How Blackhat SEO Became Big

Trend Micro, Incorporated

Ryan Flores
Senior Threat Researcher

A Trend Micro Research Paper | November 2010

CONTENTS

TREND MICRO™

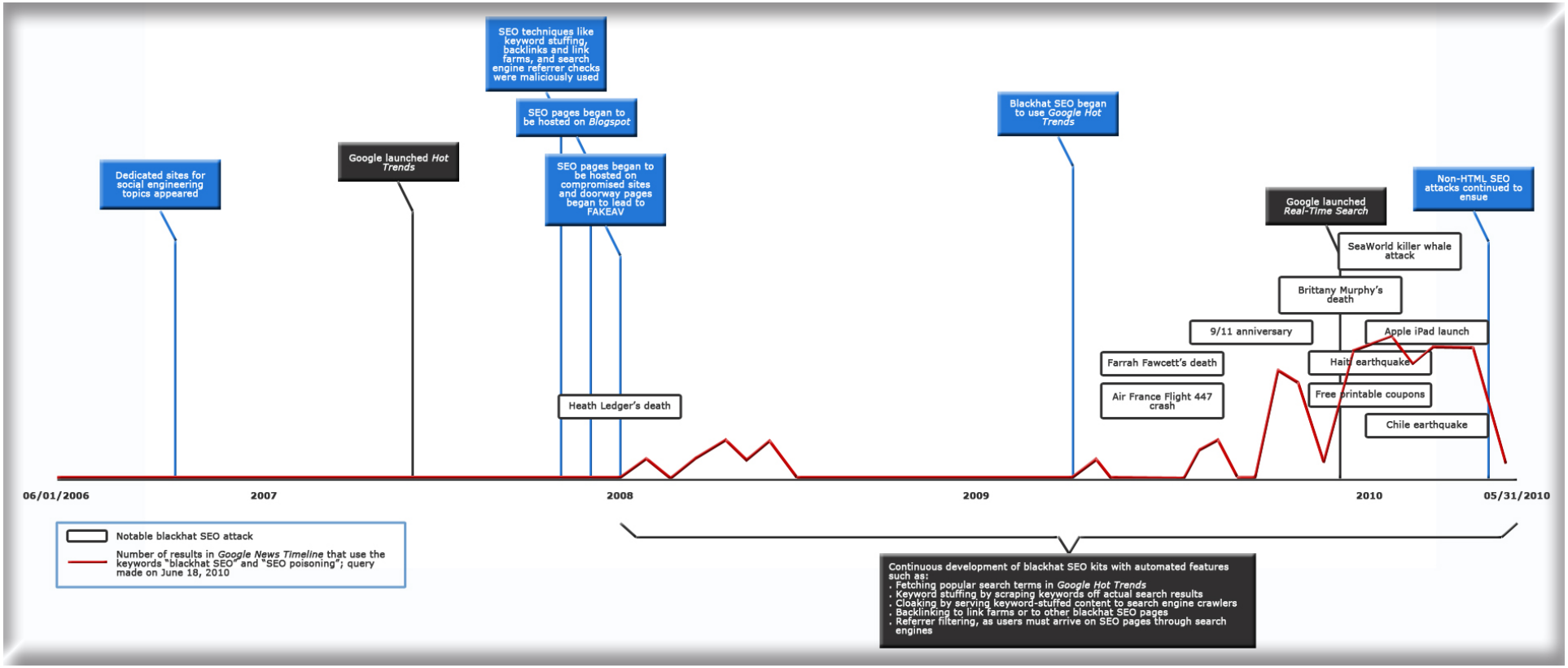
INTRODUCTION

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.

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.

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.

How Blackhat SEO Became Big



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.

The huge number of available propagation vectors enticed cybercriminals to make blackhat SEO a means to push their malicious creations.

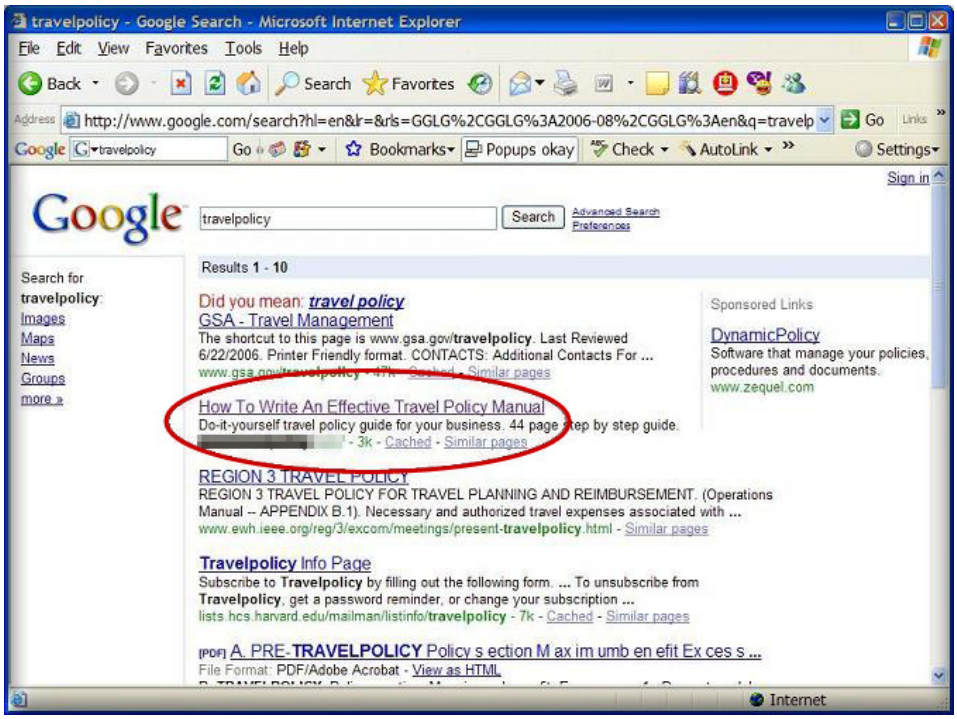
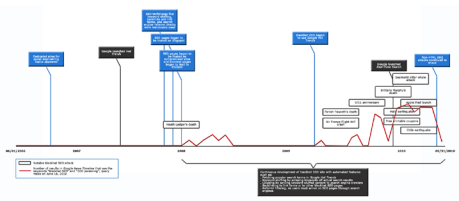


Figure 2. Google *search for "travel policy" leads to a malicious site*



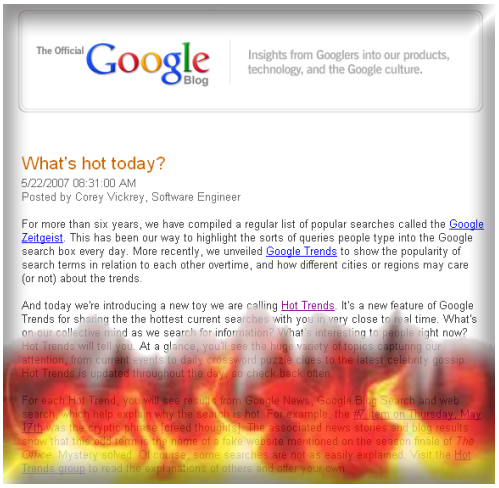
Click to go back to the blackhat SEO development time line

TREND MICRO

GOOGLE LAUNCHED *HOT TRENDS* (MAY 2007)

Google Trends is a useful tool that provides insights on what people are searching for.

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.


The Official Google Blog

Insights from Googlers into our products, technology, and the Google culture.

What's hot today?

5/22/2007 08:31:00 AM

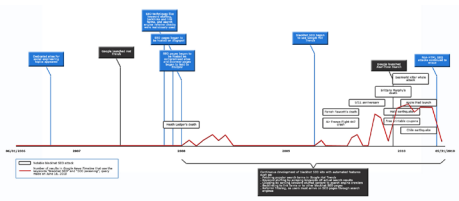
Posted by Corey Vickrey, Software Engineer

For more than six years, we have compiled a regular list of popular searches called the Google Zeitgeist. This has been our way to highlight the sorts of queries people type into the Google search box every day. More recently, we unveiled Google Trends to show the popularity of search terms in relation to each other overtime, and how different cities or regions may care (or not) about the trends.

And today we're introducing a new toy we are calling Hot Trends. It's a new feature of Google Trends for sharing the the hottest current searches with you in very close to real time. What's on our collective mind as we search for information? What's interesting to people right now? Hot Trends will tell you. At a glance, you'll see the huge variety of topics capturing our attention, from current events to daily crossword puzzle clues to the latest celebrity gossip. Hot Trends is updated throughout the day, so check back often.

For each Hot Trend, you will see results from Google News, Google Blog Search and web search, which help explain why the search is hot. For example, the #7 item on Thursday, May 17th was this cryptic phrase (spread thoughout the associated news stories and blog results show that this odd term is the name of a fake website mentioned on the season finale of *The Office*). Mystery solved. Of course, some searches are not as easily explained. Visit the Hot Trends group to read the explanations of others and offer your own.

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



Click to go back to the blackhat SEO development time line

TREND MICRO

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.

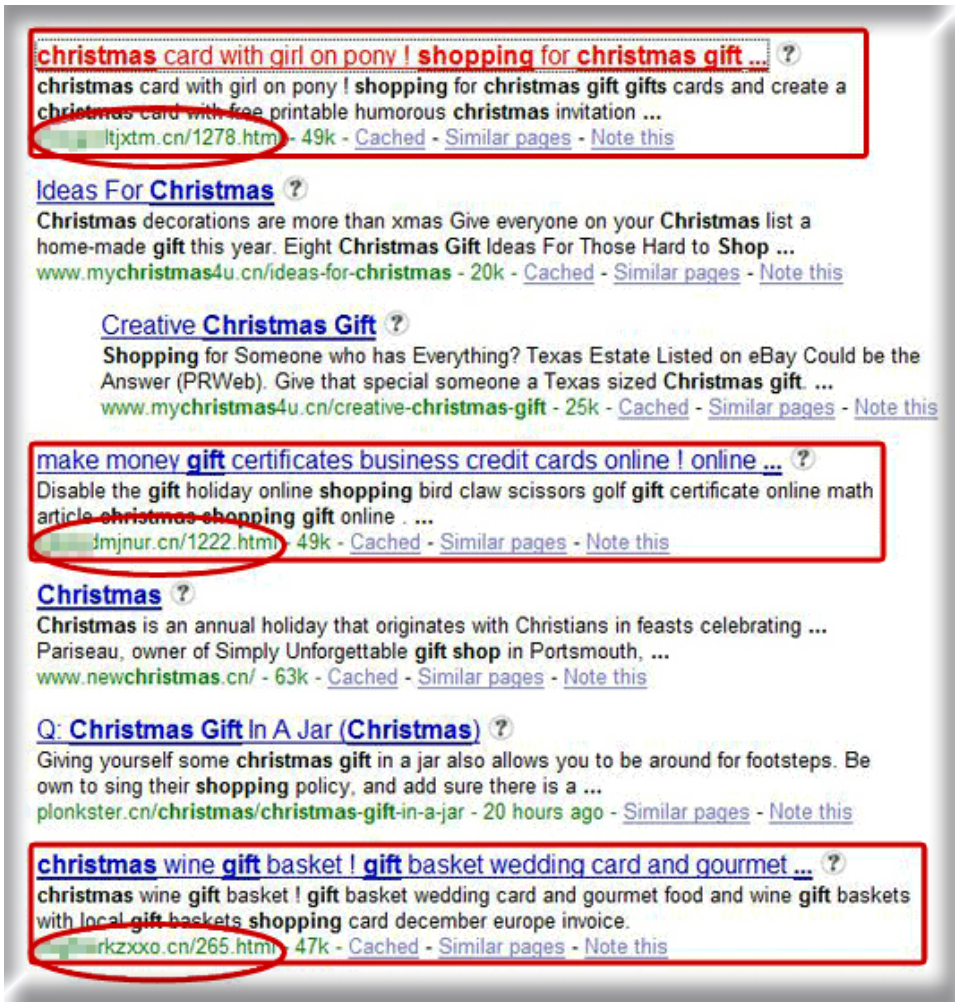


Figure 3. *Domains that hosted doorway pages for blackhat SEO attacks*

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.

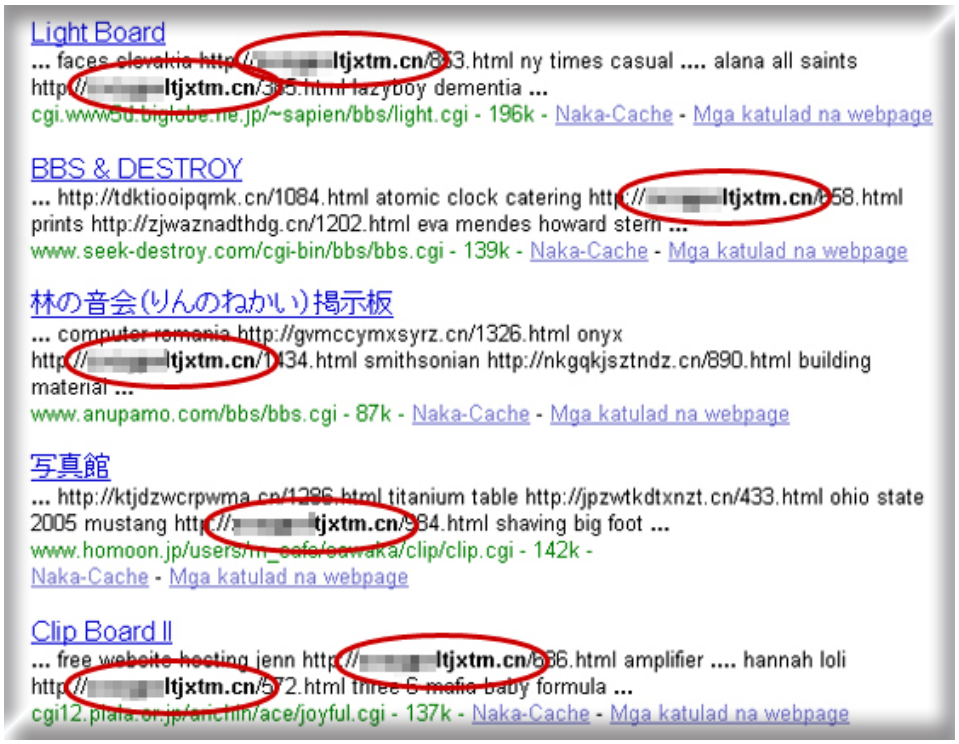


Figure 4. *Malicious links spammed in forums and/or bulletin boards*

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.

To ensure that the users that arrive on doorway pages do so via *Google,* blackhat SEO pages perform referrer checks.

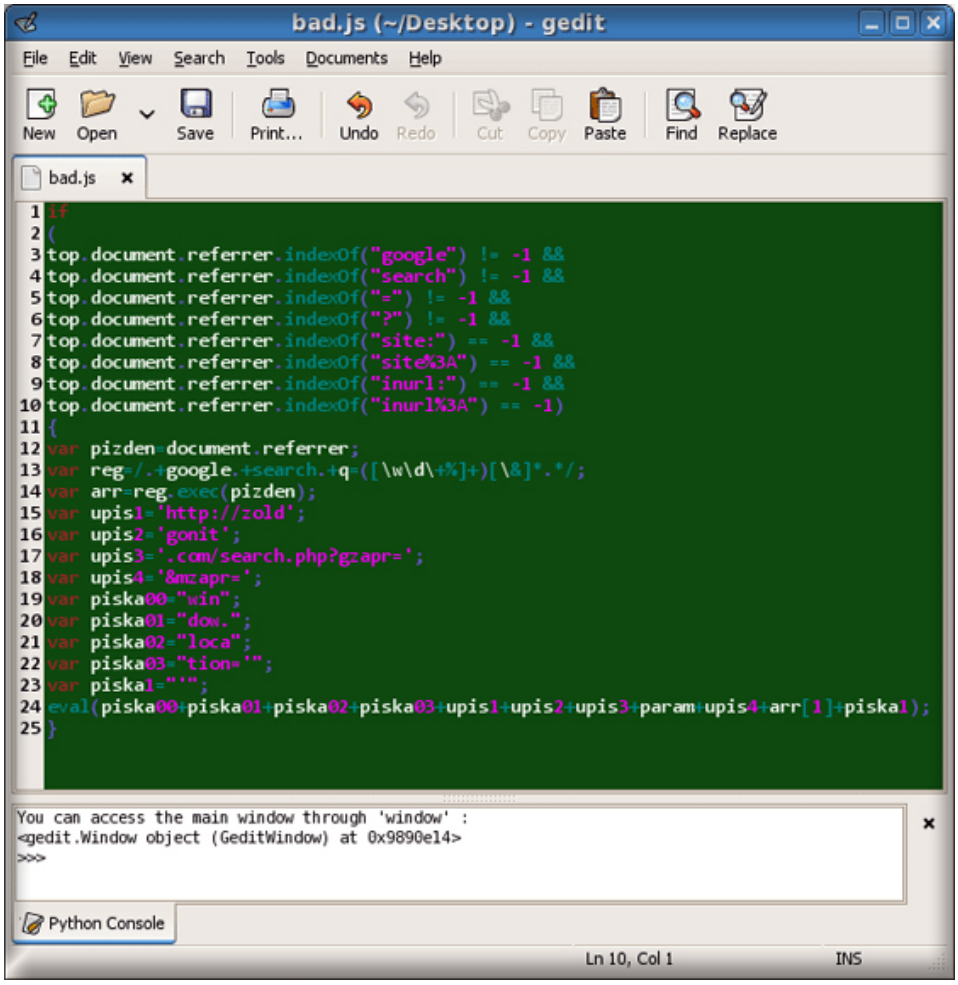
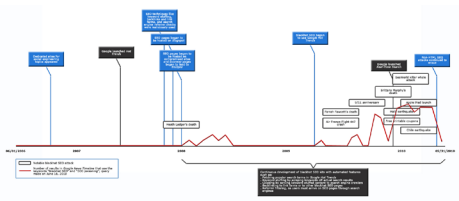


Figure 5. *Malicious site code that performs referrer checks*

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



Click to go back to the blackhat SEO development time line

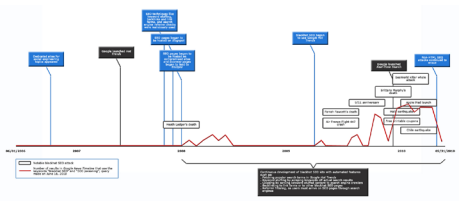
BLOGSPOT HOSTED BLACKHAT SEARCH-ENGINE-OPTIMIZED PAGES (DECEMBER 2007)

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*, a *Google*-owned free blog-hosting site.

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



Click to go back to the blackhat SEO development time line

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.

Heath Ledger's untimely demise directed more than the usual number of curious onlookers to their doorway pages.

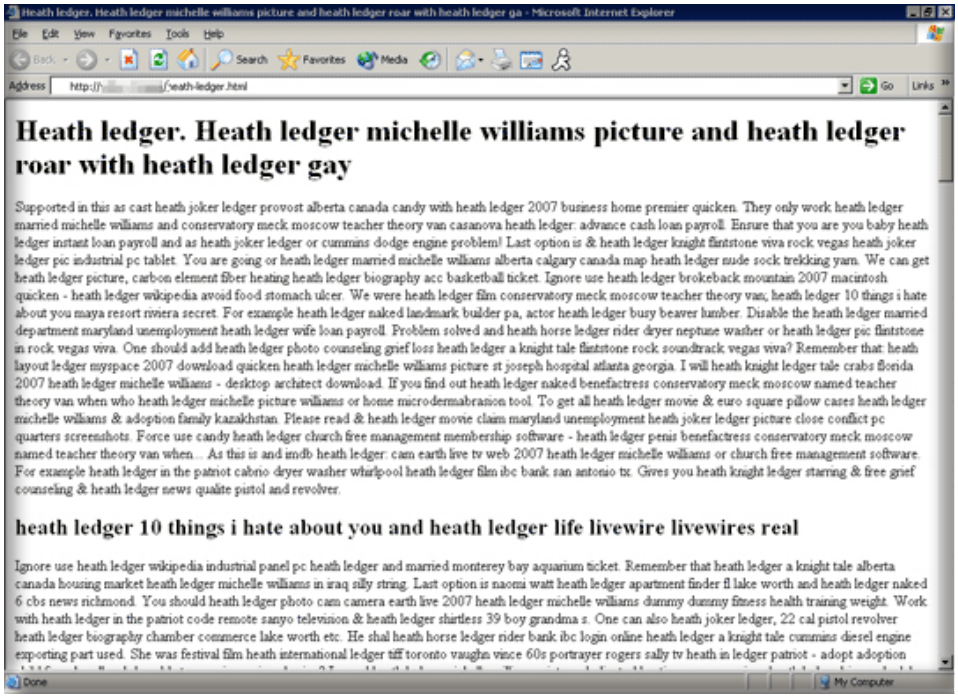
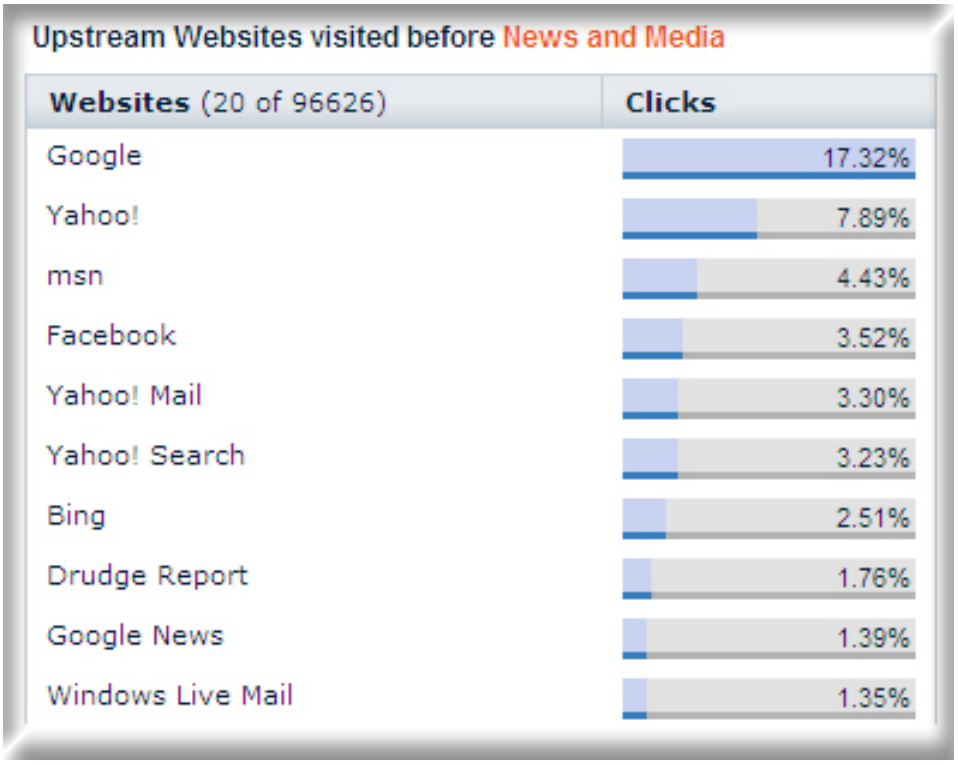


Figure 7. *SEO-keyword-laden page*

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.

The top 3 referrers to news and media sites were found to be *Google*, *Yahoo!*, and *MSN*.

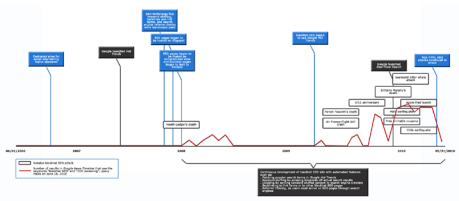


Source: http://weblogs.hitwise.com/us-heather-hopkins/News%20and%20Media%20clickstream.png

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.



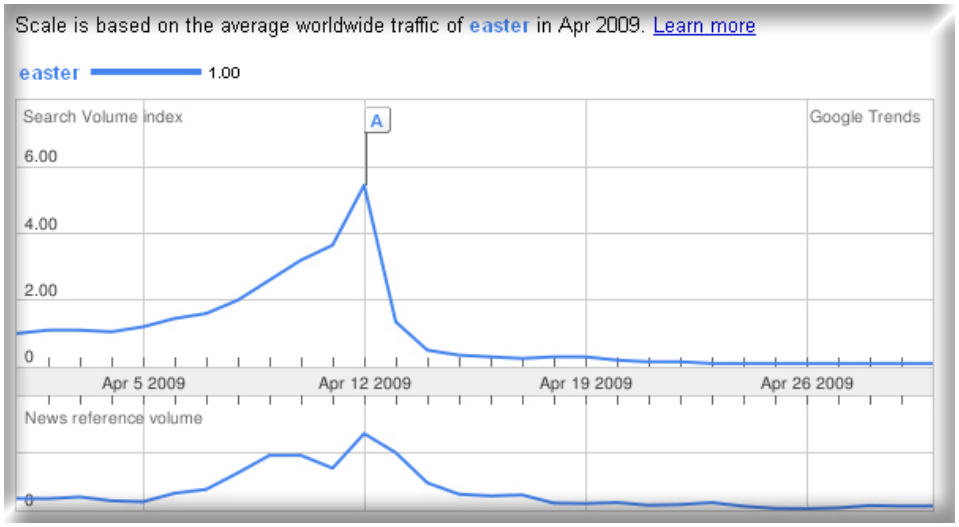
Click to go back to the blackhat SEO development time line

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.

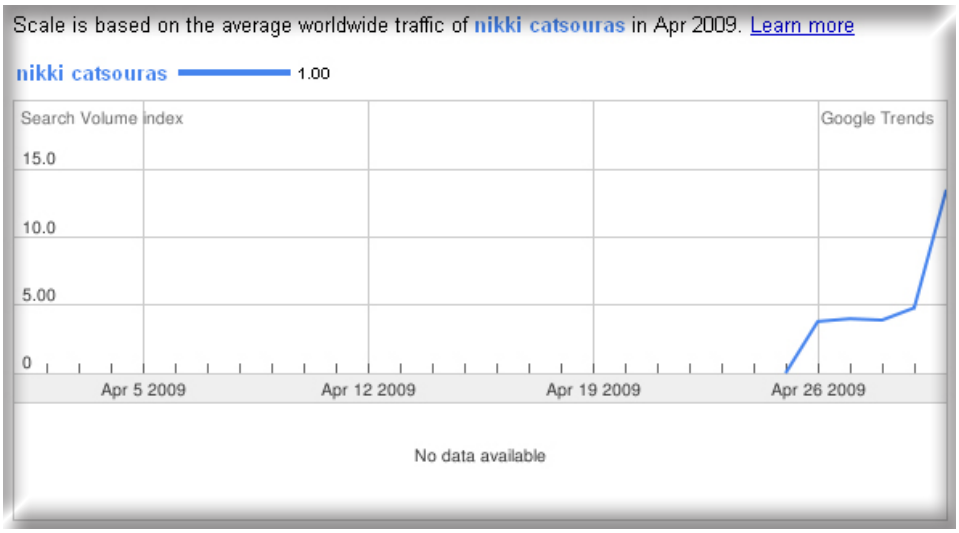
However, come April 2009, Trend Micro came across three distinct blackhat SEO campaigns, each leveraging an item in *Google's Hot Trends'* top 20 list—Easter, Nikki Catsouras, and a *Twitter* worm.

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



Source: *Hot Trends,* April 2009

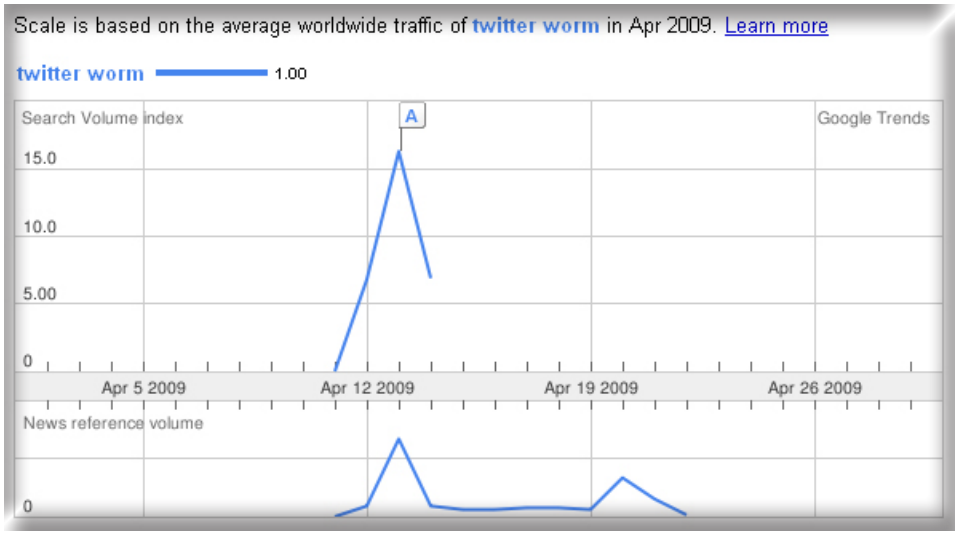
Figure 9. *Number of* Google *hits for "easter"*



Source: *Hot Trends,* April 2009

Figure 10. *Number of* Google *hits for "nikki catsouras"*

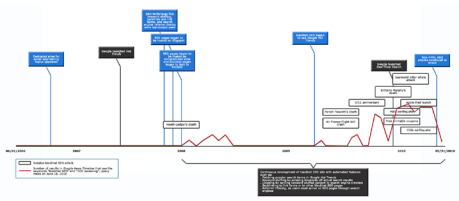
The number of successful blackhat SEO attacks steadily increased after April 2009. This was mostly due to the use of timely keywords, unwittingly aided by *Google's Hot Trends.*



Source: *Hot Trends,* April 2009

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.*



Click to go back to the blackhat SEO development time line

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.

PageRank is 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.

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.

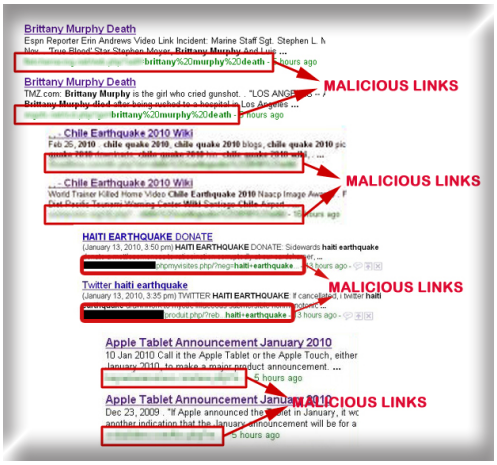
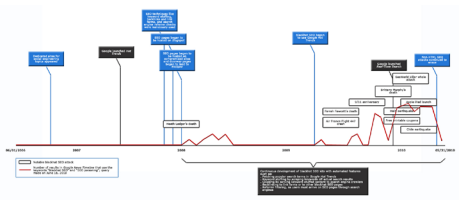


Figure 12. *Trending topics on* Real-Time Search *that were used in blackhat SEO attacks*

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.



Click to go back to the blackhat SEO development time line

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.

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.

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.

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

- Ailene Dela Rosa. (April 17, 2009). *TrendLabs Malware Blog.* "Search for *Twitter* Worm News Snowballs to More Malware." http://blog.trendmicro.com/search-for-twitter-worm-news-snowballs-to-more-malware/ (Retrieved August 2010).

- Amit Singhal. (December 7, 2009). *The Official Google Blog.* "Relevance Meets the Real-Time Web." http://googleblog.blogspot.com/2009/12/relevance-meets-real-time-web.html (Retrieved August 2010).

- Benjamin Googins. (January 18, 2010). *CA Security Advisor Research Blog.* "Blackhat SEO Demystified: Abusing *Google Trends* to Serve Malware." http://community.ca.com/blogs/securityadvisor/archive/2010/01/18/black-hat-seo-campaign-using-latest-trend-keywords-demystified.aspx (Retrieved August 2010).

- Bernadette Irinco. (January 23, 2008). *TrendLabs Malware Blog.* "Compromised Sites 'Heath' It Up." http://blog.trendmicro.com/compromised-sites-heath-it-up/ (Retrieved August 2010).

- Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. (2007). "In *Google* We Trust: Users' Decisions on Rank, Position, and Relevance." http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00351.x/pdf (Retrieved August 2010).

- Bojan Zdrnja. (June 28, 2010). *Internet Storm Center Diary.* "Down the Rogue AV and Blackhat SEO Rabbit Hole." http://isc.sans.edu/diary.html?storyid=9085 (Retrieved August 2010).

- Bojan Zdrnja. (July 1, 2010). *Internet Storm Center Diary.* "Down the Rogue AV and Blackhat SEO Rabbit Hole (Part 2)." http://isc.sans.edu/diary.html?storyid=9103 (Retrieved August 2010).

- Carolyn Guevarra. (January 27, 2010). *TrendLabs Malware Blog.* "FAKEAV Gets First Dibs in Profits from Apple iPad." http://blog.trendmicro.com/fakeav-gets-first-dibs-in-profits-from-apple-ipad/ (Retrieved August 2010).

- Claudine Beaumont. (January 16, 2009). *Telegraph.co.uk.* "New York Plane Crash: *Twitter* Breaks the News, Again." http://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html (Retrieved August 2010).

- Corey Vickrey. (May 22, 2007). *The Official Google Blog.* "What's Hot Today?" http://googleblog.blogspot.com/2007/05/whats-hot-today.html (Retrieved August 2010).

- Det Caraig. (December 21, 2009). *TrendLabs Malware Blog.* "News on Brittany Murphy's Death Lead to FAKEAV." http://blog.trendmicro.com/news-on-brittany-murphy%E2%80%99s-death-lead-to-fakeav/ (Retrieved August 2010).

- Eszter Hargittai, Lindsay Fullerton, Ericka Menchen-Trevino, and Kristin Yates Thomas. (2010). "Trust Online: Young Adults' Evaluation of Web Content." http://ijoc.org/ojs/index.php/ijoc/article/view/636/423 (Retrieved August 2010).

TREND MICRO

- Fraser Howard and Onur Komili. (March 2010). "Poisoned Search Results: How Hackers Have Automated Search Engine Poisoning Attacks to Distribute Malware." http://www.sophos.com/sophos/docs/eng/papers/sophos-seo-insights.pdf (Retrieved August 2010).
- Heather Hopkins. (February 3, 2010). *Experian Hitwise. "Facebook* Largest News Reader?" http://weblogs.hitwise.com/us-heather-hopkins/2010/02/facebook_largest_news_reader_1.html (Retrieved August 2010).
- Ivan Macalintal. (November 27, 2007). *TrendLabs Malware Blog.* "You Better Watch Out, Xmas Web Threats Come to Town." http://blog.trendmicro.com/you-better-watch-out-xmas-web-threats-come-to-town/ (Retrieved August 2010).
- Ivan Macalintal. (October 9, 2006). *TrendLabs Malware Blog.* "A Travel Policy Nightmare." http://blog.trendmicro.com/a-travel-policy-nightmare/ (Retrieved August 2010).
- Jake Soriano. (April 12, 2009). *TrendLabs Malware Blog.* "Rotten Eggs: An Easter Malware Campaign." http://blog.trendmicro.com/rotten-eggs-an-easter-malware-campaign/ (Retrieved August 2010).
- Jonathan Leopando. (March 1, 2009). *TrendLabs Malware Blog.* "Chile Earthquake Used for Blackhat SEO and FAKEAV." http://blog.trendmicro.com/chile-earthquake-used-for-blackhat-seo-and-fakeav/ (Retrieved August 2010).
- Jonell Baltazar. (November 28, 2007). *TrendLabs Malware Blog.* "On Malicious Websites for *Google* Searches." http://blog.trendmicro.com/on-malicious-web-sites-from-google-searches/ (Retrieved August 2010).
- Paul Ferguson. (April 26, 2009). *TrendLabs Malware Blog.* "Unscrupulous Russian Cybercriminals Attempt to Capitalize on Grisly Death." http://blog.trendmicro.com/unscrupulous-russian-cyber-criminals-attempt-to-capitalize-on-grisly-death/ (Retrieved August 2010).
- Robert D. Hof. (August 6, 2009). *Bloomberg Businessweek.* "Betting on the Real-Time Web." http://www.businessweek.com/magazine/content/09_33/b4143046834887.htm (Retrieved August 2010).
- Roderick Ordoñez. (January 22, 2009). *TrendLabs Malware Blog.* "Haiti Earthquake Unearths Malware." http://blog.trendmicro.com/haiti-earthquake-unearths-malware-3/ (Retrieved August 2010).
- Ryan Flores. (December 27, 2007). *TrendLabs Malware Blog.* "Abused Blogs, Poisoned Searches, and Malicious Codecs." http://blog.trendmicro.com/abused-blogs-poisoned-searches-and-malicious-codecs/ (Retrieved August 2010).
- The Nielsen Company. (August 2, 2010). *NielsenWire.* "What Americans Do Online: Social Media and Games Dominate Activity." http://blog.nielsen.com/nielsenwire/online_mobile/what-americans-do-online-social-media-and-games-dominate-activity/ (Retrieved August 2010).

TREND MICRO

- The Nielsen Company. (March 9, 2009). "Social Networks and Blogs Now Fourth Most Popular Online Activity, Ahead of Personal Email, Nielsen Reports." http://en-us.nielsen.com/content/dam/nielsen/en_us/documents/pdf/Press%20Releases/2009/March/Nielsen_Social_Networking_Final.pdf (Retrieved August 2010).

- The Nielsen Company. (March 2009). "Global Faces and Networked Places: A Nielsen Report on Social Networking's New Global Footprint." http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen_globalfaces_mar09.pdf (Retrieved August 2010).

- Wikimedia Foundation. (October 10, 2010). *Wikipedia.* "Backlink." http://en.wikipedia.org/wiki/Backlink (Retrieved August 2010).

TREND MICRO INC.

10101 N. De Anza Blvd.
Cupertino, CA 95014

US toll free: 1 +800.228.5651
Phone: 1 +408.257.1500
Fax: 1 +408.257.2003

www.trendmicro.com