Click Fraud Detection using Practical Memetics

Defcon 15, Aug 3-5, 2007

Broward Horne

http://www.realmeme.com

Overview

- Advertising Click Fraud
- Botnets (Agents Of Click Fraud)
- S-Curve
- Original Meme Theory (Dawkins)
- Empirical Meme Theory (Me! Ho!)
- Expanded Meme Mining Model
- MySpace Example
- IAFF.com Example
- Meme Seepage Theory
- Botnet Proof
- Gaming Conjectures
- Extrapolations

Click Fraud - Definition

From Wikipedia -

“When a person, script or computer program imitates a legitimate user to generate a “charge per click” of an advertised product”

Is Click Fraud relatively unknown? Poll

Google Adsense Model

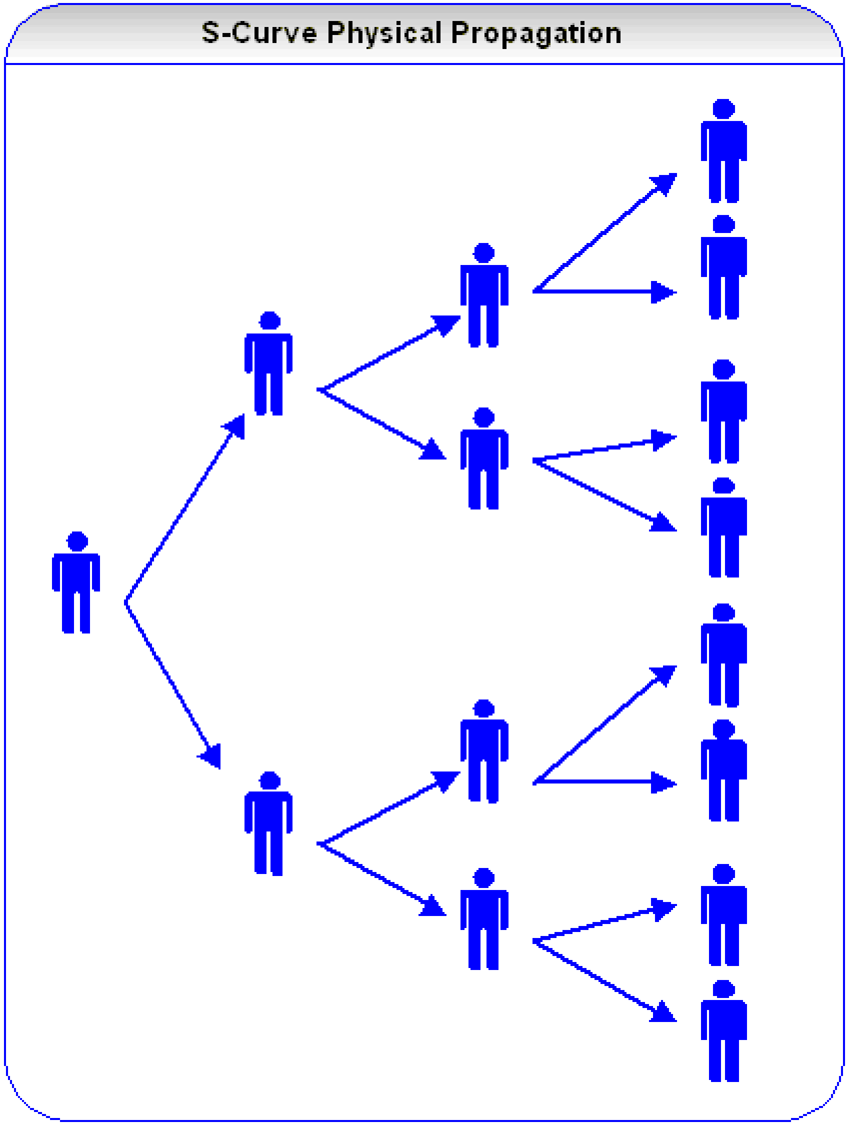
- www.google.com/adsense
- Click fraud risk

Botnets

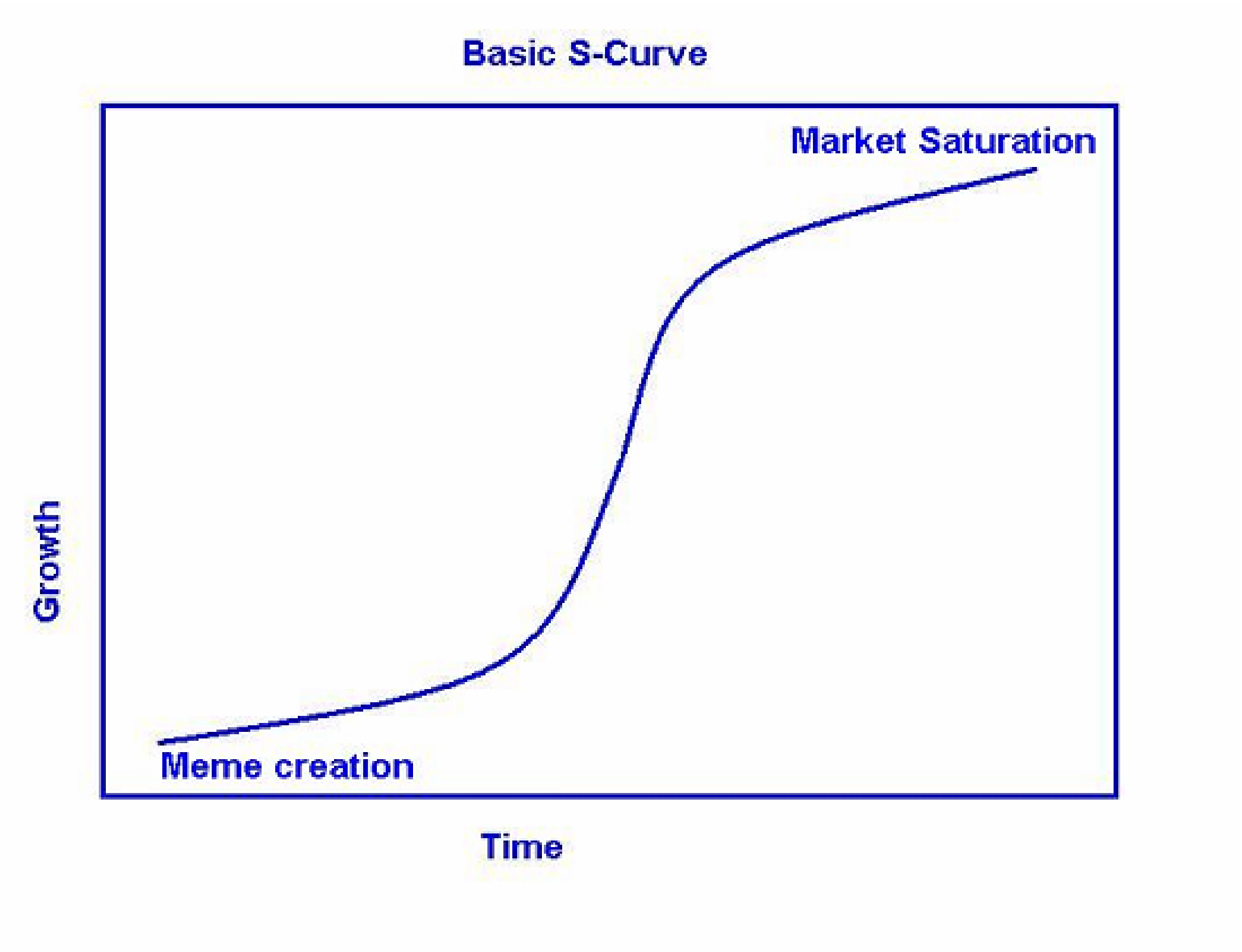
- From Wikipedia-

  “a collection of compromised computers (“zombies”) running… under a common command-and-control infrastructure”

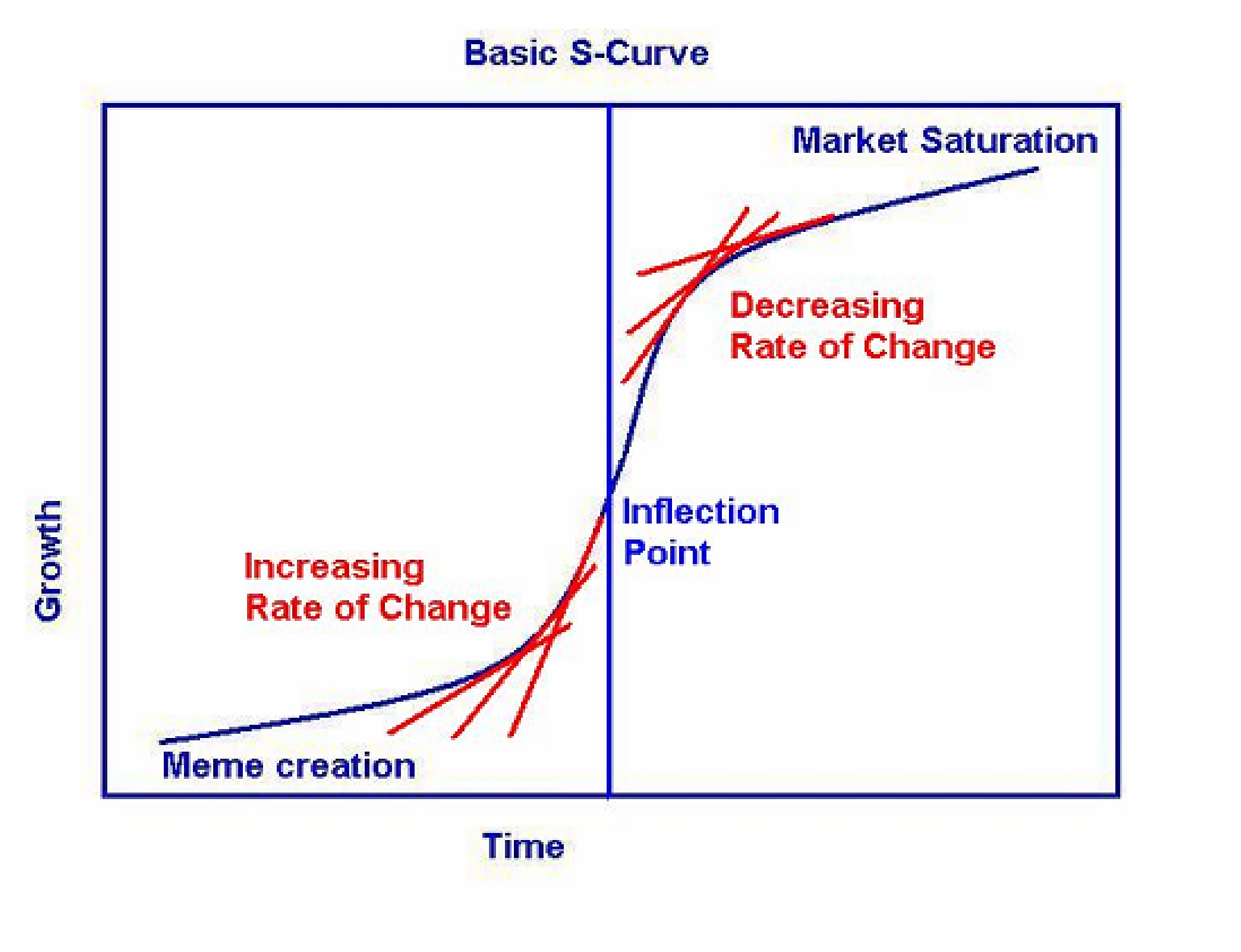
S-Curve (Physical Model)



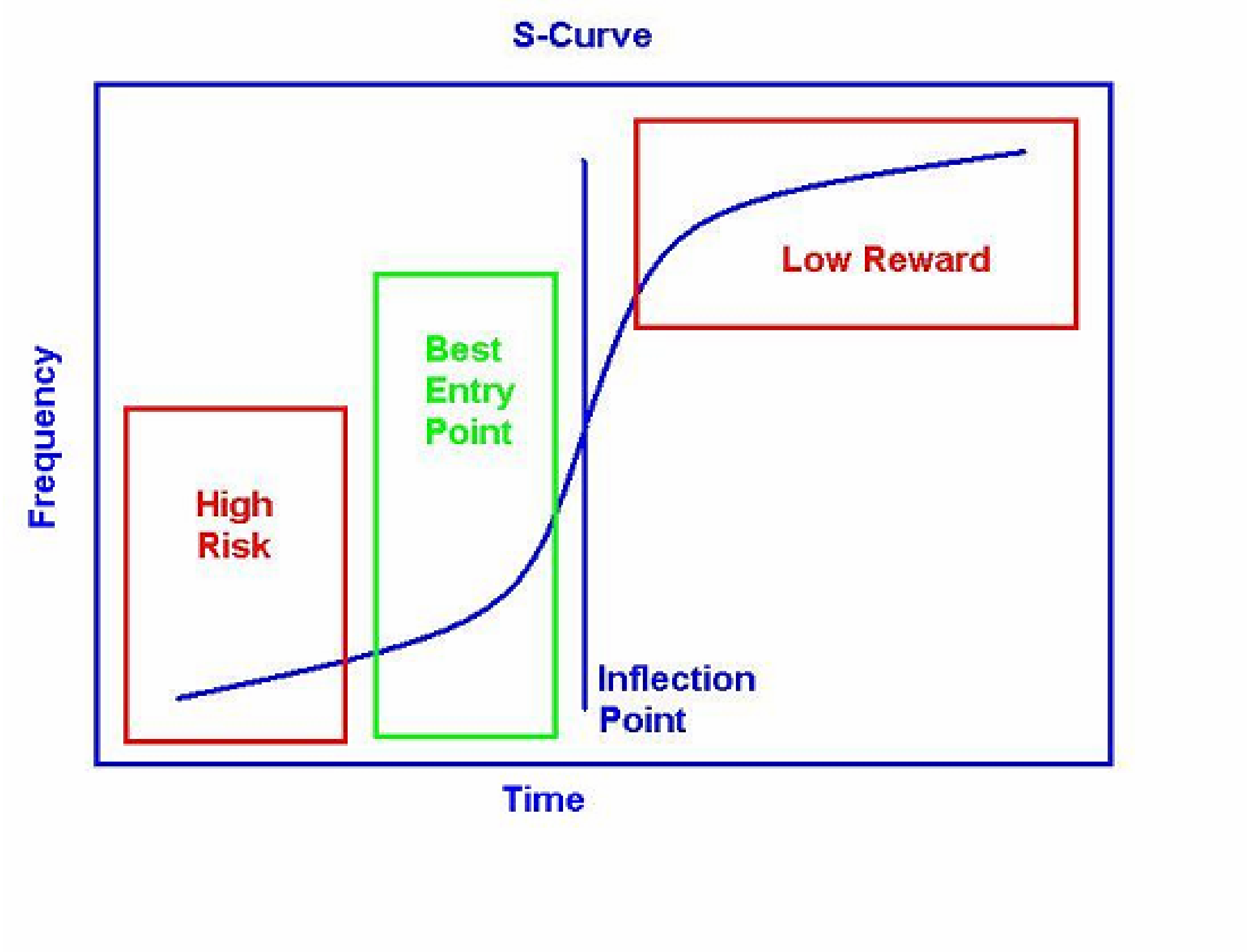
S-Curve (Math Model)



S-Curve Rate-Of-Change



S-Curve Strategy



Dawkins Meme Theory

- Dawkins coined the term “meme” in 1976
- An idea like “I want a tattoo”
- Wikipedia definition -

“a unit of cultural information that propagates from one mind to another as a theoretical unit of cultural evolution”

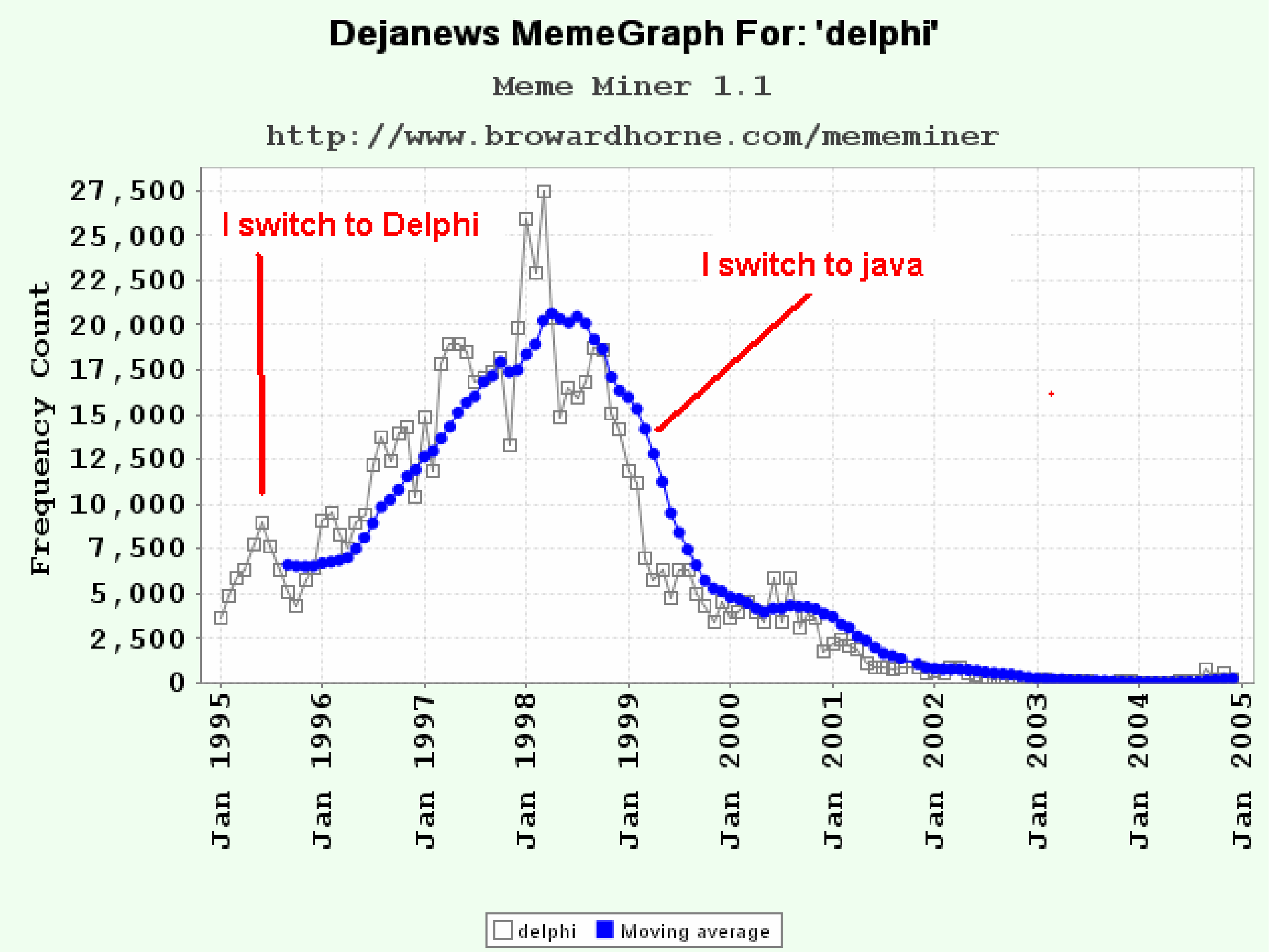
Ideosphere

- The sum of all memes in circulation
- The “Global Human Consciousness”
- The Internet has a subset of the Ideosphere

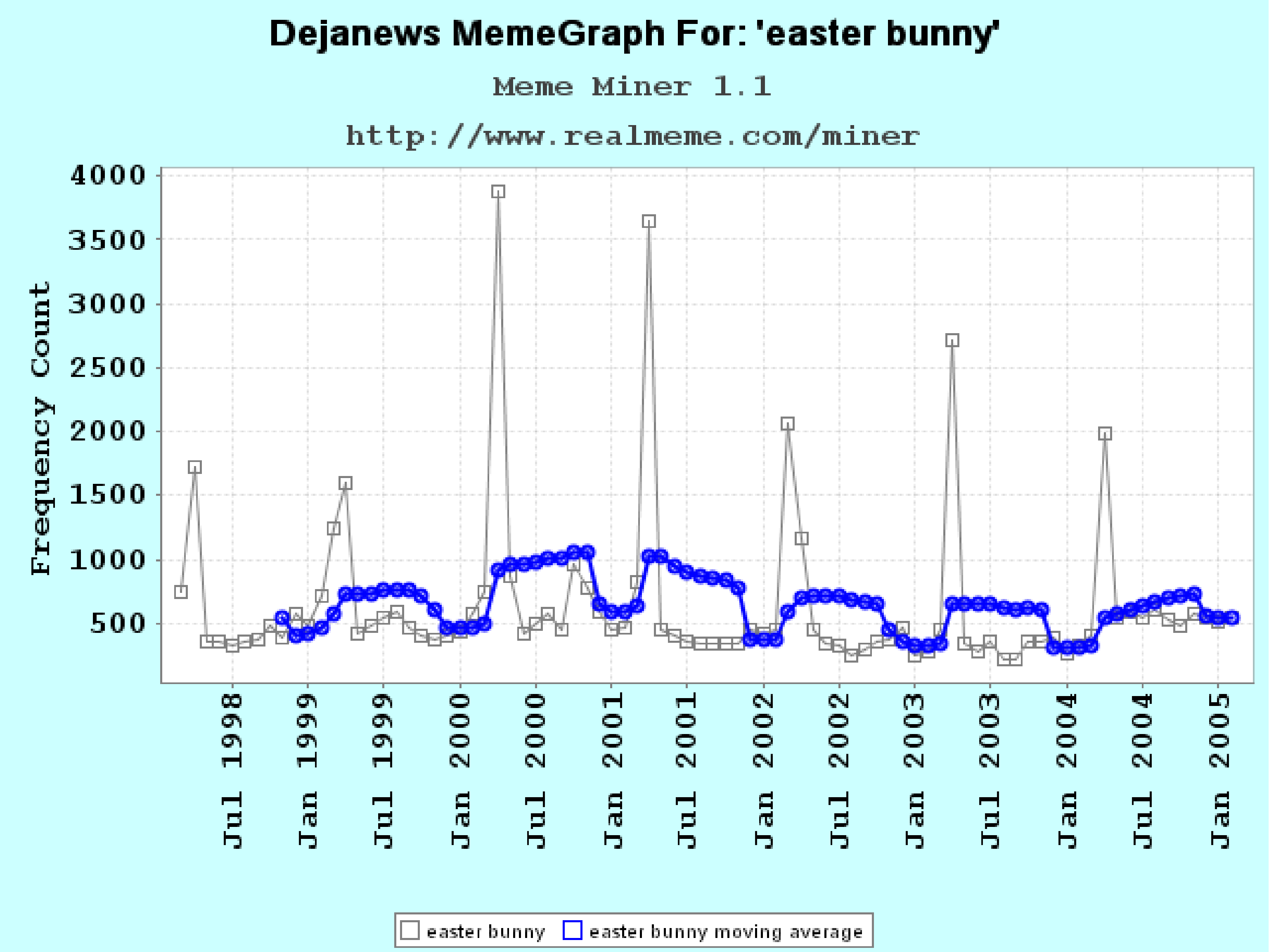
Empirical Meme Theory (Me!)

- Original Meme Theory
- Keywords As Proxy For A Meme
- Electron Flow (E = I x R) (Networks)
- S-Curve
- Meme Miner (Dejanews.com)
- Google Trends tool
- Blogpulse.com

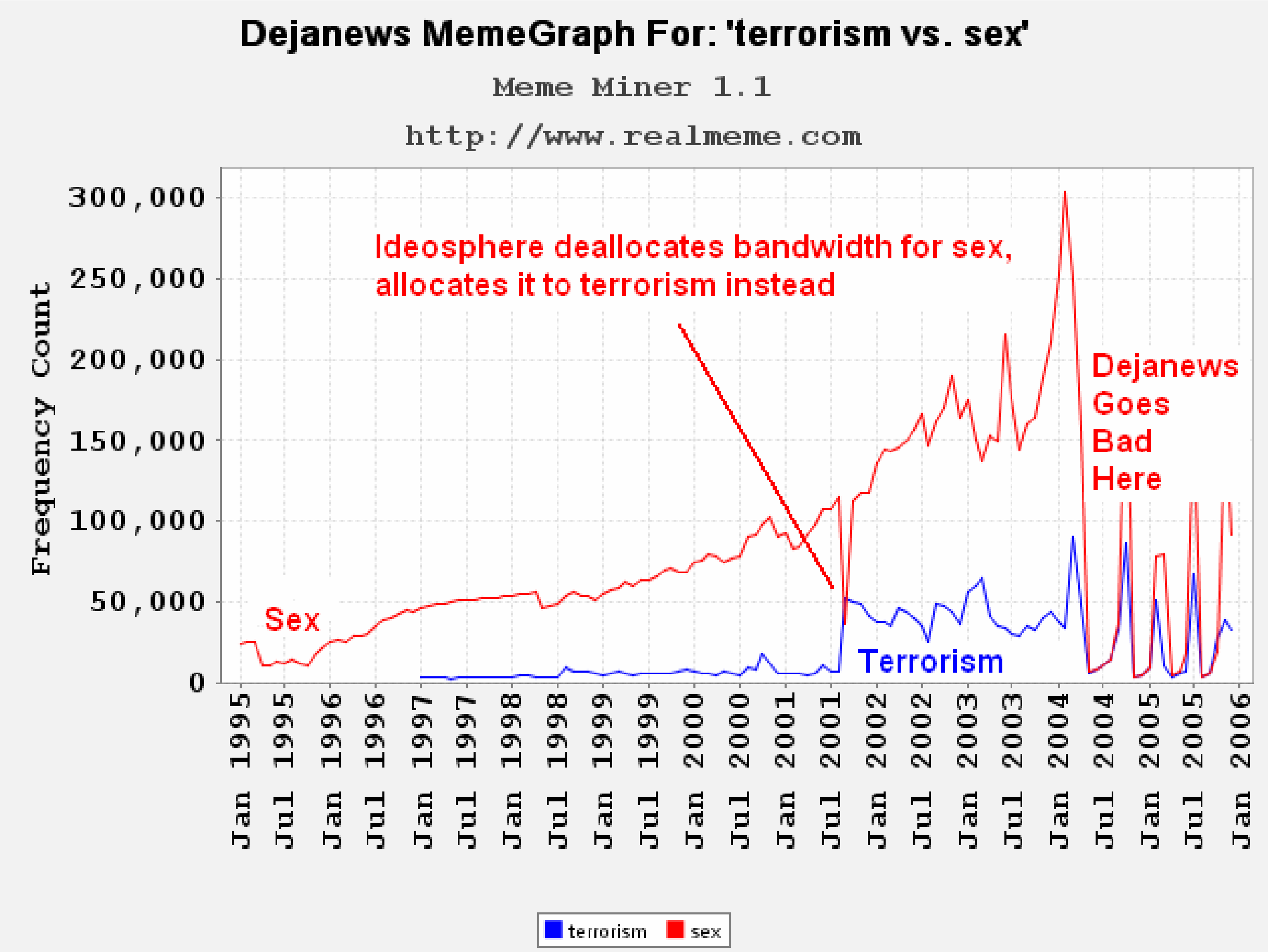
Example: Delphi



Example: Easter Bunny



Example: Sex & Terrorism



Ideosphere deallocates bandwidth for sex, allocates it to terrorism instead

Dejanews Goes Bad Here

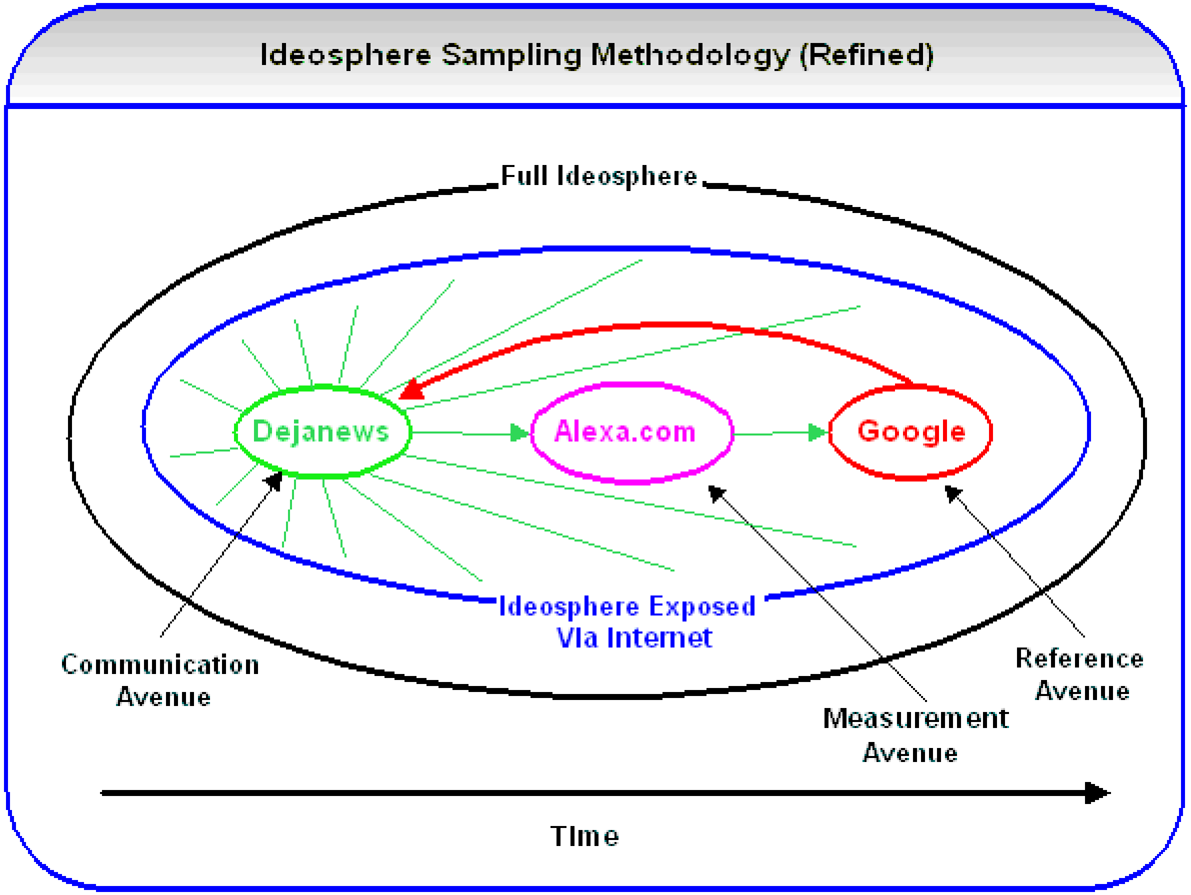
Meme Assumptions

- **Memes propagate as an S-Curve**
- **Memes propagate to most sites but at different amplitudes and latencies**

Meme Miner Inadequacies

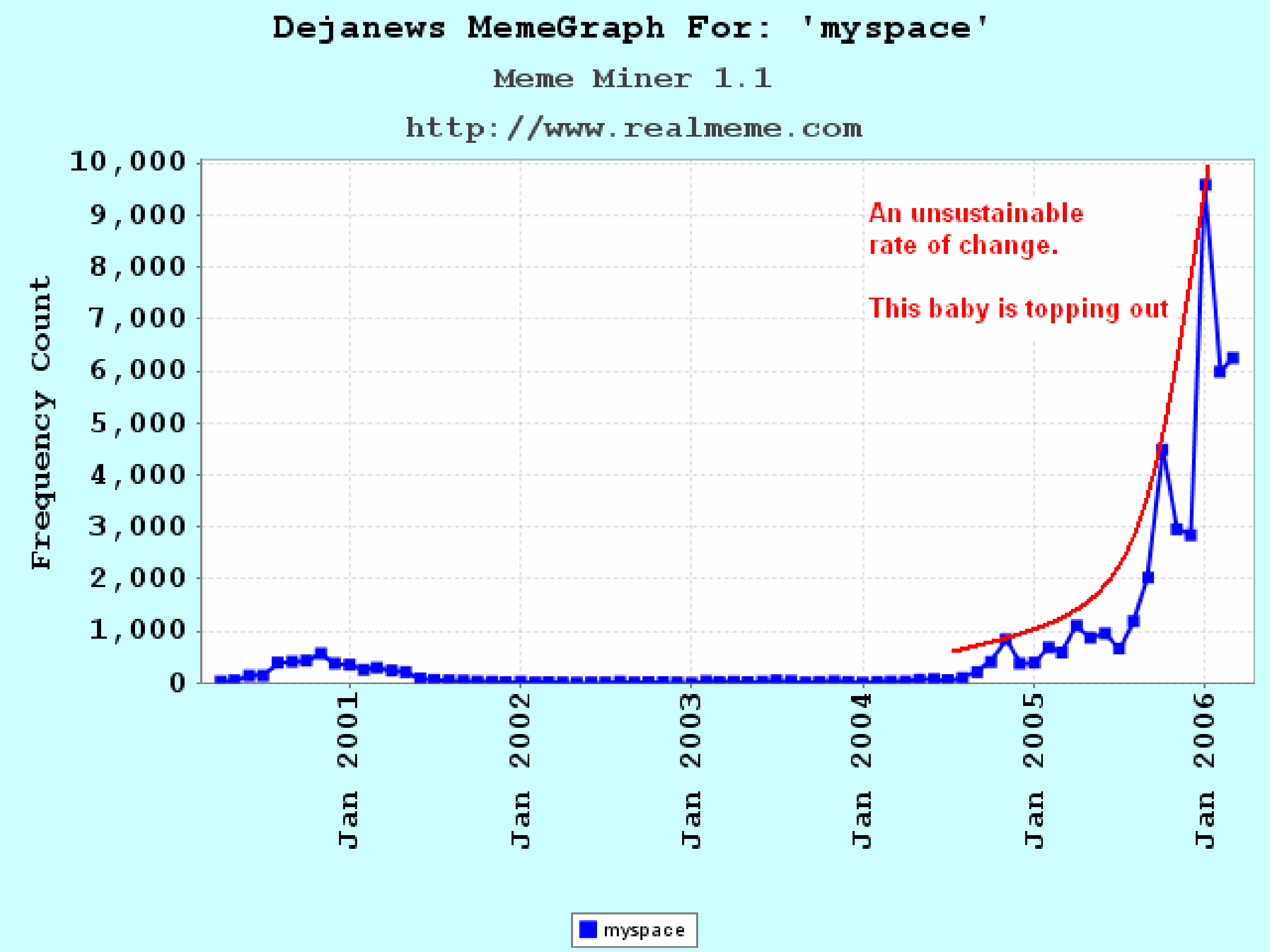
- Single source
- Dejanews.com indexing revised by Google
- Older technology losing favor

Expanded Mining Model



First tested on “MySpace” meme…

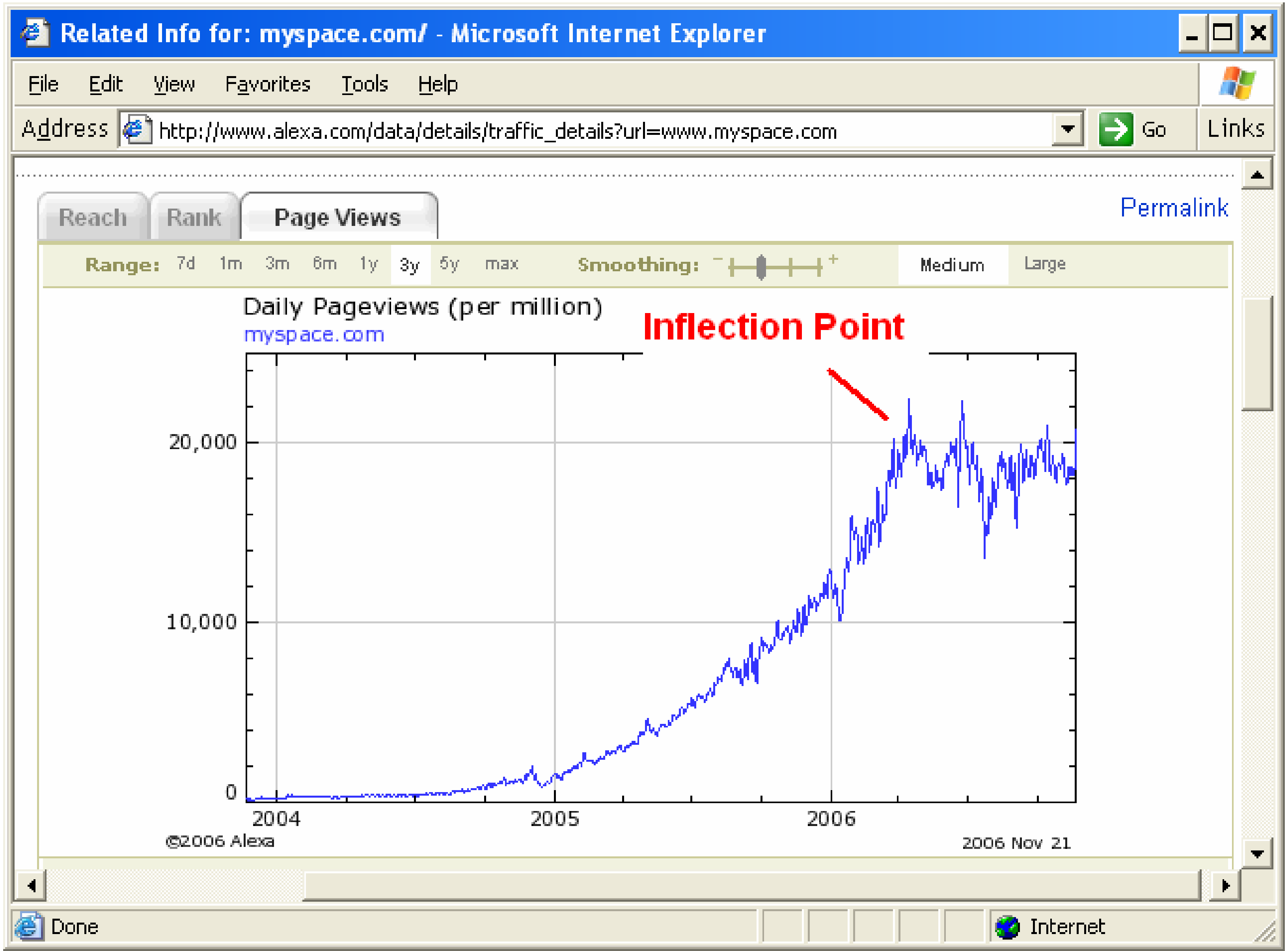
MySpace Meme (Dejanews)



An unsustainable
rate of change.

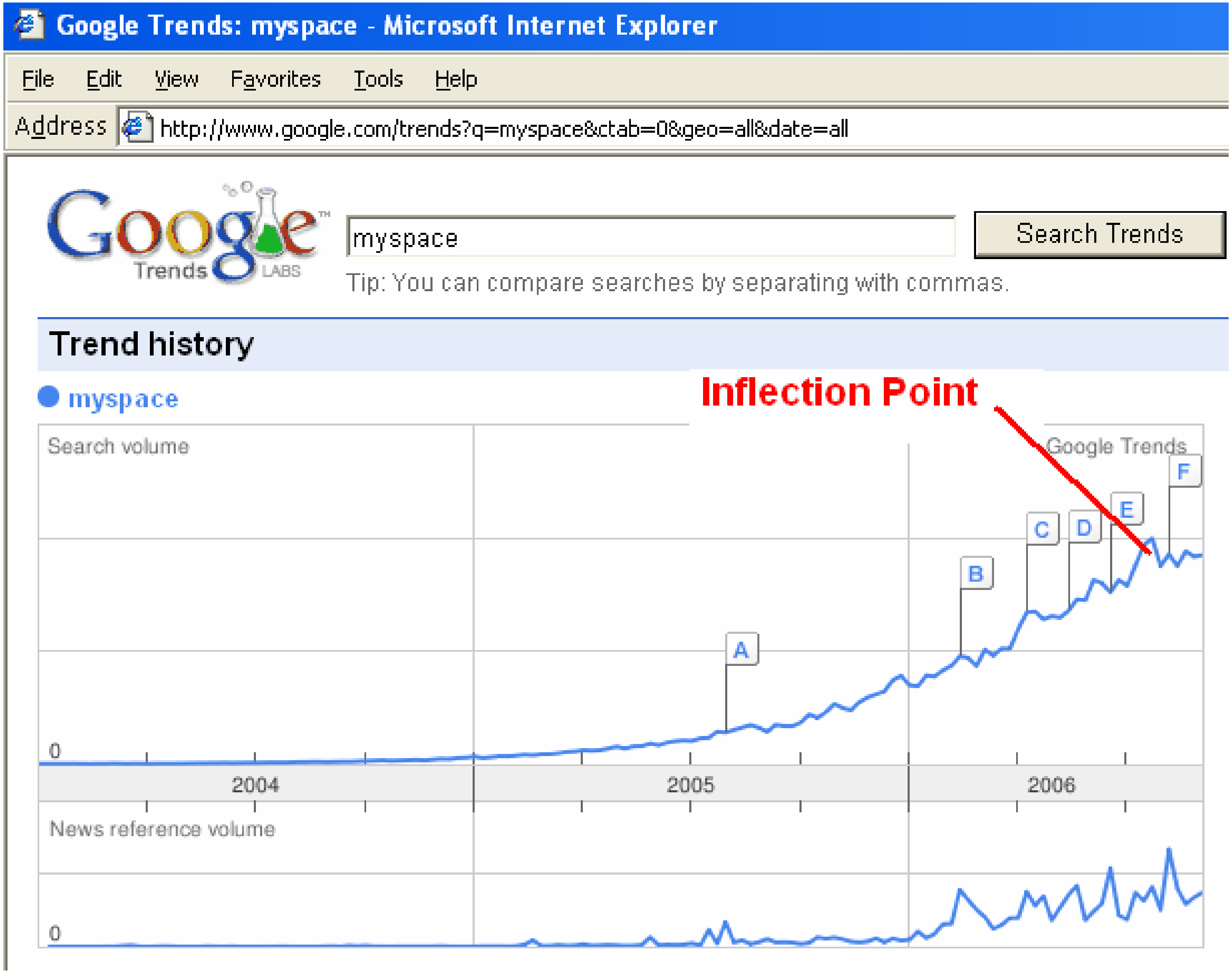
This baby is topping out

MySpace Meme (Alexa)



Inflection Point

MySpace Meme (Google)



Google Trends: myspace - Microsoft Internet Explorer

Address http://www.google.com/trends?q=myspace&ctab=0&geo=all&date=all

Inflection Point

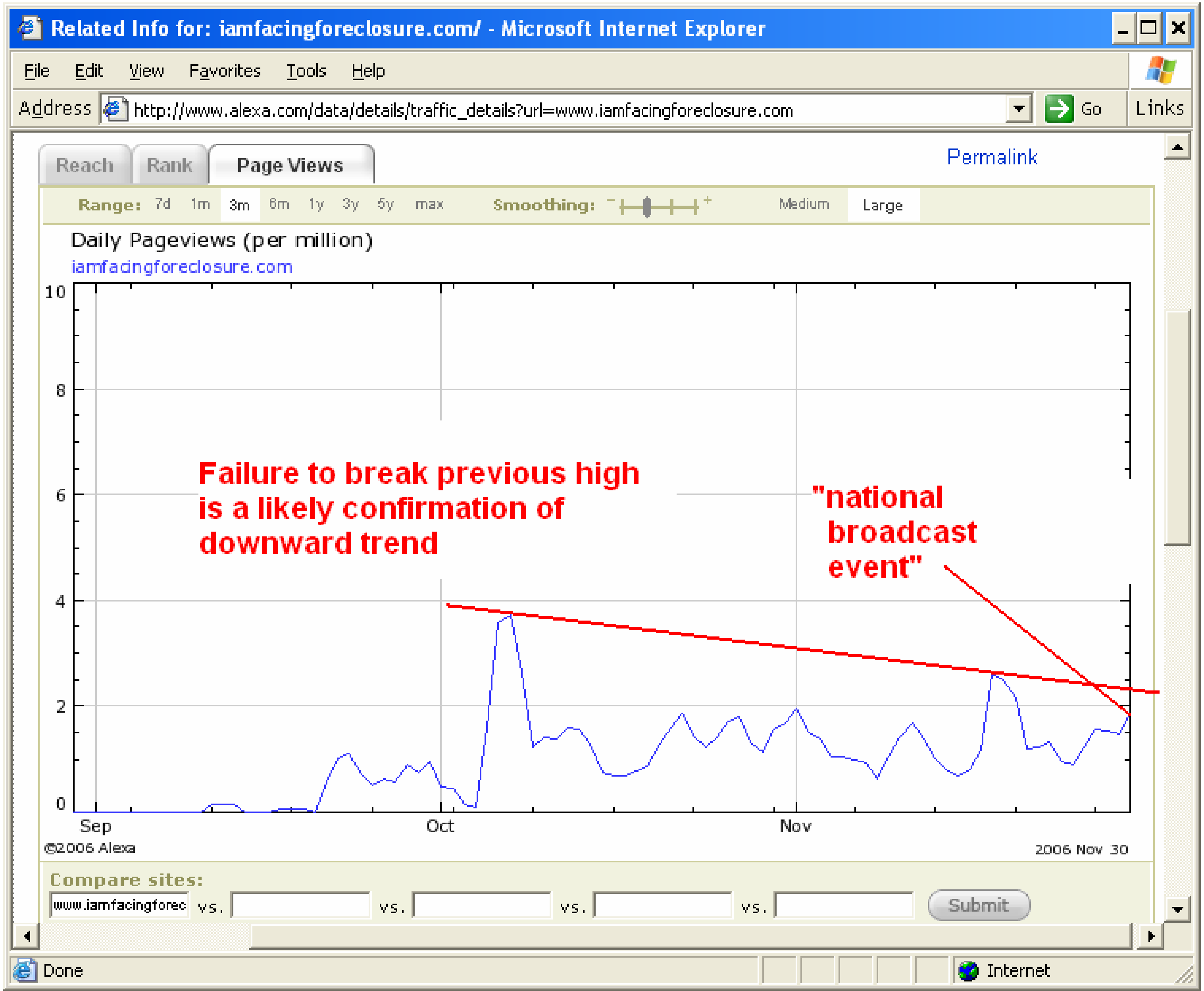
The New Model Worked…

For the “MySpace” meme, so I tested it against another case, a new site with a high growth rate…

“IAmFacingForeclosure.com” (IAFF.com)

But the results were different…

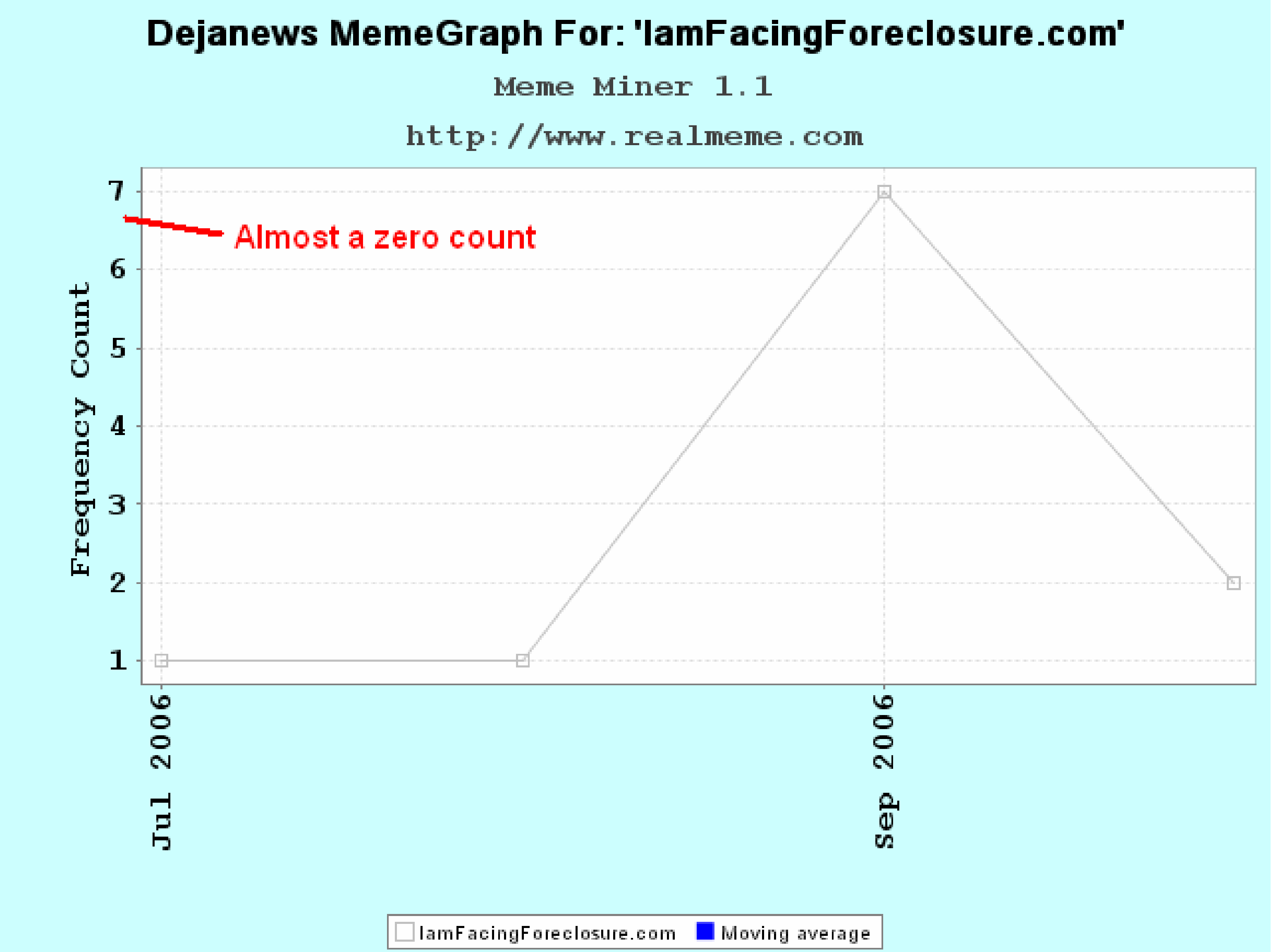
IAFF Meme (Alexa)


Related Info for: iamfacingforeclosure.com/ - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.alexa.com/data/details/traffic_details?url=www.iamfacingforeclosure.com
Go
Links
Reach
Rank
Page Views
Permalink
Range: 7d 1m 3m 6m 1y 3y 5y max
Smoothing: − +
Medium
Large
Daily Pageviews (per million)
iamfacingforeclosure.com
10
8
6
4
2
0
Sep
Oct
Nov
©2006 Alexa
2006 Nov 30
Compare sites:
www.iamfacingforec
vs.
vs.
vs.
vs.
Submit
Done
Internet

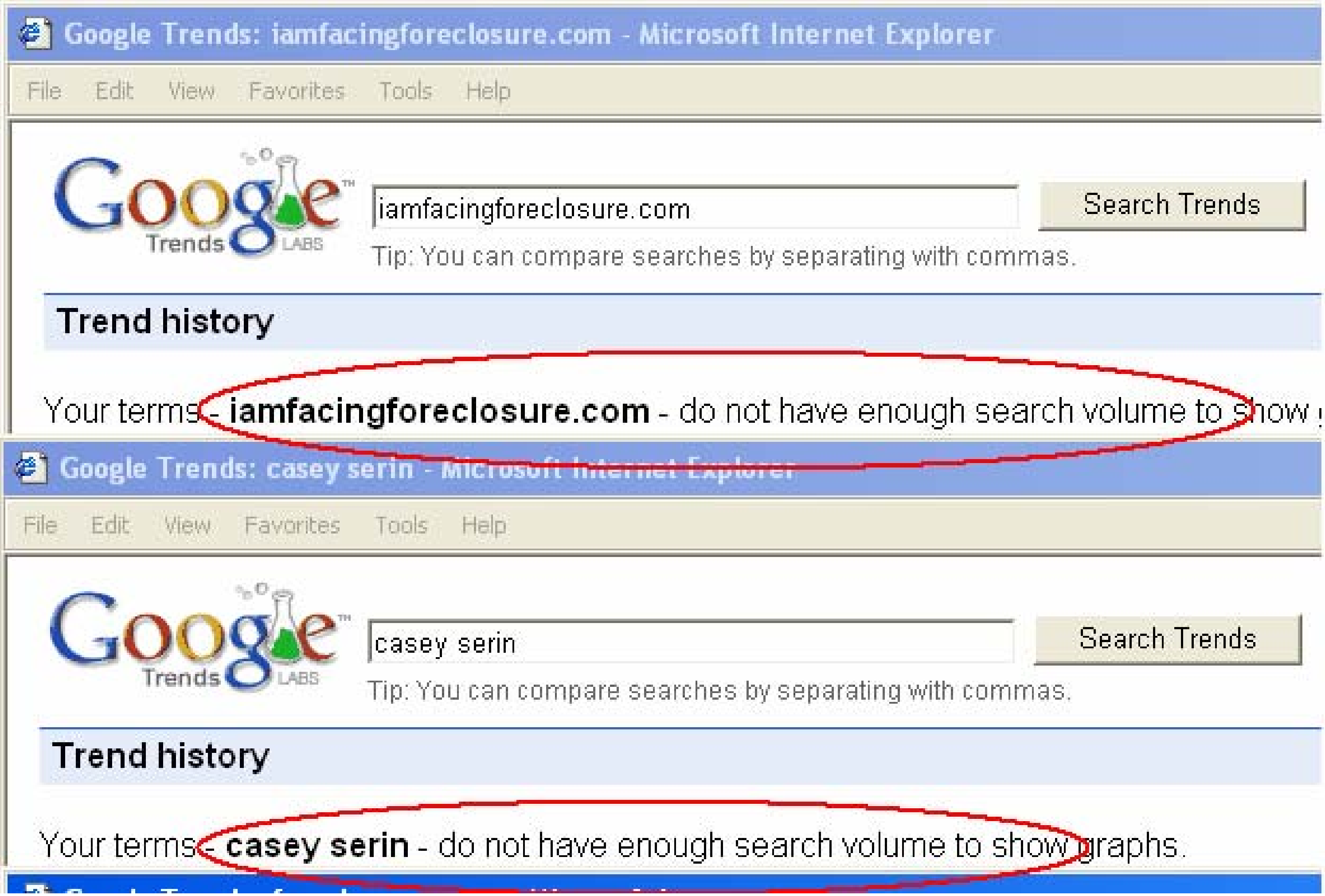
Failure to break previous high is a likely confirmation of downward trend

"national broadcast event"

IAFF Meme (Dejanews)



IAFF Meme (Google)


Google Trends: iamfacingforeclosure.com - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Google Trends LABS
Search Trends
Trend history

iamfacingforeclosure.com

Tip: You can compare searches by separating with commas.

Your terms - **iamfacingforeclosure.com** - do not have enough search volume to show

Google Trends: casey serin - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Google Trends LABS
Search Trends
Trend history

casey serin

Tip: You can compare searches by separating with commas.

Your terms - **casey serin** - do not have enough search volume to show graphs.

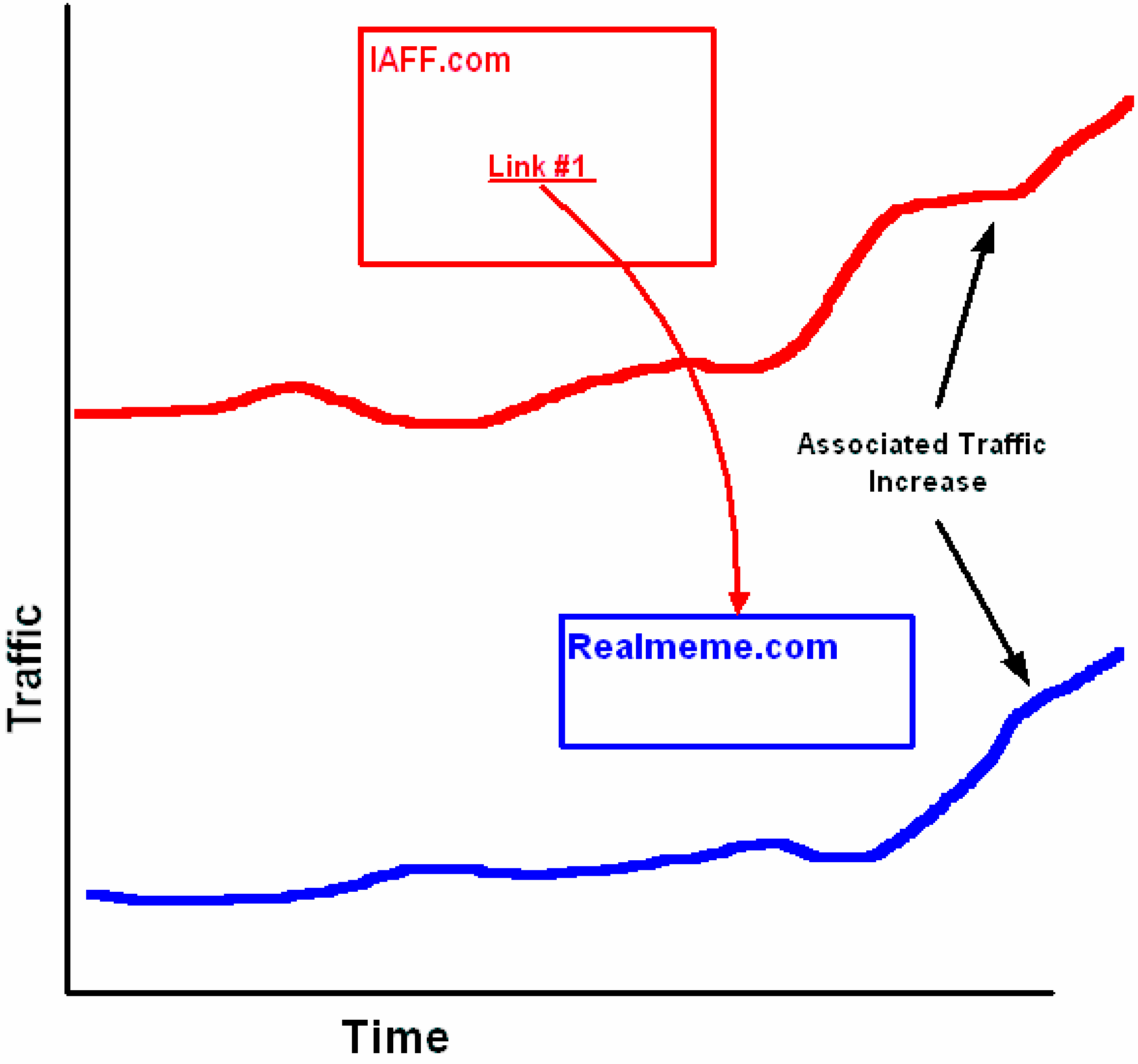
Meme Seepage Theory

Remember our Meme assumptions? Memes propagate as an S-curve and across most sites but with varying amplitudes and latencies.

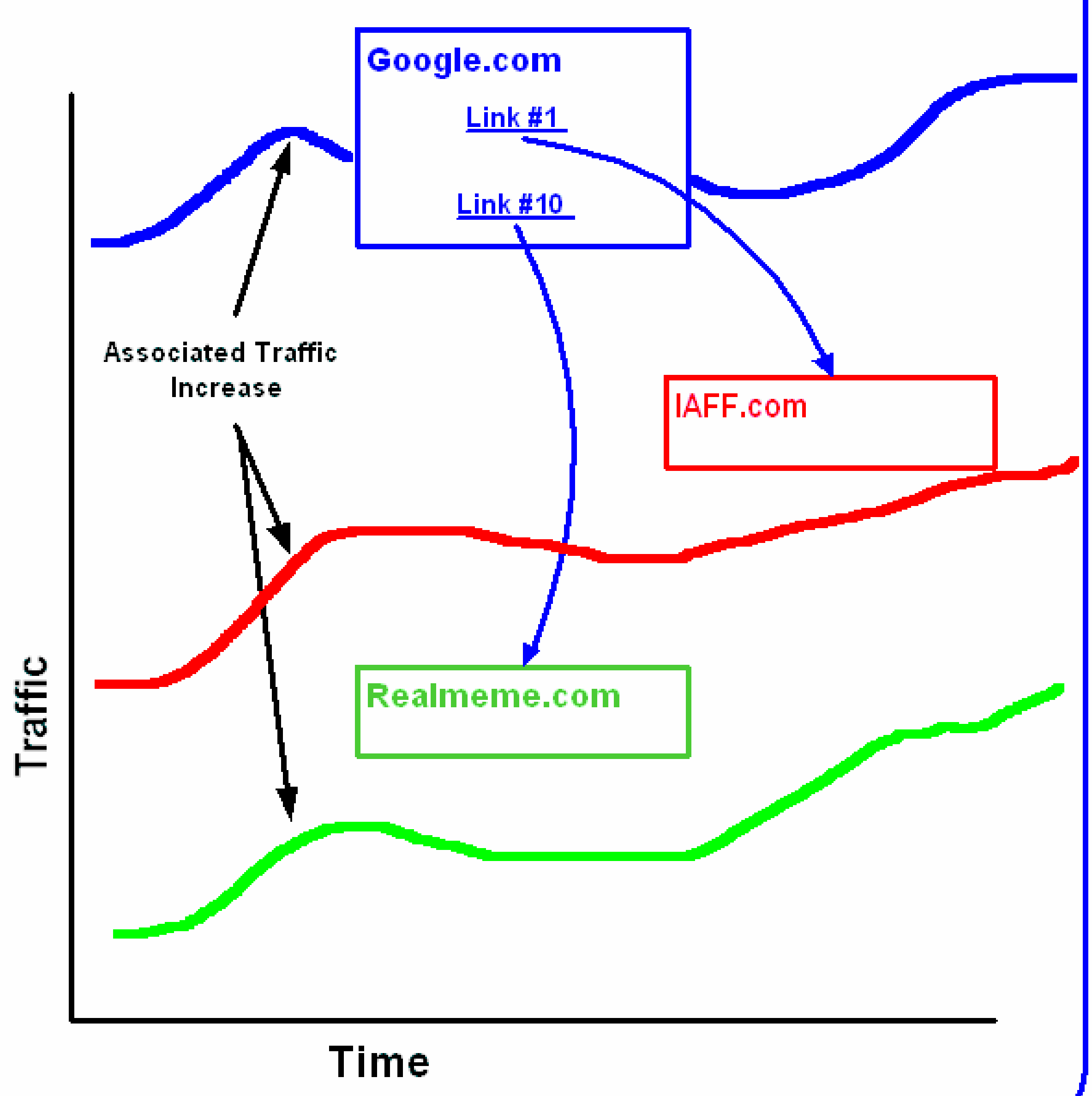
If traffic increases to a primary site, then traffic to linked sites should increase proportionally (more or less)

If traffic increases to a primary site, then traffic to reference sites like Google should increase proportionally… and other sites in the top 10 result list should experience a lesser but measurable increase in traffic…

Primary Meme Seepage



Secondary Meme Seepage



So I Experimented..

I posted a link directly to IAFF.com, to tap off a slice of IAFF.com’s traffic via meme seepage. Theoretically, a doubling of IAFF traffic should produce an equal ratio of redirected traffic to my own site, RealMeme.com

But the results were wrong again…

And I Experimented Again…

Alexa showed IAFF.com with a 25% increase in traffic but my site experienced no concurrent increase.

I was surprised and I posted the results to IAFF.com.

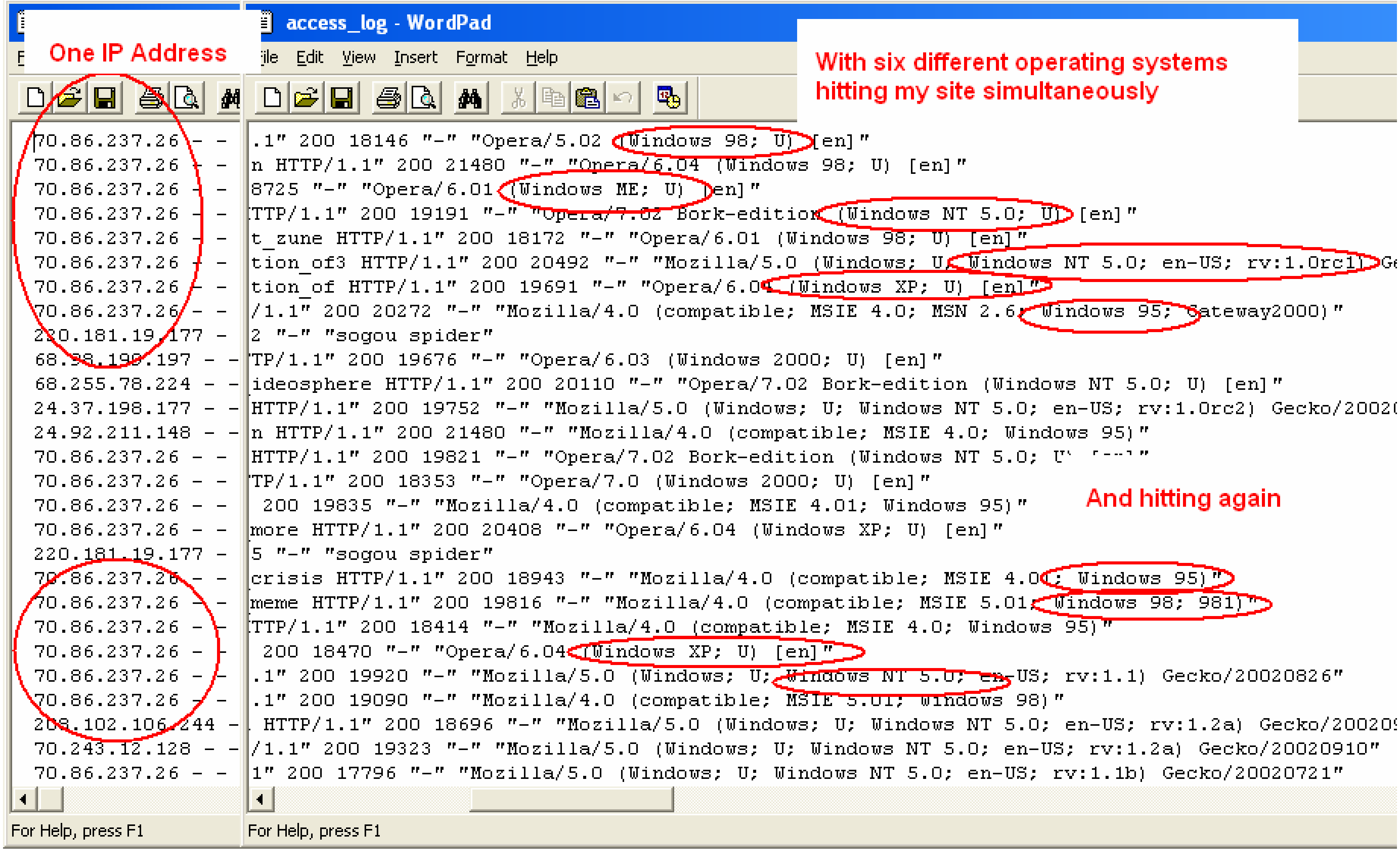
And a few days later, I did experience an anomalous increase in traffic but it didn’t match IAFF.com’s Alexa traffic delta. Here’s what hit my site…

Botnet Proof

The following page hits are logs from my website.
It’s clear that they were artifically produced…

- Too many simultaneous operating systems per IP.
- The traffic is too dense and changeover too abrupt.
- Too many 2-page hits (my traffic is 90% 1-page)
- The page hits don't follow a human click flow.
- All hits are bookmarks, no blog entry points
- Most bookmarks are older

Botnet Logs (Same IP)



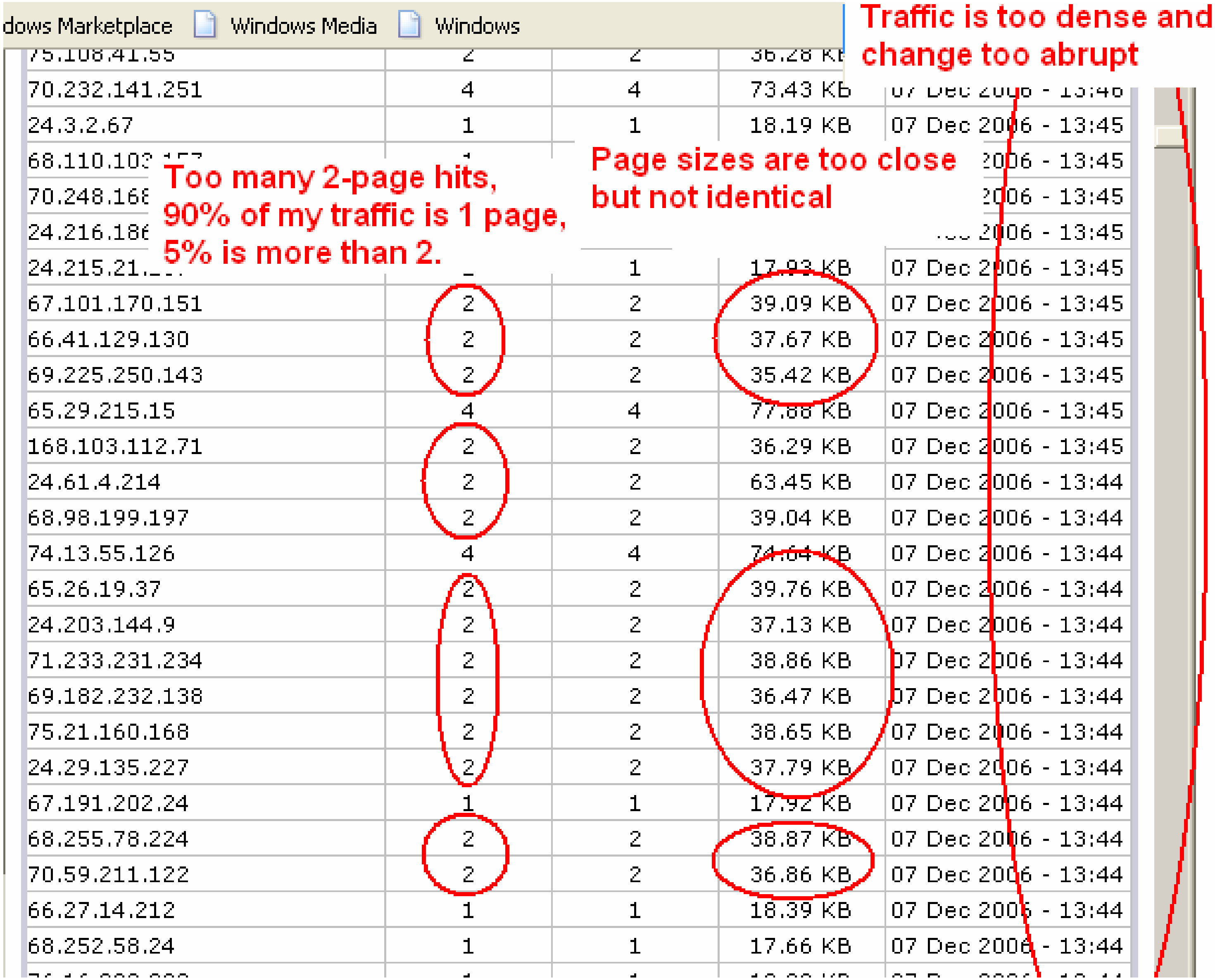
One IP Address

With six different operating systems hitting my site simultaneously

And hitting again

```
70.86.237.26 - -   .1" 200 18146 "-" "Opera/5.02 (Windows 98; U) [en]"
70.86.237.26 - -   n HTTP/1.1" 200 21480 "-" "Opera/6.04 (Windows 98; U) [en]"
70.86.237.26 - -   8725 "-" "Opera/6.01 (Windows ME; U) [en]"
70.86.237.26 - -   TTP/1.1" 200 19191 "-" "Opera/7.02 Bork-edition (Windows NT 5.0; U) [en]"
70.86.237.26 - -   t_zune HTTP/1.1" 200 18172 "-" "Opera/6.01 (Windows 98; U) [en]"
70.86.237.26 - -   tion_of3 HTTP/1.1" 200 20492 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.0rc1) Ge
70.86.237.26 - -   tion_of HTTP/1.1" 200 19691 "-" "Opera/6.04 (Windows XP; U) [en]"
70.86.237.26 - -   /1.1" 200 20272 "-" "Mozilla/4.0 (compatible; MSIE 4.0; MSN 2.6; Windows 95; Gateway2000)"
220.181.19.177 -   2 "-" "sogou spider"
68.38.190.197 - -  TP/1.1" 200 19676 "-" "Opera/6.03 (Windows 2000; U) [en]"
68.255.78.224 - -  ideosphere HTTP/1.1" 200 20110 "-" "Opera/7.02 Bork-edition (Windows NT 5.0; U) [en]"
24.37.198.177 - -  HTTP/1.1" 200 19752 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.0rc2) Gecko/2002
24.92.211.148 - -  n HTTP/1.1" 200 21480 "-" "Mozilla/4.0 (compatible; MSIE 4.0; Windows 95)"
70.86.237.26 - -   HTTP/1.1" 200 19821 "-" "Opera/7.02 Bork-edition (Windows NT 5.0; U) [en]"
70.86.237.26 - -   TP/1.1" 200 18353 "-" "Opera/7.0 (Windows 2000; U) [en]"
70.86.237.26 - -    200 19835 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows 95)"
70.86.237.26 - -   more HTTP/1.1" 200 20408 "-" "Opera/6.04 (Windows XP; U) [en]"
220.181.19.177 -   5 "-" "sogou spider"
70.86.237.26 - -   crisis HTTP/1.1" 200 18943 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows 95)"
70.86.237.26 - -   meme HTTP/1.1" 200 19816 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows 98; 981)"
70.86.237.26 - -   TTP/1.1" 200 18414 "-" "Mozilla/4.0 (compatible; MSIE 4.0; Windows 95)"
70.86.237.26 - -    200 18470 "-" "Opera/6.04 (Windows XP; U) [en]"
70.86.237.26 - -   .1" 200 19920 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.1) Gecko/20020826"
70.86.237.26 - -   .1" 200 19090 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows 98)"
208.102.106.244 -   HTTP/1.1" 200 18696 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2a) Gecko/20020
70.243.12.128 - -  /1.1" 200 19323 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2a) Gecko/20020910"
70.86.237.26 - -   1" 200 17796 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.1b) Gecko/20020721"
```

Botnet Logs (Density)



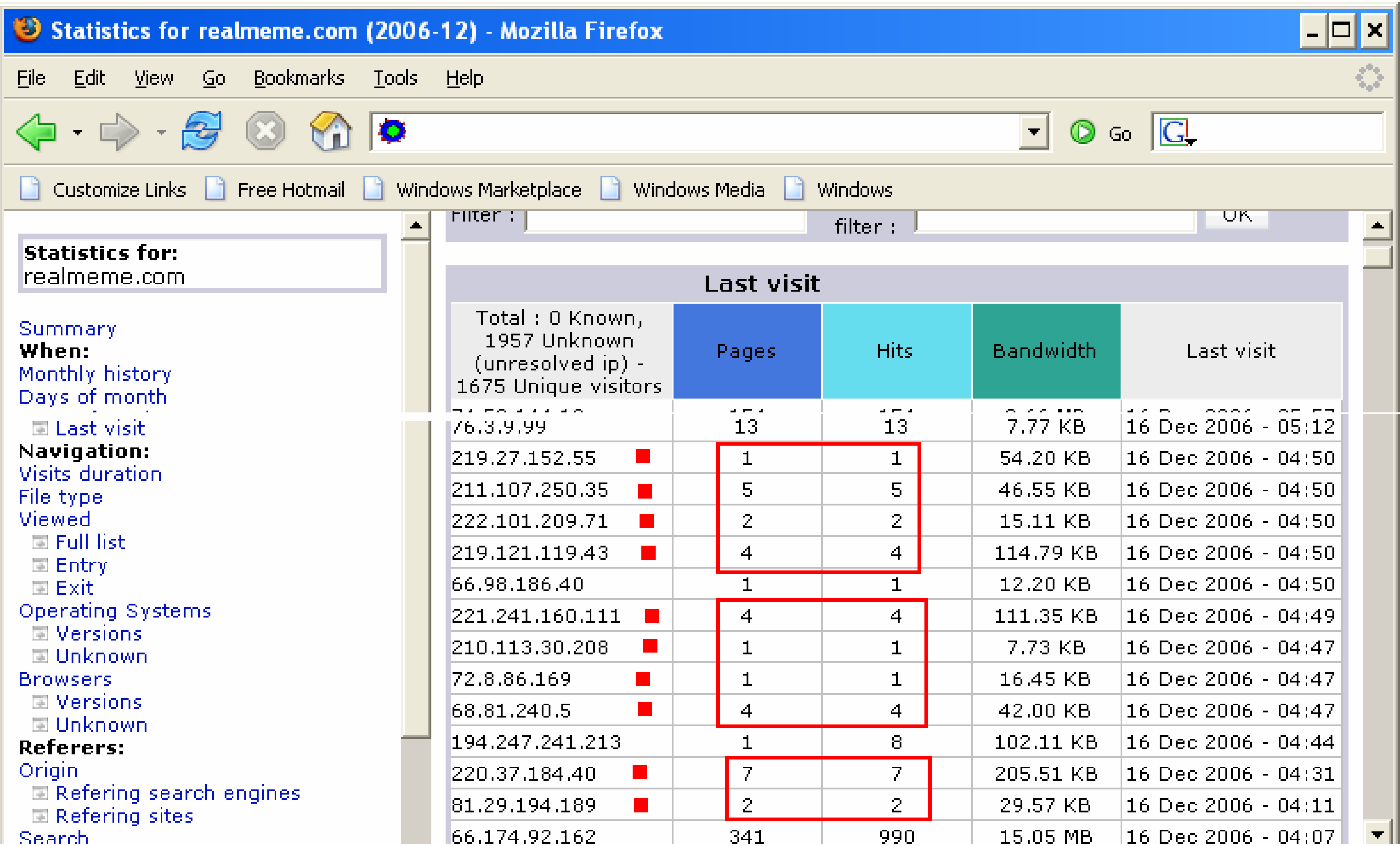
75.108.41.55	2	2	36.28 KB	[illegible]
70.232.141.251	4	4	73.43 KB	07 Dec 2006 - 13:46
24.3.2.67	1	1	18.19 KB	07 Dec 2006 - 13:45
68.110.103.157	1	[illegible]	[illegible]	2006 - 13:45
70.248.168	[illegible]	[illegible]	[illegible]	2006 - 13:45
24.216.186	[illegible]	[illegible]	[illegible]	2006 - 13:45
24.215.21	1	1	17.93 KB	07 Dec 2006 - 13:45
67.101.170.151	2	2	39.09 KB	07 Dec 2006 - 13:45
66.41.129.130	2	2	37.67 KB	07 Dec 2006 - 13:45
69.225.250.143	2	2	35.42 KB	07 Dec 2006 - 13:45
65.29.215.15	4	4	77.88 KB	07 Dec 2006 - 13:45
168.103.112.71	2	2	36.29 KB	07 Dec 2006 - 13:45
24.61.4.214	2	2	63.45 KB	07 Dec 2006 - 13:44
68.98.199.197	2	2	39.04 KB	07 Dec 2006 - 13:44
74.13.55.126	4	4	74.64 KB	07 Dec 2006 - 13:44
65.26.19.37	2	2	39.76 KB	07 Dec 2006 - 13:44
24.203.144.9	2	2	37.13 KB	07 Dec 2006 - 13:44
71.233.231.234	2	2	38.86 KB	07 Dec 2006 - 13:44
69.182.232.138	2	2	36.47 KB	07 Dec 2006 - 13:44
75.21.160.168	2	2	38.65 KB	07 Dec 2006 - 13:44
24.29.135.227	2	2	37.79 KB	07 Dec 2006 - 13:44
67.191.202.24	1	1	17.92 KB	07 Dec 2006 - 13:44
68.255.78.224	2	2	38.87 KB	07 Dec 2006 - 13:44
70.59.211.122	2	2	36.86 KB	07 Dec 2006 - 13:44
66.27.14.212	1	1	18.39 KB	07 Dec 2006 - 13:44
68.252.58.24	1	1	17.66 KB	07 Dec 2006 - 13:44

Traffic is too dense and change too abrupt

Too many 2-page hits, 90% of my traffic is 1 page, 5% is more than 2.

Page sizes are too close but not identical

Botnet Logs (Pages)



Statistics for realmeme.com (2006-12) - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
Go
Customize Links
Free Hotmail
Windows Marketplace
Windows Media
Windows
Filter :
filter :
OK

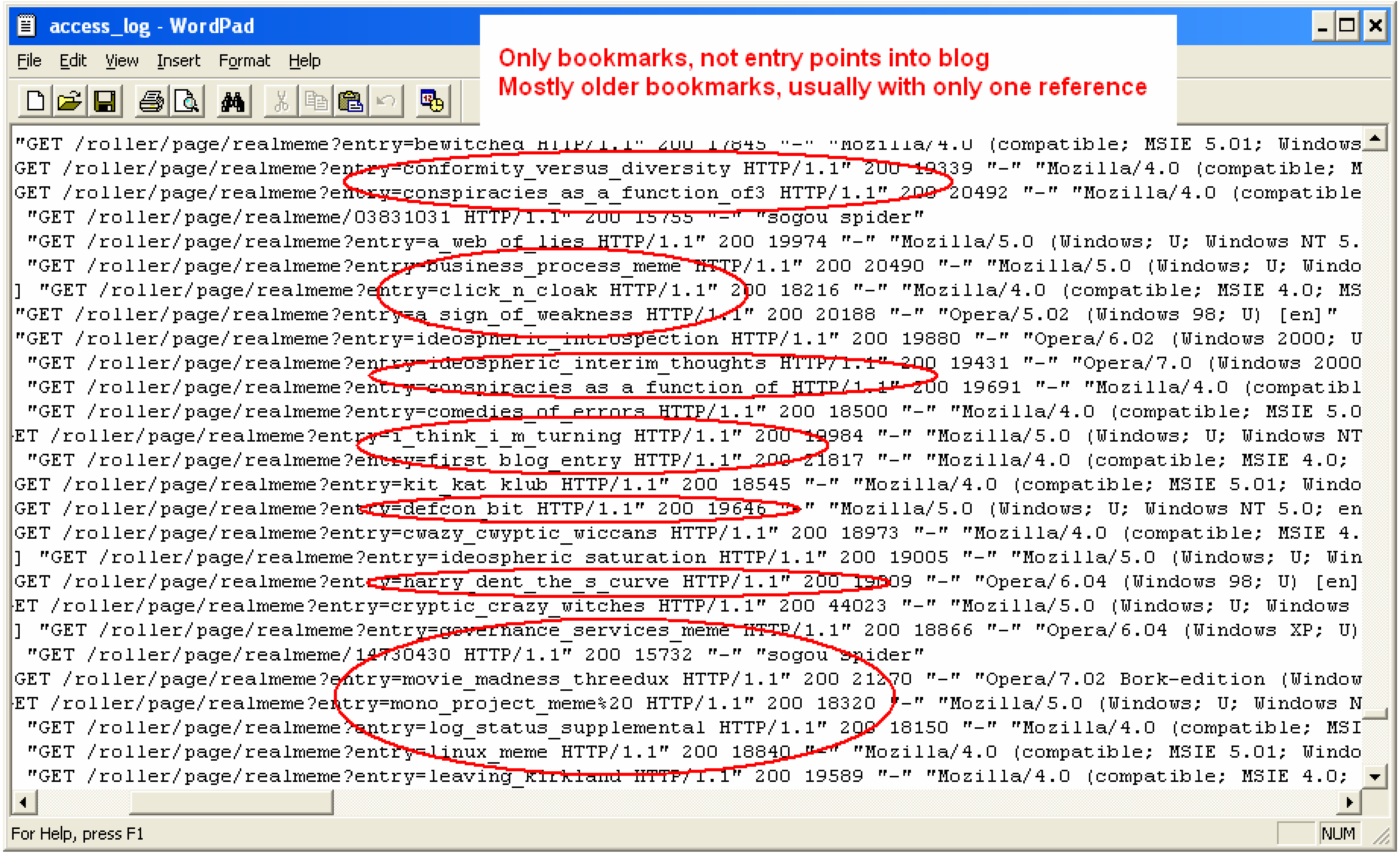
Statistics for:
realmeme.com

Summary
When:
Monthly history
Days of month
Last visit
Navigation:
Visits duration
File type
Viewed
Full list
Entry
Exit
Operating Systems
Versions
Unknown
Browsers
Versions
Unknown
Referers:
Origin
Refering search engines
Refering sites
Search

Last visit

Total : 0 Known, 1957 Unknown (unresolved ip) - 1675 Unique visitors	Pages	Hits	Bandwidth	Last visit
71.53.144.10	151	151	3.66 MB	16 Dec 2006 - 05:57
76.3.9.99	13	13	7.77 KB	16 Dec 2006 - 05:12
219.27.152.55 🟥	1	1	54.20 KB	16 Dec 2006 - 04:50
211.107.250.35 🟥	5	5	46.55 KB	16 Dec 2006 - 04:50
222.101.209.71 🟥	2	2	15.11 KB	16 Dec 2006 - 04:50
219.121.119.43 🟥	4	4	114.79 KB	16 Dec 2006 - 04:50
66.98.186.40	1	1	12.20 KB	16 Dec 2006 - 04:50
221.241.160.111 🟥	4	4	111.35 KB	16 Dec 2006 - 04:49
210.113.30.208 🟥	1	1	7.73 KB	16 Dec 2006 - 04:47
72.8.86.169 🟥	1	1	16.45 KB	16 Dec 2006 - 04:47
68.81.240.5 🟥	4	4	42.00 KB	16 Dec 2006 - 04:47
194.247.241.213	1	8	102.11 KB	16 Dec 2006 - 04:44
220.37.184.40 🟥	7	7	205.51 KB	16 Dec 2006 - 04:31
81.29.194.189 🟥	2	2	29.57 KB	16 Dec 2006 - 04:11
66.174.92.162	341	990	15.05 MB	16 Dec 2006 - 04:07

Botnet Logs (Entry Points)



Only bookmarks, not entry points into blog
Mostly older bookmarks, usually with only one reference

```
"GET /roller/page/realmeme?entry=bewitched HTTP/1.1" 200 17845 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows
GET /roller/page/realmeme?entry=conformity_versus_diversity HTTP/1.1" 200 19339 "-" "Mozilla/4.0 (compatible; M
GET /roller/page/realmeme?entry=conspiracies_as_a_function_of3 HTTP/1.1" 200 20492 "-" "Mozilla/4.0 (compatible
 "GET /roller/page/realmeme/03831031 HTTP/1.1" 200 15755 "-" "sogou spider"
 "GET /roller/page/realmeme?entry=a_web_of_lies HTTP/1.1" 200 19974 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.
 "GET /roller/page/realmeme?entry=business_process_meme HTTP/1.1" 200 20490 "-" "Mozilla/5.0 (Windows; U; Windo
] "GET /roller/page/realmeme?entry=click_n_cloak HTTP/1.1" 200 18216 "-" "Mozilla/4.0 (compatible; MSIE 4.0; MS
"GET /roller/page/realmeme?entry=a_sign_of_weakness HTTP/1.1" 200 20188 "-" "Opera/5.02 (Windows 98; U) [en]"
"GET /roller/page/realmeme?entry=ideospheric_introspection HTTP/1.1" 200 19880 "-" "Opera/6.02 (Windows 2000; U
 "GET /roller/page/realmeme?entry=ideospheric_interim_thoughts HTTP/1.1" 200 19431 "-" "Opera/7.0 (Windows 2000
 "GET /roller/page/realmeme?entry=conspiracies_as_a_function_of HTTP/1.1" 200 19691 "-" "Mozilla/4.0 (compatibl
 "GET /roller/page/realmeme?entry=comedies_of_errors HTTP/1.1" 200 18500 "-" "Mozilla/4.0 (compatible; MSIE 5.0
ET /roller/page/realmeme?entry=i_think_i_m_turning HTTP/1.1" 200 19984 "-" "Mozilla/5.0 (Windows; U; Windows NT
 "GET /roller/page/realmeme?entry=first_blog_entry HTTP/1.1" 200 21817 "-" "Mozilla/4.0 (compatible; MSIE 4.0;
GET /roller/page/realmeme?entry=kit_kat_klub HTTP/1.1" 200 18545 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windo
GET /roller/page/realmeme?entry=defcon_bit HTTP/1.1" 200 19646 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en
GET /roller/page/realmeme?entry=cwazy_cwyptic_wiccans HTTP/1.1" 200 18973 "-" "Mozilla/4.0 (compatible; MSIE 4.
] "GET /roller/page/realmeme?entry=ideospheric_saturation HTTP/1.1" 200 19005 "-" "Mozilla/5.0 (Windows; U; Win
GET /roller/page/realmeme?entry=harry_dent_the_s_curve HTTP/1.1" 200 19909 "-" "Opera/6.04 (Windows 98; U) [en]
ET /roller/page/realmeme?entry=cryptic_crazy_witches HTTP/1.1" 200 44023 "-" "Mozilla/5.0 (Windows; U; Windows
] "GET /roller/page/realmeme?entry=governance_services_meme HTTP/1.1" 200 18866 "-" "Opera/6.04 (Windows XP; U)
 "GET /roller/page/realmeme/14730430 HTTP/1.1" 200 15732 "-" "sogou spider"
GET /roller/page/realmeme?entry=movie_madness_threedux HTTP/1.1" 200 21270 "-" "Opera/7.02 Bork-edition (Window
ET /roller/page/realmeme?entry=mono_project_meme%20 HTTP/1.1" 200 18320 "-" "Mozilla/5.0 (Windows; U; Windows N
 "GET /roller/page/realmeme?entry=log_status_supplemental HTTP/1.1" 200 18150 "-" "Mozilla/4.0 (compatible; MSI
 "GET /roller/page/realmeme?entry=linux_meme HTTP/1.1" 200 18840 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windo
 "GET /roller/page/realmeme?entry=leaving_kirkland HTTP/1.1" 200 19589 "-" "Mozilla/4.0 (compatible; MSIE 4.0;
```

Botnet Epiphany

Okay, I’m not the smartest guy in the world
but I eventually figured out that this new traffic
was generated by bots.

But why?

So I tried another experiment…

Secondary Seepage Failure

I have a confession. My website was designed specifically for Google rankings and it’s been surprisingly successful (my Defcon 16 presentation! Ho!) So I decided…

to induce a secondary seepage from IAFF.com to my site via Google…

Binding

I bound my site to IAFF.com by posting an IAFF.com analysis which was indexed by Google. At one point, I was the #7 Google result for

“IAmFacingForeclosure.com”

So now I’m getting a slice of traffic directly from IAFF.com AND from Google’s search results for IAFF.com

Binding Results

Once again, I saw major anomalies between IAFF.com’s claimed traffic and the induced seepage to my site during “a major television event”.

Can I prove fraud? No.

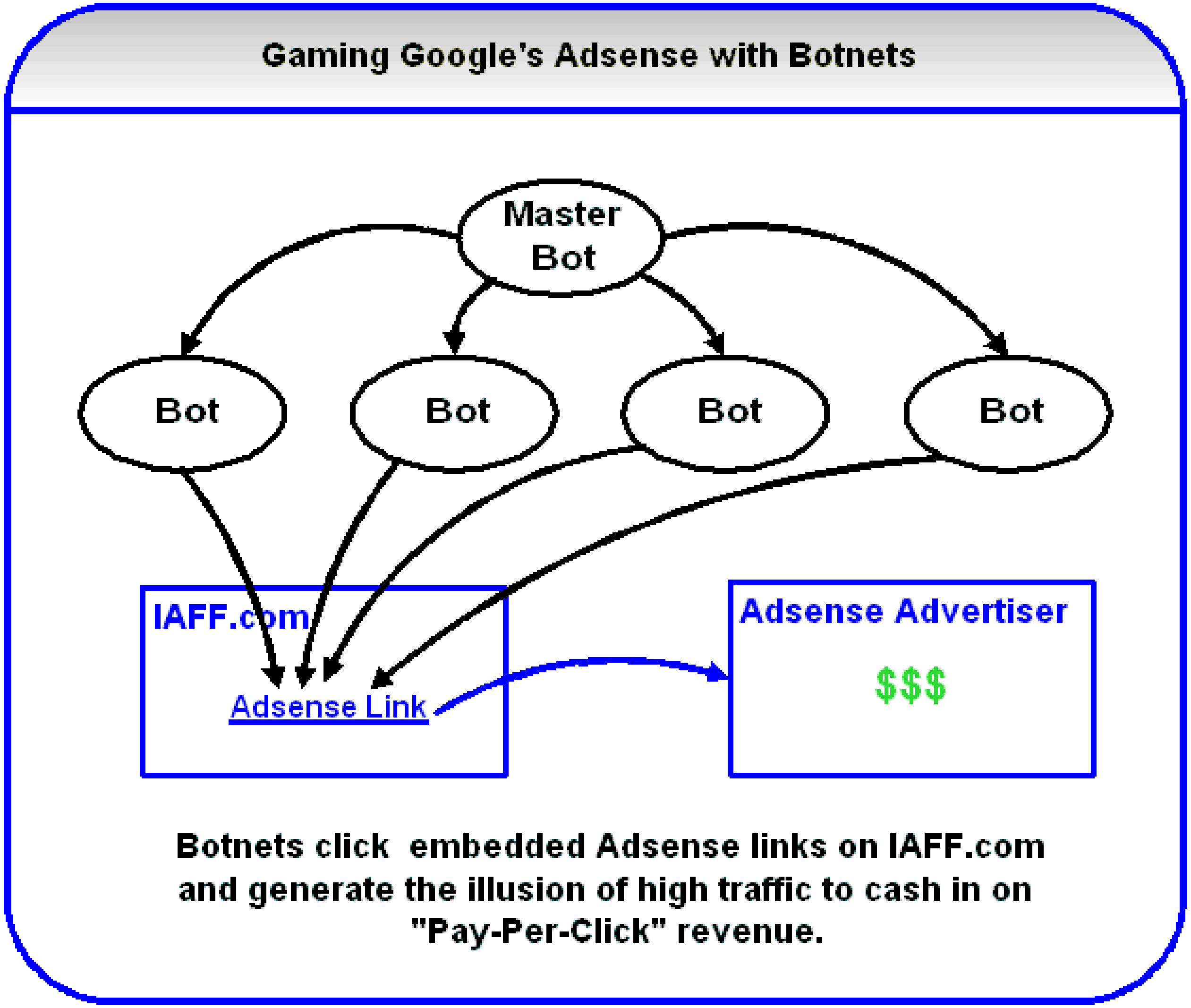
But I don’t need to. I’m not Google or a Google advertiser.

My Theory

The botnets are mimicking meme seepage by generating traffic to linked secondary sites.

At first, I couldn’t figure out why. But as I worked out a methodology to expose botnet manipulation, I realized that LACK OF SEEPAGE is a major red flag. After all, that’s how I found these anomalies to begin with.

Gaming Google



Gaming Google's Adsense with Botnets

Master Bot

Bot

Bot

Bot

Bot

IAFF.com

Adsense Link

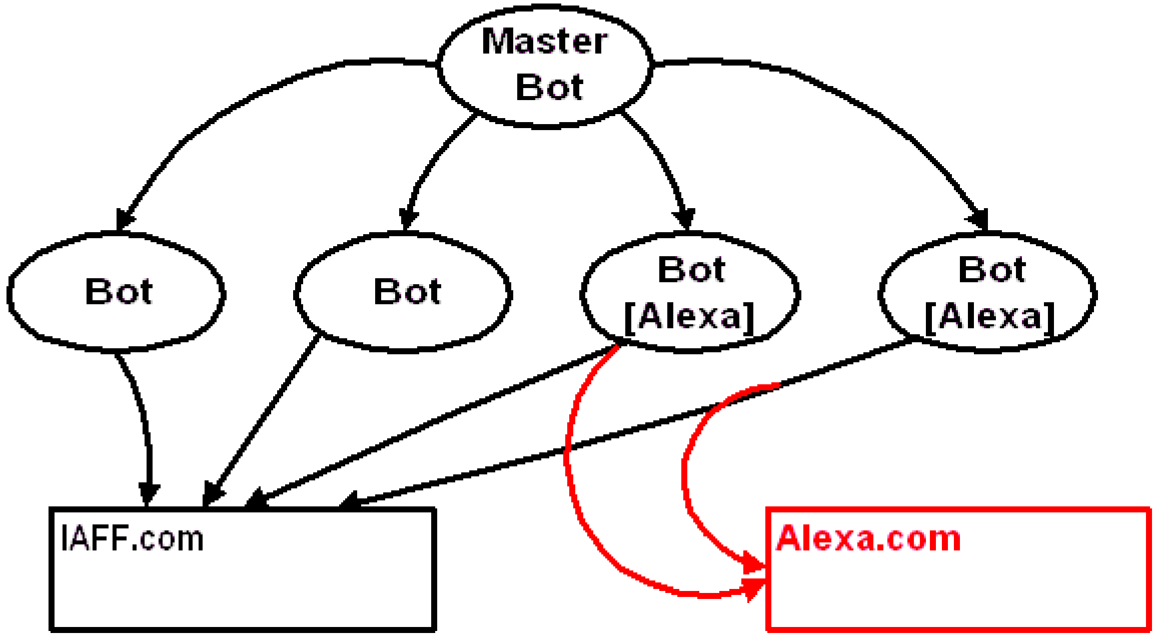
Adsense Advertiser

$$$

Botnets click embedded Adsense links on IAFF.com and generate the illusion of high traffic to cash in on "Pay-Per-Click" revenue.

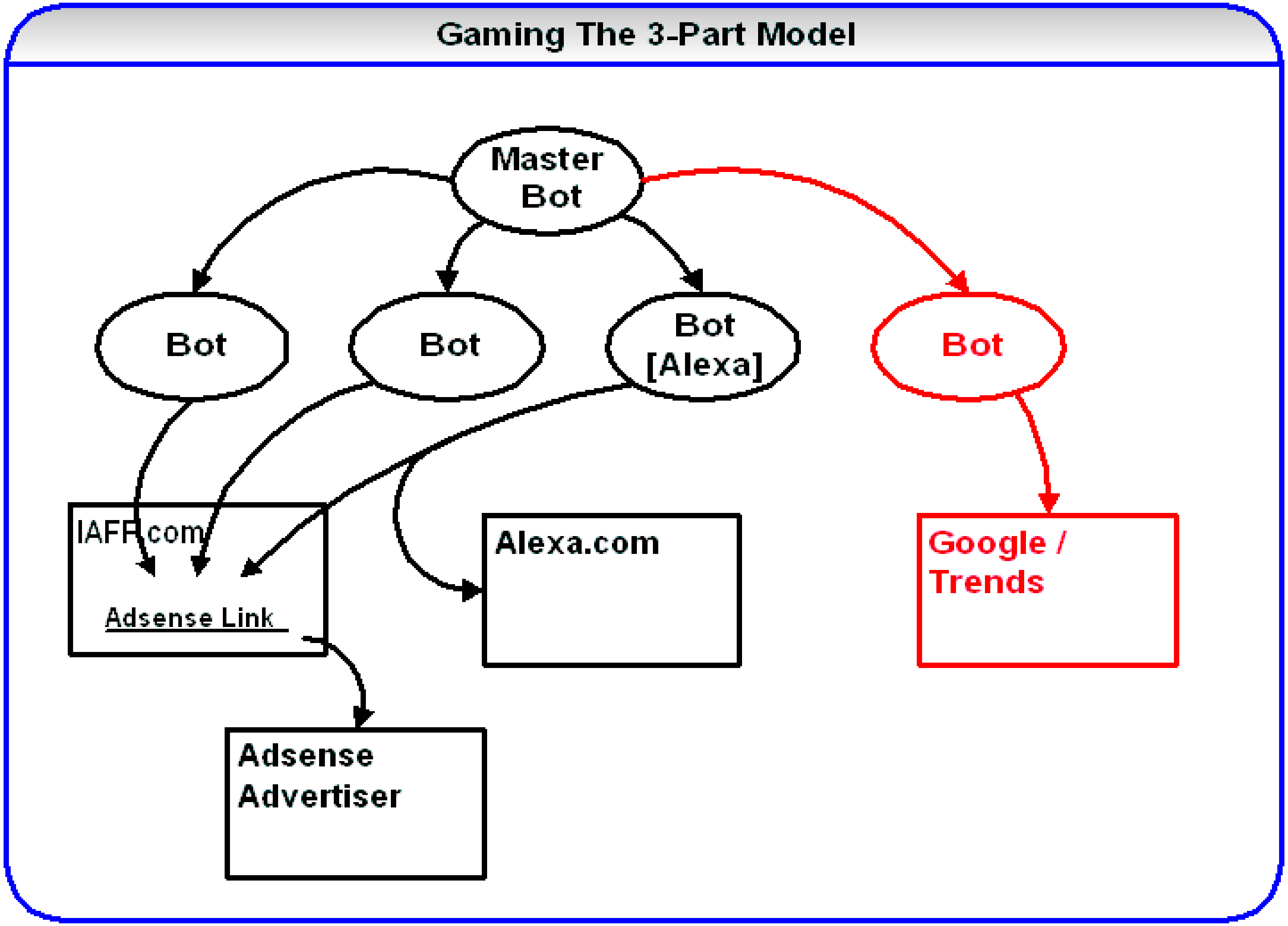
Gaming Alexa

Gaming Alexa.com with Botnets

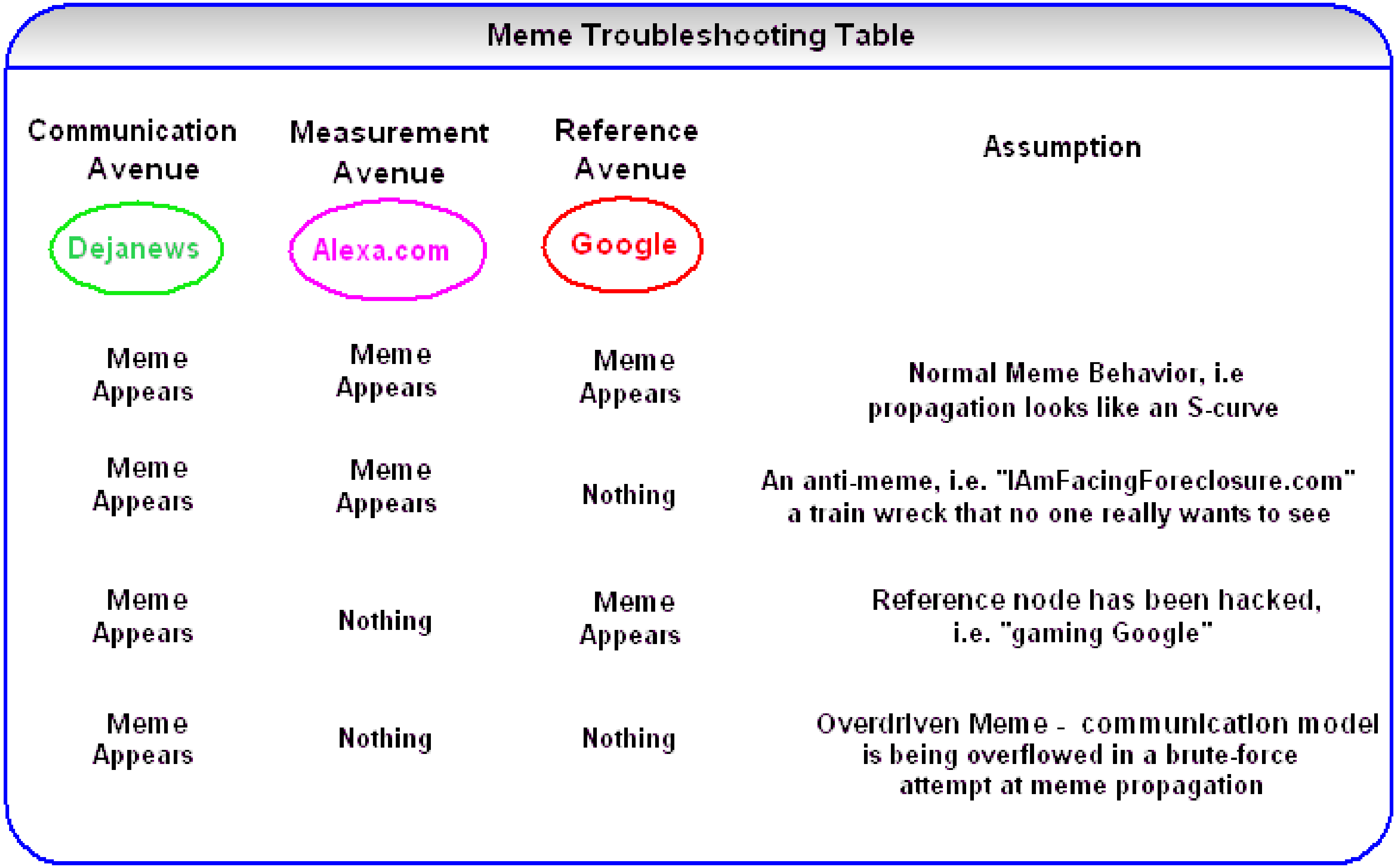


A certain percentage of hijacked Bot systems have Alexa toolbars. These systems generate an illusion of increased IAFF.com traffic to Alexa.com

Gaming My Miner Model



Meme Troubleshooting Table



Meme Troubleshooting Table

Communication Avenue	Measurement Avenue	Reference Avenue	Assumption
Dejanews	Alexa.com	Google	
Meme Appears	Meme Appears	Meme Appears	Normal Meme Behavior, i.e propagation looks like an S-curve
Meme Appears	Meme Appears	Nothing	An anti-meme, i.e. "IAmFacingForeclosure.com" a train wreck that no one really wants to see
Meme Appears	Nothing	Meme Appears	Reference node has been hacked, i.e. "gaming Google"
Meme Appears	Nothing	Nothing	Overdriven Meme - communication model is being overflowed in a brute-force attempt at meme propagation

Humans versus Bots

Needed: a pervasive, immutable quality which is detectable in humans but which bots can never duplicate.

What is it?

Humans actually buy advertised products.

Conclusion

- Click fraud is more pervasive than reported
- It can be detected with memetic analysis…
- Botnets are a serious problem
- It will eventually become almost impossible to detect sophisticated bots.
- What will Google do?