CONTENTS

Introduction ............................................................................................................................................ 3

Dedicated Sites for Specific Social Engineering Topics Were Launched (October 2006) ....... 5

Google Launched Hot Trends (May 2007)................................................................................................ 6

Malicious Use of SEO Techniques Sprouted (November 2007)....................................................... 7

Blogspot Hosted Blackhat Search-Engine-Optimized Pages (December 2007).......................... 10

Compromised Sites Played Host to Blackhat SEO (January 2008)................................................ 11

Blackhat SEO Started to Use Hot Trends (March 2009).................................................................... 13

Google Launched Real-Time Search (December 2009).................................................................... 15

Looking into the Future: The Real-Time Web...................................................................................... 16

Conclusion............................................................................................................................................... 17

References........................................................................................................................................... 18
INTRODUCTION

Search engine optimization (SEO), a specialized domain once limited to online marketing experts and promoters, is now being extensively used by cybercriminals to promote unwanted or malicious sites. SEO has become so big that the term "blackhat SEO" now usually refers to search-engine-optimized pages that lead to malicious sites rather than to the use of blackhat methods to artificially obtain higher search result ranking.

This research paper will not delve into the technical details of the latest blackhat SEO kits, as CA, Sophos, and SANS have already published well-written analyses on these. Rather, this paper explores how blackhat SEO has evolved over time as well as the causal events that most probably contributed to the increase in number and effectiveness of blackhat SEO techniques.
Note: Clicking each blue or gray text box in the diagram above leads to specific pages in the paper where each topic is discussed in more detail.

Figure 1. Blackhat SEO development time line
DEDICATED SITES FOR SPECIFIC SOCIAL ENGINEERING TOPICS WERE LAUNCHED (OCTOBER 2006)

In 2006, STRATION and NUWAR variants polluted mailboxes, ZLOB variants came disguised as media codecs, SOHANAD variants arrived via instant messages (IMs), and Internet Explorer (IE) and Microsoft Office became ripe with vulnerabilities. The huge number of available propagation vectors enticed cybercriminals to make blackhat SEO a means to push their malicious creations.

Back then, however, some enterprising cybercriminals created several niche sites to push malware. One example of such was a “travel policy” site that was specifically created to install a backdoor onto users’ systems via drive-by downloads. Placing second in Google, this malicious site proved that with the use of the proper keywords, a simple search can turn into a series of unfortunate events for unsuspecting users.

Figure 2. Google search for “travel policy” leads to a malicious site
How Blackhat SEO Became Big

GOOGLE LAUNCHED HOT TRENDS (MAY 2007)

Google, in its official blog, announced the launch of Hot Trends (now known as Google Trends), which shares the “hottest current searches with users in very close to real-time.” This means that the search strings in Google Trends are actual search strings a lot of people are interested in during a particular hour or day.

Google Trends is a useful tool that provides insights on what people are searching for. More importantly, however, it also shows how people are using Google and what events trigger them to conduct an online search.

Unfortunately, however, cybercriminals can also use these insights to design better social engineering ploys and malicious blackhat SEO campaigns.
How Blackhat SEO Became Big

MALICIOUS USE OF SEO TECHNIQUES SPROUTED (NOVEMBER 2007)

By this time, a lot of bloggers and Web masters were already aware of how Google works and how to get high page ranking. Both the Google Analytics Blog and Google Webmaster Central Blog are more than a year old and offer tips on improving a page’s ranking.

This shows that though using SEO techniques were no longer new, using them to spread malware was. The security industry was jolted by the discovery of several dozen domains that hosted doorway pages stuffed with keywords specifically designed for blackhat SEO attacks on November 2007.

On November 2007, the security industry was jolted by the discovery of several dozen domains that hosted doorway pages stuffed with keywords specifically designed for blackhat SEO attacks.

Figure 3. Domains that hosted doorway pages for blackhat SEO attacks
This blackhat SEO campaign, however, did not just stop at keyword stuffing to increase a malicious page's ranking. The cybercriminals behind it also spammed malicious links that led to doorway pages in several online forums and/or bulletin boards in order to further improve the malicious sites' ranking by way of backlinks.

**Backlinks are incoming links to a website or Web page.**

**Figure 4.** Malicious links spammed in forums and/or bulletin boards
To ensure that the users that arrive on doorway pages do so via Google, blackhat SEO pages perform a referrer check. In addition, these doorway pages also disallow `inurl:` and `site:` queries, as these advanced Google queries are commonly used by security researchers to look for malicious or blackhat search-engine-optimized sites.

![Figure 5. Malicious site code that performs referrer checks](image)

The particular blackhat SEO campaign featured here was highly successful and became the blueprint for future blackhat SEO campaigns and toolkits.
How Blackhat SEO Became Big

BLOGSPOT HOSTED BLACKHAT SEARCH-ENGINE-OPTIMIZED PAGES (DECEMBER 2007)

As comprehensive as the November 2007 blackhat SEO attack was, the cybercriminals behind it made one glaring mistake. They used purchased domains to host the blackhat SEO pages, which was fine if they were creating a valid site but did not make sense if all they needed was a doorway page to lure unsuspecting users to their specially crafted malicious sites.

So, less than a month after the November 2007 massive blackhat SEO campaign, dozens of blackhat SEO doorway pages were discovered in Blogspot.

Less than a month after the November 2007 massive blackhat SEO campaign, dozens of blackhat SEO doorway pages were discovered in Blogspot.

Figure 6. Blackhat SEO doorway page hosted on Blogspot
How Blackhat SEO Became Big

COMPROMISED SITES PLAYED HOST TO BLACKHAT SEO (JANUARY 2008)

After realizing that doorway pages should be free, the cybercriminals, after only about a month, decided that hosting blackhat SEO pages on compromised sites was the way to go.

Several blackhat SEO pages, discovered some time in January 2008, were stuffed with “Heath Ledger”- and “Michelle Williams”-related keywords. This proved to be a lucky break for cybercriminals, as Heath Ledger’s untimely demise directed more than the usual number of curious onlookers to their doorway pages.

This incident marked the first time Trend Micro noticed how blackhat SEO pages were used as doorways to FAKEAV download sites. Before this, blackhat SEO pages only either led to exploit-ridden sites that installed Trojans/backdoor programs into systems or to fake codec download sites.
On the user front, this incident revealed two usual behaviors that the cybercriminals could use to their advantage. First, it revealed that users often turned to online search engines to seek out the latest breaking news. This means that users relied on search engines instead of directly going to established news sites for news, as confirmed by a Hitwise blog entry in February 2010.


**Figure 8. Upstream websites users visited before news and media sites**

The top 3 referrers to news and media sites were found to be Google, Yahoo!, and MSN. All of which had online search functions. This shows that users trust online search engines to lead them to the news they are looking for, which leads us to the second user behavior that cybercriminals exploit—users trust the results they obtain from online search engines.

Two different research studies confirmed that users were likely to trust and click the top results Google returns, regardless of relevance to the information they are looking for.
BLACKHAT SEO STARTED TO USE HOT TRENDS (MARCH 2009)

Before April 2009, the success of blackhat SEO campaigns remained sporadic, as their launch mostly relied on scheduled events or holidays.

However, come April 2009, Trend Micro came across three distinct blackhat SEO campaigns, each leveraging an item in Google’s Hot Trends’ top 20 list—Easter, Nikki Catsouras, and a Twitter worm.

![Figure 9. Number of Google hits for “easter”](source)

![Figure 10. Number of Google hits for “nikki catsouras”](source)

Before April 2009, the success of blackhat SEO campaigns remained sporadic, as their launch mostly relied on scheduled events or holidays.
The number of successful blackhat SEO attacks steadily increased after April 2009. This was mostly due to the use of timely keywords, unwittingly aided by Google’s Hot Trends.

Figure 11. Number of Google hits for “twitter worm”

The number of successful blackhat SEO attacks steadily increased after April 2009. This was mostly due to the use of timely keywords, unwittingly aided by Google’s Hot Trends.
**GOOGLE LAUNCHED REAL-TIME SEARCH**
**(DECEMBER 2009)**

One of Google’s best-kept secrets is PageRank, a simple idea that shows that the more sites link to a site makes that site more relevant and allows it to rise in terms of page ranking. To rank pages, however, Google must first crawl and index a website. This site should have linked sites in order to get a high page ranking.

While PageRank ensures relevance, however, it fails to cover newly published content a user may be looking for simply because there are very few sites linking to the new content.

Then came Twitter and Facebook. Although these are already the top microblogging and social networking sites, these still pose real threats to Google through the sheer amount of searchable real-time information their users share.

The plane crash over the Hudson River in January 2009 highlights Twitter’s dominance in providing information on breaking news. The first Tweet reporting the plane crash appeared 15 minutes before any news site reported the event. Google’s Hot Trends took more than an hour to show keywords related to the said crash.

This is exactly the type of instance Google wants to address with the release of **Real-Time Search**. By including Twitter, Facebook, blogs, and other user-generated feeds in search results and by modifying PageRank to become less restrictive, Google aims to provide fresh content with regard to breaking news or trending topics, in particular.

However, as shown in Figure 1 earlier, the number of reported blackhat SEO attacks dramatically increased around the same time Google released Real-Time Search. In fact, performing a Google search on notable and trending events such as Brittany Murphy’s death, the Haiti and Chile earthquakes, and even iPad’s launch turned up blackhat SEO pages as results.

Once again, Google unwittingly provided a conducive environment for the proliferation of blackhat SEO pages. To be fair though, Google has begun filtering search results, tagging blackhat SEO pages with a “This site may harm your computer” notice a few months after Real-Time Search went live.
LOOKING INTO THE FUTURE: THE REAL-TIME WEB

Conducting searches is one of the top activities people do online, allowing it to become the main traffic driver to sites for years now. Blackhat SEO is, however, just one of the means by which cybercriminals use online searches to drive traffic to their malicious sites.

As early as last year, social networking has been posing a threat to conducting online searches in terms of becoming the fastest-growing online activity.

In a more recently released Nielsen study, however, social networking has become the top-ranking online activity, ushering in the era of “the real-time Web.”

The real-time Web is a catchphrase that depicts the Twitter and Facebook phenomenon that made streams of user information available to others in real-time. These information streams have been pegged as a next-generation gold mine that can possibly give rise to the “next Google.”

The type of information users share on the real-time Web is personal and unique. Links leading to blogs, news articles, pictures, and videos are usually shared on a user’s social network, which effectively serves as an endorsement of the content created by a friend or contact, which Facebook’s “Like” function, for instance, further enforces.

Malware such as KOOBFACE variants have effectively been using this “endorsement model” to infect thousands of systems. Lately, survey spam that trick users into performing a series of steps designed to make them “Like” the spam page have been littering Facebook.

In the very near future, we can expect cybercriminals to fully utilize the real-time Web as a traffic driver to their nefarious sites. This will be very similar to how blackhat SEO is currently being used to drive traffic to FAKEAV pages.
CONCLUSION

The emergence of blackhat SEO as a preferred method to distribute malware demonstrates the complex interaction between online services, user behaviors, and cybercriminals’ opportunistic nature.

This research paper presented how blackhat SEO evolved from its static HTML hit-or-miss days into using dynamic toolkits. These toolkits enable malicious sites to constantly get high page ranking.

In the end, blackhat SEO is just a means by which cybercriminals drive traffic to their malicious sites. The emergence of the real-time Web may, therefore, give cybercriminals a new venue to promote their malicious pages.
REFERENCES


How Blackhat SEO Became Big


How Blackhat SEO Became Big

