Fidelity Investments

Fidelity Investments eBusiness FeB-T Engineering

E-Commerce and Reliable Systems

Design for Stability

Presentation Objectives

Fidelity Investments

- Describe successful E-Commerce environments
- Describe system and application behavior
- Describe system and application components
- Discuss holistic design approach
- Discuss true stability
  - Difficult to achieve in the E-Commerce marketplace
  - Worth the effort

E-Commerce Landscape

Fidelity Investments

- Participants in the Internet marketplace are interested in the electronic buying and selling of goods and services
- Participants always want their marketplace objectives to be met
- The bad parts of the Internet marketplace is are the ones customers remember
- Lack of understanding of the marketplace component technologies blocks successful participant experience
- The Internet marketplace is only as good as it’s weakest link
- The Internet marketplace is a collection of services, each with a different level of stability
- The responsibility of creating a stable Internet marketplace lays on those who are empowered to effect change: you make the difference

Customer Landscape

Fidelity Investments

- **Customer connection**
  Ability of customer to connect to Internet Service Provider (ISP) to access the internet
- **Internet peering points**
  Ability to traverse the Internet mesh from ISP to other ISP's
- **Firewall and proxy support**
  Ability to open sockets and route traffic through a secure intermediary
- **Internet latency**
  Ability to send and receive data from a customer, through ISP routers, to the server, and back in a timely fashion
- **Browser versions and supported functions**
  Ability of customer browser to execute the HTML, DHTML, Java applets or Javascript code returned from the server

Behavior Definitions

Fidelity Investments

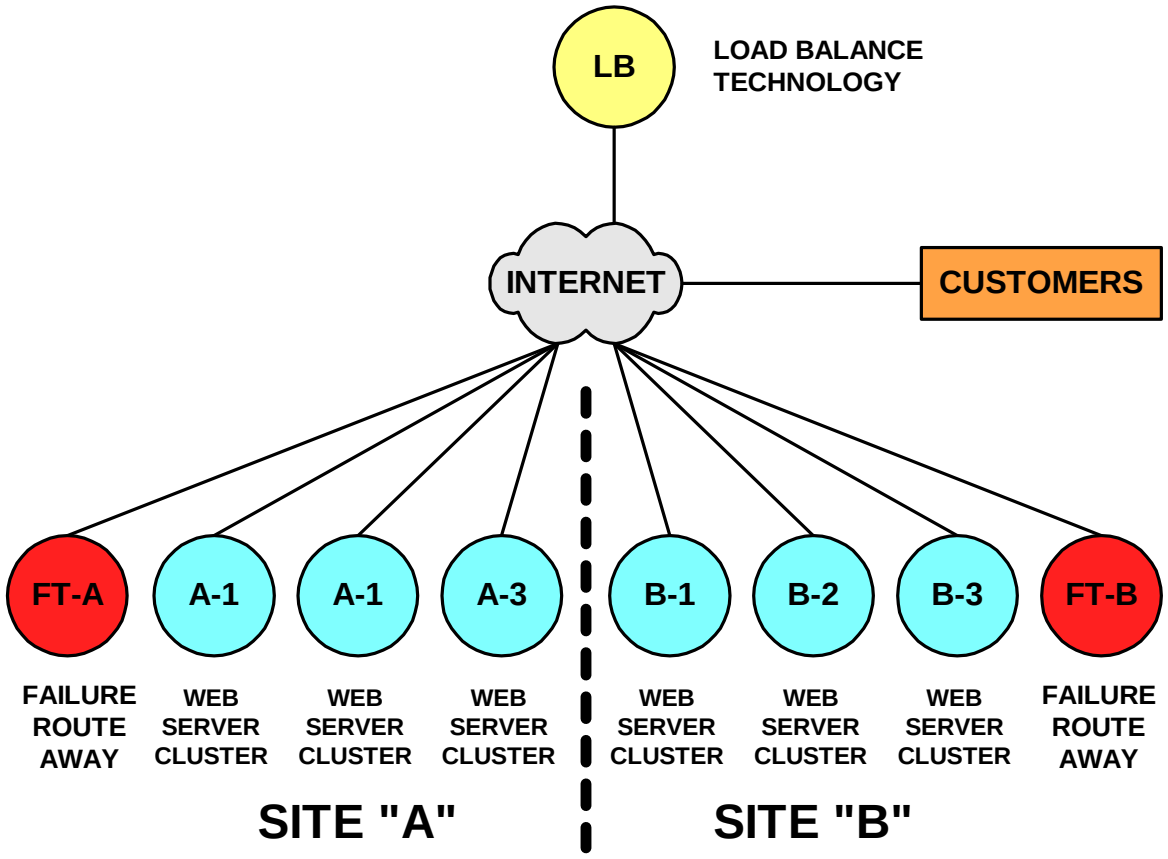
- **Availability**
  The ability to always provide a response to a request.
- **Supportability**
  The ability to administrate, monitor and disable services
- **Performance**
  The ability to respond to a request in a timely manner
- **Capacity**
  The ability to provide responses to all simultaneous requests made to a site
- **Usability**
  The ability for customers to easily navigate to the desired target
- **Maintainability**
  The ability to install and upgrade services
- **Extensibility**
  The ability to add features as the services evolve
- **Feature Set**
  The ability to provide the services customers request

System Redundancy

Fidelity Investments

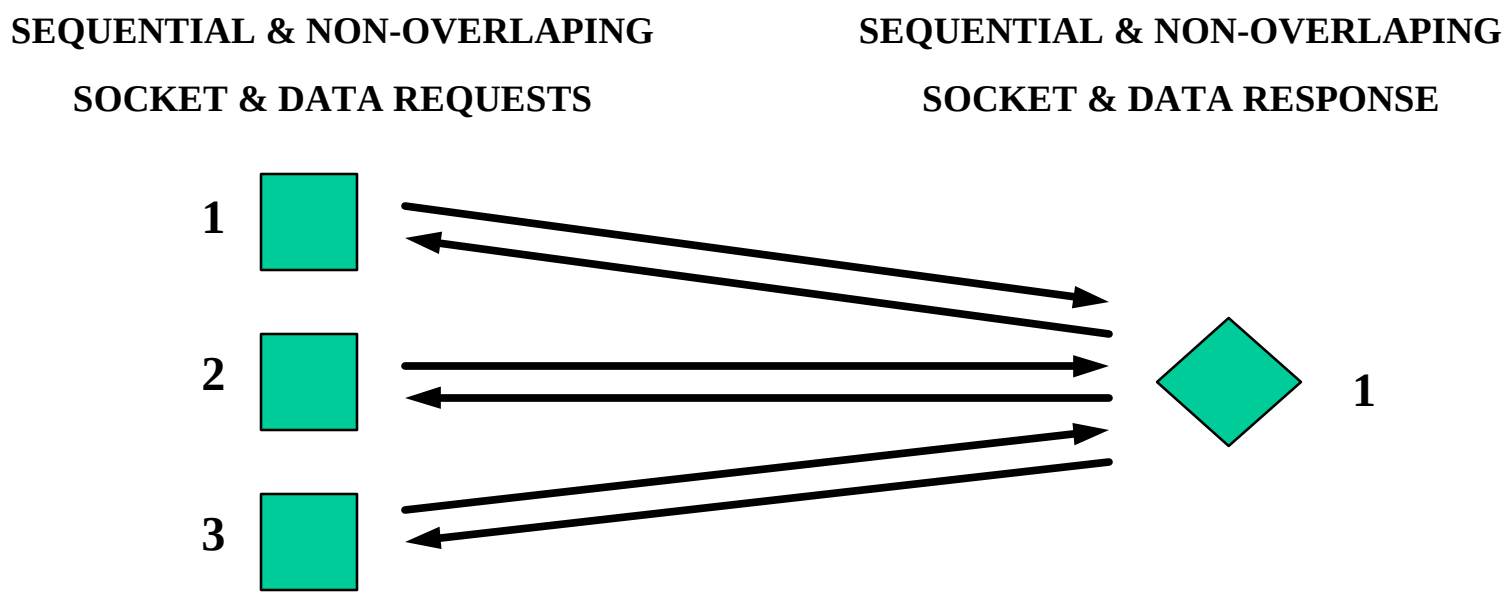
- **Multi-site design**
  - Install identical systems in multiple and geographically disperse locations
  - Distribute customers to all geographic sites
- **Multi-system design**
  - Customer accesses multiple systems of the same type
  - Customer is sent to a backup system when primary system fails
- **Data Replication**
  - One read-write master to many replicas
  - Synchronization time across replicas
  - Real time updates versus batch cycle updates

Customer Load Balancing



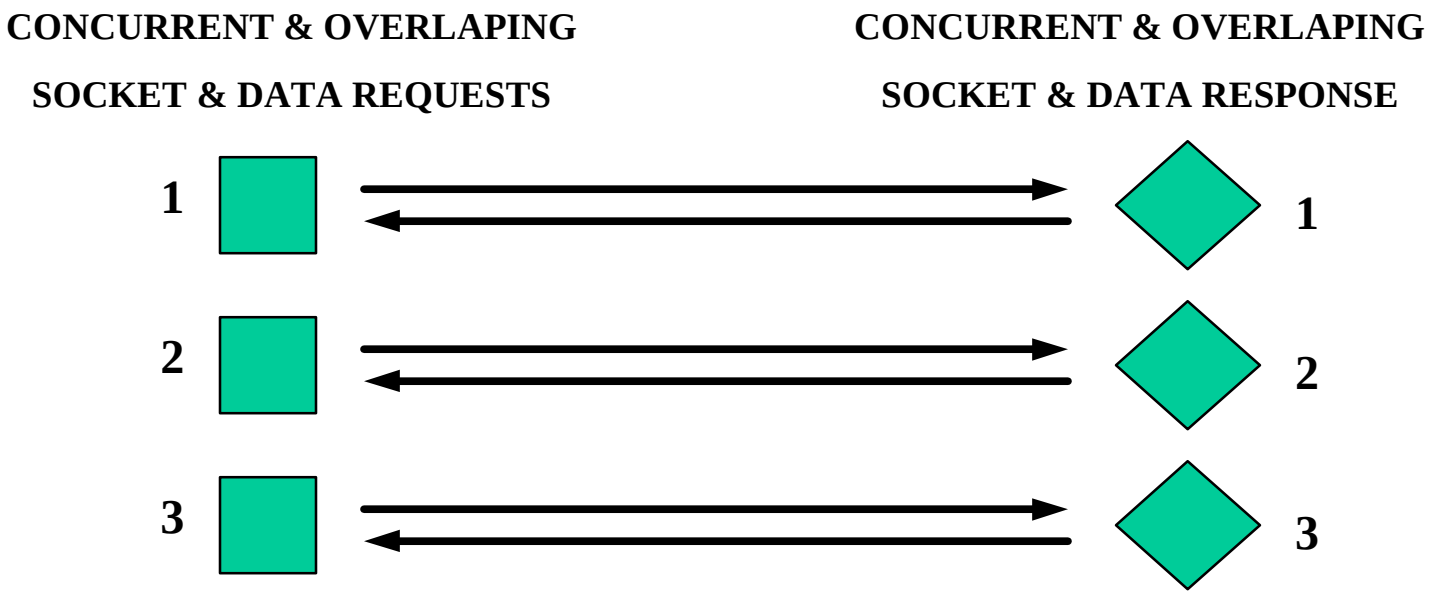
Transient TCP Connections

- Connections are established, data is requested, data is delivered, connection is immediately terminated
- A small number of sockets can service a large number of requests if requests are short enough to support arrival rates
- Socket setup overhead may overload system
- Socket address limitations may cause unavailability



Persistent TCP Connections

- Connections are established, data is requested, data is delivered, connection remains for subsequent requests until thread connection is terminated
- The number of requests that can be serviced equals the number of threaded sockets available
- Thread limitations may cause unavailability



Application Threads of Execution

Fidelity Investments

- **Race condition**
  One thread changes shared resources and causes another thread to fail
- **Deadly embrace**
  Two threads block and wait for each other to complete
- **Exhaustion**
  All threads are waiting to complete and no new threads are available
- **Starvation**
  Attempts are made to start a new thread when no new threads are available
- **Pooling**
  A larger number of thread requestors successfully share a smaller number of actual threads

Application Thread Model Example

Fidelity Investments

PRODUCT "A"		SECURE/443		
ZONE A	WEB	AT	EP	TT
	HOST A	1024	1	1024
		1024	**1**	**1024**
ZONE B	GW	AT	EP	TT
	HOST B-1	128	4	512
	HOST B-2	128	4	512
		256	**8**	**1024**
ZONE C	TX	AT	EP	TT
	HOST C	128	8	1024
		128	**8**	**1024**

PRODUCT "B"		SECURE/443		
ZONE A	WEB	AT	EP	TT
	HOST A	32	4	128
		32	**4**	**128**
ZONE B	GW	AT	EP	TT
	HOST B-1	32	2	64
	HOST B-2	32	2	64
		64	**4**	**128**
ZONE C	TX	AT	EP	TT
	HOST C	32	2	64
		32	**2**	**64**

ZONE C / BACKEND		AUTH			CACHE			WATCHDOG			STATS		
		AT	EP	TT/FL	AT	EP	TT	AT	EP	TT	AT	EP	TT
	HOST C	64	4	256	64	1	64	64	1	64	64	1	64
		64	**4**	**256**	**64**	**1**	**64**	**64**	**1**	**64**	**64**	**1**	**64**

	TX-1			TX-2			TX-3			TX-4		
	AT	EP	TT/FD	AT	EP	TT/FD	AT	EP	TT/FD	AT	EP	TT/FL	NC	NS
	64	8	512	64	4	256	64	4	256	64	2	128	2	4
	64	**8**	**512**	**64**	**4**	**256**	**64**	**4**	**256**	**64**	**2**	**128**		**4**

IN FLIGHT TRANSACTION ESTIMATES AND TOTAL TPS FOR SERVER TYPES

	TAT	FLT	TPS	FAC	
WEB SERVER	1024	10	102	100%	ZONE A
GW	1024	10	102	100%	ZONE B
TX	1024	10	102	100%	ZONE C
AUTH	256	3	85	100%	ZONE C
TX-1	512	7	73	100%	ZONE C
TX-2	256	4	64	100%	ZONE C
TX-3	256	4	64	100%	ZONE C
TX-4	128	4	32	100%	ZONE C

PERCENTAGE DISTRIBUTION FOR BACK END SERVERS

	TAT	TEP	TDS	TLS	TT
AUTH	25%	29%	50%	0%	**40%**
TX-1	25%	7%	13%	0%	**10%**
TX-2	25%	14%	25%	0%	**20%**
TX-3	25%	14%	0%	100%	**20%**
TX-4	25%	14%	25%	0%	**20%**
	125%	79%	113%	100%	110%

ALLOCATIONS FOR BACK END SERVERS

TAT	TEP	TDS	TLS	TT
256	28	1024	256	**1280**
256	**28**	**1024**	**256**	**1280**

KEY:

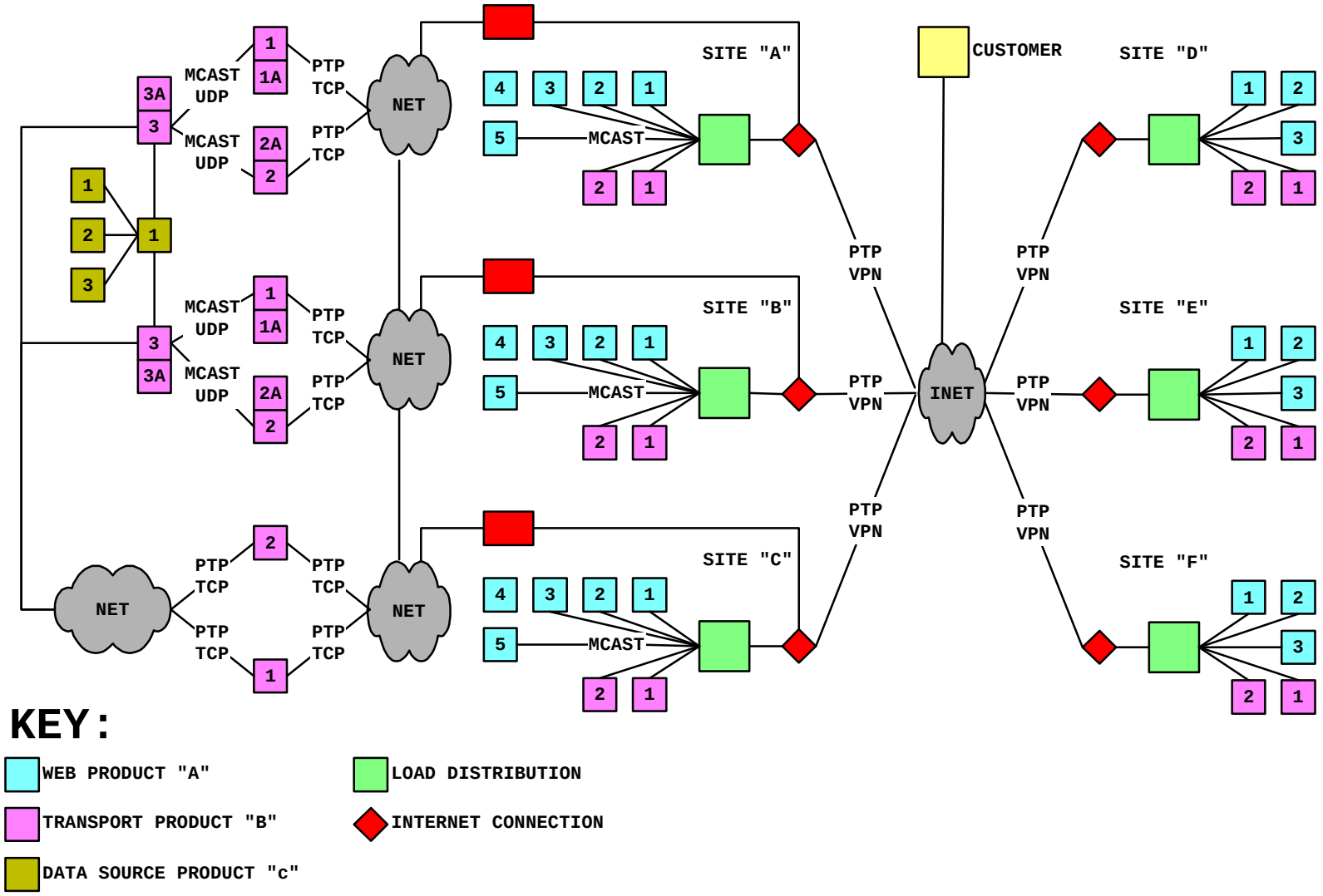
TAT	TOTAL APPLICATION THREADS ALLOCATED TO SERVER
FLT	TOTAL INFLIGHT TRANSACTION TIME FOR SERVER
TPS	TOTAL TRANSACTIONS PER SECOND FOR SERVER
FAC	EFFICIENCY FACTOR FOR SERVER
AT	Application Threads
EP	Servers Processes
TT/FD	Total Threads / Persistent Sockets
TT/FL	Total Threads / Transient Sockets
NC	Config File Settings
NS	Sockets

Instrumentation

Fidelity Investments

- **Dynamic debug levels**
  Run-time setting of server software to log different levels of detail about what is happening inside the program
- **Concurrency levels**
  The number of active connections at the server
- **Operating system statistics**
  Low-level system internals such as CPU activity, RAM use, disk activity, number of active threads
- **Response times**
  The time taken to respond to various types within a fixed time period
- **Request type profiling and histograms**
  Charts and analysis of quantities and arrival rates of various request types

Prototypical Distributed Computing Model

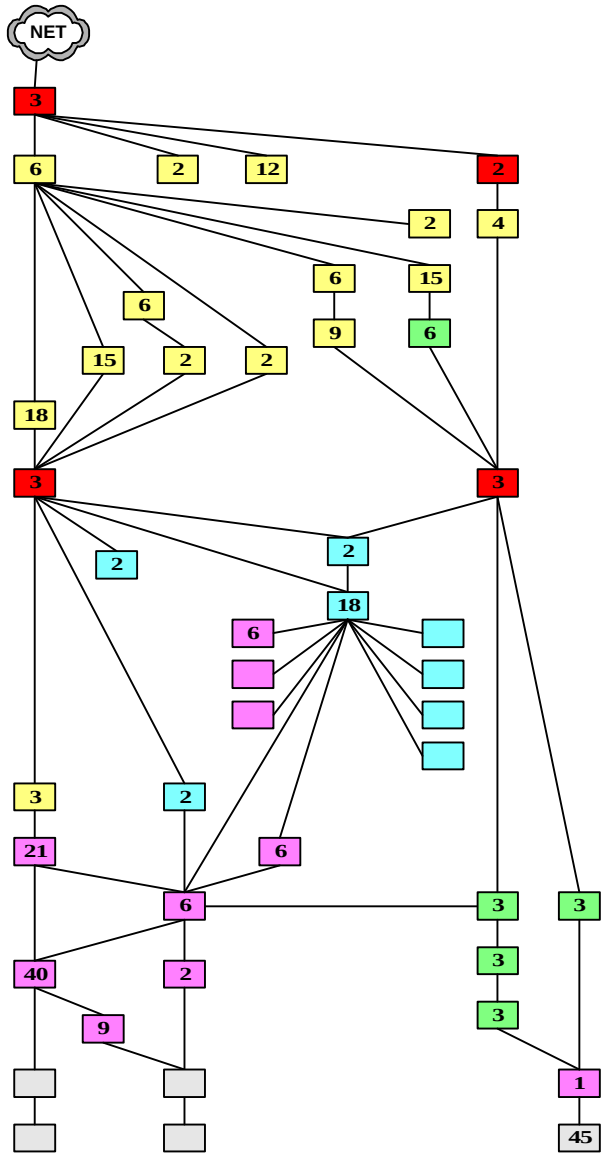


Resource Exhaustion

Fidelity Investments

- Too many active processes to allocate CPU time
- No free TCP sockets
- No free threads of execution
- No free memory or swap space
- No free disk space
- No free file descriptors
- No free network bandwidth
- Too many scheduled processes

Prototypical Tiered Computing Model



Performance & Capacity

Fidelity Investments

- **Little's Law**

  n = number of customers in the system (in flight request)

  t = mean time customers are in the system (response time)

  r = customer arrival rate (new requests)

- **Max number of queued in-flight requests**

  n = t * r

- **Max number of new requests per second**

  r = n / t

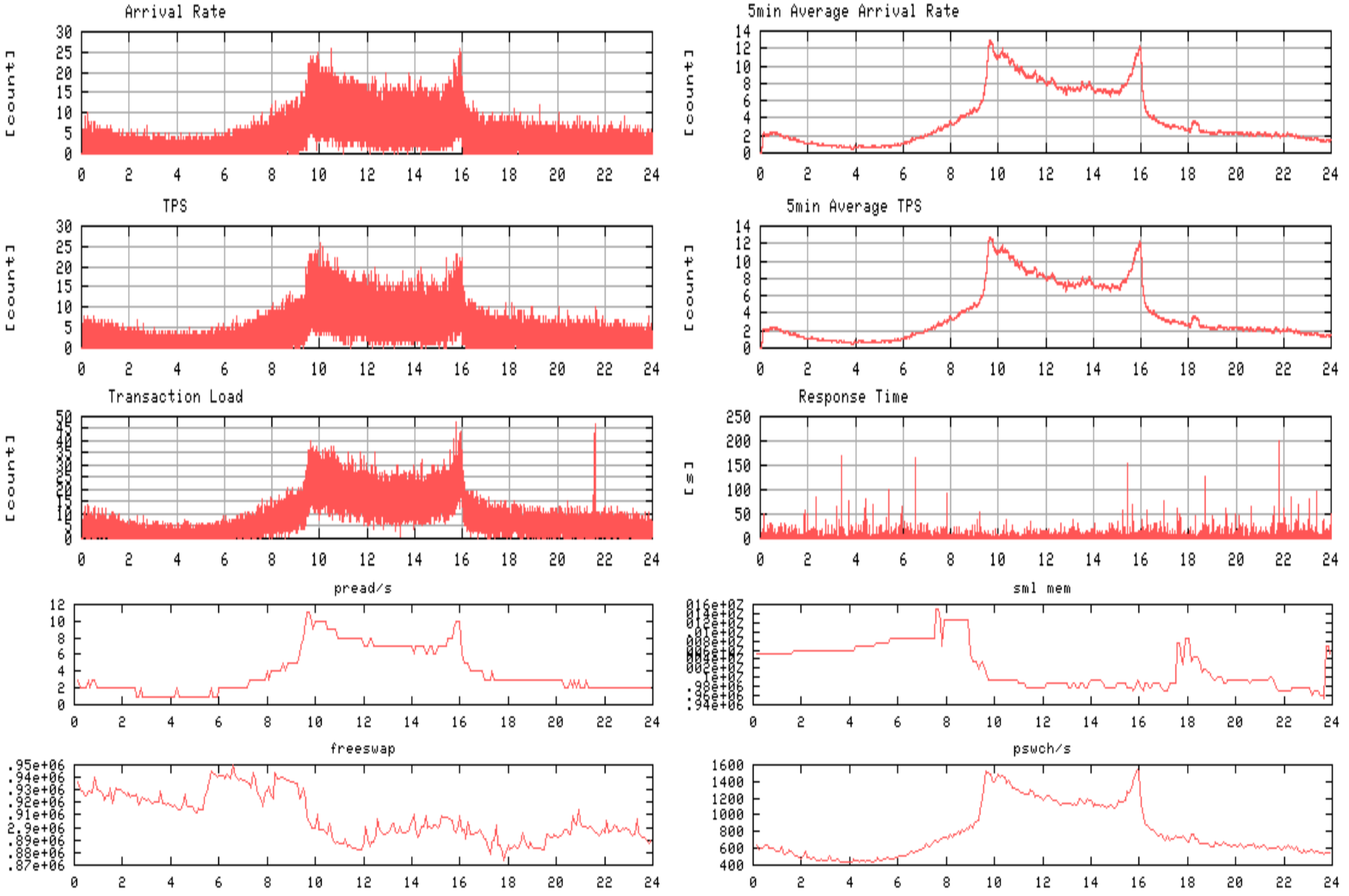
- **Moral**

  Given a finite set of resources, the slower the response time, the smaller the queue of customers that can be supported

- **Choice**

  Use fewer resources versus provide more resources

Typical User and System Histograms



TPS Requirements Model

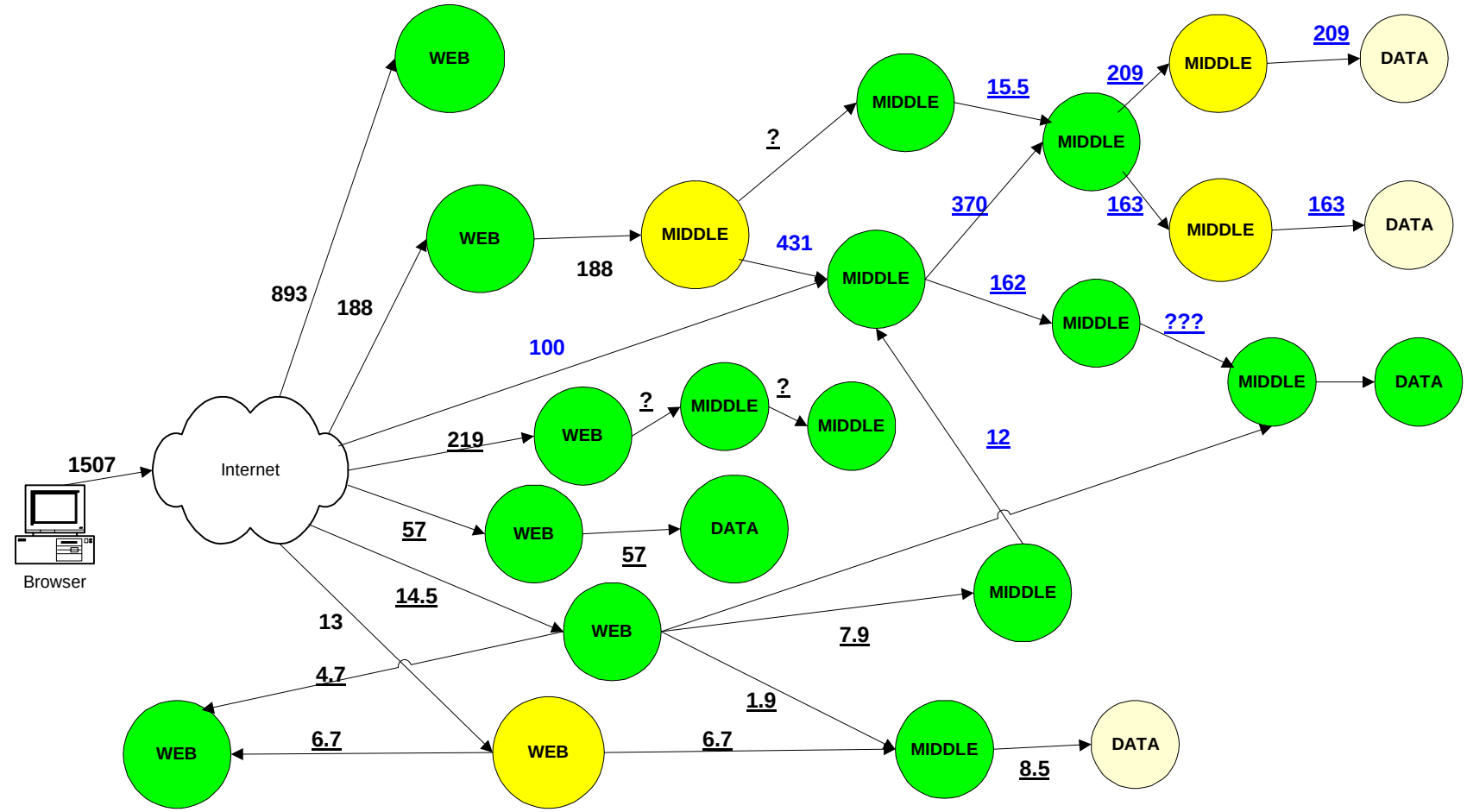
Fidelity Investments

- Goal: Maintain 4:1 Headroom @ 0-8 Sec Delay

Server Type	Number of Servers	Current Total TPS	Total TPS Req'd @ 4:1	Per Server TPS Req'd @ 4:1	Current TPS Capacity Per Server
Web	3	32	125	41.8	420 (10:1)
Broker	3	32	125	41.8	240 (4.8:1)
Data	1	80	322	322	1353 (4.1:1)

Avg Data Resp Time (sec)	Concurrent In-Flight Requests	Concurrent Data Requests	Concurrent Data Connections (min - max)
1	125	288	125 – 288
2	250	576	250 – 576
4	500	1152	500 – 1152
8	1000	2304	1000 – 2304
10	1250	2880	1250 – 2880
15	1875	4320	1875 - 4320

Capacity Relationship Model



Middle numbers inside the bubbles represent total capacity in WEB equivalent TPS for all bubbles ahead of MIDDLE and MIDDLE VPS for all bubbles behind MIDDLE. Bottom numbers inside the bubbles represent the headroom for a load of 180 TPS and 60,000 users. Links connecting components show the TPS (black) or VPS (blue) for traffic flowing between components based on a load of 180 WEB TPS and 60,000 users. Underscored numbers are estimated.

Conclusions

Fidelity Investments

- The entire system must be available when customers want to do business on E-Commerce sites
- The success of E-Commerce sites is impacted by Performance, Capacity, Availability, Security, Supportability, Usability, Maintainability, Extensibility, Feature Set & Cost
- Do not wait until the end of the process to consider the whole problem space: do it now!
- Do not sacrifice quality and stability for speed of delivery: the customer will pay the price, and will not return to your site!
- There is no better qualified person than you to work through and solve E-Commerce stability solutions

Reading

Fidelity Investments

- Little laws for utility processes and waiting times in queues
  By Serfozo, R.F.
  Queueing Systems Theory and Applications (Sept. 1994)
  vol.17, no.1-2, p. 137-81.
- RFC 1180 - A TCP/IP Tutorial
  http://info.internet.isi.edu/in-notes/rfc/files/rfc1180.txt
- Sun Performance and Tuning: SPARC and Solaris
  By Adrian Cockcroft
  280 pages; ISBN 0-13-149642-5
  http://sunsite.queensu.ca/sunmicro/sunpress/books/Cockcroft/Cockcroft.html