

# Top 5 Things to know about Blackhat Analytics by Phil Pearce (What, Why, How, Then & Now)

Intro

Background

5 things

- What is it?
- Why, do I need to know?
- How do I check? (to see if I am exposed)
- Then: prevention & cures
- Now: takeaways & final thoughts

Questions

# About Me

- Online marketing and WA for 5 years
- Ecommerce and Finance background
- Passionate about Advanced Analytics & pushing the boundaries of Tools & Techniques.
- More:
  - <http://twitter.com/philpearce>
  - <http://www.linkedin.com/in/philpearce>

# Expertise & Certifications

## Active on the following forums

- [Yahoo Web Analytics group](#)
- [Google Analytics forum](#) (alias PPC\_guru - Top Contributor - 542 answers)
- [Old Google Analytics forum](#) (17 star answers)
- [Web Analytics Exchange](#)
- <http://webanalyticsassociation.org/members/?id=7312794>

## Awards

\* **Best web analytics report** - *Runner-up* - Avinash Blog 2008

Phil gets his prize for the best effort, not only not only did he create an admirable custom report that was valuable but he even created images of the mocks. *Bravo!* <http://www.kaushik.net/avinash/2008/06/contest-winners-the-best-web-analytics-report.html>

**Web analytics Haiku poem** - *Honourable mention* - eMetrics Washington 2009

## Certifications:

- Google Analytics IQ (Grade 98%), Google Adwords Professional (Grade: 89%)
- Yahoo Search Ambassador (Grade: 90%), MSN AdExcellence (Grade 86%)
- WAA basecamp - Applying web analytics - 1 day course

# Background

- Blackhat SEO (Google guidelines, EU hacking laws)
- Blackhat PPC (Adwords ToS, Advertising & Trading Standards, FSA, Patents Office)
- ... Blackhat Analytics

# What: Definition

*Intentional act of **distorting, deleting, unethically using, or hijacking** WA data using technical or legal loopholes; with the goal of making **financial gains** or obtaining a **competitive advantage**.*

# Metaphors

- BHA is like playing poker when your opponent knows your hand.
- BHA is like having a nice neat tidy Google Garden, then a mole comes and messes up the lawn!
- BHA is like cheating at an exam
- BHA is like SEO 2years ago
- BHA is the new Arnie, it will be back.
- BHA is like knowing you are paying less per click than your competitor, and will win a sustained bid war.
- BHA is driving at 80 miles per hour in daylight, when your competitors are driving in the dark.

# Examples (shortened list)

- Log spam (backlink spam, fake referral or fake user-agent spam)
- Hidden noscript backlinks spam & pagerank dilution
- Conversion Fraud
- CPA under (or over)bidding
- Behavioural malware
- Monitoring of Competitor visits
- Blocking of competitive intelligence tools
- Browser History CSS hacks (e.g. beencounter)
- GA phishing & Spear-Phishing (e.g Google insight says "GA login" rapidly increasing searches)
- DataDisruption.js scripts:
  - appends the word "xxx" on to all referring keywords, and changes source="other source"
  - randomly changes all referring KWs to a pre-set of 10 phrases
  - track spider visit & resulting in data junk captured
  - 0% bounce rate and default 2 pageviews or more
  - triggers 10+ tracking calls & changing all external links to example.com
- GA ping-pong
- Auto opt-in FireFox extension tracking & Mobile App tracking
- Flash cookie respawn
- Fake ga.js & GA virus (aka Gumblar/Beladen)
- Cookie buffer overflow via iframe
- DNS poisoning - resulting in rerouting of /ga.js to malicious server
- Router spam - auto opt-in of ads on 404 & DNS error pages + dropping of 3<sup>rd</sup> party cookies
- WiFi packet sniffing of utm.gif requests

# Why, do I need to know?

- GA “3 Day Data Death” ToS Notice
- Free WA packages unable to remove PII without deleting whole GA accounts!
- Loss of Business insight and disconnected from the Business pulse
- Bad Data decisions or uninformed decisions
- Brand Damage or embarrassment, and loss of customer trust
- Loss key business assets or removal of Barrier to entry, CPC bid inflation
- Legal implications & fines (£2,000 upto 500,000)

## How do I check? - to see if I am exposed? (part1)

- GA data audit:
  - Hostname check
  - ISP location
  - City
  - IP + UserAgent using special filter
  - URL check using &limit=50000
- Excel tool for mining URL parameters
- Advanced segment reports import for above filters.
- Check Scheduled reports
- Number of Admin users

## How do I check? - to see if I am exposed? (part2)

- GA onsite audit
  - View source code (control-u)
  - Using wasp to look for PII`s in CustomSetVar
  - Xenu site crawl & url export
  - Pages to check - Login pages, password reset pages, email newsletter links.
  - Site:mydomain.com inurl:"utm" inurl:"gmail"
  - Check robot.txt for noindex pages passing PII`s

# Prevention & Cures (part1)

- GA filters:
  - Hostname include filter: `(^|\.)yourdomain.com$`
  - ISP location exclude Ask.com bot: `^(iac search and media europe ltd|iac search media inc)$`
  - Top content report - Contains box:  
email|add|postcode|zipcode|tel etc using `&&limit=50000`
  - IP+useragent filter in GA
  - Check data stored in User-defined, CustomFields and Event fields
  - Check all GA profiles including Raw Data profile for PII`s.

# Prevention & Cures (part2)

- Enable capture of IIS or apache Raw server logs, urchin tracking pixel backup or get Yahoo Analytics account as secondary analytics package.
- Check privacy policy present and if processing PII register with ICO (cost £35)
- FireFox noscript extension & Keep pdf reader updated
- IE8 noframes tag / nosniff
- Use Secure-FTP for global footer changes or lockdown FTP to fixed IP`s, enable php change logging.
- CPA audits (GA vs Affiliate report)
- Ping [www.google-analytics.com/ga.js](http://www.google-analytics.com/ga.js) and [ssl.google-analytics.com/ga.js](http://ssl.google-analytics.com/ga.js)
- Use Certified Ethical Analyst, GAAC or GA qualified individual

# Takeaways & final thoughts (discussion)

- How much is your data worth?
- Are you in a very competitive industry?
- Do you have more than 3 Admin users, more than 10 scheduled reports
- Is sensitive backoffice data or urls being accidentally tracked?
- Can you afford to drive traffic in the dark with no insight?
- When was the last time you audited your WA install?
- Should there be a registrar of Approved Analytics consultants (like medicine), and non-compliance results in removal of industry accolade?
- As no rules exist, who should make the rules and enforce them?
- What is ethical analytics and what is unorthodox, misguided or accidental?

Beer Time...