakticky nesprávny pokročilý operátor ako ďalší hľadaný termín.

Hľadaný termín má rovnakú syntax ako termíny v bežne zadávanom dotaze, čo znamená, že spolu s pokročilým operátorom môžeme použiť jeden vyhľadávací termín, frázu tak isto ako aj Booleovské operátory. Musíme ale opäť dať pozor na medzery medzi operátorom, dvojbodkou a začiatočnými úvodzovkami, resp. pri Booleovských operátoroch aby sme ich neumiestnili tak, že by sme nimi oddelili dvojbodku. Príklad:

```
intitle:"Index of" -intetext:topgear
```

V jednom dotaze je možné kombinovať aj viacero pokročilých opeátorov, niektoré sa však dajú kombinovať lepšie než iné, niektoré dokonca vôbec. Taktiež nie všetky operátory sa dajú použiť kdekoľvek. Niektoré sa dajú použiť iba v sekcii Web, niektoré zase iba v sekcii Skupiny (Groups).

Príklady platných dotazov s pokročilými operátormi:

```
intitle:google
```

Vráti stránky, ktoré majú vo svojom titulku slovo Google.

```
intitle:"index of" "top gear".
```

Vráti nám stránky, ktoré majú v titulku frázu `index of` a okrem toho sa niekde (v titulku, v texte, v URL...) vyskytuje fráza `top gear`. Google interpretuje medzeru v dotaze ako koniec hľadaného termínu pre pokročilý operátor a pokračuje spracovaním ďalšej časti dotazu. Teda operátor *intitle* sa vzťahuje iba na frázu `index of` a už nie na `top gear`.

V nasledujúcom texte sa pokúsim podať akýsi sumár najčastejšie používaných pokročilých operátorov v technikách Google hackingu.

Intitle, allintitle

Prostredníctvom operátoru *intitle* obmedzujeme hľadanie na titulok stránky. Ten sa dá vyjadriť ako text umiestnený vo vnútri tagu TITLE v HTML dokumente.

```
intitle:"index of" "top gear"
```

vráti stránky s frázou `index of` v titulku a `top gear` niekde v stránke.

Allintitle hľadá stránky, kde sa každý zo zadaných termínov za operátorom objavuje v titulku stránky. Funguje v podstate ako použitie *intitle*: pred každým termínom. Kombinácia operátoru *allintitle* s inými pokročilými operátormi však nemusí byť úplne bezproblémová, preto by sme mali v čo najväčšej miere obmedziť jeho používanie.

Intext

Pátra po reťazci v texte stránky. Je vhodný, ak vieme že hľadaný text sa má nachádzať iba v texte stránky, poprípade ak niektoré výrazy sú príliš bežné v odkazoch alebo URL adresách.

```
intext:"yahoo.com"
```

Opäť existuje varianta *allintext*, ktorá však nie je veľmi vhodná na použitie s podobných dôvodov ako pri operátore *allintitle*.

Site

Operátor *site* umožňuje pátrať len po tých stránkach, ktoré hosťujú na konkrétnom serveri, alebo pochádzajú z nejakej konkrétnej domény najvyššej úrovne. Parametre pre operátor musia končiť platným názvom domény najvyššej úrovne, inak nedostaneme žiadne výsledky, ako napríklad pri hľadaní

```
site:apple.store
```

Korektné hľadanie by vyzeralo nasledovne:

```
site:apple.store.com
```

Ak teda máme predstavu aké informácie chceme hľadať a vieme, že sa nachádzajú napríklad na stránkach MFF Karlovej univerzity, jedna z možností je dotaz:

```
site:mff.cuni.cz "harmonogram akademickeho roku"
```

Operátor *site* sa dá bezproblémovo kombinovať s inými vyhľadávacími operátormi.

Inurl

Obmedzuje hľadanie na URL stránky. Operátor *inurl* je vo veľkej miere používaný, pretože umožňuje vyhľadávať podadresáre, čo nám operátor site zabezpečiť nedokázal.

```
site:apple.store.com inurl:webobjects
```

Inurl operátor dobre poslúži aj na nájdenie stránok vyhľadávania a nápovedy, keďže ich koncepcia je do istej miery obvyklá [6].

```
inurl:help
```

```
inurl:search
```

Filetype

Google dokáže hľadať aj oveľa viac než len webové stránky. Pomocou operátora *filetype* môžeme vyhľadávať súbory všelijakých typov, konkrétne teda operátor *filetype* hľadá stránky končiace na danú príponu súboru, ktorá je súčasťou URL. Hlavné typy súborov, ktoré Google hľadá, sú zhrnuté v tabuľke 1 [7].

To ale nie sú všetky typy súborov, ktoré Google dokáže vyhľadať. Vo svojej databáze má informácie o omnoho väčšom množstve prípon, prehľadá akýkoľvek typ stránky s akoukoľvek príponou, ale nemusí byť schopný vyhľadať neznámy typ súboru.

Link

Operátor *link* umožňuje vyhľadávať stránky, ktoré majú odkazy na iné stránky. V argumente sa nezadáva hľadaný termín, ale URL alebo názov serveru. V rámci techník zužovania vyhľadávania pamätajme na to, že dlhé URL sú konkrétnejšie, a preto môžu vracať menej výsledkov.

Numrange

Operátor *numrange* dokáže vyhľadávať čísla v zadanom rozsahu, vyžaduje na to dva parametre, dolnú a hornú hranicu intervalu oddelených pomlčkou. Ak hľadáme číslo 36, bude na to stačiť dotaz v tvare `numrange:35-37`

Existuje aj skrátená verzia operátoru, v ktorej stačí namiesto *numrange* zadať obidve čísla oddelené dvomi bodkami. Skrátená verzia hore uvedeného dotazu by potom mala tvar `35..37`. Operátor sa dá kombinovať aj s ostatnými operátormi a vyhľadávacími termínmi

a je veľmi silný a nebezpečný, ak si ho do svojho arzenálu zaradí hacker so zlými úmyslami [4].

Typ súboru	**Prípona súboru**
Adobe Portable Document Format	Pdf
Adobe PostScript	Ps
Lotus 1-2-3	wk1, wk2, wk3, wk4, wk5, wki, wks, wku
Lotus WordPro	Lwp
MacWrite	Mw
Microsoft Excel	Xls
Microsoft PowerPoint	Ppt
Microsoft Word	Doc
Microsoft Works	wks, wps, wdb
Microsoft Write	Wri
Rich Text Format	Rtf
Shockwave Flash	Swf
Text	ans, txt

Tabuľka 1. Hlavné typy súborov, ktoré Google vyhľadáva

Cache

Operátor *cache* slúži na zobrazenie archivovanej verzie stránky. Google ukladá časti stránok, ktoré predtým preliezol, takže k nim máme prístup na stránke výsledkov cez odkaz *Archív* (Podrobnejšie sa problematike Google archívu budeme venovať v kapitole 2). Ak by sme chceli rovno prejsť na archivovanú verziu nejakej stránky, stačí v dotaze Googlu použiť operátor *cache*. Napríklad:

```
cache:idnes.cz
```

Po prípadnom zadaní nekompletnej URL alebo názvu hostiteľa môže Google vrátiť nepredvídateľné výsledky. Pri zadaní neplatného názvu hostiteľa alebo URL sa odošle

dotaz ako hľadanie fráze. Operátor sa nemôže používať s inými operátormi, ani vyhľadávacími termínmi.

V skratke ešte spomeniem operátor *related*, ktorý nájde stránky podobné zadanej stránke a operátor *info*, ktorý zobrazí súhrnné informácie o webe a zobrazí odkazy na iné hľadania Googlu. Podrobný prehľad ďalších pokročilých operátorov je možné na Internete veľmi ľahko nájsť (napríklad na *http://www.googleguide.com/advanced_operators.html*).

2 ZÁKLADY HACKINGU GOOGLOM

Na to, aby bola obrana proti Google hackerom pátrajúcich po rôznych citlivých informáciách účinná, musíme poznať zbrane ich arzenálu. V tejto kapitole sa pokúsim priblížiť vybrané techniky a metódy Google hackingu, ich motiváciu a tiež spôsob hľadania a využitia určitého druhu informácií. Na začiatok uvádzam tabuľku s prehľadom jednotlivých techník (tabuľka 2), popisu a využitím najzaujímavejších z nich sa budem ďalej venovať.

Technika	Podrobnejšie informácie
Využitie Google archívu	anonymita
Výpisy adresárov a pátranie po súboroch	konfiguračné súbory systémové logy
Získavanie informácií o sieťach a systémoch – profil webového serveru	pomocou výpisov adresárov hlásenia o chybách prihlasovacie portály štandardné stránky sieťové zariadenia webové utility
Databázy a Google	hlásenia o chybách prihlasovacie portály
Pátranie po užívateľských menách, heslách, osobných dátach a dôverných informáciách	
Lokalizácia exploitov a hľadanie cieľov	
Google a automatizovaný zber informácií	automatizované získavanie dát zo zdroja parsovanie, obrusovanie a ďalšie spracovanie získaných dát
Google služby a ich možnosti	AJAX Search API Google Kalendár Google vlastný vyhľadávač
Získavanie „inak“ zaujímavých dát	video, hudba, software...

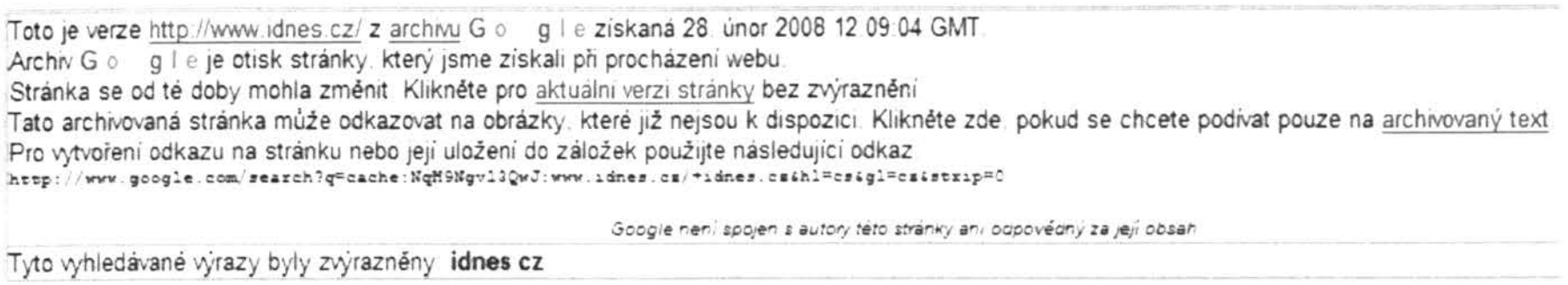
Tabuľka. 2 Techniky Google hackingu

2.1 Anonymita s Google archívom

Povaha aktivity hackerou zväčša vyžaduje určitý stupeň anonymity, preto sa neraz stáva ich prioritou číslo jedna. Obecne všetka komunikácia medzi našim systémom a vzdialeným serverom, či už sa jedná o prezeranie stránok alebo hľadanie citlivých dokumentov, môže byť zaznamenávaná. K účelu zabezpečenia anonymity pri prezeraní webovej stránky cieľa, na ktorý sa útočník zameriava, sa využívajú najmä proxy serveri, avšak do určitej miery k tomu môže dobre poslúžiť aj Google.

Google oplýva z tohto pohľadu veľmi užitočnou vlastnosťou – archívom (cache). Ako náhle prejde nejakú stránku alebo dokument, skoro vždy môžeme počítať s tým, že k dispozícii bude kópia tejto stránky alebo dokumentu. Samozrejme nevýhodou tohto prístupu je fakt, že sa ku kópii môže dostať hypotetický útočník, aj keď pôvodný zdroj už dávno neexistuje. Ďalšou tienistou stránkou archívu je to, že hackeri môžu prehľadať celý náš web bez toho, aby odoslali čo len jediný paket na náš webový server. Následkom toho je, že server nemôže nič zapísať do súborov protokolov (logov). Teda ak by boli odcudzené nejaké citlivé dáta, my o tom nemusíme ani len tušiť.

Googlom archivované stránky sú prístupné cez odkaz *Archív (cache)* na stránke výsledkov vyhľadávania, alebo ich môžeme vyhľadať pomocou pokročilého operátora *cache*. Čo sa vlastne deje za scénou pri prezeraní archivovanej verzie nejakej stránky? V prvom rade sa nám zobrazí akési záhlavie, úvodný text stránky Google pre archivovaný dokument, ktorý môžeme vidieť na obrázku 5. Tu si všimnime upozornenie, že archivovaná stránka sa môže odkazovať na obrázky, ktoré už možno nie sú dostupné. To okrem iného znamená, že ak si prezeráme nejakú archívnu kópiu webovej stránky, začneme vlastne sťahovať obrázky priamo zo samotného serveru, na ktorom je umiestená pôvodná stránka. Ak sme sa pokúšali o anonymitu tým, že prezeráme archívnu kópiu Googlu, nielenže sme ju nedosiahli, ale dokonca náš prehliadač informoval webový server o tom, že sa pokúšame prezerať archivovanú verziu stránky. Existuje však spôsob, ako načítať pôvodnú stránku bez toho, aby sme komunikovali s externým serverom a aby naša komunikácia prebiehala iba so serverom Googlu. Ak URL adresu, ktorá odkazuje na archivovanú verziu stránky mierne modifikujeme a na jej koniec pripojíme parameter *&strip=1*, zaistíme tým zobrazenie iba archivovaného textu bez všetkých externých odkazov. URL adresu s nastaveným parametrom *strip* na 1 môžeme vidieť aj po kliknutí na odkaz *„archivovaný text“* v úvodnom texte archivovanej verzie stránky.


Toto je verze http://www.idnes.cz/ z archivu G o g l e získaná 28. únor 2008 12:09:04 GMT.
Archiv G o g l e je otisk stránky, který jsme získali při procházení webu.
Stránka se od té doby mohla změnit. Klikněte pro aktuální verzi stránky bez zvýraznění
Tato archivovaná stránka může odkazovat na obrázky, které již nejsou k dispozici. Klikněte zde, pokud se chcete podívat pouze na archivovaný text
Pro vytvoření odkazu na stránku nebo její uložení do záložek použijte následující odkaz
http://www.google.com/search?q=cache:NqM9Ngv13QwJ:www.idnes.cz/+idnes.cz&hl=cs&gl=cz&strip=0

Google není spojen s autory této stránky ani odpovědný za její obsah

Tyto vyhledávané výrazy byly zvýrazněny: **idnes cz**

Obr. 5. Úvodný text archívnej verzie stránky

Ak by sme na prezeranie testovanej stránky použili anonymný proxy server, dostal by webový server iba IP adresu tohoto proxy servera a nie našu skutočnú IP adresu, pretože proxy server predstavuje akúsi medzistanicu pri putovaní dát, ktoré užívateľ odosiela a príjma. Pri pátraní po proxy serveroch vie byť Google opäť nápomocný. Dotazom

```
inurl:"nph-proxy.cgi" "Start browsing through this CGI-based proxy"
```

je možné naraziť na pár použiteľných CGI proxy serverov, ktoré fungujú na webovom serveri, navonok sa chovajú ako klasický HTTP proxy a dobre poslúžia ako čiastočná ochrana. Defaultná stránka CGI proxy býva zväčša uložená v súbore s názvom *nph-proxy* a obsahuje text *„Start browsing through this CGI-based proxy by entering a URL below"*. Tieto informácie viedli k zostaveniu vyššie uvedeného dotazu a zistenie podobných informácií býva zárodkom aj ďalších hľadaní podobného charakteru.

2.2 Výpisy adresárov a pátranie po súboroch

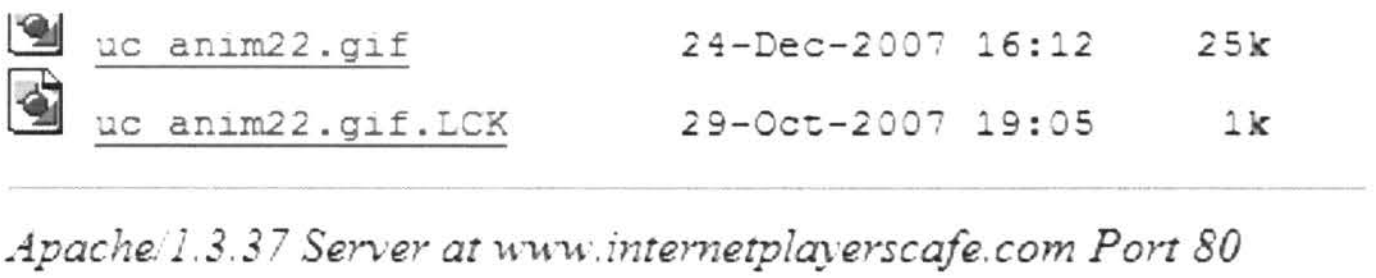
Výpis adresára je určitý druh stránky, ktorá obsahuje zoznam súborov a adresárov nachádzajúcich sa na danom servery. Na takýchto stránkach je možné prechádzať z adresára do adresára jednoduchým kliknutím na ich odkaz a tiež sťahovať a prezerať súbory v nich obsiahnuté.


Index of /images

Name	Last modified	Size	Description
Parent Directory	20-Feb-2008 13:31	-	
3heads.gif	24-Dec-2007 16:05	49k	
3oval.jpg	24-Dec-2007 16:05	89k	
3oval.jpg.LCK	04-Nov-2007 10:57	1k	
Bentley II.jpg	24-Dec-2007 16:05	57k	
Bentley.jpg	24-Dec-2007 16:05	16k	

Obr. 6. Ukážka stránky s výpisom adresára

Trpia však rôznymi nedostatkami. V prvom rade nie sú bezpečné. Nijak nebránia užívateľovi v sťahovaní súborov alebo v prístupe k iným adresárom. Typicky obsahujú titulok a niektoré pätičku (príklad vidíme na obrázku 7), v ktorej sa zobrazuje verzia webového servera, čo je pre útočníka cenná informácia ak sa chce dozvedieť nejaké detaily o web serveri pri vytváraní jeho profilu . Tiež sa často zobrazujú omylom, keď chýba alebo je neplatný úvodný súbor stránky (*index.html* a pod.)



Obr. 7. Pätička s verziou webového serveru

Vypátrať pomocou Googlu výpisy adresárov nie je vôbec zložité, väčšina takýchto stránok začína reťazcom „Index of“, ktorý sa nachádza aj v ich titulkoch. Teda na nájdenie takýchto stránok nám postačuje dotaz `intitle:"Index of"`, ktorý nájde všetky stránky s titulkom *Index of.* Tento dotaz však vráti aj veľké množstvo stránok, ktoré sú pre nás irelevantné a vôbec nevyzerajú ako stránky po ktorých pátrame. V tejto chvíli príde na rad zužovanie výsledkov a naskytne sa priestor na variácie nášho prvotného dotazu. Už dotazy ako

```
intitle:"index of" "parent directory"
intitle:"index of" +name +size
```

nám zabezpečia oveľa rozumnejšie výsledky.

Pri pátraní po konkrétnom adresári môžeme náš dotaz ľubovoľne modifikovať a skúšať, či vráti relevantné výsledky. Napríklad:

```
intitle:Index.of.admin
intitle:index.of inurl:admin
```

Vo výpisoch adresárov je tiež možné nájsť konkrétne súbory a to tak, že budeme hľadať „index of“ v titulku a názov súboru v texte webovej stránky, napríklad:

```
intitle:"index of" "bentley.jpg"
```

Výsledok tohto dotazu vidíme na obrázku 8.

Obr. 8. Výsledky špecifického vyhľadávania konkrétnych súborov

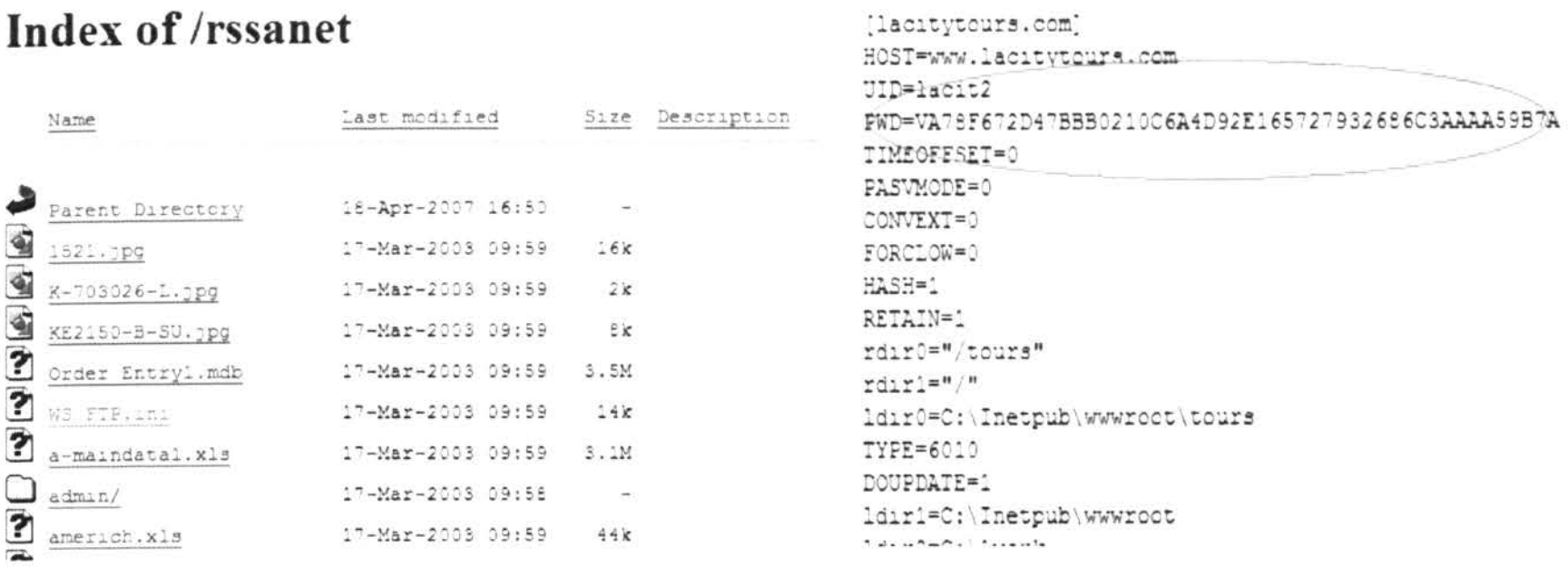
Pochopiteľne obrázky áut Bentley útočník hľadať asi nebude. Zoberme si však ako príklad program WS_FTP, relatívne známy a používaný FTP klient, ktorý tak ako väčšina užívateľských programov umožňuje zapamätať si heslá pre účty. Svoju konfiguráciu a informácie o užívateľských účtoch ukladá do súboru *ws_ftp.ini* a každý, kto získa prístup ku konfigurácii FTP klienta má prístup aj ku jeho zdrojom. Heslá ukladané v súbore *ws_ftp.ini* sú síce šifrované, ale ak má útočník k dispozícii konfiguračný súbor, môže použiť nástroje pre dešifrovanie hesla, alebo jednoducho nainštalovať program WS_FTP a spustiť ho s nájdenou konfiguráciou. A k veľkému množstvu konfiguračných súborov klienta WS_FTP sa môžeme dopátrať napríklad pomocou dotazov:

```
intitle:"index of" "parent directory" "ws_ftp.ini"
filetype:ini ws_ftp pwd
filetyp:ini inurl:ws_ftp.ini
```

Výsledky našej práce sú zobrazené na obrázku 9, ktorý som získal približne v marci 2008 a je na ňom zobrazený ako aj výpis adresára so súborom *ws_ftp.ini*, tak aj obsah tohto súboru so zvýrazneným zašifrovaným heslom.

Hľadanie konkrétneho súboru teda môže prebiehať rôznymi spôsobmi. Pátranie pomocou *filetype* a *inurl* je občas výhodnejšie aj preto, že daný súbor sa podarí vypátrať bez ohľadu na to, ako naň Google narazil a taktiež sa môže stať, že výpisy adresárov, ktoré

nájde `index.of` nepovolia prístup k súboru. Okrem toho výpisy adresárov nie sú zase až také bežné, ale dokážu poskytnúť priaznivé výsledky a dajú sa relatívne ľahko vypátrať.



Obr. 9. Konfiguračný súbor programu WS_FTP s vyznačeným zašifrovaným heslom

2.3 Informácie o sieťach a systémoch

Účinnému útoku na počítačový systém zväčša predchádza rozpoznanie cieľa, vytvorenie jeho akéhosi profilu. Tento proces zahŕňa okrem iného skenovanie počítačov, otestovanie fungujúcich služieb, typu operačného systému a verzie služieb programu. K tomuto účelu sa obvykle používa nejaký skener typu *Nmap*, no existuje ešte iná možnosť. Správcovia systémov občas inštalujú www servery, ktoré za behu generujú štatistiky o práci serveru, obsahujú zoznamy spustených procesov, alebo aj systémové záznamy (logy) [8]. Ak sa Googlu útočník spýta na napríklad na štatistiky programu *phpSystem* (alebo rôznych iných programov) dotazom

```
"Generated by phpSystem"
```

dostane sa mu do rúk množstvo podrobností o systéme. Ďalšie možnosti dotazov na štatistiky a informácie vytvárané populárnymi programami by mohli vyzerať podobne, príklady sú uvedené v tabuľke 3 [8].

Získanie informácií takéhoto druhu môže útočníkovi pomôcť s použitím príslušných nástrojov určených na prienik do systému alebo exploitov, preto by správcovia systémov pri použití programov umožňujúcich monitorovanie zdrojov počítačov mali dbať na to, aby prístup k nim bol chránený. V poslednej dobe sa veľa výskytov takýchto stránok obmedzilo, systémový administrátori sú obozretnejší a to je len dobre.

Dotaz	Typ informácií
`"Generated by phpSystem"`	typ a verzia operačného systému, hrdwarová konfigurácia, prihlásení užívatelia, otvorené spojenia...
`"This summary was generated by wwwstat"`	štatistiky práce WWW serveru
`"These statistics were produced by getstats"`	štatistiky práce WWW serveru
`intitle:"Statistics of" "advanced web statistics"`	štatistiky práce WWW serveru, informácie o návštevníkoch
`intitle:"Multimon UPS status page"`	štatistiky práce UPS zariadení
`intitle:"Apache::Status" (inurl:server-status | inurl:status.html | inurl:apache.html)`	verzia serveru, typ operačného systému, zoznam procesov

Tabuľka 3. Programy vytvárajúce štatistiky o fungovaní systému

Predstavme si ďalej situáciu, že v nejakom bežne používanom programe sa vyskytne bezpečnostná diera. Hypotetický útočník chce nájsť nejaké počítače s týmto programom, aby sa mohol pokúsiť na ne zaútočiť (napríklad má k dispozícii účinný exploit). K tomuto účelu môže okrem rôznych skenerov dobre poslúžiť aj Google. Keby sa spomínaná bezpečnostná diera týkala napríklad serveru Microsoft IIS verzie 5.0, výsledkom dotazu

```
"Microsoft-IIS/5.0 Server at" intitle:index.of
```

by boli odkazy na hľadané servery.

Príklady ďalších dotazov na iné verzie serverov demonštruje tabuľka 4 [8]. Všimnime si, že informácie takéhoto druhu väčšinou prezrádzajú stránky s výpisom adresárov, preto aj pátranie po nich využíva túto techniku.

Dôvodom úspešného hľadania je fakt, že niektoré servery pridávajú do určitých dynamicky generovaných stránok titulky obsahujúce svoje meno a verziu. Táto informácia sa sama o sebe nepovažuje za nebezpečnú, ale za istých okolností môže mať pre útočníka veľký význam a Google sa tak stáva vyhľadávačom potenciálnych obetí.

Dotaz	Server
"Apache/1.3.28 Server at" intitle:index.of	Apache 1.3.28
„Apache/2.0 Server at" intitle:index.of	Apache 2.0
"Apache/* Server at" intitle:index.of	ľubovoľná verzia Apache
"Microsoft-IIS/* Server at" intitle:index.of	ľubovoľná verzia Microsoft Internet Information Services
"HP Apache-based Web Server/*" intitle:index.of	ľubovoľná verzia serveru HP

Tabuľka 4. Dotazy na rôzne typy www serverov

Rovnako nebezpečné a účinné môže byť pátranie po stránkach hlásení o chybách webového serveru. Tie môžu poskytnúť nielen cenné informácie, ale taktiež je možné ich využiť k identifikácii softwaru použitého na serveri. Dobrým spôsobom, ako také chyby nájsť, je zistiť, aký druh hlásenia pri chybách konkrétny server generuje. Analýzou týchto chybových hlásení je možné sa dopátrať k určitým charakteristickým rysom, pomocou ktorých budeme schopný vytvárať unikátne vyhľadávacie frázy. Napríklad súbor generovaný serverom typu IIS zobrazujúci chybovú správu typu 400 (dokument nedostupný), môže byť lokalizovaný pomocou frázy

```
intitle:"The page cannot be found" „Internet Information Services"
```

Pri vytváraní profilu webového serveru určite útočníkovi dobre poslúžia aj „vstupné brány" k webovému serveru - prihlasovacie portály. Sú to stránky, ktoré pyšne zobrazujú dialóg pre užívateľské meno a heslo. Sú navrhnuté, aby umožnili prístup ku konkrétnym schopnostiam alebo funkciám webu až po tom, čo sa užívateľ prihlási. Okrem iného slúžia aj pri pátraní po odkazoch a dokumentácii, čo sa môže ukázať účinné pre útok. A navyše, ak má útočník po ruke účinný exploit proti konkrétnemu softwaru a tento software prevádzkuje prihlasovací portál, útočník je schopný pomocou vhodných dotazov Googlu vypátrať potenciálne vhodné ciele. Niekoľko dotazov pátrajúcich po prihlasovacích portáloch je zhrnutých v tabuľke 5.

Dotaz	Prihlasovací portál
`inurl:"login.asp"`	Užívateľ všeobecne
`inurl:"admin/login.asp"`	Administrátor všeobecne
`inurl:login.cfm`	ColdFusion
`intitle:novell intitle:webaccess "copyright * - * novell"`	Novell Groupwise
`inurl:"exchange/logon.asp"`	MS Outlook Web Access

Tabuľka 5. Dotazy pátrajúce po prihlasovacích portáloch

2.4 Databázy a Google

V poslednej dobe sa intenzívne sústreďuje pozornosť na bezpečnosť webových databázových aplikácií, najmä na software komunikácie s databázou. Dokladajú to aj stále častejšie diskusie o téme injektáže SQL, alebo nedávny masový útok na Microsoft SQL databázy (apríl 2008), ktorý zasiahol bezmála pol milióna webových serverov, čo indikuje, že zameranie útočníkov na databázy sa skutočne stupňuje. Útočník však obvykle nepoužije Google k tomu, aby sa vlámal do databázy alebo aby poškodil nejakú aplikáciu pracujúcu s databázou – hackeri Googlu skôr skladajú útržky informácií, ktoré unikli z potenciálne zraniteľných webov a použijú ich k výberu vhodného cieľa a neskoršiemu premyslenému útoku na daný cieľ.

O stránkach zvaných prihlasovacie portály som sa už v krátkosti zmienil v kapitole 2.3. Sú akousi „vstupnou bránou“ do nejakého webového serveru, resp. webovej aplikácie. Ich potenciál poskytovať cenné informácie je značný aj v prípade, že sa jedná o databázové prihlasovacie portály. Obvykle bývajú dobre zabezpečené, čo ešte viac priťahuje skúmavé pohľady webových útočníkov. Príklad databázového prihlasovacieho portálu môžeme vidieť na obrázku 10. Bez ohľadu na silu zabezpečenia, už len existencia prihlasovacieho portálu sprostredkováva útočníkovi letmý pohľad na typ software a hardware, na ktorý by sa dalo zamerať.

Obr. 10. Databázový prihlasovací portál

Cenné pre útočníka môžu byť aj hlásenia o chybách – dajú sa využiť ako pri tvorbe profilov všetkého druhu, tak aj vo fáze zhromažďovania informácií. Taktiež hrajú dôležitú úlohu pri detekcii a vytváraní profilov databázových systémov - je z nich možné získať dáta o systéme, konfigurácii a štruktúre databázy. Opäť uvádzam pre ilustráciu niekoľko dotazov, nájdeme ich v tabuľke 6 [9].

Dotaz	Popis
intitle:"Execution of this script not permitted"	„Cgiwrap“ chybové hlásenie odhaľuje administrátorský e-mail, porty..
intitle:"htsearch error" ht://Dig error	Môže odhaliť administrátorský e-mail, adresárovú štruktúru..
"Warning: mysql_connect(): Access denied for user: '*@*' "on line" -help –forum	Odhaľuje prihlásenia k databázi, ktoré boli kôli nejakej príčine odmietnuté
"access denied for user" "using password"	SQL chybová správa môže odhaliť názvy súborov, cesty, mená funkcií...

Tabuľka 6. Dotazy pátrajúce po chybových správach databáz

Jediný spôsobom ochrany systémov pred verejným informovaním o chybách je predovšetkým rýchle odstránenie chýb a ak existuje tá možnosť, tak aj nastavenie programu tak, aby boli informácie o chybách zapisované do pre tento účel špeciálne určených súborov a neposielané na stránky dostupné užívateľom. Je však nutné pamätať, že

aj keď budú chyby odstránené rýchlo (a tým pádom vlastne stránky zobrazované Googlom neaktuálne), útočník si bude môcť prezrieť kópiu stránky uchovávanú v Google archíve.

2.5 Pátranie po užívateľských menách, heslách

Mechanizmy pre overovanie totožnosti obvykle chránia informácie pomocou užívateľského mena a hesla. Na sieti je možné nájsť ako aj prístupové mená, ktoré sú často podhodnocované, lebo tvoria tú menej dôležitú polovicu pre väčšinu systémov overovania totožnosti, no dajú sa využiť v sociálnom inžinierstve, tak aj heslá k rôznym službám. Je to spôsobené najmä neznalosťou užívateľov umiestňujúcich tieto dôverné informácie na verejne prístupné miesta, ale tiež nedbalosťou vývojárov programov, ktorí buď nedostatočne chránia dáta užívateľov, alebo neinformujú o nutnosti zmeny štandardných nastavení svojich produktov.

Pri pátraní po užívateľských menách už dotaz ako `"your username is"` môže podať pomocnú ruku. V jeho výsledkoch by sa našli rôzne portály popisujúce proces vytárania užívateľských mien. Útočník si teda užívateľské meno zostaví pomocou týchto informácií a údajov, ktoré pozbieral z iných zdrojov.

Ďalším nepriamym spôsobom získavania užívateľských mien je pátranie po prihlasovacích portáloch, ktorých súčasťou bývajú odkazy na rôzne stránky užívateľskej podpory, napríklad pomocou dotazu:

```
login | logon
```

Prirodzene ciest pre nájdenie užívateľských mien pomocou Googlu je mnoho a jednou z ďalších môžu byť štatistky programov, ktoré kontrolujú aktivity daného webu. Rôzne druhy informácií o používaní webového serveru zobrazuje napríklad program Webalizer. Jeho výstupné súbory sa dajú vypátrať pomocou dotazov ako:

```
intext:webalizer intext:"total usernames" intext:"Usage Statistics for"
```

V niektorých prípadoch sú však zobrazované užívateľské mená už neplatné alebo neaktuálne, ale v tomto prípade je nápomocný stĺpec „Visits“, ktorý uvádza počet použití užívateľského účtu behom monitorovaného obdobia [4].

Užívateľské mená sa `však` môžu vyskytovať aj na bežných stránkach a k výsledkom sa dá dopátrať aj bádaním vo výsledkoch dotazu:

```
username | userid | employee.ID | "your username is"
```

Niekoľko ďalších príkladov funkčných dotazov je zhruntých v tabuľke 7.

Dotaz	Popis
`inurl:admin inurl:userlist`	súbory so zoznamom užívateľov všeobecne
`inurl:admin filetype:asp inurl:userlist`	súbory so zoznamom užívateľov všeobecne
`filetype:reg intext:"internet account manager"`	MS Internet Account Manager môže prezradiť užívateľské mená
`filetype:reg reg HKEY_CURRENT_USER username`	exportované registre Windows môžu obsahovať užívateľské mená a ďalšie informácie
`filetype:log username putty`	klientské protokoly PUTTY SSH - môžu prezradiť užívateľské mená a informácie o serveri

Tabuľka 7. Dotazy pátrajúce po užívateľských menách

Prístupové heslá sú jedným z najcitlivejších informácií vyskytujúcich sa na webe. Mali by sa dobre chrániť, no aj napriek tomu množstvo ľudí ukladá heslá do súborov a umiesťuje ich na zdroje dosiahnuteľné z Internetu. Dokonca mnoho z nich plní funkciu správcov sietí, vďaka čomu narastá ich nebezpečnosť.

Konkrétne pravidlá pre ich vyhľadávanie je však dosť ťažké naformulovať, no útočníci bývajú veľmi kreatívny. Dobré výsledky sa dajú dosiahnuť s kombináciou slov *account, users, admin, administrators, passwd, password* v spojení s typmi súborov .xls, .txt, .doc, .mdf, .pdf. Taktiež je dobré venovať pozornosť na adresáre obsahujúce v mene slová *admin, backup* a podobne, napríklad `inurl:admin intitle:index.of`

Takéto všeobecné techniky hľadania sú veľmi účinné, ak sú zamerané na nejaký konkrétny server a kombinujú sa teda ešte s operátorom *site*.

Väčšinou sa heslá na webe vyskytujú v zašifrovanej podobe, ako demonštruje obrázok 11, ktorý sa podarilo získať približne v apríli 2007, no ani to nie je dostatočný obranný mechanizmus. V mnohých prípadoch stačí tieto heslá vložiť do nejakého špeciálneho programu (napríklad John the Ripper alebo Cain and Abel), ktorý heslo dešifruje a zobrazí v podobe čistého textu. Takéto programy dokážu heslo zložené zo 4 znakov ASCII tabuľky prelomiť dokonca behom niekoľkých sekúnd [22]. Samozrejme, čas nevyhnutný

k prelomeniu hesla závisí na jeho dĺžke a rozmanitosti použitých znakov. Už zistenie hesla dĺžky 8 znakov by za použitia techniky brute force, založenej na testovaní všetkých možných reťazcov o zvolenej dĺžke nad zvolenou znakovou sadou, trvalo podstatne dlhšie, dokonca až niekoľko rokov (v závislosti na zvolenej znakovej sade).

Dotaz

```
filetype:pwd inurl:(service | authors | administrators)
inurl:_vti_pvt
```

kombinuje hľadanie pre niektoré z podporných súborov MS FrontPage.

-FrontPage- eracinelli:o5Q7L/O/aECw.
-FrontPage- eracinelli:o5Q7L/O/aECw.
/_vti_pvt/service pwd - 1k - Archiv - Podobné stránky

-FrontPage- contractor:9RPTxXVk.eTGM
-FrontPage- contractor:9RPTxXVk.eTGM.
org/_vti_pvt/service pwd - 1k - Archiv - Podobné stránky

-FrontPage- lakeregioncreations:iNGdly8lDlk76 .fake:invalid
Format souboru: nerozpoznáno - Zobrazit jako HTML
-FrontPage-. lakeregioncreations:iNGdly8lDlk76 .fake:invalid.
com/tmp/_vti_pvt/service pwd - Podobné stránky

-FrontPage- ekendall:bYld1Sr73NLKo louisa:5zm94d7cdDFiQ
-FrontPage- ekendall:bYld1Sr73NLKo louisa:5zm94d7cdDFiQ.
org/garderobe/_vti_pvt/service pwd - 1k - Archiv - Podobné stránky

Obr. 11. Zašifrované heslá získané približne v apríli 2008

V niektorých prípadoch sa dajú dokonca nájsť súbory so všetkými potrebnými informáciami: užívateľské mená, nezašifrované heslá, a tiež hostitelia, ktorý pomocou týchto údajov overujú totožnosť užívateľov. Túto skutočnosť dokladá obrázok 12.

name: = "michellepopp"; password: = "6126268"; URL: = "pass-3.htm ...
name: = "michellepopp"; password: = "6126268"; URL: = "pass-3.htm"; name: = "popp";
password: = "popp123popp"; URL: = "/popp/index.html"; END_FILE.
com/POPP/PASSWORD.LOG - 1k - Archiv - Podobné stránky

name: = "1234"; password: = "1234"; URL: = "demo.html"; name ...
Format souboru: nerozpoznáno - Zobrazit jako HTML
name: = "1234";. password: = "1234";. URL: = "demo.html";. name: = "1234";. password: =
"1234";. URL: = "done.html";. END_FILE.
.cz/test/pass/passmaster/password.log - Podobné stránky

Obr. 12. Užívateľské mená, nezašifrované heslá, názvy hostiteľov (apríl 2008)

Dá sa však očakávať, že hesiel nachádzajúcich sa v archíve Googlu nebude veľa, no ak natrafíme na nejaký výsledok, už len tento fakt prezrádza úroveň zabezpečenia serveru

a ten sa pochopiteľne môže stať terčom ďalšieho útoku. Ukážkové dotazy na dáta spojené s heslami sú uvedené v tabuľke 8 [4,8].

Dotaz	Výsledok
`„http://*:*@www" site`	heslá na stránky *site*, zapísané v podobe http://username:password@www...
`filetype:bak inurl:"htaccess | passwd | shadow | htusers"`	záložné kópie súborov, v ktorých sa môžu nachádzať informácie o menách užívateľov a heslách
`filetype:mdb inurl:"account | users | admin | administrators | passwd | passwords"`	súbory typu mdb, ktoré môžu obsahovať informácie o heslách
`intitle:"index of" pwd.db`	súbory pwd.db môžu obsahovať mená užívateľov a zašifrované heslá
`inurl:admin inurl:backup intitle:index.of`	adresáre obsahujúce v mene slová admin a backup
`"Index of/" "Parent Directory" "ws_ftp.ini" filetype:ini ws_ftp PWD`	konfiguračné súbory programu WS_FTP, ktoré môžu obsahovať heslá pre prístup k FTP serverom

Tabuľka 8. Heslá – ukážkové dotazy

Iba výnimočne sa dá dostať k požadovaným aktuálnym informáciám jednoducho a väčšina hodnotných nálezov vyžaduje trpezlivosť, kreativitu, inteligenciu a tiež trochu šťastia. To platí rovnako pre hľadanie mien a hesiel, tak aj pre ostatné dôverné informácie.

2.6 Osobné dáta a dôverné informácie

Bohužiaľ veľmi často sa stáva, že aj rôzne dôverné dokumenty sú umiestňované na verejne prístupné miesta, alebo posielané bez patričného zabezpečenia. Stačí, keď útočník istým spôsobom vypátra náš životopis posielaný napríklad pri hľadaní práce a okamžite zistí našu adresu, telefón, e-mail, referencie a podobne. Na Internet je skutočne možné životopisy nájsť v hojnom počte, napríklad dotaz

```
intitle:"curriculum vitae" "phone *" "address *" "e-mail"
```

vráti zaujímavé výsledky.

Keďže mnoho užívateľov Internetu vytvára rôzne elektronické adresáre, aj v nich sa dajú nájsť mená, čísla telefónov a e-mailové adresy. S použitím sociálneho inžinierstva je taktiež tieto informácie možné zneužiť, najmä ak sa týkajú ľudí v rámci nejakej uzatvorenej skupiny, napríklad firmy.

Ukážkové dotazy na dokumenty s vysokou pravdepodobnosťou výskytu osobných dát a dôverných informácií nájdeme v tabuľke 9.

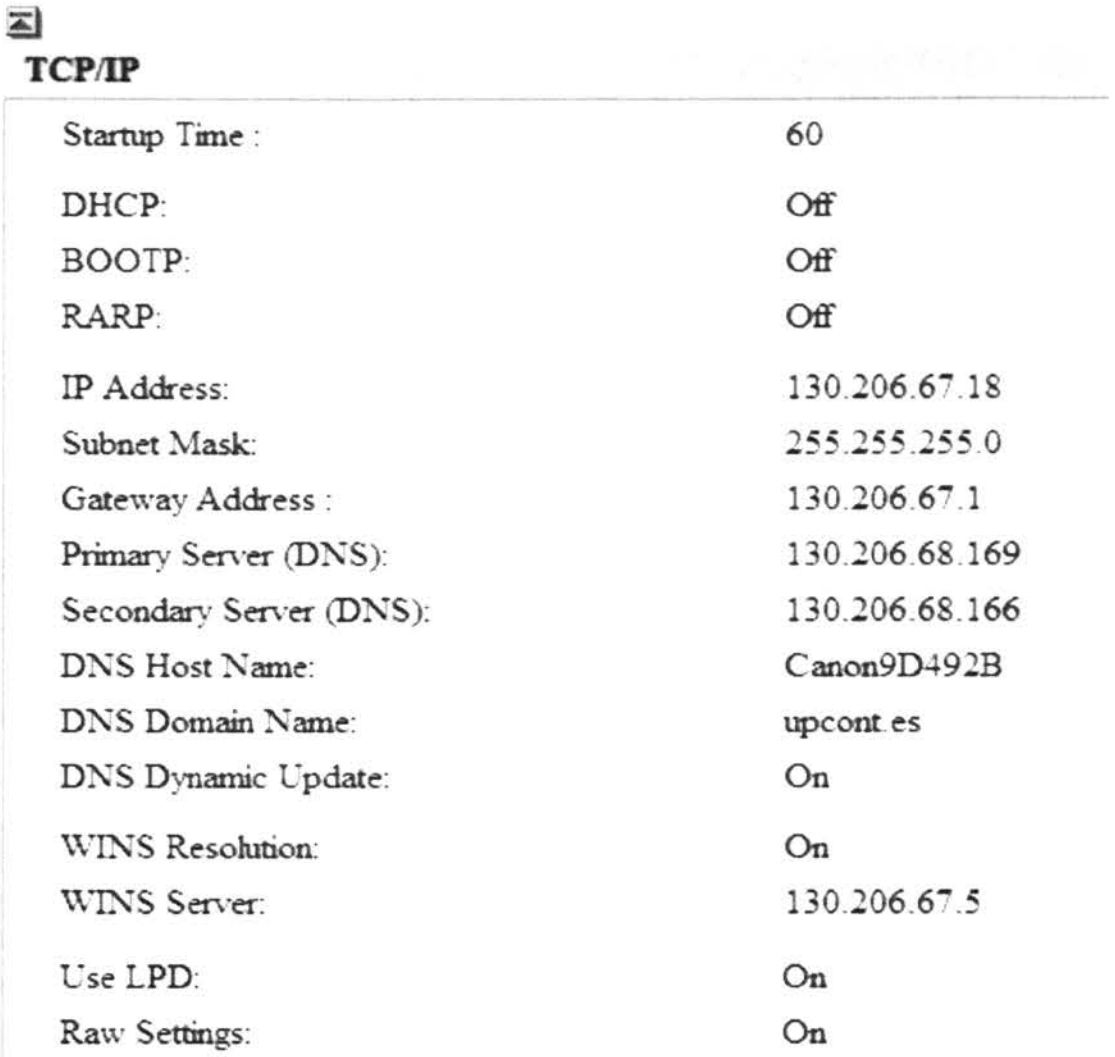
Dotaz	**Typ dôvernej informácie**
`filetype:xls inurl:email.xls`	súbory email.xls, ktoré môžu obsahovať elektronické dáta
`"not for distribution" confidental`	dokumenty so stupňom utajenia confidental
`filetype:ctt "msn"`	zoznam kontaktov MSN

Tabuľka 9. Dotazy pátrajúce po osobných dátach

Na zabránenie úniku osobných informácií rovnako ako v prípade hesiel môžeme jedine zachovávať ostražitosť a mať prehľad o zverejnených dátach. Firmy a inštitúcie by mali vytvoriť a dodržiavať príslušné procedúry a postupy popisujúce vnútorný obeh informácií.

2.7 Sieťové zariadenia

Nie je nijak neobvyklé, že v prostredí počítačových sietí existujú zariadenia so svojou vlastnou webovou stránkou určenou pre konfiguráciu a ovládanie daného zariadenia. Niektorí správcovia však bezpečnosť takých zariadení, ako sieťové tlačiarne alebo web kamery neberú vážne, a tým pádom vypátranie ich stránky predstavuje určité riziko, pretože často obsahujú informácie o cieľovej sieti, ako môžeme vidieť na obrázku 13. Tento druh informácií môže zohrávať dôležitú úlohu pri mapovaní siete, na ktorú sa útočník zameral.



TCP/IP

Startup Time :	60
DHCP:	Off
BOOTP:	Off
RARP:	Off
IP Address:	130.206.67.18
Subnet Mask:	255.255.255.0
Gateway Address :	130.206.67.1
Primary Server (DNS):	130.206.68.169
Secondary Server (DNS):	130.206.68.166
DNS Host Name:	Canon9D492B
DNS Domain Name:	upcont.es
DNS Dynamic Update:	On
WINS Resolution:	On
WINS Server:	130.206.67.5
Use LPD:	On
Raw Settings:	On

Obr. 13. Informácie o sieti zo stránky tlačiarne Canon ImageReady nájdenej Googlom v marci 2008

Na webovú kameru pripojenú na sieť sa vo väčšine prípadov nazerá skôr ako na zdroj zábavy, než na niečo ohrozujúce bezpečnosť. Spoločnosť Netscape bola kedysi známa tým, že poskytovala zákazníkom pohľad na sídlo firmy a jej okolie [8]. No nie je ťažké si predstaviť zneužitie dát získaných z web kamery napríklad na priemyselnú špionáž, plánovanie prepadnutia a podobe. Navyše webové kamery sú často umiestnené mimo objektu, čo je pre obhliadku terénu priamo ideálne.

Podobne sieťové tlačiarne sú zdrojom ohromného množstva informácií. Zobrazujú informácie o okolitej sieti a navyše veľa z takýchto zariadení sa nachádza v prednastavenej konfigurácii, ktorá umožňuje jednoduché zmocnenie sa takéhoto zariadenia, v horších prípadoch útočník môže byť schopný sledovať úlohy pre tlač alebo dokonca posielať vlastné sieťové informácie. Tabuľka 10 zhŕňa niekoľko dotazov, výsledkom ktorých sú stránky sieťových zariadení.

Dotaz	Zariadenie
`intitle:"Live View/ - AXIS"`	AXIS Video Live Camera
`inurl:"viewerframe?mode="`	Kamera Panasonic Network
`intitle:liveapplet inurl:LvAppl`	Sieťová kamera Canon
`intitle:"WJ-NT104 Main Page"`	Sieťová kamera Panasonic
`"powered by webcamXP" "Pro\Broadcast"`	WebcamXP
`intitle:"remote ui:top page"`	Tlačiareň Canon ImageReady
`inurl:sts_index.cgi`	Kopírky RICOH
`intitle:"Sipura.SPA.Configuration" -.pdf`	VoIP zariadenia

Tabuľka 10. Príklady dotazov pátrajúcich po sieťových zariadeniach

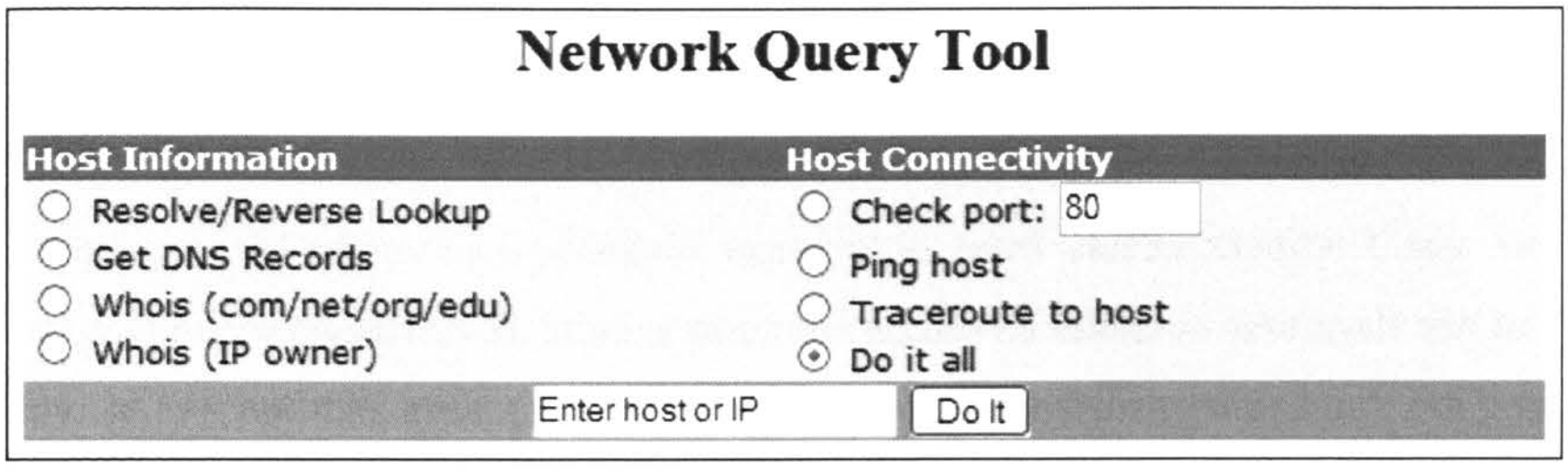
2.8 Webové utility nepochádzajúce z Googlu

Google je veľmi účinný a flexibilný nástroj, ale prirodzene nevie všetko. Pri procese mapovania siete a získavaní poznatkov o určitom cieli či potenciálnej obeti je často efektívne využiť aj iné prostriedky, mimo Google. Napríklad úlohy ako *ping*, *whois*, utility *traceroute*, skenovanie portov a podobne – na všetky tieto funkcie existuje mnoho nástrojov. Jedným z nich je aj nástroj zvaný Network Query Tool (NQT), zobrazený na obrázku 14. NQT je webová aplikácia, čo znamená, že akýkoľvek užívateľ, ktorý má prístup na túto stránku, môže využívať jej funkcie proti akémukoľvek cieľu. Štandardná inštalácia umožňuje vyhľadávať IP názvy hostiteľov, vydávať dotazy DNS, vykonávať dotazy *whois*, overovať aktivitu na konkrétnych portoch či zisťovať trasy paketov. Je to veľmi šikovný nástroj aj z toho dôvodu, že jeho funkcie pochádzajú z webu hosťujúceho aplikáciu NQT, čo znamená, že webový server maskuje skutočnú adresu užívateľa, do určitej miery je teda zabezpečená anonymita.

Pri pátraní po serveroch, ktoré prevádzkujú NQT, opäť nastupuje na scénu Google. Program NQT býva obvykle uložený v súbore s názvom *nqt.php* a vo svojej štandardnej konfigurácii zobrazuje titulok „Network Query Tool“. Takže jednoduchý dotaz

```
inurl:nqt.php intitle:"Network Query Tool"
```

nám vráti určité výsledky, v ktorých je možné sa dopátrať k fungujúcemu programu NQT.



Obr. 14. Program Network Query Tool

2.9 Exploity a ich využitie

Pri prenikaní, nabúravaní, jednoducho pri násilných vniknutiach do systémov a cieľov, na ktoré sa hacker zameriava, využíva rôznorodé nástroje. Do tejto výbavy neodmysliteľne patria časti kódu, ktorým sa anglicky hovorí expoity. Exploit (z rovnakého francúzskeho slova s významom „úspech"), je kúsok softwaru, dát, alebo postupnosť príkazov, využívajúca výhody chyby, slabosti alebo zraniteľnosti s cieľom spôsobiť neúmyselné alebo nečakané správanie softwaru, hardwaru alebo čohokoľvek elektronického [10]. Napomáhajú teda zaútočiť na konkrétny cieľ a vďaka tomu, že šírením podobného kódu sa zaoberá množstvo stránok, schopnosťami Googlu sa po nich pátra celkom ľahko.

Jedným so spôsobov, ako nájsť kód exploitov, je zamerať sa na príponu súboru zdrojového kódu a hľadať konkrétny obsah vo vnútri kódu. Veľké množstvo exploitov je napríklad napísaných v programovacom jazyku C, kde sa ako prípona so zdrojovým kódom obvykle používa .c. Dotaz `filetype:c exploit`, vráti okolo 10 000 výsledkov a väčšina z nich bude ukazovať presne na ten druh programov, ktoré hľadáme. Ak vezmeme do úvahy, že toto by mohli byť najviac navštevované weby hosťujúce zdrojový kód C a obsahujúce slovo exploit, tento zoznam môže poslúžiť ako základ pre tvorbu zoznamu obľúbených verejných webov s exploitami. Potom je už iba nás, akým spôsobom (v závislosti na operačnom systéme a schopnostiach) sa rozhodneme jednotlivé weby izolovať z vyhľadávacej stránky Googlu.

Iný spôsob, ktorým sa dá dopátrať ku kódu exploitu, je založený na znalosti bežne používaných reťazcov v samom zdrojovom kóde [9]. Napríklad, množstvo programov

jazyka C obsahuje štandardné knižnicové funkcie pre vstupno-výstupné operácie pridané vo vnútri zdrojového kódu pomocou príkazu *include*, ako napríklad *#include <stdio.h>*. Dotaz v tvare `"#include <stdio.h> exploit"` by vypátral zdrojový kód C obsahujúci slovo exploit, pričom sa však neobmedzuje iba na hľadanie dokumentov s príponou .c, ale ponúkne nám aj zachytený kód v HTML stránkach.

Účinným spôsobom na vypátranie exploitu je nová služba Google Code Search (*www.google.com/codesearch*), ktorá je primárne určená na hľadanie verejných zdrojových kódov. Je to v podstate akási prirodzená alternatíva k predošlým technikám, ktorá ponúka zaujímavé rozšírené možnosti. Predovšetkým je to povolenie dotazov s regulárnymi výrazmi a unikátne pokročilé operátori (*file*, *package*, *license* a pod.). Pre zaujímavosť uvádzam na obrázku 15 výsledky zobrazené po zadaní dotazu `#include <stdio.h>` do Code Search.

Obr. 15. Google Code Search

Ak sa pokúšame nájsť kód napísaný v jazyku C alebo C++, pomocou `lang:c` alebo `lang:c++` sa dopátrame k túženému výsledku. Aj keď by sa mohlo zdať, že je to veľmi podobné hľadaniu pomocou prípony súboru, nie je tomu tak. Služba Code Search totiž pracuje pokročilejšie, rozhodnutie, v akom progamovacom jazyku bol kód napísaný padne až po dôkladnej analýze zdrojového kódu (bez ohľadu na príponu súboru) [9].

Ako dokazujú výskumy mnohých bádateľov a bloggerov, Google Code Search môže byť taktiež využitý na lokalizáciu potenciálne zraniteľného softwaru. Pár príkladov uvádzam v tabuľke 11 [9].

Dotaz	**Popis zraniteľnosti**
lang:php (echo|print).*\$_(GET|POST| COOKIE|REQUEST)	XSS (cross-side scripting) zraniteľnosť
.*mysql_query\(.*\$_(GET|POST).*	Možná SQL injektáž
lang:php (system|popen|shell_exec|exec)\s*\(\$_(GET|POST|COOKIE| REQUEST).*\)	Umožnenie vzdialeného vykonania kódu
lang:php echo.*\$_SERVER\ ['PHP_SELF']	XSS zraniteľnosť
lang:php "WHERE username='$_"	SQL injektáž

Tabuľka 11. Dotazy pre Google Code Search pátrajúce po zraniteľnom kóde

Útočníci stále častejšie využívajú Google, aby vypátrali webové ciele, ktoré sú zraniteľné konkrétnymi exploitami, alebo sú náchylné na konkrétny typ útoku. Pátranie po cieľoch najčastejšie prebieha cez demonštračné stránky, alebo zdrojové kódy. Demonštračné stránky bývajú generované príslušnou aplikáciou automaticky a obsahujú špecifický text. Útok takéhoto druhu sa snaží využiť bezpečnostné problémy určitej verzie aplikácie a demonštračné stránky často tento druh informácie poskytujú. Konkrétnych príkladov hľadania zraniteľných webových aplikácií a stránok existuje množstvo, jedným z najväčších zdrojov takýchto dotazov je Google Hacking Database (GHDB), o ktorej ešte padne zmienka v kapitole 3.

Zaujímavým prípadom z tohto pohľadu je masívny útok na SQL databázy, ktorý sa prehnal Internetom koncom apríla 2008 a zasiahol viac ako pol milióna webov, medzi ktorými sa objavili aj weby s obrovskou návštevnosťou a štátne weby z USA a Veľkej Británie (dokonca bol zasiahnutý aj server OSN). Jednalo sa v podstate o bežný SQL Injection útok, teda slabinu útočníkovi umožňujúcu pracovať s databázou webového serveru, ktorá najčastejšie obsahuje podstatné dáta použité na stránkach, vrátane prihlasovacích údajov či publikovaných textov. Ako odrazový mostík pre útok bol použitý Google a práve jeden z dotazov, odhaľujúci chybu SQL databáze rovnakú pre niekoľko

stoviek tisíc webov. Vďaka získanému zoznamu nebol pre útočníkov problém vytvoriť robota, ktorý chybu otestoval na funkčnosť a zneužil [11].

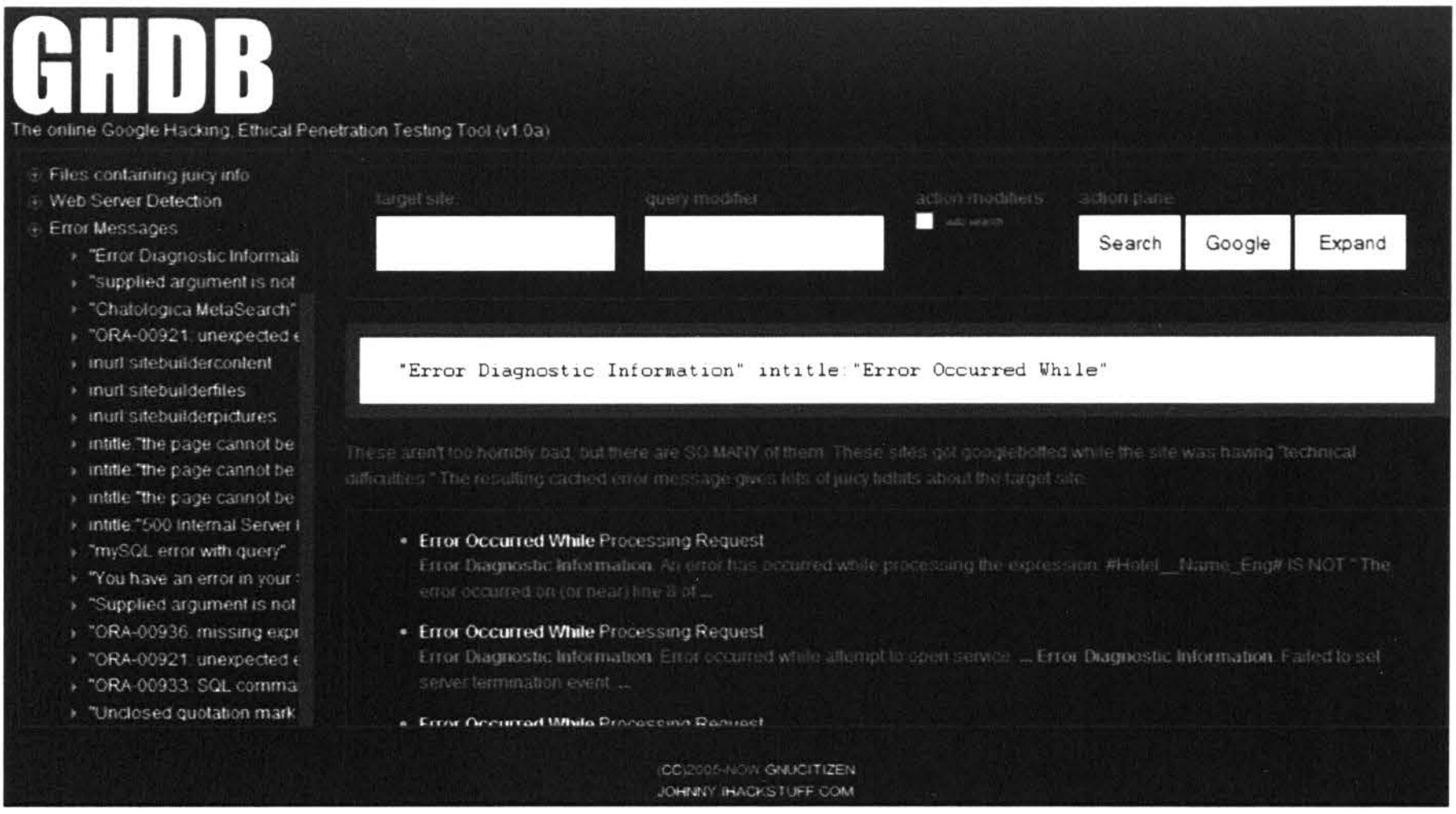
2.10 AJAX Search API - prekvapivé možnosti Google služieb

AJAX Search API je zaujímavá služba Googlu rozširujúca možnosti vyhľadávania. Jej zámerom bolo nahradiť staršiu vyhľadávaciu službu založenú na protokole SOAP, ktorej podpora bola pred určitým časom skončená. Primárny cieľ je umožniť externým stránkam hosťovať Googlom dodávané pomôcky buď v rámci, alebo aj mimo hosťujúcej stránky. Týka sa to vyhľadávania video klipov, máp, blogov, kníh a to všetko na jednom mieste, čo je veľmi užitočné, pretože získavame okamžitú spätnú väzbu cez celú Google platformu. Pre použitie AJAX Search API musíme získať API (Application Programming Interface – rozhranie pre programovanie aplikácií) kľúč, jeho generácia je možná na domovskej stránke *http://code.google.com/apis/ajaxsearch.* Nebudem zachádzať do podrobností, ale uvediem aspoň základné myšlienky, ako je možné využiť túto službu, tak trochu neštandardným spôsobom. S trochou znalostí JavaScriptu je jednoduché vytvoriť vlastné vyhľadávacie enginy pre kladné, ale aj škodlivé účely. Prvý krok je s pomocou prostriedkov na monitorovanie sieťovej prevádzky (ako je napr. rozšírenie LiveHTTP Headers prehliadača Mozilla Firefox) zistiť, že so službou Googlu sa komunikuje prostredníctvom URL adresy podobnej tejto:

```
www.google.com/uds/GwebSearch?callback=our_callback&context=0&rsz=lar
ge&q=GHDB&key=internal&v=1.0
```

Povšimnime si najmä parameter *callback*, čo je funkcia JavaScriptu a parameter *key*, ktorého hodnota bude vygenerovaný API kľúč. Modifikáciou tejto URL a jej zakomponovaním do skriptu môžeme jednoducho komunikovať a získavať požiadavky zo služby Googlu GwebSearch. Táto technika spolu s technikou scrapingu (získavanie, škrabanie dát z web stránky) je nosnou myšlienkou on-line nástroja Google Hacking Database (obrázok 16). Úspešnú implementáciu tohoto projektu nájdeme na adrese *http://www.gnucitizen.org/ghdb/.* Táto aplikácia v podstate dynamicky extrahuje všetky informácie z GHDB Johnnyho Longa a prezentuje ich v úhľadnej, grafickej forme. Je možné prehliadať každú kategóriu z databázy a vybraním dotazu získame okamžitú a „živú“ spätnú väzbu – zobrazenie výsledkov, ktoré sú zabezpečené pomocou rozhrania Google AJAX Search API. Toto všetko ešte v kombinácii s ďalšou službou Googlu,

Google Co-op (Vlastný vyhľadávač) otvára kreatívnym útočníkom nové možnosti pre vytváranie ďalších nebezpečných nástrojov.



Obr. 16. Online nástroj GHDB (GNUCITIZEN)

2.11 Zaujímavé dáta

Možnosti využitia Googlu sú rozsiahle, o čom sme sa mohli presvedčiť pri jeho využití ako prostriedku mapovania siete či zisťovania slabín možného terča útoku. Google však dokáže ísť ešte ďalej a poskytnúť zaujímavé informácie aj pre takpovediac bežných užívateľov informačných technológií.

Tvorbou správne upravených dotazov sa pre nás prostredníctvom Googlu môže stať voľne prístupná napríklad hudba, video, elektronické knihy, sériové čísla softwaru, aplikácie... Možností je nepreberne, dôležitá je najmä fantázia a kreativita užívateľa a navyše takýchto, treba však podotknúť vo väčšine prípadov nelegálnych kópií dát, na Internete neustále pribúda.

V našom pátraní po takomto type dát by sme sa samozrejme mohli obrátiť na peer-to-peer siete, prečo však neskúsiť využiť Google a jeho obrovskú databázu, ktorá nám napovie, kde takéto súbory hľadať. V našom pátraní môžeme skúsiť dotazy [12]:

```
-inurl:(htm|html|php) intitle:"index of" +"last modified" +"parent
directory" +description +size +(wma|mp3) "Interpret"
```

```
-inurl:(htm|html|php) intitle:"index of" (avi|mpg|mov|wmv) "Názov Filmu"
```

Zámenou termínu Interpret a Názov Filmu za požadovanú hľadanú hudbu alebo film docielime toho, že Google nám vo výsledkoch poskytne množstvo zaujímavých dát na stiahnutie. Možnosťou hľadania potenciálnych zdrojov filmov alebo hudby by boli aj dotazy [13]:

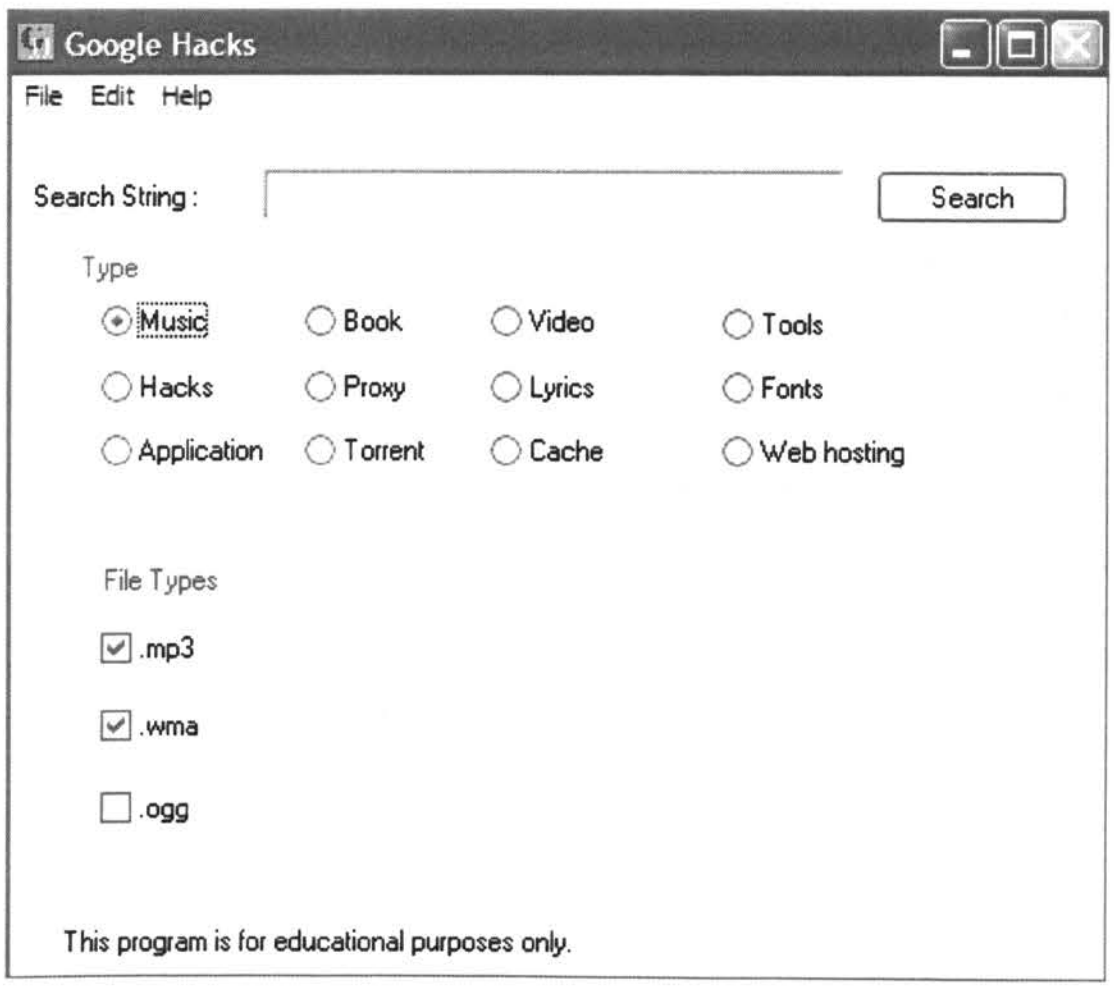
```
"parent directory" DVDRip -xxx -html -htm -php -opendivx -md5
"parent directory" MP3 -xxx -html -htm -php -opendivx -md5 -md5sums
```

Opäť môžeme zmeniť názov adresára a sledovať, k akým výsledkom sa dopátrame. Rovnako užitočný vie byť dotaz:

```
inurl:adobe filetype:iso
```

ktorý hľadá takzvané ISO obrazy zvyčajne obsahujúce inštalačné súbory k softwaru. Aj tu samozrejme platí, že argumenty operátorov *inurl* a *filetype* sú variabilné.

Na Internete sa objavujú aj malé šikovné programy, ktoré zjednodušujú prácu a pátranie po dátach takéhoto typu. Jedným z nich je program Google Hacks (obrázok 17), môžeme ho nájsť na adrese *http://code.google.com/p/googlehacks/* a mal by sa využívať najmä na vzdelávacie účely.



Obr. 17. Prostredie programu Google Hacks

3 OBRANA PROTI HACKEROM GOOGLU

Google hacking ponúka široké možnosti. Pochopenie základných techník a taktických manévrov Google hackerov je dôležité pre adekvátnu obranu a ochranu pred rôznymi „zlými chlapcami", ktorý sa snažia získať a zneužiť citlivé informácie. Google hacking nie je iba získavanie dôverných a citlivých informácií využitím možností, ktoré nám vyhľadávač Google ponúka, ale aj vniknutie do problematiky ochrany vlastných súborov a dát práve porozumením metód vedúcich k napadnutiu systému či prelomeniu bezpečnostného mechanizmu. Na tomto mieste sa posnažím podať konkrétne informácie ako zabrániť danému typu úniku informácií, ako napraviť už existujúci únik, jednoducho ako ochrániť náš web pred útokmi takéhoto druhu.

Trochu netechnickou, no nemenej dôležitou metódou z hľadiska bezpečnosti webu je účinná bezpečnostná politika. Je dôležité pochopiť, aká užitočná je prísna politika ohľadne publikovania rôznych dát na Internete, preto si myslím, že stojí za zmienku, aj keď sa jej podrobnejšie nebudeme venovať.

Údaje, ktoré by mali podliehať ochrane sa často na webe objavujú nevedomky, alebo z nepozornosti a vystavujú ich na dosah samotní užívatelia. Myslieť by sme mali hlavne na to, že webový server je primárne určený pre ukladanie dát určených širokej verejnosti. Ak je naozaj dôležité úplné súkromie nejakých informácií, mali by sa radšej presunúť na iné miesto, napríklad intranet alebo špecializovaný server, ktorého jedinou úlohou je poskytovať informácie bezpečným spôsobom [4].

Taktiež nie je veľmi dobrý nápad prideľovať verejnému webovému serveru odlišné role na základe rôznych prístupových úrovní. Dôvod je jednoduchý. Ak by sa nachádzali citlivé informácie na webovom serveri hneď vedľa verejných informácií, potom by pri napadnutí tohto serveru mohlo ľahko dôjsť k úniku citlivých informácií. Navyše netreba podceňovať ani ľudský faktor, pretože užívateľ môže ľahko skopírovať dáta z jedného súboru do iného, čo by mohlo mať za následok neúčinnosť nejakého obranného mechanizmu založeného na adresároch [4].

3.1 Výpisy adresárov

S rizikami výpisu adresárov sme sa už zoznámili. Užívateľom webu dovoľujú vidieť väčšinu súborov z nejakého adresára a často zobrazujú okrem súborov aj iné položky, ktoré

môžu byť zneužité. Z tohto dôvodu by mali byť výpisy adresárov vždy vypnuté, pokiaľ nie je naším cieľom umožniť prácu so súbormi v štýle FTP. Výpisy adresárov sa u niektorých webových serverov zobrazujú v prípade, ak chýba takzvaný indexovací súbor, najčastejšie pomenovaný *index.html*, *index.htm*, *default.asp* a podobne a je definovaný v konfigurácii daného serveru. Tento súbor by sa mal teda nachádzať v každom adresári, ktorý má užívateľovi zobraziť nejakú úvodnú stránku. Možnosťou vypnutia zobrazovania adresárov je aj vytvorenie súboru *.htaccess* v koreňovom adresári, v ktorom je potrebné pre tento účel nadefinovať pred slovo *Indexes* pomlčku, alebo znamienko mínus:

Options – Indexes

3.2 Blokovanie indexovacích robotov

Weboví roboti sú pomocníci internetových vyhľadávačov, ktorý automatizovane prechádzajú weby a indexujú nájdené informácie. Google na tento účel využíva robota s názvom Googlebot. Inštrukcie pre robotov sa nachádzajú v súbore *robots.txt*, ktorého štruktúra je štandardizovaná a roboti by mali jeho obsah, resp. obmedzenia v ňom obsiahnuté rešpektovať. Umožňuje definovať, ktoré súbory a adresáre majú byť mimo dosah týchto webových robotov, teda ktoré časti webu majú zostať skryté pred vyhľadávacími službami. Avšak je potrebné si uvedomiť, že niektorí webový roboti môžu ignorovať súbor *robots.txt*, ale väčšina tých s „dobrou povesťou“ tento súbor rešpektuje [14]. A ďalej, súbor *robots.txt* je verejne prístupný súbor a hocikto teda môže nahliadnuť do jeho obsahu. A skutočne, hackeri túto možnosť aj využívajú a je to jedným z ich trikov, aby získali predstavu o tom, ako sú na serveri rozdelené súbory a adresáre. Dotazom Googlu

```
inurl:robots.txt filetype:txt
```

sa môžeme presvedčiť, koľko existuje webov s prehľadaným súborom *robots.txt*.

Súbor *robots.txt* musí byť umiestnený v koreňovom adresári webového serveru a musí mať povolené také nastavenie prístupu, aby ho webový server mohol čítať. Je to textový súbor a každý jeho riadok okrem komentára musí začínať jedným z dvoch príkazov, *user-agent* alebo *disallow*. Pole *user-agent* obsahuje názov alebo typ robota, riadok disallow určuje, na čo sa robot nesmie dívať [14]. Pre robota Googlu by teda riadok *user-agent* vypadal takto:

```
User-agent: Googlebot
```

V poli *user-agent* sa dá použiť zástupný znak *, ktorým smerujeme pokyny všetkým robotom. Obsah súboru *robots.txt*, ktorý by zakazoval všetkým robotom indexovať všetky časti webu by vyzeral:

```
User-agent: *
Disallow: /
```

V prípade, že chceme povoliť nejakého konkrétneho robota, jednoducho mu nič nezakazujeme a príkaz *disallow* sa nastaví na „nič“, teda mu povoľujeme všetko (v zmysle indexovania adresárov). Príklad, kedy sa vylúčia všetci roboti okrem robota GoodBot:

```
User-Agent: *
Disallow: /
User-Agent: GoodBot
Disallow:
```

Štandardy súboru *robots.txt* zverejnené na stránke *www.robotstxt.org* deklarujú, že používanie regulárnych výrazov a zástupných znakov (až na výnimku „*“ v poli *User-Agent*) nie je podporované ani v riadku *user-agent*, ani *disallow*. Robot Googlu však podporuje používanie prípon. Vzor pre *disallow* môže používať hviezdičku v zmysle zástupného znaku pre ľubovoľný počet znakov. Okrem toho sa používa aj znak $ indikujúci koniec názvu. Teda ak by sme chceli Googlebotovi zabrániť, aby prechádzal naše PDF dokumenty a navyše aj adresár */tmp*, obsah súboru *robots.txt* by vyzeral nasledovne [14]:

```
User-Agent: Googlebot
Disallow: /tmp
Disallow: /*.pdf$
```

V prípade, že súbor *robots.txt* nevytvoríme, alebo bude prázdny, bude to znamenať, že sme poskytli všetkým robotom právo prechádzať a indexovať celú našu adresárovú štruktúru.

Ako náhle máme súbor *robots.txt* vytvorený a správne umiestnený, môžeme overiť jeho platnosť validátorom, ktorý nájdeme pomocou Googlu, alebo napríklad na stránkach *www.sxw.org.uk/computing/robots/check.html* [9].

3.3 META Tagy

V niektorých situáciách sa nám môže stať, že súbor robots.txt nemôžeme, alebo nechceme vytvárať. V tom prípade je obmedzenie robotov možné použitím nasledujúceho META tagu v hlavičke HTML dokumentu [4]:

```
<meta name="robots" content="noindex, nofollow">
```

Týmto špeciálnym META tagom hovoríme všetkým spolupracujúcim robotom, aby neidnexovali obsah stránky (hodnota `noindex` v položke *content*) a taktiež aby pri prehľadávaní stránky nesledovali odkazy vyskytujúce sa na stránke (`nofollow`).

Týmto prípadom však využitie špeciálnych META tagov ako metódy ochrany na našej stránke nemusí končiť. Môže sa nám vyskytnúť situácia, kedy chceme, aby vyhľadávacie roboty mali prístup a prehľadali stránku, avšak súčasne nechceme, aby Google vytváral archivovanú verziu stránky, teda aby sa vo výsledkoch vyhľadávania pri tejto stránky zobrazoval odkaz *Archív*. O bezpečnostných rizikách Google archívu už tiež bola reč. Zakázanie uloženia archivovanej verzie stránky teda prevedieme pomocou tagu:

```
<meta name="robots" content="noarchive">
```

Samozrejme aj tu platí, že túto našu reštrikciu budú dodržiavať iba spolupracujúci roboti.

Ak chceme zabrániť iba Googlu, aby daný dokument archivoval, uvedieme v sekcii HEAD dokumentu nasledovný riadok kódu.

```
<meta name="googlebot" content="noarchive">
```

Ak však už došlo k archivácii kópie určitej stránky, ktorú sme zabudli zabezpečiť, dá sa z Google archívu následne odstrániť.

A ešte jeden tip z oblasti META tagov. Vložením tagu

```
<meta name="googlebot" content="nosnippet">
```

opäť do hlavičky HTML dokumentu zabezpečíme, aby Google nezobrazoval fragmenty obsahu. Fragment obsahu je kúsok textu nachádzajúci sa na stránke výsledkov pod názvom nájdeného dokumentu a niekedy je našim cieľom zamedziť jeho zobrazovanie (napríklad v prípade autorizovaného prístupu k obsahu dokumentu). Zaujímavým vedľadjším efektom meta-značky `NOSNIPPET` je to, že Google nebude daný dokument ani archivovať [4].

3.4 Štandardné nastavenia

Už vieme, že aj s minimálnym vynaložením úsilia je možné sa dopracovať k nejakým štandardným stránkam, frázam, titulkom stránok, programom a dokumentáciám. Tieto súbory väčšinou obsahujú určité charakteristické rysy, podľa ktorých nie je ťažké vystopovať a zistiť, aký software je na konkrétnom servery použitý, aké je štandardné nastavenie programu, aký typ zabezpečenia je zvolený. Ak je útočník vybavený napríklad exploitom pre danú verziu softwaru, zvyšuje sa riziko úspechu jeho útoku, ako bolo už niekoľkokrát spomínané. Preto by sa mali odstraňovať spomínané položky z všetkého webového softwaru, ktorý sa na server inštaluje. Osvedčenou bezpečnostnou praktikou je odstránenie štandardných účtov a hesiel a tiež všetkých inštalačných skriptov alebo programov, ktoré boli dodané spoločne so softwarom.

Veľmi dôležitou zložkou bezpečnosti systémov je aj pravidelná aktualizácia a záplatovanie systémov, na ktorú by sa nemalo zabúdať.

3.5 Hľadanie bezpečnostných rizík – hacknutie vlastného webu

Pre získanie predstavy o potenciálnych bezpečnostných rizikách vlastného webu existuje jedna účinná metóda – pokúsiť sa naň podniknúť útok. Týmto spôsobom môžeme získať predstavu o slabých miestach nášho systému a podniknúť určité kroky k ich odstráneniu. Prirodzene jediný človek nemôže mať znalosti a schopnosti všetkých potenciálnych útočníkov, takže sa na tento spôsob ochrany nemôžeme spoľahnúť ako na jediný. V oblasti hackingu Googlom existujú nástroje pre automatizované skenovanie bezpečnostných rizík a slabín, ktoré nám sprostredkujú pohľad z perspektívy toho, ako samotný Google vidí náš web. Existujú však aj manuálne metódy, ktoré dobre poslúžia v prípade, že náš testovaný web nie je príliš rozľahlý. Pomocou operátoru *site* môžeme zistiť, čo všetko Google indexuje a dotaz `site:nasadomena.cz` by vypísal všetky Googlom archivované stránky zo serveru *nasadomena.cz*. Prostredníctvom týchto výsledkov si môžeme overiť, či skutočne všetky nájdené stránky sú určené pre verejnosť a neobsahujú citlivé informácie. Takéto ručné prehľadávanie je však namáhavé a časovo náročné, a preto je často vhodné použiť spomínaný proces automatizácie.

3.5.1 Automatizácia vyhľadávania a automatizačné nástroje

Existuje niekoľko spôsobov, ktorými sa dá hľadanie Googlom automatizovať, avšak Google oficiálne povoľuje automatizáciu dotazov iba prostredníctvom Google API (Application Programming Interface). Niektoré z ďalej spomínaných nástrojov spoliehajú na SOAP (Simple Object Access Protocol) API, ktoré však Google prestalo podporovať v dôsledku zavedenia AJAX API. Ak disponujeme starým SOAP API licenčným kľúčom, patríme medzi šťastlivcov, pretože tento kľúč stále funguje s nástrojmi, ktoré ho vyžadujú. Ak ho nemáme, stále je možnosť dohľadať ho pomocou Googlu, alebo zvážiť použitie programu Aura od spoločnosti SensePost (*www.sensepost.com/research/aura*) ako alternatívnu možnosť. Následne je potrebné upozorniť, že automatizačné nástroje, ktoré nevyžadujú zadanie licenčného kľúča môžu fungovať v rozpore s tým, ako Google definuje podmienky používania svojich služieb (Terms of Services). Viacej informácií o týchto podmienkach je dostupných na *www.google.com/accounts/tos* [9].

Google Hacking Database

Už sme sa oboznámili s princípom získavania citlivých údajov z databázy Googlu a tiež boli predstavené techniky tvorby dotazov, ktoré na tento účel slúžia. Dalo by sa povedať, že Google Hacking Database (GHDB) je ich najznámejším (aj keď v súčasnosti nie príliš aktuálnym) úložiskom a jej zakladateľ nie je nikto iný ako Johnny Long, uznávaný šepcialista na informačnú bezpečnosť, ktorý o bezpečnosti sietí a hackingu Googlom už aj prednášal na niekoľkých konferenciách o bezpečnosti počítačov po celom svete. Na pravidelnom rozširovaní databázy sa však môže podieľať ktorýkoľvek dobrovoľník a nadšenec, v súčasnej dobe zahŕňa takmer 1500 dotazov, ktoré ďalej separuje do viacerých kategórií, ako napríklad [15]:

- *Informačné správy a zraniteľnosť*
- *Chybové hlášky*
- *Súbory obsahujúce zaujímavé informácie*
- *Súbory obsahujúce heslá*
- *Stránky s prihlasovacím formulármi*

Tento depozitár hackerských techník realizovateľných pomocou Googlu, hosťovaný na *http://johnny.ihackstuff.com*, spomínam aj z dôvodu, že mnohé z automatizačných nástrojov ho vo veľkej miere využívajú a podporujú.

Charakteristika súčasných automatizačných nástrojov

Prvým z použiteľných nástrojov, o ktorom sa zmienim je **Gooscan**. Vytvoril ho Johnny Long, je založený na Linuxe, ponúka dávkové hľadanie Googlom a využíva výťažky z Google Hacking Database. Nebol vytvorený tak, aby spolupracoval s API Googlu, takže porušuje podmienky prevádzky služieb a je teda na rozhodnutí každého z nás, či chceme zámerne porušovať tieto podmienky pri zisťovaní prípadných únikov z nášho webu. Ak sa nakoniec rozhodneme použiť tento, alebo jemu podobný nástroj, ktorý podmienky prevádzky porušuje, Google môže zablokovať niektoré rozsahy IP adries a znemožniť používanie jeho vyhľadávacieho enginu [4].

Ďalšou aplikáciou je **Athena**, určená pre užívateľov operačného systému Windows. Rovnako ako Gooscan nie je založená na API Googlu a jej použitie je taktiež v rozpore s podmienkami používania služieb Googlu. Môžeme ju nájsť na adrese http://snakeoillabs.com/ a predpokladom na používanie Atheny je nutnosť mať nainštalovaný .NET Framework. Na vykonávanie základného vyhľadávania potrebujeme načítať nejaký XML súbor (môžeme stiahnuť alebo vytvoriť vlastný), štandardne sa dodáva aj súbor obsahujúci dotazy nájdené v GHDB.

Wikto od spoločnosti SensePost (www.sensepost.com) je úžasný webový skener s mnohými možnosťami. Nás prirodzene zaujíma predovšetkým aspekt Google dotazovania. Program funguje tak, že zadáme cieľ nášho skenovania a poprípade nejaké detailné informácie o serveri. Následne sa dostaneme do rozhrania, kde sa vyžaduje zadanie Google API kľúča. Táto sekcia je tak trochu chyták, lebo ako som už spomínal, Google prestal vydávať nové SOAP API licenčné kľúče. Ak ho však vlastníme, nič nám nebráni použiť ho, ak nie, možnosťou „obídenia" je použitie programu s názvom Aura od spoločnosti SensePost. Potom sa už ocitneme na hlavnej obrazovke, odkiaľ je možné realizovať samotné dotazy proti nášmu cieľu, teda testovanej stránke. Fáza skenovania sa opäť spolieha na GHDB, ktorej aktuálnu verziu do aplikácie načítame a tým umožníme otestovať našu stránku potenciálne škodlivými Google dotazmi. Wikto okrem toho ponúka

aj možnosť manuálneho zadávania dotazov a hľadania určitých typov súborov nachádzajúcich sa na cieľovom serveri.

Zaujímavým nástrojom na skenovanie reťazcov z GHDB je aj **Goolag Scanner**. Na svedomí ho má hackerská skupina Cult of The Death Cow, nájdeme ho na web adrese http://goolag.org a je to software, ktorý umožňuje jednoduché a ľahké použitie GHDB pre skenovanie vybranej web stránky. Práve Goolag Scanner bol inšpiráciou pre spomínaný on-line nástroj Gnucitizien GHDB využívajúci AJAX API. Každopádne je ďalšou aplikáciou, ktorej použitie je možné aj laikmi bez bližšej znalosti systému vyhľadávačov a bez akejkoľvek námahy [16].

Na záver tejto stručnej charakteristiky som si nechal dva relatívne nové nástroje, a to **Google Rower** a **Google Site Indexer**. Obidva sa dajú stiahnuť zo stránky *http://www.tankedgenius.com*. Google Rower funguje ako rozšírenie do prehliadača Mozilla Firefox (ale tiež aj ako samostatný program) a na rozšírenie vyhľadávania používa techniku „brute force". Google Site Indexer používa špeciálne operátory *site* a *inurl* na vytvorenie akejsi súborovej a adresárovej mapy cieľovej web stránky. Posielaním dotazov ako `site:tankedgenius.com` Google Site Indexer inkrementálne prejde všetky stránky, ktoré predtým zaindexoval Google [9].

3.6 Okamžité odstránenie z Google indexu

V prípade, že máme skontrolovaný náš web a zaznamenali sme potenciálne úniky informácií, nabáda sa otázka – čo robiť teraz?

V prvom rade je potrebné tento obsah z nášho webu čo najrýchlejšie odstrániť. Je vhodné aj zistenie zdroja úniku a zaistenie, že sa úniky v budúcnosti nebudú opakovať. Google prevádzkuje výbornú stránku pomáhajúcu so zodpovedaním niektorých najčastejšie kladených dotazov z hľadiska webmastera, umiestnenú na webovej adrese *www.google.com/webmasters* [9].

Týmto však náš problém nemusí byť úplne vyriešený. Google má archívnu kópiu nášho úniku informácií, ktorá len čaká na zneužitie. Možnosťou jej odstránenia je systém automatického odstránenia URL na adrese *http://www.google.com/webmasters/tools /removals*. Tento nástoj, ktorého stránku môžeme vidieť na obrázku 18, nás prevedie sériou otázok určených na overenie vlastníctva obsahu stránky a rozhodnutie, aký obsah sa

pokúšame odstrániť. Po zdarnom absolvovaní niektorej z možností, by sme sa ešte mali uistiť, že proces odstránenia URL bol naozaj úspešný a k nežiaducemu obsahu sa nie je možné dopátrať.


Google @gmail.com Můj účet Napověda

Nástroj žádosti o odstranění webové stránky

Moje žádosti o odstranění › Vytvořit novou žádost

Co byste si přáli odstranit?

Potřebujete z Google odstranit citlivé informace? Narazili jste v našich výsledcích vyhledávání na slepý odkaz? Chcete nám pomoci zlepšit filtr Bezpečného vyhledávání? Pro odeslání vaší žádosti vyberte jednu z níže uvedených možností.

- (•) Informace nebo obrázek, které se objevují ve výsledcích vyhledávání na Google
- ( ) Zastaralý nebo neaktivní odkaz ve výsledcích vyhledávání Google, který vrací chybu 404 (nenalezeno) - nenalezeno nebo chybu 410 (ukončeno)
- ( ) Na našich filtrovaných výsledcích Bezpečného vyhledávání se objevuje nepatřičná webová stránka nebo obrázek. Co je to Bezpečné vyhledávání?

Zrušit | Další »

Pokud jste webmaster a chtěli byste odstranit obsah svých stránek z našeho indexu, začněte s naším nástrojem pro odstraňování adres URL pro webmastery v nástrojích pro webmastery.

© 2008 Google Inc. - Ochrana osobních údajů

Obr. 18. Nástroj automatického odstránenia webovej stránky

3.7 Filtrovanie vyhľadávacích fráz

Prirodzene vývojári Googlu nespia na vavrínoch a aktívne sa snažia prispieť k zvýšeniu bezpečnosti Googlu a vyhľadávania. Implementácia filtrácie známych fráz, ktoré slúžia k vyhľadávaniu potenciálnych obetí napadnuteľných rôznymi internetovými vírusmi a červami je určite krokom dopredu. Táto metóda blokovania vyhľadávacích dotazov sa taktiež zameriava na rozšírené dotazy pátrajúce po citlivých informáciách, ako napríklad čísla kreditných kariet, ktoré by mohli vážnou mierou poškodiť mnohých užívateľov.

Tento detekčný systém je samozrejme chvályhodný počin, no treba upozorniť na to, že v niektorých prípadoch nie veľmi účinný. A to z jednoduchého dôvodu. Vyhľadávacie dotazy môžu byť veľmi ľahko modifikované, napríklad zmenením poradia kľúčových slov, veľkosti písmen a podobne, čo má za následok oklamanie a znefunkčnenie tohto systému, no na význam hľadania tieto modifikácie často veľký vplyv nemajú.

4 GOOGLE, HACKING A ČESKÉ PROSTREDIE

Na mnohých miestach v mojej práci som sa snažil poukázať na ohromný potenciál Googlu. Ten samozrejme nezostáva nepovšimnutý a aj preto nie je núdza o stále nové a nové spôsoby jeho využitia, resp. zneužitia. Aj v českom prostredí sme boli svedkami úspešných pokusov o využitie Googlu k „nekalým účelom".

Zapriahnutie potenciálu tohto vyhľadávača v niečí prospech (a v neprospech niekoho iného) sme v našich podmienkach mohli najviditeľnejšie badať približne v roku 2006 v podobe takzvaných Google bômb, ktoré „dopadli" na predných českých politikov a vládne inštitúcie. Čo to vlastne Google bomba je?

Google bomba označuje techniku umožňujúcu ovplyvniť internetový vyhľadávač, ktorý potom vracia nerelevantné výsledky [17]. Algoritmus vyhľadávačov je veľmi komplexná záležitosť, musí zohľadňovať radu faktorov a zbierať množstvo informácií o stránkach. Okrem obsahu webu sa obvykle zaznamenávajú aj údaje o odkazoch z iných stránok a obsah týchto odkazov potom hrá svoju úlohu pri zostavovaní výsledkov vyhľadávania. Ak sa teda objaví rada stránok, ktoré budú odkazovať na *www.opensource.cz* s textom odkazu napríklad „český server o open source", bude Google reagovať na hľadanie fráze `"český open source"` odkazom práve na tento server. Tento postup je samozrejme správny, vyhľadávač zohľadňuje aj informácie, ktoré o odkaze uvedú ďalšie zdroje, ale bohužiaľ pomerne dobre zneužiteľný a výroba Google bomby je z toho dôvodu do istej miery jednoduchá.

Aj preto sme sa mohli stretnúť s tým, že po zadaní vyhľadávacích fráz ako `"namyšlenej buran"`, `"senilní ješita"` alebo `"prasopes"` Google vrátil odkazy na stránky vrcholných predstaviteľov českej vlády. Autor jednoducho vytvoril na internetových stránkach pod konkrétnym označením odkaz vedúci na politikov web a keď bolo podobných odkazov viacej na ďalších stránkach, vyhľadávač vo výsledkoch prisúdil cieľovej stránke na dané slovo väčšiu relevanciu, dôveryhodnosť a následne vyššie umiestenie. Medzi ďalšie známe Google bomby v českom prostredí patria [18]:

`"ministerstvo brutality"`

`"Velký bratr"`

`"zločinci a vrazi"`

Google bomby sa začali vyskytovať v prostredí Internetu už od roku 2001 a neubránili sa im ani známe osobnosti v zahraničí [17]. Medzi dotknutých patrili napríklad aj prezident USA, alebo taliansky premiér [19]. Český predstavitelia vlády boli týmito útokmi postihnutí najmä v spomínanom roku 2006 a na útoky panovali rôzne postoje. Od rezolútneho nesúhlasu s žiadosťou, aby Google okamžite stiahol či zakázal podobné odkazy, až po názor, kedy sa incident bral ako žart a sloboda prejavu.

V súčasnosti už podobné „bomby" nevybuchujú, pretože Google sa k ním od roku 2007 postavil čelom a rozhodol sa ich riešiť. Aj keď ich nepovažuje za nijak vážny problém, pretože podobné žartíky sa podľa oficiálneho Google blogu vraj neobjavujú v masovom merítku a priemerný užívateľ na ne prakticky nenarazí, začali sa bomby obracať proti nemu. Stále viacej ľudí si myslelo, že sa jedná o názory Googlu, alebo že sú výsledky pre tieto dotazy ručne upravované, a preto Google prišiel s algoritmom, ktorý minimalizuje dopad Google bômb a detekuje ich v rôznych jazykoch [17].

Výrazy ako „prasopes" alebo „namyšlenej buran" už teda na bomby nereagujú a namiesto odkazov na politikov a iné oficiálne stránky tak dnes dostaneme relevantné odkazy predovšetkým na články o Google bombách. Avšak Internet skrýva netušené možnosti, a preto nie je vylúčené, že niekto príde s ďalším podobným princípom.

Český hacker

V súvislosti s českým prostredím mi nedá nespomenúť aj postoje českého hackera ako takého. V prvom rade však treba povedať, že termín hacking je v dnešnej dobe chápaný značne rozsiahle. Pod pojmom hackera si ľudia často predstavujú nebezpečného kriminálnika snažiaceho sa nabúrať do Pentagonu, kradnúceho a predávajúceho informácie za milióny dolárov a ohrozujúceho celosvetový mier. Hlavne v médiách je tento pojem často prekrúcaný a zamieňaný za označenia odlišných príslušníkov kyberpriestoru. Keby sme chceli vyvodiť nejaký všeobecný profil hackera, na základe dostupných informácií z literatúry a Internetu (napríklad z uznávaného Slovníka moderného hackera Jargon File [20]), mohli by sme ho zhrnúť: Jedná sa o veľmi dobre znalosťami vybaveného a kreatívne mysliaceho užívateľa, ktorý nachádza uspokojenie v objavovaní skrytých detailov v informačných a telekomunikačných systémoch, predovšetkým ich zabezpečení a zraniteľnosti. Hackera by sme teda mohli charakterizovať ako inteligentného človeka so zapálením v určitom obore. Má radosť z novo nahromadených skúseností a vedomostí. To,

čo robí, sa snaží robiť čo najlepšie a ak rieši nejaký problém, tak to skúša zo všetkých strán, kým ho nevyrieši.

Nemôžeme generalizovať, no českí hackeri sú tak trochu svojskí a valná väčšina z nich sa nesnaží nabúravať stránky inštitúcií a firiem, len aby zabili voľný čas. „Každý hack by mal mať zmysel a pokiaľ možno Ti niečo priniesť," tvrdí jeden z nich [21]. Teda tvor menom český profi hacker, aj keď to určite nie je celkom presné označenie, skutočne existuje a motiváciou jeho činnosti je veľmi často profit.

5 GOOGLE HACKING A OSTATNÉ VYHĽADÁVAČE

Okrem Googlu tvoria na celosvetovom trhu vyhľadávačov prim hlavne Yahoo, MSN a čínsky Baidu. Okrem nich existuje aj mnoho ďalších, menších vyhľadávačov. Prečo sa teda na poli vyhľadávania citlivých informácií ustálilo spojenie „Google hacking"? Je to preto, že použitie alternatívnych vyhľadávačov z hľadiska „hackingu" je dosť diskutabilné, až nevhodné.

V prvom rade chýbajúce aplikačné rozhranie potrebné pre automatizáciu testu slabín či napadnuteľných cieľov znemožňuje na tieto účely použitie akéhokoľvek iného vyhľadávača, než je práve MSN, Yahoo a Google, ktoré v súčasnej dobe ako jediné disponujú API rozhraním.

Aj keď je vyhľadávací algoritmus Googlu silný a účinný, čo sa týka indexácie počtu stránok, v tomto ohľade ho prekonáva jeho najväčší konkurent Yahoo. Avšak v prehláseniach o počte indexovaných stránok treba byť obozretný. V takýchto štatistikách sú často zahrnuté aj duplicitné výsledky, čo má v praxi za následok negatívny dopad na prehľadnosť a použiteľnosť získaných výsledkov. V unikátnosti výsledkov reálnych dotazov sa umiestňuje na prvom mieste Google, čím sa stáva pre množstvo používateľov tou správnou voľbou pre vyhľadávanie na internete.

Ďalším dôležitým kritériom pri posudzovaní kvality jednotlivých indexovacích systémov je práca s pokročilými operátormi. Tú Yahoo ponúka a dokonca sa pre mnohé pokročilé operátory ustálili rovnaké názvy ako v prípade Googlu, ale existujú určité odlišnosti, ktoré Yahoo v oblasti hackingu značne znevýhodňujú. Jedná sa hlavne o spôsob implementácie hľadania špeciálnych súborov. Pre hľadanie špeciálnych súborov pomocou Yahoo nemôžeme použiť vyhradený operátor, ale je potrebné modifikovať celú štruktúru dotazu, navyše Yahoo dokáže indexovať iba najbežnejšie formáty súborov, ako html, pdf, ppt, xls, doc, txt a xml. Teda pri použití Yahoo by sme pri niektorých dotazoch len ťažko hľadali adekvátne vyhľadávacie reťazce použité v Googli.

Predpokladom pre dosiahnutie čo najväčšieho počtu relevantných výsledkov je najmä použitie vyhľadávača s kvalitnými vyhľadávacími algoritmami a mohutnosť databázy indexovaných dokumentov. V kombinácii s možnosťou automatizácie vyhľadávania teda Google predstavuje ideálny prostriedok v oblasti vyhľadávania citlivých informácii a aj preto sa ustálil termín „Google hacking".

ZÁVER

Google hacking svoj najväčší rozmach zaznamenáva najmä od roku 2005 vďaka praotcovi Johnnymu Longovi, ktorý ako prvý upozornil na jeho potenciál a skryté nebezpečenstvo. Google hacking patrí medzi frekventovane rozoberané tematiky v oblasti bezpečnosti, dalo by sa povedať stáva sa fenoménom poslednej doby aj vďaka nenáročnosti – hackerom Googlu sa môže stať osoba s minimálnymi vstupnými znalosťami, no o to väčšie nebezpečenstvo predstavuje. Diskusie sa vedú aj na tému či by Google hacking mal byť vôbec označovaný za „hacking" v pravom slova zmysle, keďže vlastne pozostáva z dobre mienených dotazov a vyhľadávania pomocou Googlu, spojeného s určitou počítačovou gramotnosťou. Podľa môjho názoru však samotné vyhľadávanie predstavuje len akési pozadie, spôsob mapovania kyperbpriestoru, kúsky do skladačky, ktoré môžu byť zúročené pri útokoch ostatných typov.

Rozmach webu a jeho obsahu so sebou prirodzene prináša aj zväčšovanie databáze Googlu a rozsahu jeho služieb, s čím je spojený aj potenciál Google hackingu do budúcnosti, keďže s tým súvisí aj nárast množstva citlivých informácií získavaných jeho prostredníctvom. Tento trend sa pravdepodobne tak skoro nezmení a teda zraniteľných cieľov a spôsobov získavania citlivých informácii bude neustále pribúdať. Záleží len na šikovnosti a kreativite hackerov, ako si poradia so zadávaním dotazov a ako získané informácie zúročia.

Dúfam, že sa mi v mojej práci podarilo poukázať na hrozbu, ktorú Google hacking predstavuje. Primárnym faktorom umožňujúcim hacking pomocou Googlu je človek. Najmä jeho nedbalosť, alebo nevedomosť spôsobujú, že sa na internete vyskytujú dôverné informácie, ktoré by mali zostať skryté. Koncepcia získavania osobných informácií a ich následného využitia v oblasti sociálneho inžinierstva teda predstavuje podľa môjho názoru do budúcnosti hrozbu, ktorú je potrebné brať na zreteľ.

ZOZNAM POUŽITEJ LITERATÚRY

[1] ISKRA, J., Google : vyhledávání, Gmail, Google Talk a další služby, Computer Press, 2006, ISBN 80-251-1043-5.

[2] VISE, D. A., MALSEED, M., Google Story, Paradigma, 2007, ISBN 978-80-7349-034-8.

[3] Internetové stránky spoločnosti ComScore.

Zdroj: http://www.comscore.com/press/release.asp?press=2018

[4] LONG, J., Google hacking, Zoner Press, 2005, ISBN 80-86815-31-5.

[5] Internetové stránky spoločnosti Google – podpora.

Zdroj: http://www.google.com/support/bin/static.py?page=faq.html&hl=cs

[6] DORNFEST, R., BAUSCH, P., CALISHAN, T., Google Hacks: Tips & Tools for Finding and Using the World's Information, O'Reilly Media, 2006, ISBN-13 978-0596527068

[7] Internetové stránky Google Frequenty Asked Questions – File Types.

Zdroj: http://www.google.cz/help/faq_filetypes.html

[8] PIOTROWSKI, M., Nebezpečný Google – vyhledávání důvěrných informací, Časopis Hakin9, 2005.

Zdroj: www.goci.xf.cz/security/ hakin9_4_2005_google_cz.pdf

[9] LONG, J., Google hacking for penetration testers, volume 2, Syngress, 2008, ISBN-13 972-1-59749-176-1.

[10] Internetové stránky Wikipedie.

Zdroj: http://en.wikipedia.org/wiki/Exploit_%28computer_security%29

[11] Internetové stránky Živé.sk.

Zdroj: http://www.zive.sk/Titulna-strana/Internetom-sa-prehnal-masivny-utok-na-SQL-databazy/sc-21-sr-1-a-277145/default.aspx

[12] Internetové stránky Google Cheat Sheet.

Zdroj: http://www.adelaider.com/google-cheat-sheet/

[13] Internetové stránky i-Hacked.com

Zdroj: http://www.i-hacked.com/content/view/23/42/

[14] Internetové stránky The Web Robots.

Zdroj: http://www.robotstxt.org

[15] Internetové stránky Google Hacking Database.

Zdroj: http://johnny.ihackstuff.com/ghdb.php

[16] Internetové stránky Synopsi Blog

Zdroj: http://blog.synopsi.com/2008-02-23/goolagorg-predstavil-google-dork-scanner

[17] Internetové stránky Root.cz

Zdroj: http://www.root.cz/clanky/google-bomby-uz-nevybuchuji/

[18] Internetové stránky Jan Ambrož

Zdroj: http://www.ambroz.org/299502_clanek.php

[19] Internetové stránky iDnes.cz

Zdroj: http://zpravy.idnes.cz/domaci.asp?r=domaci&c=A060315_153840_domaci_ton

[20] Internetové stránky Jargon File

Zdroj: http://www.catb.org/jargon/html/H/hacker.html

[21] Internetové stránky Novinky.cz

Zdroj: http://www.novinky.cz/clanek/120035-jak-vydelava-cesky-hacker.html

[22] Internetové stránky Reboot.cz

Zdroj: http://reboot.cz/howto/hacking/hacker-zlomi-heslo-za-10-sec/articles.html?id=620