**Telcordia™
Technologies**

Performance from Experience

What is Actually Taking Place on Web Sites: Lessons from Web Server Logs

Mark Rosenstein

mbr@research.telcordia.com

An SAIC Company

Overview

- Web Server Logs Are a Valuable Resource
  - Tell you about your web site
  - Tell you about your business
- Web Server Logs Are Tricky to Analyze
  - Even without explicit session information - still valuable
  - Even with explicit session information - caching still makes it tricky!
- Be Very, Very, Very Suspicious of Reported Site Information
- Case Study
  - Focus on HTTP and HTML interactions
  - Additional Case Studies, especially related to caching, in Paper
- Discussion

Telcordia.
Technologies
Performance from Experience
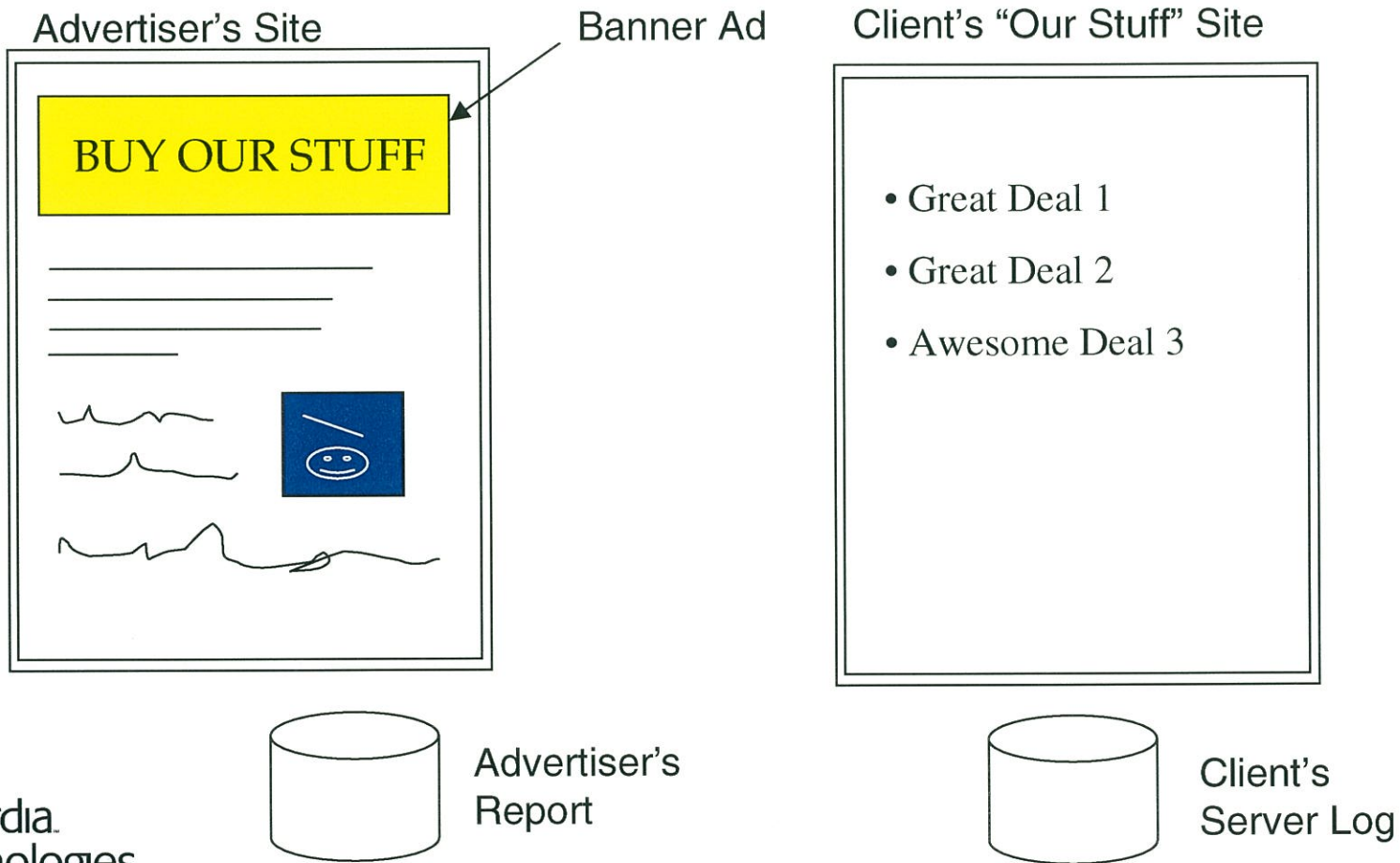
Useful Info About Your Web Site

- Example of Evaluating Design

- 3 images: 394, 524, 16,478 bytes
- A priori - all should be downloaded the same number of times
- Reality - large image only downloaded about 2/3 of others
  - Click ahead
  - Stop button

Telcordia. Technologies

Performance from Experience

Useful Info About Your Business

- Example of Expanding Customer Base

- 3 clicks down in hierarchy - customers ask what's required
- Nearly 50% of customers using definition of form item help

Telcordia.
Technologies

Performance from Experience

Tricky, Tricky, Tricky

- Banner Click-throughs - only 54% showed up in Server Logs

Advertiser's Site   Banner Ad   Client's "Our Stuff" Site

BUY OUR STUFF

- Great Deal 1
- Great Deal 2
- Awesome Deal 3

Advertiser's Report

Client's Server Log

Reported Aggregate Site Information

- Scully slide here
- Comment on unpeopled world

- worried about what commercial web analysis tools report - No Ground truth
- Ketsup bottle vs. Harry Potter
- LA Museum
  - missing sit
  - 3gut earn
  - turn out to be housy
  - point isn't software available - point is no one looking

TRUST NO ONE

AGENT SCULLY

Deep Throat
"Erlenmeyer Flask"

Contents of Web Server Log (One "Hit")

Originating IP:	198.81.129.99
Timestamp:	[12/Oct/2000:05:28:50 -0500]
HTTP Command & Protocol Version:	"GET / HTTP/1.0"
Status Code:	200
Bytes Transferred:	7717
Browser:	"Mozilla/4.51 [en] (Win95; I)"
Referer URI:	"http://www.your-website-here.com/"

Telcordia. Technologies

Performance from Experience

Case Study: Where Did All the Customers Go?

- Decrease in Business at E-Commerce Site
- Majority of Customers are Referred from Main Site
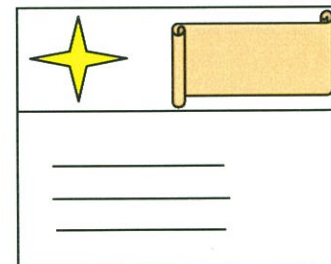- Huge Decrease in Refers - from Referer Log Entry

Main Site

E-Commerce Site

Great Buy 1

Gotta Have 2

Great Stuff 3

Buy Now!
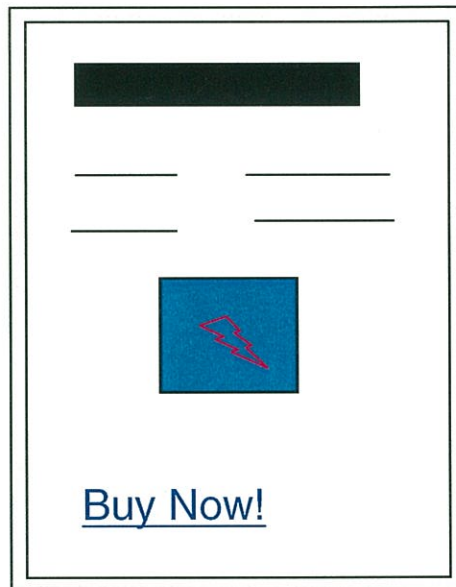
moderate sized site
- megabites not Gigabyes
- 10'sof k hits/day not millions

Referer Information - The <img> case
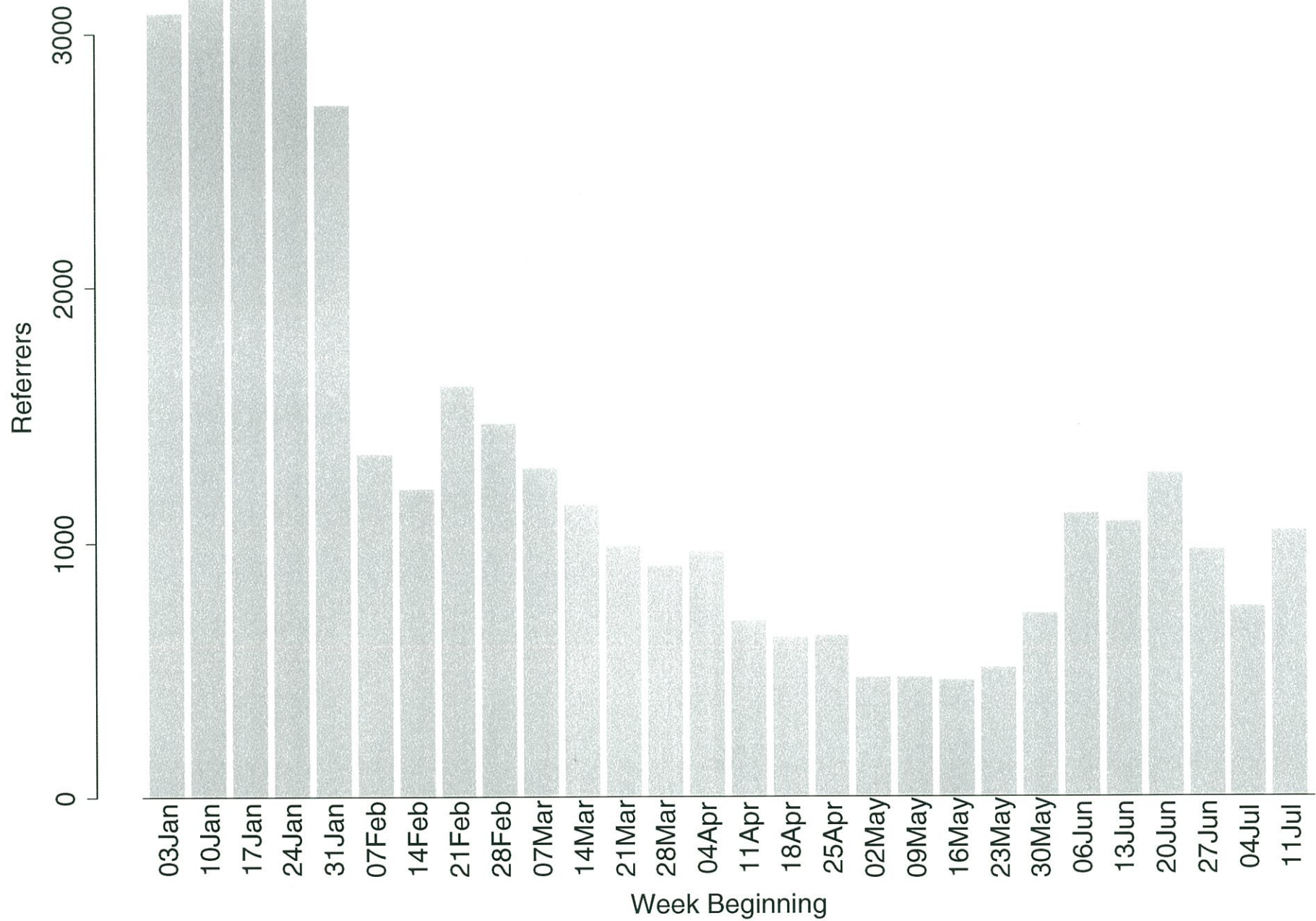
Main Site

E- Commerce Site

...

<H2>A Title</H2>

<img src="/images/star.jpg">

<img src="/images/scroll.gif">

...

- **What Does this imply for Server Log?**

Good thing

Telcordia.
Technologies
Performance from Experience

Referrers by Week

Referer Info - Server Redirect

- You Type: http://www.amazon.com
- You See:
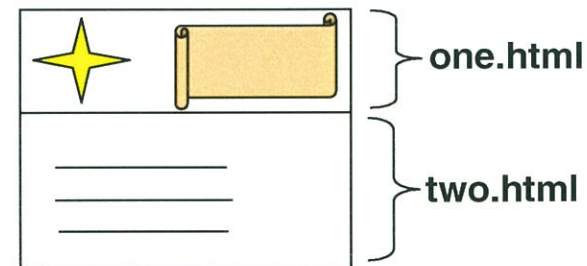 http://www.amazon.com/exec/obidos/subst/home/home.html/10
  3-6946079-9063016
- What Happened
  - Server responded to request with 302 and new location
  - Browser requested page from new location
- What Does this imply for Server Log?

Telcordia
Technologies

Performance from Experience

Referer Info - Frames

Main Site

E- Commerce Site

one.html

two.html

Buy Now

```
<frameset>
<frame src="one.html">
<frame src="two.html">
</frameset>
```
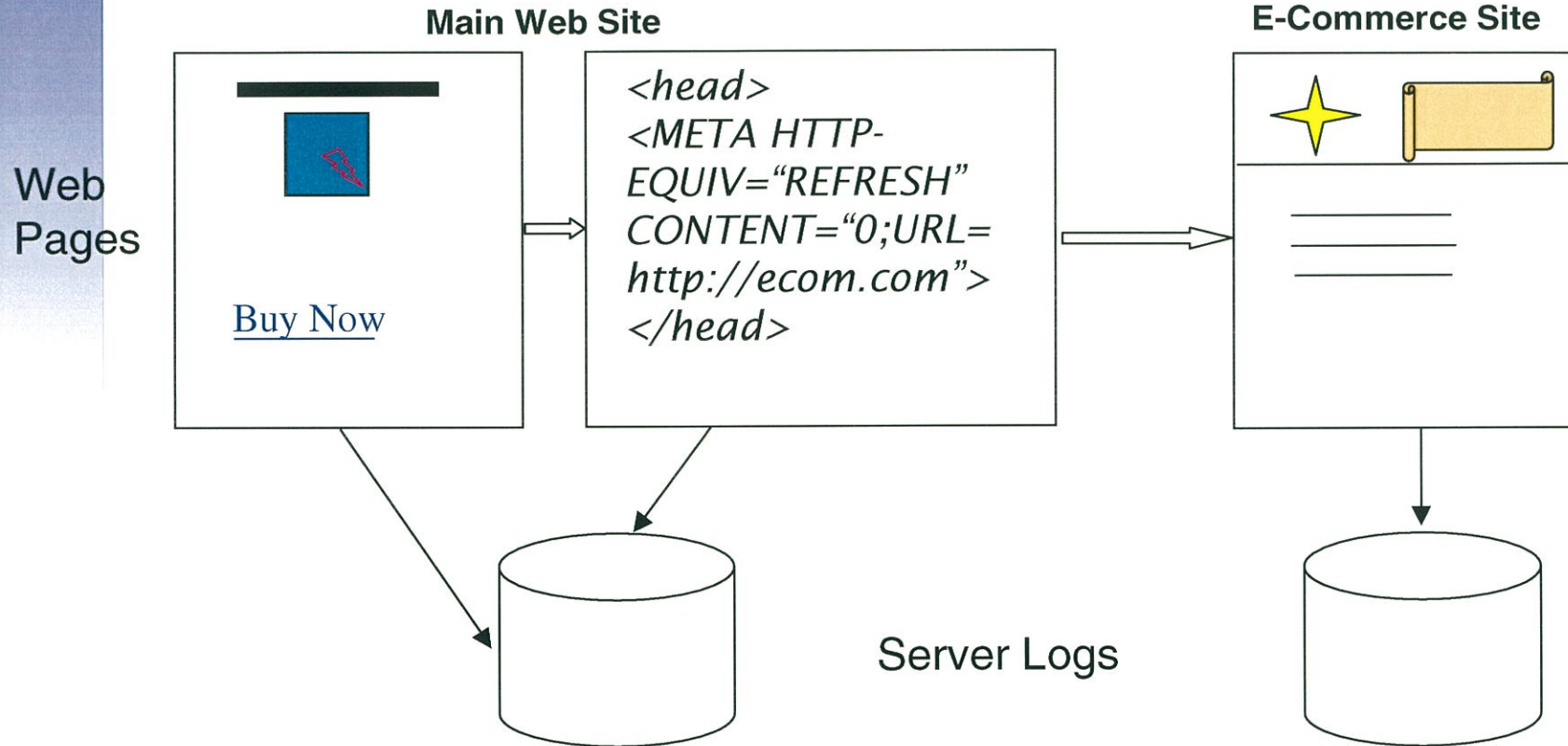
- What Does This Imply for Server Log?

  - 2 brands of browsers
    - one consistently reports 3 refers to Mainsite
  - second - early versions report 3refers as well
    - later versions - one report
    - could of seperate by browser field except -

Referer Info - Browser Pull

Main Web Site

E-Commerce Site

Web Pages

Buy Now

```
<head>
<META HTTP-
EQUIV="REFRESH"
CONTENT="0;URL=
http://ecom.com">
</head>
```

Server Logs

Only META sends referer on.

Result

- Massive Undercounting - From Browser Pull
- Inconsistent Counting - From Frames

Ironically redesign helped.

Telcordia Technologies

Performance from Experience

Caching

- Important to Understand
- expires=0
  - Some commercial tools
  - Bad Idea, given use of **Back** button (up to 50% of navigation acts)
- Proposed java applet partial solution in Paper

Telcordia.
Technologies

Performance from Experience

Discussion

- Web Server Logs Are a Valuable Resource
- Web Server Logs Are Tricky to Analyze
- Be Very, Very, Very, Suspicious of Reported Site Information
- Work at Protocol Standards Level Would Make Things Easier

Telcordia
Technologies

Performance from Experience

Comparison of Filtering out Measurement Sites by Hour of Day