z/TPF V1.1

TPF Users Group Spring 2008

TPF Shared PR/SM Performance Considerations

Robert Blackburn Ph.D.

AIM Enterprise Platform Software
IBM z/Transaction Processing Facility Enterprise Edition 1.1.0

Preliminaries

- **Assume no dedicated LPARs or just subtract their CPs from the CEC total**
- **Assume no capping of LPARs**
- **'Wait Completion = No' is very strongly recommended**
  - Event driven dispatching by a wait or interrupt
  - Can fully utilize CEC
- **'Wait Completion = Yes' is not recommended**
  - Processing weights always in effect – how scheduled
  - LPAR is given entire dispatch interval even if it sits in wait state
  - Significant response time increase
  - Very slight ITR gain

Preliminaries

- **Choosing PR/SM run time**
  - quite robust
  - 7 to 13 mills is fine
  - Too small – larger overhead
  - Too large – response time issues
- **The busy period**
  - Mean busy period length, denoted E(B)= E(s)/(1-p)
    - where E(s) = mean service time = instructions per IO (often in TPF systems)
    - p = utilization
  - e.g. at p=.8 and E(s)=10K we have E(B) = 10K / .2 = 50K
  - With 300 MIPs CP this is 50K / 300E6 = 1/6 mill
  - Thus mean busy period much smaller than PR/SM run time
    - Reason for above robustness
    - z/OS might have E(s) = 500K not 10K

Weights determine share as p -> 1

- **Weights are a way to specify the portion of the CEC to which the LPAR is entitled**
- **Weights come into play if and only if the number of logical CPs being dispatched is greater than the number of physical CPs available**
  – the PR/SM history is a relatively short time interval
- **Low priority logical processor may be preempted if an IO interrupt is pending for a higher priority logical processor**
- **Higher priority**
  – Further behind in its share
  – Not as far ahead in its share

Virtual CP considerations

- **Number of logical CPs in LPAR determine upper bound for CPU usage**
  - e.g. if CEC is 6way and LPAR is 4way then LPAR max usage of CEC < .66(4/6) regardless of weight
- **Never define more logical CPs than needed in the peak time of day/day of month**
  - Worthwhile to determine your peak period
  - Very large fluctuations in workload intensity may need high number of CPs

Virtual to Real --- Customer Measured Effect

- **Old CEC was a real 4way (R=4)**
  - TPF shared 3way(V=3)
  - VM shared 2way(V=2)
  - V=3+2
  - V/R = 5/4 = 1.25
- **New CEC was a real 3way**
  - V/R = 5/3 = 1.66
  - Performance was not as predicted
- **Changes on the new CEC**
  - defined VM as 1way
  - Now V/R = 4/3 = 1.33 near the original 1.25
- **Result: 4 to 5% performance improvement**
  - Brought new CEC to performance expectations

Details on V/R Effect

- **z990 and z9 specific**
- **L1(Instruction and data) 256K**
  - minimal value for just dispatched LPAR
    - most of its previous entries gone
- **L2 is 32M per book(8 CPs)**
  - Entries will persist across dispatches
- **TLB and TLB2**
  - TLB 512 entries
  - TLB2 512/4K (CRSTE and PTE)
    - Keep entries for several LPARs in TLB2 at same time
    - Significant performance gain when numerous images running
    - PTLB only done for those entries formed by currently active LPAR
- **As increase the number of virtual CPs the L2 and TLB2 become essentially smaller for each LPAR**
  - Wait/sec
  - Cache footprint
  - Partitioned cache has lower hit ratio than shared cache
    - get lower bound on performance
    - use the square root rule to estimate effect
  - Calculate incremental misses to memory

Routing Weights – customer example

- **This effect becomes more important as TPF MPs share the CEC**
- **CEC has 3 real CPs**
  - TPF LPAR1 has 3 shared CPs
  - TPF LPAR2 has 1 shared CP
- **Deliberately LPAR1 has weights set higher than it ever uses**
  - even over a short period, say, 5 minutes
- **RESULT: LPAR1 essentially gets all the CPU it needs and LPAR2 gets the remainder**
  - Essentially a priority queue

Weights ---Problem/Explanation

- **LPAR1 p= .6**
- **LPAR2 p=.8**
  - Input list queues of 1200
- **CEC utilization**
  - (3x.6 + .8)/3 = .86
  - Why is LPAR2 acting as if its utilization=1?
- **Assume independence of 3 virtual CPs in LPAR1**
  - P(all 3 CPs busy) = $.6^3$ = .22
  - P(at least one CP available) = 1-.22 = .78
  - This is very close to actual LPAR2 util of .8
- **Thus LPAR1 leaves 1.2 CPs of power unused that LPAR2 can not fully use**
  - with only 1 defined virtual CP

Weights - solution

- **Define LPAR2 as a 2way**
- **Can calculate P(at least 2 CPs free) by LPAR1**
  - Calculate with binomial distribution
  - Answer is .288 + .064 = .352
- **Sufficient for LPAR2 to fully utilize 1 CP**
- **Customer must balance V/R costs vs ability to exploit free cycles**
  - Effect is significantly lessened with larger LPAR MPs
  - E.g. 6^{10} = .006 = P(10 are busy)
- **Note this priority situation is somewhat unusual when both LPARs are TPF**
  - Common with other LPAR being z/OS or VM

Multiple TPF Shared LPARs in a CEC

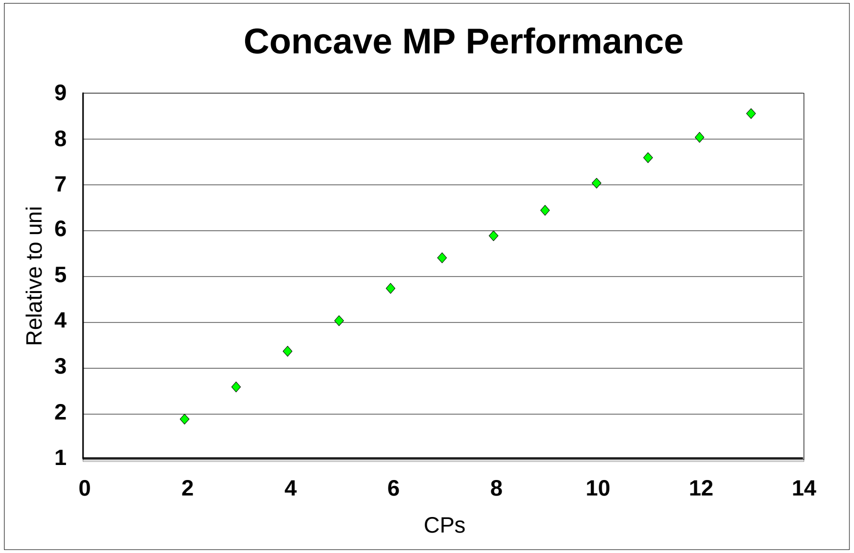
- **Very different than a single TPF LPAR with one/several nonTPF systems acting as MIPS soaks**
- **Customers accepted that VM could suffer significantly reduced MIPS as TPF increased its utilization**
- **With 2 or more TPF LPARs competing**
  - generally not a lower priority TPF
- **Thus total CEC utilization is now the critical factor**
  - It is as important as native CEC utilization used to be
  - z/TPF LPAR can measure the entire CEC utilization
  - TPF4.1 LPAR can not

Issues involved with large CEC performance analysis having multiple LPARs

- **Crossing several generations of machine design**
  - especially nonIBM to IBM
- **Large change in MP level**
  - e.g. from 8way to 3way or reverse
- **Type of TPF workloads**
  - RES, Fares, FEP
- **Potential TPF workload changes**
- **Using throughput measures other than the TPF ITRRs**
  - Gartner, LSPR etc
- **Type of LPARs sharing the CEC**
- **TPF wait/sec is much larger than other operating systems**
  - p(1-p) / E(s)
  - E(s) is much smaller for TPF systems

Performance estimate of TPF LPAR under large CEC

- **TPF LPAR is 8w on 13w CEC**
- **MP performance is concave**
  - chord of any 2 points on the graph lies below the graph
- **8w = 5.9 (power of uni)**
- **13w = 8.58**
- **(8/13)13w = 5.28**
- **Thus linear estimate can significantly differ from the actual point**
- **Best estimate is near the actual MP value**
  - More work needed here



References

- **May/July 2004 IBM Journal of R&D**
  – www.research.ibm.com
- **System z10 PR/SM Planning Guide**

Trademarks

- Update the following as appropriate. refer to http://www.ibm.com/legal/copytrade.shtml and don't leave this line in
- IBM, xxx and xxxx are trademarks of International Business Machines Corporation in the United States, other countries, or both.
- delete any of the following which are not used. Update the Windows trademark to reflect only those items used. delete this line
- Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
- Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
- Intel, Intel Inside (logos), MMX, Celeron, Intel Centrino, Intel Xeon, Itanium, Pentium and Pentium III Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States, other countries, or both.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Linux is a trademark of Linus Torvalds in the United States, other countries, or both.
- Other company, product, or service names may be trademarks or service marks of others.
- Notes
- Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
- All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
- This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
- All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
- Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.
- Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.
- This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.