

UNITE 2009

Managing Service Delivery, Workloads, and Capacity Burn on Metered and Non-Metered Systems

UNITE Conference
Minneapolis, MN
11 November 2009

Session GE4028
Guy Bonney
Bob Morrow

MGS, Inc.

- Software Engineering, Product Development & Professional Services firm founded in 1986
- We solve business problems with:
 - Products: SightLine™, CheckOut, MGSWEB Web Services, Deliver, C.A.T.T. , SecureCATT, and others
 - Professional Services
 - ❖ IT Management Planning
 - ❖ Capacity Planning and Management
 - ❖ Consulting and Technical Services including Performance Management and Hardware-Software-Network Integration
 - ❖ Application Development Services including Java/J2EE development and platform rehosting
 - ❖ Training Services
 - Software Engineering Services on ClearPath MCP, Windows, and UNIX platforms.

Agenda

- Concepts
- Defining Workloads
 - Identify
 - Characterize
 - Establish Service Levels
- Managing Service Delivery
- Managing Capacity Burn

Service Delivery Concepts

- Whether using ITIL®, COBIT®, ISO/IEC 20000, Quality Management, or other integrated process framework they all include areas similar to the following:
 - Service Level Management
 - Financial Management of IT Services
 - Capacity and Performance Management
 - IT Service Continuity Management
 - Availability Management
- This session discusses service level management and capacity management

What is Service Delivery?

- Delivery of defined IT services to requestors (typically users) at:
 - an expected performance level,
 - with expected reliability and availability, and
 - within budget constraints.

What is Performance?

- Meeting Expectations.
- Complete Processing a Unit of Work in the Expected Time.
 - On-line transactions - Response time
 - Batch runs - Elapsed time/Deadlines
 - “On-Line Batch” - Elapsed (Response) time
- Processing work requires Capacity.

Capacity Defined

- Ability to process work
- Maximum work that can be done while meeting Performance objectives (service levels)
- Performance cannot be achieved without sufficient Capacity to process the Work
- Performance and Capacity are inextricably linked

Managing Service Delivery

- Know the Workload(s)
- Know the Service Requirements (performance)
- Know the Demand Pattern
- KNOWLEDGE ENABLES:
 - Planning for Capacity needs
 - Managing Services to Budget

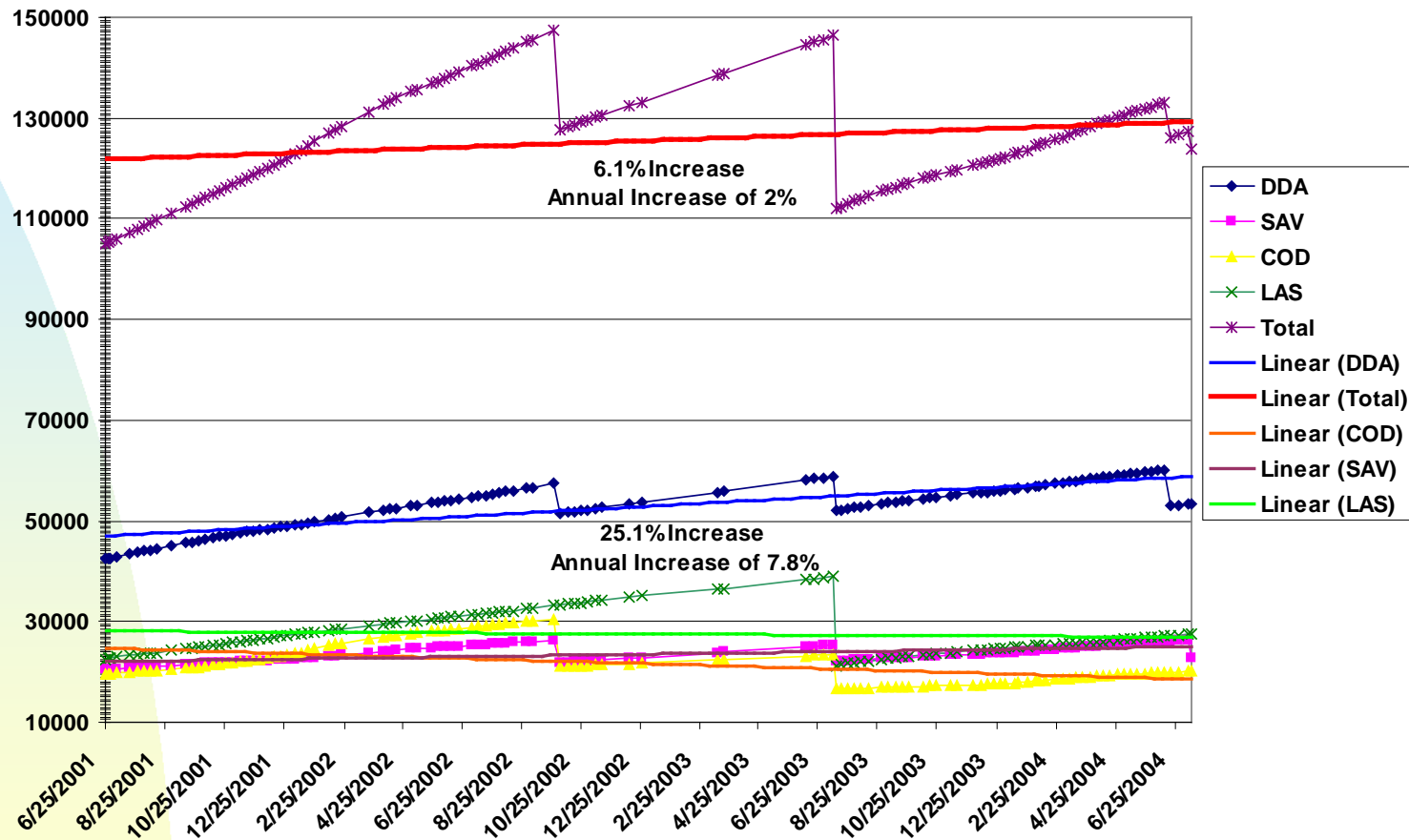
Define and Characterize Workloads

- Workload Identification
 - Identify programs associated with a business function:
 - ❖ Payroll
 - ❖ Accounting (A/R, A/P, G/L)
 - ❖ Order Entry
 - ❖ Outpatient Admissions
 - Determine whether there are on-line and batch elements and when they are active on the system
 - For on-line workloads:
 - ❖ Identify COMS programs, TIP transactions, or IIS entities
 - For batch workloads:
 - ❖ Identify major batch processes
 - ❖ Is it a single-threaded process or parallel processing?

Define and Characterize Workloads

- Equate business volume indicators to identified workloads
 - Changes in the business environment affect services
 - Business volumes are natural business units (NBU's)
 - ❖ Orders processed
 - ❖ Occupied hospital beds
 - ❖ Outpatient admissions
 - ❖ Student registrations
 - System transaction volumes are natural forecasting units (NFU's).
 - ❖ Inquiry, Update, or Order Entry transactions
 - ❖ Admissions, patient accounts, and discharge inquiry and update transactions

Business Volume History



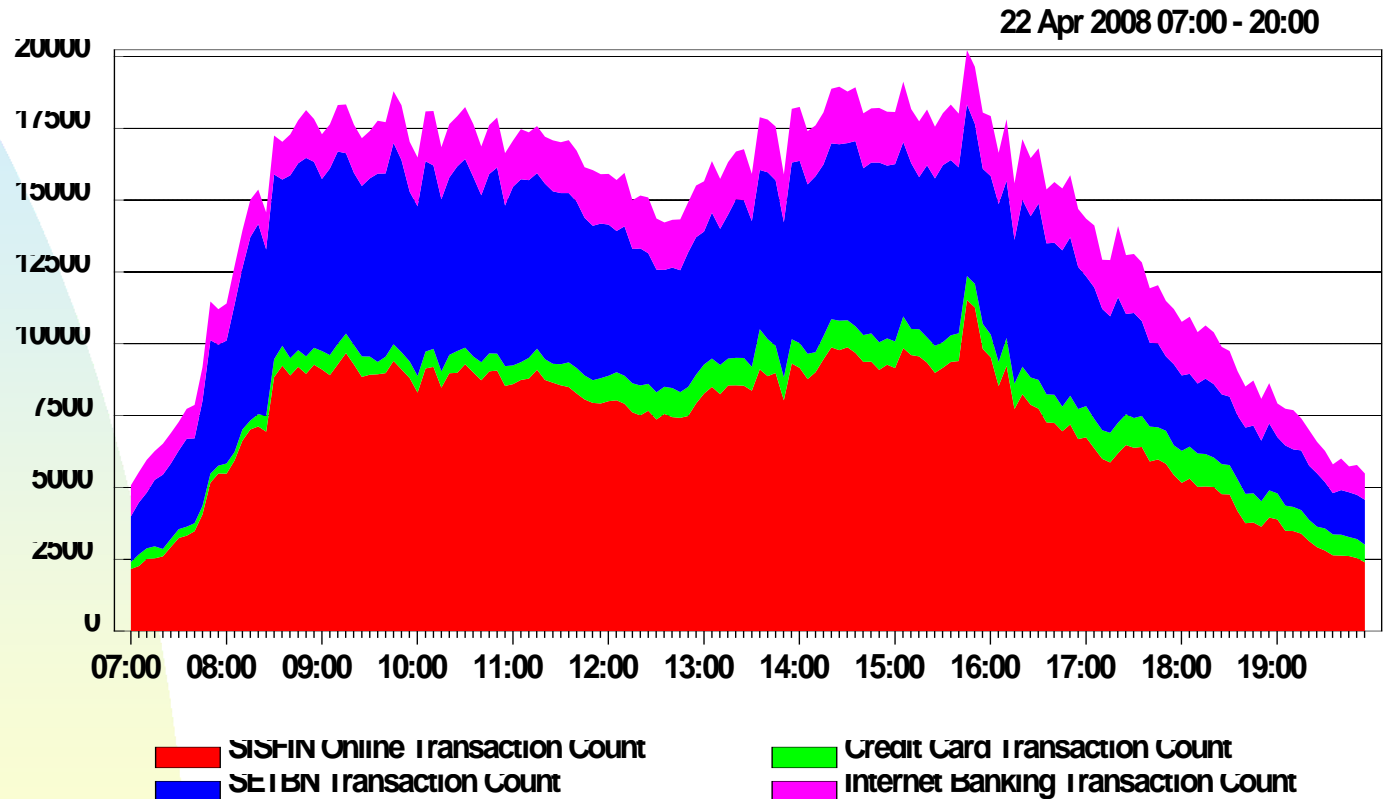
Define and Characterize Workloads

- Identify and document service level indicators for each principle workload
 - Number job requests processed
 - Required end-user response time
 - Report deadlines
 - Service errors
 - Number of service requests

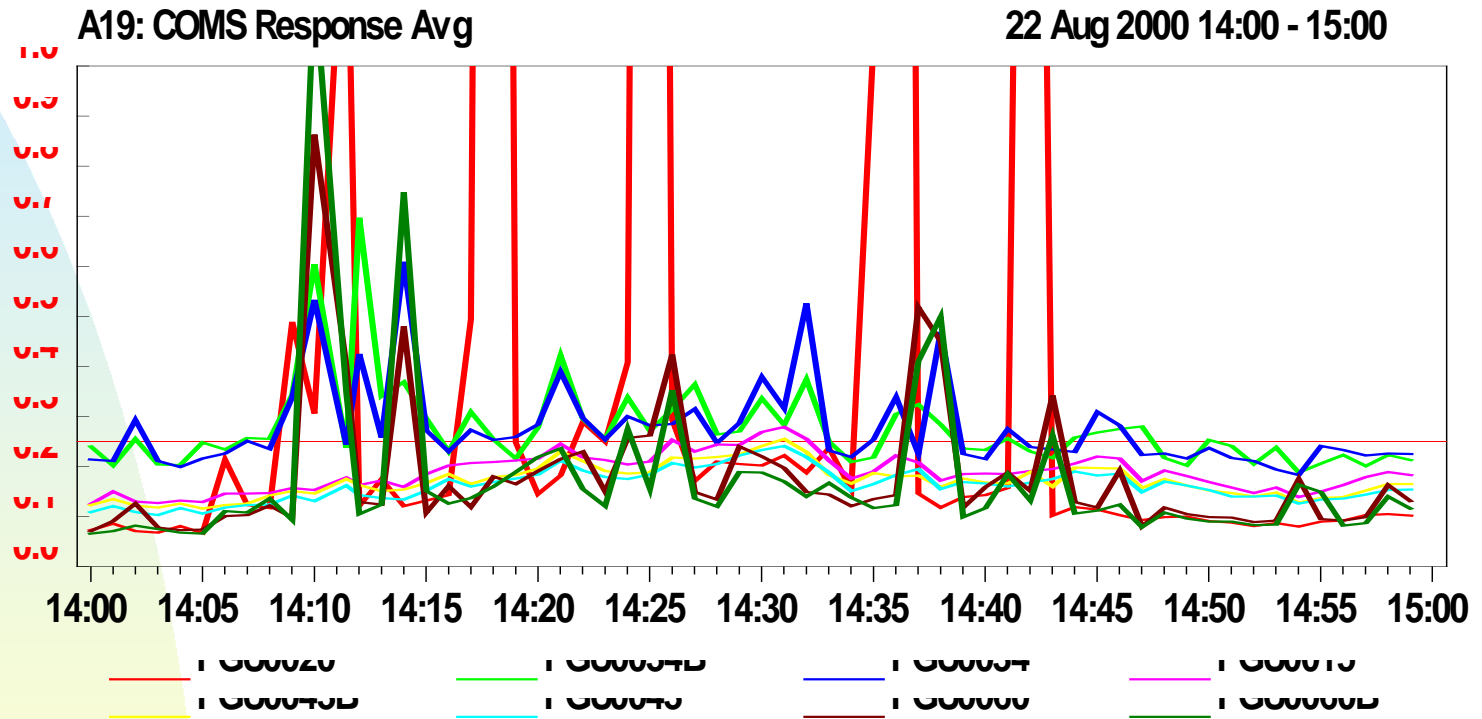
Perform a Workload Service Level Analysis

- Workload Resource Analysis
 - Determine the heaviest resource per workload
 - Determine resource contention for each workload
 - Workload Analysis
 - ❖ Workload Resource Times
 - Interval * CPU% 10 seconds
 - Interval * ReadyQ% 13 seconds
 - I/O time * IO% 5 seconds
 - Unknown (DMS, etc) 2 seconds
 - ❖ Sample Interval 30 seconds
 - ❖ Other contention indicators
 - OtherPbit %
 - CPU Stretch

On-Line Workload Profile

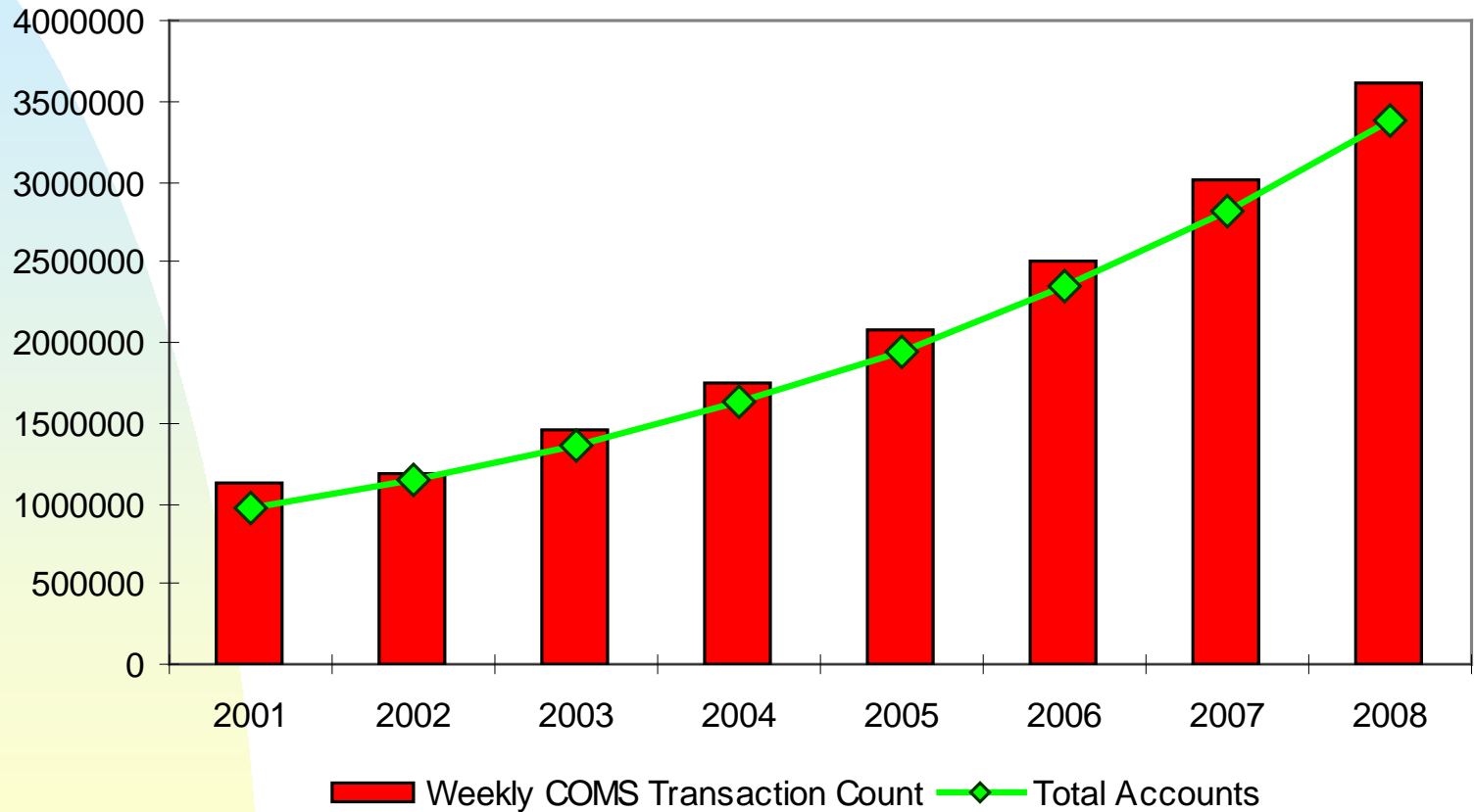


On-Line Response Time



On-Line Workload Projection

Average Weekly Transaction Counts vs. Total Accounts

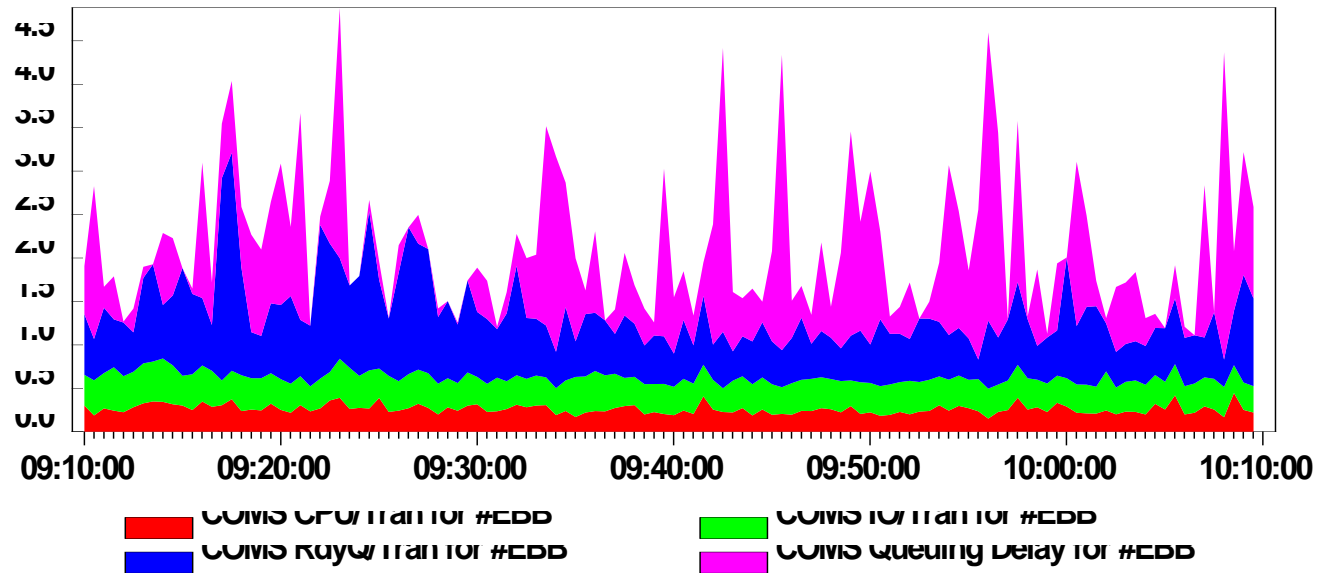


Perform a Workload Service Level Analysis

- Transaction Resource Analysis
 - Determine the heaviest resource per program
 - Determine resource contention for each program
 - Reconcile where response time is spent
 - Transaction Analysis Breakdown
 - Average Response time 900 ms
 - ❖ CPU time 150 ms
 - ❖ Readyq time 300 ms
 - ❖ I/O time 200 ms
 - ❖ DMS/COMS overhead 250 ms
 - COMS Program Q-Depth .5 transactions
 - DMS BTR Delay 150 ms

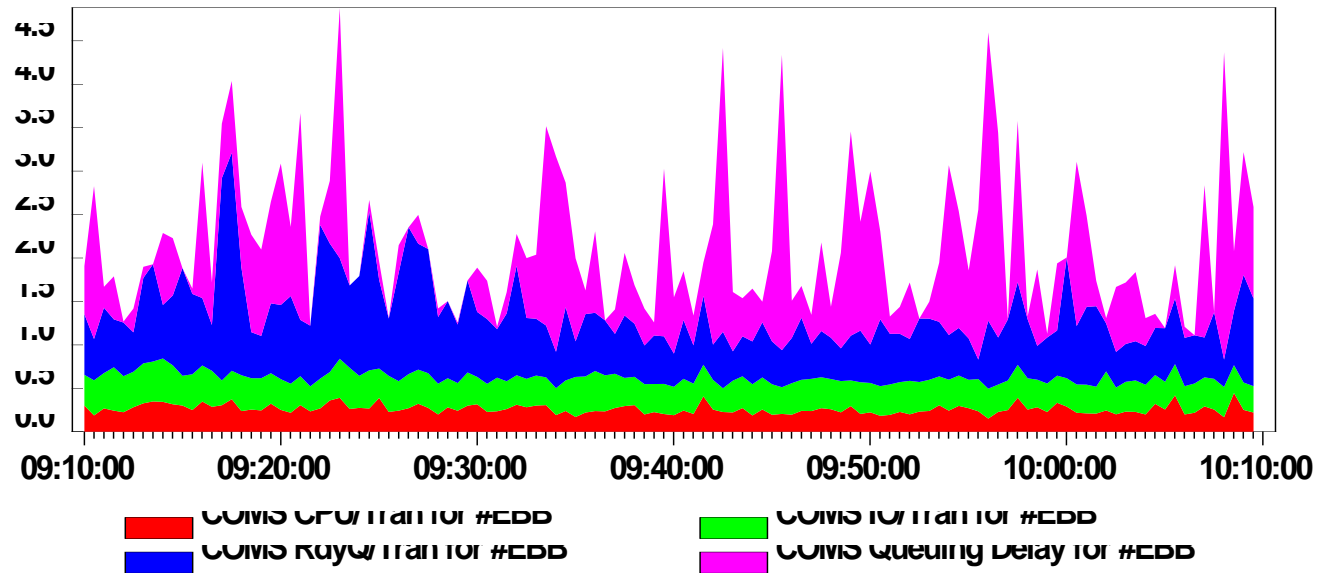
Workload Service Level Analysis Response Time

1 Jul 2004 09:10:00 - 10:10:00

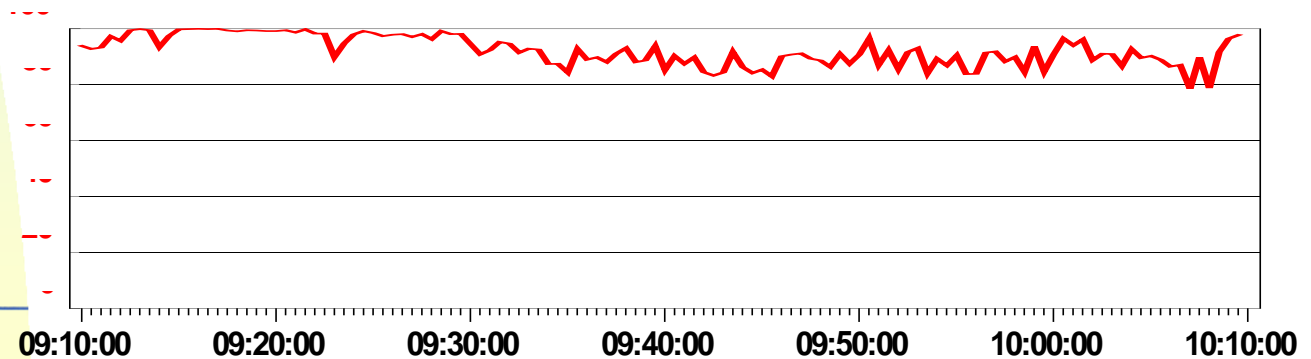


Workload Service Level Analysis Response Time

1 Jul 2004 09:10:00 - 10:10:00



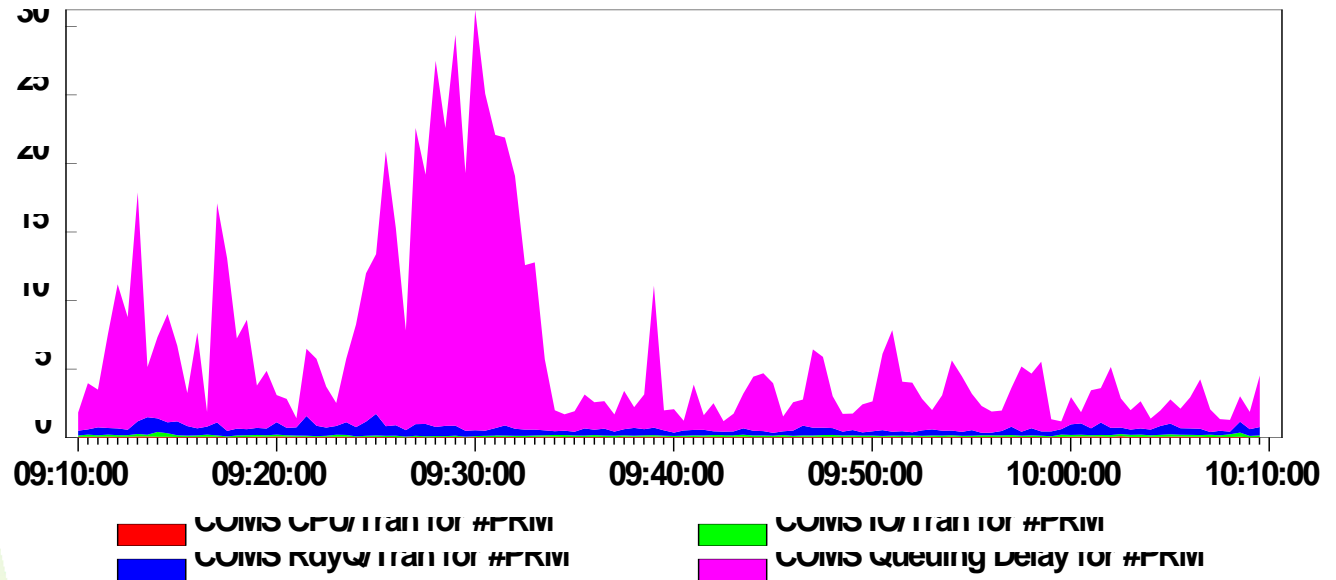
1 Jul 2004 09:10:00 - 10:10:00



Workload Service Level Analysis

Response Time

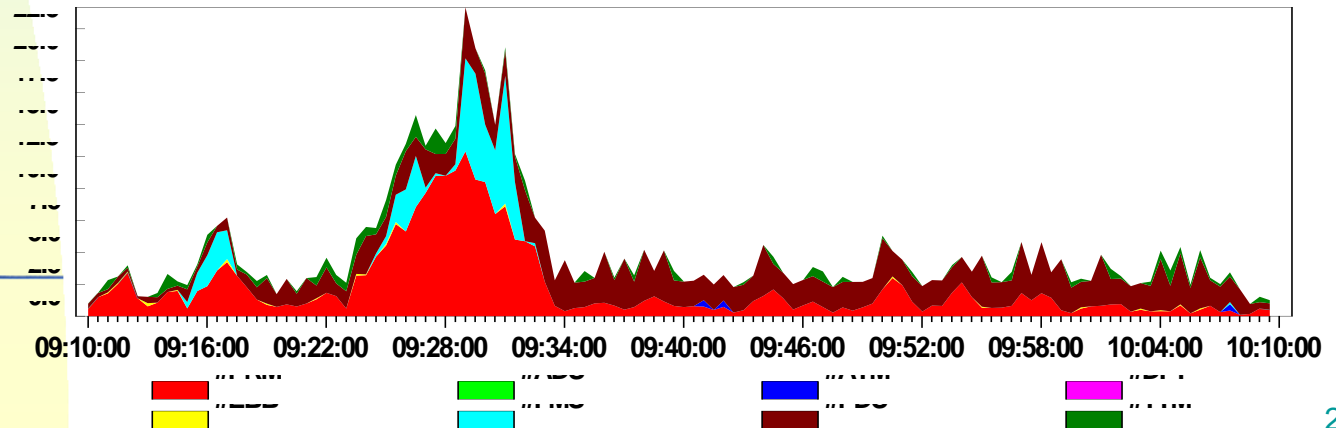
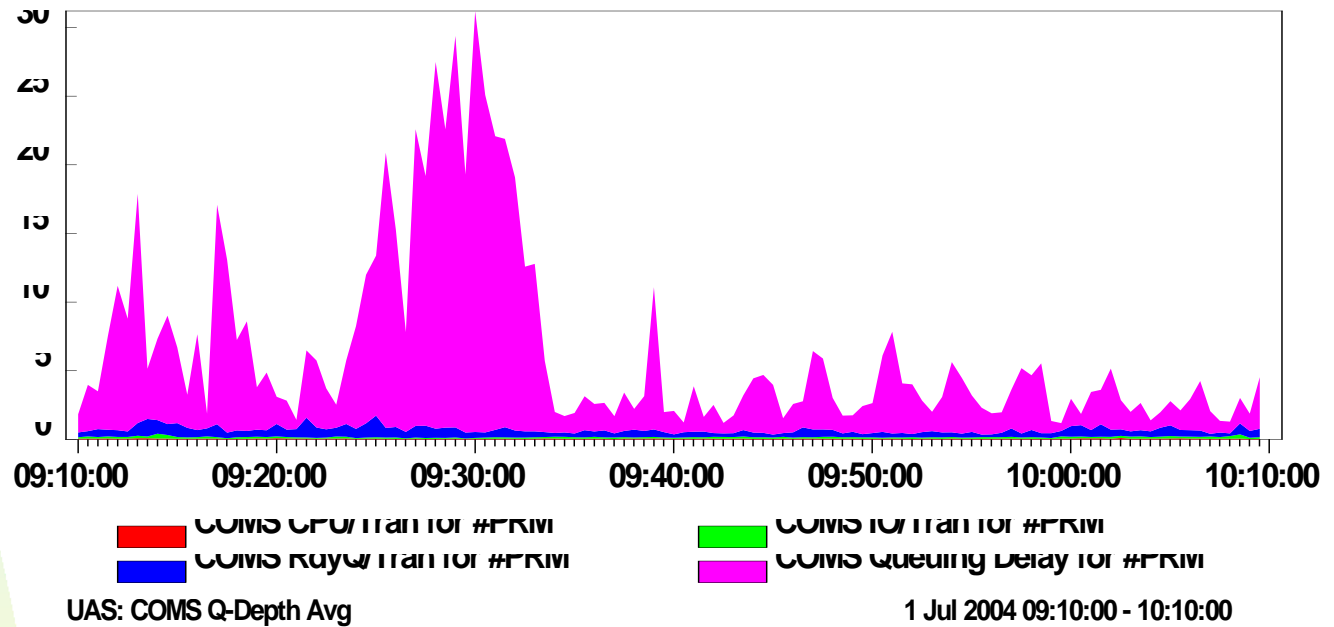
1 Jul 2004 09:10:00 - 10:10:00



Workload Service Level Analysis

Response Time

1 Jul 2004 09:10:00 - 10:10:00



Perform a Workload Service Level Analysis

■ Batch Workload Resource Analysis

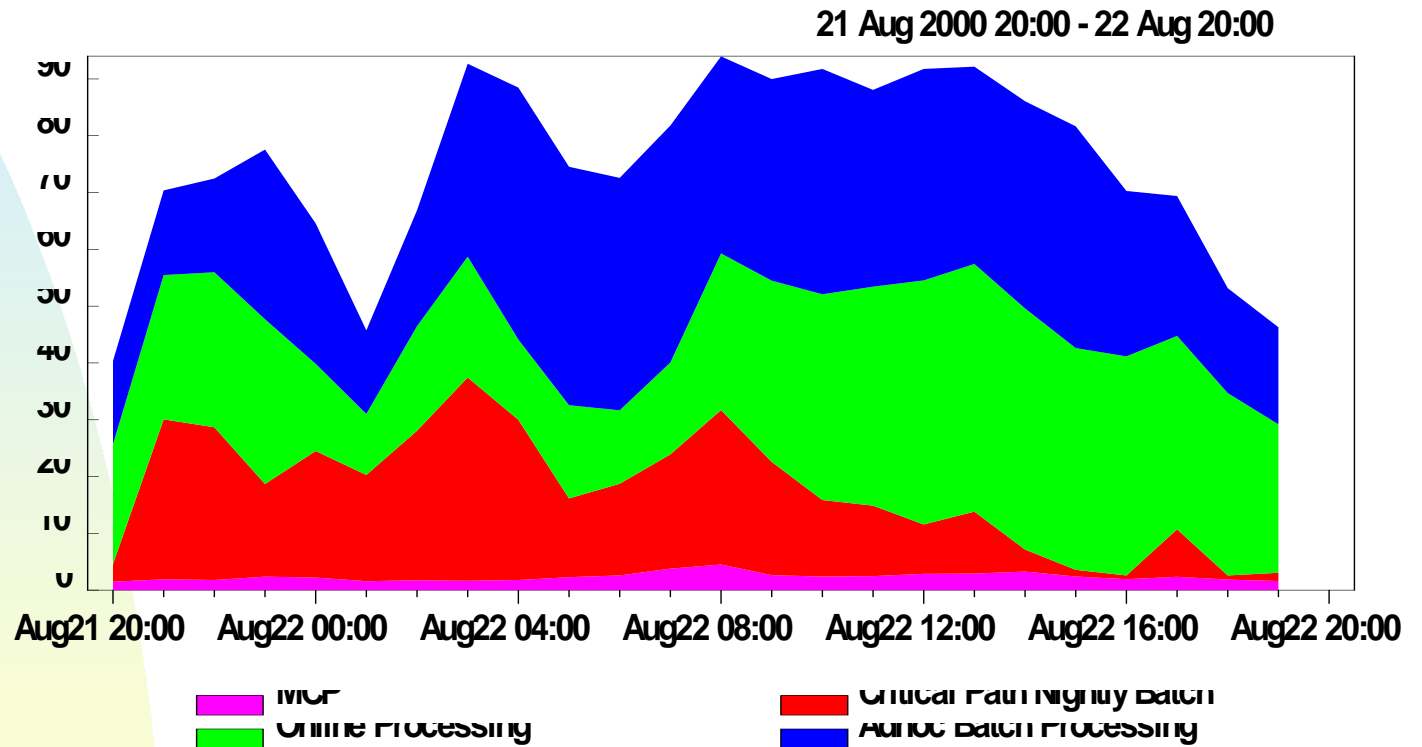
- Must utilize SWA vs. SPA performance data
- Identify the programs with the highest resource times
- Determine resource contention per program
- Reconcile where elapse time is spent
- Batch elapsed analysis breakdown

❖ Average elapsed time	8100 seconds (2:30)
❖ CPU time * Program Count	500 seconds
❖ Readyq time * Program Count	2500 seconds
❖ I/O time * Program Count	1000 seconds
❖ Unknown overhead	4100 seconds
• DMS BTR Delay * BTR WaitCnt	500 seconds
• Program RSVP	3000 seconds
• Other Delay	600 seconds

Workload Distribution

- Deadlines
 - Identify processing constraints
 - Identify input availability constraints
 - Identify output delivery requirements
- Map Requirements vs. Resources
 - Foundation/constant demands
 - Variable demands
 - Plot on a timeline with deadlines
- Determine Demand Peaks

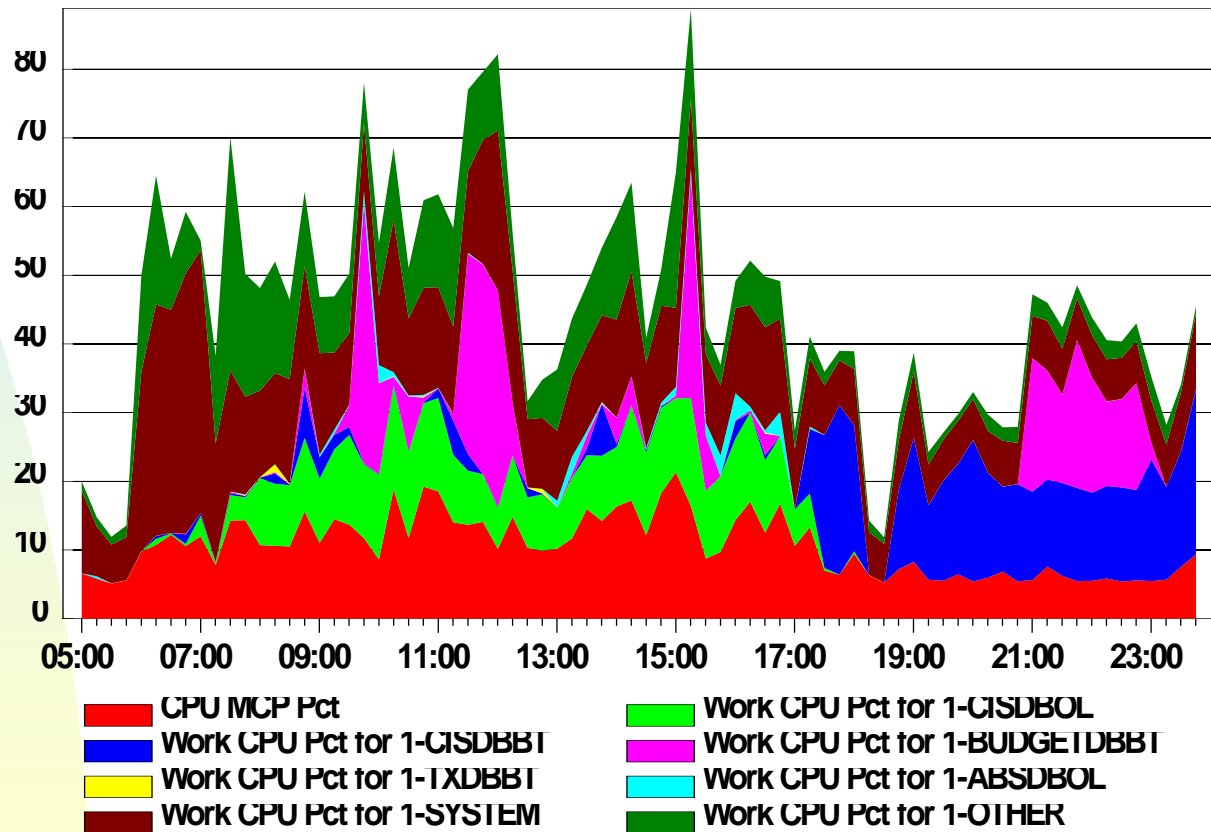
Workload CPU Demand by Type



Application CPU Distribution

Quarter Hour Averages - CPU Busy

29 Nov 2005 05:00 - 30 Nov 00:00

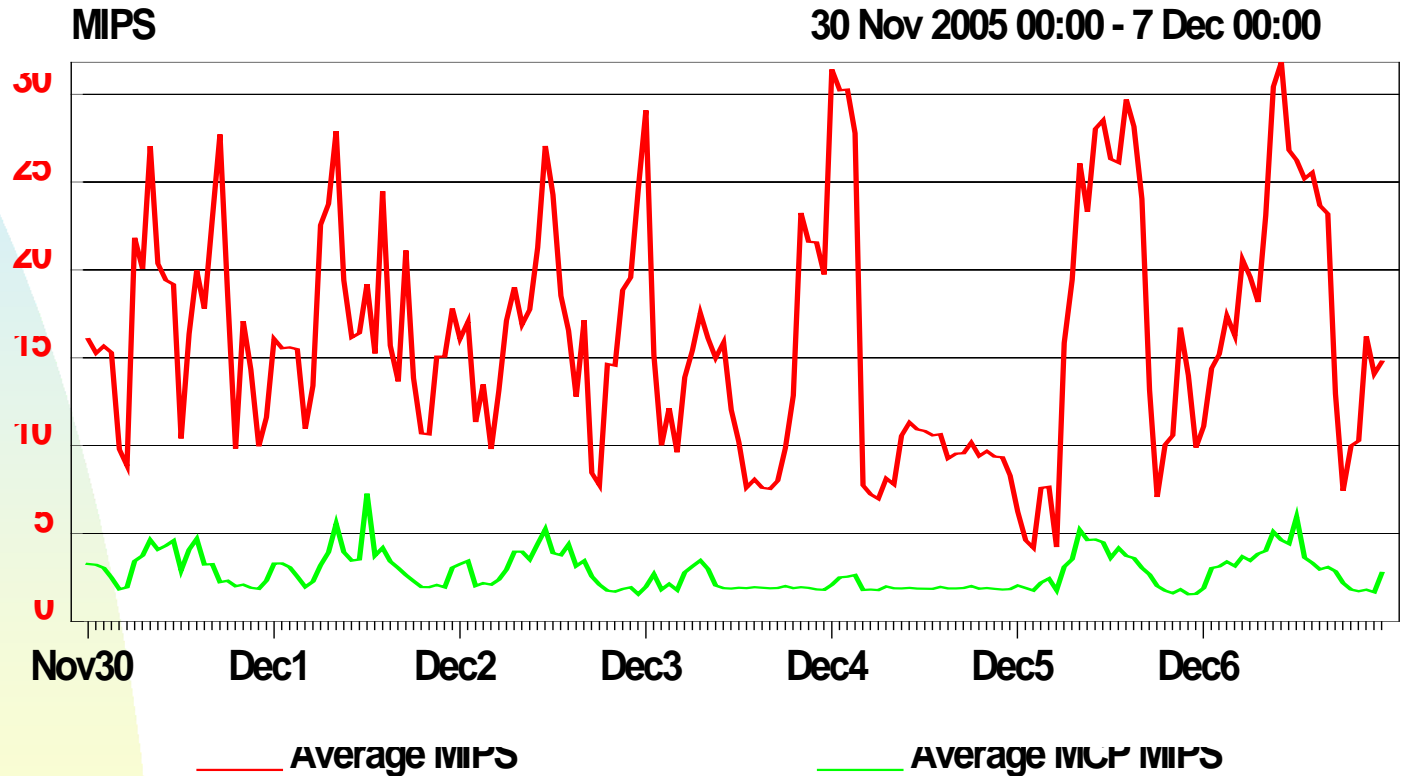


Management Methodology

- Establish baseline
 - Performance requirements (objectives)
 - Capacity usage
 - Workload volume
- Maintain performance/capacity history
- Perform trend analysis
 - Capacity usage
 - Workload volume
- Projections
 - Future capacity usage
 - Potential hardware solutions
- Alerting and periodic reporting
 - Capacity consumption vs plan
 - Performance vs service levels
 - Capacity usage alerts

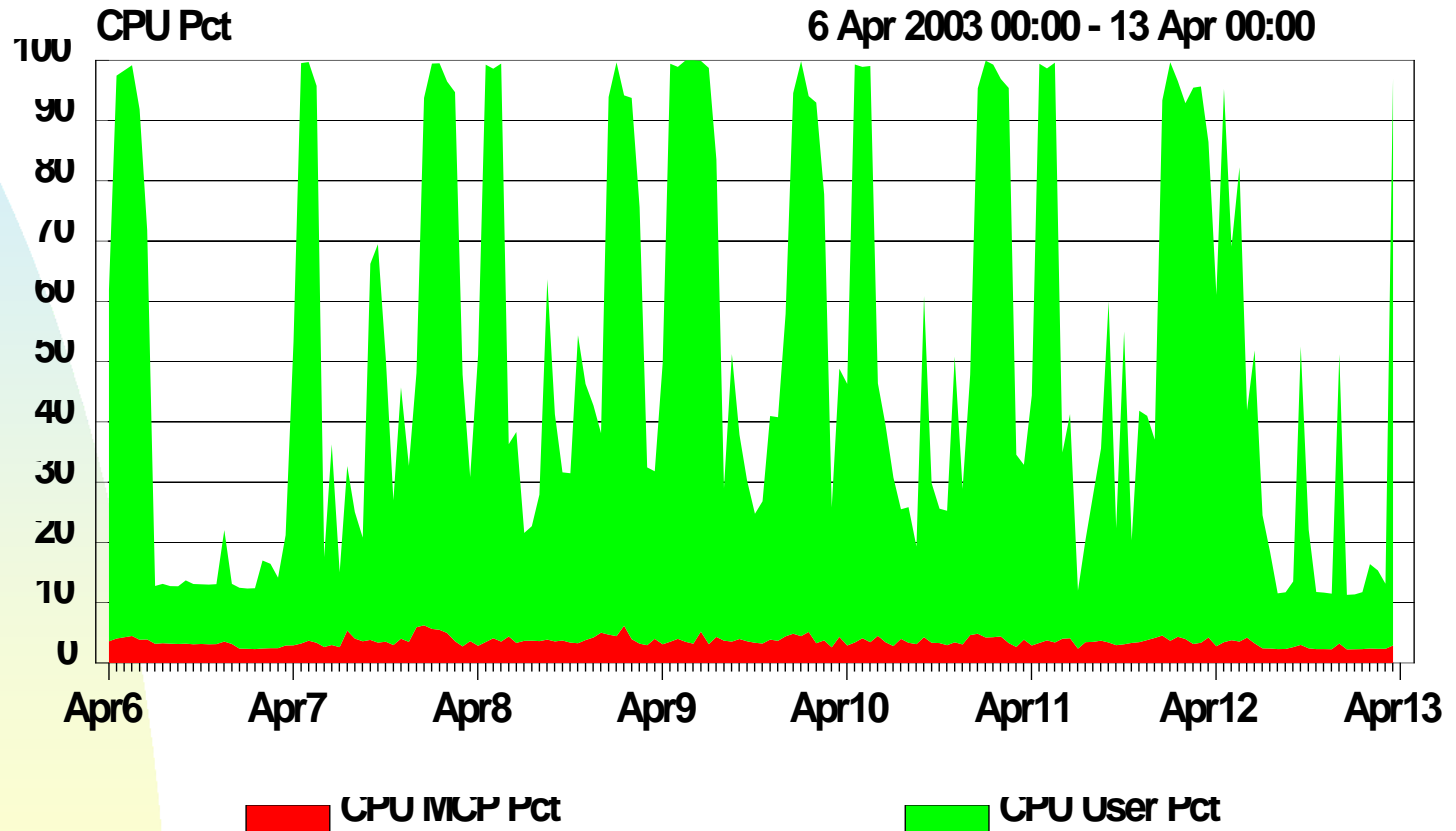
Management Methodology

Monitor Capacity Usage



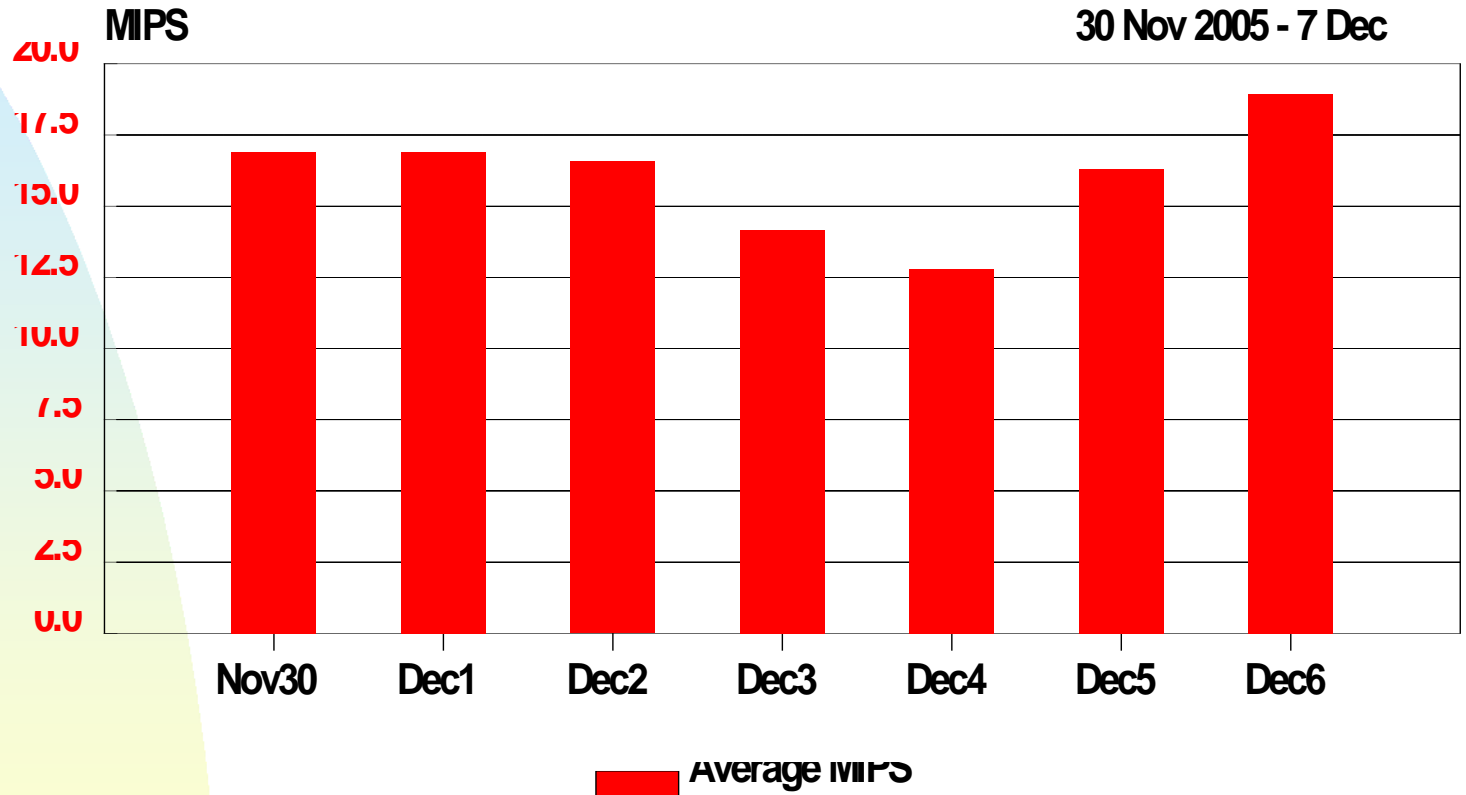
Management Methodology

Monitor CPU Busy Percent



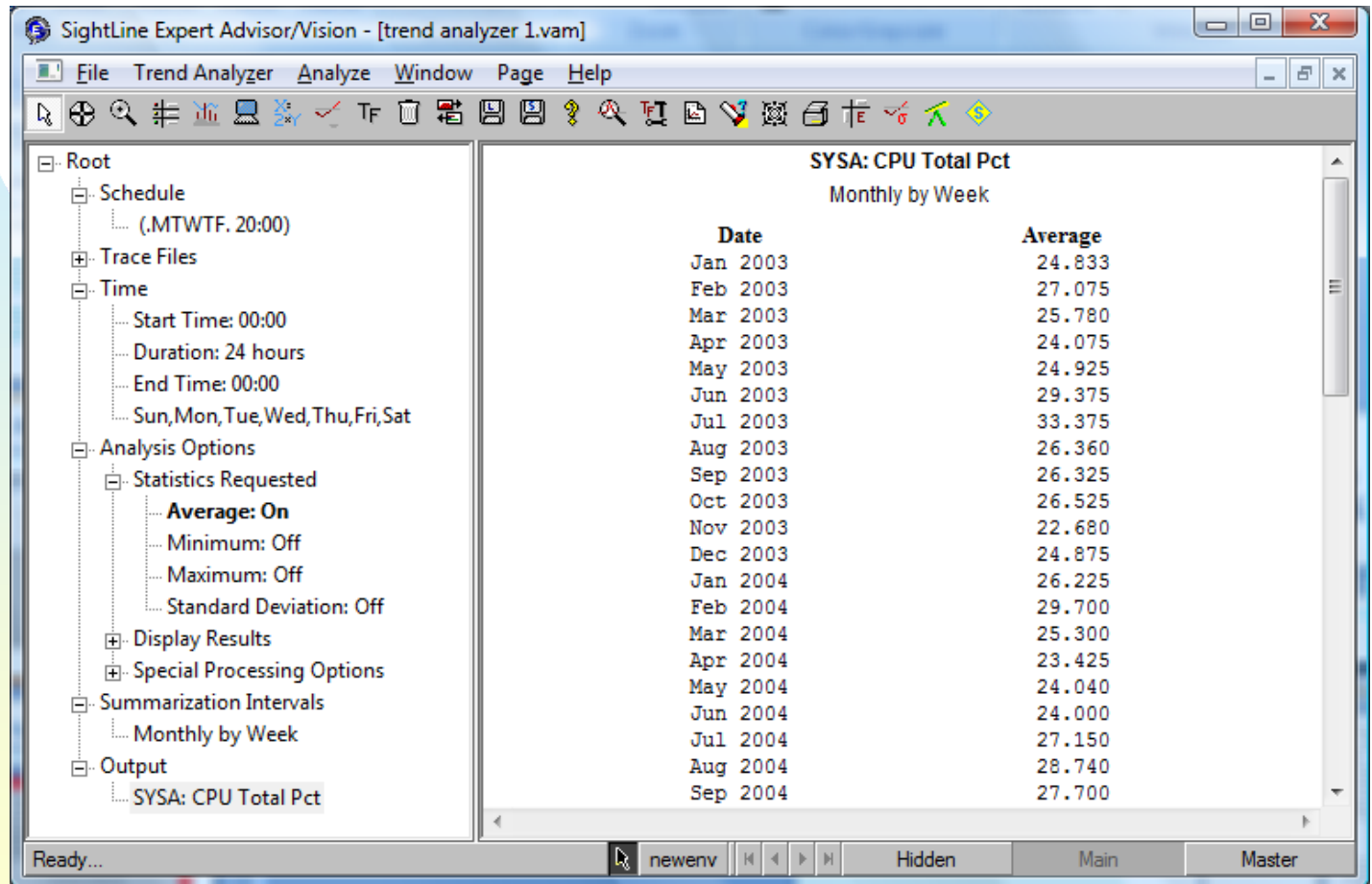
Management Methodology

Demand Pattern - Capacity by Day



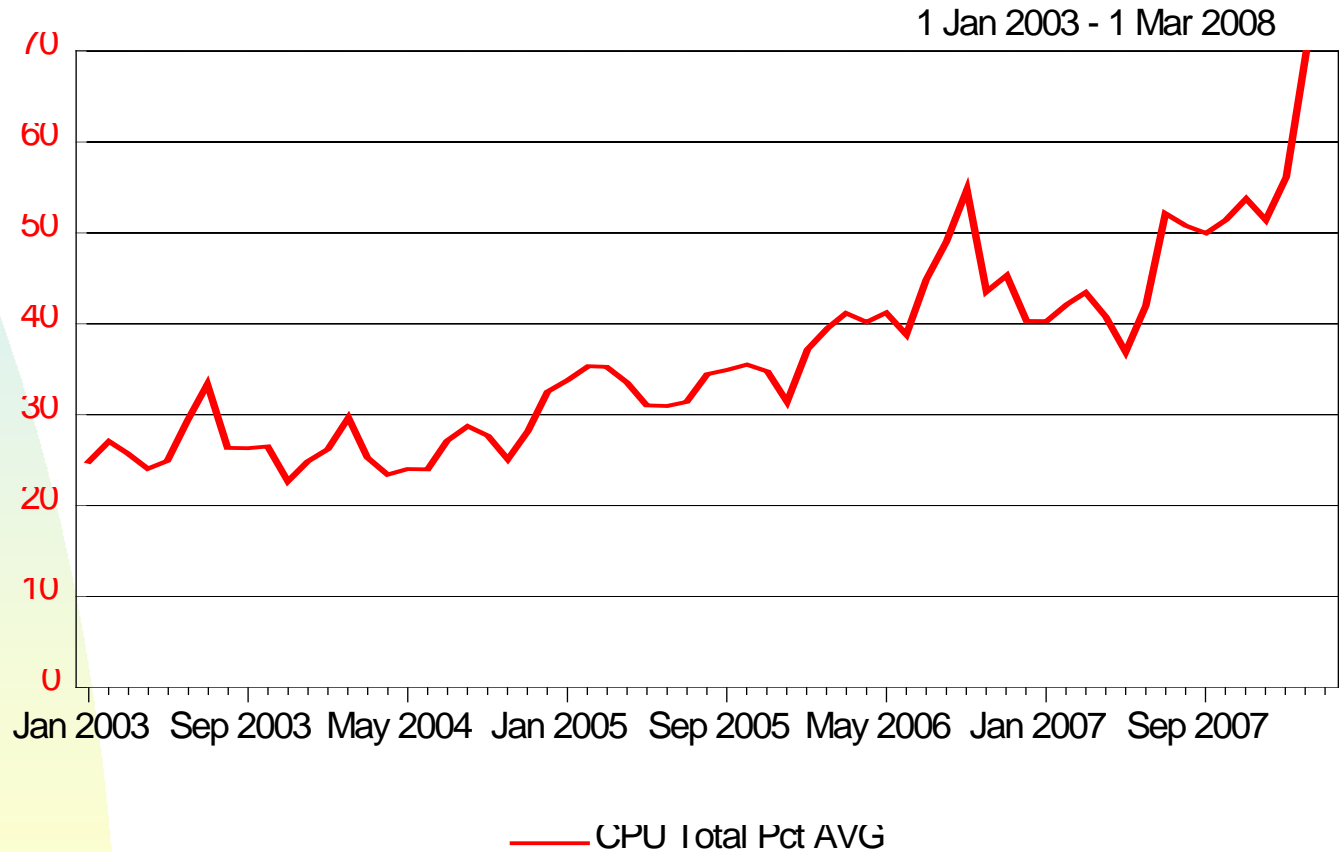
Management Methodology

Trend Analysis



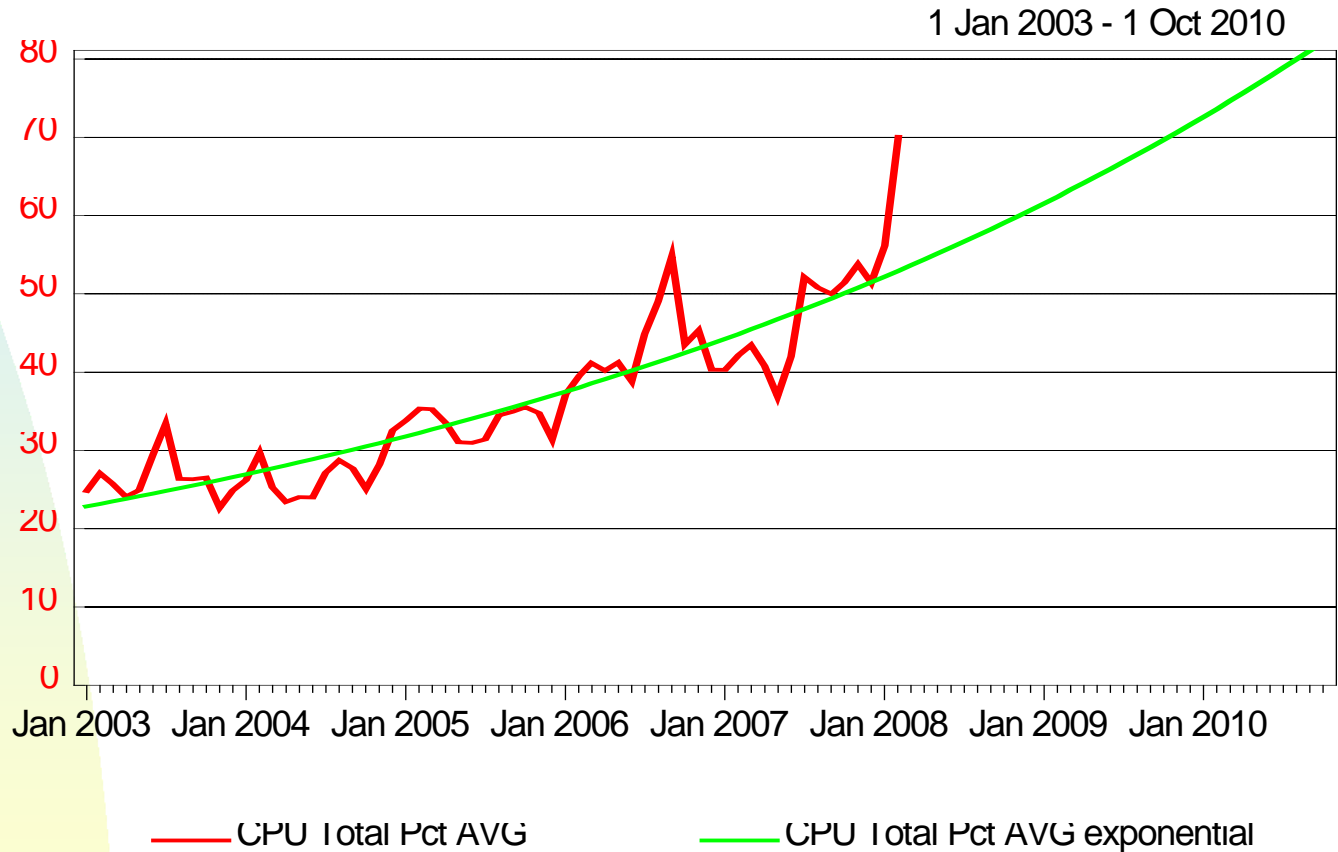
Management Methodology

Trending Data



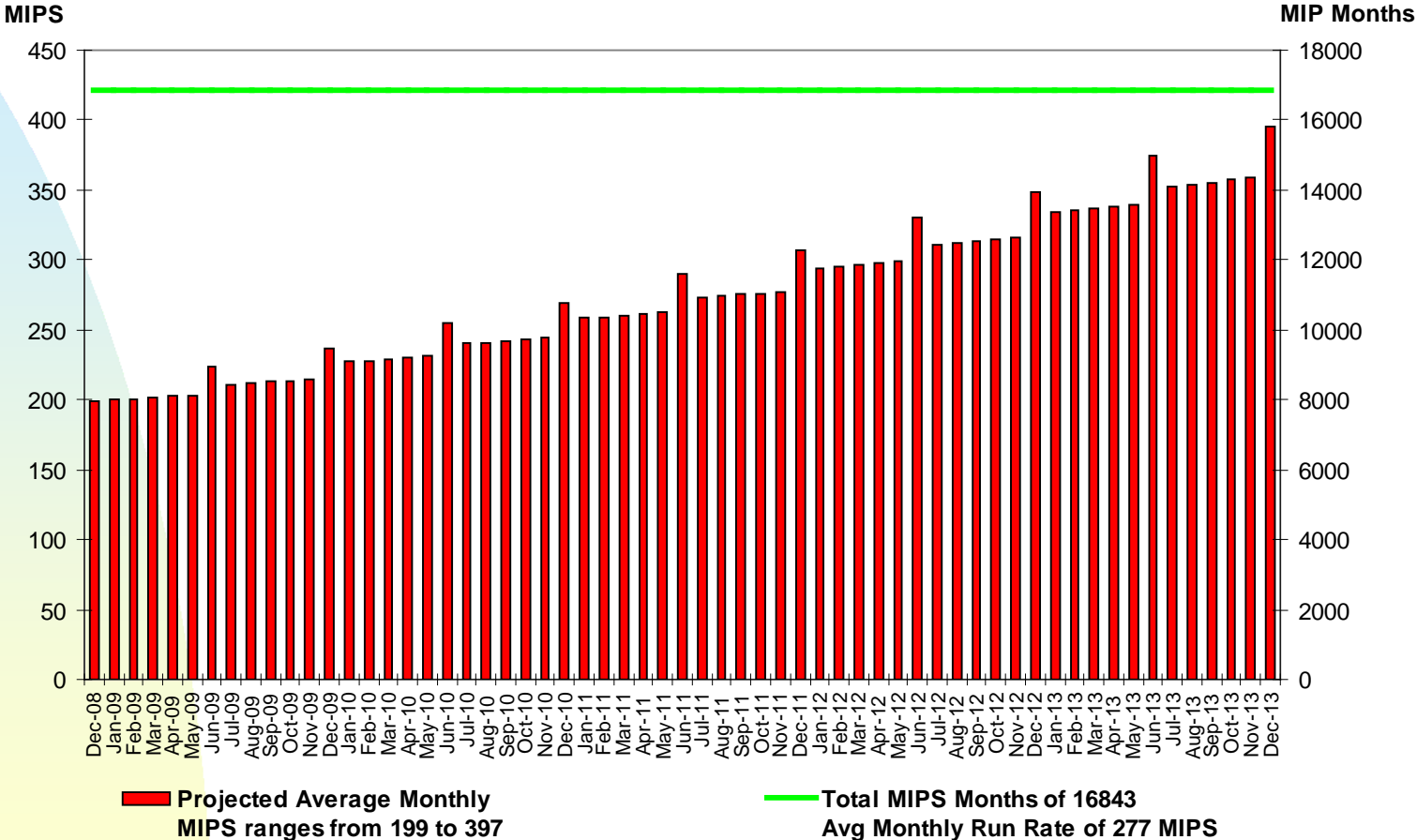
Management Methodology

Trend Projection



Managing Capacity

Capacity Usage Projection



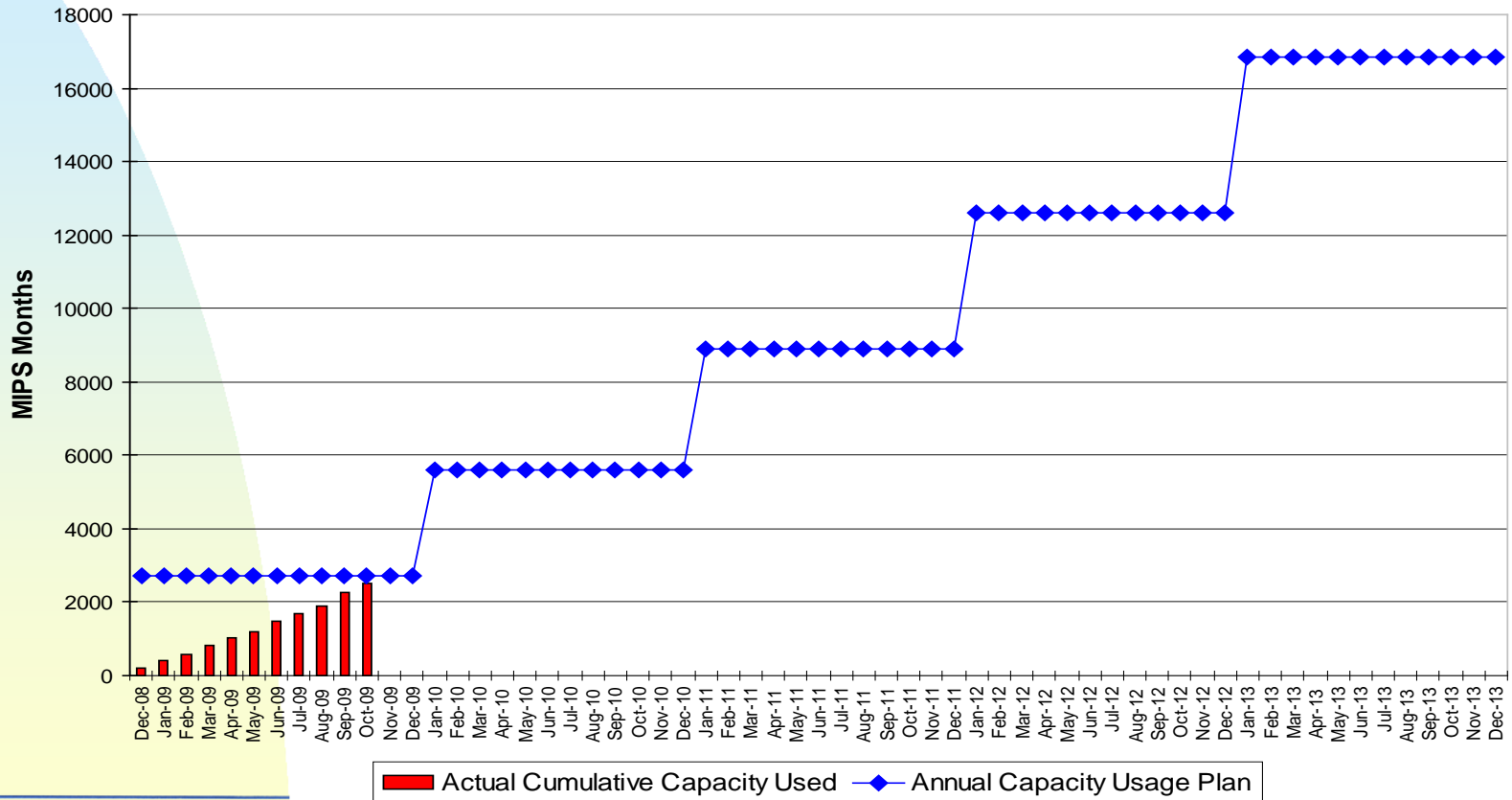
Managing Capacity

Metering Capacity Usage Plan

Managing Capacity

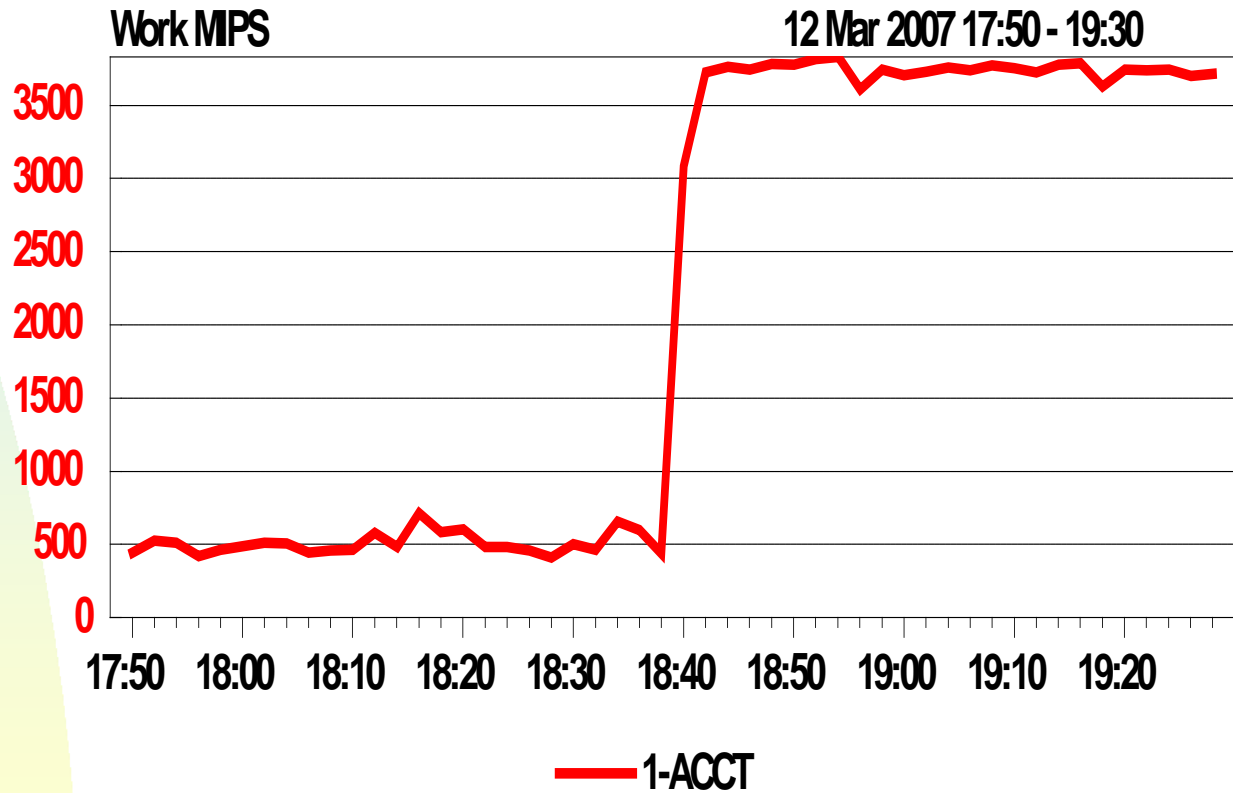
Metering Capacity Used

Actual Cumulative Capacity Usage



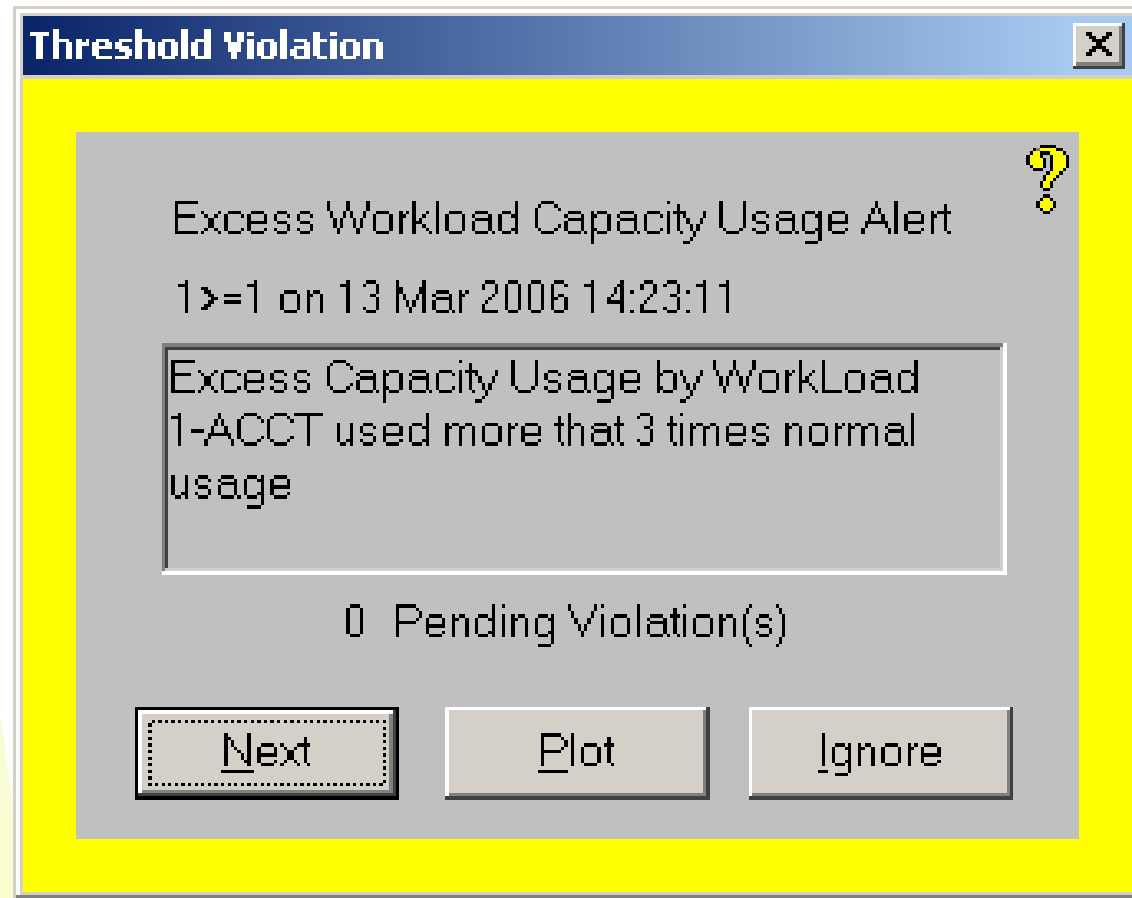
Managing Capacity

Excessive Workload Capacity Usage

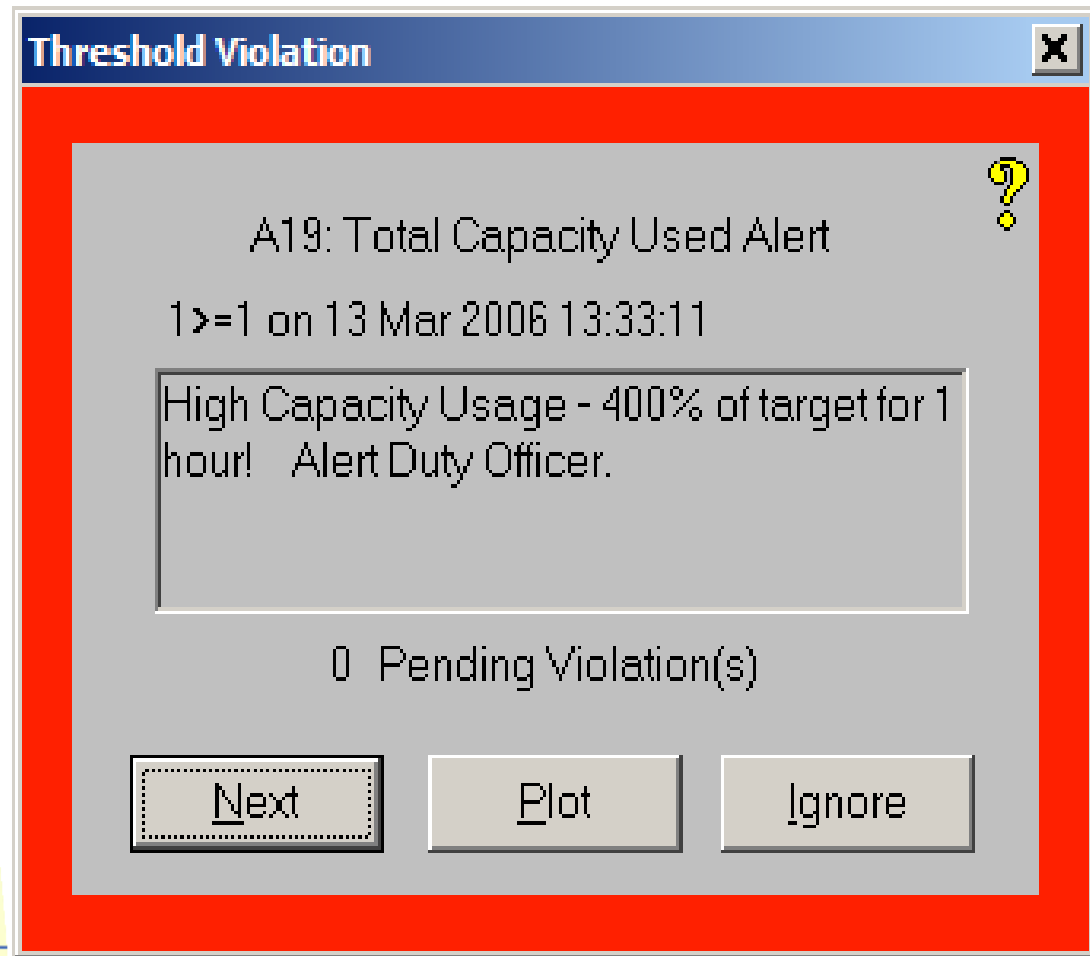


Managing Capacity

Excessive Workload Usage Alarm

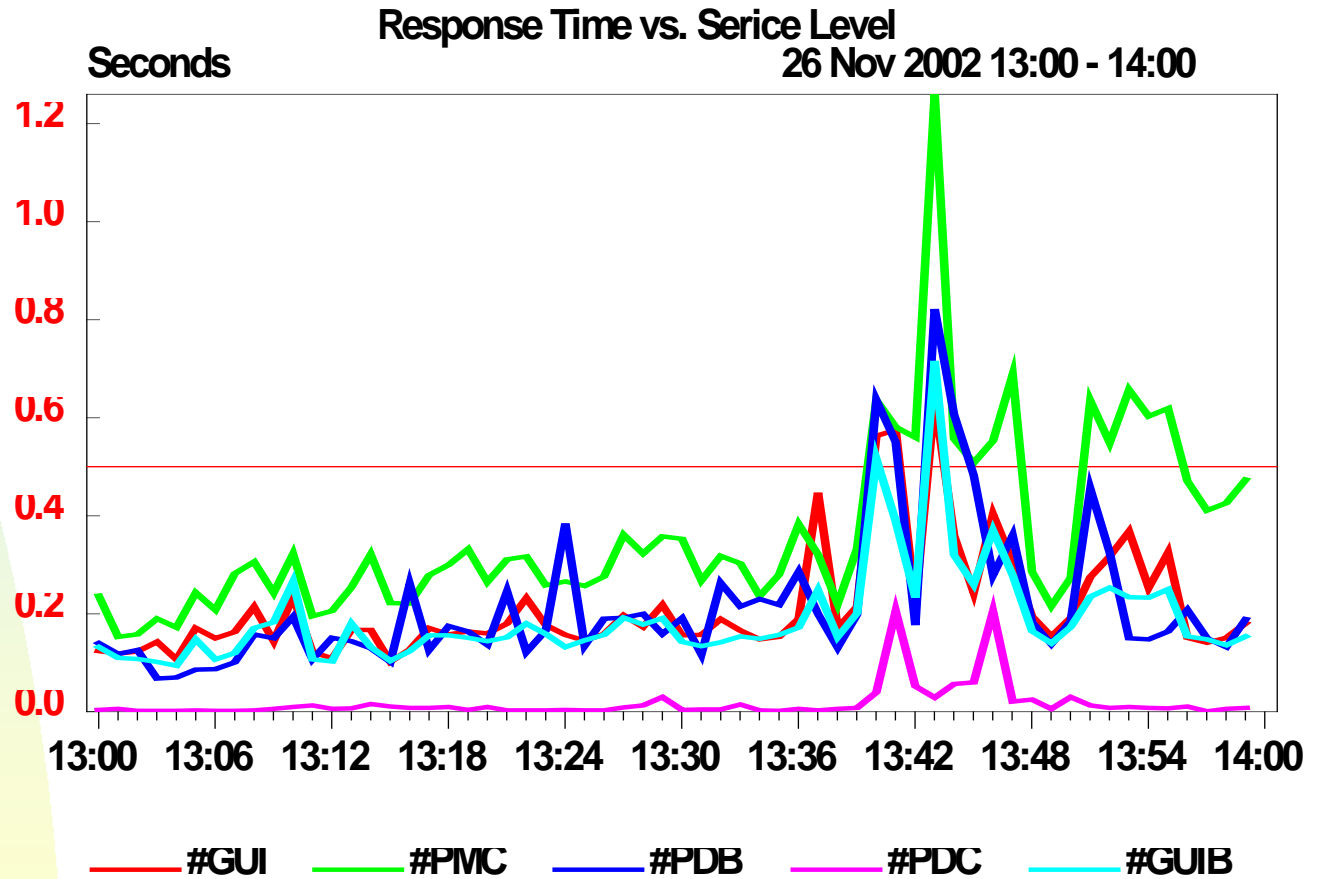


Managing Capacity Problem Alerting



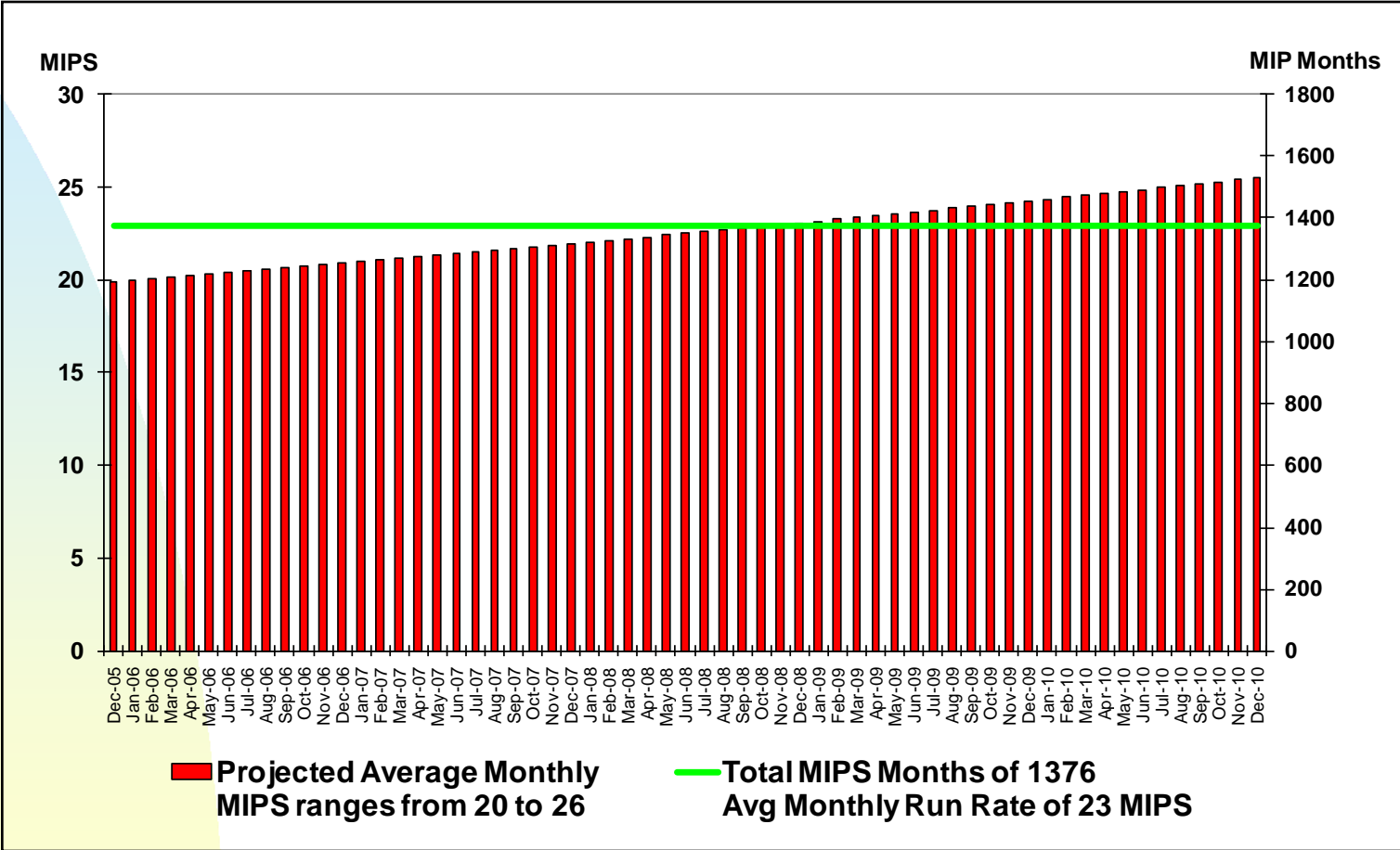
Managing Capacity

Performance vs. Service Levels



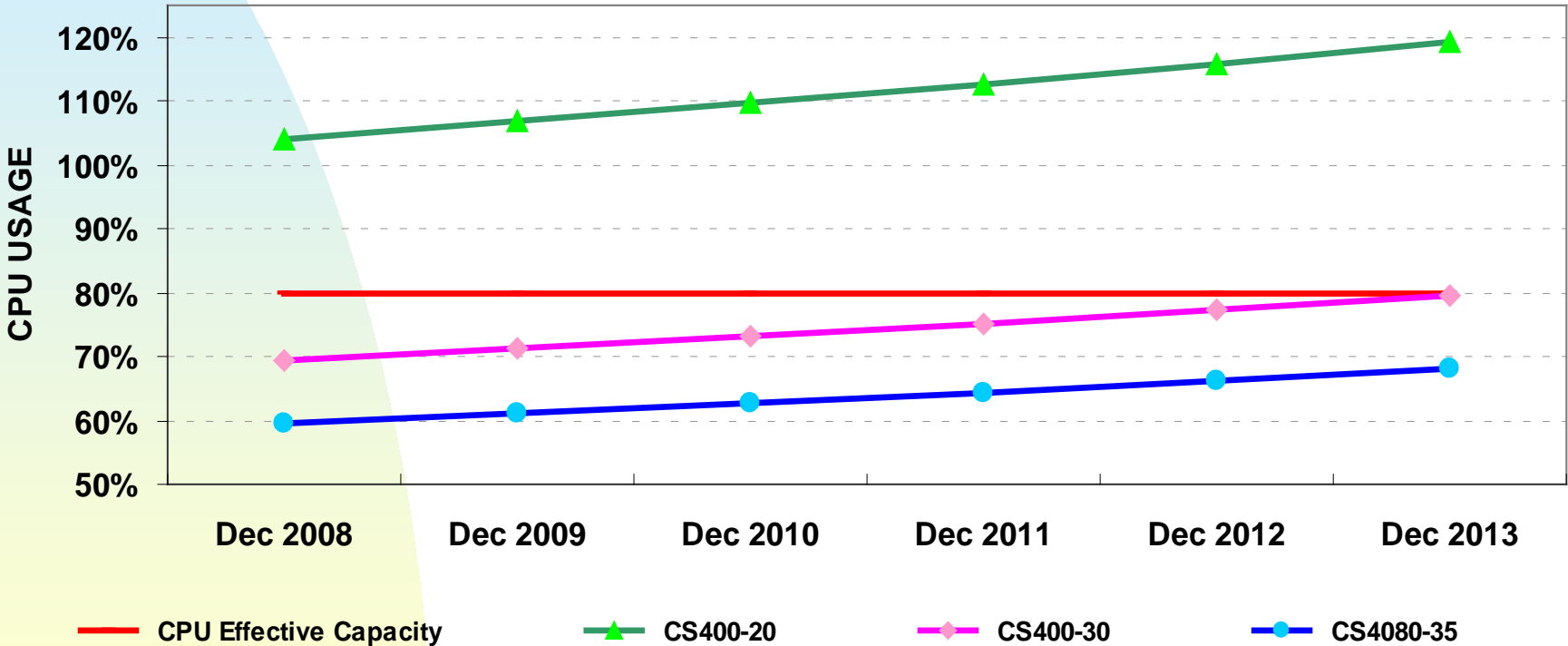
To Meter or Not to Meter

Metered System Option



To Meter or Not to Meter Pay for Peak Option

CPU REQUIREMENT PROJECTION
Pay for Peak Replacement Systems' CPU Utilization vs Effective Capacity



To Meter or Not to Meter Considerations

- Consistent Workload?
- Small to Medium Load?
- Gradual Growth?
- User Based Licensing?
 - Few software licenses needed
- → Non-metered solution may be less expensive – do detailed financial analysis – note that Libra 4000 changes the HW/SW expense ratio.

Questions?

- Thank you for your attention
- Are there any questions?

Note that this presentation will be available for download today at:

www.mgsinc.com/download.html