

Capping, Capping, and Capping: A Comparison of Hard and Soft-capping Controls

Horst Sinram

IBM Germany Research & Development


04 Aug 2014

Session 15719



#SHAREorg



Copyright (c) 2014 by SHARE Inc.  Except where otherwise noted, this work is licensed under <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Agenda

- **Overview of capping types**
- Initial capping
- Absolute capping
- Defined capacity & group capacity
- Resource group capping
- 4HRA management
- Additional Material

Reasons you would consider capping techniques...

- For technical reasons:
 - Protect LPARs against other LPARs
 - Influence capacity-based workload routing
 - Guarantee unused CPC processor capacity
 - Protect workloads (sets of service classes) against other workloads
- For non-technical reasons:
 - Limit software cost
 - Capacity limit for one or more LPARs
 - Four hour rolling average (4HRA) consumption
 - Control gradient of 4HRA
- Impact of capping needs to be monitored and accepted
- Cap limits should be adjusted as appropriate
 - Watch your SLAs

Comparison of LPAR capping types

Type of capping	Scope	Specification in terms of	Proc types	Stable SU/MSU limit under configuration changes	Suitable to technically separate LPARs or groups of LPARs	Defined
Initial (hard capping)	LPAR	LPAR share of CPC capacity	Any	-	+	SE/HMC
Absolute capping	LPAR	Fractional #processors		○	+	
Defined capacity (DC, soft capping)	LPAR	MSU	CP	+	-	
LPAR group capacity (GC, soft capping)	Group of LPARs	MSU		+	-	
Resource group capping	Groups of service classes in Sysplex or per LPAR	Unweighted CPU SU/sec, fraction of LPAR share, or fractional #processors	CP*	+	N/A	WLM Policy
Logical configuration	LPAR	Integer #processors	Any	○	+ but coarse grain	HMC+ OS



PR/SM controlled



WLM controlled, PR/SM enforced



WLM controlled

Which capping techniques may be combined?

Type of capping	Initial (hard capping)	Absolute capping	Defined capacity (soft capping)	LPAR group capacity (soft capping)	Resource group capping
Initial (hard capping)		+	-	-	+
Absolute capping			+	+	+
Defined capacity (soft capping)				+	+
LPAR group capacity (soft capping)					+
Resource group capping					

Agenda

- Overview of capping types
- **Initial capping**
- Absolute capping
- Defined capacity & group capacity
- Resource group capping
- 4HRA management

- Additional Material

Initial capping (aka “hard capping”)

- Defined to PR/SM per processor type. Managed by PR/SM through limiting the processor time available to the LP's logical processors
- The LPAR capacity is capped to LPAR share of CPC shared capacity

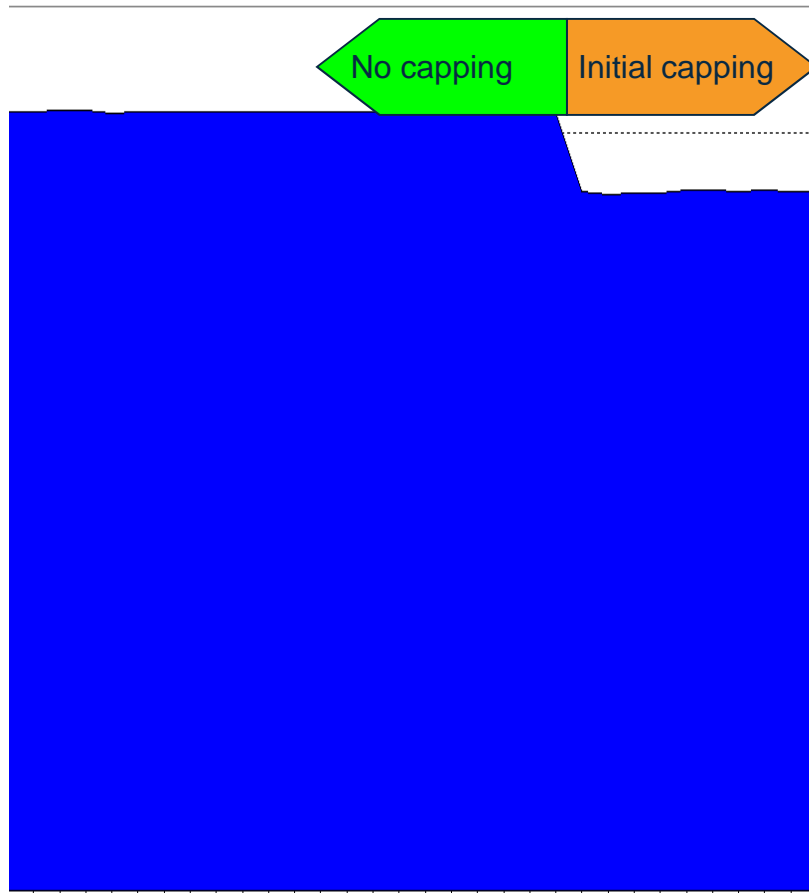
$$\text{LPAR}_i \text{ share} = \frac{\text{Weight}_i}{\sum \text{Weight}_j}$$

All activated LPARs

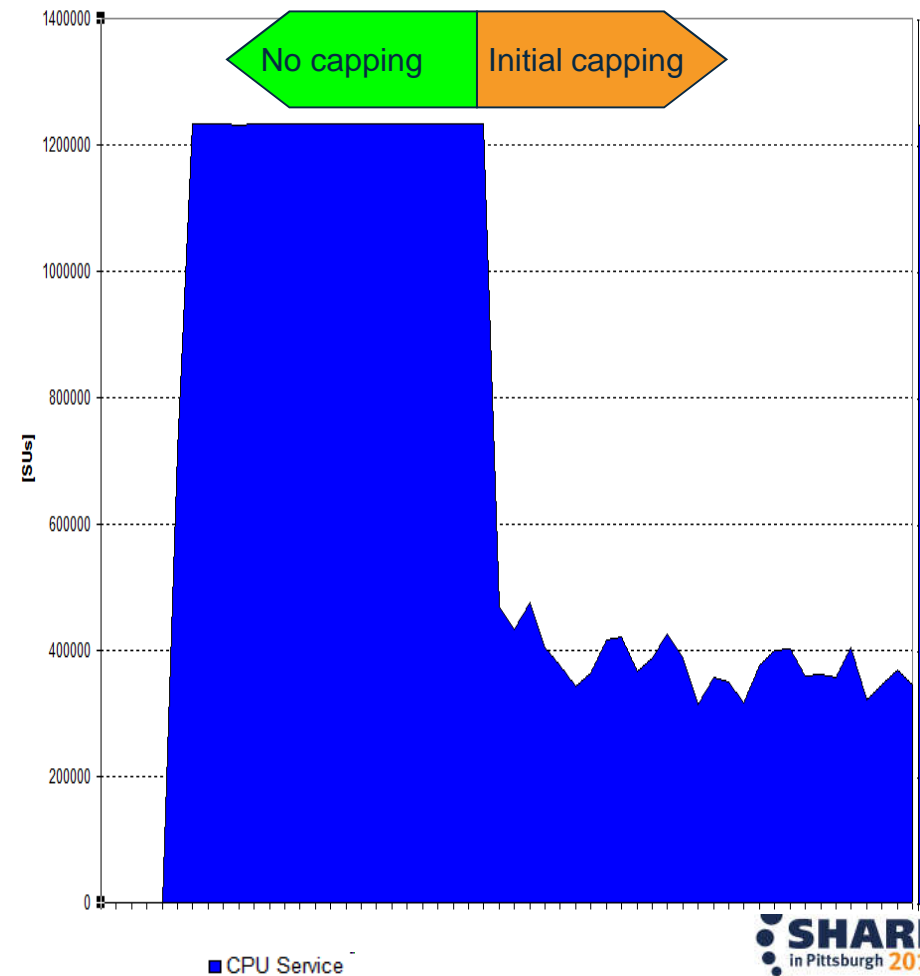
- LPAR weight is distributed across online CPs of the given type
- With HiperDispatch=NO an LP's share is divided by the number of online logical CPs
 - Capping is done on a logical CP basis.
May result in over capping if not all LCPs can be utilized
 - Consider following example:
zEC12-732, 10 CPs online, Share=5.6%, low CPC utilization
Workload: 2 TCBs

Initial Capping with HiperDispatch=Yes vs. No

Service Consumption
Initial capping with 2 work units, HD=YES, Service Class: CPUHIGH Period: 1

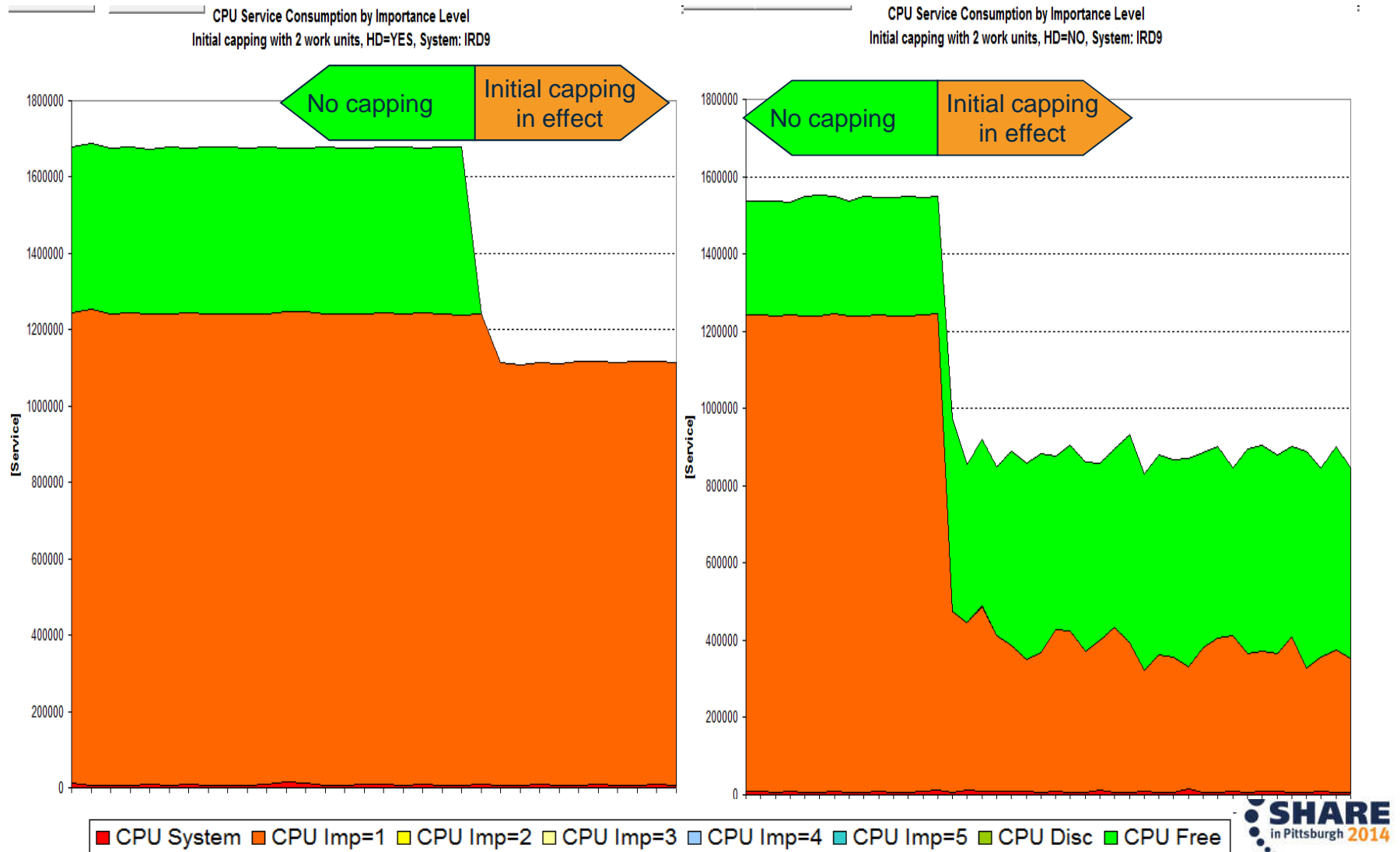


Service Consumption
Initial capping with 2 work units, HD=NO, Service Class: CPUHIGH Period: 1



■ CPU Service

Initial Capping with HiperDispatch=Yes vs. No



Initial Capping with HiperDispatch=Yes vs. No CPU Activity Reports

CPU	2827	CPC CAPACITY	3665	SEQUENCE CODE	0000000000
MODEL	732	CHANGE REASON=NONE	HIPERDISPATCH=YES		
H/W MODEL	H43				
---CPU---		----- TIME % -----			LOG PROC
NUM	TYPE	ONLINE	LPAR BUSY	MVS BUSY	PARKED
0	CP	100.00	89.12	97.67	0.00
1	CP	100.00	87.50	97.83	0.00
2	CP	100.00	2.51	82.33	96.54
3	CP	100.00	1.87	63.68	96.54
4	CP	100.00	0.01	-----	100.00
5	CP	100.00	0.01	-----	100.00
6	CP	100.00	0.01	-----	100.00
7	CP	100.00	0.01	-----	100.00
A	CP	100.00	0.01	-----	100.00
B	CP	100.00	0.01	-----	100.00
TOTAL/AVERAGE			18.10	96.92	180.4

LOG PROC	
SHARE %	
100.0	HIGH
80.4	MED

With HiperDispatch=Yes
the high/medium
processors receive a
higher processor share.

MODEL	732	CHANGE REASON=NONE	HIPERDISPATCH=NO				
H/W MODEL	H43						
---CPU---	----- TIME % -----					LOG PROC	
NUM	TYPE	ONLINE	LPAR BUSY	MVS BUSY	PARKED	SHARE %	
0	CP	100.00	14.61	54.28	-----	18.0	
1	CP	100.00	13.00	46.80	-----	18.0	
2	CP	100.00	10.71	31.82	-----	18.0	
3	CP	100.00	6.77	18.55	-----	18.0	
4	CP	100.00	4.22	6.44	-----	18.0	
5	CP	100.00	4.87	13.16	-----	18.0	
6	CP	100.00	1.75	2.72	-----	18.0	
7	CP	100.00	4.54	13.05	-----	18.0	
A	CP	100.00	4.02	10.40	-----	18.0	
B	CP	100.00	3.08	6.88	-----	18.0	
TOTAL/AVERAGE			6.76	20.41		180.0	

LOG PROC	
SHARE %	
18.0	
18.0	
18.0	
18.0	
18.0	
18.0	
18.0	
18.0	
18.0	
18.0	

Stability of initial cap limits

- The effective limit for an initial cap changes when...
 - The initial weight of the capped LPAR is changed
 - LPARs are de/activated or the total weight changes due to initial weight changes
 - Temporary capacity is de/activated
 - CBU, On/Off CoD...

Agenda

- Overview of capping types
- Initial capping
- **Absolute capping**
- Defined capacity & group capacity
- Resource group capping
- 4HRA management

- Additional Material

Absolute Capping Limit

- Defined to PR/SM per processor type. Managed by PR/SM through limiting the number of PR/SM time slices available to the LP's logical processors
- Specification in terms of (fractional) number of processors per processor type
 - E.g., 3.75 CPs
- Introduced with zEC12 GA2
- Primarily intended for non z/OS images
- Can be specified independently from the LPAR weight
 - But recommended to specify absolute cap above weight
 - WLM algorithms consider weight

Absolute Capping Limit

- Unlike initial capping absolute capping may be used *concurrently* with defined capacity and/or group capacity management
 - The respective minimum becomes effective.
 - WLM/SRM is aware of the absolute cap, e.g. for routing decisions.
 - $RCTIMGWU = \text{MIN}(\text{absolute cap, defined capacity, group cap})$ when all capping types are in effect
 - RMF provides RCTIMGWU in SMF70WLA
 - In addition, SMF70HW_Cap_Limit value in hundredths of CPUs

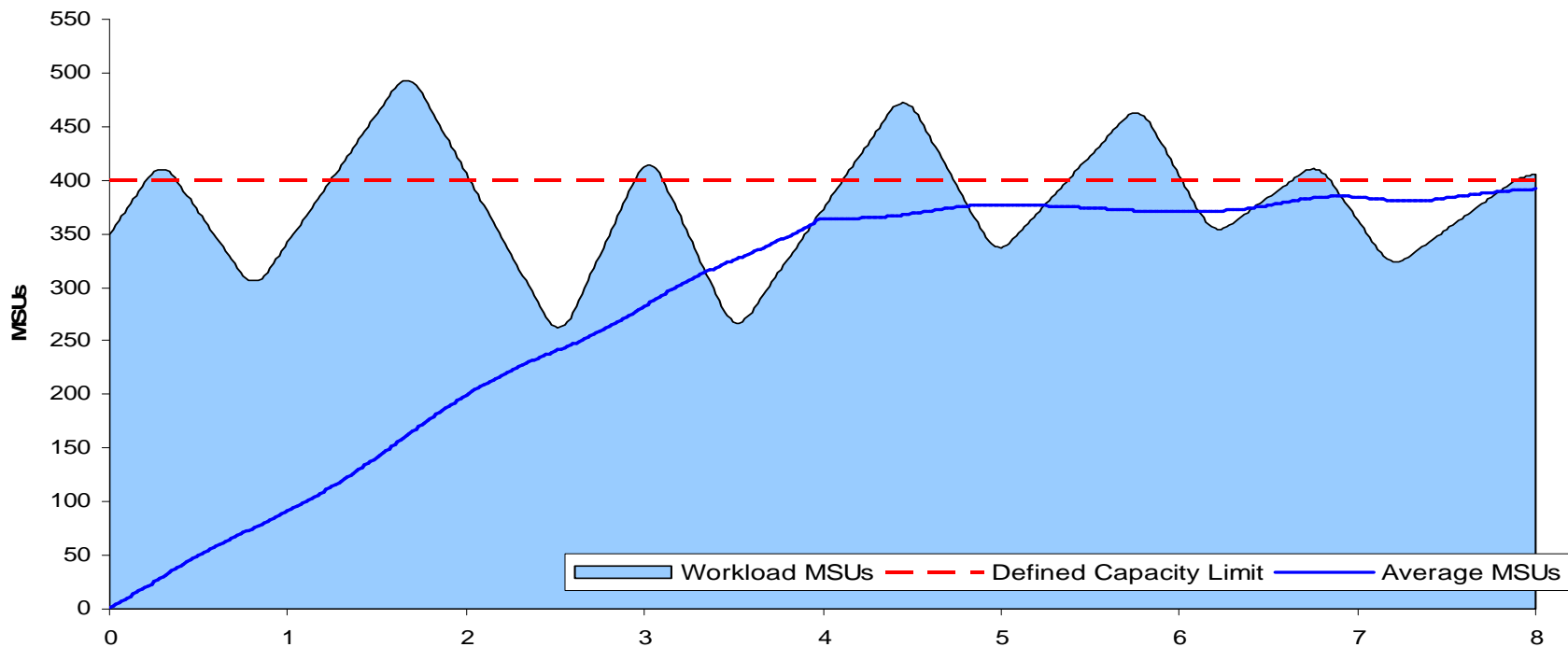
Stability of absolute cap limits

- The effective limit for an absolute cap changes significantly when
 - The absolute cap value of the capped LPAR is changed
 - Temporary capacity is de/activated AND the capacity level (processor speed) changes
 - I.e., general purpose processor CBU, On/Off CoD to/from subcapacity models
- The effective MSU rating for an absolute cap changes when the physical configuration changes

Agenda

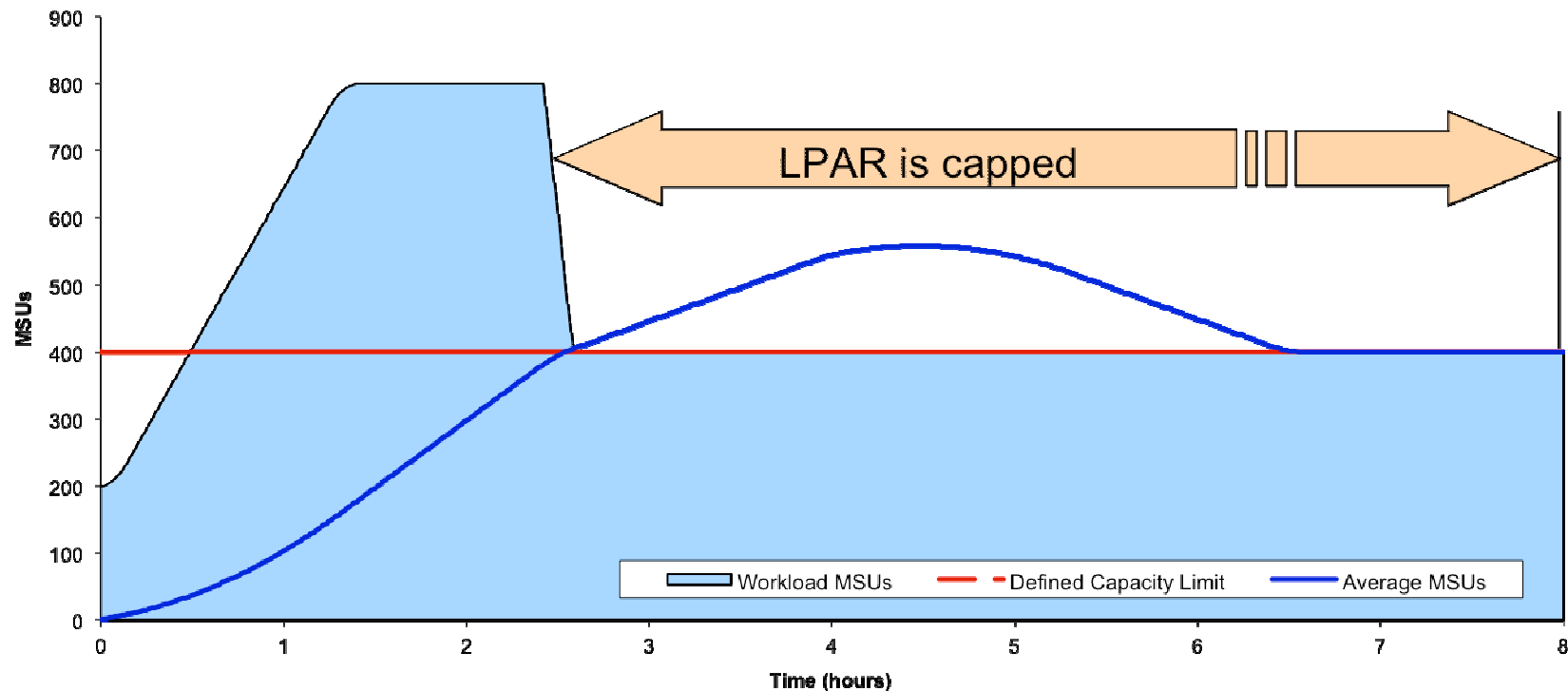
- Overview of capping types
- Initial capping
- Absolute capping
- **Defined capacity & group capacity**
- Resource group capping
- 4HRA management
- Additional Material

4 Hour Rolling Average



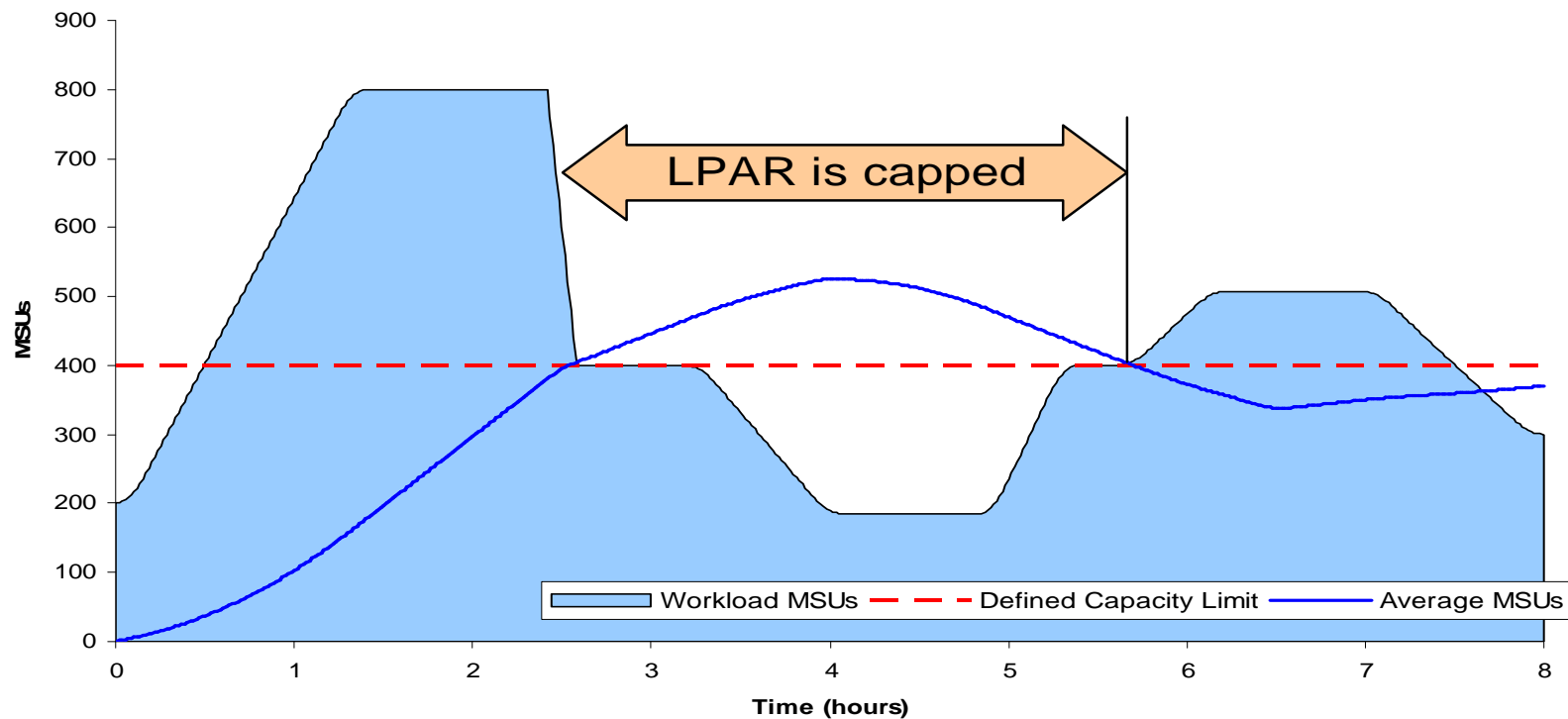
- Average consumption in LPAR in last 4h (rolling)
- MSU \equiv “Million Service Units per hour”
 \neq Service Units $\cdot 3600 / 1000000$
- Tracked as array of 48 intervals of 5 min = 4h

LPAR Capping



- An LP is –soft- capped when the 4HRA exceeds the defined capacity limit
- It remains capped until the 4HRA is below the defined limit
- When capped, the consumption is limited to the defined limit
- WLM advises PR/SM how to cap the LP

End of capping phase



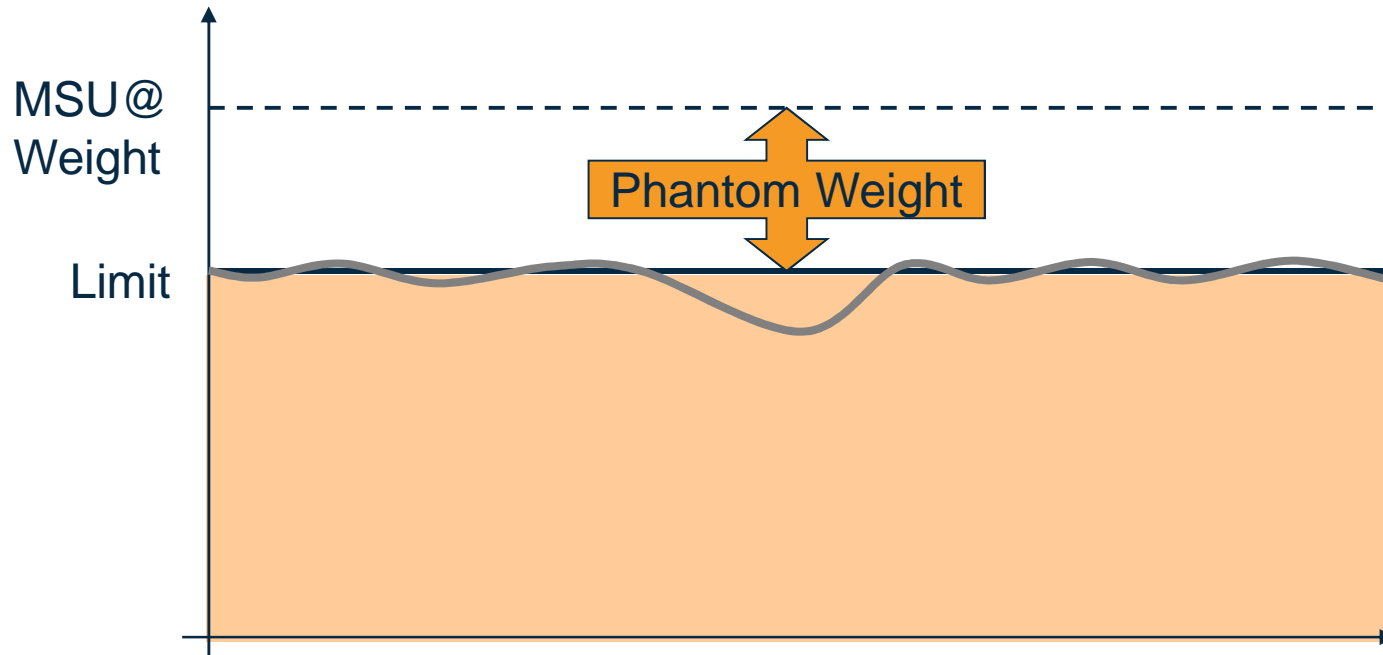
- Capping ends when the 4 hour average is below the softcap

Underlying soft capping techniques

- Historically, PR/SM algorithms were designed to cap a partition at its weight.
- Therefore, WLM and PR/SM use particular interfaces to cap a partition to an arbitrary MSU figure

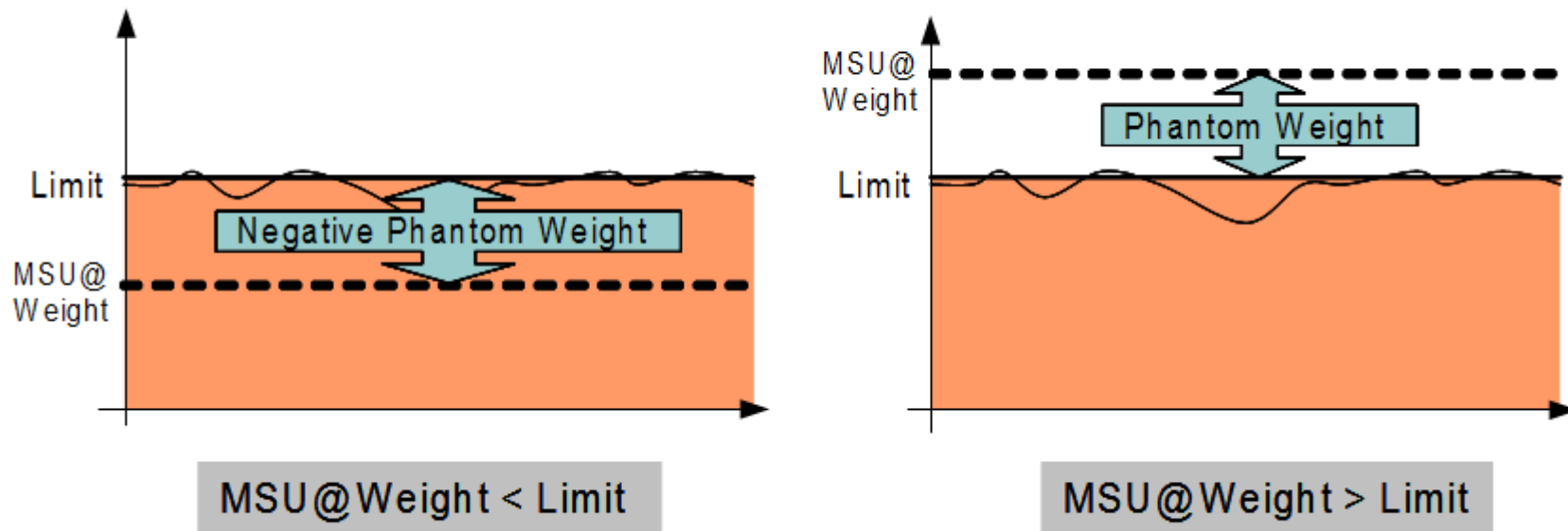
Weight vs. defined capacity limit	Hardware/Software level	Selected capping technique
MSU@weight > limit	Any	Phantom weight
MSU@weight ≤ limit	zEC12 GA2 and z/OS V2.1 or later	Negative phantom weight
	Other	Pattern capping

Phantom weight



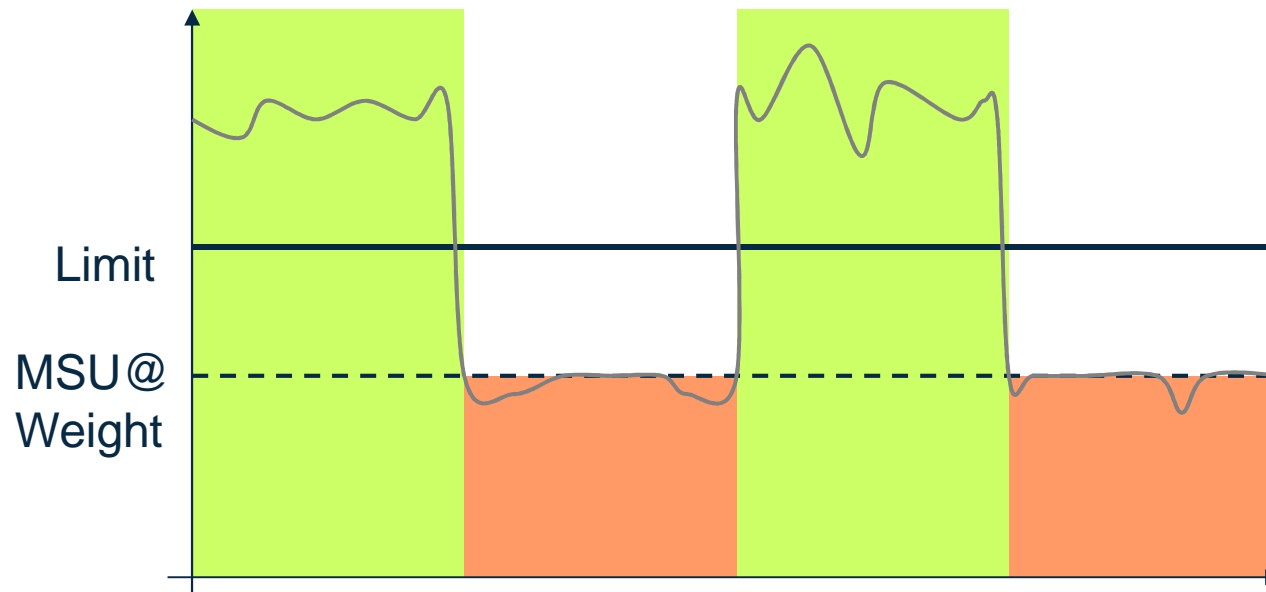
- Phantom weight is used to modify the PR/SM share of an LPAR
- WLM does not change a phantom weight as long as the limit and configuration do not change
➔ smooth capping

Capping with phantom weight



- zEC12 with z/OS V2.1 and above support not only positive but also negative phantom weights.
 - Note: While a positive phantom weight changes the PR/SM priority of a partition, a negative phantom does not elevate the PR/SM dispatching priority.

Cap pattern

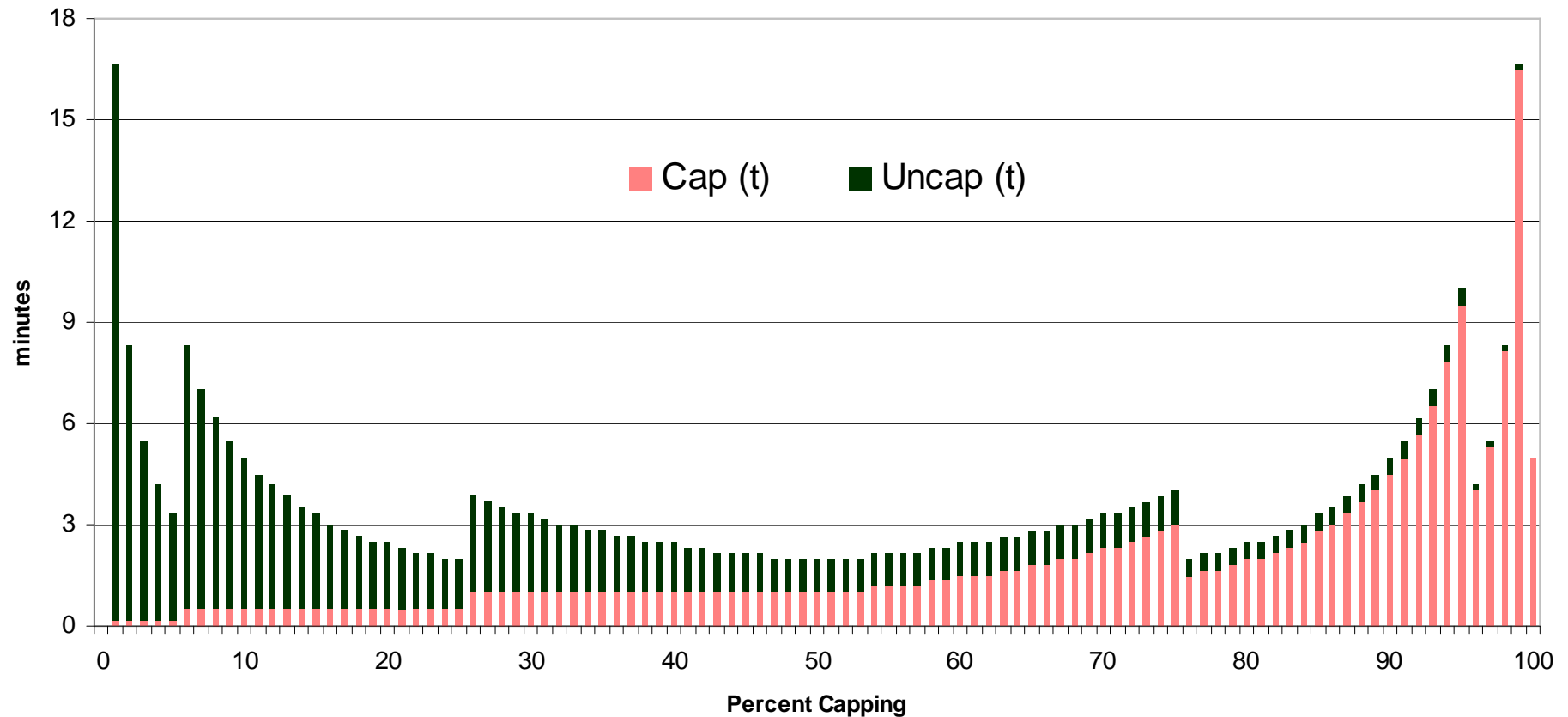


Prior to negative phantom weights WLM set up a cap pattern:
Alternating periods of

- LP capped to MSU@Weight, and
- LP uncapped

On average the MSU limit is enforced.

Cap pattern length

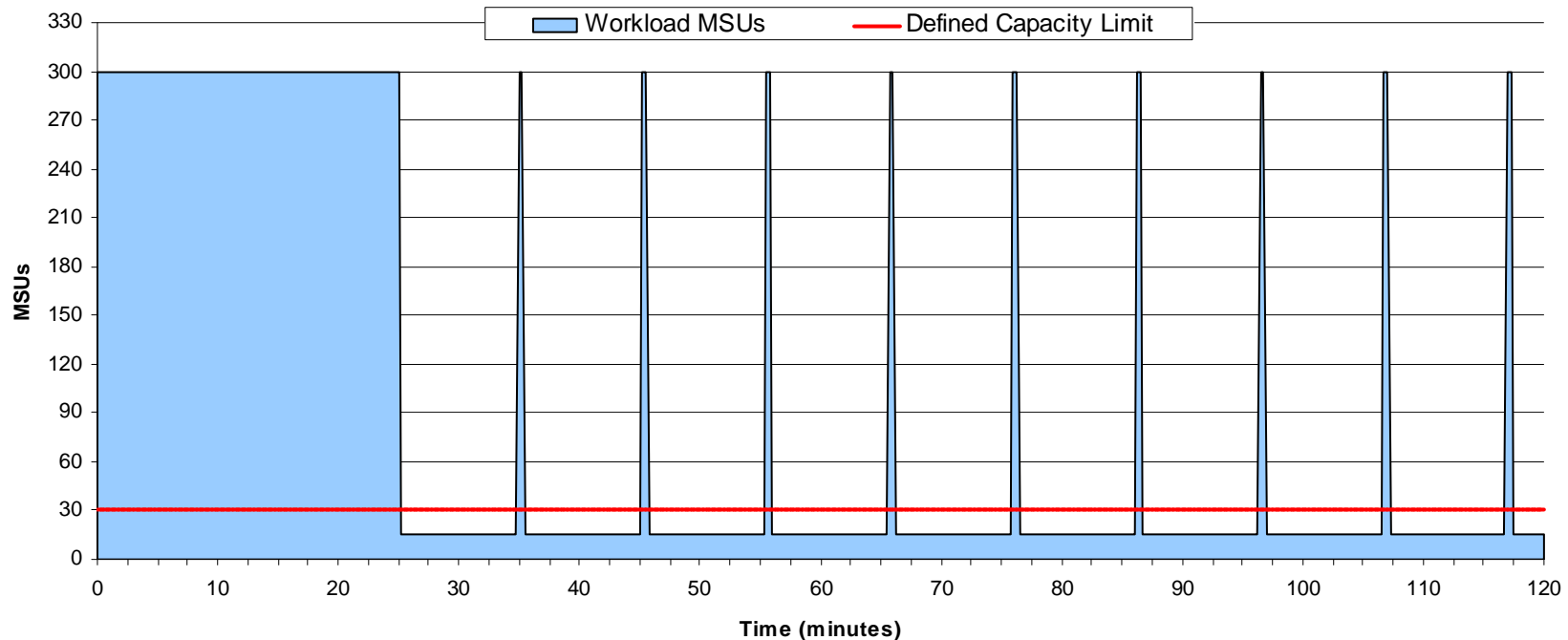


The LPAR cap pattern changes usually at an order of a few minutes.

The extreme cases are

- 01% WLM capping = 10 sec capped / 16.5min uncapped
- 99% WLM capping = 10 sec uncapped/ 16.5min capped

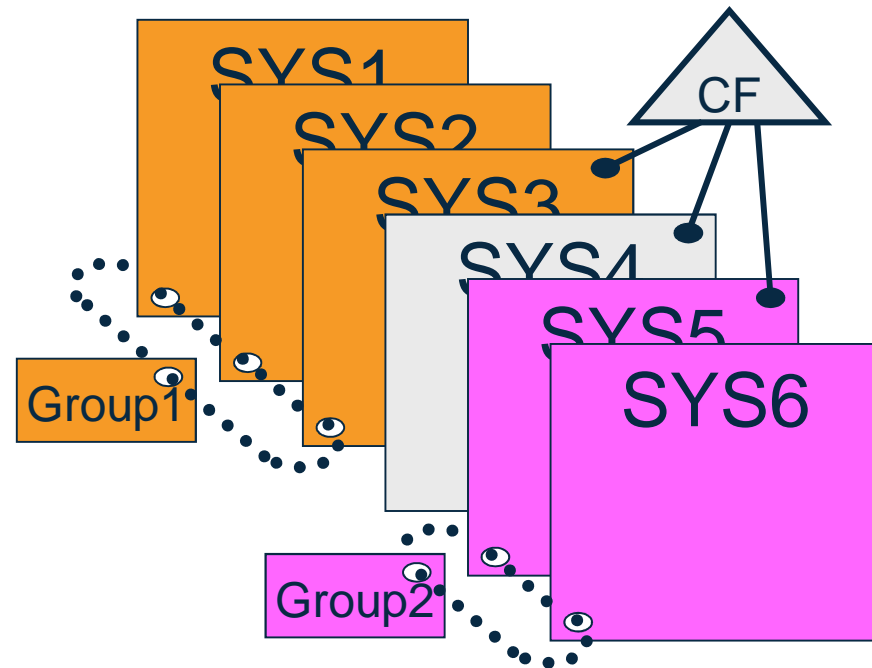
Extreme capping pattern



- May lead to complains about extreme patterns
- Occurs when weight is low but limit is high
- Cap pattern caps to $MSU@Weight$

Group Capping

An LPAR capacity group can be used to enforce a MSU limit for a set of one or more LPARs.



- A capacity group is limited to a single CPC but independent from the Sysplex
- A system can be joined to one group at most
- A system will not join or will leave the capacity group when requirements not met
 - Namely, initial capping must not be active

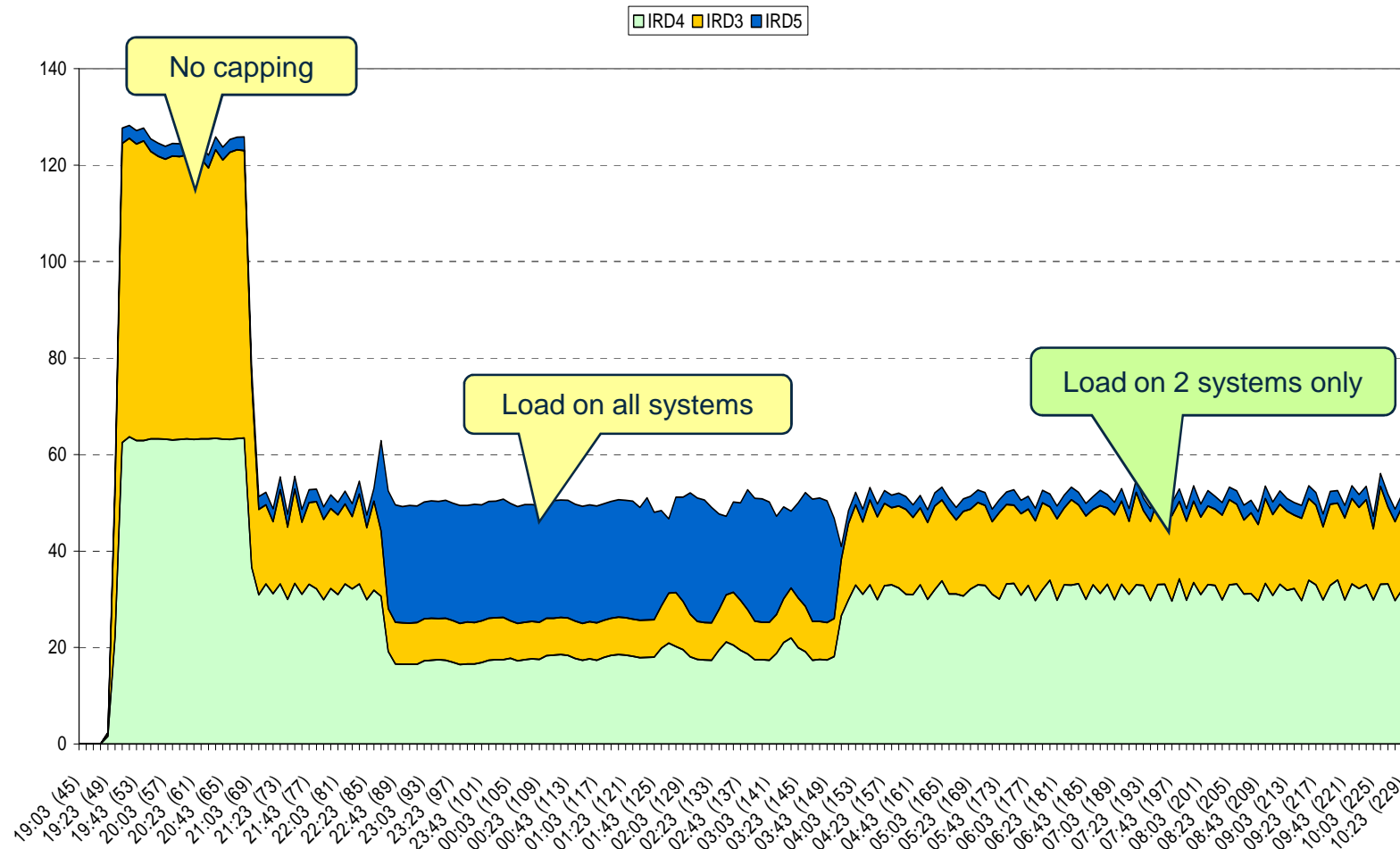
Group capping example

System	Weight	DC (MSU)	GC (MSU)	Initial GC Share (MSU)	Donation at full demand (MSU)	GC Entitle ment (MSU)
SYS1	600	-	400	200	-	240
SYS2	300	-		100	-	120
SYS3	300	40		100	60	40

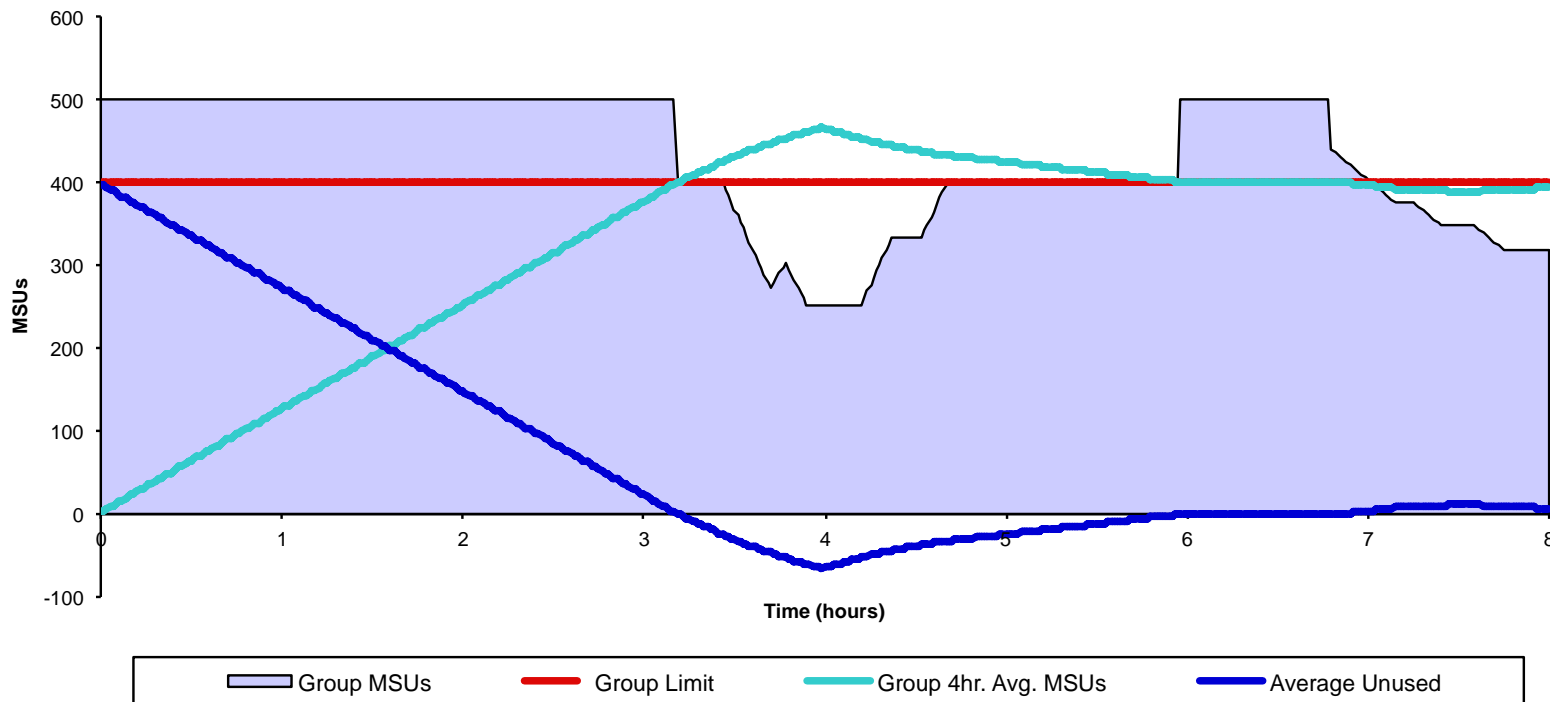
- The share of a group member is based on its *weight*
 - IRD with zEC12 GA2 & z/OS V2.1: initial weight
 - IRD in prior environments: current weight
- Unused capacity is donated to other group members
- The minimum of DC and GC entitlement is used for capping an LPAR

Group Capping behavior

Actual MSU values



Unused vector (group capping)



- Group capacity is tracked via an **unused** group capacity array of 48 intervals of 5 min
- Group capping is active when average unused group capacity negative
- Each system tracks unused capacity while joined to a capacity group
 - Not synchronized upon group changes: systems may have a different view for up to 4h

RMF: Partition Data Report

P A R T I T I O N D A T A R E P O R T																	
z/OS V1R12				SYSTEM ID SYS1				DATE 10/13/10				INTERVAL 14.59.678					
				RPT VERSION V1R12 RMF				TIME 09.30.00				CYCLE 1.000 SECONDS					
MVS PARTITION NAME				SYS1				NUMBER OF PHYSICAL PROCESSORS				9		GROUP NAME		N/A	
IMAGE CAPACITY				100				CP				7		LIMIT		N/A	
NUMBER OF CONFIGURED PARTITIONS				9				ICF				2		AVAILABLE		N/A	
WAIT COMPLETION				NO													
DISPATCH INTERVAL				DYNAMIC													

----- PARTITION DATA -----							-- LOGICAL PARTITION PROCESSOR DATA --				-- AVERAGE PROCESSOR UTILIZATION PERCENTAGES --					
-----MSU----- -CAPPING--							PROCESSOR- ----DISPATCH TIME DATA----				LOGICAL PROCESSORS		--- PHYSICAL PROCESSORS ---			
NAME	S	WGT	DEF	ACT	DEF	WLM%	NUM	TYPE	EFFECTIVE	TOTAL	EFFECTIVE	TOTAL	LPAR	MGMT	EFFECTIVE	TOTAL
SYS1	A	20											0.01		4.24	4.25
SYS2	A	1											0.01		0.34	0.35
SYS3	A	10											0.02		3.41	3.43
SYS4	A	300											0.01		68.68	68.69
SYS5	A	200											0.01		23.02	23.03
													0.05			0.05
TOTAL													0.11		99.69	99.80
CFC1	A	DEF											0.01		99.95	99.96
CFC2	A	DEF											0.00		0.00	0.00
PHYSICAL													0.03			0.03
TOTAL													0.04		99.95	99.99

----- PARTITION DATA -----						
-----MSU----- -CAPPING--						
NAME	S	WGT	DEF	ACT	DEF	WLM%
			1	2	3	4
SYS1	A	20	100	10	NO	62.2
SYS2	A	1	0	1	YES	0.0
SYS3	A	10	5	8	NO	3.3
SYS4	A	300	95	155	NO	0.0
SYS5	A	200	50	52	NO	0.0

RMF: Partition Data Report

1. **MSU DEF** DC limit for this partition in MSU as specified on HMC
2. **MSU ACT** Actual avg. MSU consumption of this LPAR
3. **CAPPING DEF** Indicates whether this partition uses initial capping
4. **CAPPING WLM%** Portion of time the LPAR was capped during the RMF interval
 - Does not necessarily imply that the cap constrained the LPAR's consumption.

RMF: CPC Capacity

RMF V1R12 CPC Capacity									
Samples: 100		System: SYS1		Date: 10/13/10		Time: 09.32.00		Range: 100 Sec	
Partition: SYS1		2094 Model 714							
CPC Capacity: 843		Weight % of Max: 68.4		4h Avg: 66		Group: N/A			
Image Capacity: 66		WLM Capping %: 5.1		4h Max: 84		Limit: N/A			
Par	1	CPC Capacity: 843		3	weight % of Max: 68.4		5	4h Avg: 66	
	2	Image Capacity: 66		4	WLM Capping %: 5.1		6	4h Max: 84	
*CP									
SYS1	66	26	NO	3.0	9.8	10.4	0.1	2.9	3.1
SYS2	77	4	NO	3.0	2.1	2.4	0.0	0.5	0.5
SYS3	0	9	NO	4.0	3.4	3.5	0.0	1.0	1.0
SYS4	0	11	NO	4.0	4.3	4.5	0.0	1.2	1.3
PHYSICAL							0.1	0.1	
*AAP									
SYS1			NO	2.0	0.2	0.4	0.2	0.2	0.4
SYS2			NO	2.0	0.2	0.4	0.2	0.2	0.4
PHYSICAL							2.7	2.7	

RMF: CPC Capacity

1. **CPC Capacity**
Total capacity of the CPC in MSU/h
2. **Image Capacity**
Maximum capacity available to this partition
3. **Weight % of Max**
Average weighting factor relative to the maximum defined weight for this partition.
4. **WLM Capping %**
Percentage of time that WLM had advised PR/SM to cap the LPAR
5. **4h Avg**
Average consumed MSU/h during the last 4 hours
6. **4h Max**
Maximum consumed MSUs during the last 4 hours

RMF: Group Capacity report

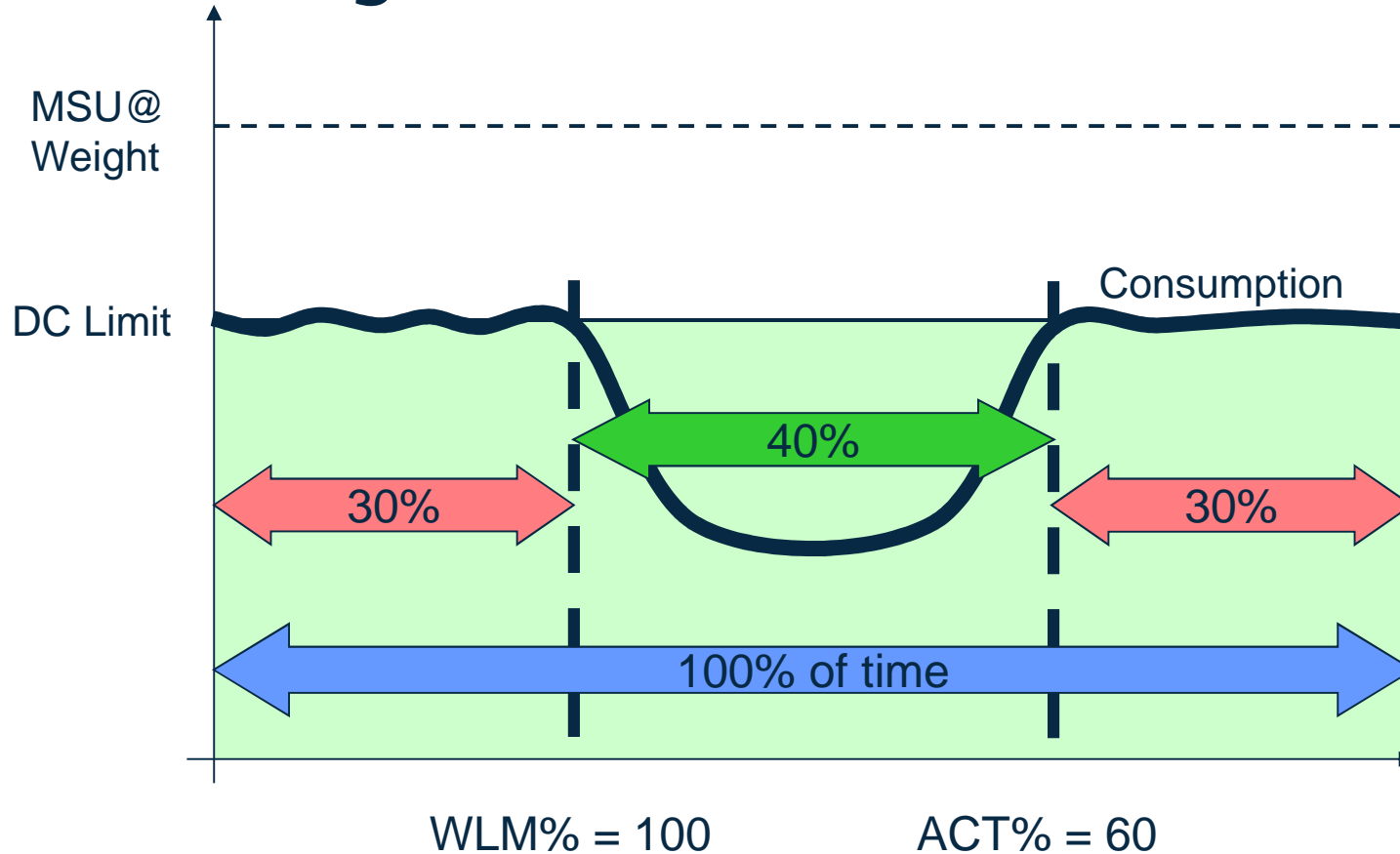


GROUP CAPACITY REPORT												
z/OS V1R12			SYSTEM ID SYS1			DATE 10/13/2010			INTERVAL 14.59.968			
			RPT VERSION V1R12 RMF			TIME 15.15.00			CYCLE 1.000 SECONDS			
----GROUP-CAPACITY----			PARTITION	SYSTEM	-- MSU --	WGT	----	CAPPING	----	- ENTITLEMENT -		
NAME	LIMIT	AVAIL			DEF	ACT		DEF	WLM%	ACT%	MINIMUM	MAXIMUM
1	2	3			4	5		6	7	8	9	10
GROUP1	1500	-22	SYS1	SYS1	80	3	600	NO	25	23	80	80
			SYS2	SYS2	80	3	500	NO	100	46	80	80
-----					-----					-----		
TOTAL					6		1100					

1. **NAME** Name of the WLM capacity group
2. **LIMIT** Group limit
3. **AVAIL** Average unused capacity in MSUs (avg. unused vector)
4. **MSU DEF** Defined capacity limit
5. **MSU ACT** Average used capacity
6. **CAPPING DEF** Tells if the initial capping is activated on HMC
7. **CAPPING WLM%** Percentage of time that WLM had set up a cap for the partition
8. **CAPPING ACT%** Percentage of time found capping actually limited the usage of processor resources for the partition
9. **MINIMUM ENT.** Minimum of the GC member share and the DC limit
10. **MAXIMUM ENT.** Minimum of the GC limit and the DC limit



Phantom weight: WLM% vs. ACT% in RMF



- RMF: WLM% is always 100 in case of phantom weight

RMF: Partition Data report

P A R T I T I O N D A T A R E P O R T									
z/OS V1R12		SYSTEM ID SYS1		DATE 10/13/2010		INTERVAL 15.00.999			
		RPT VERSION V1R12 RMF		TIME 13.30.00		CYCLE 1.000 SECONDS			
MVS PARTITION NAME	SYS1	NUMBER OF PHYSICAL PROCESSORS		16	GROUP NAME	1	GROUP1		
IMAGE CAPACITY	120	CP		8	LIMIT	2	200 *	4	
NUMBER OF CONFIGURED PARTITIONS	6	AAP		2	AVAILABLE	3	64		
WAIT COMPLETION	NO	IFL		5					
DISPATCH INTERVAL	DYNAMIC	ICF		1					
		IIP		0					

1. GROUP NAME

Name of the WLM capacity group

2. LIMIT

Group limit

3. AVAILABLE

Average unused capacity in MSU (i.e., avg. unused vector)

4. ★

When present, indicates the partition has been in the capping group for less than 4h

RMF Data Portal

RMF Data Portal - Mozilla Firefox: IBM Edition

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

RMF Monitor III Data Portal for z/OS

Explore Overview My View Home

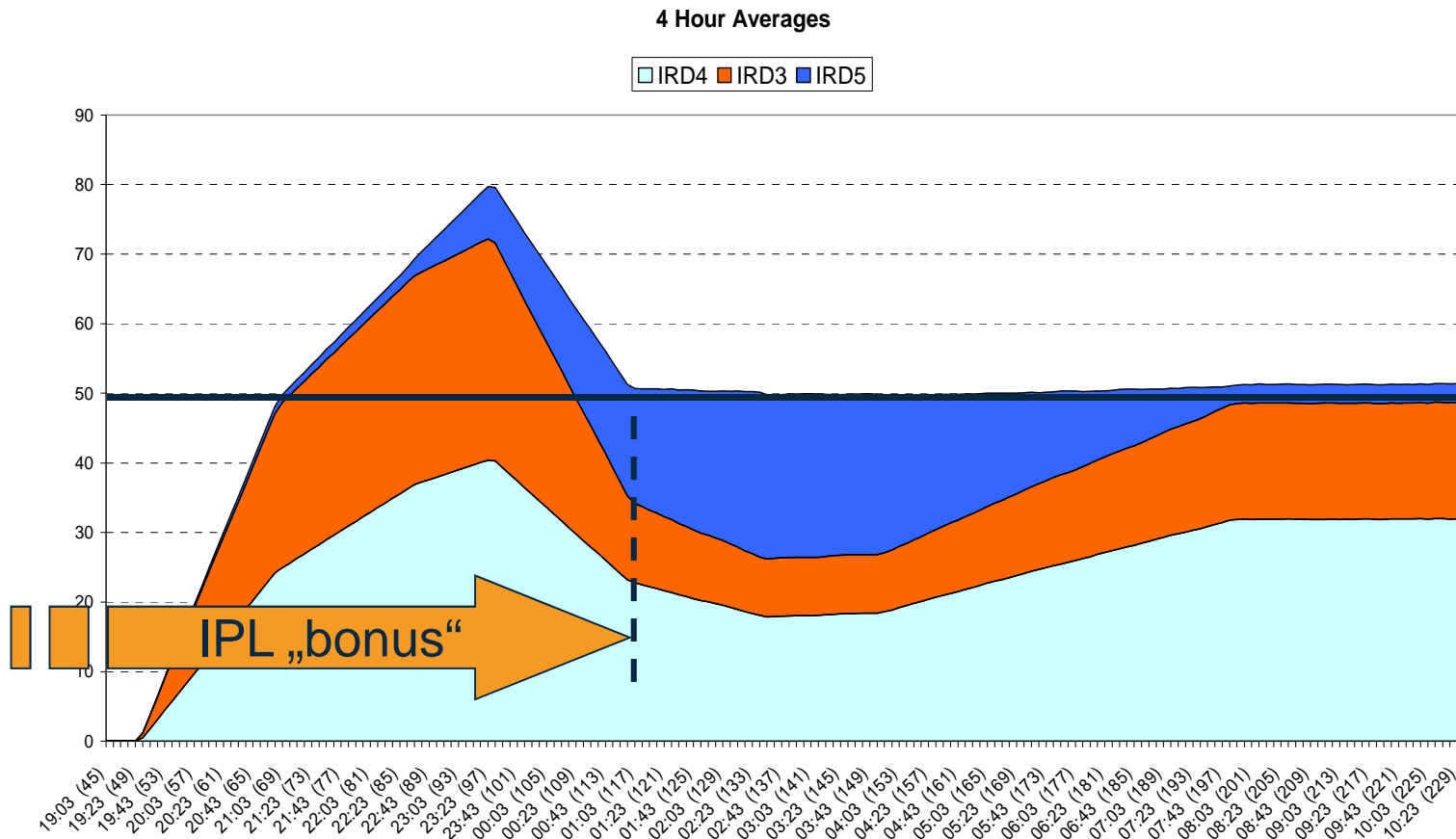
RMF Report [,TRX2,MVS_IMAGE] : CPC (Central Processor Complex)

Time Range: 03/18/2009 08:46:00 - 03/18/2009 08:47:00

Partition Name: TRX2	CPU Type: 2097	CPU Model: 704	CPC Capacity (MSU/h): 401
Weight % of Max: 19.9	4h MSU Average: 2	Capacity Group Name: RMFGRP	Image Capacity: 60
WLM Capping %: 0.0	4h MSU Maximum: 3	Capacity Group Limit: 150	Less than 4h in Capacity Group: N
Proj Time until Capping: 14400	Proj Time until Group Capping: 14400	4h Unused Group Capacity Average: 142	CPC sequence number: 000000000001EBAE
# CP Processors: 4	# ICF+IFL+AAP Processors: 0	# AAP Processors: 1	# ICF Processors: 2
# IFL Processors: 18	# IIP processors: 1	Configured Partitions: 58	Wait Completion: NO
% Capacity Used: 7	# Dedicated CPs: 0	# Dedicated AAPs: 0	# Dedicated IIPs: 0
# Shared physical CPs: 4	# Shared physical AAPs: 1	# Shared physical IIPs: 1	Vary CPU management available: NO
WLM LPAR management enabled: YES	Physical Total % of shared CPs: 5.1	Physical Total % of shared AAPs: 0.0	Physical Total % of shared IIPs: 0.0
Physical Total % of shared ICFs: 61.1	Physical Total % of shared IFLs: 0.0		

Many capping related fields are available in RMF Monitor III Data Portal

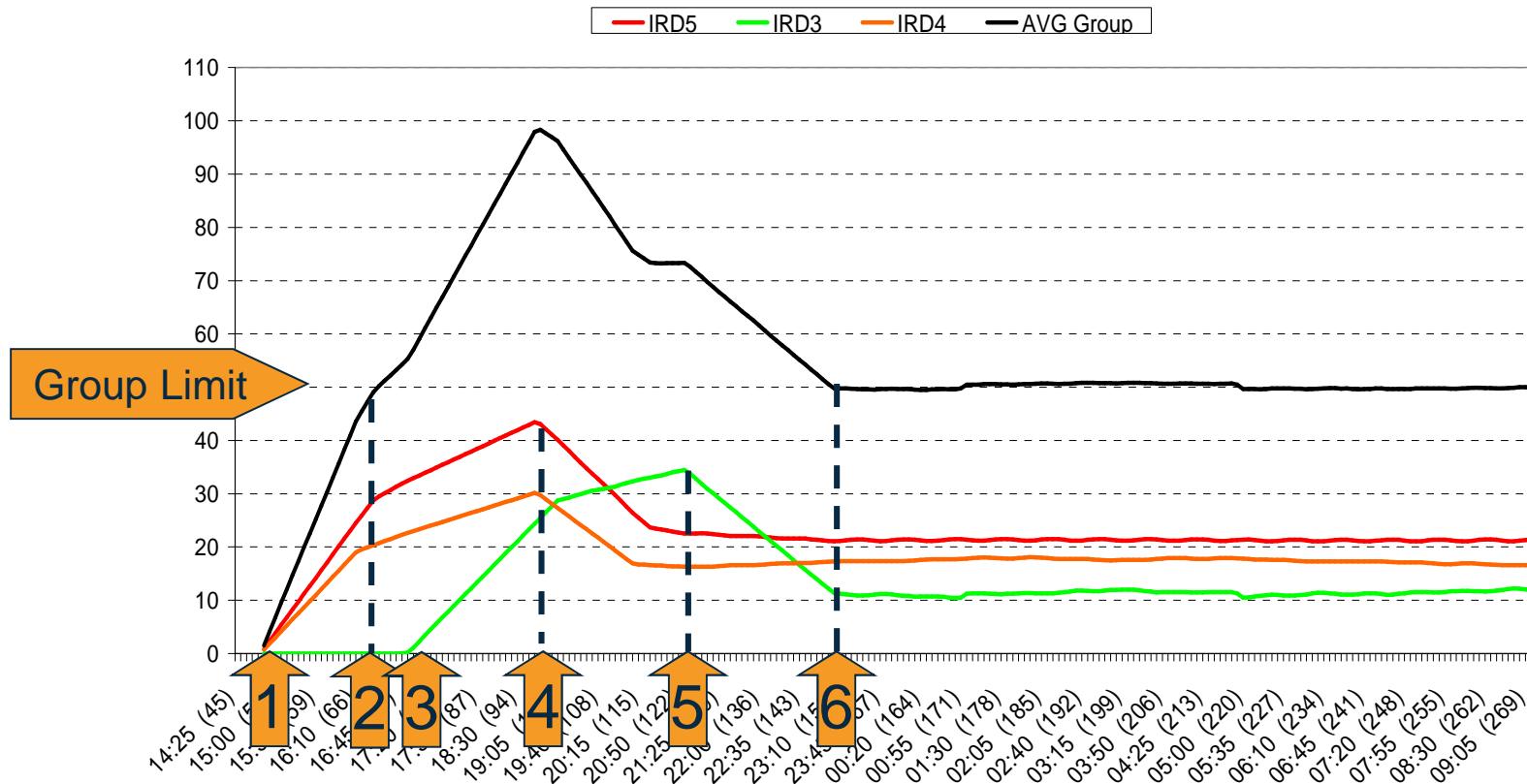
4 hour rolling average at IPL



Average is always for 4 hours even when the IPL was less than 4 hours ago

A member joins the capacity group

4 Hour Averages



1. Workloads begin on IRD4 & 5
2. Group limit reached
3. System IRD3 joins group

4. IRD4 & 5: Four hours since (1.)
5. IRD3: Four hours since (3.). All systems have same GC view.
6. Group Avg. = Group limit

Capping and HiperDispatch

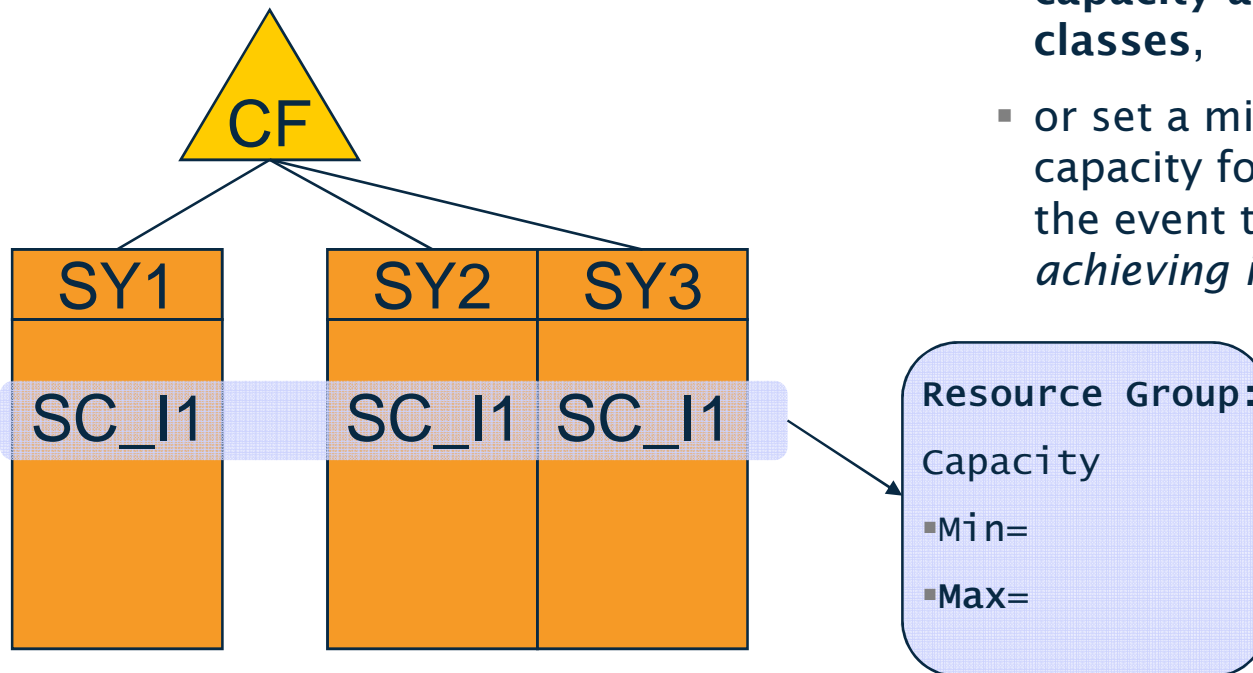
- z/OS may park vertical low (VL) processors when
 - Initial capping, or
 - Capping through cap patterns or –positive- phantom weight occurs.
 - Rationale: LPAR consumption is limited to its weight
 - Can affect CPU delays and execution velocity

Agenda

- Overview of capping types
- Initial capping
- Absolute capping
- Defined capacity & group capacity
- **Resource group capping**
- 4HRA management
- Additional Material

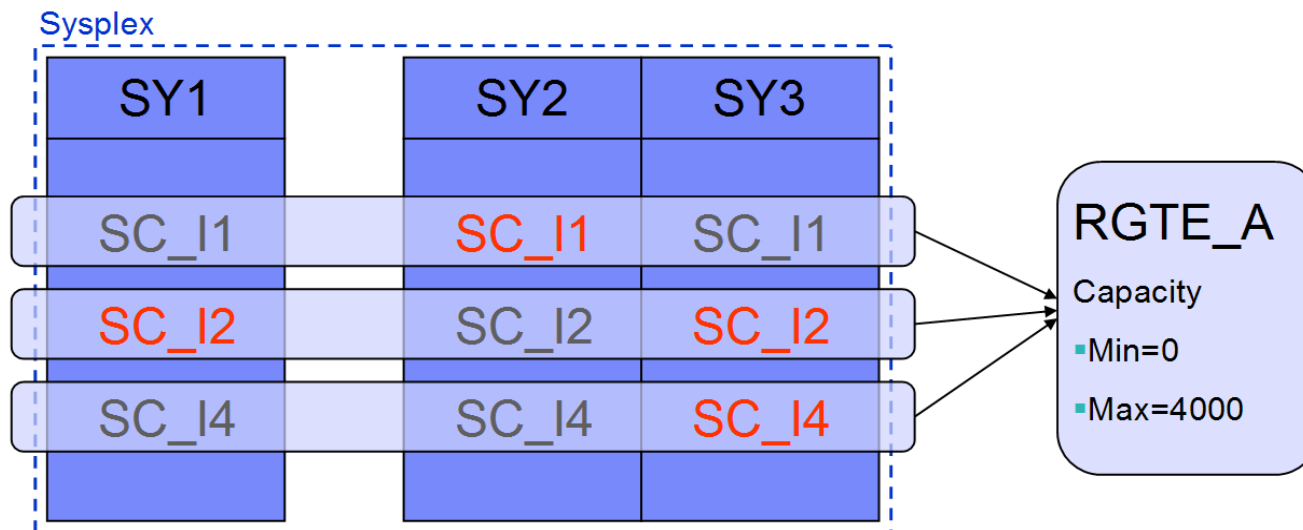
What is a Resource Group?

- Resource groups are a means to limit or protect work *when proper classification, goals and importance are not sufficient.*
- A Resource Group is associated to one or more Service Classes
- Defines the service that the related Service Class(es) are managed to. Either
 - limit the amount of processing capacity available to the service classes,**
 - or set a minimum processing capacity for the service classes in the event that the work is *not achieving its goals*



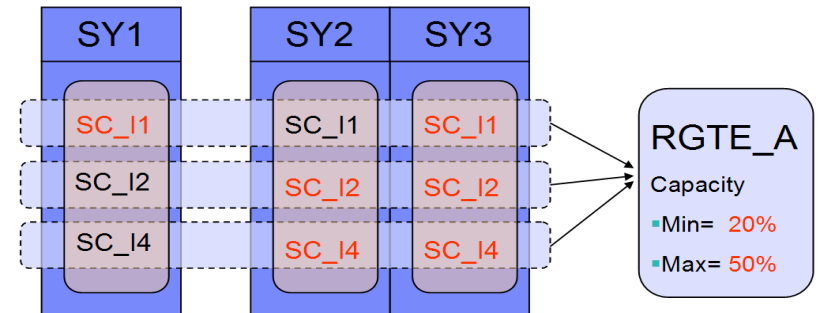
Type 1 Resource Groups

- Sysplex-wide defined in **unweighted service units per second**
 - “Unweighted” or “raw” meaning that the CPU and SRB service definition coefficients are not applied
- Sysplex-wide managed
- General Considerations
 - Multiple service classes may be assigned to a resource group
 - Different utilizations on the different systems and mix of importance levels make it difficult to predict actual consumption
 - Systems may have different capacities



Type 2 and 3 Resource Groups

- Sysplex-wide defined, but definition applies to each system
- Managed by each system
- General Considerations
 - Multiple service classes can be assigned to a resource group but this has no sysplex-wide effect
 - Definition is based on one of two possible units:
 - **Type 2: Percentage of LPAR capacity**
 - **Type 3: In number of processors (100 = 1 CP)**



Locating LPAR SU/sec Numbers

The service units that

- The Service Unit information can be located in the “z/OS MVS Planning: Workload Management” [manual](#) CPU Capacity Table
- Or on IBM Resource Link <https://ibm.biz/BdFHFv> :

IBM zEnterprise EC12

Processor	STIDP Type	STSI Model Name	CPs	SU/SEC	SRMsec/RealSec
2827-701	2827	701	1	78048.7805	1811.5932
2827-702	2827	702	2	73394.4954	1811.5932
2827-703	2827	703	3	71428.5714	1811.5932
2827-704	2827	704	4	69868.9956	1811.5932
2827-705	2827	705	5	68085.1064	1811.5932
2827-706	2827	706	6	66945.6067	1811.5932
2827-707	2827	707	7	65843.6214	1811.5932

A 4-way LPAR on a zEC12 model 7xx server can deliver approx.
 $4 * 69869$
~ 279476 SU/sec

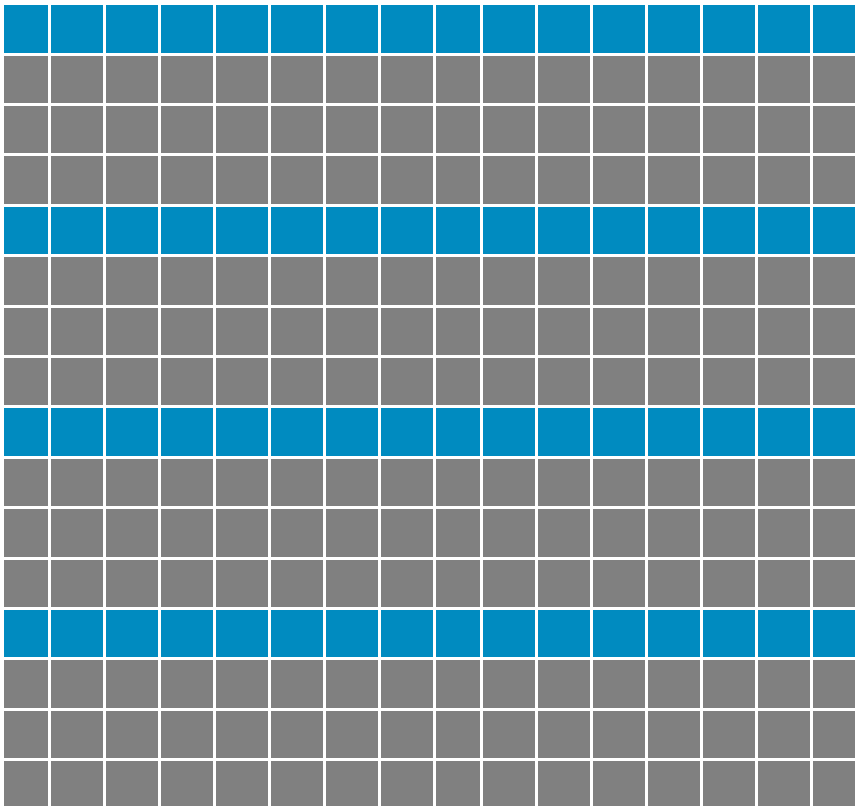
Resource Group Maximum Management

- If work in a resource group is consuming CPU above the specified maximum capacity, the system throttles back (CAPs) the associated work to slow down the rate of resource consumption.
- To cap work, WLM calls Supervisor to mark the work unit nondispatchable.
 - The ASCBUWND bit is set in address spaces and ENCBUWND bit is set in enclaves to indicate that the unit of work is not dispatchable due to resource group capping.

Resource Group Management

- To implement capping, the elapsed time is divided into 256 or 64 (pre-z/OS V2.1) slices. Each cap slice then represents $1/256^{\text{th}}$ or $1/64^{\text{th}}$ of the total elapsed time.
- Dispatchable units from address spaces or enclaves belonging to a resource group are made nondispatchable during some slices in order to reduce access to the CPU to enforce the resource group maximum.
- The time where address spaces or enclaves in a resource group are set non-dispatchable is called a **CAP SLICE**.
- The time where address spaces or enclaves in a resource group are set dispatchable is called an **AWAKE SLICE**.

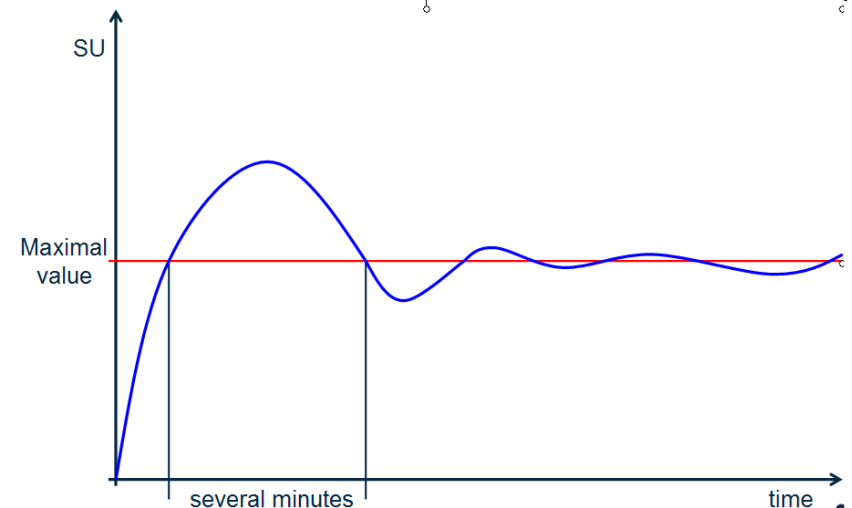
Resource Group Maximum continued...



- This table is an example of a cap pattern with 64 awake slices and 192 cap slices.
- The active slices are distributed equally over the pattern

Resource Group Maximum continued...

- Every 10 seconds the policy adjustment code re-evaluates the resource groups and adjusts the cap pattern accordingly
- The forecast for the next 10 seconds is based on the average data from the last minute
- Because of the 1 minute average data, during a ramp up period, the max may be exceeded. Also, during periods of workload oscillation WLM may tend to under cap on the up swing but over cap when the workload is dropping off.



Resource Group Maximum continued...

Under certain conditions work may continue consuming service even while being capped

- Any locked work will continue to be dispatched as long as the lock is held
 - Check promoted times in RMF workload activity report
- The region control task is exempt from this nondispatchability.
- The address space will not be marked nondispatchable until the next dispatch.

Resource Group Considerations with zAAP/zIIPs

- Resource Groups are managed based on their general purpose processor consumption (TCB+SRB)
- Difficult to predict result of assigning RGs to service classes that execute on specialty processors
 - Especially when IFAHONORPRIORITY=YES or IIPHONORPRIORITY=YES is in effect.

1	9	17	25	33	41	49	57
2	10	18	26	34	42	50	58
3	11	19	27	35	43	51	59
4	12	20	28	36	44	52	60
5	13	21	29	37	45	53	61
6	14	22	30	38	46	54	62
7	15	23	31	39	47	55	63
8	16	24	32	40	48	56	64

Other considerations for Resource Groups

- **Not valid for transaction oriented work, such as CICS or IMS transactions.**
 - In order to assign a minimum or maximum capacity to CICS or IMS transactions, the region service classes can be assigned to a resource group.
 - Such interactive work can respond harshly to CPU bottlenecks: Evaluate what cap level can be tolerated
- **Given the combination of the goals, the importance level, and the resource capacity, some goals may not be achievable when capacity is restricted.**
- Unless there is a specific need for limiting or protecting capacity for a group of work, it is best to not define resource groups and to just let workload management manage the processor resources to meet performance goals.

Identifying Resource Group Capping

- In the RMF Workload Activity report, RG capping is identified in the Execution Delays section as CAP delays
- CAP delays may also be incurred by service classes that have not been associated with resource groups
 → Discretionary Goal Management (DGM)

```

GOAL: EXECUTION VELOCITY 20.0%      VELOCITY MIGRATION:      I/O MGMT  93.9%      INIT MGMT 90.1%

SYSTEM
RESPONSE TIME EX  PERF  AVG  --EXEC USING%--  ----- EXEC DELAYS % ----- -USING%-
VEL%  INDX  ADRSP  CPU AAP IIP I/O  TOT CPU CAP I/O  CRY CNT

SYS1      --N/A--  93.9  0.2   0.0   46 N/A N/A  43  5.8 4.3 1.2 0.3  0.0 0.0

```

Discretionary Goal Management (DGM)

- Allows an *eligible over-achieving* service class to donate CPU to a discretionary period
 - Objective is to improve service that discretionary periods receive when no non-discretionary periods need help and goals are vastly overachieved
- The donation is implemented through resource group capping.
- To be considered as a donor a period must meet several requirements, including
 - Not a member of a Resource Group (RG)
 - Non-aggressive goal:
 - If it has a velocity goal, the goal must be ≤ 30
 - If it has a response time goal, the goal must be > 60 sec
 - The performance index PI must be < 0.7
- If a period should never donate due to DGM, define appropriately:
 - Velocity goal > 30 or response time goal ≤ 60 sec, or
 - Define resource group with MIN=MAX=0 and associate service classes to be protected with that RG

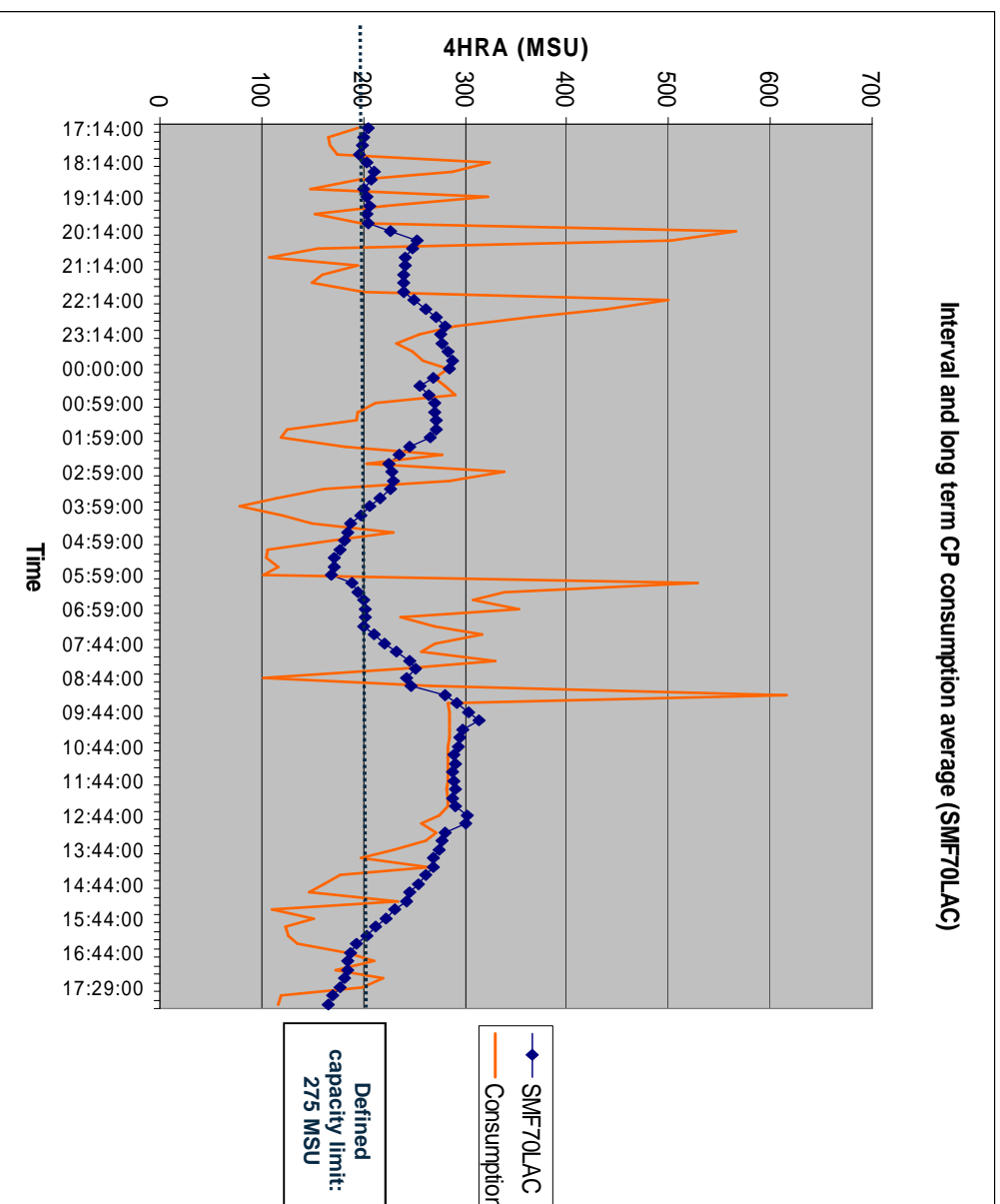
Agenda

- Overview of capping types
- Initial capping
- Absolute capping
- Defined capacity & group capacity
- Resource group capping
- **4HRA management**
- Additional Material

4HRA business aspects

- Peak value of MIN(4HRA, defined capacity limit) over billing period determines software charges
 - 4HRA peaks may exceed the defined limit
- Periods of low utilization can be used to “save” capacity for subsequent peak times
 - No capping when 4HRA < limit
- Utilization peaks drive up the 4HRA
- From a cost perspective it may be desirable to **limit the peak consumption**
- Seek for technical means to
 - Limit consumption (→peak consumption)
 - Primarily of less important work
 - Also during -previously uncapped- periods
 - Maintain service levels, responsiveness and system integrity
 - Especially for important work

Interval consumption and the 4 hour rolling average: A sample day



Capacity Provisioning Capabilities Overview

- The Capacity Provisioning Manager (CPM) can control additional capacity on IBM zEC12, z196, or z10 (plus BC10 and later)
 - Number of temporary zAAPs or zIIPs
 - Temporary general purpose capacity
- Considers different capacity levels (i.e. effective processor speeds) for subcapacity processors (general purpose capacity)
 - Can advise on logical processors
 - **Defined capacity and group capacity limits**
 - Can control one or more IBM zEnterprise or System z10 servers
 - Including multiple Sysplexes
 - Provides commands to control z196 and later static power save mode
 - Provides commands to control temporary IFLs



CPM allows for different types of provisioning requests:

- Manually at the z/OS console through Capacity Provisioning Manager commands
- Via user defined policy at specified schedules
- Via user defined policy by observing workload performance on z/OS

Policy Approach

The Capacity Provisioning policy defines the circumstances under which additional capacity may be provisioned:

- Three “dimensions” of criteria considered:
 - **When** is provisioning allowed
 - **Which** work qualifies for provisioning
 - **How much** additional capacity may be activated
- These criteria are specified as “rules” in the policy:

```
If
{ in the specified time interval
  the specified work “suffers”
}
Then up to
{ - the defined additional capacity
  may be activated
}
```

- The specified rules and conditions are named and may be activated or deactivated selectively by operator commands

Capacity Provisioning Policy Strategies... for cost optimization with LPAR defined capacity

- Baseline defined capacity (DC) limit relatively low
 - but still realistic for periods of low to average utilization
- Use CPM rules to increase DC limit when needed
 - Time conditions without workload conditions:
Unconditionally provision full rule scope
 - Time & workload conditions:
Allow for higher DC limits as required by workload
- Can differentiate between systems, or service definitions
- Group capacity analogously
- See following scenario

Capacity Provisioning Policy Sample... ... with LPAR defined capacity (1)

- Two workloads that may warrant higher DC limits

Maximum Processor Scope	Logical Processor Scope	Maximum Defined Capacity Scope	Maximum Group Capacity Scope	Rules
<input checked="" type="checkbox"/> <input type="checkbox"/> Actions ▼				
Name Filter	Description Filter	Default Status Filter		
<input type="checkbox"/> WeekNight	Weekdays DC pre midnight batch	<input checked="" type="checkbox"/> Enabled		
<input type="checkbox"/> WeekdayDC	Weekdays DC for online work	<input checked="" type="checkbox"/> Enabled		

- WeekdayDC rule scope allows for up to +300 (additional) MSU:

Processor Scope	Defined Capacity Scope	Group Capacity Scope	Conditions
<input checked="" type="checkbox"/> <input type="checkbox"/> Actions ▼			
System Filter	Sysplex Filter	Max. Increase (MSU) Filter	
<input type="checkbox"/> SYS1	PLEX1	300	

Capacity Provisioning Policy Sample... ... with LPAR defined capacity (2)

- Rule is enabled for all weekdays prime time

Nonrecurring Time Conditions

Recurring Time Conditions

Workload Conditions

Actions

	Name Filter	Start Date Filter	End Date Filter	Mon Filter	Tue Filter	Wed Filter	Thu Filter	Fri Filter	Sat Filter	Sun Filter	Start Time ▲ Filter	Deadline Filter	End Time Filter
<div></div>	AllWeekD	Jan 2, 2014	Dec 31, 2014	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>	7:45 AM	6:00 PM	6:30 PM

- Workload is defined by specific service classes

Importance Filters

Included Service Classes

Excluded Service Classes

☒☐

Actions ▼

	Service Definition Filter	Service Policy Filter	Service Class Filter	Period Filter	Provisioning PI Filter	Provisioning Duration (Minutes) Filter	Deprovisioning PI Filter	Deprovisioning Duration (Minutes) Filter
<input type="checkbox"/>	Any service definition	Any service policy	DB2HI	1	1.4	2	1.1	10
<input checked="" type="checkbox"/>	Any service definition	Any service policy	ONLSTC	1	1.5	5	1.1	10

Capacity Provisioning Policy Sample... ... with LPAR defined capacity (3)

- Similarly, another rule is defined to cover a batch workload
 - Up to +70 MSU for a single batch service class

Nonrecurring Time Conditions

Recurring Time Conditions

Workload Conditions

Actions

	Name Filter	Start Date Filter	End Date Filter	Mon Filter	Tue Filter	Wed Filter	Thu Filter	Fri Filter	Sat Filter	Sun Filter	Start Time Filter	Deadline Filter	End Time Filter
<div></div>	AllWeekN	Jan 2, 2014	Dec 31, 2014	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>	8:00 PM	10:00 PM	10:00 PM

Importance Filters									
Included Service Classes									
Excluded Service Classes									
Actions									
	Service Definition Filter	Service Policy Filter	Service Class Filter	Period Filter	Provisioning PI Filter	Provisioning Duration (Minutes) Filter	Deprovisioning PI Filter	Deprovisioning Duration (Minutes) Filter	PI Scope Filter
<input type="checkbox"/>	Any service definition	Any service policy	BATCRIT	1	1.8	5	1.3	10	System

z/OS Capacity Provisioning Documentation

- For more information contact: IBMCPM@de.ibm.com
- *z/OS Capacity Provisioning: Introduction and Update for z/OS V2.1, SHARE in Anaheim, Session 14210, 8/2013*
- Website <http://www.ibm.com/systems/z/os/zos/features/cpm>
- *z/OS MVS Capacity Provisioning User's Guide, SC34-2661, at <http://publibz.boulder.ibm.com/epubs/pdf/iea3u100.pdf>*
- *ITSO Redbook: System z10 Enterprise Class Capacity on Demand, SG24-7504 <http://www.redbooks.ibm.com/abstracts/sg247504.html?Open>*



z/OS Workload Management

– More Information –

- z/OS WLM Homepage:

<http://www.ibm.com/systems/z/os/zos/features/wlm/>

- Inside WLM: <https://ibm.biz/BdF4L4>
- WLM Capping Technologies: <https://ibm.biz/BdF4Lr>

- z/OS MVS documentation

- z/OS MVS Planning: Workload Management:

<http://publibz.boulder.ibm.com/epubs/pdf/iea2w1c0.pdf>

- z/OS MVS Programming: Workload Management Services:

<http://publibz.boulder.ibm.com/epubs/pdf/iea2w2c0.pdf>

- *IBM Redbooks publications:*

- System Programmer's Guide to: Workload Manager:

<http://publib-b.boulder.ibm.com/abstracts/sq246472.html?Open>

- ABCs of z/OS System Programming Volume 12

<http://publib-b.boulder.ibm.com/abstracts/sq247621.html?Open>

Workload Manager

Welcome to WLM/SRM



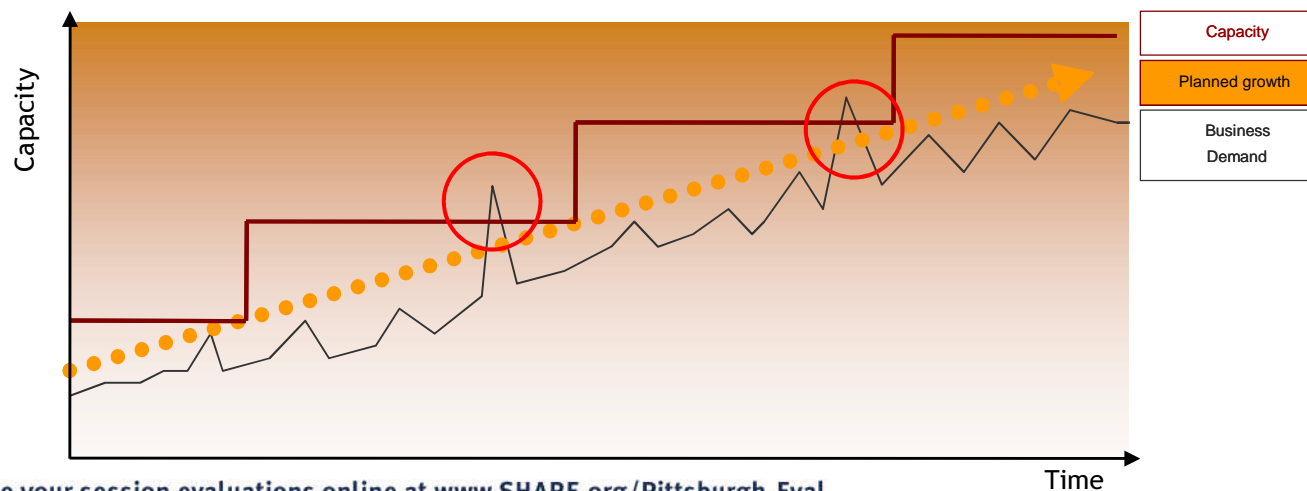
Agenda

- Overview of capping types
- Initial capping
- Absolute capping
- Defined capacity & group capacity
- Resource group capping
- 4HRA management
- **Additional Material**

IBM z/OS Capacity Provisioning Basics



- **Contained in z/OS base component free of charge**
 - Requires a monitoring component, such as z/OS RMF, or equivalent
 - Base element since z/OS V1.9
- **Exploits on System z On/Off Capacity on Demand Feature**
 - IBM zEnterprise System z10 or later
 - If On/Off CoD is not used CPM “analysis” mode may be used for monitoring and alerts
- **Exploits Defined Capacity and Group Capacity**
 - Defined Capacity with IBM System z10 or later
 - Group Capacity with IBM zEnterprise z196 or later



धन्यवाद

Hindi

多謝

Traditional Chinese

ขอบพระคุณ

Thai

Спасибо

Russian

Gracias

Spanish

Thank You

English

Obrigado

Brazilian Portuguese

شكراً

Arabic

Grazie

Italian

多谢

Simplified Chinese

Danke
German

Bedankt

Dutch

Merci
French

நன்றி

Tamil

ありがとうございました

Japanese

감사합니다

Korean