

CassBeth

E-Commerce

January 17, 2002

FredsSpot

*Earthlings
United*

*GiftsTo
Humanity*

*Outerspace
Shopper*

*Martian
Shopper*

ElizAndra

Contents

- Introductions
- What is E-Commerce
- The Process
- Findings & Insights
- Automation
- Scale of the Internet
- Demos
- Technologies, spin-offs, & assets

FredsSpot

*Outerspace
Shopper*

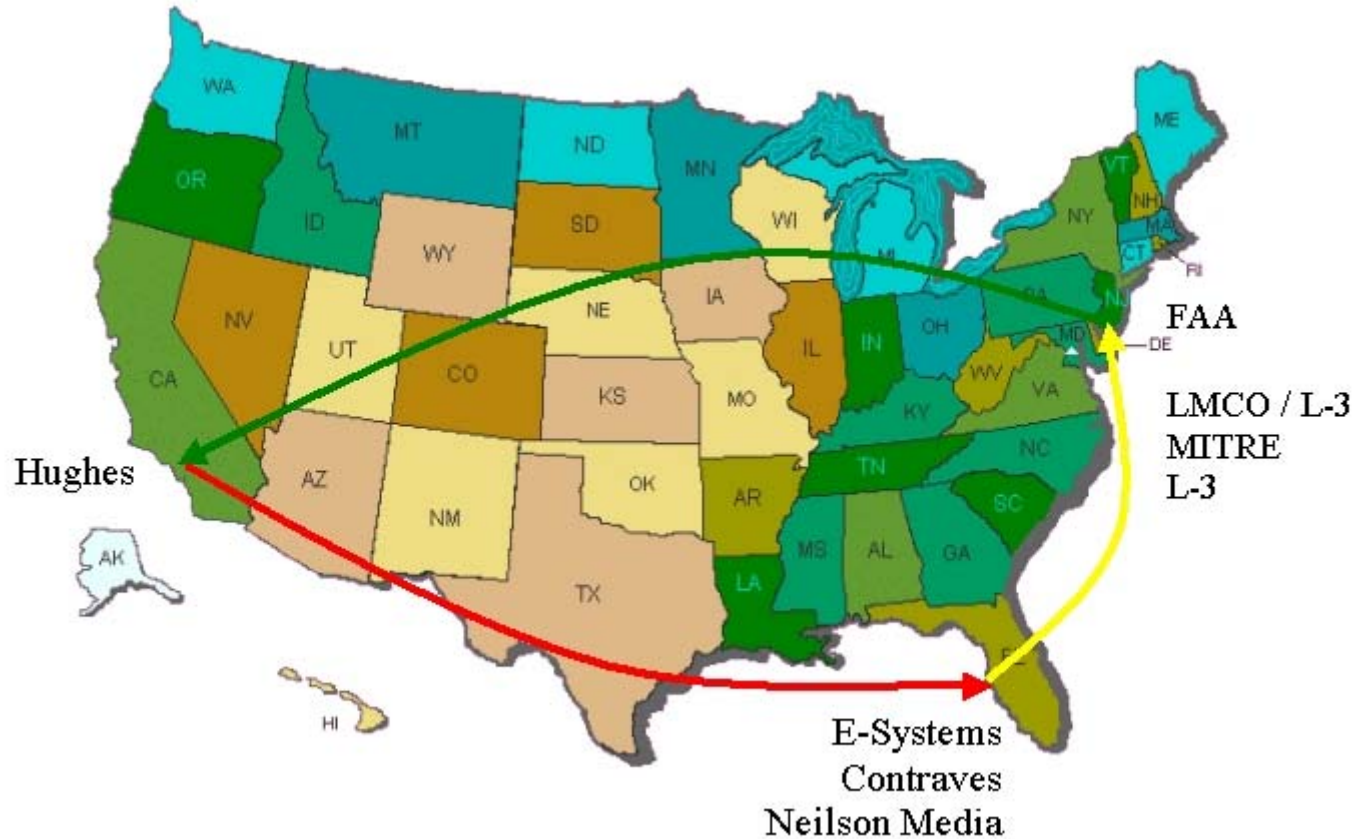
*Earthlings
United*

*Martian
Shopper*

*GiftsTo
Humanity*

ElizAndra

Introduction



Introduction

- CassBeth established in August 1997
 - Early goals were to study internet technologies, involve family in a group activity, and perhaps make some money
- CassBeth is now a family of web sites
 - ElizAndra, FredsSpot, GiftsToHumanity
 - MartianShopper, OuterSpaceShopper, EarthlingsUnited
- Traffic exceeds 1.5 million page views per year

Some Systems

- Air Defense System
- Air Traffic Control System
- Transportation System
- Political System
- Financial System
- Banking System
- Health Care System
- E-Commerce System ???

Personal Insight

- Infrastructure systems are always operational - simulations exist in offline lab settings (ATC)
 - failure is immediately detected
- Military systems are always in simulation mode - operational settings exist in unique isolated situations (Air Defense)
 - failure is not always apparent in the confusion of war
- How do you create / modify / expand an infrastructure system that is always online?

CassBeth

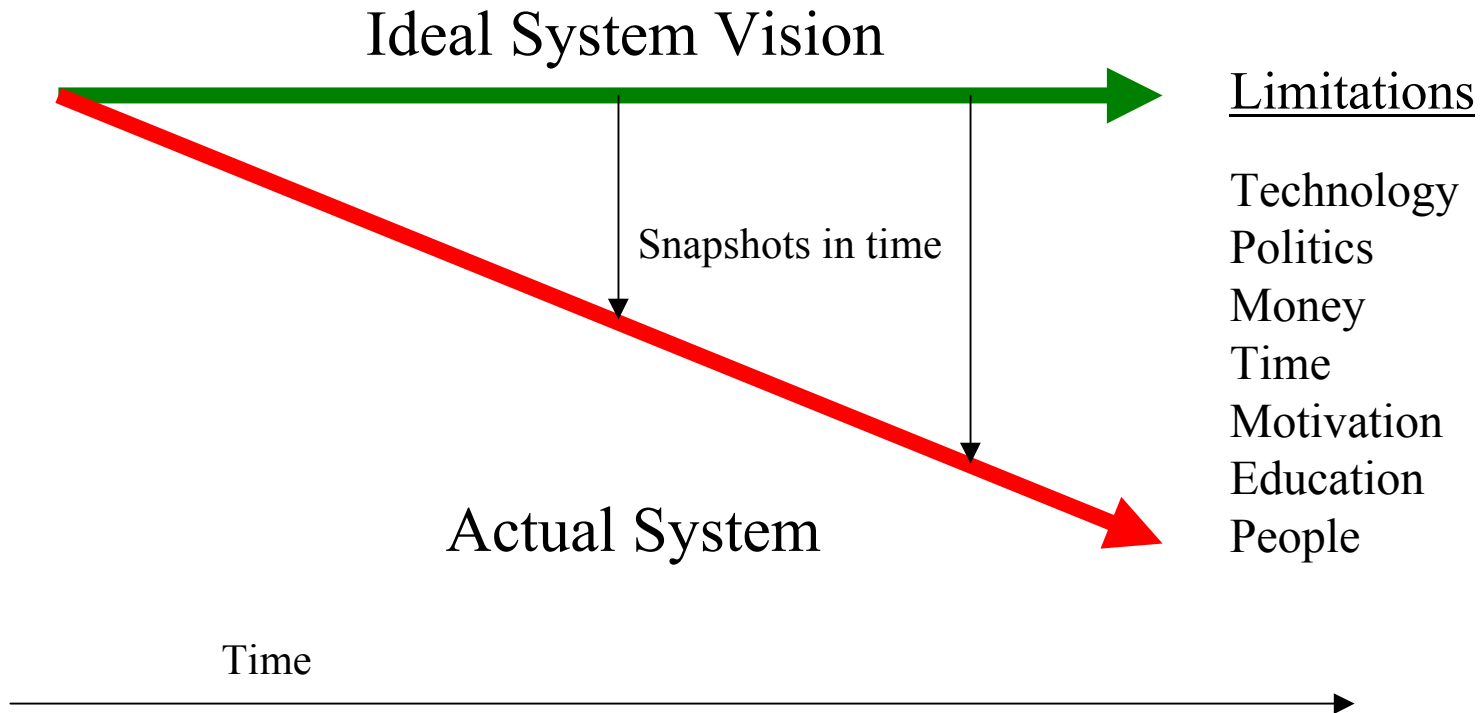
What is E-Commerce

Infrastructure System

Stage?



System Engineering



System Engineer - who am I

- Creator, keeper, and champion of the system architecture
- Search for the forest and leave the trees to others
- Look for key requirements and issues that are system architecture drivers
- Use the scientific method to solve technical problems that optimize the architecture
- Digest the findings of specialists who provide data critical to the architecture / system vision

System Engineer - who am I

- Communicate the vision to everyone of every skill level and background
- Constantly ask what if & look for break throughs that move system closer to ideal or to a new ideal
- Study the past & present of the system or similar systems & postulate a future state of being for the new system
- Immerse in technologies & postulate what could be

Creating Systems

- Approach #1: Identify the mission / needs / requirements and then attempt to create it with current, modified, or new technologies, machines, and processes
- Approach #2: Let technology evolve on its own, taking its path of least resistance and when someone decides, throw it into real world operational settings, letting the human participants morph it into their world

So What Did We Do?

- Formulated final vision of the system
 - create an effective business based 100% on Internet technologies
- Created a live model - realizing it is infrastructure
- Decomposed the system into 5 broad functions
 - gather content
 - author pages
 - publish to web
 - submit to search engines
 - track results
- Created processes / techniques / tools for each of the 5 broad functions

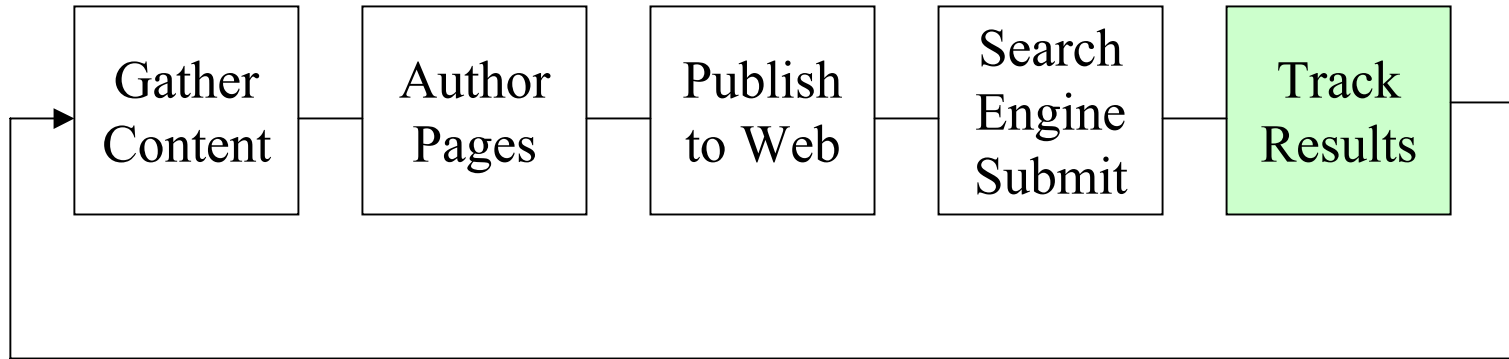
Businesses Attempted

- Web Site Development & Web Page Hosting
 - Retailers, Professionals, Businesses
- Internal Product Sales
 - Limoges, Stained Glass, Tiffany Lamps, Electronics, Eggs, etc
- Advertising
- Amazon Associate Program
- Other Associate Programs
 - Cdnw, E-Toys, LinkShare, Befree, Commision Junction
- Multiplication of the winner via similar sites

Quick Observation

- I have access to more compute power in my home today than the ATC and Air Defense Systems I use to work on
- If I chose to, I have more access to and can pay for more bandwidth than the ATC and Air Defense Systems I use to work on
- Computers and Communications are slowly becoming like ops amps - infinite gain and bandwidth

Process From Day 1



Optimize for \$\$\$

Critical Role of Logs

- The log files told us the winners instantly
- We were never able to guess the market even though we really tried
- You can never guess the market, you can test it and move out after a valid test
- A web site with proper logging will test the market within 5 days of getting traffic
 - In the old days before the search engines fell victim to the profit mantra, the result from publishing to conclusion was < 10 days

Significant Early Insights

- This is Not TV or Radio
 - Physics limited the number of radio and TV stations
- The Internet allows for hundreds of millions of transmitters
 - People will not bookmark the universe or type www.xyz.com
 - Whoever established the ground rules for the Internet and the rules of the search engines was a genius. These simple rules make millions of transmitters possible today.
- There is traffic that results in sales & the traffic is very focused

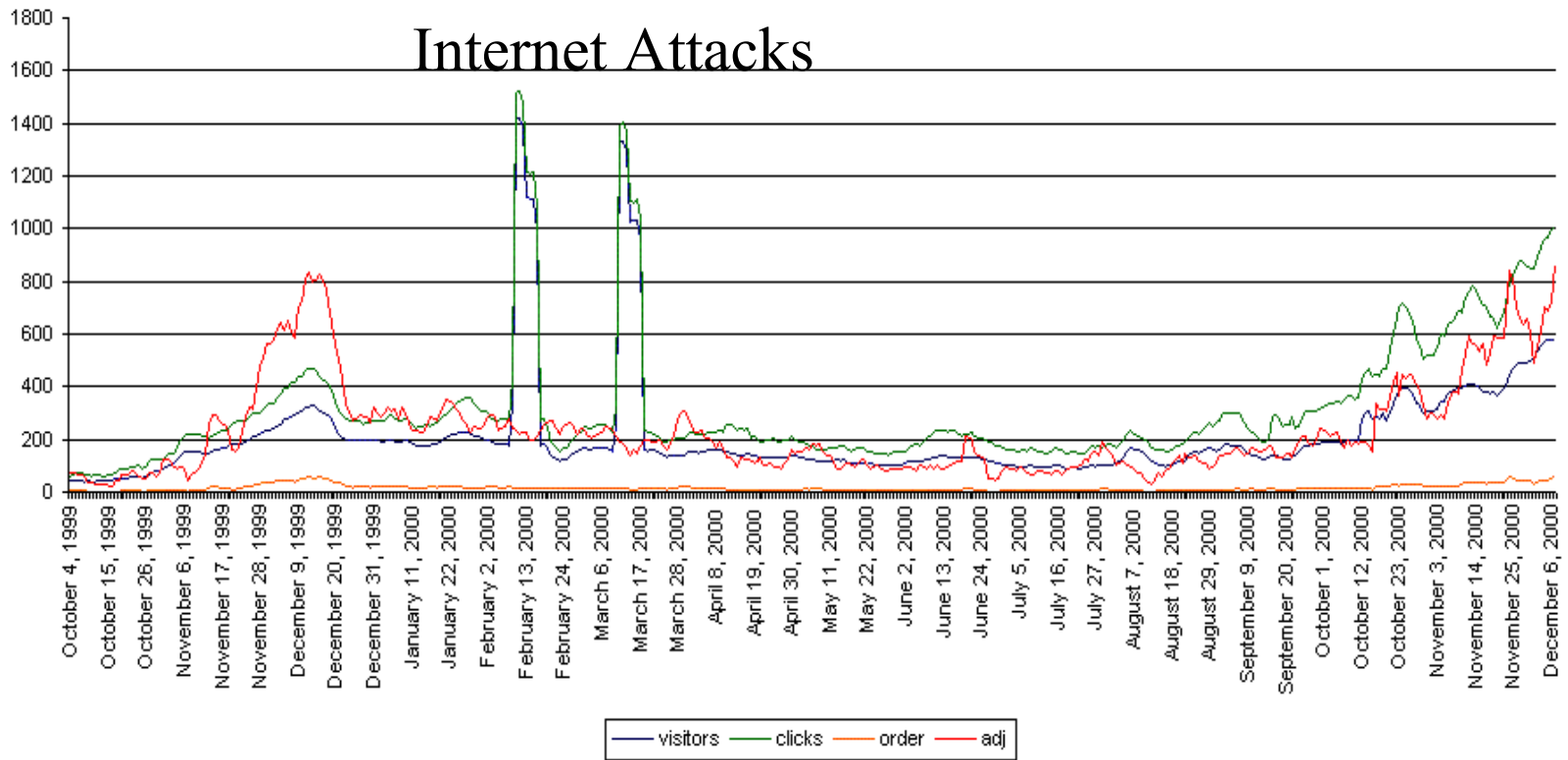
Web Findings

- People arrive based on a theme
- People rarely depart a theme
- Search engines require 5-10 pages per theme
- Cross fertilization increases search engine results
- Cross fertilization helps in sales - confidence
- Banner advertising does not work
- Madison Ave advertising significantly increases sales of unrelated items

Broad Findings

- People are flakes
- No one wants to pay for non physical labor services
 - Business is about hard boiled eggs
- Everyone thinks the Internet is simple & trivial
 - Small business owner: I'll have my high school kid build a site
 - Fortune 500 executive view: DOT-Bombs, hahahaha
- Distributors are not dependable
 - Out of stock, late deliveries, unwilling to work
- Distributors are not prepared for what is coming
 - Amazon could become Earth's first second tier distributor
 - Amazon could break franchises & allow main street rebirth

Sample Amazon Log



Amazon

- Amazon.com Invented Something Great
 - Cassbeth.com delivered over \$10,000 worth of business to amazon.com in 98. If cassbeth.com is a typical good associate (i.e. top 20%), then the associates delivered $150,000 * .2 * 10,000 = \300 million in 1998 sales for a 4% to 5% commission.
- Amazon has over 600,000 associates today
- Amazon holds the patents
 - 1. 6,144,958 System and method for correcting spelling errors in search queries
 - 2. 6,064,980 System and methods for collaborative recommendations
 - 3. 6,029,141 Internet-based customer referral system
 - 4. 6,006,225 Refining search queries by the suggestion of correlated terms from prior searches
 - 5. 6,003,024 System and method for selecting rows from dimensional databases
 - 6. 5,999,924 Method and apparatus for producing sequenced queries
 - 7. 5,963,949 Method for data gathering around forms and search barriers
 - 8. 5,960,411 Method and system for placing a purchase order via a communications network
 - 9. 5,727,163 Secure method for communicating credit card data when placing an order on a non-secure network
 - 10. 5,715,399 Secure method and system for communicating a list of credit card numbers over a non-secure network

Invest in Amazon?

- Amazon will get all the customers in the end
 - Then why are sales increasing dramatically?
- Why will people come to your silly site
 - Who knows and why do I care
 - Business is about hard boiled eggs
- Amazon will go out of business
 - Oh well
- Amazon will dump the associates
 - And lose 80% of their revenue stream? Oh well.

Hard Boiled Eggs

- You try it, if it works and you like it, you keep doing it
- You try it, if it fails but you like it, stop doing it if you need the cash
- You try it, if it works but you hate it, stop doing it because life is too short



=> This is Business <=

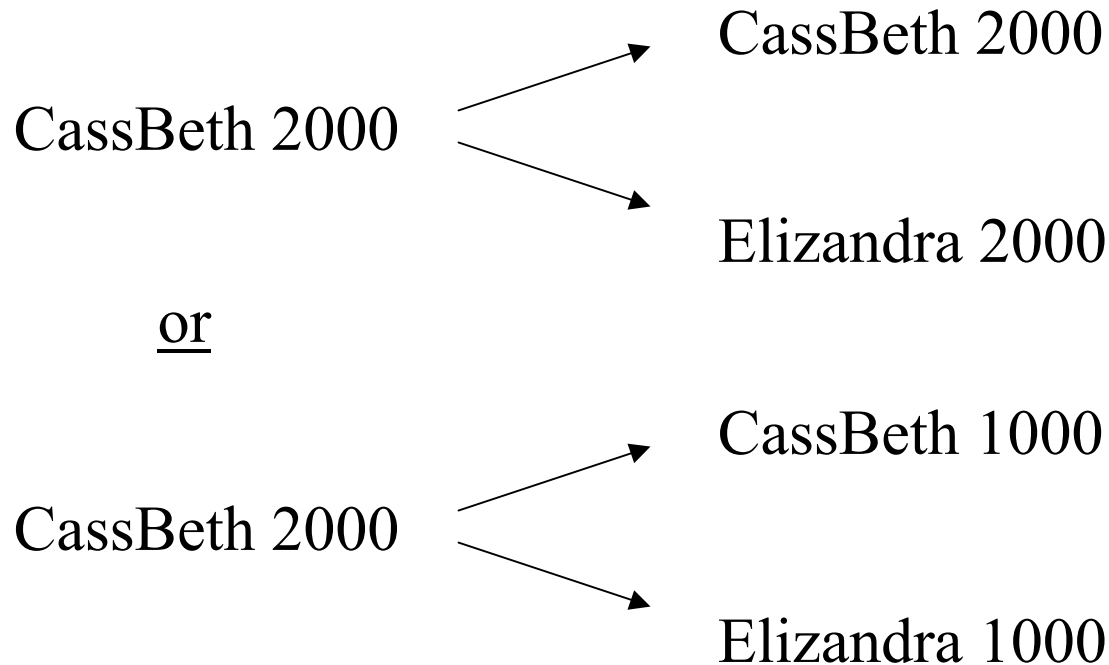
Our Hard Boiled Eggs



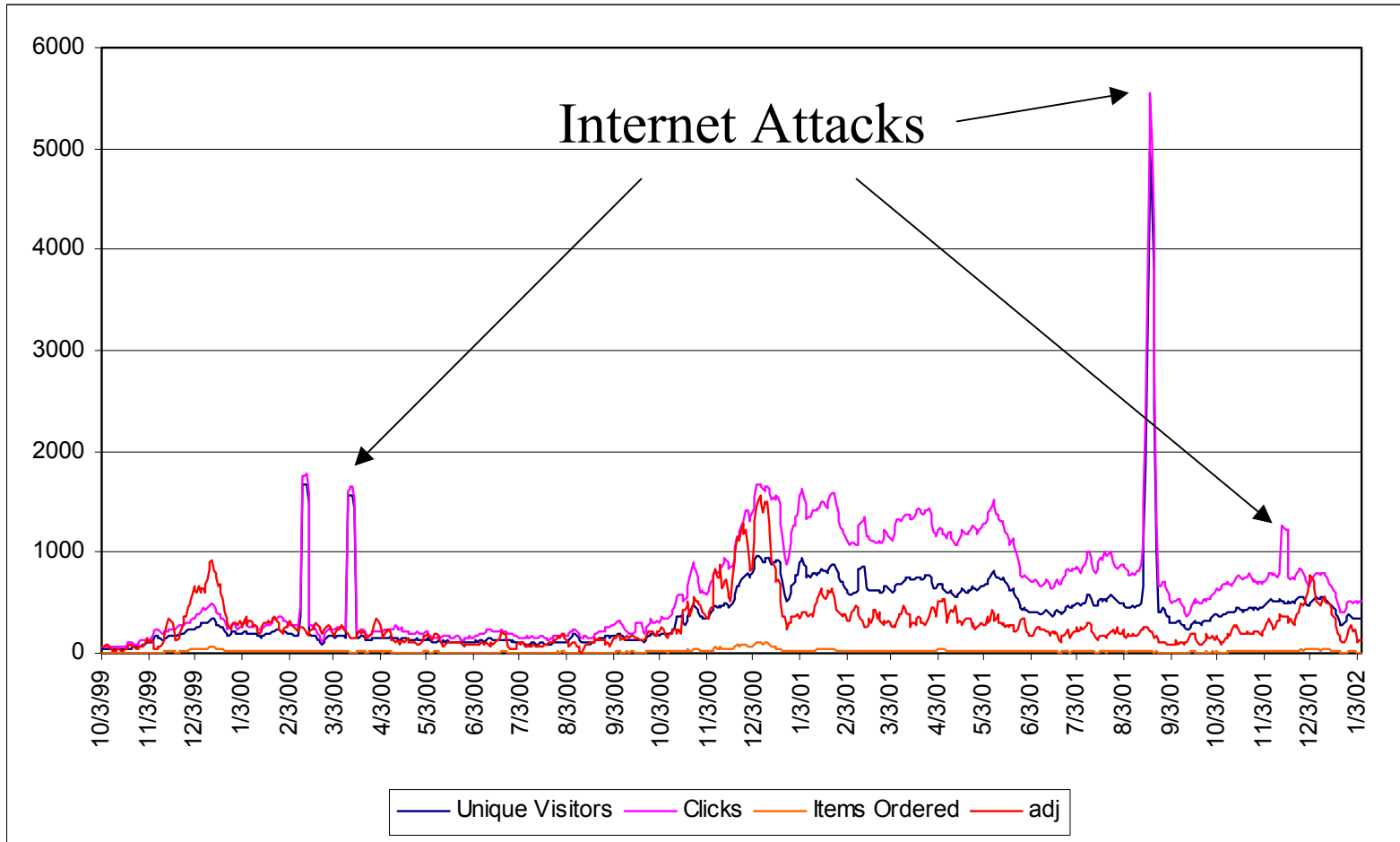
- Resell Amazon products
- Multiply with additional sites
- Use massive automation to build large sites
- Track results & look for traffic conflict between sites, if any
 - Results are unclear because of political instability at the time the new sites were launched in 4Q 2000
 - Results are still unclear because of the economic down turn in 2001
 - September 11, 2001

Channel Conflict

The great experiment



The Experiment



Automation Evolution

Phase	1	2	3	4
Content	Catalogs	Catalogs	Catalogs	Catalogs
Author	AOL Press	AOL Press	AOL Press	AOL Press
Publish	Manual ftp	Manual ftp	Manual ftp	Manual ftp
SE Submit	Manual	Manual	Manual	Semi Auto
Track	None	Counters	Logs Text	Logs Tools

Phase	5	6	7	Future
Content	Catalogs	Man mine	Automated	Automated
Author	AOL Press	Automated	Automated	Automated
Publish	Manual ftp	Libraries	Libraries	Automated
SE Submit	Automated	Automated	Automated	Automated
Track	Custom	Custom	Custom	Custom

The Work

- Daily
 - » Check and delete much e-mail
- Weekly
 - » Examine real time log analysis reports
 - » Examine Amazon sales and traffic reports
- Monthly
 - » Examine Amazon online sales reports
 - » Download Amazon monthly sales and traffic reports
 - » Update sales projection reports
 - » Download log files from all sites
 - » Delete old log files from all sites
 - » Prepare status report

The Work

- Quarterly
 - » Update product libraries
 - » Update shops on all sites
 - » Update Perl tools
 - » Examine Amazon online and e-mail sales reports
- Yearly
 - » Taxes (Mar)
 - » Annual numbers (Jan)
 - » Examine Amazon online sales reports
 - » Strategy and redirection (summer)

Scale of Internet

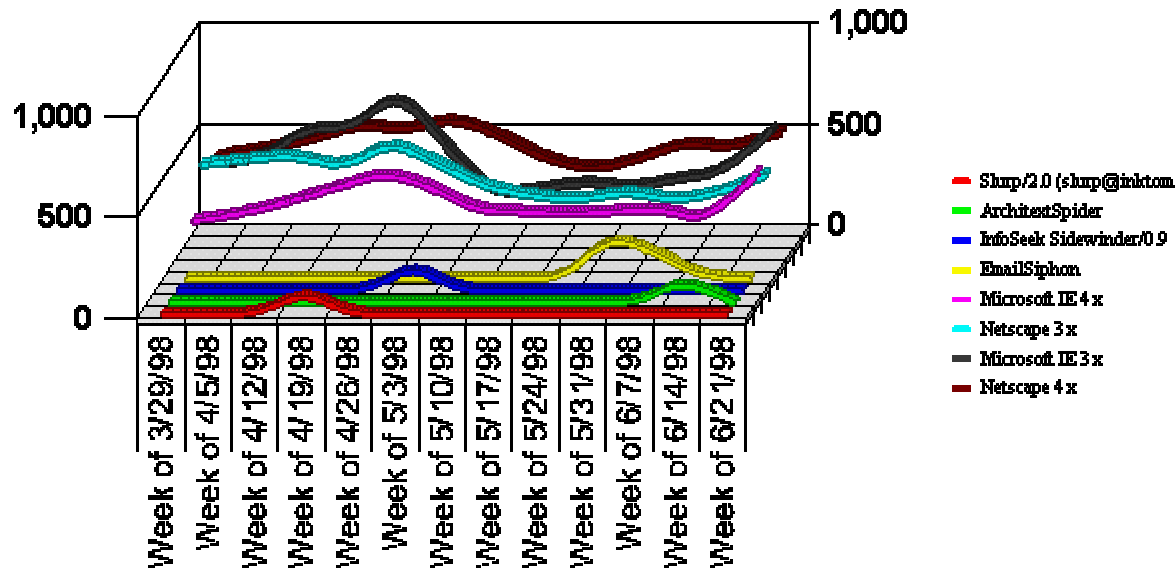
- CassBeth averaged a visitor every 3 seconds in 4th quarter of 2000
 - Traffic exceeded 10 gig bytes in December
- Other six sites combined traffic equaled CassBeth
- Six additional Domains registered & installed between Oct and Dec 2000
 - Over 1000 pages on CassBeth
 - Over 400 pages per site on other 6 sites
 - 45,000 to 60,000 products per domain
 - 35 to 45 megabytes per domain
 - 100 megabytes CassBeth

Scale of Internet

- 4,000 to 5,000 MIPS used to create a new shop in less than 1 minute
- 200-500 MIPS used to spider new shop content
- Shop building metrics
 - Manual content gathering previously took 1-2 person weeks while automated approach takes less than 30 seconds to start and 10-20 minutes for machines to complete
 - Manual Authoring previously took 1-3 person weeks while automated approach takes less than 1 minute
 - Manual search engine submission 3-6 person weeks while automated approach takes less than 30 seconds to start and 1 hour for machines to complete

Scale of Internet

Most Popular Browsers Per Week



Scale & Perspectives

- In the early 80's I was asked to create a new architecture that would not fail & support growth & technology insertion into the next century
- The solution was a Highly distributed system with over 300 computers per facility with 30 facilities across the USA - way too risky, IBM MFs won
- Never did I / we visualize hundreds of millions of computers connected together into a single machine
 - » I am truly humbled today

Technologies - Demos

- Order, shipping, & accounting system
- Visitor analysis
 - tail, seanalysis, searchlog, happyvisitors, graph
- Automated search engine submission
- Automated shop building
- Automated content library gathering
- Automated library distribution

All are 100% Internet Based

Demo Highlights

- All reports maintain context
 - Allows analysts to modify analysis with each new insight or to submit the same data for the same analysis & comparison with previous runs
- Data can be completely unstructured
 - Unlike database solutions which are highly structured & therefore very brittle leading to breaks which require changes
- The GUI is effortless & provides hyperlinks
 - Visualization is key to analysis & previously GUIs were more than half the battle
- Its ubiquitous
 - Available to anyone with access to the Internet

Technology Spin-offs

- To MITRE for free as employee to support NEXCOM radio data reduction & analysis
- To MITRE for free as employee to support URET DART log analysis & regression testing
- To L-3 for free as employee to support IAT source code analysis, req's tracking to the source code level, and testing

Spin-off Results

- Pattern searches on globs of unstructured data can be more useful than database approaches
 - Simple search engine can return web page or glob of data
- Approach was not practical just 2-3 years ago
 - All machines used were / are in the 500 - 1000 MIPS range
- Each solution had completely different reporting needs and the HTML browser GUI worked
- Completely different data was analyzed
 - log data from NEXCOM, NAS and L-3 all very different systems
 - source code and requirements

Spin-off Observations

- In the past companies built & owned their tools
 - Separated themselves from the pack and wannabes
- COTS solutions for many of the custom tools surfaced in the 80's and 90's
- IMHO experience shows that there are always problems with the COTS tools
 - If there are internal processes they usually need to change to accommodate the COTS introduction (last 5% is 90% of the work)
 - If there are no processes, no one knows what to do with the tools

Spin-off Observations

- The Internet and its underlying technologies may make internal tools practical again
 - No license fees so they are available to everyone in company
 - Desired processes are not modified to accommodate the tool(s)
 - Company can separate itself from the pack once again
- IMHO existing tools & techniques developed decades ago and still in use today stand to benefit significantly with the introduction of simple Internet technologies (PERL, server, browser)
 - DART log and URET

Current Assets

- Registered domain names
- Current content
 - site organization
 - themes / shops
- Good will - previous customers
- Search engine indexing
- Existing site traffic
- Product libraries

Technology Assets

- Contact, order, shipping, accounting (no value)
- Logging and log analysis (not bad - no value)
- Automated search engine submission (no value)
- Automated shop building (very powerful)
- Automated content gathering (very powerful)
- Actual technology transfer / spin-offs

Elements of a Web Site

- HTML content
 - text, graphics, audio, images
- CGI processing usually in PERL
 - adds functionality to the site such as forms processing
- SSI processing
 - allows for common content to be included across many pages
- JAVA or JAVA script
 - adds functionality to the site such as display processing
- A web server connected to the backbone
 - APACHE is the leader and its free

Small Site Hosting

- Ability to run CGI scripts
- Ability to run server side includes (SSI)
 - work with .html, .shtml, and exec commands
- Ability to use your own .htaccess file
- Ability to FTP content using third party tools
- Ability to Telnet using third party tools
 - some require secure telnet, which is painful, but ok because they also provide control panels

Small Site Hosting

- 5-10 GB of transfer per month
- Daily backups
- Your own IP address
- Its nice to run as yourself, not as a special user `<web>` when your pages are accessed
 - this simplifies privilege issues
- UNIX or any of its derivatives

Small Site Hosting

- Its nice to have https secure processing
 - but you probably won't use it
- Its nice to have SQL and shopping carts
 - but you probably won't use it, especially if you read the CassBeth report

Links

Tutorial

www.earthlingsunited.com/webdev

E-Commerce

www.cassbeth.com/custsrvc/weblessons98.doc

Group

groups.yahoo.com/group/Amazon-Affiliates/messages

Portal

www.elizandra.com/

Portal

www.fredsspot.com/

Portal

www.giftstohumanity.com/

Portal

www.martianshoppr.com/

Portal

www.outerspaceshopper.com/

Portal

www.earthlingsunited.com/

Portal

www.cassbeth.com/