

# Network Application Accelerators at Major League Baseball



(keeping MLB.com above the “Mendoza line”)



# **“Who is this guy?”**

- **I’m Ryan Nelson, Director of Operations for Major League Baseball Advanced Media (MLBAM)**
- **MLBAM is the “Internet Division” of MLB:**
  - **All 30 Clubs (cubs.com, yankees.com, devilrays.com...)**
  - **League websites, worldseries.com, etc...**
  - **Increasing amounts of other stuff**
    - **Minor League Baseball**
    - **Other Sports**
    - **Other Entertainment websites/Streaming Events**



**This Presentation is copyrighted by the  
Commissioner of Major League Baseball.**

**The accounts and descriptions of this  
presentation, and the events described  
herein may not be duplicated, transcribed,  
or rebroadcast in any form without the  
express written permission of Major League  
Baseball.\***

**\* Just kidding.**



# Nature of web traffic at MLB.com

- **Seasonal:**
  - November is very quiet.
  - April is insane
- **Very “dense” site**
  - Lots of small components
    - Box Scores
    - Calendars, Schedules, Stats tables
    - Basically, lots of hits/page
  - “Bursty” traffic

# VERY “bursty” traffic

- **MLB Gameday is basically a self-written denial-of-service attack we wrote against ourselves**
  - **Users are synchronized with game events.**
    - **“next batter” means 500,000 browsers ask for his mugshot at the same time**
  - **7:05pm is showtime**
  - **Day games are far worse... people watch from work, online**
    - **1st round playoffs mean important day games**
  - **Scoreboard page reloads**
  - **Ticketing on-sale days**
  - **Contests/All-Star Balloting**
  - **RSS feeds / AJAX stuff**

# When I got there:

("I", meaning "me, and all the hard-working people I work with")

- **Site was entirely dynamic**
  - **Everything but the images!**
- **Site was uniformly balanced across about 40 web/app servers**
- **Layout of site content was already well-established**
  - (translation: "hard to get changed")
- **Did not use CDN**
  - **Couldn't! Everything was dynamic, not cacheable**
- **SSL accelerator hardware cards in every web server**

And... the site went down. A lot.



The screenshot shows the CNET News.com website interface. At the top, there's a yellow header with the CNET logo and 'NEWS.com'. Below that are navigation tabs for 'Today on CNET', 'News', 'Reviews', 'Compare prices', 'How-to', and 'Deals'. A secondary navigation bar lists categories like 'Today on News', 'Business Tech', 'Cutting Edge', 'Access', 'Threats', 'Media 2.0', 'Markets', and 'Digital'. A search bar is located below the navigation. The main article is titled 'MLB.com on the bench' by Cecily Barnes, published on April 2, 2001. The article text discusses the MLB.com website going offline on opening day. On the right side, there's a sidebar with a 'DEALS' section and a 'THE BIG P' section listing items like Sony, Star, and Nokia.

**c|net NEWS.com**

Today on CNET | **News** | Reviews | Compare prices | How-to | De

Today on News | Business Tech | Cutting Edge | Access | Threats | **Media 2.0** | Markets | Digital

Search:

## MLB.com on the bench

By [Cecily Barnes](#)  
Staff Writer  
Published: April 2, 2001, 12:40 PM PDT

 TalkBack  E-mail  Print

**It's opening day for baseball season, and there's already been a foul ball.**

Major League Baseball's MLB.com Web site fell offline Monday morning as enthused fans logged on in droves to hear audio play of their favorite games.

"We seem to have been down for a brief period this morning, and we're attributing that to the several gazillion people who came to our site for opening day," said MLB spokesman Jim Gallagher.

The slip-up comes one week after the league announced it would begin [charging](#) fans to hear audio broadcasts of baseball games. That deal was announced in conjunction with RealNetworks. Consumers will be charged

**DEALS**

**THE BIG P**

- [Sony](#)
- [Star](#)
- [Nokia](#)



**Obviously,  
this presentation is about  
“How Application  
Accelerators Saved the Day”**

# What “Application Accelerators” Are: (my definition, for the purposes of this talk)

- **Layer-7 network hardware devices**
  - They “speak” at the application layer of networking
    - “what URL did they ask for?”
    - “what ‘Host:’ header did they send?”
    - “is the ‘I-Am-the-CEO’ cookie set?”
  - Not just HTTP
    - SMTP/POP
    - FTP
    - SQL?
    - SSL-ified versions of protocols like this
  - Manipulate things, in layer-7, or act on what they see there.

# Load Balancing with Layer 7

- **Direct requests to various servers based on URL-specific criteria**
- **Separate and protect web “applications” from each other**
  - **Use the web server that’s best for the job**
  - **Protect the Cubs pages from a Red Sox ticketing surge**
  - **Keep the homepages/subscriptions/important-stuff UP. Let the broken stuff break, but in its own sandbox.**
    - **That way, we can fix the broken section, rather than having to fix the whole site.**
  - **Send cacheable requests to a cache**
    - **Flattens the “all-dynamic” site**
      - **Need cache headers, though!**

Continued...

- **HTTP keep-alive magic**
  - Saves up to 30% of CPU previously spent on TCP opens/closes
  - Makes things “feel” much faster -- really!
- **HTTP Compression**
  - Saves \$\$\$
  - Speeds up user experience
- **Centralized logging**
  - No more collecting logs
  - Logs grow in one place, in real time
- **SSL in front, HTTP in back**
  - Less server load
  - Easier to debug things
- **Surge Protection**
  - TCP Syn-Flood and DOS protection
  - “Virtual Waiting Room” stuff
  - Priority queuing

**AND MOST IMPORTANTLY...**



# **You can bolt all these features onto your already existing site, without huge changes**

- **That's a huge part of these appliances' appeal**
- **They let you respond to problems you'd not anticipated**

- **Before MLB, I'd used very early models of Netscaler's accelerators at a previous employer.**
- **Uses then:**
  - **HTTP keep-alives**
    - **Reduced latency in the "time of slow modems"**
  - **Cache redirection**
  - **Centralized logging**

# At MLB:

- **Got them “in the door” just for SSL-gateway**
  - **Before:**
    - **Paying ~ \$60,000/yr for support contracts on SSL hardware**
    - **Plus the cost of one certificate per server (\$20 - \$800 each, depending on provider)**
  - **Cost savings in first year alone paid for the first H/A pair we bought**
  - **...Then we started using them for separation of services**
  - **...then cache redirection**

# A word about cache efficiency:

- **With Layer 7 load balancing, you can do “URL hashing” as an LB method**
  - This means “URLs that look like (*this*) go to (*This server*)” (until it fails)
  - **This makes caches work WAY BETTER**
    - All kinds of caches
      - http cache accelerators
      - NFS disk caches
      - On-disk-controller cache / SAN cache
      - JSP/PHP/Python compiled caches
      - Web servers in-memory file cache
- **Bottom line, web servers work much better (usually) when they get the same requests over and over**

# A word about CDN's:

- **We use Akamai a lot for our http content**
  - **“doesn't that mean you don't need App Accelerators?”**
  - **Actually, Akamai nodes seem to really like talking to App Accelerators**
    - **Usually, it's all strict HTTP 1.1**
    - **Compression**
      - **Saves bandwidth and \$\$\$**
    - **Connection re-use**
    - **Content switching means you can give priority to your CDN's nodes!**

## **Summary of neat and/or life-saving features:**

- **Centralized logging**
- **Isolation/Protection of web site functional areas**
- **Traffic surge protection**
- **HTTP multiplexing**
- **Compression**
- **SSL**
- **You can bolt it onto an existing infrastructure**
  - **A really big deal if you work at a place like mine**

# Questions?

- **My contact info:**
  - **Ryan Nelson**
  - **[ryan.nelson@mlb.com](mailto:ryan.nelson@mlb.com)**