z/OS Performance Spotlight – Some Top Things You May Not Know

aka Peter and Scott's Tips and Tidbits

Peter Enrico & Scott Chapman

Enterprise Performance Strategies, Inc.

performance.questions@epstrategies.com

Contact, Copyright, and Trademarks

Questions?

Send email to performance.questions@EPStrategies.com, or visit our website at https://www.epstrategies.com or http://www.pivotor.com.

Copyright Notice:

Trademarks:

Enterprise Performance Strategies, Inc. presentation materials contain trademarks and registered trademarks of several companies.

The following are trademarks of Enterprise Performance Strategies, Inc.: **Health Check®, Reductions®, Pivotor®**

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries: IBM®, z/OS®, zSeries®, WebSphere®, CICS®, DB2®, S390®, WebSphere Application Server®, and many others.

Other trademarks and registered trademarks may exist in this presentation

Abstract

- During this session, Peter Enrico and Scott Chapman will discuss a variety of z/OS performance measurement, analysis, and tuning techniques that may not be commonly known or are not often discussed.

- The key objective of this presentation is to provide the attendee with information they can bring back to their shop and conduct some analysis or tuning exercises. A secondary objective of this session is to help the attendee learn more about the z/OS environment, and how things work. This session is sure to be highly educational!

EPS: We do z/OS performance…

- Pivotor - Reporting and analysis software and services
  - Not just reporting, but analysis-based reporting based on our expertise
- Education and instruction
  - We have taught our z/OS performance workshops all over the world
- Consulting
  - Performance war rooms: concentrated, highly productive group discussions and analysis
- Information
  - We present around the world and participate in online forums
  - Many presentations available: https://www.pivotor.com/content.html

z/OS Performance workshops available

During these workshops you will be analyzing your own data!

- Essential z/OS Performance Tuning
  - October 3-7, 2022
- WLM Performance and Re-evaluating Goals
  - September 12-16, 2022
- Parallel Sysplex and z/OS Performance Tuning
  - Next on is winter 2023
- Also… please make sure you are signed up for our free monthly z/OS educational webinars! (email contact@epstrategies.com)

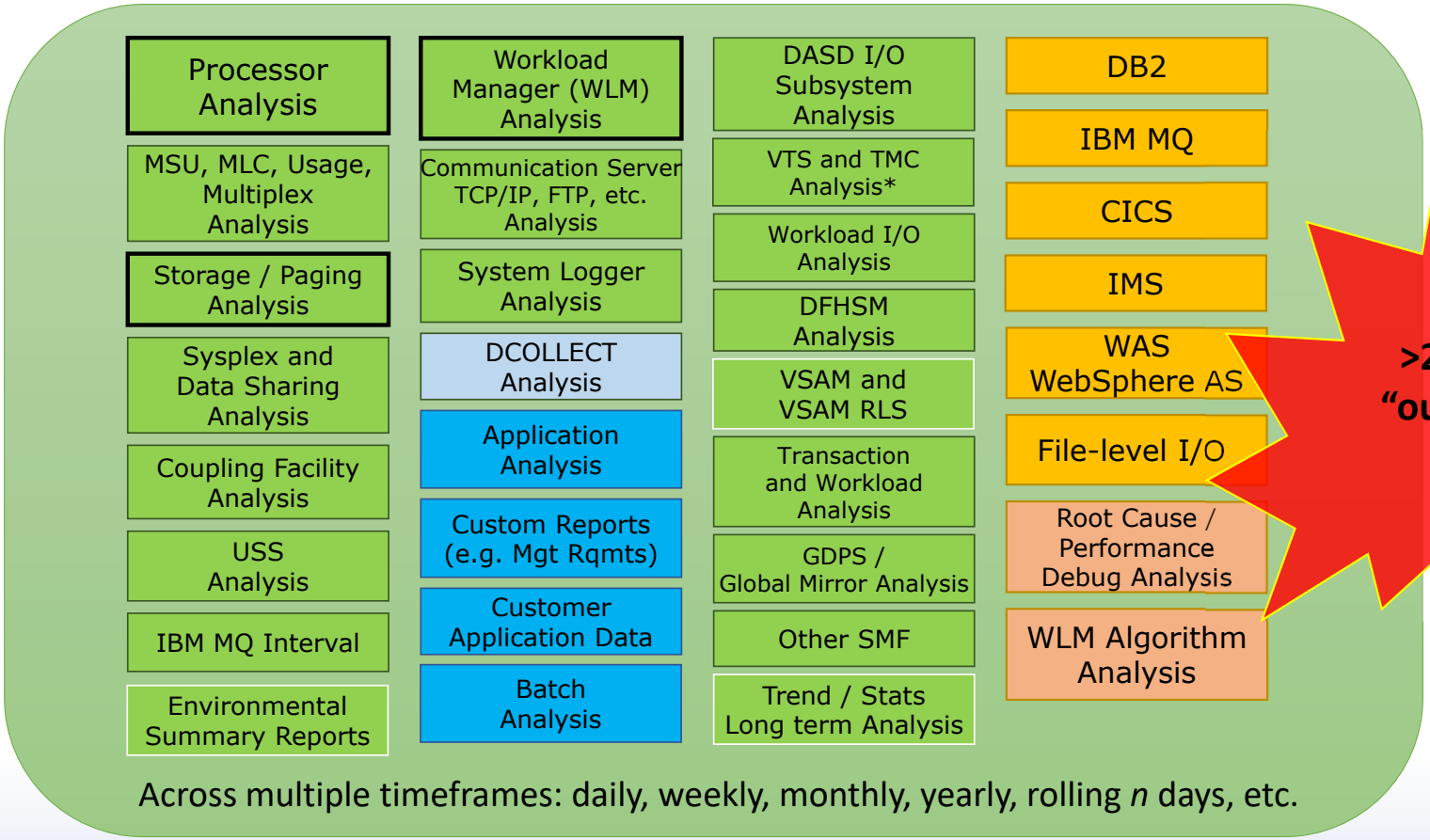
EPS presentations this week

What	Who	When	Where
PSP: z/OS Performance Tuning - Some Top Things You May Not Know	Peter Enrico Scott Chapman	Tue 1:15	Delaware A
Planning Your Next Mainframe Processor Upgrade	Scott Chapman	Tue 2:45	Franklin C
z/OS Performance Risk Management: Easy Things To Do To Reduce the Risk of Bad Performance	Scott Chapman	Wed 10:30	Franklin C
Pinpointing Transient Performance Problems with SMF 98 & 99	Peter Enrico	Thu 8:00	Franklin A
WLM's Algorithms - How WLM Works	Peter Enrico	Thu 1:15	Franklin C

Pivotor – Intelligent Reporting

- Pivotor is our data reporting tool & service designed specifically for z/OS performance reporting
  - Designed and used by z/OS performance experts
  - Processes data from SMF, DCOLLECT, and customer sources
  - Contains hundreds of z/OS performance reports “out of the box”
  - Designed to be easy to use and manage
  - Reports are organized into logical and searchable report sets
  - Features include intelligent exceptions, drill down, search, canned analysis, and so much more
  - Built in expanded helps to help foster education

Comprehensive Report Sets for Immediate Performance Analysis

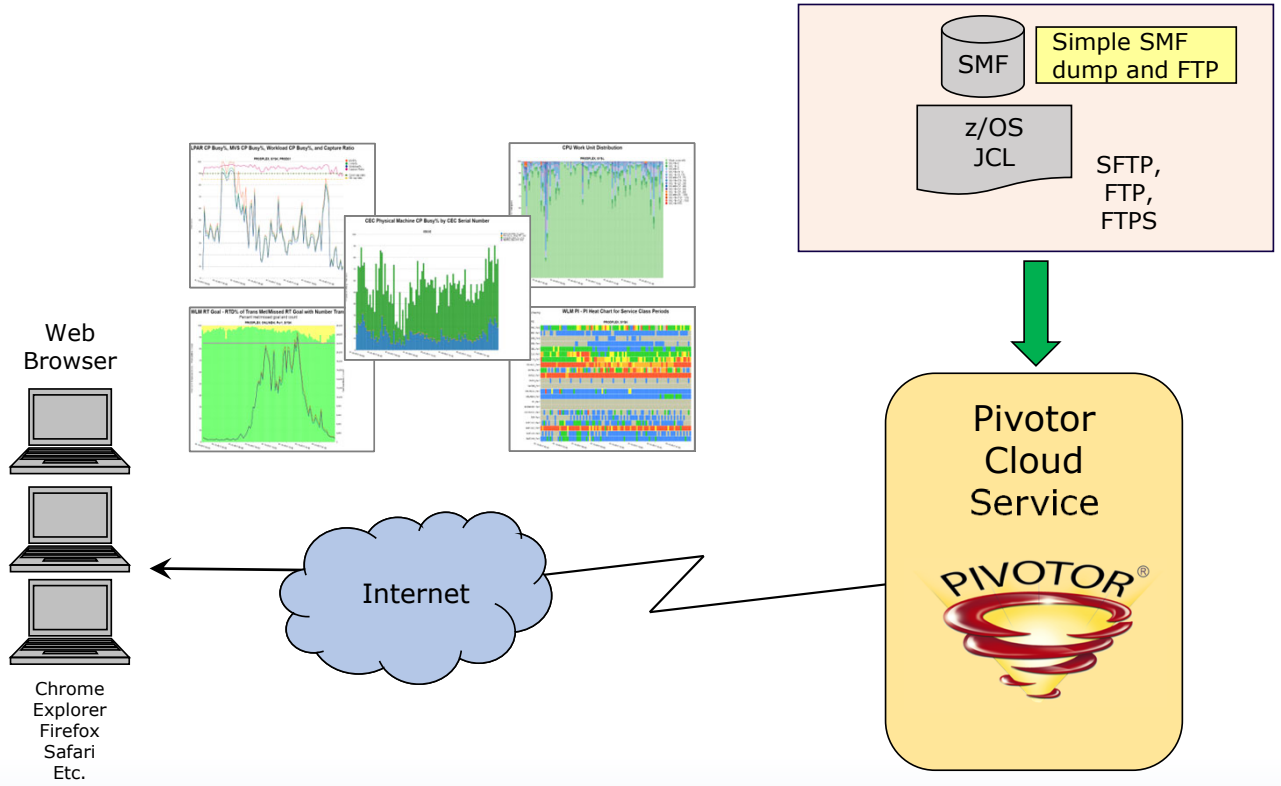


Pivotor Software as a Solution (SaaS)

- Pivotor is offered as both a SaaS or local install
- When SaaS:

SaaS Includes:

- Formal yearly cursory review / discussion
- Ability to ask us performance questions, or for us to look at a particular problem or concern. (support@epstrategies.com)
- We can occasionally look in on your data and performance
- We can participate in performance debug with IBM, or other vendors



- Pivotor pricing is clear and affordable

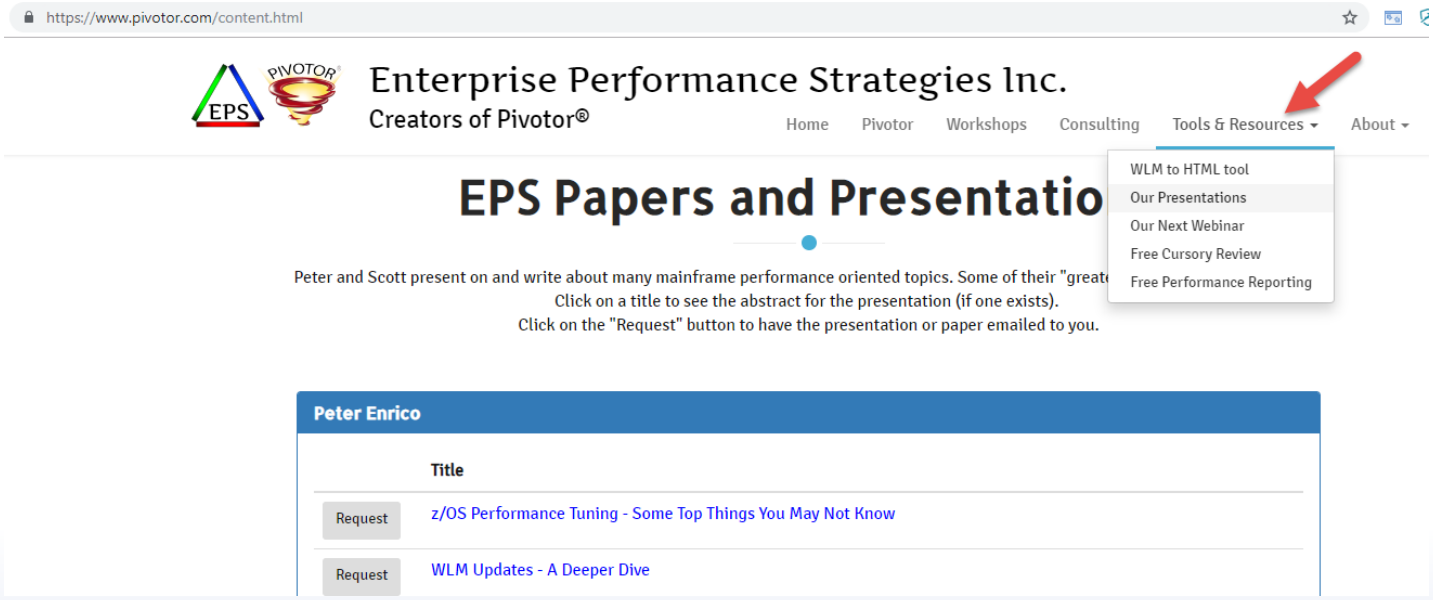
Like what you see?

- The z/OS Performance Graphs you see here come from Pivotor™
- If you just a free cursory review of your environment, let us know!
  - We're always happy to process a day's worth of data and show you the results
  - See also: http://pivotor.com/cursoryReview.html
- We also have a free Pivotor offering available as well
  - 1 System, SMF 70-72 only, 7 Day retention
  - That still encompasses over 100 reports!

More Free Things!

- On our web site click on Tools & Resources to access:
  - WLM to HTML Tool
    - Get your WLM policy in a useful and usable HTML format
  - Our Presentations
    - Lots of great content from the past few years

https://www.epstrategies.com/
https://www.pivotor.com/
(Same site behind both URLs)

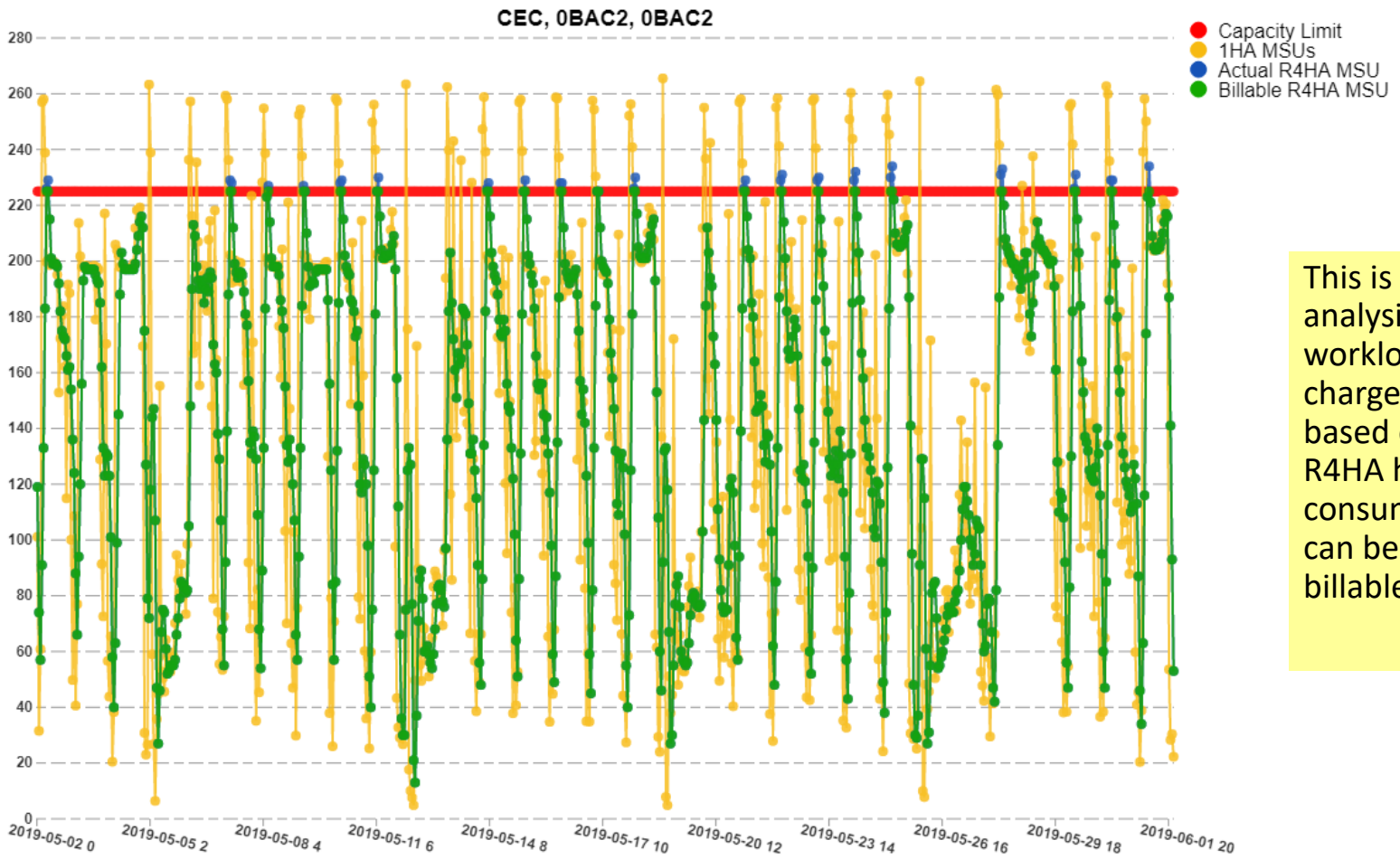


Tailored Fit Pricing: Performance Perspective

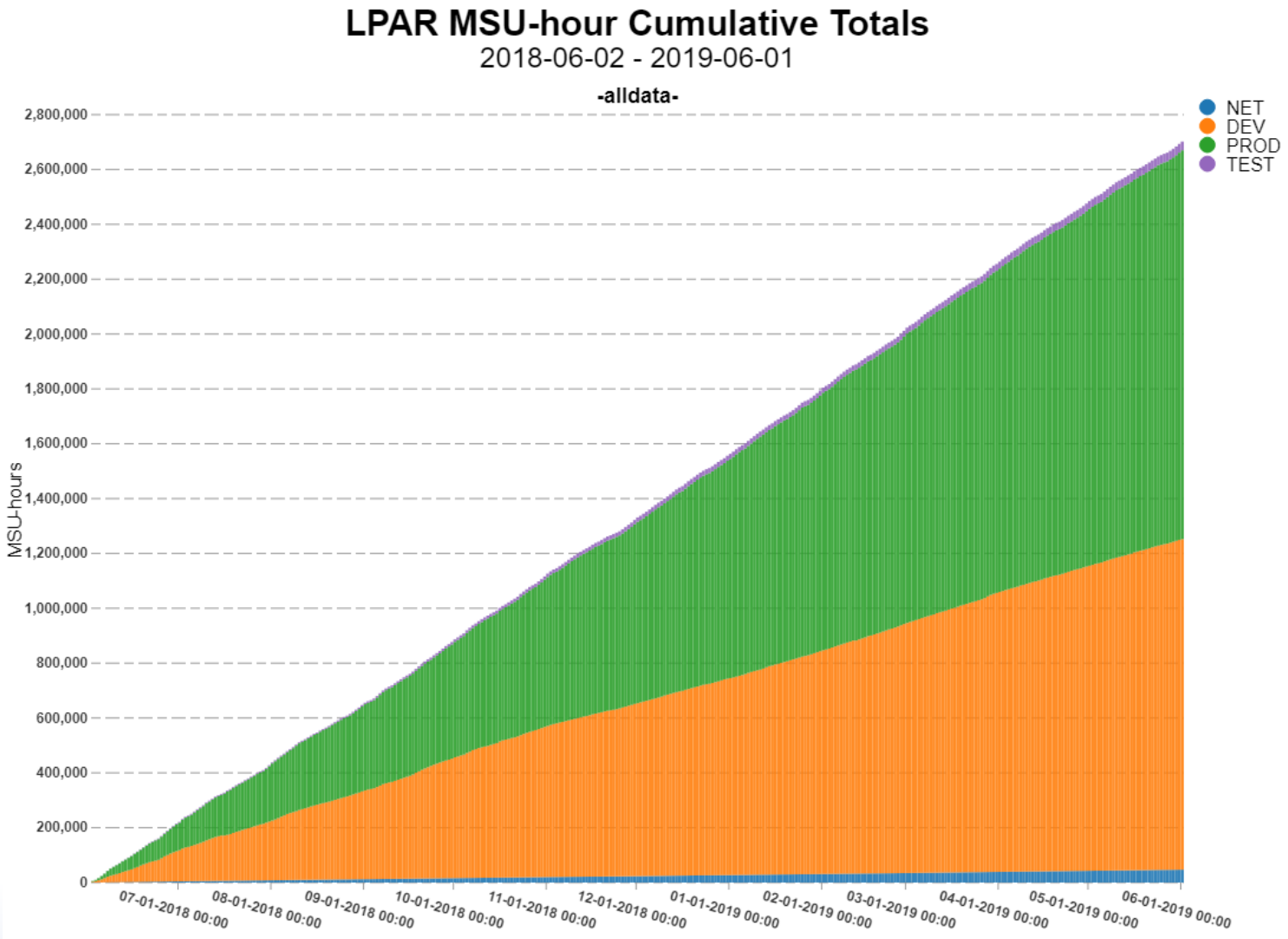
Tailored Fit Pricing: The Basics

- See 2019 announcement:
  http://www-01.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_ca/4/897/ENUS219-014/index.html
- Optional pricing method that does away with R4HA
- Enterprise Capacity Solution
  - Charge is based on total installed capacity
- Enterprise Consumption Solution (great: both abbreviated ECS)
  - Charge is based on total annual consumed MSU-hours
  - "MSU-hours" abbreviated "MSUs" in most places in the doc
  - Because MSUs = Millions of SUs/hour
- Charges based on past 12 months + agreed-upon reduced rate for growth
  - Price per MSU = last 12 months charges / last 12 months total MSUs
  - "at least 50% MLC price/performance for incremental MSU consumed above their baseline"
  - Agreements individually negotiated

MSU Averages Comparisons

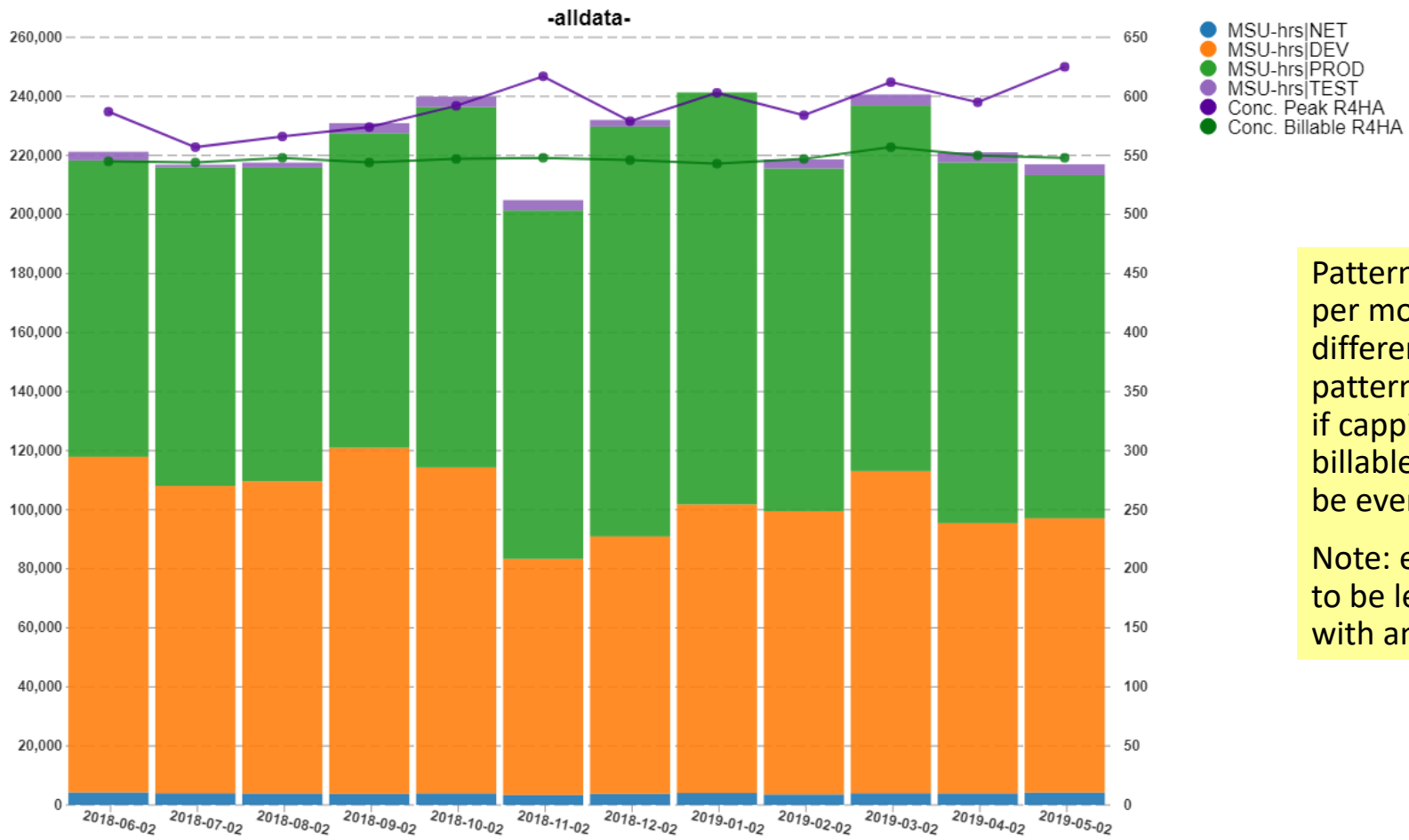


This is a typical monthly analysis with R4HA workload license charges. License cost based on peak billable R4HA hourly MSU consumption. Capping can be used to limit billable MSUs.



With TFP, all utilization contributes to the software cost.

MSU-hour Totals by MLC Month



Pattern of total usage per month is likely much different than the pattern of the R4HA and if capping is in place, billable R4HA usage may be even more different.

Note: expect TFP billing to be level each month with annual true-up.

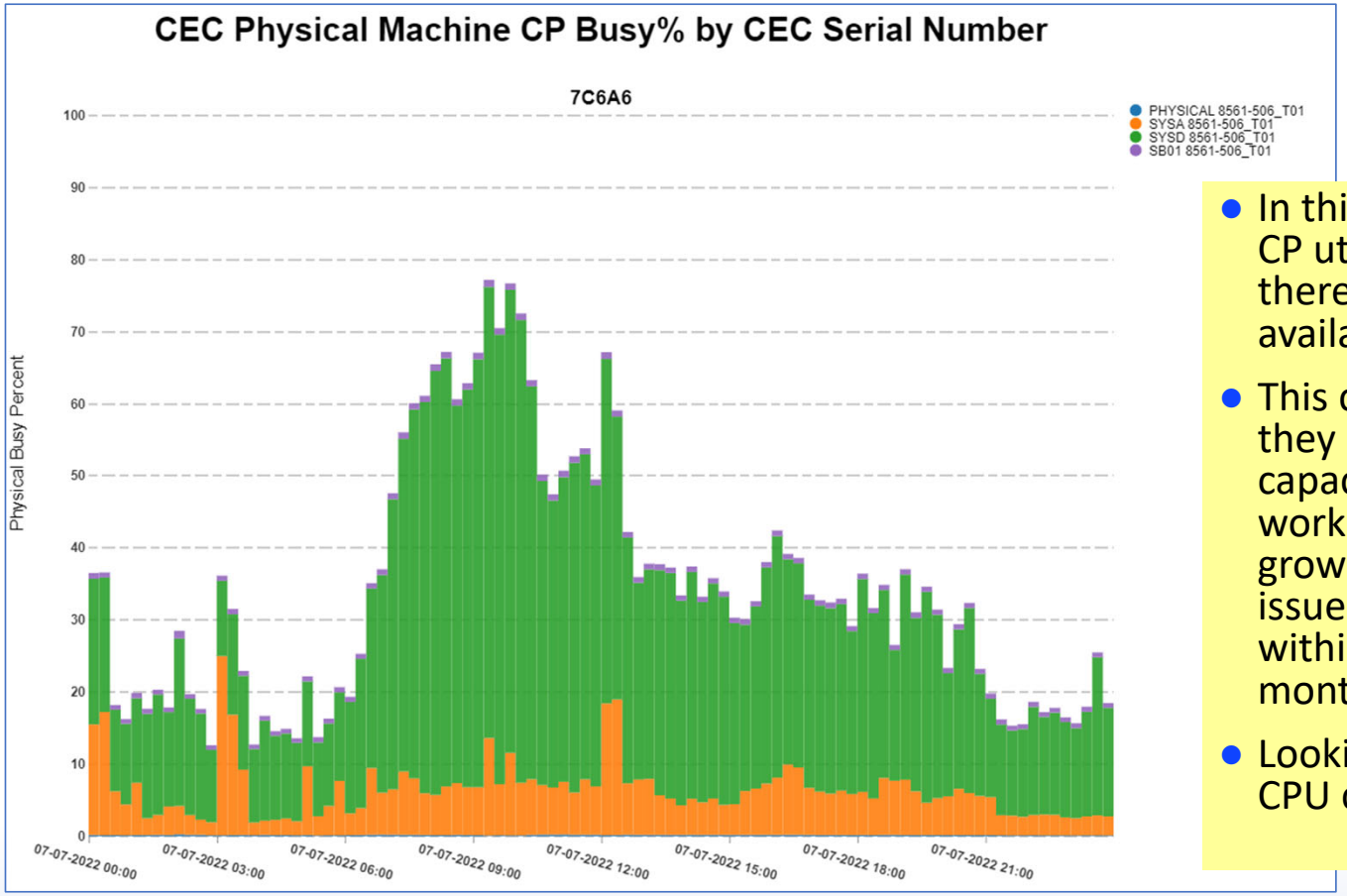
TFP Performance Implications

- Many performance & capacity decisions are driven by software costs so you need to understand how software costs are calculated
- For IBM MLC peak monthly R4HA is most commonly used today
  - So many performance questions are around managing those peaks
  - Work outside the peaks don't impact software costs so common cost savings plan is to move work from the peaks to lower-utilization times
- With TFP Consumption Solution all hours count, so potentially more need for performance tuning a wider variety of time frames
- With TFP Capacity Solution you're managing performance such that your workload fits within your capacity

See also the slides from our June Webinar: "Is Tailored Fit Pricing a Good Fit?" (Given by Scott with special guest Al Sherkow)

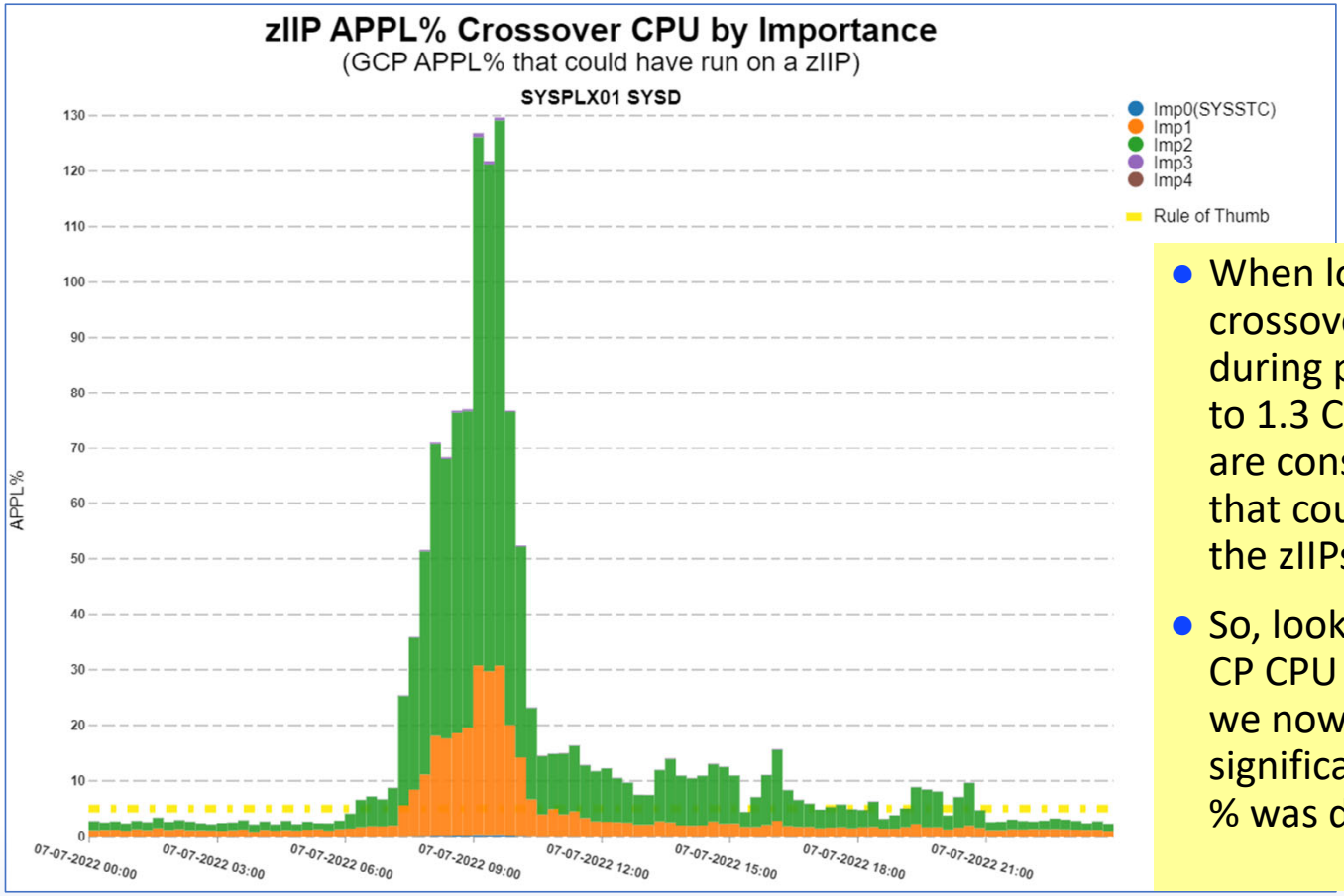
Crossover Can Drive Up Costs

Reminder: zIIP crossover can drive up costs - CP physical utilizations



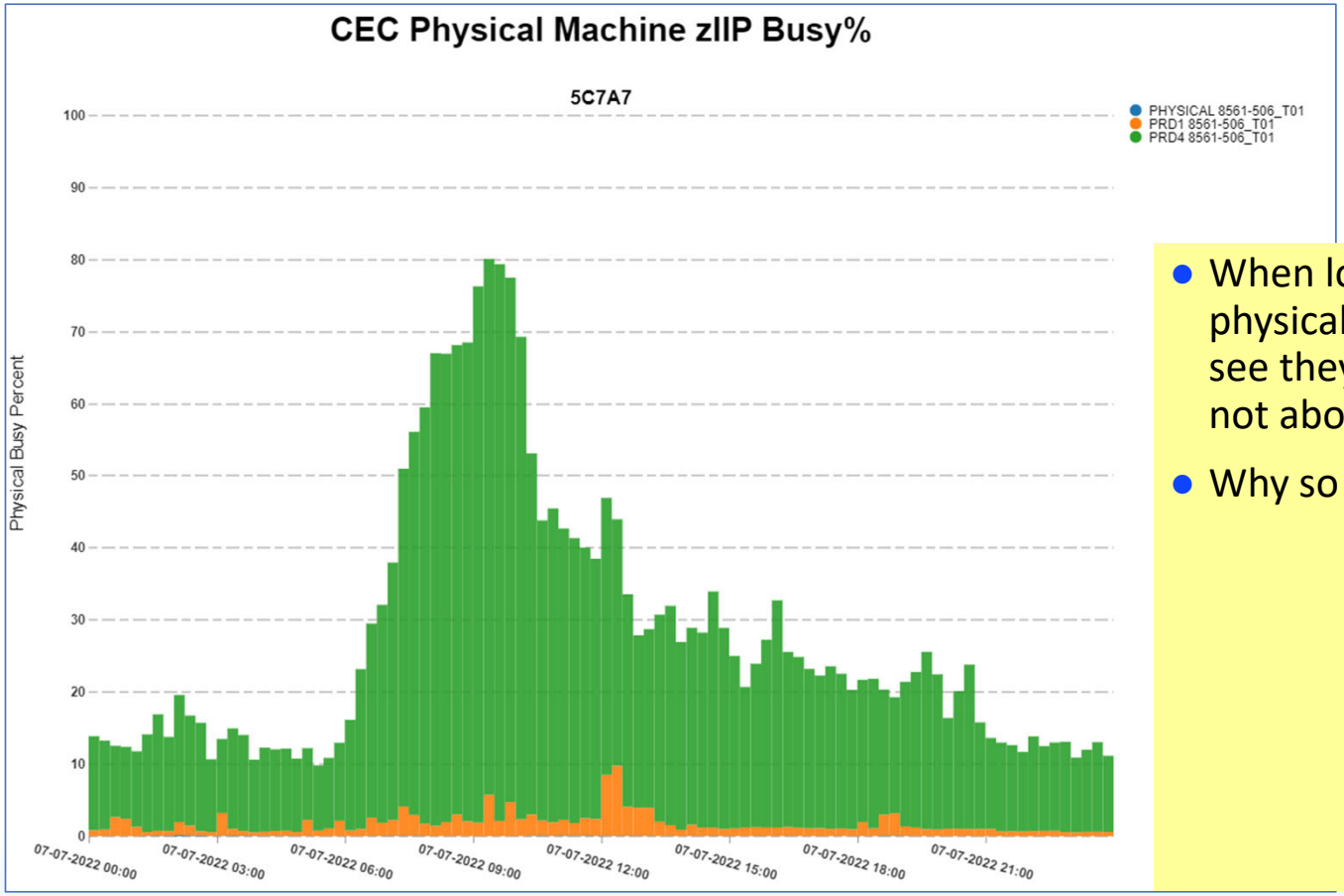
- In this example, the CP utilizations show there is some available capacity.
- This customer felt they had enough CP capacity, but workloads are growing. Capacity issued expected within the next few months.
- Looking to head off CP CPU capacity issues.

Reminder: zIIP crossover can drive up costs - Amount of crossover



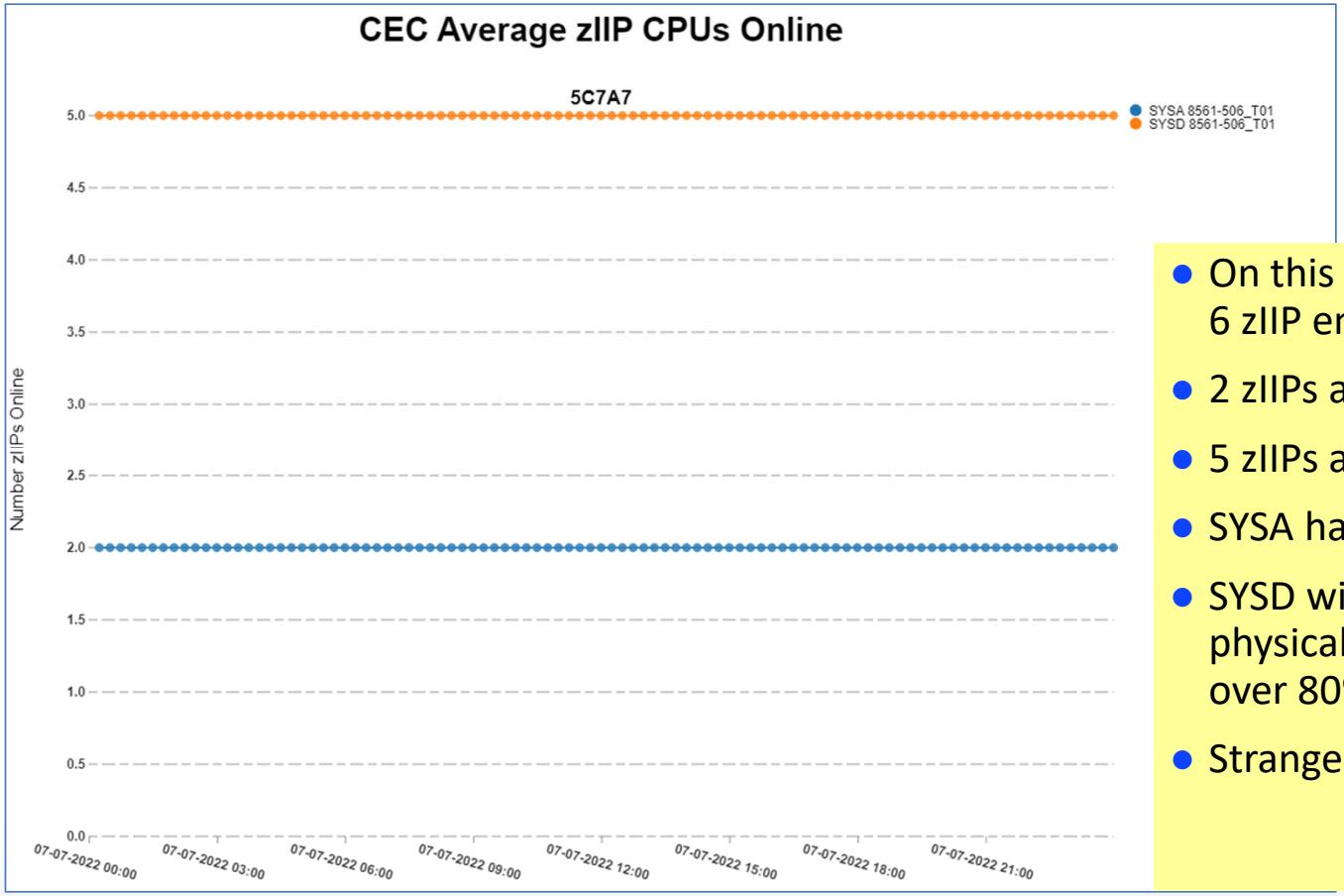
- When looking at zIIP crossover, we see that during peak periods, up to 1.3 CP CPUs of capacity are consumed with work that could have run on the zIIPs.
- So, looking back on the CP CPU utilization chart, we now know a significant portion of CPU % was due to crossover.

Reminder: zIIP crossover can drive up costs - zIIP physical utilizations



- When looking at zIIP physical utilizations, we see they are high, but still not above 80%.
- Why so much crossover?

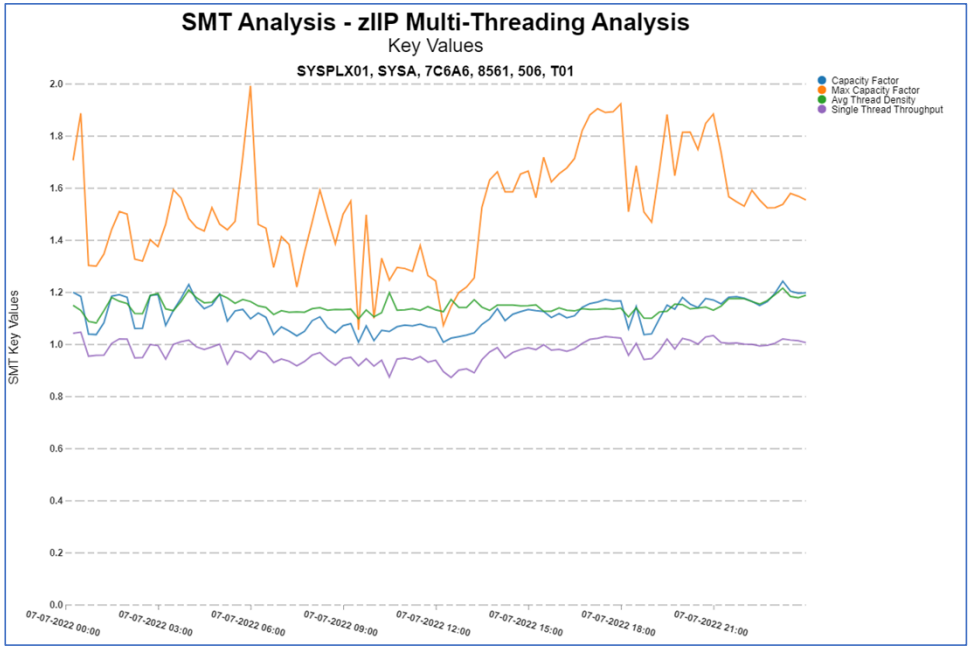
Reminder: zIIP crossover can drive up costs - zIIP configured to each partition



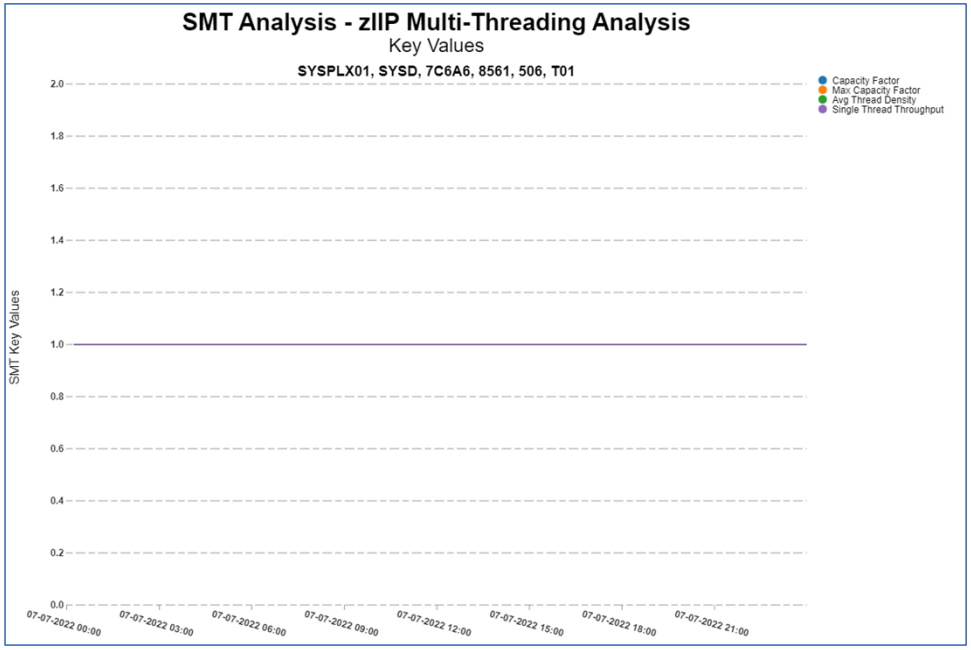
- On this machine there are 6 zIIP engines
- 2 zIIPs assigned to SYSA
- 5 zIIPs assigned to SYSD
- SYSA has low zIIP usage
- SYSD will never have a physical zIIP utilization over 80%
- Strange configuration

Reminder: zIIP crossover can drive up costs - SMT

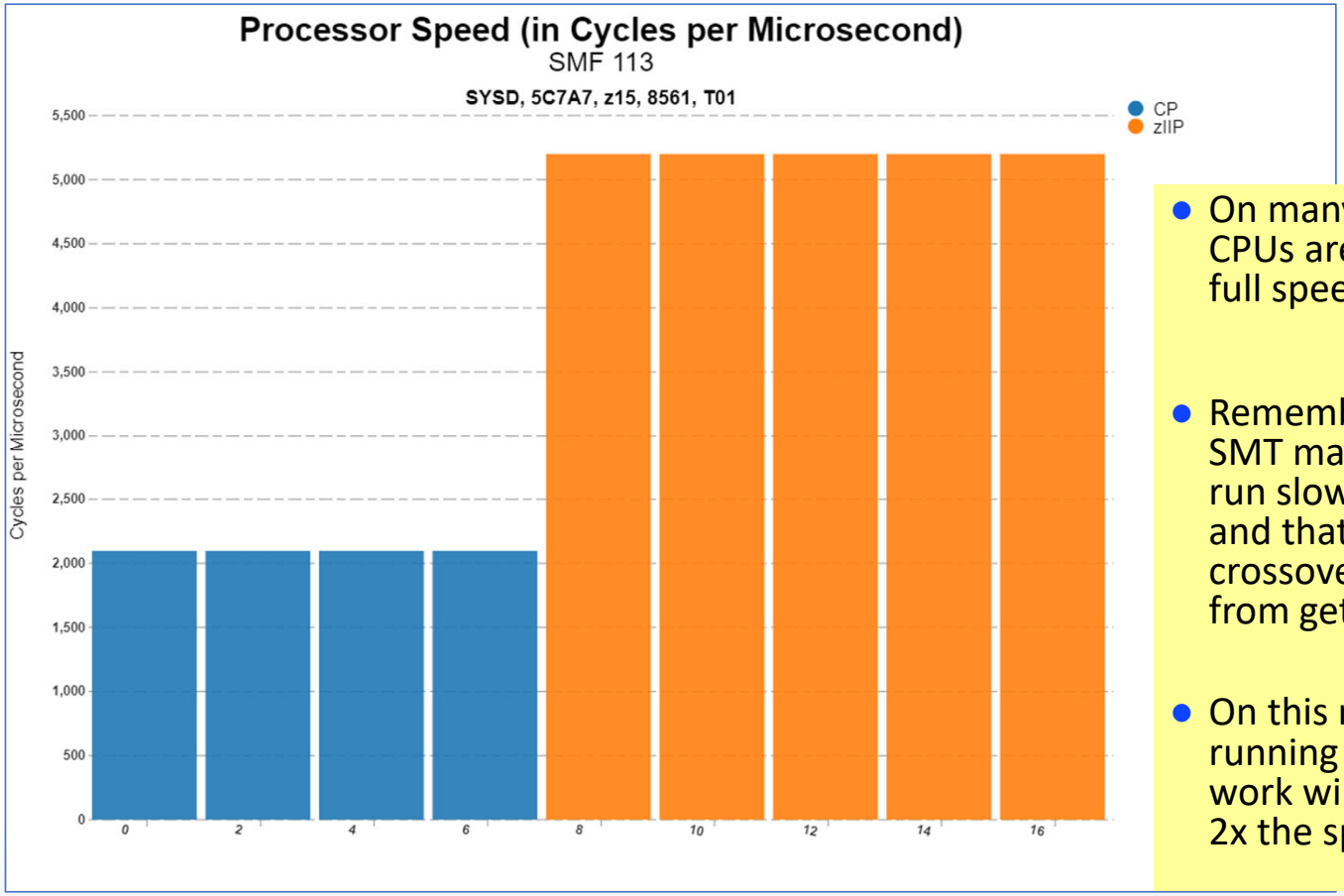
SYSA – SMT turned on



SYSD – SMT turned off (Strange)



Reminder: zIIP crossover can drive up costs - Processor Speeds



- On many machines the CP CPUs are not running at full speed.
- Remember that although SMT may make the work run slower that non-SMT, and that preventing crossover may delay work from getting dispatched…
- On this machine, once running on the zIIPs, the work will run at more than 2x the speed

Reminder: zIIP crossover can drive up costs - Recommendations

- In this example, a consulting group recommended
  - Don't give biggest partition access to all the zIIPs
  - Make sure small, low zIIP usage, partition has at least 2 zIIPs
  - Don't turn on SMT for big partition because SMT slows the work down

- Our recommendation
  - Give big partition access to all zIIPs
  - Turn on SMT on big partition.

    Even though work running on SMT zIIPs is slower than non-SMT zIIPs, the crossover is much worse since
    - Consuming CP CPU capacity, MSUs, etc
    - The CP CPUs are less than half the speed of zIIPS

Reminder: Potential Crossover solutions

- Buy more zIIP(s)
  - Always best from a performance perspective
  - They may be cheap relative to the software costs they offset
- ***Ensure zIIPs are optimally configured to LPARs, and weighted correctly***
- Use HONOR PRIORITY=NO on specific service classes
  - Prevents zIIP-eligible work in those service classes from crossing over to the GCPs
- Increasing ZIIPAWMT may be useful in some limited cases
  - Usually not recommended but we have seen situations where it was useful
- Run less zIIP work
  - This seems counter-productive: we're generally heading towards more zIIP work, not less
- Enable SMT
  - More capacity, although likely impact to individual thread performance

SRB – System Recovery Boost

System Recover Boost

- Overall goal: provide additional capacity and performance to better recover from certain planned or unplanned events
  - Shutdown faster
  - Startup faster
  - Faster DR site switch
  - Faster sysplex recovery
- Requires z15, or higher, machine
- No additional charge for basic System Recovery Boost
  - Optional extra cost item discussed in just a moment

Types of boost

- Speed Boost: *GP engines run as full speed engines*
  - Applicable only to sub-capacity systems
  - I.E. this is of no use to 8561-7xx and 8562-Z0x systems
  - Only boosted LPARs run as full speed!

- zIIP Boost: *zIIPs will run work not eligible to run on zIIPs*
  - I.E. ***any*** workload may be dispatched to a zIIP
  - Reserved (but physically available) zIIPs brought online to the boosted LPAR
  - Note that zIIPs always run full speed
  - What if you want more zIIPs for zIIP boost?

Boost classes = triggering events

Class	Where	Duration	WLM Work Routing
Shutdown	Single system	30 minutes	Routes work away from boosted system
IPL	Single system	60 minutes	Routes work to boosted system
Recovery Process	Multiple systems	<5 minutes	No change

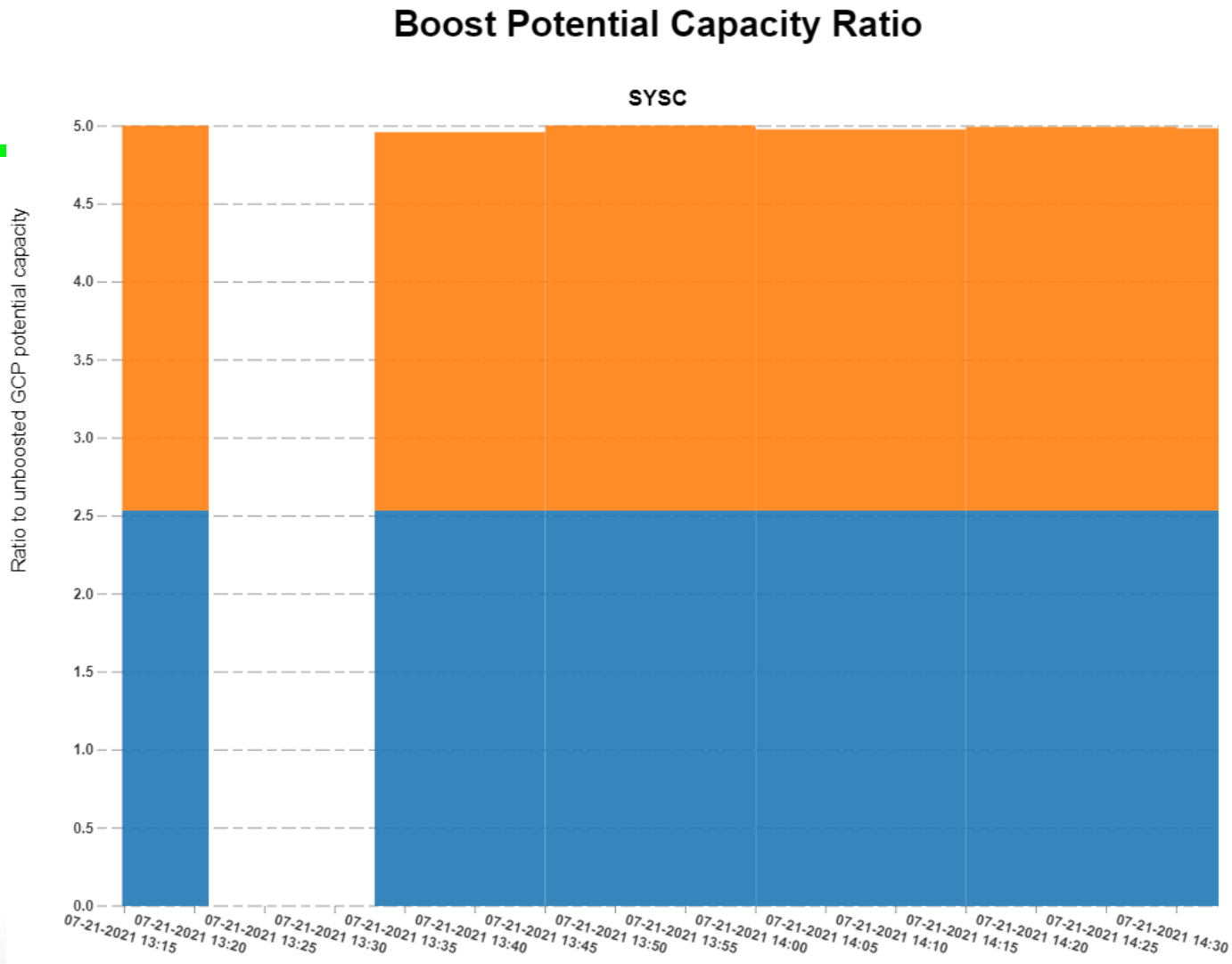
- Recovery Processes
  - CF data sharing member recovery
    - Triggered by disconnection from lock structure while locks were held
  - Sysplex partitioning
    - System removed from a sysplex
  - CF structure recovery
    - Structure rebuild, duplex failover, or reduplexing
  - HyperSwap
    - Recovery from storage controller failover
  - Note: z/OS 2.5 added additional boosts

LPARs being boosted

- Seems obvious: they get more work done!
- But may need changes to prepare them to get more work done
  - Maybe: define reserved (offline) zIIPs to be brought online during boost period
  - May need to add zIIP weight to be able to access newly online zIIPs
- Increased zIIP usage could cause some zIIP work to cross over to the GPs
  - Work with HONORPRIORITY=NO prevented from crossing over
  - So *may* want to consider changing *from* HONORPRIORITY=NO during IPL boost
    - Probably unnecessary in most cases
    - Recovery process boosts likely short enough to not be a major concern
    - Work should be routed away from the LPAR during shutdown boost anyways
- Boosted LPARs over-achieving work won't be capped to help discretionary
- New SMF interval started when boost starts/ends

LPARs *not* being boosted

- These LPARs *could* be negatively impacted
- CPU cache effectiveness may be impacted by boosted LPARs (at least theoretically)
  - Speed boosted LPARs do more work per unit of time on the GPs
  - zIIP boosted LPARs may drive more work to zIIPs
  - zIIP caps ignored during boost periods
- Higher physical zIIP utilization may impact non-boosted LPAR's ability to get work dispatched on zIIPs
  - Could potentially lead to more crossover
  - zIIP caps ignored but weight enforcement still applies
  - May need to change relative weights during boost periods, if trying to protect unboosted LPARs
- Resource Group caps with sysplex scope don't count work running on boosted LPARs
  - May allow more than expected work to run on unboosted systems
- Conversely:
  - If boosted systems consume less of GP capacity, might help non-boosted systems



How much extra capacity did we get from the boost? Potentially SYSC could have done ~5x the GP work compared to what it could have done un-boosted!

This report was tricky because it considers how much the other LPARs are using hence the slight variations you see in the ratios.

Things that limit SRB's effectiveness

- Full speed GCPs eliminates speed boost
- Busy zIIPs
  - Less available capacity to the boosted LPAR
- Single physical zIIP shared among several LPARs
  - PR/SM will still move the zIIP between LPARs limiting the time the boosted LPAR might be able to use it
- Few or no zIIPs available to the boosted LPAR

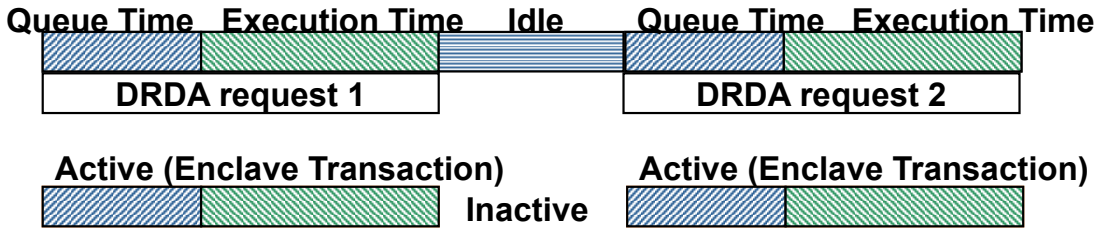
Scott's Opinion: In the 2020s, almost every machine should have at least 2 zIIPs!

Consider Velocity Goals for DB2 High Performance DBAT

What is DB2 DBAT? (Very high level)

- A high-performance DBAT is a database access thread that stays associated with a remote connection at transaction boundaries, rather than being pooled.
  - Basically, inactive connection support
  - DBAT connection terminated after 200 uses

- Non-DBAT

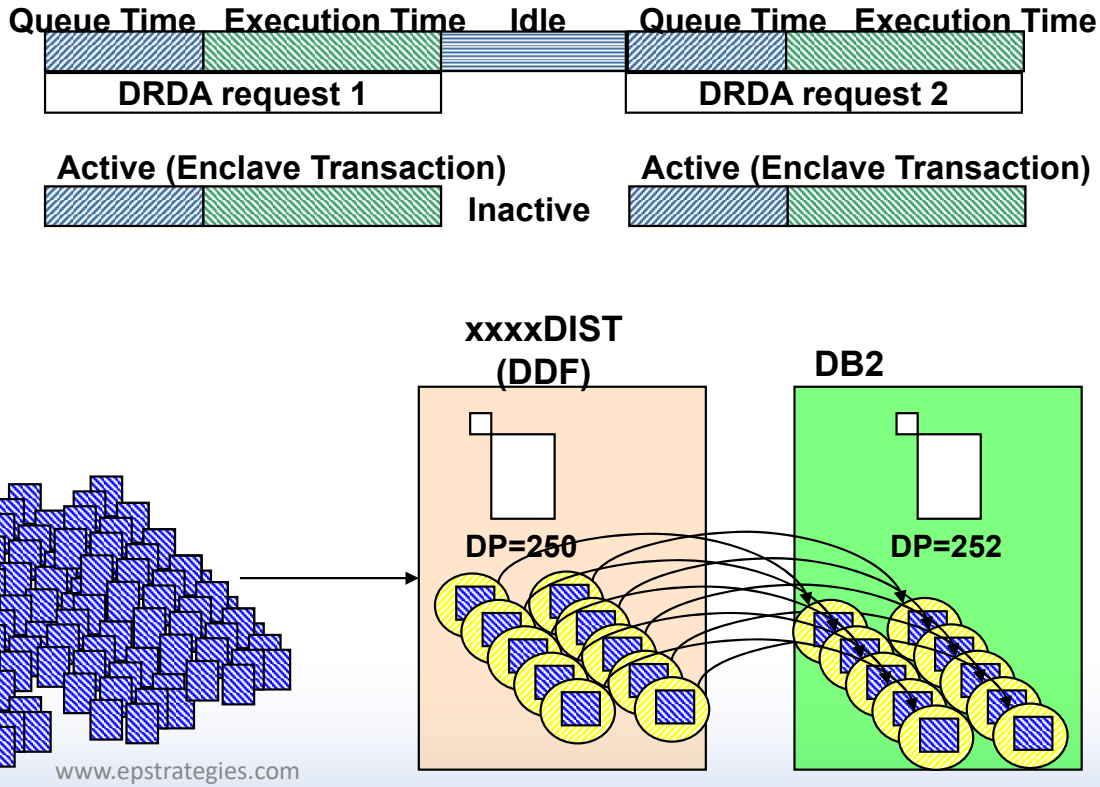


- DBAT

Traditional / Typical DB2 DDF Enclave

- DDF transactions are managed by WLM as independent enclaves
  - The life of the enclave is the life of the transaction

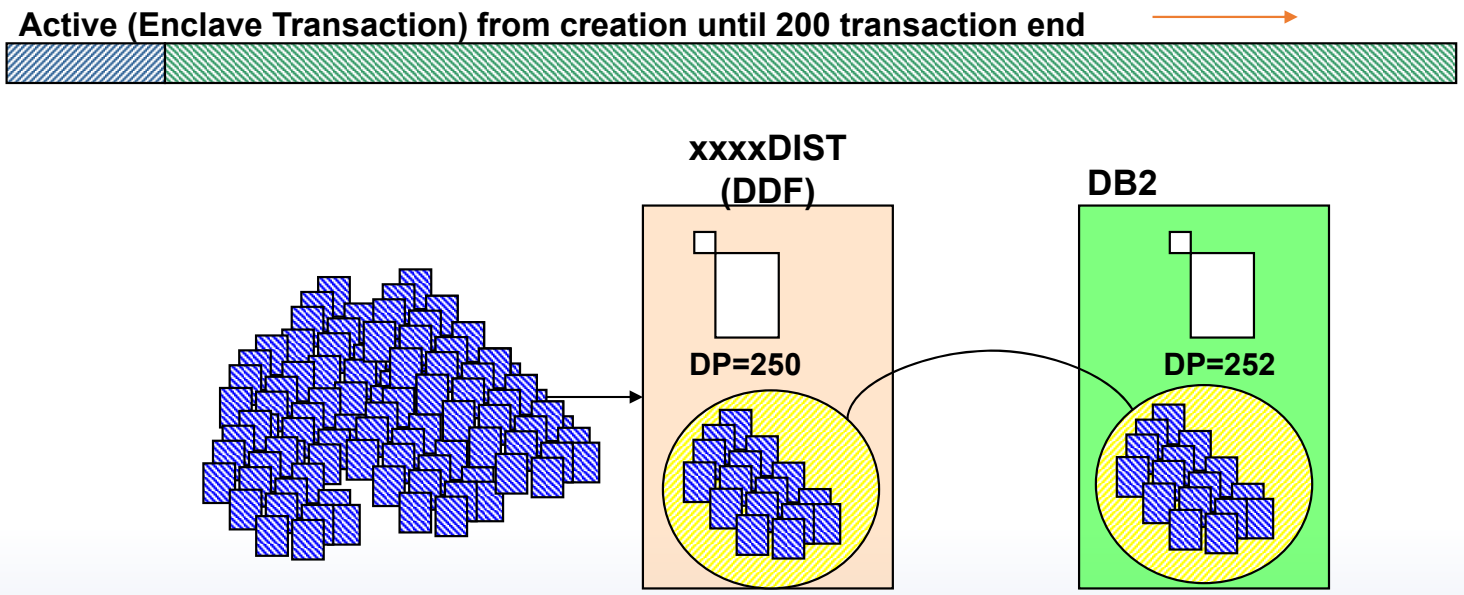
- In this case, 200 transactions come in, and for each transaction an enclave is created
- When each transaction ends, the enclave for that transaction is deleted, and an ended transaction and response time are posted to the WLM Service Class period
- WLM response time goals are best since each transaction is measured and considered by WLM



DBAT DB2 DDF Enclave w/ PH34378

- The DBAT transactions still run in independent enclaves
  - But now the life of the enclave is for as long as up to 200 transactions in the enclave
  - Thus, you may have 200 ended transactions, but from a WLM point-of-view, the 200 transactions are part of a longer running enclave

- In this case, 200 transactions come in, and all are associated with a single enclave.
- After the last of the 200 transactions end, enclave is deleted, and one ended transaction is posted to the WLM service class period. Response time is from start of first transaction to end of last ended transaction
- Velocity goals may be better



Recommendations for DB2 DBAT Workloads

- Find out from your DB2 group if they are using High Performance DBATs
- If so, consider segregating these DBAT trans to separate WLM service classes
  - Assign those service classes velocity goals

Capture Ratio is Still Important

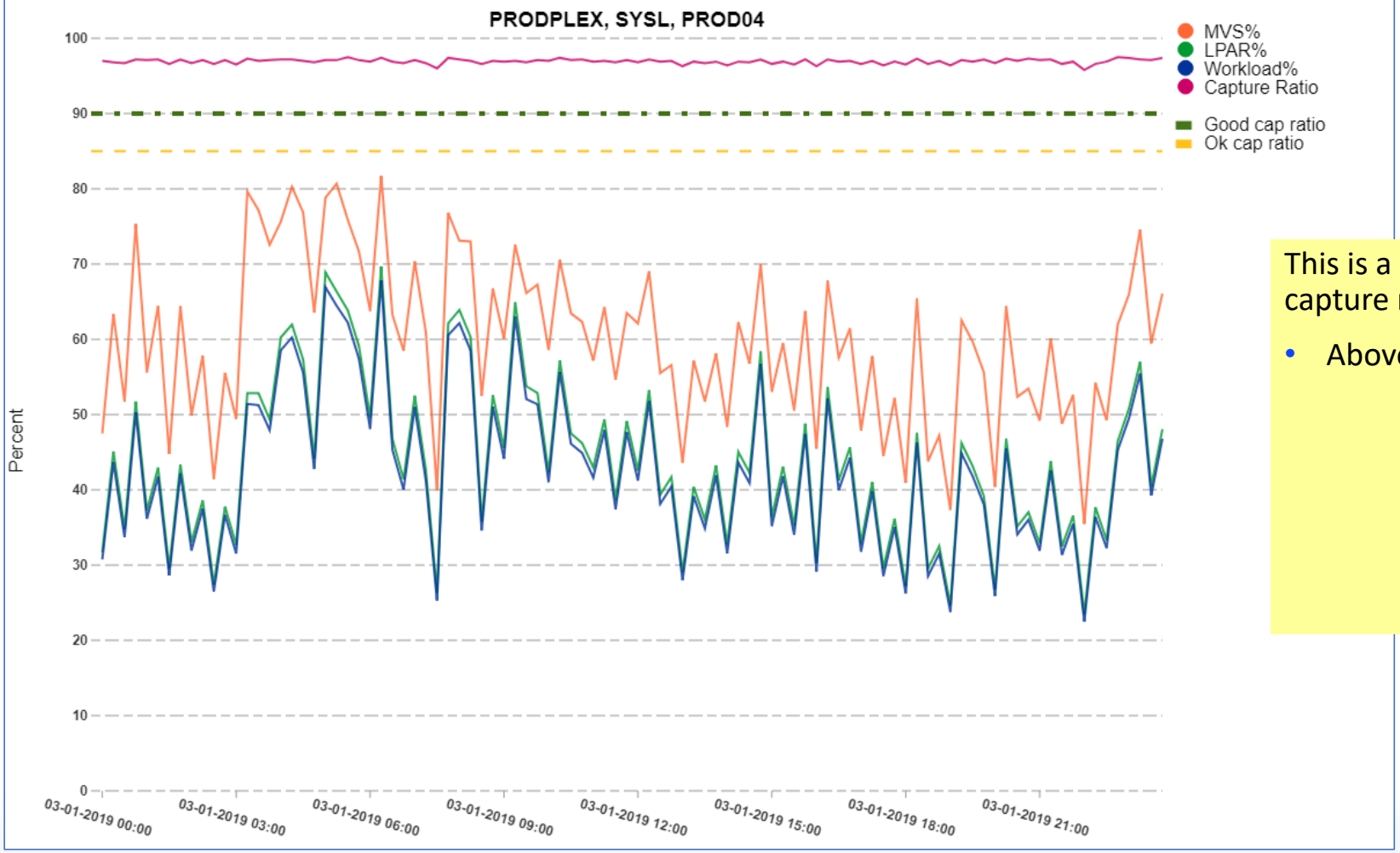
Capture Ratio

- Captured time = CPU time attributed to particular service class periods in the type 72 records
- Capture Ratio = $\frac{Captured\ time}{Effective\ dispatch\ time}$

Target Capture Ratio: >90%

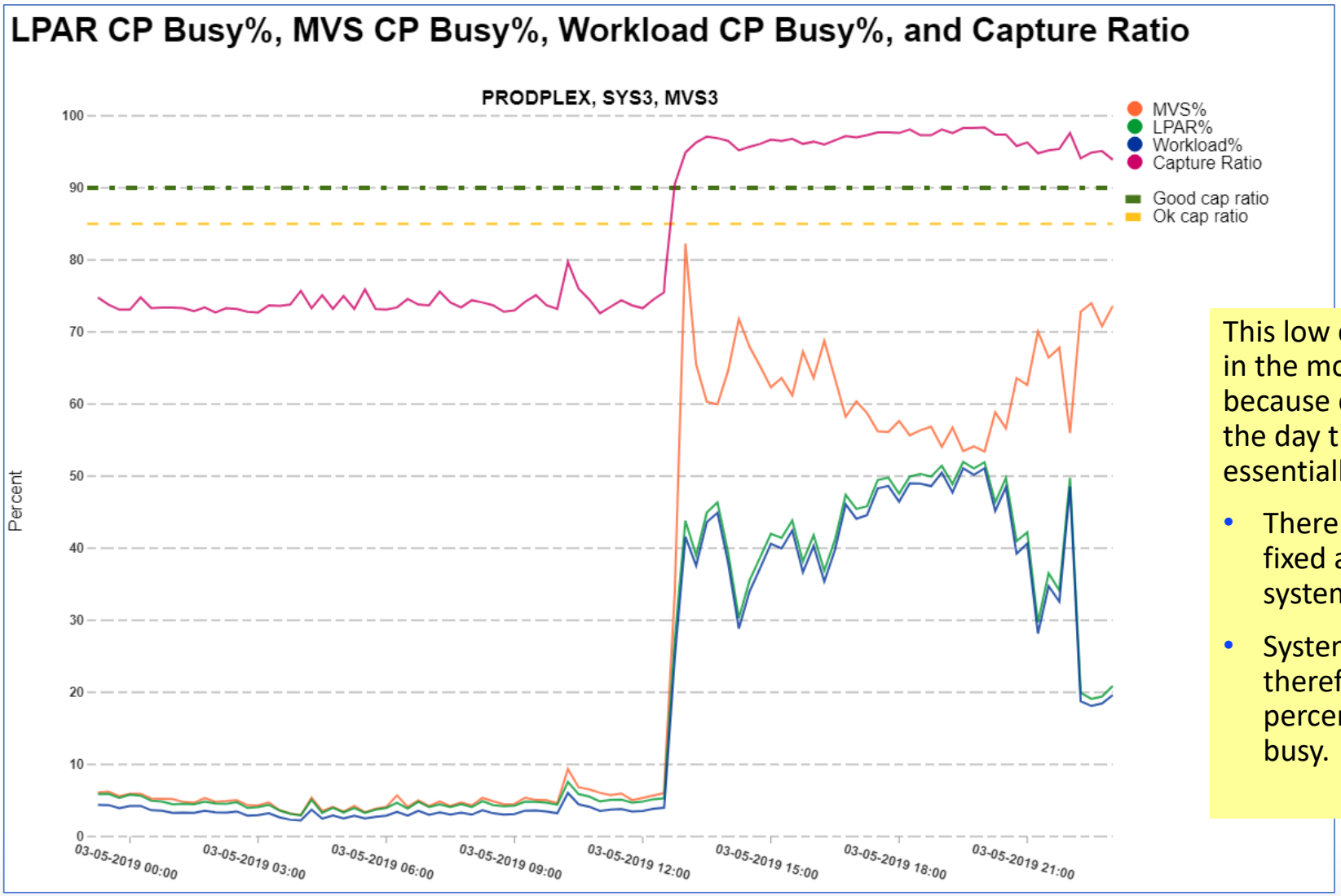
- Uncaptured time = $(Captured\ time) - (Effective\ dispatch\ time)$
- Some uncaptured time is normal and is because z/OS cannot directly attribute some system-related work to a specific workload
  - Periodic system management
  - Interrupts
  - Contention
  - Paging
  - SLIP traps and system traces
  - Inefficient allocation process

LPAR CP Busy%, MVS CP Busy%, Workload CP Busy%, and Capture Ratio



This is a pretty ideal capture ratio:

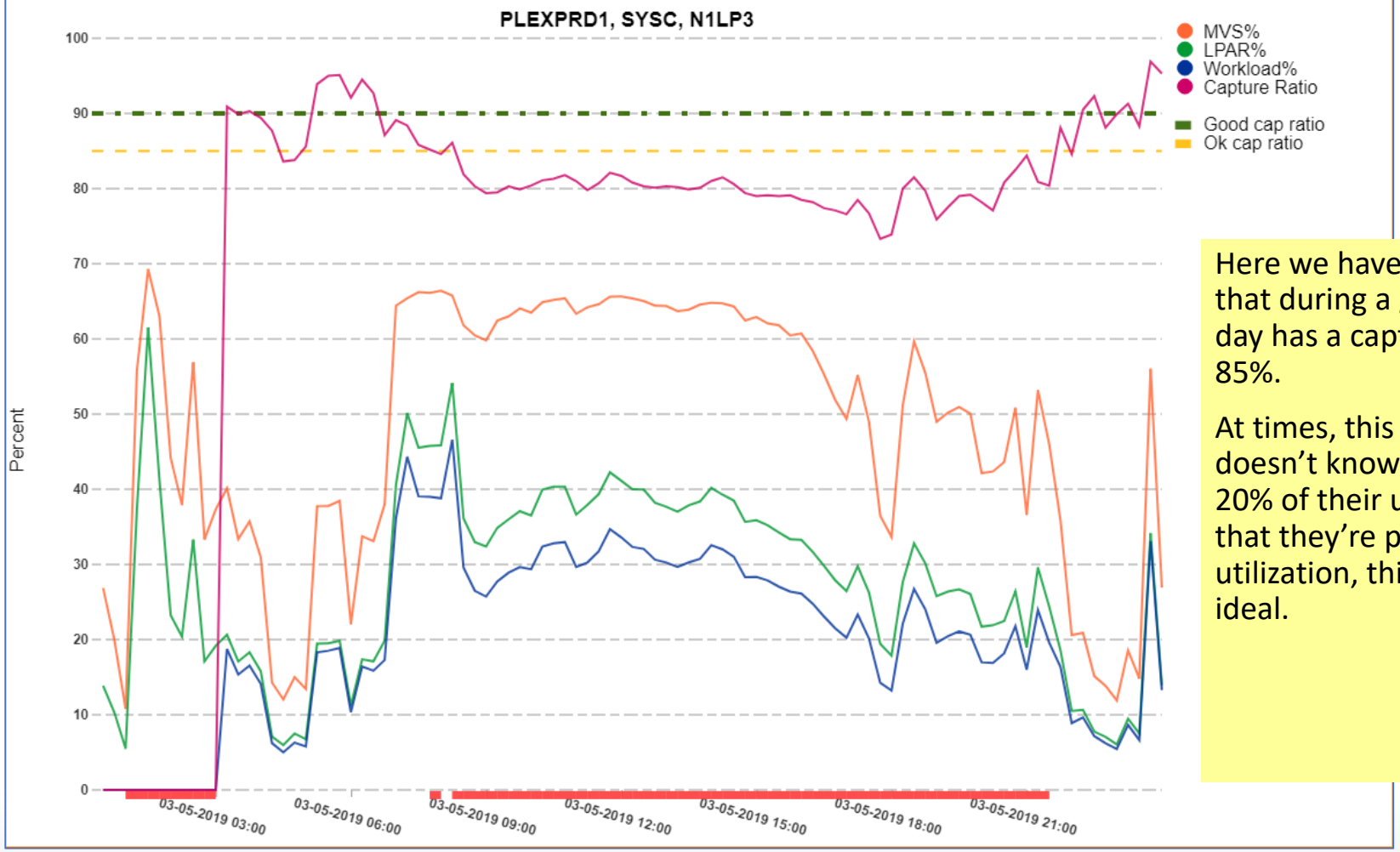
- Above 95% all day!



This low capture ration in the morning is ok because during part of the day the system was essentially idle

- There will always be a fixed amount of system overhead.
- System overheads are therefore a larger percentage of the low busy.

LPAR CP Busy%, MVS CP Busy%, Workload CP Busy%, and Capture Ratio



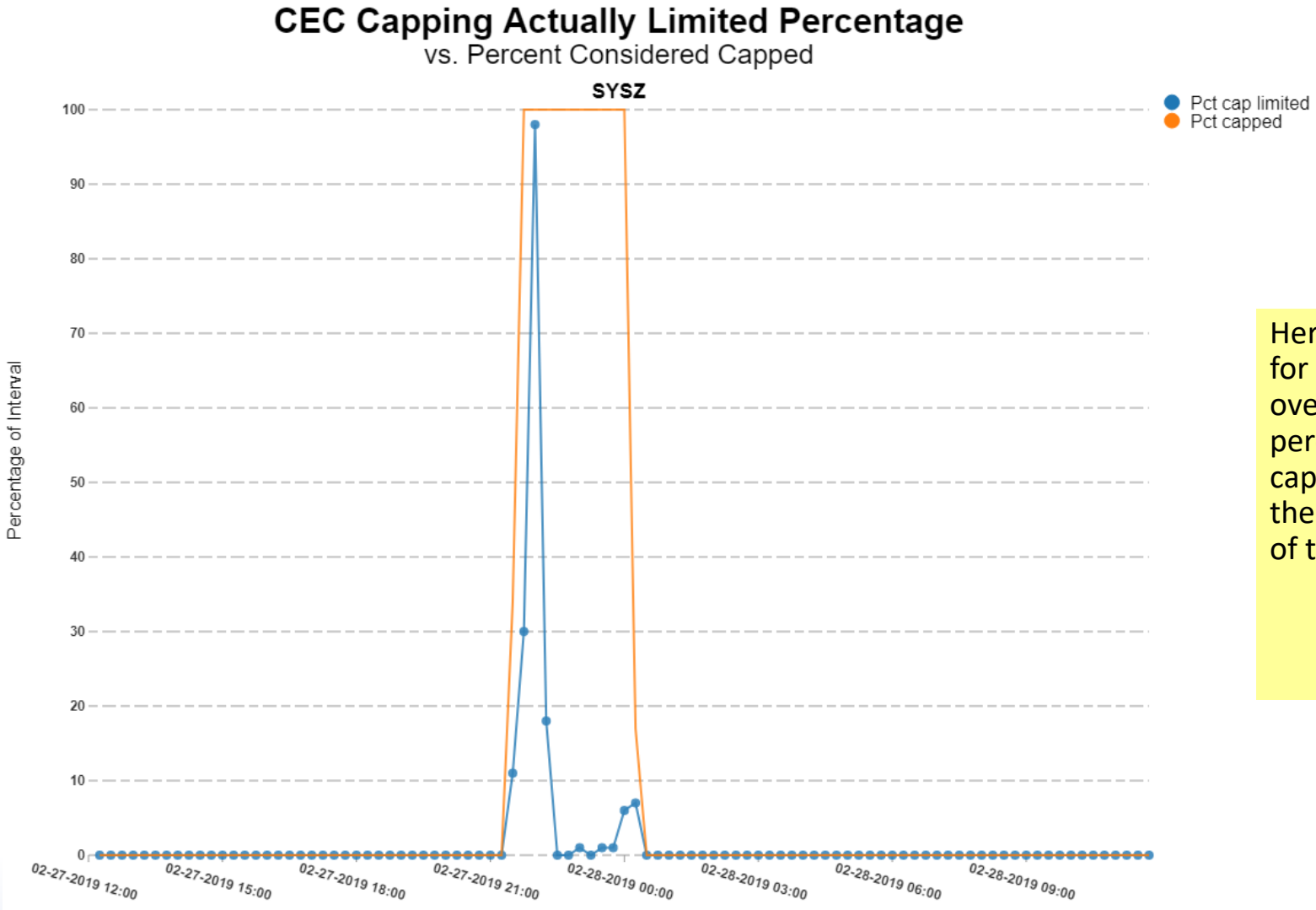
Here we have a busy LPAR that during a good part of the day has a capture ratio below 85%.

At times, this customer doesn't know what is driving 20% of their utilization. Given that they're paying for that utilization, this is less than ideal.

Debugging Capture Ratio Issues

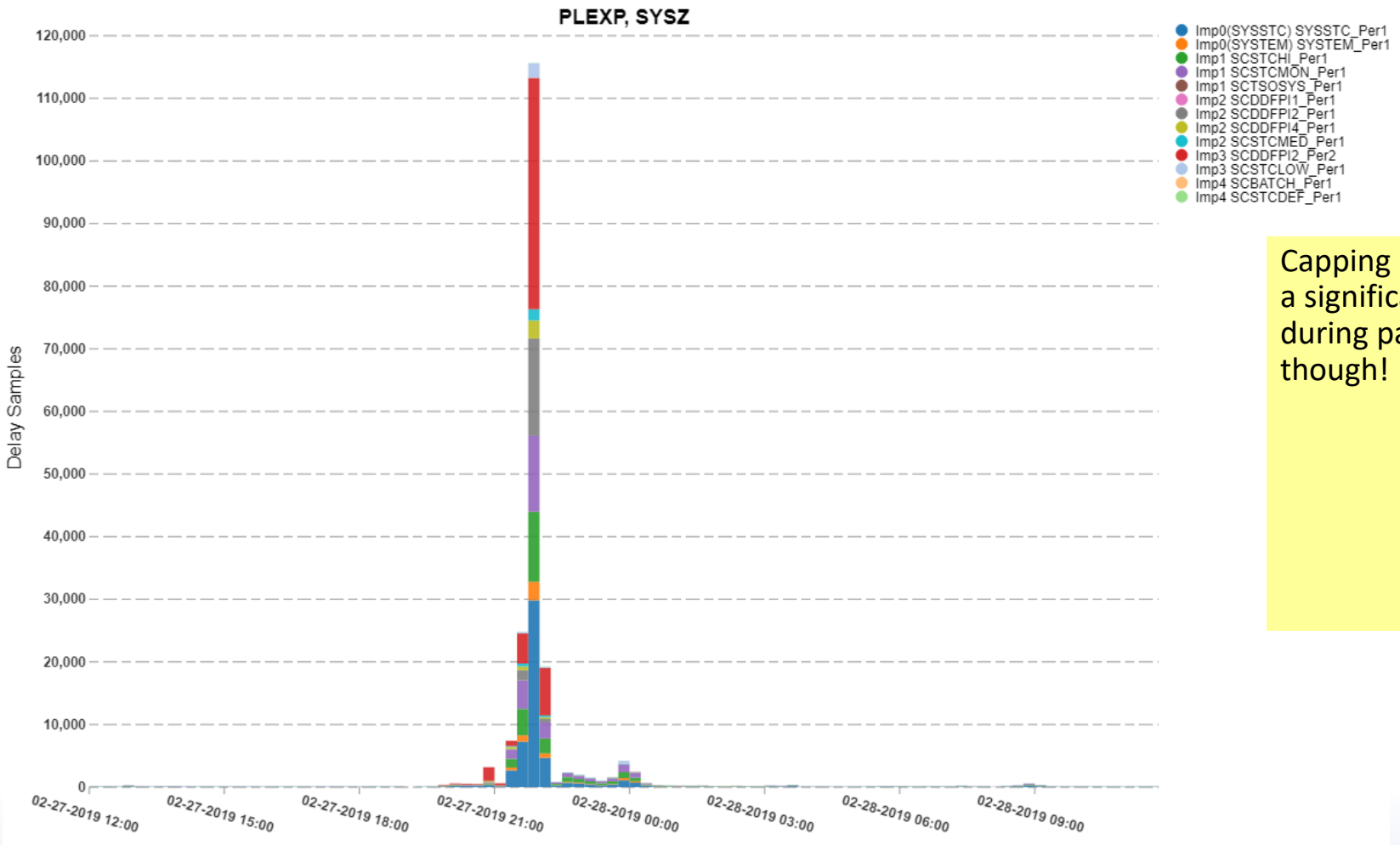
- Debugging capture ratio issues are not easy
  - If knowing what was contributing to capture ratios was easy, IBM would be measuring potential causes in the SMF data
- Some ideas
  - See if the capture ratio correlates to the LPAR utilization
  - Find out what work / batch / jobs / transactions are running during specific periods of time of interest
  - Sometimes this highlights certain applications or system activity that may be using some system service that results in uncaptured time
- Not easy, but remember, you are being billed on uncaptured MSUs

Capping May or May Not Be Impactful

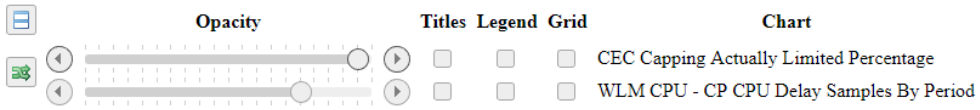


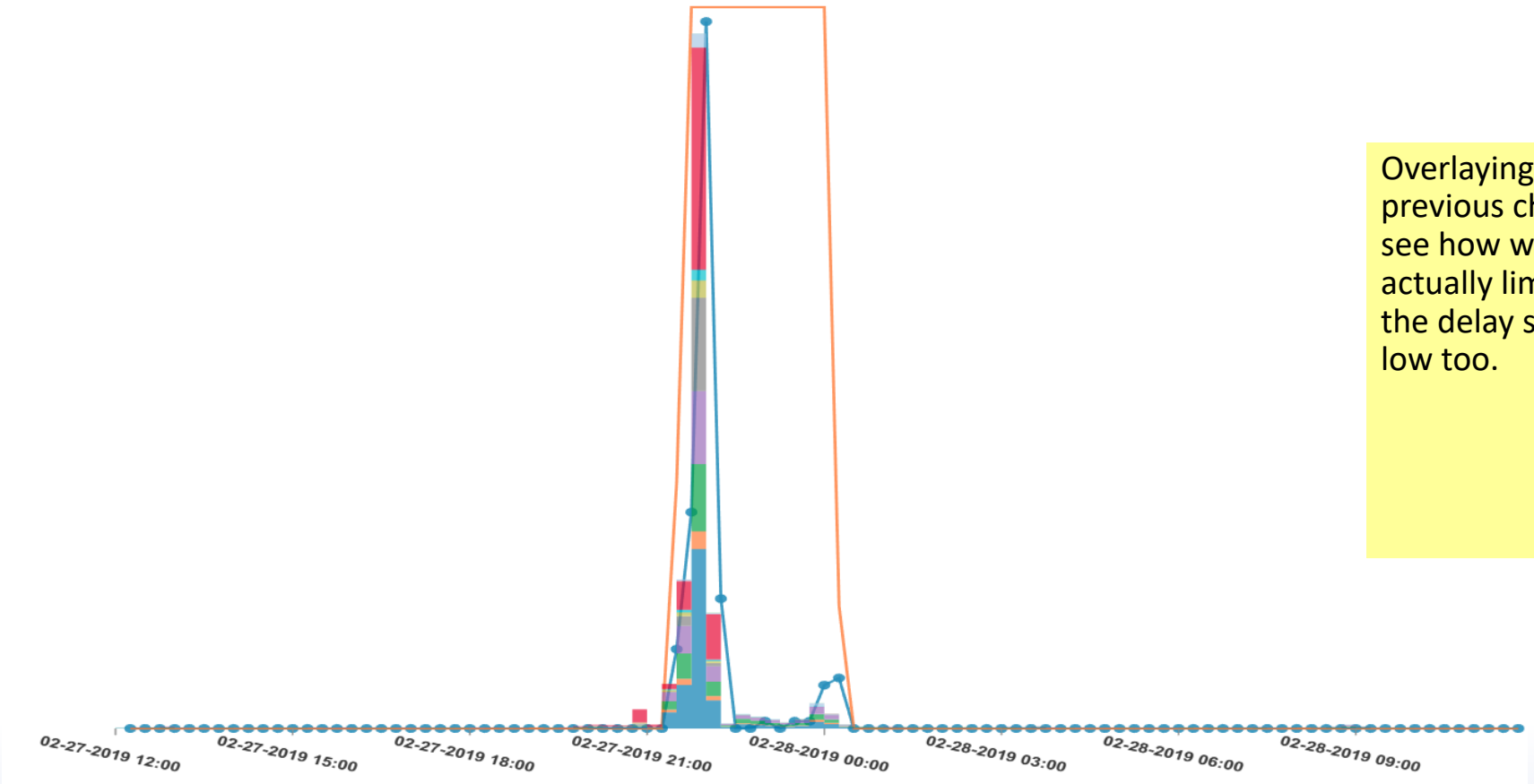
Here SYSZ was capped for a few hours overnight. But the percentage of time the cap was actually limiting the LPAR was not 100% of the capping interval.

WLM CPU - CP CPU Delay Samples By Period

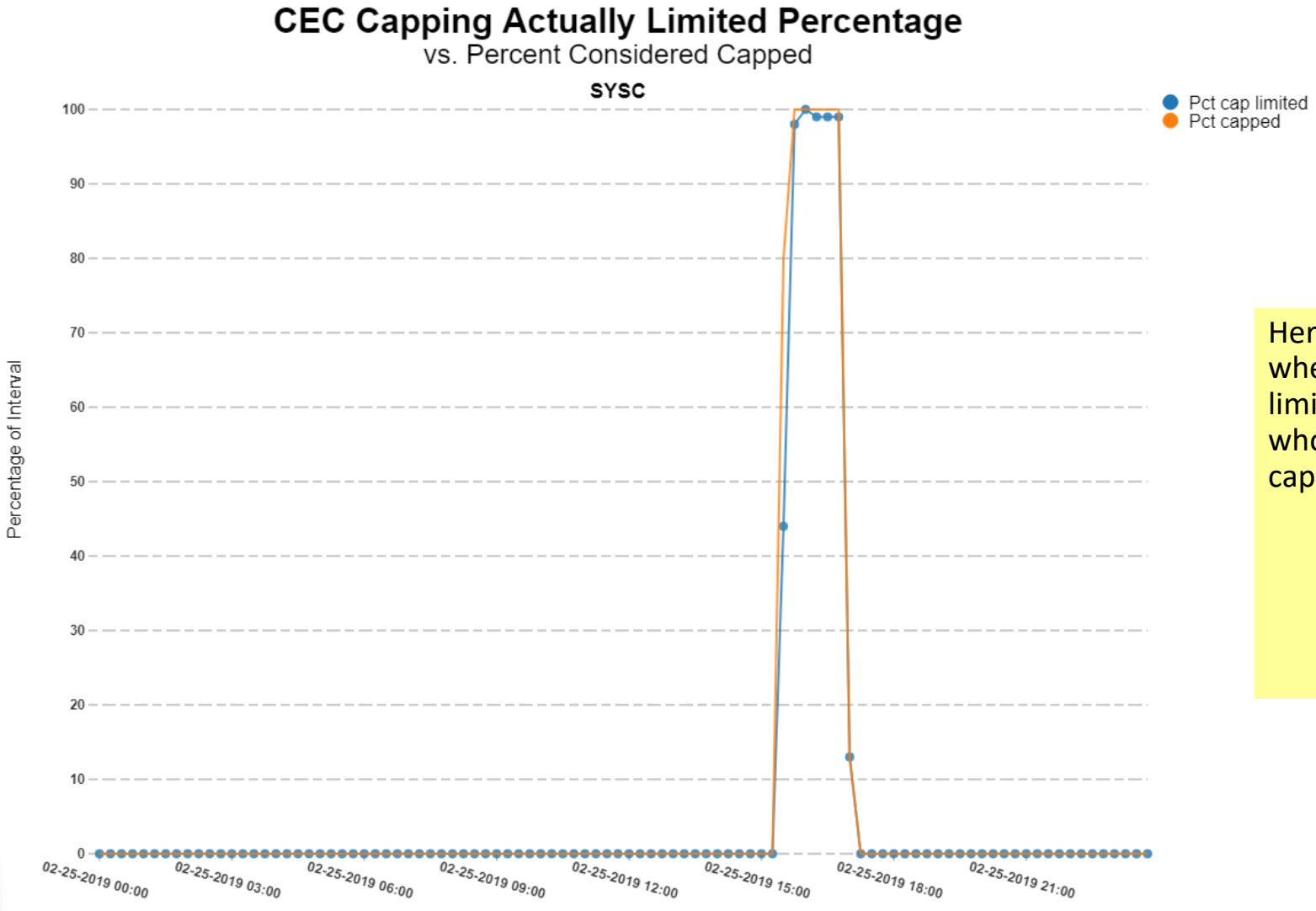


Capping clearly did have a significant impact during part of that time though!



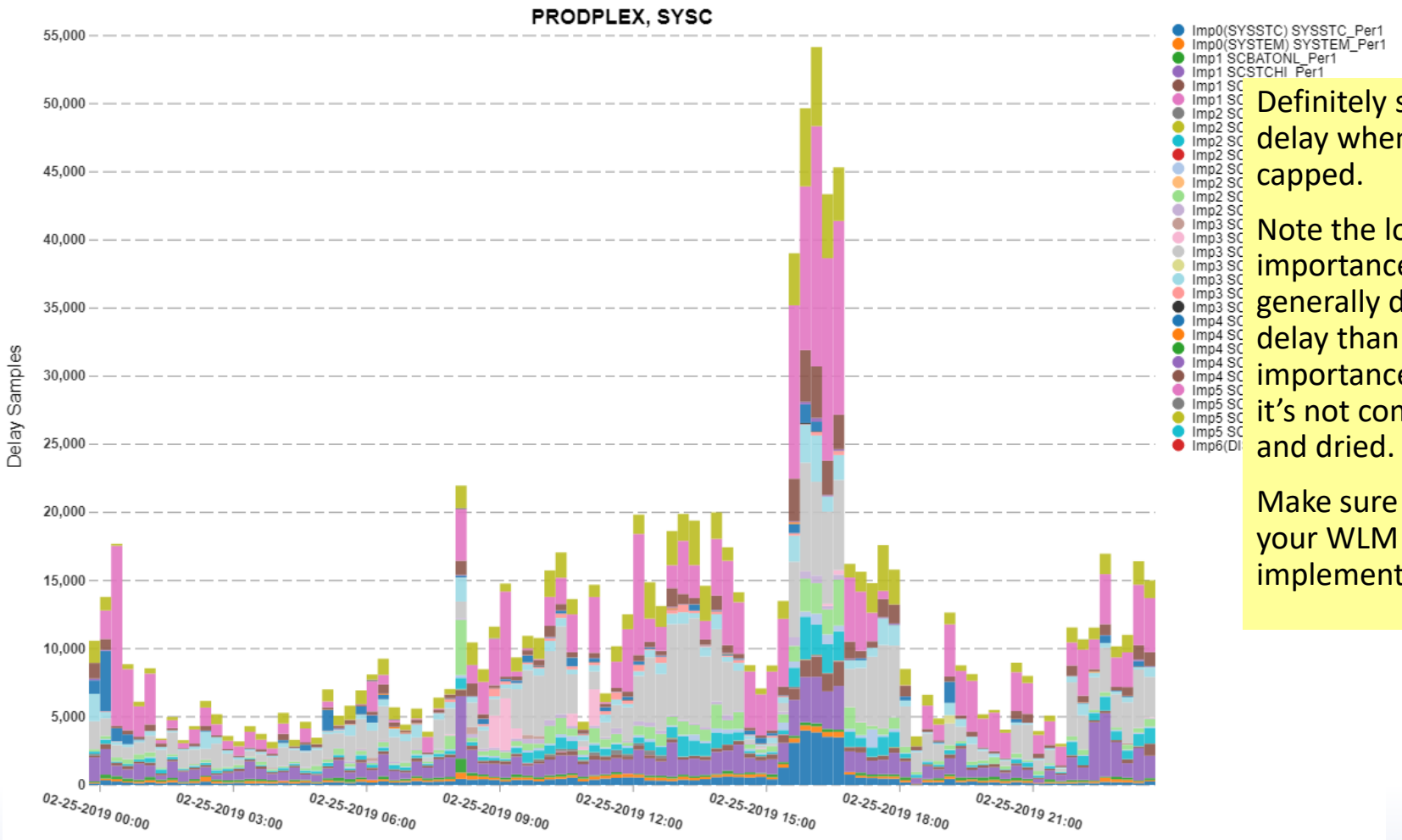


Overlaying the two previous charts we can see how when capping actually limited is low, the delay samples are low too.



Here we see an example where the percent limited is essentially the whole time the LPAR was capped.

WLM CPU - CP CPU Delay Samples By Period



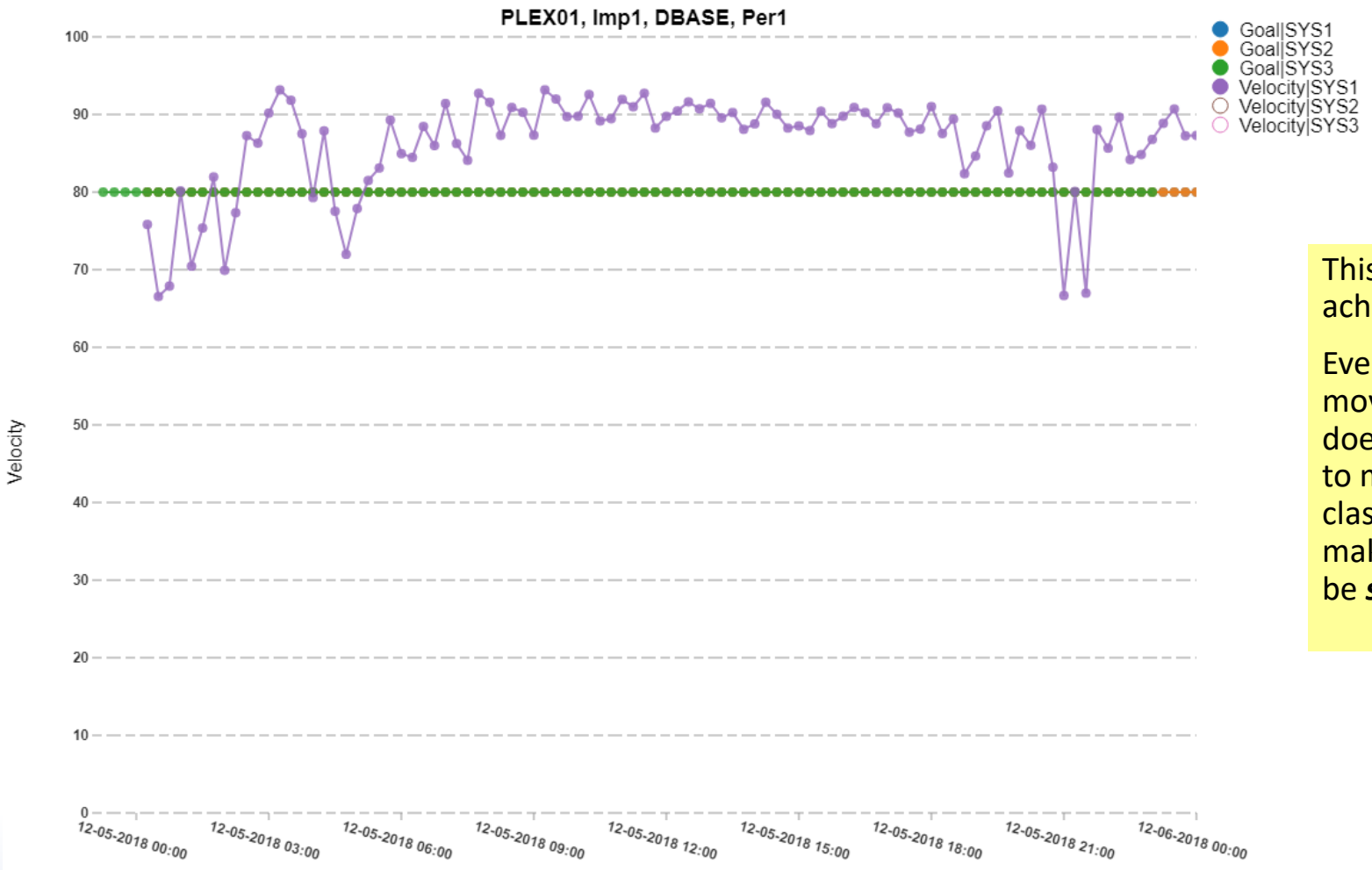
Definitely saw more CPU delay when they were capped.

Note the lower importance workloads generally did see more delay than the higher importance work, but it's not completely cut and dried.

Make sure to review your WLM policy before implementing capping.

Beware High Velocities

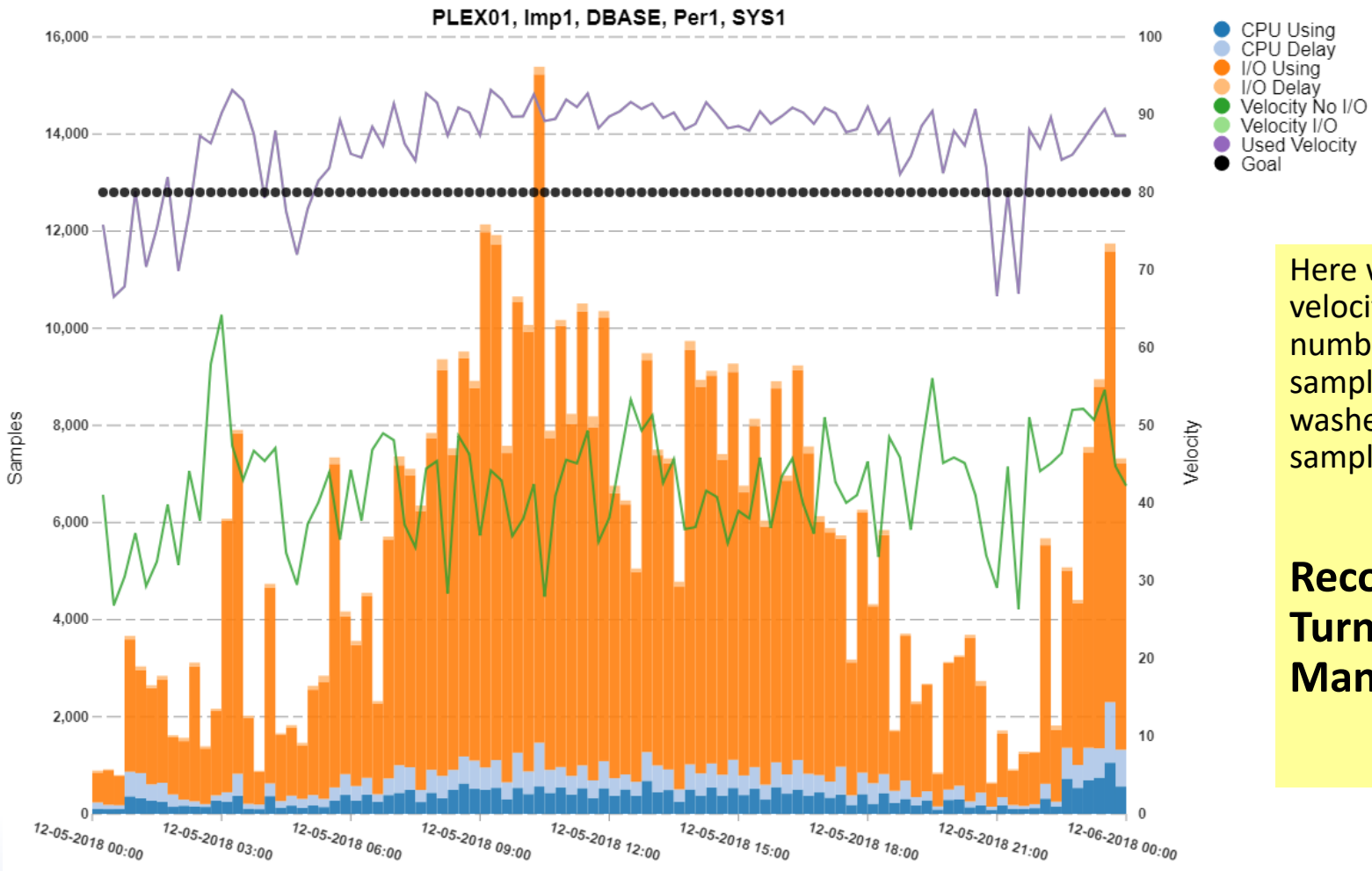
WLM Velocity Goal - Achieved Velocity Across Sysplex



This SCP is easily over-achieving its goal.

Even if the goal was moved to 90 or 95, WLM doesn't really have room to manage this service class because it looks to make changes that will be ***significant***.

CPU & I/O Contribution to Velocity



Here we see why the velocity is so high—the number of I/O using samples completely washes out the CPU samples.

Recommendation: Turn off I/O Priority Management

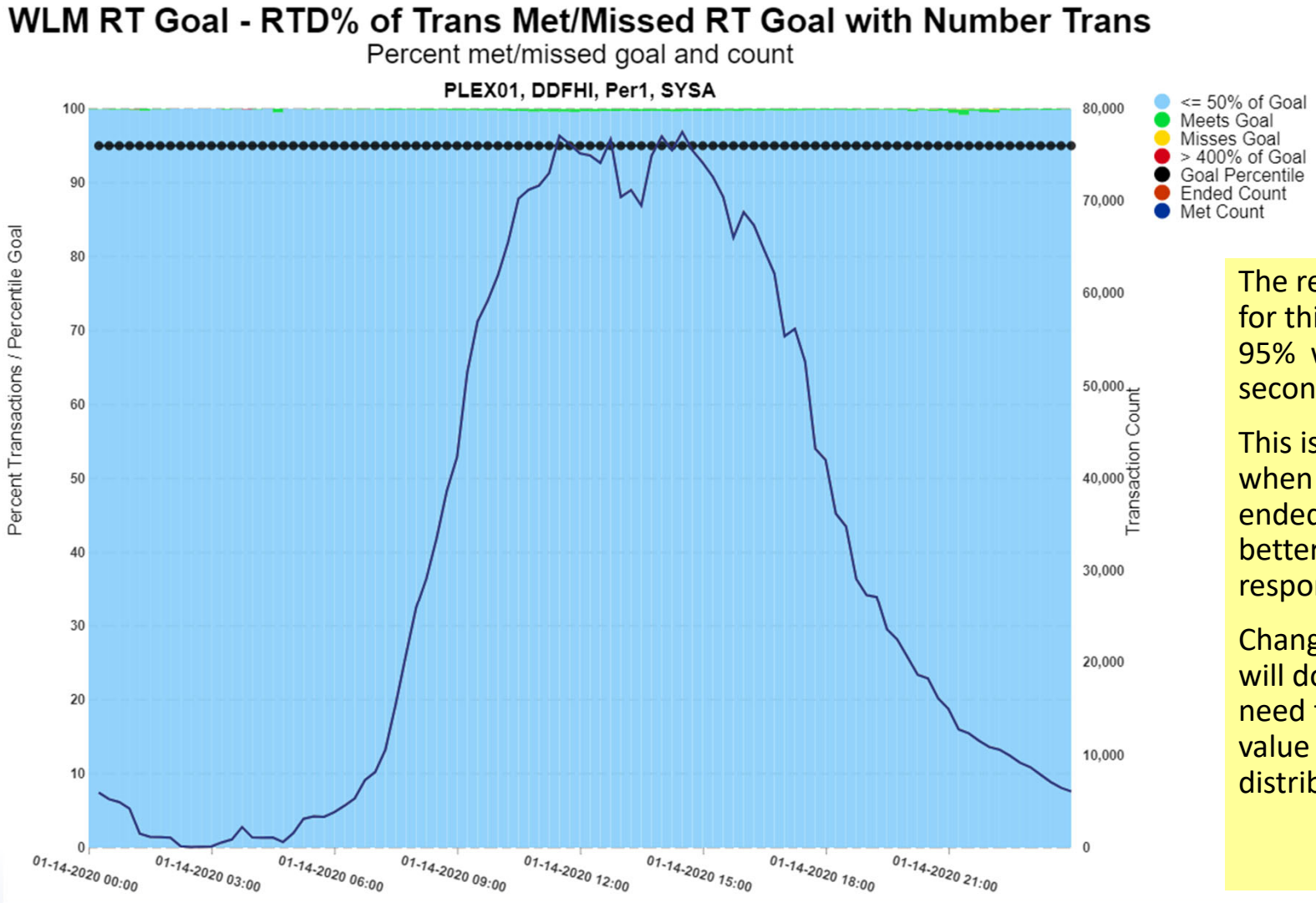
Less Than 0.015 Response Time Goals Allowed

Response time goals <0.015 seconds now allowed

- Problem: For many years the minimum response time goal objective allowed was 0.015 seconds (i.e. 15 milliseconds)
  - Restriction was due to rounding and way WLM maintains response time distributions
  - On today's high-speed processors and smaller transactions, 0.015 second goal may be too 'easy' of a goal to meet for some workloads
  - Easy goals at high importance levels can be bad
  - Especially in high processor capacity constrained environments
  - New minimum is 0.001 second response time

- As a reminder : WLM allows for two types of response time goals
  - Average response time goals
    - Example: Achieve a 0.05 second response time on average
    - Recommended not to use
  - Percentile response time goal
    - Example: Ensure that 90% of the transactions complete within 0.05 seconds
    - Allows for management of the transactions to be based on typical transactions rather than outlier transactions

WLM RT Goal - RTD% of Trans Met/Missed RT Goal with Number Trans

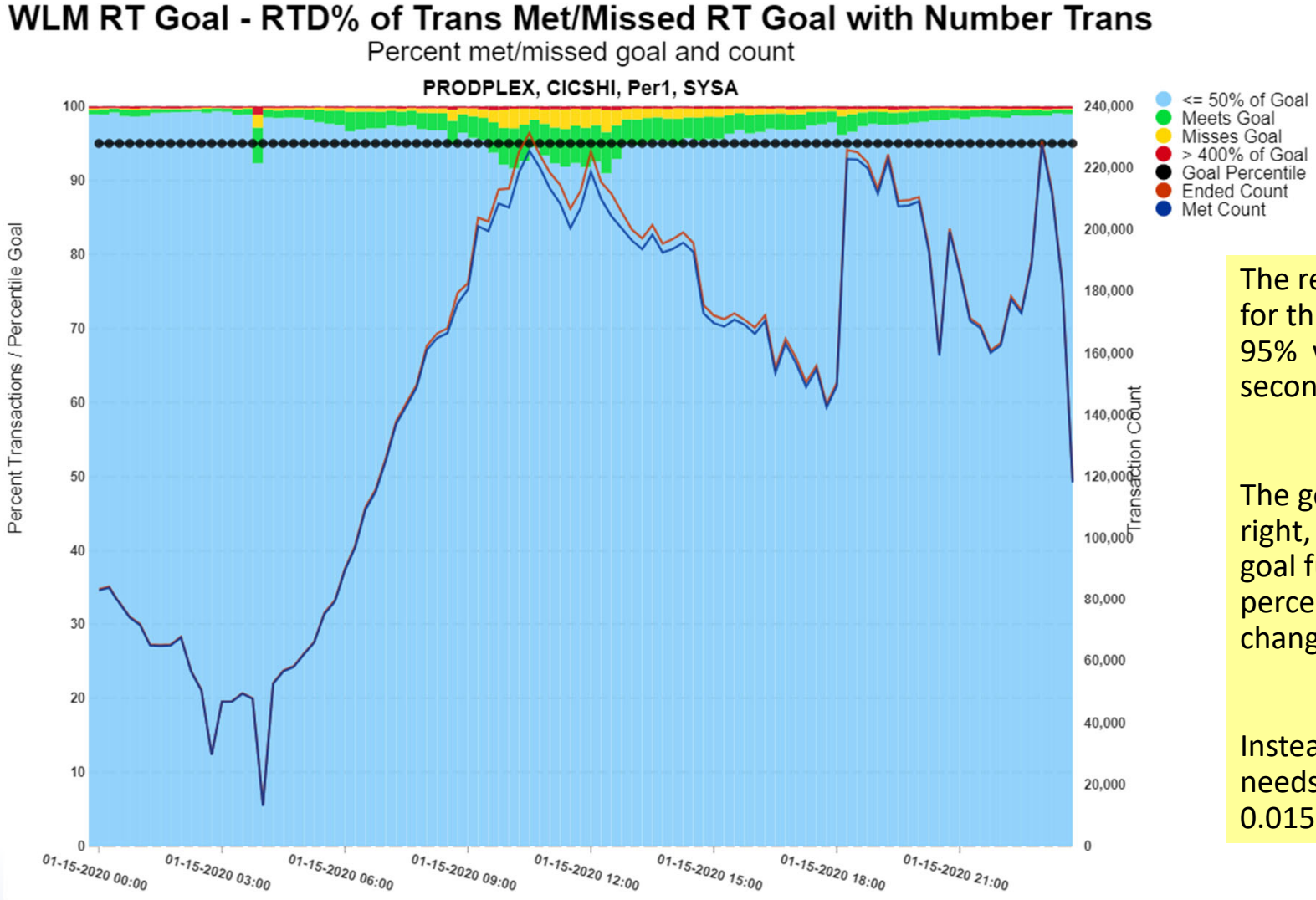


The response time goal for this service class is 95% within 0.015 seconds.

This is an example of when nearly 100% of the ended transactions did better than 50% of the response time objective.

Changing the percentile will do nothing. You need to change the goal value to influence the distribution.

WLM RT Goal - RTD% of Trans Met/Missed RT Goal with Number Trans

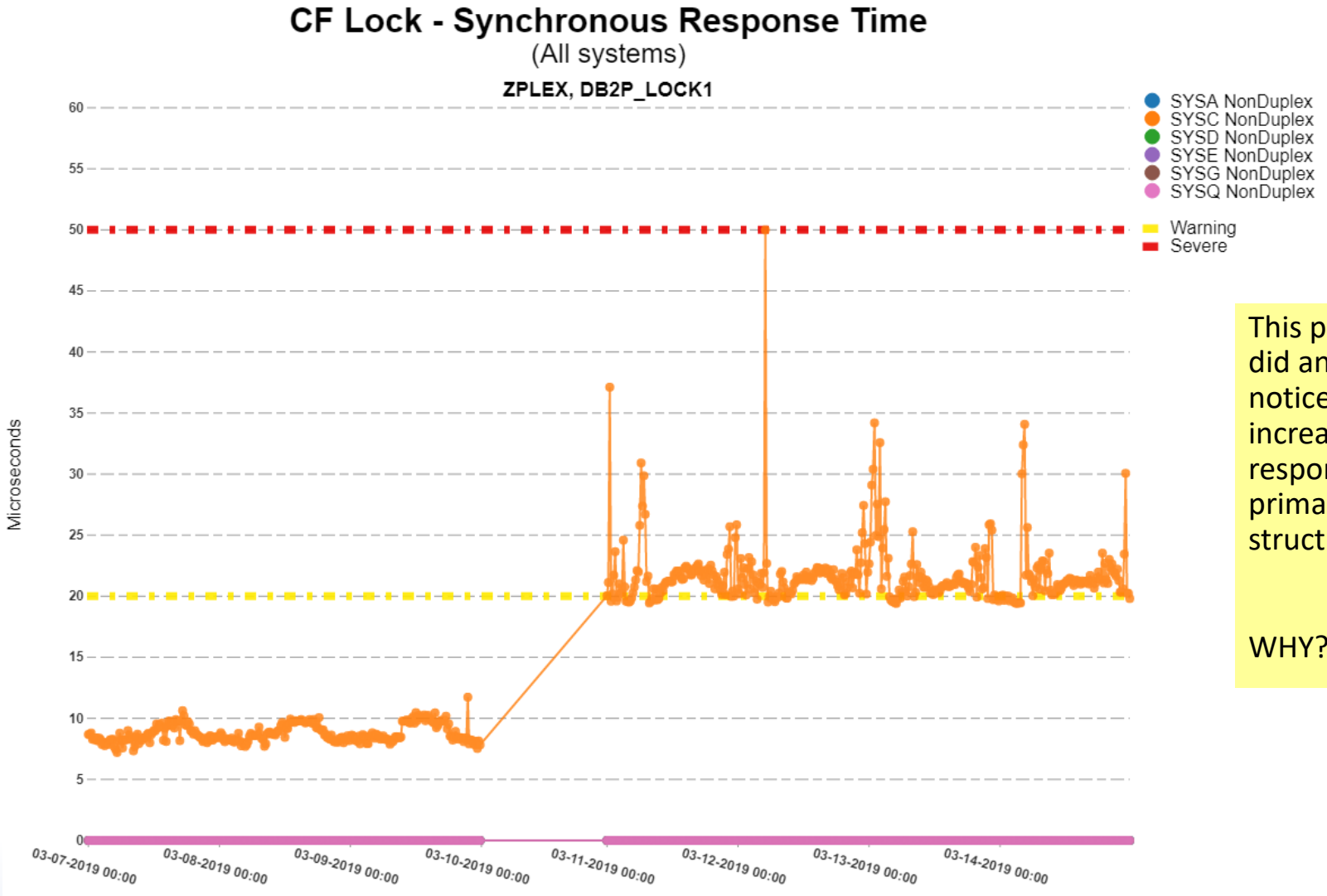


The response time goal for this service class is 95% within 0.009 seconds.

The goal is just about right, but to tune the goal further, the percentile should not change.

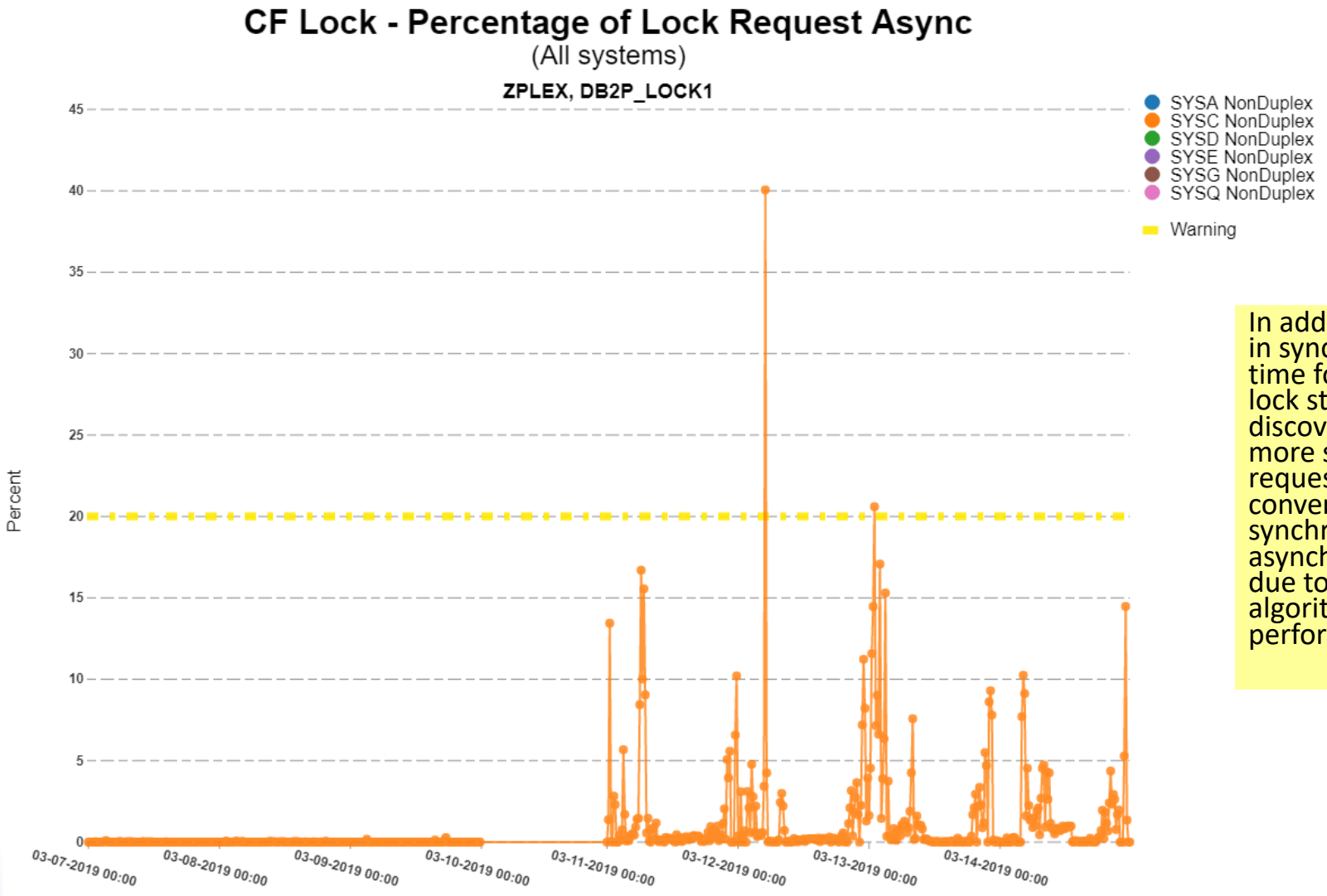
Instead the goal value needs to be less than 0.015 seconds.

Only use long-range CF Links if you need to

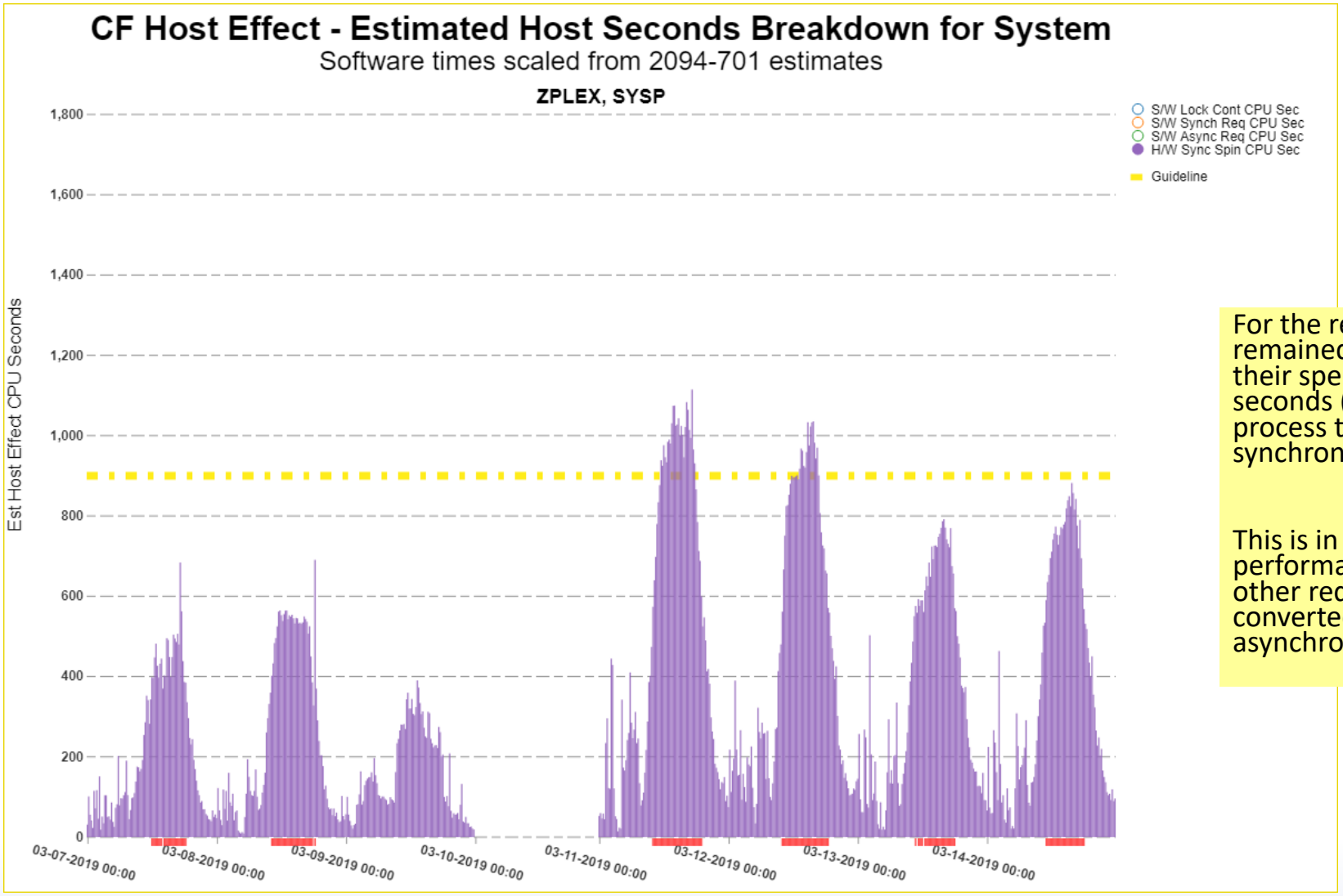


This particular customer did an upgrade and noticed a marked increase in Synchronous response times for their primary DB2 lock structure.

WHY?!?!?!



In addition to an increase in synchronous response time for their primary DB2 lock structure, it was discovered that more and more synchronous requests were being converted from synchronous to asynchronous requests due to XES heuristic algorithm. Another sign performance problems.



For the requests that remained synchronous, their spent more CPU seconds (i.e. MSUs) to process these remaining synchronous requests.

This is in addition to the performance hit to those other requests that were converted to asynchronous.

- CF Link technology regularly changes
- Generally speaking, there are faster short-distance technologies vs. slower long-distance technologies
  - Shorter/faster: 10s of meters and Gigabytes/second
  - Longer/slower: 10s of kilometers and Gigabits/second

High End servers	Signal Rate	Relative Speed	Relative Speed 2	z196	zEC12	z13	z14
IC	Memory	5x	5-45x	Yes	Yes	Yes	Yes
ISC-3	2Gb/s	1x	1x	Yes	Yes*	No	No
12X IFB	6 GB/s	3.8x	12-25x	Yes	Yes	Yes	Yes*
1X IFB	5 Gb/s	1.6x	2x	Yes	Yes	Yes	Yes*
ICA SR	8 GB/s	5x	17.5-30x			Yes	Yes
CE LR	10 Gb/s	3.2x?	3.5x			Yes – GA2	Yes

Mid-Range servers	Signal Rate	Relative Speed	Relative Speed 2	z114	zBC12	Z13s	z14 zR1
IC	Memory	5x	5-45x	Yes	Yes	Yes	Yes
ISC-3	2Gb/s	1x	1x	Yes	Yes – Limited	No	
12X IFB	6 GB/s	3.8x	12-25x	Yes	Yes	Yes*	No
1X IFB	5 Gb/s	1.6x	2x	Yes	Yes	Yes*	No
ICA SR	8 GB/s	5x	17.5-30x			Yes	Yes
CE LR	10 Gb/s	3.2x?	3.5x			Yes – GA2	Yes

(* last generation to support)
Some source - Coupling Faciltiy Configuration Options, David Raften, 2018
Note different relative effective transfer speeds based on 2 different sources

Wrap-up

- We hope you enjoyed this and that you've learned something
- Let us know if you like this potpourri of topics format
- We'll be around now and all week for questions

- Questions?

- Please visit our website: www.epstrategies.com
  - Past presentations
  - WLM to HTML tool
  - More information about Pivotor
  - Future educational webinars

Your feedback is important!

Submit a session evaluation for each session you attend:

SHARE mobile app -or- www.share.org/evaluation