An Application Performance Framework

THE Open GROUP

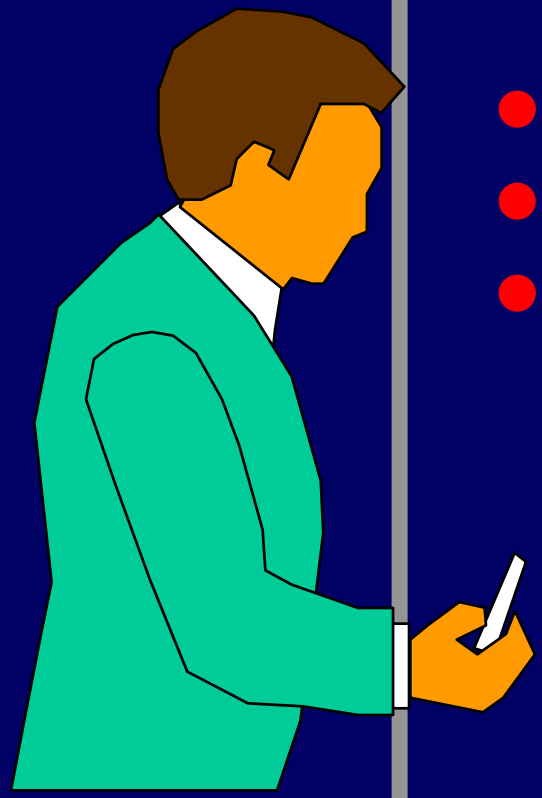
Boundaryless Information Flow Conference
Application Quality / Resource Management
Washington DC, October 23, 2003

NetForecast®

Peter Sevcik
NetForecast, Inc.
977 Seminole Trail #317
Charlottesville, VA 22901

434 249 1310
www.netforecast.com

Outline



- Information Performance Framework
- Linking Performance to Business
- Performance Metrics

Disclaimer

NetForecast®

- This is not the Open Group's Application Quality / Resource Management (AQRM) Framework
- This is a framework that NetForecast uses to help enterprises make decisions regarding improving application performance
- This is work-in-progress which means that we are constantly trying to make it better

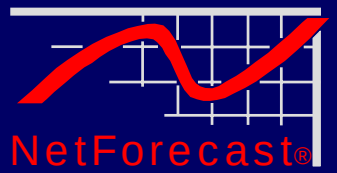
Reference
Documentation of this framework:
A Framework for Enterprise Application Performance, BCR, November 2003

Putting Performance in Perspective

NetForecast®

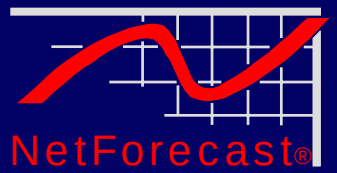
- A subset of the enterprise IT market, called Enterprise Performance Management (EPM) ensures that all the aspects of a business perform properly
  - Finance
  - Customer Relations Management (CRM)
  - Supply Chain Management (SCM)
  - Human resources
  - EPM is a $15B market with about $2B spent on software
- The question is, if EPM watches the enterprise, what watches EPM?

Everyone is in the Application Performance Game



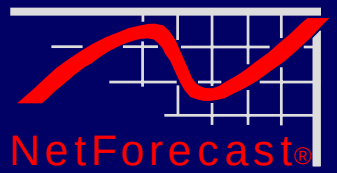
- If you think that the diet business is big!
  - A Google search on Application Performance yields twice as many hits as a search on Diet Plan
- Every vendor of information technology lays some claim to improving performance
- The “pure” performance market is large
  - 30 vendors measure
  - 70 vendors improve
- Since there are so many players, there are many views
  - Some are confusing, inconsistent, conflicting, and incomplete

Defining Application Performance



- We can take two approaches to the problem of defining performance
  - Focus the definition to one precise view and tell all the others to stop calling their view “performance”
  - Explain how all the views are parts of some larger construct and let all players have a role
- We are taking the latter path by describing a performance framework
- A framework should provide a way by which to understand the many views of performance and help guide intelligent choices

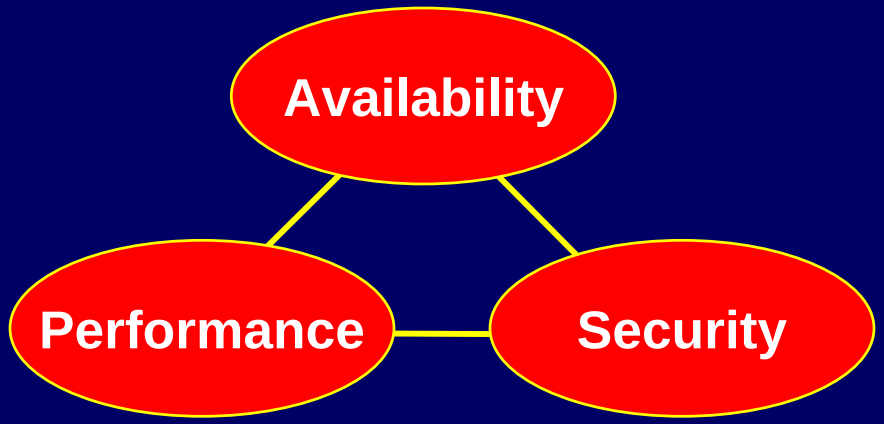
The Goals of this Application Performance Framework



- **Comprehensive**
  - Try to cover all the aspects of performance
  - Each major aspect is called a performance function
- **Clear**
  - Define the performance functions without using the word “performance”
  - The sum of all functions equals performance
- **Uniform**
  - Define metrics in each function that are appropriate to that function
  - However, normalize the metrics so they can be compared across functions
- **Useful**
  - Make sure that the different needs of different applications are accounted for
  - Define metrics and methodologies that can help make practical decisions
- **Valuable**
  - Insure that information technology is supporting the business
  - Have the methodology show the linkage of performance to business goals

Management: Three Legs of a Stool

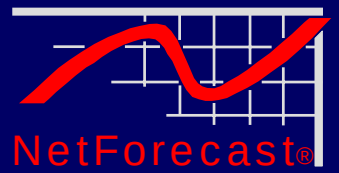
NetForecast®



- Availability – Making sure the system is working
- Performance – Making sure it is working properly
- Security – Making sure it is safe

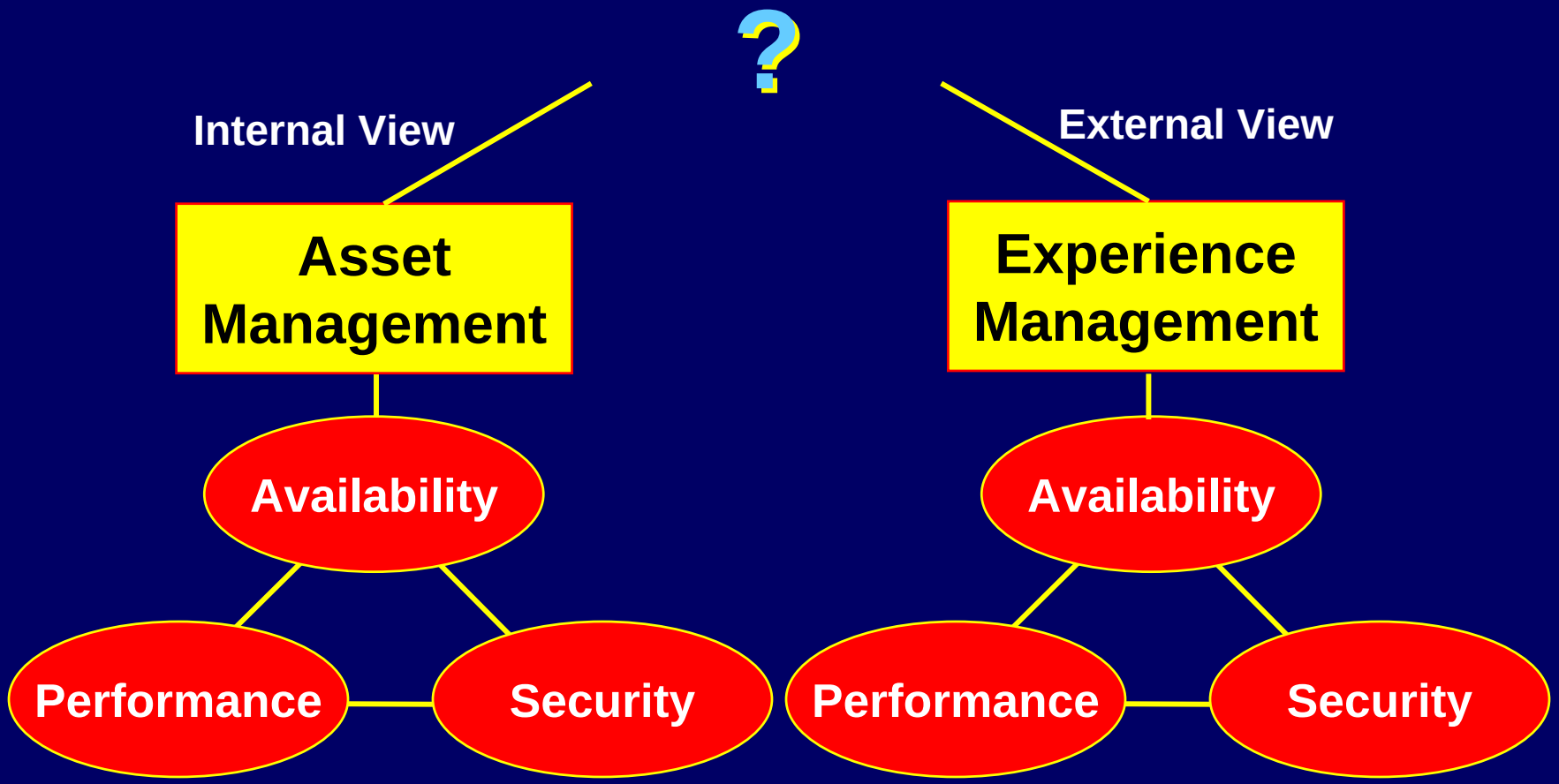
- Each of these goals conflicts with the other two
- All are needed as a unified solution to business goals

Two Major Objectives for Network Business System Management



- **Asset Management (managing the delivery assets)**
  - Link the system goals to the internal needs of the business
  - Insure that the budgets are met
  - Get the best efficiently out of the assets
  - Control or decrease costs
- **Experience Management (managing user experience)**
  - Link the system goals to the external needs of the business
  - Insure that users are happy
  - Provide the highest value to the users
  - Maintain or increase revenue

Current Broken Framework



Using the same term for the key functions has led to confusion!
Performance is also re-used as an aspect of availability and security

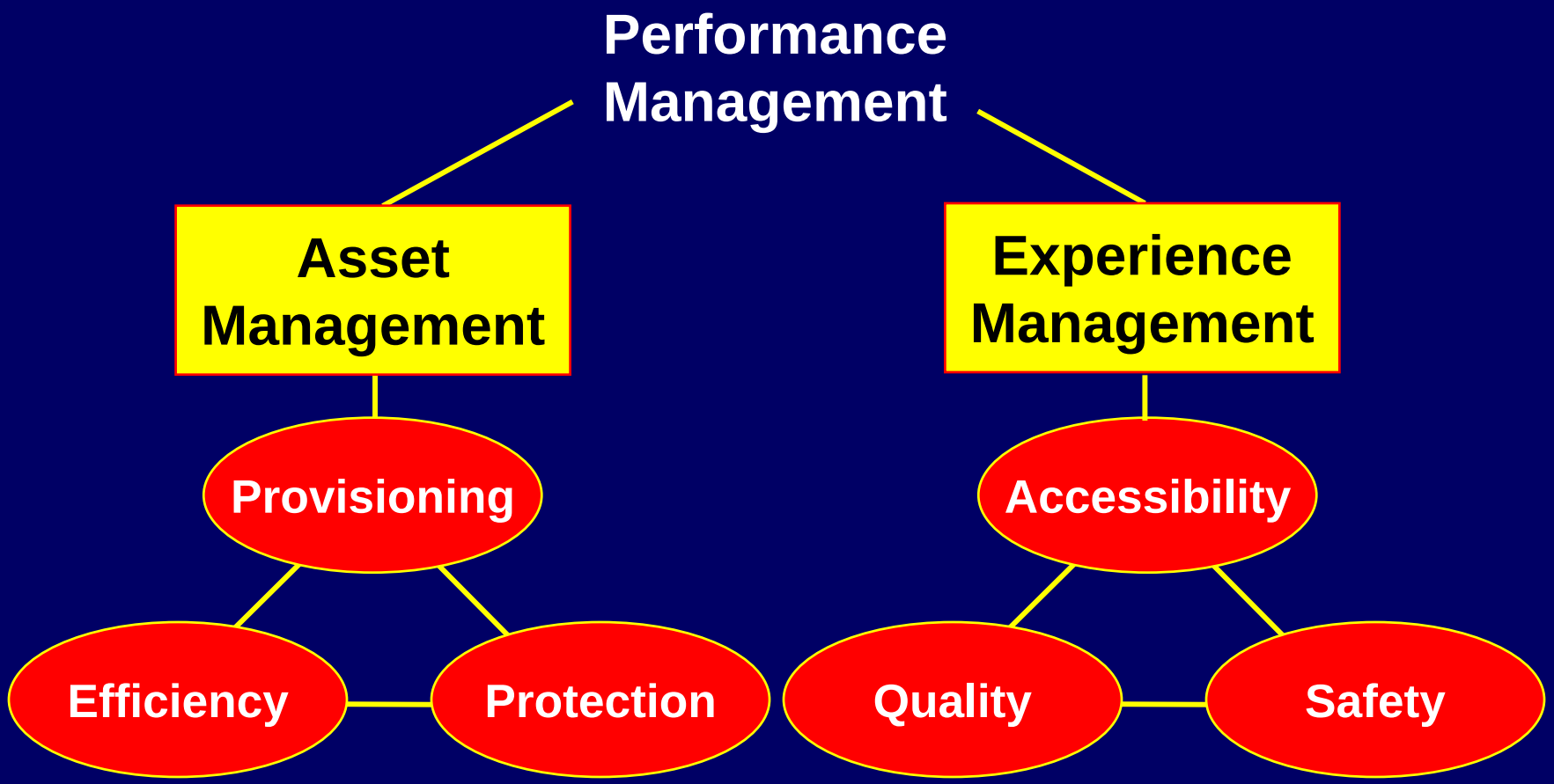
Significant Differences

- Asset Management
  - Availability – All of the assets are working
  - Performance – Efficient utilization of the assets
  - Security – Making sure the system is safe from attack
- Experience Management
  - Availability – Users have access to the resources they need
  - Performance – The users have a quality experience
  - Security – Making sure that the users are safe

Better Terms

- Asset Management
  - Provisioning
  - Efficiency
  - Protection
- Experience Management
  - Accessibility
  - Quality
  - Safety

Performance Framework



Asset Performance Functions

NetForecast®

- **Provisioning**
  - The ability of the system to establish new service or recover failed service
  - Discovery, topology maps, alarms, uptime, routing stability, fail-over
- **Efficiency**
  - The ability of the system to make the best utilization of the assets that provide the service
  - Aggregate traffic, Utilization of key components, users per server, users per Mbps bandwidth
- **Protection**
  - The ability of the system to protect itself from malicious or unauthorized use that would degrade the asset's effectiveness
  - Firewalls, DOS Protection, logging, VPN

Experience Performance Functions

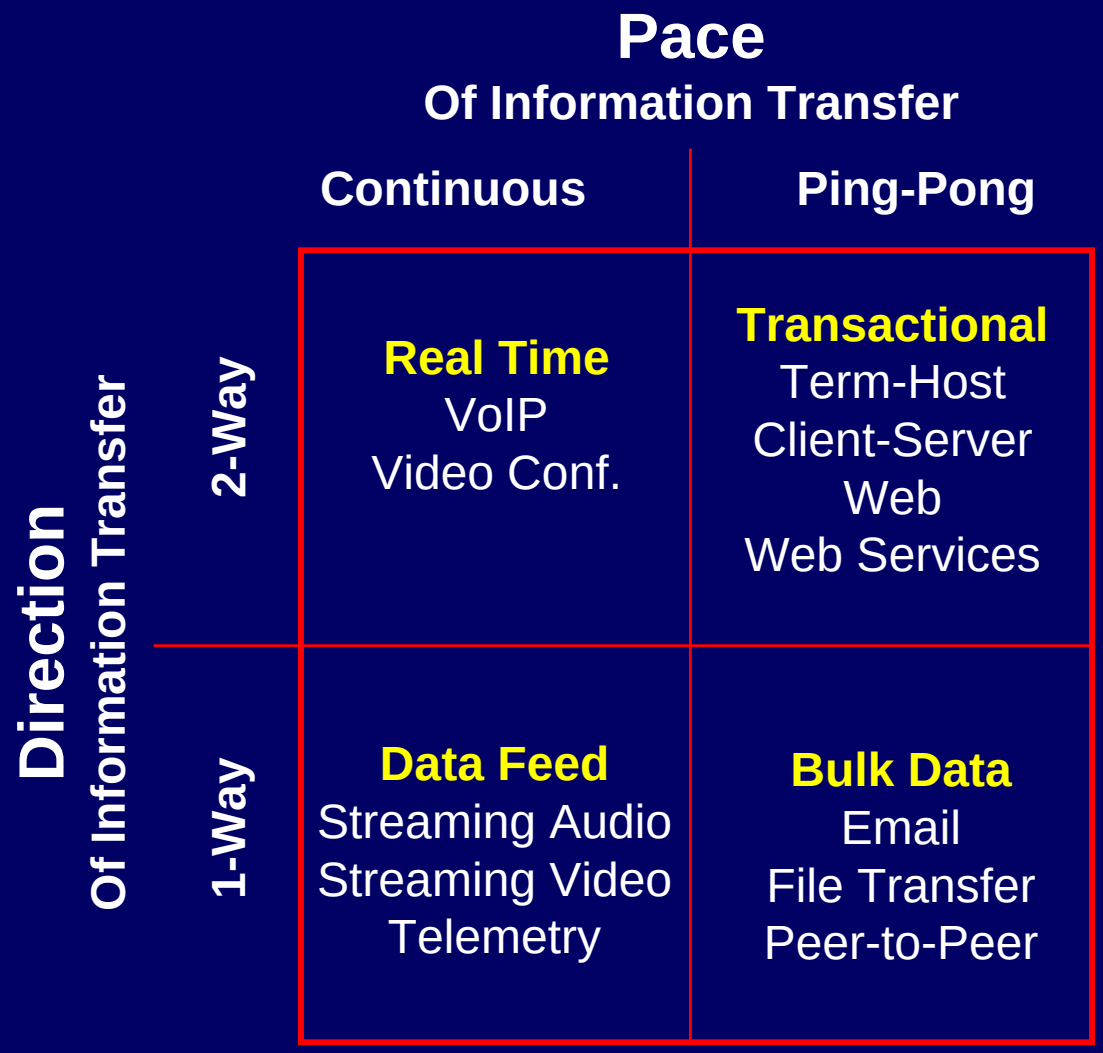
- **Accessibility**
  - **The ability of the system to provide access to its authorized users**
  - **Local access, Connectivity, Filtering effects**
- **Quality**
  - **The quality of the technical aspects of the user's experience with the system**
  - **Transactions: Response time, Voice: MOS**
- **Safety**
  - **Level of comfort the user has when interacting with the system, for example**
    - Spam interferes with the email experience
    - Popups interfere with the application experience
    - Viruses disable the user's machine
    - Keystroke recording changes the user's behavior
    - Privacy concerns hinder the user's ability to enter data
  - **Privacy, Identity protection, credit protection, anti-spam, anti-virus, pop-up blocking**

The Other Big Difference

NetForecast®

- **Asset Management requires no application input**
  - **The best ratings occur when there are no users!**
  - **If all the boxes are working and connectivity exists, then the system is operating properly**
  - **The fundamental assumption is that if the system is working properly then all the users that want service are getting service**
    - No need to ask the users if they agree
- **Experience Management is application specific**
  - **Any rating must relate to an end-user's view of the application**
    - Requires defining application tasks and user group as a pair
  - **If end-to-end task response time is under a target value, then the user will have a satisfactory experience**
    - Requires flow-oriented measurement
  - **Any report of performance quality must be differentiated by user group and application**
    - Talking to users is essential

The Application Space

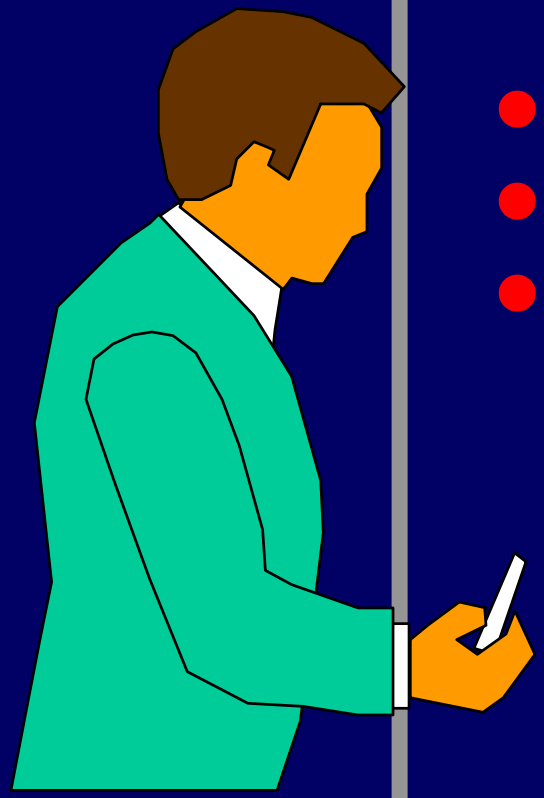


The Performance Framework

NetForecast®

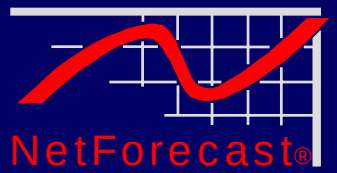
		Asset Management			Experience Management		
		Provisioning	Efficiency	Protection	Accessibility	Quality	Safety
Real Time	Voice over IP						
	Video Conference						
Transactional	Terminal-Host						
	Client-Server						
	Web						
	Web Services						
Data Feed	Streaming Audio						
	Streaming Video						
	Telemetry						
Bulk Data	Email						
	Peer-Peer						
	File Transfer						

Outline



- Information Performance Framework
- Linking Performance to Business
- Performance Metrics

Network Business Systems

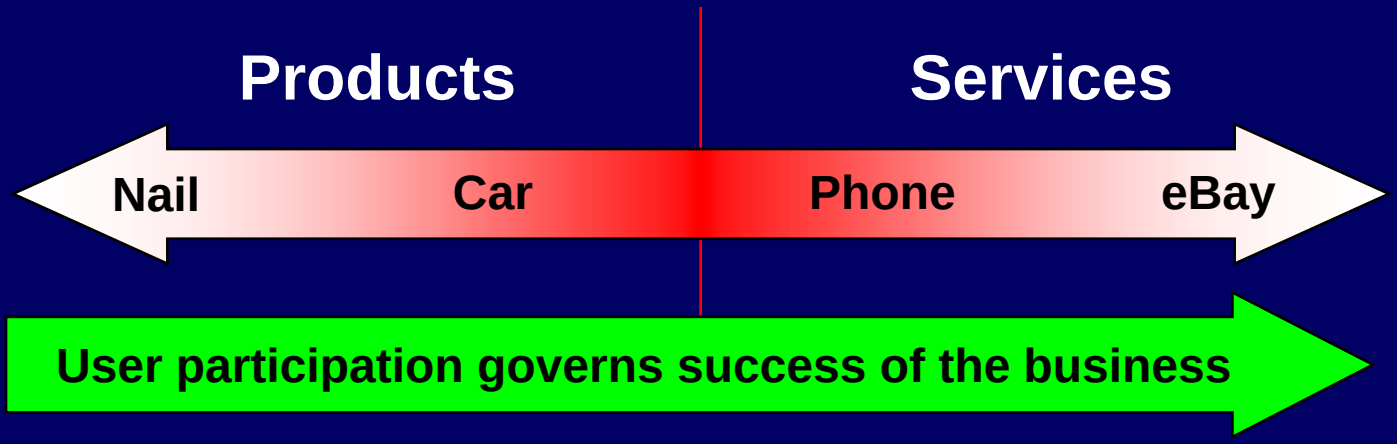


- **Definition: a business that operates over a network**
  - **The business can be partially or completely delivered on the network**
    - GE customer support for a refrigerator
    - Amazon.com
    - Schwab online trading
  - **Users can be employees or customers**
  - **The network can be internal or external (Internet)**
- **Key characteristics**
  - **Distributed nature of the system makes it complex**
  - **Many computers and devices involved**
  - **Many administrations run separate groups of computers and devices**
  - **Geographically spread out**
  - **Hard to envision the complete system**
  - **The system delivers a service of a service in support of a product**
  - **The consumers of the service are often separated from the service providers**
  - **The user is an integral part of the system**

The Unique Aspect of a Network Business

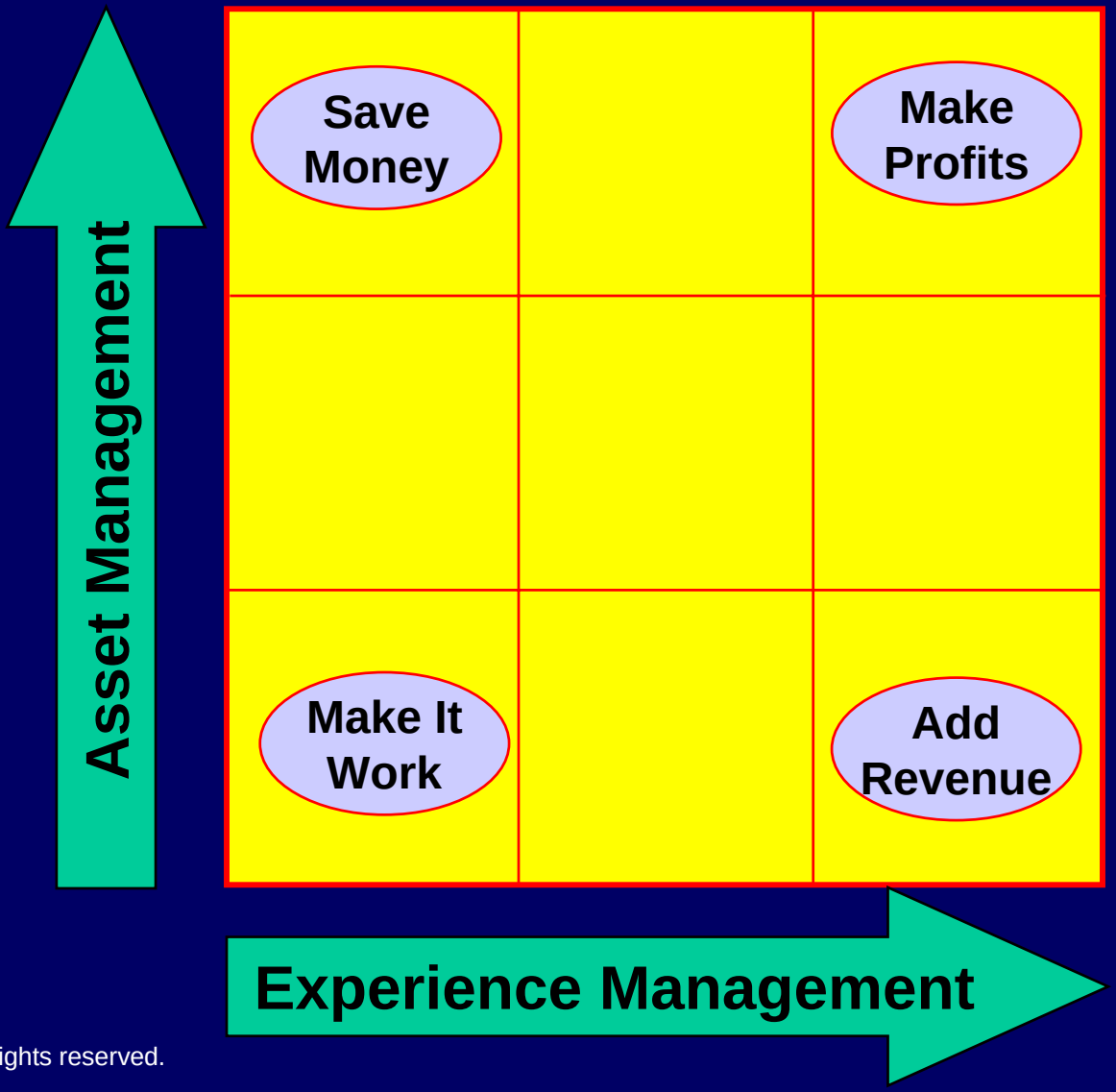
NetForecast®

- The users are not just customers they are participants in the business
- Good user experience is key to success

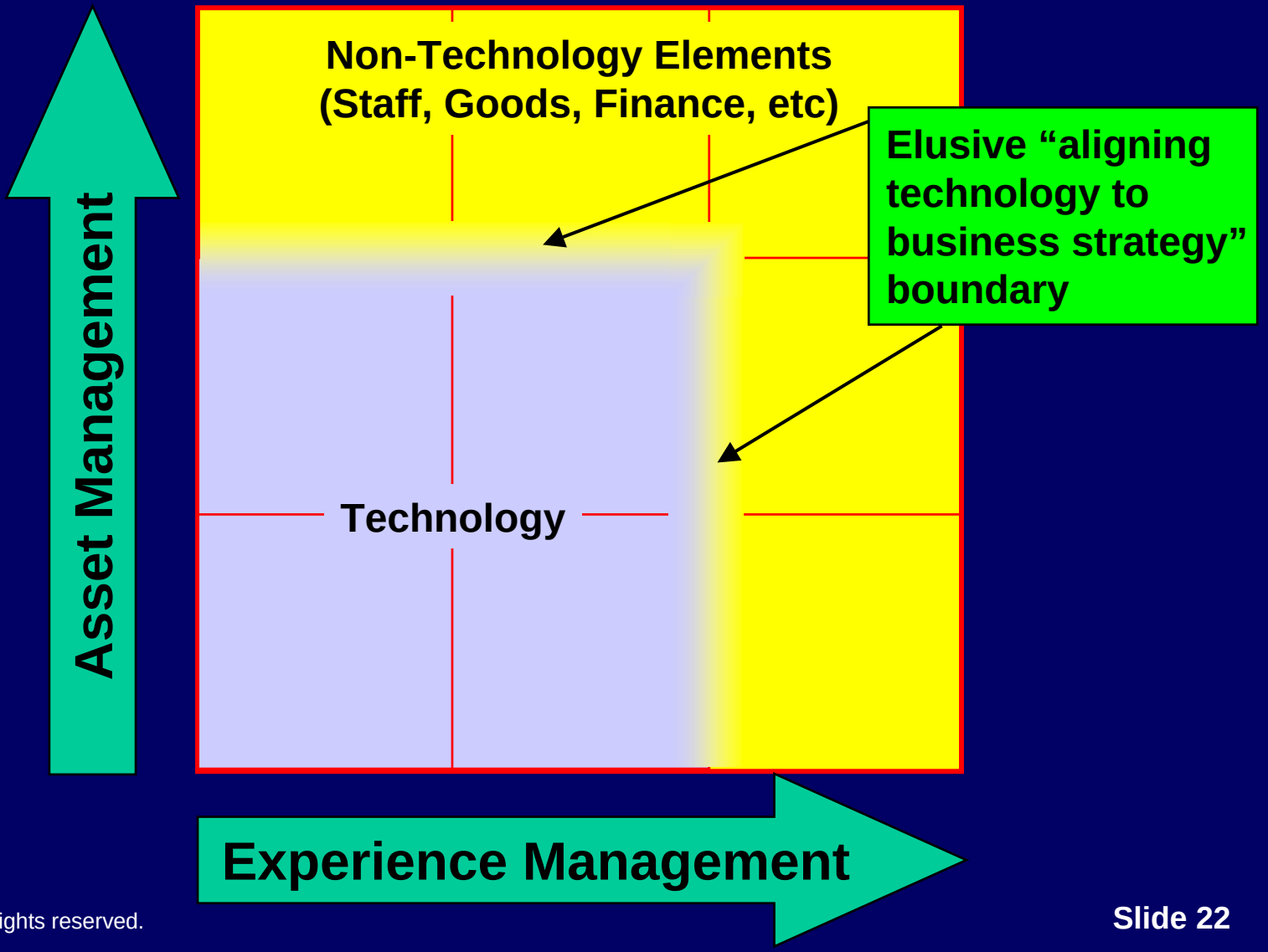


The Two Sides of Business Management

NetForecast®

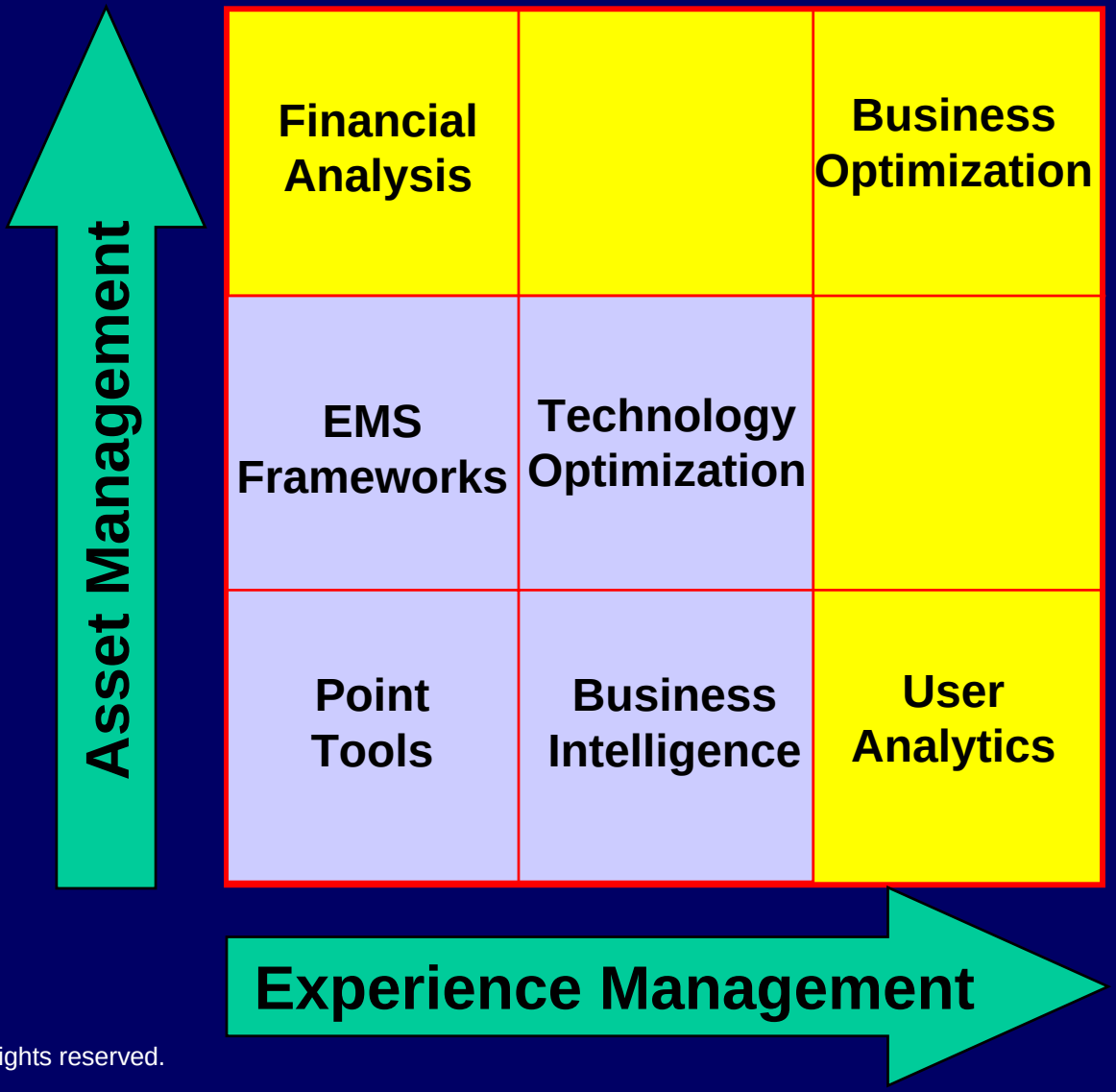


The Performance Management Framework

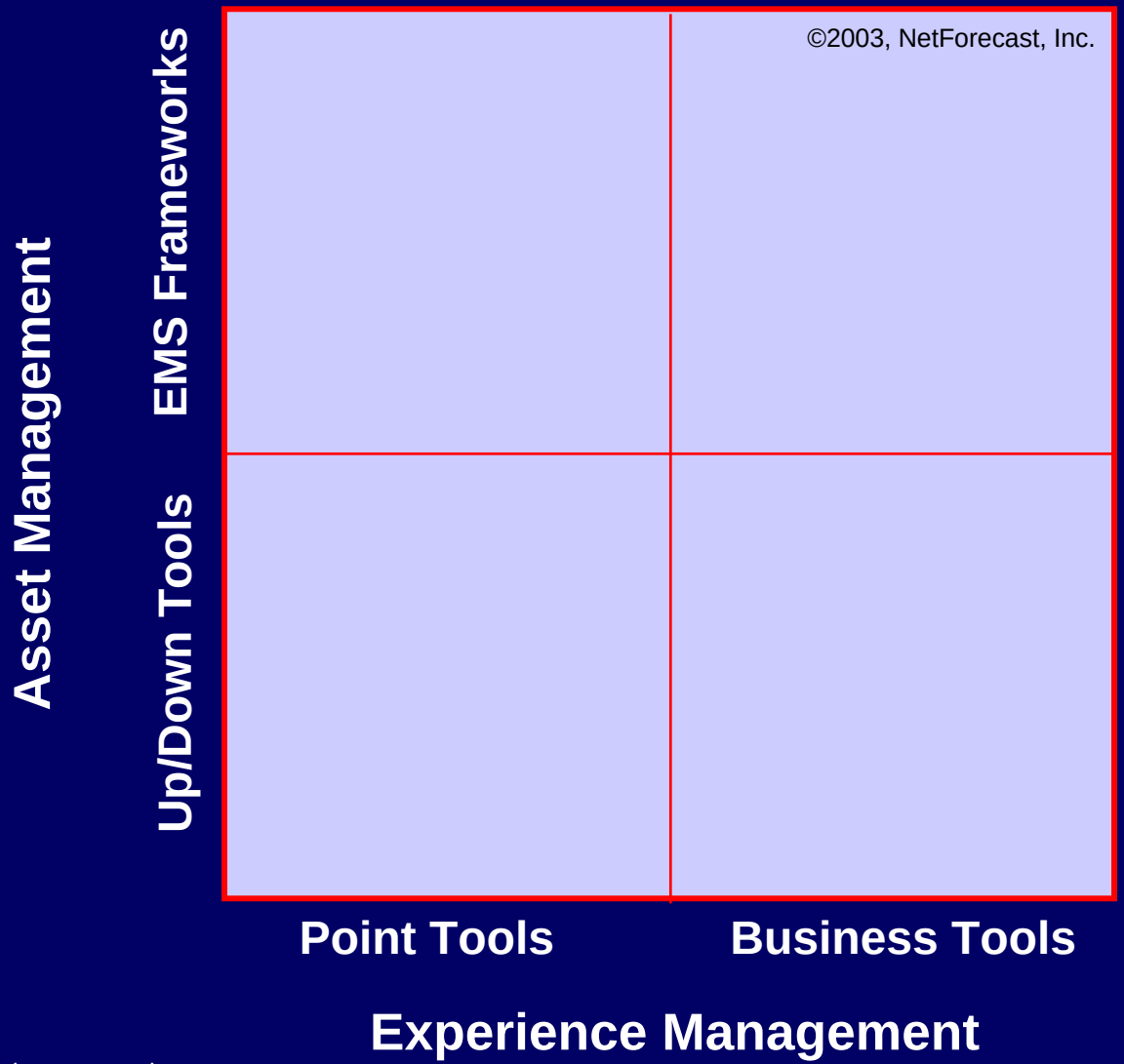


The Performance Management Framework

NetForecast®



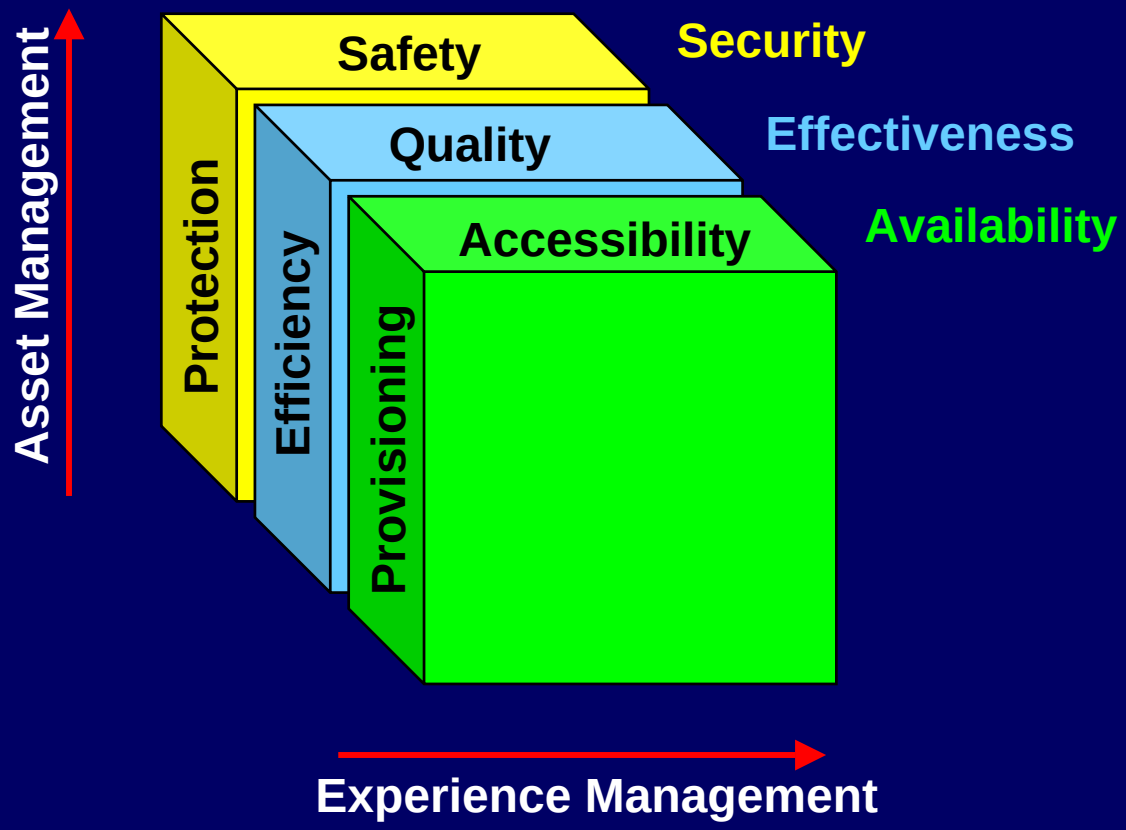
Focusing on Managing Technology



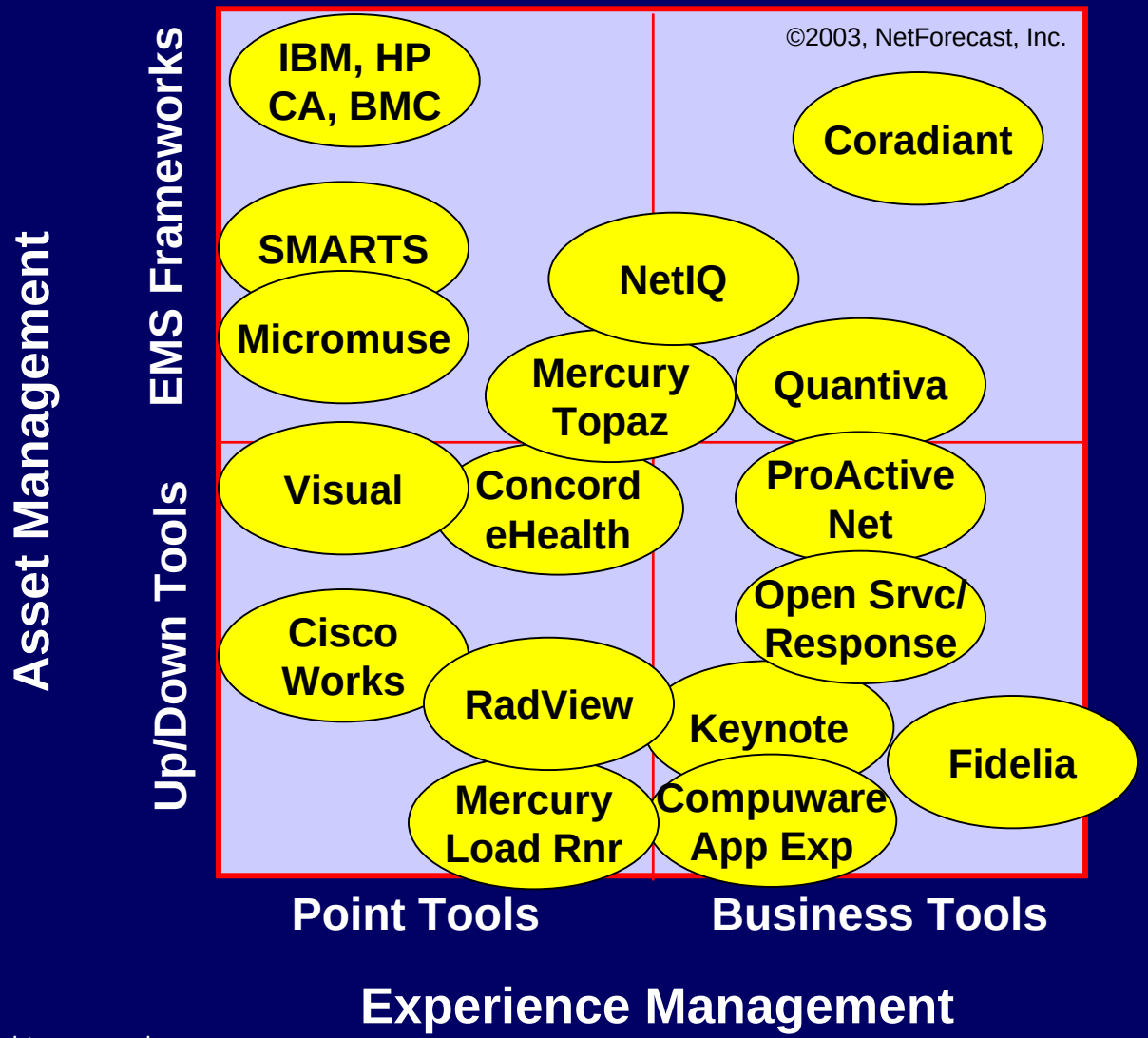
Multiple Dimensions of Performance

NetForecast®

The Performance Management Cube

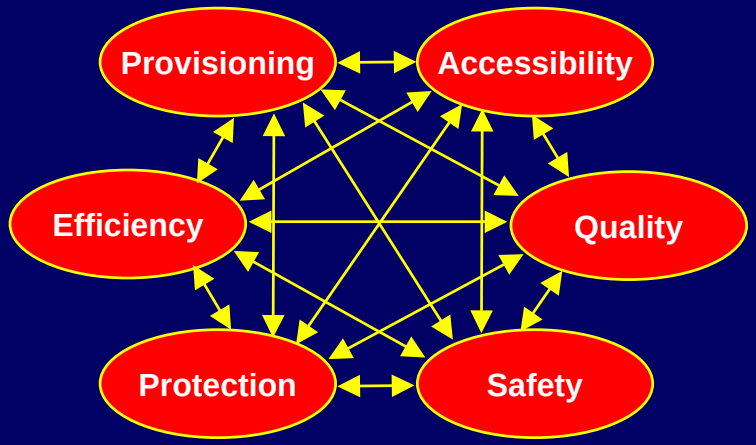


Some Vendors for Illustration

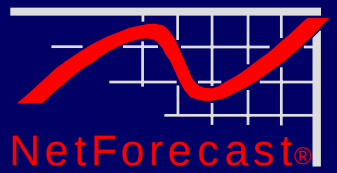


The Next Business-Technology Challenge

- Optimizing the whole is a complex challenge
  - There are 30 interdependencies (arrows) in the framework!
  - This is the next research challenge
- How to manage the relationships
  - Each function needs a uniform metric
  - From 0 (fail) to 1 (perfect)
  - Optimizing performance means balancing the metrics
- All the functions interact with each other
  - Improving one may hurt another



Complete Performance Optimization



- Imagine the following scenario
  - Improve enterprise asset security (protection) with a new device
- Applying the performance framework
  - Index values for before and after the new security device
  - For a Web application used by the enterprise partners

	Before	After	
Provisioning	.98	.95	adding another box lowered it
Efficiency	.60	.55	lowered user traffic
Protection	.80	.97	improved as planned
Accessibility	.96	.80	users have a harder time getting in
Quality	.92	.85	response time suffers (encryption)
Safety	.80	.80	users get no security benefit

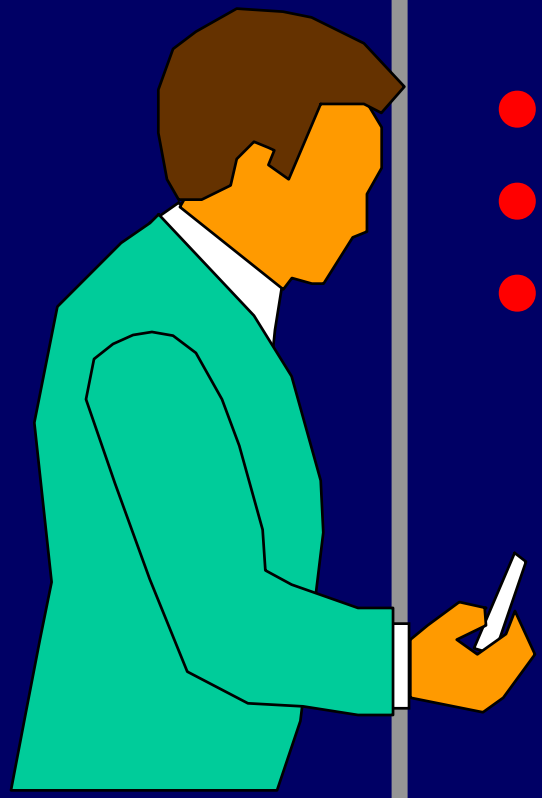
- Was this a good change for the enterprise?

Applying the Framework Today

NetForecast®

		Asset Management			Experience Management		
		Provisioning	Efficiency	Protection	Accessibility	Quality	Safety
Real Time	Voice over IP						
	Video Conference						
Transactional	Terminal-Host	*Impact of a new Web content compression device*					
	Client-Server						
	Web	–	+	0	–	+	0
	Web Services						
Data Feed	Streaming Audio						
	Streaming Video						
	Telemetry						
Bulk Data	Email						
	Peer-Peer						
	File Transfer						

Outline



- Information Performance Framework
- Linking Performance to Business
- Performance Metrics

Defining the Metrics

NetForecast®

- **From the goals for this process, the metrics must be uniform and useful**
  - **Each metric can start with unique attributes but must be normalized into a scale of 0 to 1**
    - 0 = complete failure
    - 1 = perfection
  - **The metrics must clearly state which part of the framework they are measuring**
- **How to proceed**
  - **Some metrics can be leveraged from accepted industry practices**
    - But normalization is still required
  - **Some metrics require new definition**
    - This is the focus of current research

Leveraging Existing Methods

NetForecast®

- **Provisioning: Up-time or the traditional system availability percentage**
  - Provisioning = min (availability of each asset used by a specific application), represented on 0-1 scale
- **Efficiency: Aggregate of server and bandwidth utilization percentages**
  - Efficiency = ave (utilization of each asset used by a specific application), represented on 0-1 scale
- **Protection: Converting actuarial risk assessment into an index of 0-1**
- **Accessibility: Very little is known about the probability that the user population has connectivity to the service when they need it**
- **Quality:**
  - **Voice has a good metric called MOS that must be converted to 0-1**
  - **Transactions: user's task response time**
    - Acceptable response time is a highly variable and not bound to a 0-1 scale
    - We propose a new Uniform Response Time Index (URTI):
    - URTI = (satisfied users + ½ tolerating users)/total user population
    - Frustrated users are not in numerator but are counted in total user population
- **Safety: Open research topic**

Where the Uniform Response Time Index (URTI) Applies

NetForecast®

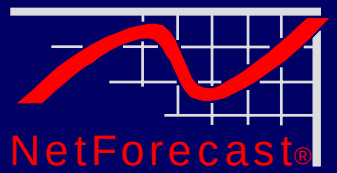
		Asset Management			Experience Management		
		Provisioning	Efficiency	Protection	Accessibility	Quality	Safety
Real Time	Voice over IP						
	Video Conference						
Transactional	Terminal-Host					URTI	
	Client-Server						
	Web						
	Web Services						
Data Feed	Streaming Audio						
	Streaming Video						
	Telemetry						
Bulk Data	Email						
	Peer-Peer						
	File Transfer						

How Users View Application Task Performance

NetForecast®

- **Satisfied**
  - User maintains concentration
  - Performance is not a factor in the user experience
  - User “budgets” time per element he will process (read, enter, reply)
  - Satisfied threshold is determined by two parameters: number of elements and interaction repetitiveness
- **Tolerating**
  - Concentration is impaired
  - Performance is now a factor in the user experience
  - User will notice how long it is taking
- **Frustrated**
  - Performance is typically called unacceptable
  - Casual user may abandon the process
  - Production user is very likely to stop working

How Users Set Their Performance Expectation



- **The NetForecast user behavior model**
  1. **A user is prepared to invest time to receive information from a computer in direct proportion to the time he or she will spend processing that information**
  2. **The user has a preset expectation for their processing time when he or she requests the data**
  3. **There are two factors that determine how much time the user puts into the preconceived personal data processing budget:**
     - Repetitiveness of the session or process
     - Interest: number of objects, fields, sentences the user will notice or read
- **Model functions**
  - **Satisfied Task Response Time = f(Personal Processing Budget)**
  - **Personal Processing Budget = f(Repetitiveness,Interest)**
  - **Therefore Satisfied Task Response Time = f(Repetitiveness,Interest)**

Counting Interest Elements

NetForecast®

- One
  - Simple check box
  - One data entry field: enter part number
- Few
  - Select among the following options
  - Expected few lines: headers of recently arrived email
- Several
  - Type your first name, last name, address, phone number
  - Information on product, prices, shipping alternatives, etc.
    - The user will typically only be interested in a few of these information fields, do not assume if you present 20, the user will read 20
- Many
  - Interesting report that is read
  - Scrolling down the page for more content

Rating Repetitiveness

- Very High
  - There are many short tasks to the process
- High
  - There are a few tasks to the process
- Low
  - Sometimes there are a few tasks, sometimes there is browsing
- Very Low
  - The user is browsing, there is no actual process being performed

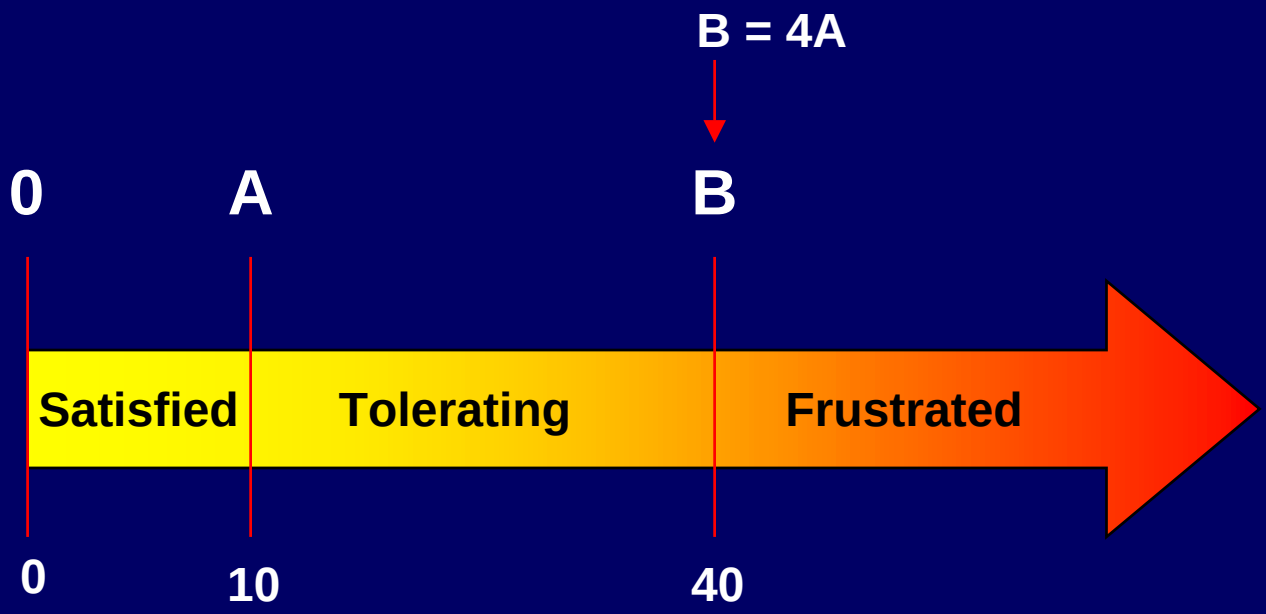
Determining the Satisfied Zone Threshold

NetForecast®

User is satisfied if task completes by X seconds

Task Repetitiveness	Number of Elements Viewed: 1	2	3	4
Very Low	4	8	12	16
Low	3	6	9	12
High	2	4	6	8
Very High	1	2	3	4

Typical Web Performance Zones



Zone boundaries for typical Web User (sec)

References

For more information on the Satisfied-Tolerating-Frustrated methodology:

Understanding How Users View Application Performance, BCR, July 2002

This Is Your Father's Performance After All!, BCR, November 2002

How Fast is Fast Enough?, BCR, March 2003

Case Study: Evaluating Several Alternatives

NetForecast®

- CRM - Customer Relationship Management
  - Highly interactive sessions supporting telemarketing
- SCM – Supply Chain Management
  - Enterprise partners interacting with product design and making commitments on their part of the process or product
- eShop – Web Commerce
  - Configuring and buying a computer on the Web
- eTrans – Web Financial Service
  - Transactions supporting the buying and selling corporate shares on the Web

Reference

For more information on this case study

Application Response Time Improvements with Transparent Turn Reduction, Sevcik and Bartlett, NetForecast Report 5066, September 2003

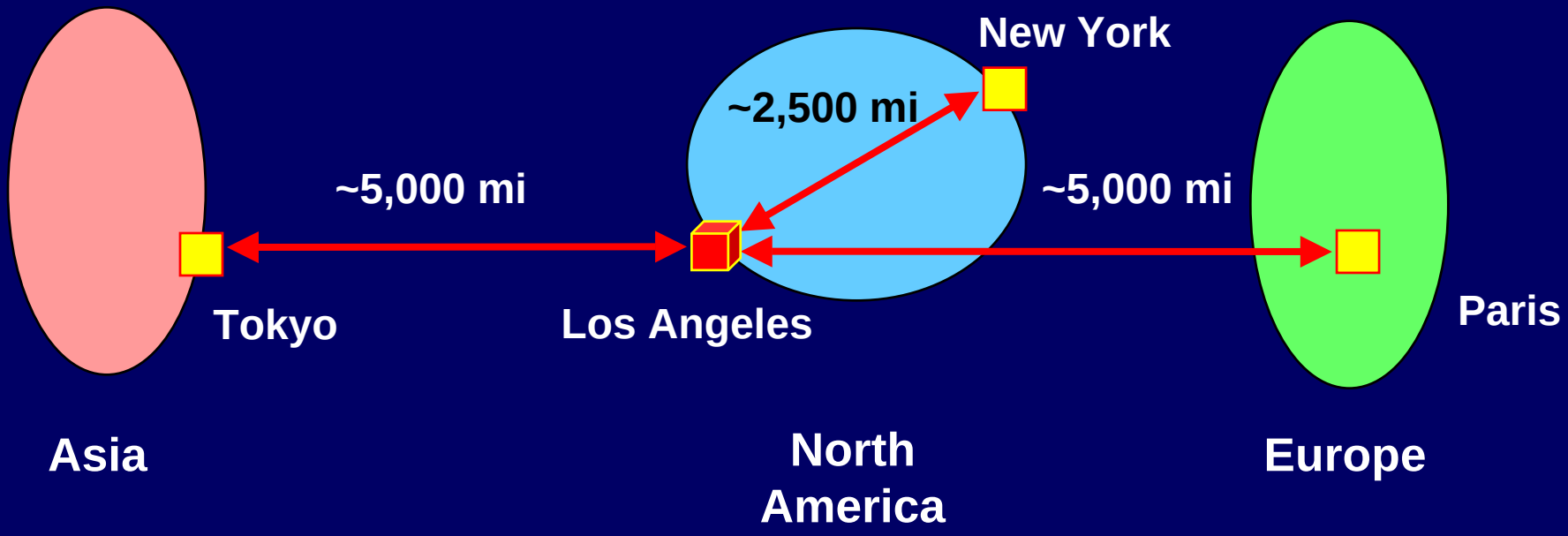
Limits of the User Satisfaction Zone (sec) by Application Class

NetForecast®

Number of Elements Viewed

Task Repetitiveness	1	2	3	4
Very Low	4	8 *Web Browsing*	12	16
Low	eTrans 3	6	9	12
High	2	SCM 4 eShop	6	8
Very High	1	2	CRM 3	4

Scenarios Modeled



What Was Modeled

- Specific performance zone thresholds (sec) for these applications

	Satisfied	Tolerating	Frustrated
SCM	<4	4-16	>16
eShop	<4	4-16	>16
eTrans	<3	3-12	>12
CRM	<3	3-12	>12

- End-to-end performance
  - Network distance for the two scenarios
  - Last mile ISP at the user
  - Broadband access for the user
- The network long-tail effect on
  - Delay
  - Loss

Reference

More information on the long tail distribution of network performance

Web Performance – Not a Simple Number, BCR, January 2003

Result: Distribution of Users by Performance Zone

NetForecast®

	At 2,500 Miles				At 5,000 Miles			
	SCM	eShop	eTrans	CRM	SCM	eShop	eTrans	CRM
Satisfied								
Direct	48%	0%	65%	0%	0%	0%	0%	0%
CDN	48%	82%	80%	0%	0%	55%	47%	0%
Comp	82%	65%	75%	0%	55%	0%	0%	0%
TTR	91%	97%	95%	65%	81%	92%	89%	0%
Tolerating								
Direct	45%	91%	28%	89%	85%	81%	85%	76%
CDN	45%	14%	15%	89%	85%	35%	42%	76%
Comp	14%	28%	19%	90%	36%	85%	87%	78%
TTR	7%	2%	3%	29%	14%	3%	7%	87%
Frustrated								
Direct	7%	9%	7%	11%	15%	19%	15%	24%
CDN	7%	4%	5%	11%	15%	10%	11%	24%
Comp	4%	7%	6%	10%	9%	15%	13%	22%
TTR	2%	2%	2%	6%	4%	4%	4%	13%

URTI for the Direct Connection

NetForecast®

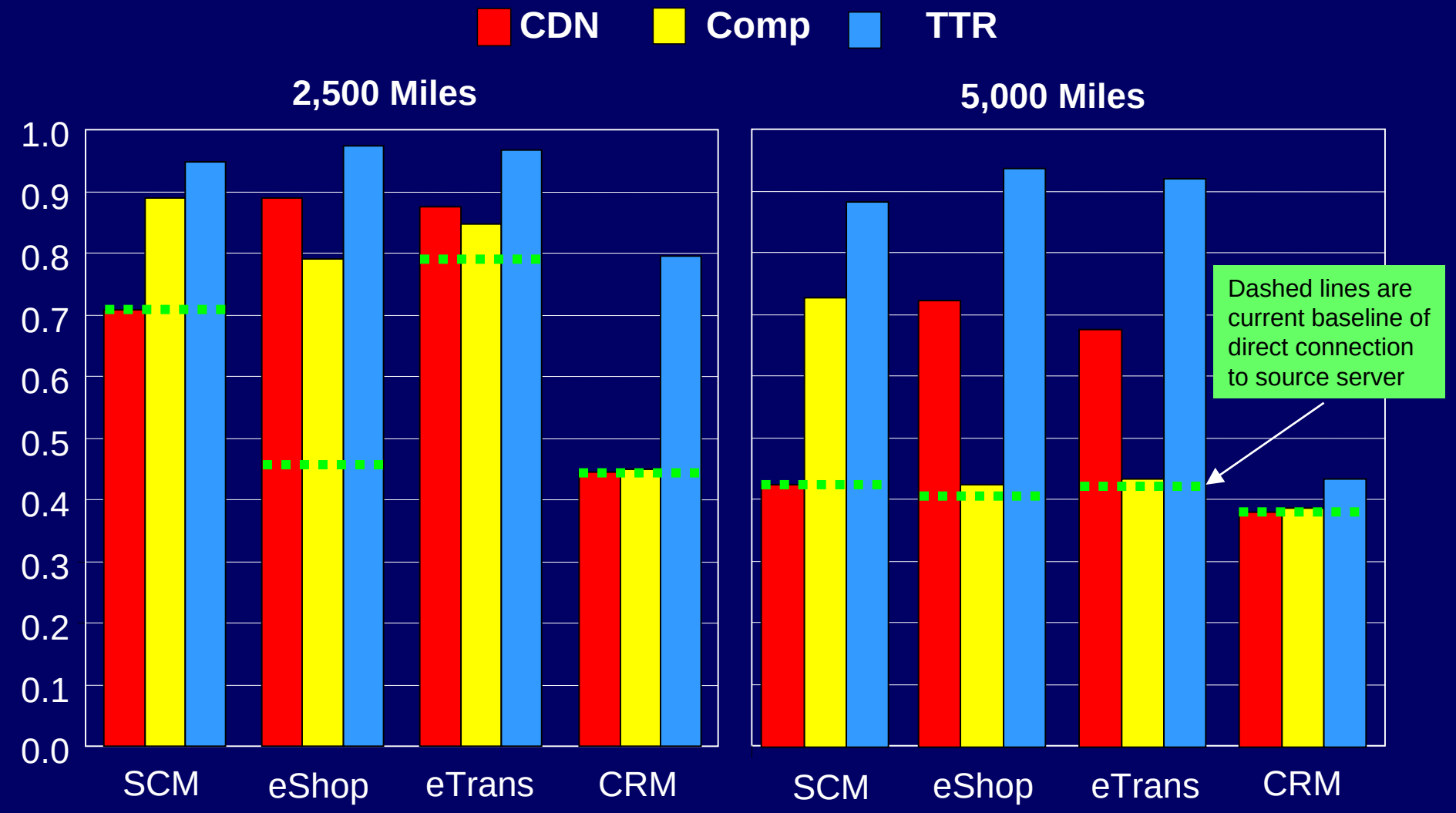
- URTI for the default direct connection to the origin server is:

- Distance of 2,500 miles
  - SCM 0.71
  - eShop 0.46
  - eTrans 0.79
  - CRM 0.45

- Distance of 5,000 miles
  - SCM 0.43
  - eShop 0.41
  - eTrans 0.43
  - CRM 0.38

Evaluating Technologies With the URTI

NetForecast®



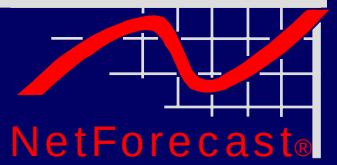
Questions

NetForecast®

1) Are the goals of the framework proper?
   - Slide 7: Comprehensive, Clear, Uniform, Useful, Valuable
2) Does the framework satisfy the goals?
   - Slide 17: Matrix of Functions and Applications
3) Do the two current metrics satisfy the framework?
   - Slide 32: Provisioning and Efficiency
4) Does the URTI metric satisfy the framework?
   - Slide 32: URTI = (satisfied users + ½ tolerating users)/total population
5) Should performance devices be classified with a simple use of the framework?
   - Slide 29 is an example.
6) Do you want to learn more or help develop the topic?

Smart Strategies From Hard Data

Thank You



Visit our Web site for the references cited and additional information:

www.netforecast.com