

از چه سرچ انجینی استفاده کنیم؟

حتماً شنیدین که افراد حوزه حریم شخصی بارها ابراز کردن که اگر از گوگل و امثالهم استفاده میکنید مهاجرت کنید به داک‌داک‌گو و امثالهم ، اما این گزاره چقدر میتونه در حریم شخصی و امنیت تأثیر گذار باشه؟ توی این مقاله کوشولو میخوایم به این موضوع بپردازیم.

الان سالها از اولین باری که من داک‌داک‌گو رو امتحان کردم میگذره، اولش یه تم سفید معمولی داشت ، و خیلی پاسخ های کوئریهاش مبتدیانه بود...
توی علم IR و RS به این قضیه میگن: Cold Start

نکته 1:

آی آر یا بازنمایی اطلاعات مخفف Information Retrieval هست و آر اس مخفف Recommender systems هست، اینا سرفصلهایی علمی هست که به این موارد میپردازه که زیر باک سرچ انجینها چخبره...

خلاصه داک‌داک‌گو اومد و خیلی ساده بود:

duckduckgo.com

به درد من نمیخورد...

بعدش چند سالی گذشت و دوستان عزیزش فیلترش کردن ، و این اواخر با مسموم کردن DNS سرور ها رکوئستهایشو بایپس کردن روی سیف-سرچ و رفع فیلترش کردن...

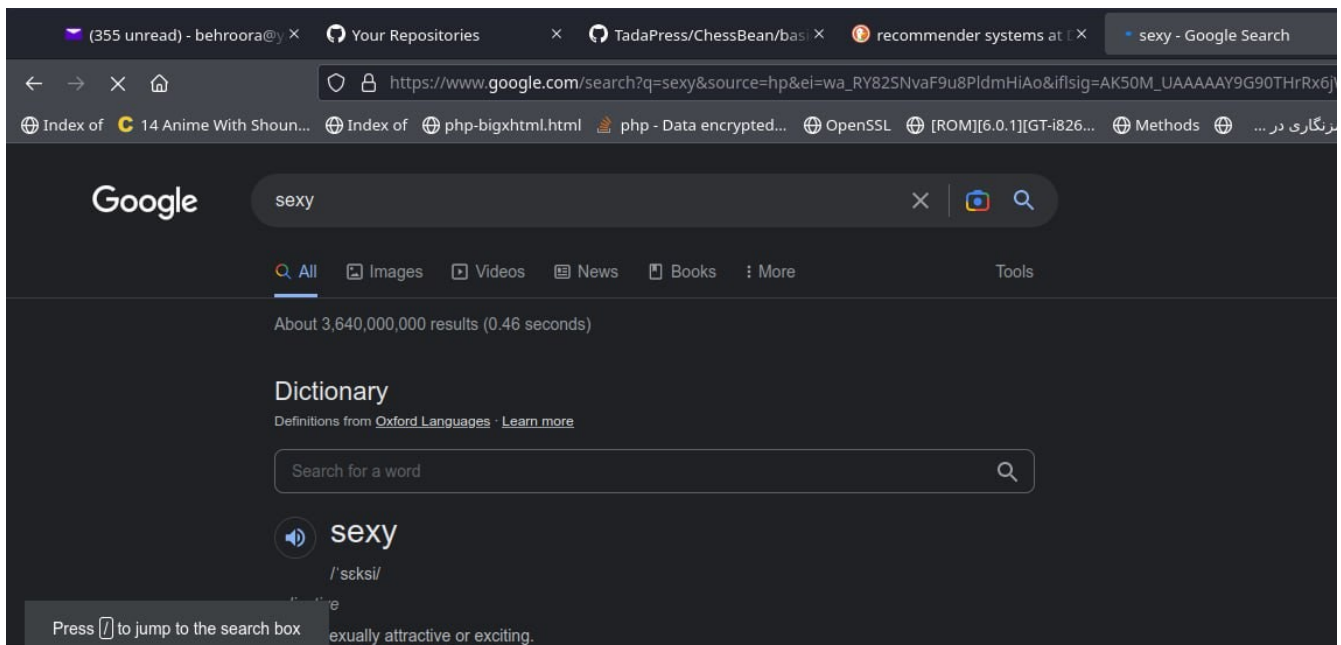
الان این روزها نتیجه های کوئریش خیلی خوب تر شده و از کلد استارت خارج شده...

+++++

داک‌داک‌گو ادعا می‌کنه که امنه ، اما این ادعا چقدر درسته؟
برای یکی مثل ما که علمش رو داریم قابل سنجشه

من یه کوئری توی گوگل میزنم و باهم بررسی میکنیم که گوگل چه اطلاعاتی از من رو توی هدرهای HTTP انکپسوله میکنه و میفرسته به سمت سرورهای خودش، بعداً همین کوئری رو توی ddg تکرار میکنیم و تست میکنیم که اونجا چه اتفاقی می‌افته.

توی گوگل:



چیزی که ما میبینیم یه سرچ ساده هست...، ولی بررسی بیشتر پیچیده‌ش میکنه:

1- اطلاعاتی که توی هدر URL به هیبت GET انکپسوله شده:

مشکله از:

```
https://www.google.com/search?
q=sexy&source=hp&ei=wa_RY82SNvaF9u8PldmHiAo&flsig=AK50M_UAAAAAY9G90T
HrRx6jWH00jz3MWt69cLMM37MN&ved=0ahUKEwiN9-
CA5eP8AhX2gv0HHZXsAaEQ4dUDCAc&uact=5&oq=sexy&gs_lcp=Cgdnd3Mtd2l6EAMyB
QgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEMgUILhCABDIFCC4QgAQyBQguEIAEMgUIAB
CABDIFCAAQgAQyBQgAEIAE0gsILhCABBDHARDRAzoICC4QgAQQ1AJQAFjuC2DTlwFoAXA
AeACAAcACiAHdCJIBBTItMi4ymAEAoAEB&scclient=gws-wiz
```

1- پارامتر کوئری :

q=sexy

2- پارامتر سورس:

source=hp

جایی هست که کوئری رو نوشتید، اینجا ایچ پی مخفف home page هست... داره به سرورش می‌گه که این بشر سایتمونو باز کرده توش سرچ کرده... یه راست از مرورگر یا پلاگین یا نیومده.

3- پارامتر ei :

ei=wa_RY82SNvaF9u8PldmHiAo

حالا کم کم داره پارامترهای مشکوک نمایان میشه :) این ie تایم استمپ دقیق _____ ق زمانی

هست که شما سرچ کردی... در هیبت انکد شده! حتی طبق افسانه ها توش میلی ثانیه هم نوشته شده!

من دیکدش میکنم، و لینک مقاله دیکد کردن + سورسکدش رو آخر میفرستم:

```
var_dump(
    time(),
    ei_decode("wa_RY82SNvaF9u8PldmHiAo")
);
```

```
int(1674687414)  
array(4) {  
[0]=>  
int(1670492097)  
[1]=>  
int(887117)  
[2]=>  
int(4261249782)  
[3]=>  
int(2701257877)  
}
```

همونطور که میبینید در قیاس با تایم استمپ فعلی کاملاً منطقیه...

و اما پارامتر های دیگه چیستن؟! میلی ثانیه؟ تاخیر؟ سرعت تایپ انگشتان من؟! به راستی چی هستن؟؟؟ از گوگل بپرسید نه از من... بخواید که یه داکيومنت کلریفای شده از همه اینها بده بیرون...

4- پارمتر iflsig

iflsig=AK50M_UAAAAAY9G90THrRx6jWH00jz3MWt69cLMM37MN

این چیست؟

راستش من حوصله تحلیل تمام حرکات گوگل خبیث رو ندارم... و هرچی هم جلوتر میریم داره وضعیت خطری تر میشه، ولی خب از یکی شنیدم که این پارامتر iflsig یه نوع تریک به مکانیزم csrf هست که ریدایرکت نوتیس رو برطرف کنه....، توکنیه که بعد از چند دقیقه اکسپایر میشه... اما جزئیات بیشتر ، از گوگل بخواین (:

نکته 2 : من CSRF و فامیلاشونو یه جایی قبلاً توضیح دادم، پیداش میکنم و مقاله‌ش میکنم میذارم اینجا...

5- پارمتر ved

ved=0ahUKEwiN9-CA5eP8AhX2gv0HHZXSAAEQ4dUDCAc

این یکی از وحشتناک ترین پارامترها هست درباره اینه که لینک چجوری کلیک شده... خودش 5 تا ساب-پارامتر داره:

ساب-پارامتر اول میگه که لینک کجای صفحه بوده،

ساب-پارامتر دوم میگه نوع لینک چی بوده،

ساب-پارامتر سوم ... میگه نتیجه سرچ کجا صفحه...

من دیگه توضیح نمیدم... خودتون بخونید... از این صفحه خیث که آپی ایرانو بسته

<https://moz.com/blog/inside-googles-ved-parameter>

6- پارامتر uact

uact=5

7- پارامتر oq

oq=sexy

8- پارامتر gs_lcp

gs_lcp=Cgdnd3Mtd2l6EAMyBQgAEIAEMgUIABCABDIFCAAQgAQyBQgAEIAEMgUILhCABDIFCC4QgAQyBQguEIAEMgUIABCABDIFCAAQgAQyBQgAEIAE0gsILhCABBDHARDRAzoICC4QgAQQ1AJQAFjuC2DT1wFoAXAAeACAACACiAHdCJIBBTItMi4ymAEAoAEB

9- پارامتر sclient

sclient=gws-wiz

راجه به این چهارتای آخر.....من فقط oq رو میدونم که Original Query بوده... یعنی اون چیزی که اول

سرچ کردی... (بعدا ممکنه تغییرش بدی)

بقیه‌ش رو من نمیدونم...

در افسانه ها اومده که اطلاعات مهمی درش نهفته هست و به نحو خبیثانه‌ای انکپسوله شده و رفته... ولی
گوگل ازشون حرفی نمیزنه...

دوستان تماااa

و ما به عنوان یه دوستار HTTP و طرفدار استاد برنرزی، میدونیم که GET محدودیت بایت داره توی ارسال داده.... فقط 8 کیلوبایت مجازه یعنی 8192 بایت ، یعنی 8192 کرکتر...

پس گوگل خبیث اطلاعات اصلیش رو اینجا جا نمیشه که بفرسته...
و اما اونارو کجا میفرسته؟!
POST؟

بریم ببینیم توی این یه سرچ چیا انکپسوله شده بوده توی پستهای HTTP :

Status	Method	Domain	File	Initiator	Type	Transferred	Size
200	GET	www.google.com	bgasy?ei=16_RY7O_Bq_97_UPvp62yAY&v=3&cs=1&asyn=fmt:jspb	m=attn,cdos,crdpf,hsm,...	json	6.10 KB	7.34 269 r
204	POST	www.google.com	gen_204?atyp=i&ei=16_RY7O_Bq_97_UPvp62yAY&dt19=3&zx=1674686425926	m=attn,cdos,crdpf,hsm,...	html	581 B	0 B 222 r
200	GET	www.google.com	fp_204?ei=16_RY7O_Bq_97_UPvp62yAY&atyp=i&zx=1674686425947&client=di	m=ANyn1,CW5FZe,Eox3...	html	761 B	0 B 237 r
200	GET	www.google.com	viewer?query=sexy&origin=https://www.google.com&cs=1	m=attn,cdos,crdpf,hsm,...	html	24.66 KB	81.4 470 r
200	GET	ssl.gstatic.com	b5685c4c7274e704e68a8f30117acdf752f62a387.png	m=C8ffd,CCowhf,CgfbTd...	png	4.35 KB	3.6 325 r
200	GET	encrypted-tbn0.gstatic.com	images?q=tbn:ANd9GcRvU0Gzvw174XKQExaxDsW2k2zo9MahbteITf0r9Tlw...	m=C8ffd,CCowhf,CgfbTd...	jpeg	6.21 KB	5.3 159 r
200	GET	encrypted-tbn0.gstatic.com	images?q=tbn:ANd9GcSMrHCIEBxczW-EVcga7RTzfzBzDMDyBNH1C5jwC5EMQ...	m=C8ffd,CCowhf,CgfbTd...	jpeg	4.21 KB	3.3 111 r
200	GET	encrypted-tbn0.gstatic.com	images?q=tbn:ANd9GcSKTExpIK5ZsW06F6bG6XpdXD_BgamTeVfNavGgzrIO&s...	m=C8ffd,CCowhf,CgfbTd...	jpeg	3.64 KB	2.7 113 r
200	GET	encrypted-tbn0.gstatic.com	images?q=tbn:ANd9GcRMaEEbBk8_5BoYo55XvIagbs-5DA0wppQYcnol7cswzA...	m=C8ffd,CCowhf,CgfbTd...	jpeg	7.44 KB	6.5 122 r
200	GET	encrypted-tbn0.gstatic.com	images?q=tbn:ANd9GcTd3-daWhV8eCnj8fCMVaqfVG3B1YydD3yBkkTBDB&s...	m=C8ffd,CCowhf,CgfbTd...	jpeg	4.54 KB	3.6 164 r
200	OPTIONS	play.google.com	log?format=json&hasfast=true&authuser=0	xhr	plain	630 B	0 B 639 r
200	GET	www.google.com	m=sy1v,sy60,uxMpu,sy1e,sy1g,byfTOB,sy1h,lsjVmc,sy2j,OTA3Ae,sy2i,COQbmf...	m=attn,cdos,crdpf,hsm,...	js	272.37 KB	934 752 r
200	GET	www.google.com	fp_204?ei=16_RY7O_Bq_97_UPvp62yAY&atyp=i&zx=1674686426656&client=di	m=ANyn1,CW5FZe,Eox3...	html	761 B	0 B 243 r
200	POST	play.google.com	log?format=json&hasfast=true&authuser=0	m=attn,cdos,crdpf,hsm,...	plain	737 B	131 871 r
200	GET	www.gstatic.com	m=byfTOB,lsjVmc,LEikZe	m=_b_tp_r604 (script)	js	13.92 KB	35.2 214 r
200	GET	www.gstatic.com	m=ws9Tlc,n73qwf,IZT63,UUJqVe,O1Gjze,xUdlpf,OTA3Ae,COQbmf,fKUV3e,aurF...	m=_b_tp_r604 (script)	js	356.88 KB	1.1 594 r

بطور کلی تا اونجایی که من میبینم 81 تا درخواست گت/پست رفته و برگشته *-
طبیعیه که چندین تا درخواست داشته باشیم... چون HTTP استیتلس هست... ولی 81 طبیعیه؟!

و اما توی اون پستها چی انکپسوله شده؟!
من نمیتونم همش رو بررسی کنم... مثلا ببینین که دومیش اینه:

POST

[https://www.google.com/gen_204?](https://www.google.com/gen_204?atyp=i&ei=wa_RY82SNvaF9u8PldmHiAo&ct=slh&v=t1&im=M&pv=0.9109014057876286&me=12:1674686414028,V,0,0,0,0:1029,h,1,1,o:2121,h,1,1,i:532,h,1,1,o:12,h,1,1,i:475,h,1,1,o:3,h,1,1,i:9,h,1,1,o:3,h,1,1,i:102,h,1,1,o:7,h,1,1,i:396,V,0,0,1920,456:126,R,1,1,0,0,1908,456:0,G,1,1,1000,188:26,V,0,0,1920,456,1:0,R,1,1,0,0,1908,456:1077,B,659:3402,e,H&zx=1674686423348)

[atyp=i&ei=wa_RY82SNvaF9u8PldmHiAo&ct=slh&v=t1&im=M&pv=0.9109014057876286&me=12:1674686414028,V,0,0,0,0:1029,h,1,1,o:2121,h,1,1,i:532,h,1,1,o:12,h,1,1,i:475,h,1,1,o:3,h,1,1,i:9,h,1,1,o:3,h,1,1,i:102,h,1,1,o:7,h,1,1,i:396,V,0,0,1920,456:126,R,1,1,0,0,1908,456:0,G,1,1,1000,188:26,V,0,0,1920,456,1:0,R,1,1,0,0,1908,456:1077,B,659:3402,e,H&zx=1674686423348](https://www.google.com/gen_204?atyp=i&ei=wa_RY82SNvaF9u8PldmHiAo&ct=slh&v=t1&im=M&pv=0.9109014057876286&me=12:1674686414028,V,0,0,0,0:1029,h,1,1,o:2121,h,1,1,i:532,h,1,1,o:12,h,1,1,i:475,h,1,1,o:3,h,1,1,i:9,h,1,1,o:3,h,1,1,i:102,h,1,1,o:7,h,1,1,i:396,V,0,0,1920,456:126,R,1,1,0,0,1908,456:0,G,1,1,1000,188:26,V,0,0,1920,456,1:0,R,1,1,0,0,1908,456:1077,B,659:3402,e,H&zx=1674686423348)

اینا چه معنی میدن؟
از گوگل پرسین.

اما اینجا قسمت مهم ماجراست، شما اگر توی گوگل روی یه لینک کلیک کنید، مستقیم ریدایرکت نمیشید به اون لینک، بلکه اول میرین به سمت سرور گوگل، اونجا یسری از اطلاعات ثبت میشه توی دیتابیس گوگل راجع به اینکه کی و کجا و با چه سشنی روی کدوم لینک کلیک کردین، و بعدش استاد شمارو ریدایرکت میکنه به مقصد.

یعنی اگر من بخوام روی یکی از نتیجه ها کلیک کنم ، هاپیر لینکی که من رو ریدایرکت میده ، مستقیما نمیره سمت داکيومنت مدنظر... بلکه اول میره سمت سرور گوگل ، بهش میگه که این طرف که اونو سرچ کرده بود ، اینقدر زمان رو فکر کرد و اسکرول کرد و با این سرعت روی فلان نتیجه کلیک کرد:))

چطوری؟ اینطوری:

این دیگه تحلیلش به عهده خودتون*-*

25 تا رکوئست رفت و برگشت... همیشه 1/3 رکوئستهای گوگل:

The screenshot shows a web browser with the DuckDuckGo search engine. The search term is "sexy". The results show the word "sexy" as an adjective with its phonetic transcription "sɛk'sē" and three definitions: "Arousing or tending to arouse sexual desire or interest.", "Highly appealing or interesting; attractive.", and "Having sexual appeal; suggestive of sex." Below the definitions is a link to "More at Wordnik".

The bottom of the screenshot shows the Chrome DevTools Network tab. It displays a list of network requests. The first request is a POST to "improving.duckduckg...". The subsequent requests are GET requests to "duckduckgo.com" for various files: "hs_firefox_v362-1?5747607&i=false&atbi=true&va=_&", "/?q=sexy&t=h_", "ProximaNova-Reg-webfont.woff2", "ProximaNova-Sbold-webfont.woff2", "b136.js", "l132.js", and "duckduckgo14.js".

Status	Method	Domain	File
200	POST	improving.duckduckg...	hs_firefox_v362-1?5747607&i=false&atbi=true&va=_&
200	GET	duckduckgo.com	/?q=sexy&t=h_
200	GET	duckduckgo.com	ProximaNova-Reg-webfont.woff2
200	GET	duckduckgo.com	ProximaNova-Sbold-webfont.woff2
200	GET	duckduckgo.com	b136.js
200	GET	duckduckgo.com	l132.js
200	GET	duckduckgo.com	duckduckgo14.js

Summary: 36 requests, 459.51 KB / 52.72 KB transferred, Finish: 4.46 s, DOMContentLoaded: 939 ms

هدر URL ش رو ببینید!

https://duckduckgo.com/?q=sexy&t=h_&ia=definition

1- پارامتر کوئری

q=sexy

2- پارامتر t

_t=h

احتمالا سورس رو نشون میده -> هومپیج

3- پارامتر

ia=definition

که چون دفینیشن کلمه کوئری رو آورده اومده... فک کنم اگه دفینیشن نداشته باشیم میشه:

ia=web

همین 3 تا تمام چیزایی هست که انکپسوله کرده...

البته توی پستهای اطلاعات بیشتری هست... ولی شما خودتون تحلیل بفرمائین.

+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=

مساله بعدی مهم این که برخلاف گوگل یا دیگر خبیثان، هایپرلینکها مستقیم میرن به نتیجه...

لاگ نمیندازن به جا:

تا همینجا میتونیم بفهمیم که چقدر ادعای داکداکگو صحت داره

من به عنوان کسی که قبلا به سرچ انجین نوشتم و یکمی با کارکردش آشنایی دارم، میدونم که حل کردن مساله کداستارت چقدر سخته ، و از طرفی اگر شما به عمد نخواید که اطلاعات رو جمع آوری کنید ، کیفیت سیستم

و خدماتتون میاد پایین...
چون شناختی از کاربر ندارین.

گوگل اینقدر خدماتش خوبه ، چون شمارو از خودتون هم بهتر میشناسه...
شرکت تجاریه و انتظاریم ازش نداریم ، اما اگر دنبال امنیت و حریم شخصی و ناشناسی هستین (که شاید
بعدا بگیم که چرا مهمه) بهتره که اول از تعویض سرچ انجینتون شروع کنید...

الان داکداکگو مکانیزم های خیلی جالبی رو پیش گرفته... مثلا اینکه خودش یه سری سرچ انجین دیگه رو
پراکسی میکنه ، و خودشو میذاره وسط ، از این طرف هویت شمارو جلوی اونا فاش نمیکنه...
با این حرکات میتونه تا حد خیلی خوبی کمبودش رو نسبت به گوگل بهبود ببخشه...
ولی احتمالا نمیتونه خیلی بهبود ببخشه ، چون اون موقع شما حق دارید که شک کنید چجوری شده که
اینقدر خوب شده اگر مسیر دیگه خبیثان رو نرفته:))

اگه علاقه مند به زیر باک داکداکگو هستین:

<https://github.com/duckduckgo>

کد :

<https://deedpolloffice.com/blog/articles/decoding-ei-parameter>

```

function ei_decode($ei)
{
    // Copyright 2013 Deed Poll Office Ltd, UK
    <https://deedpolloffice.com>
    // Licensed under Apache Licence v2.0
    <http://apache.org/licenses/LICENSE-2.0>
    $ei = base64_decode(str_replace(array('_', '-'), array('+', '/'),
    $ei));
    if (!preg_match('/^
        (.{4})
        ((?:[\x80-\xff]*[\0-\x7f])+
        $/sx', $ei, $matches)) return false; // Non-valid ei value
    $ret = array();
    $val = 0;
    foreach (str_split($matches[1]) as $i => $c)
        $val = PHP_INT_SIZE < 5 && function_exists('bcadd') ?
            bcadd($val, bcmul(ord($c), bcpow(2, $i * 8))) :
            $val + (ord($c) << $i * 8);
    $ret[0] = $val;
    preg_match_all('/[\\x80-\\xff]*[\\0-\\x7f]/', $matches[2], $matches,
    PREG_SET_ORDER);
    foreach ($matches as $j => $match) {
        $val = 0;
        foreach (str_split($match[0]) as $i => $c)
            $val = PHP_INT_SIZE < 8 && function_exists('bcadd') ?
                bcadd($val, bcmul(ord($c) & 0x7f, bcpow(2, $i * 7))) :
                $val + ((ord($c) & 0x7f) << $i * 7);
        $ret[$j + 1] = $val;
    }
    return $ret;
}

```

منابع:

<https://github.com/duckduckgo>

<https://help.duckduckgo.com/privacy/t/>

<https://help.duckduckgo.com/duckduckgo-help-pages/settings/params/>

<http://stackoverflow.com/questions/2659952/ddg#2659995>

<https://stackoverflow.com/questions/69660435/what-are-the-components-of-a-google-com-url-string>

<https://stackoverflow.com/questions/70866734/what-does-the-ved-parameter-in-a-google-search-refer-to>

<https://cs50.stackexchange.com/questions/38839/cs50w-project-0-im-feeling-lucky>

<https://deedpolloffice.com/blog/articles/decoding-ei-parameter>

<https://stackoverflow.com/questions/18584386/what-does-ei-mean-in-the-google-homepage-url-https-www-google-co-in-gws-rd#20753179>