usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Boosting GPU Virtualization Performance with Hybrid Shadow Page Tables

Yaozu Dong

Mochi Xue

Xiao Zheng

Jiajun Wang

Zhengwei Qi

Haibing Guan

Shanghai Jiao Tong University

Intel Corporation

GPU Usage

- Gaming (2D/3D graphic)
- HD video hardware decoding
- High performance computing

High Performance Computing shifts computation-intensive workloads to cloud environment.

Machine learning	Molecular dynamics simulations	Media transcoding

A new computing paradigm: GPU Cloud

gHyvi

An optimized GPU virtualization scheme based on gVirt.

GMedia: A media transcoding benchmark	Relaxed Shadow Page Table	Adaptive Hybrid Page Table Shadowing Policies

Up to 13x performance of gVirt and 85% of native.

GPU benchmarks

2D/3D graphic

For OpenGL and DirectX commands
Such as 3DMark, PassMark

GPGPU

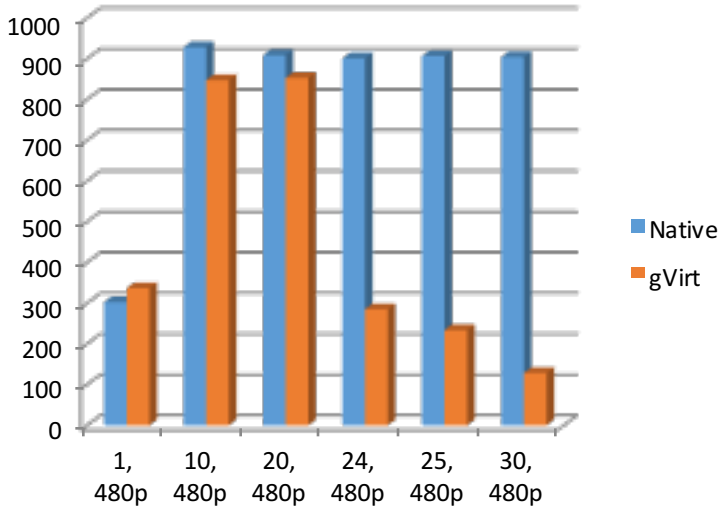
For CUDA and OpenCL commands
Such as Rodinia [ATC 2013], Parboil

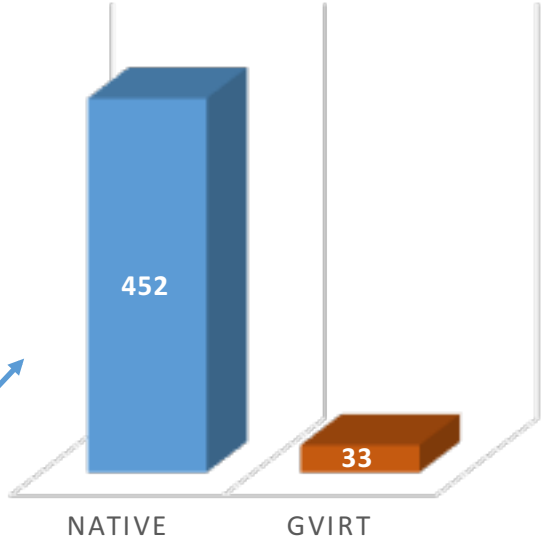
GMedia

A hardware media transcoding benchmark based on Intel's MSDK(Media Software Development Kit).

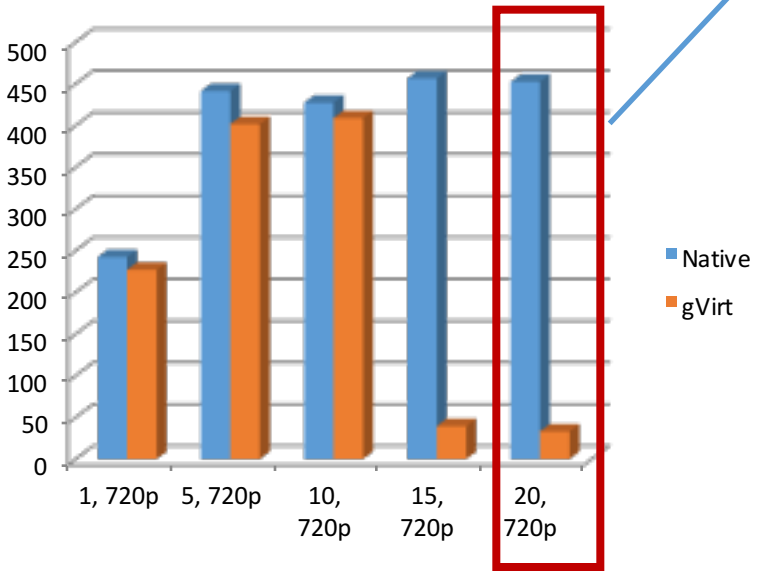
- MSDK provides APIs for hardware acceleration.
- A wrapper invokes media functions form MSDK.
- Test cases can run with assigned threads and settings.
- The FPS results reflect the performance.

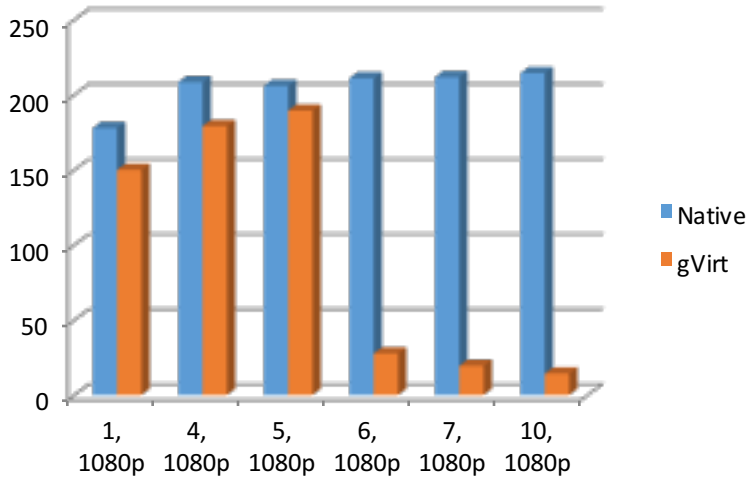
Massive update issue



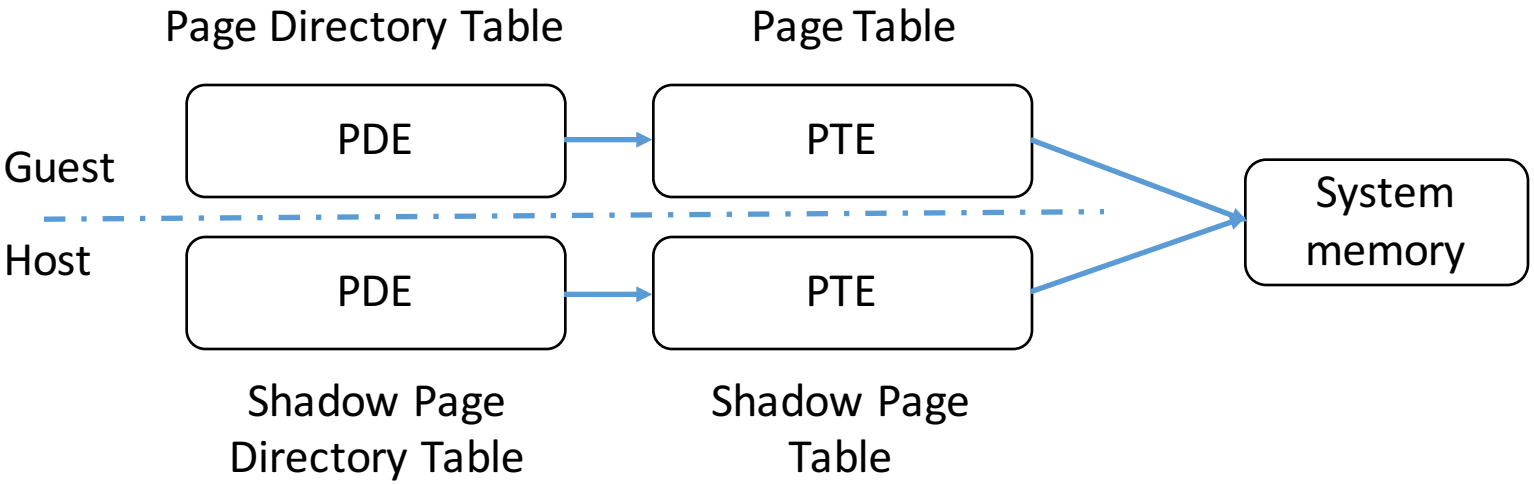


93% performance degrade.



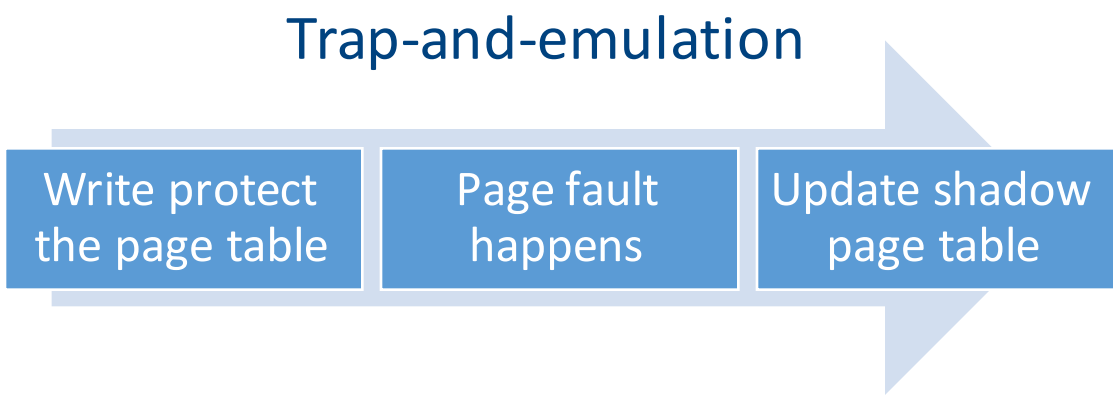


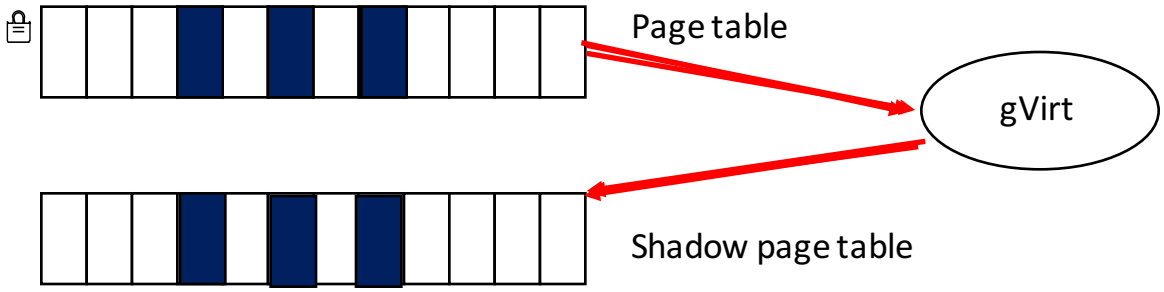
Shadow page table



- 512 * 1024 * 4KB = 2GB
- Resource partition
- User space isolation

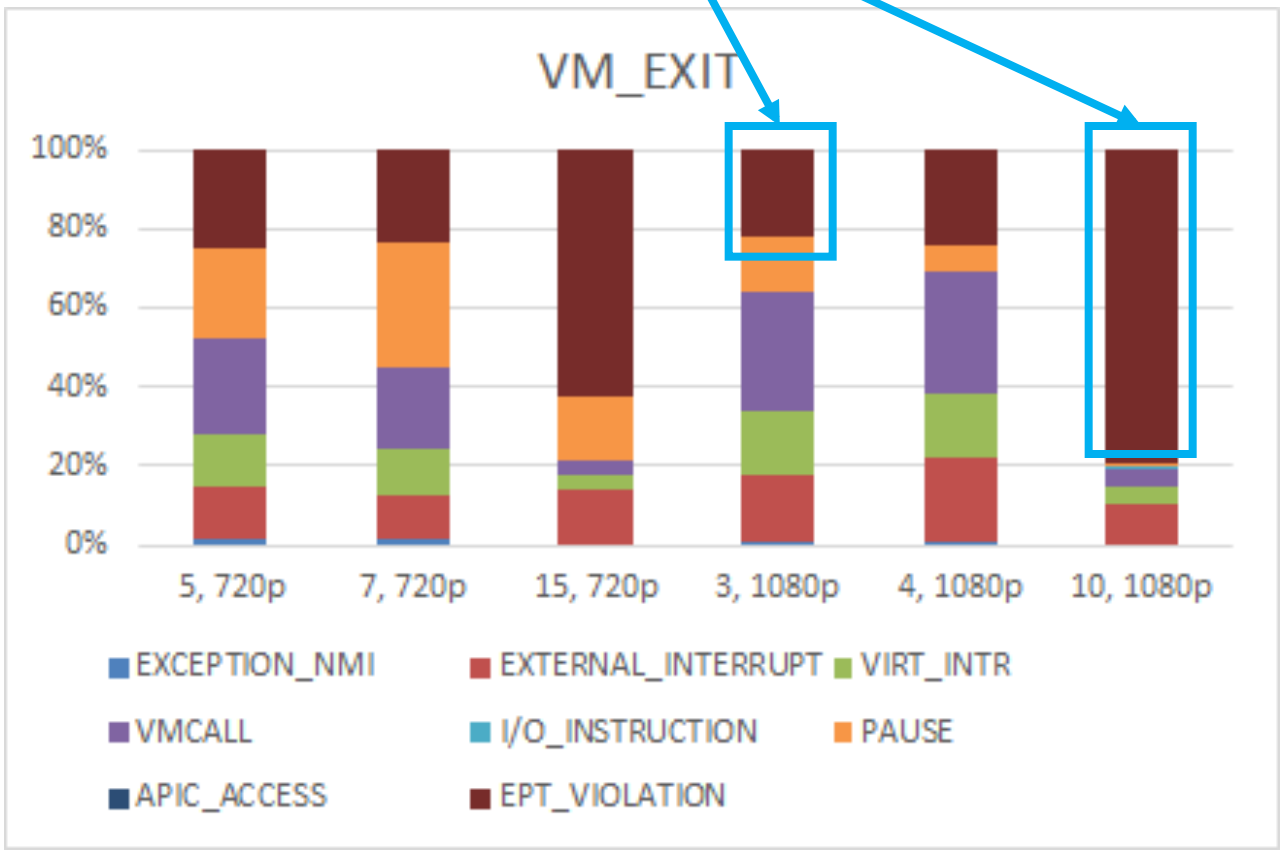
Strict shadow page table





Case profiling

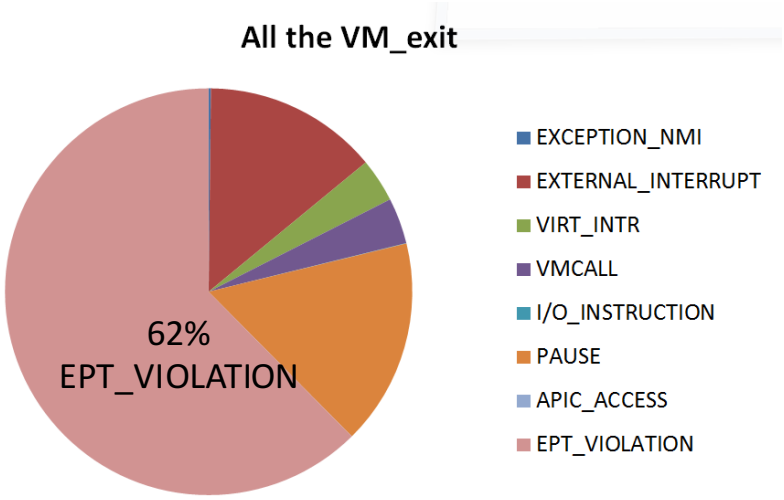
Proportion changes
21.43% →79.45%

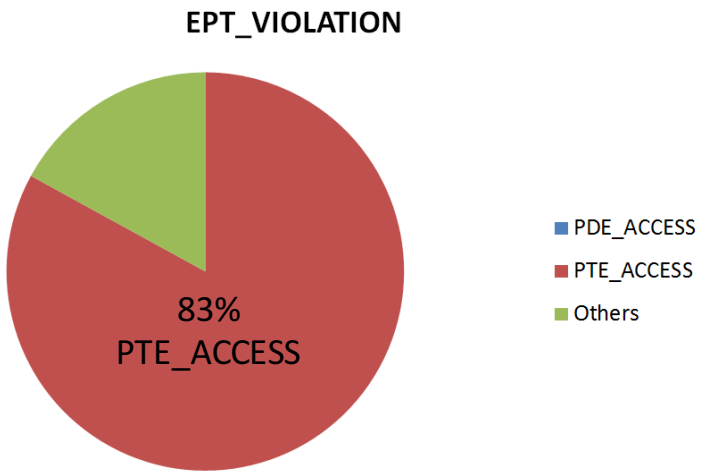


Breakdown of EPT_VIOLATION

Take “15thread-720p” as an example:

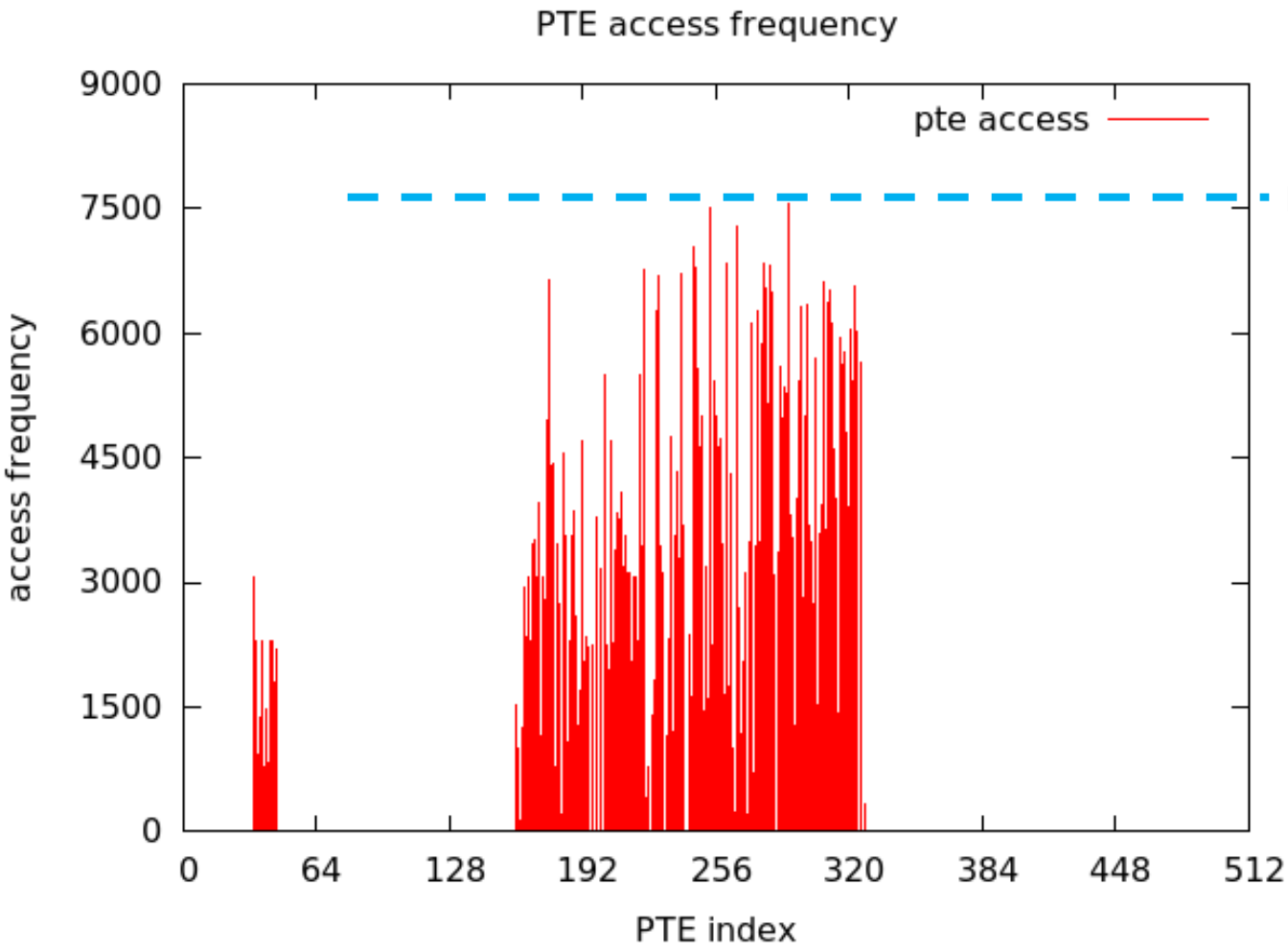
- 62% of all VM_exit is due to EPT_VIOLATION
- 83% of EPT_VIOLATION is caused by PTE access





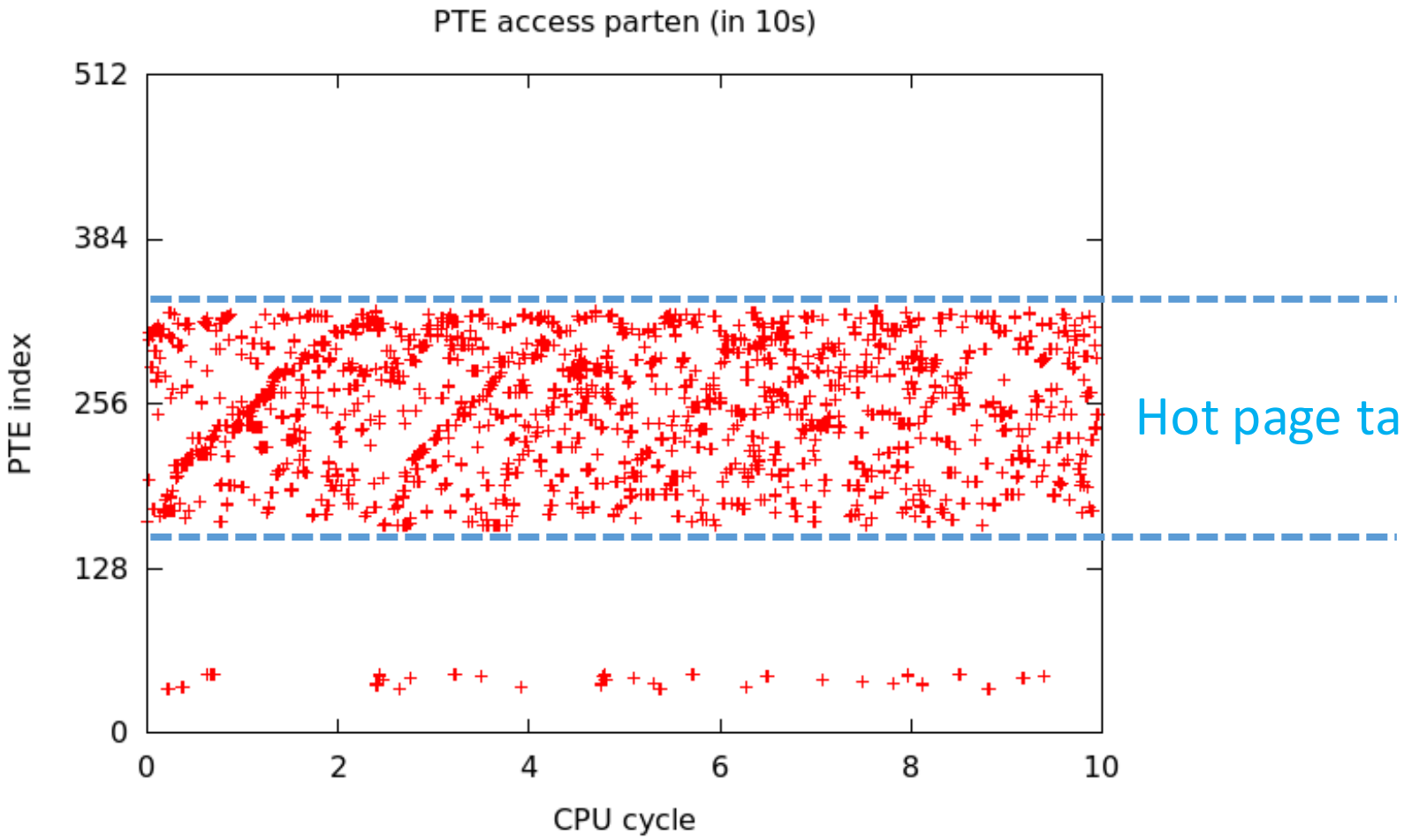
Frequency

The frequency of update in 10s.

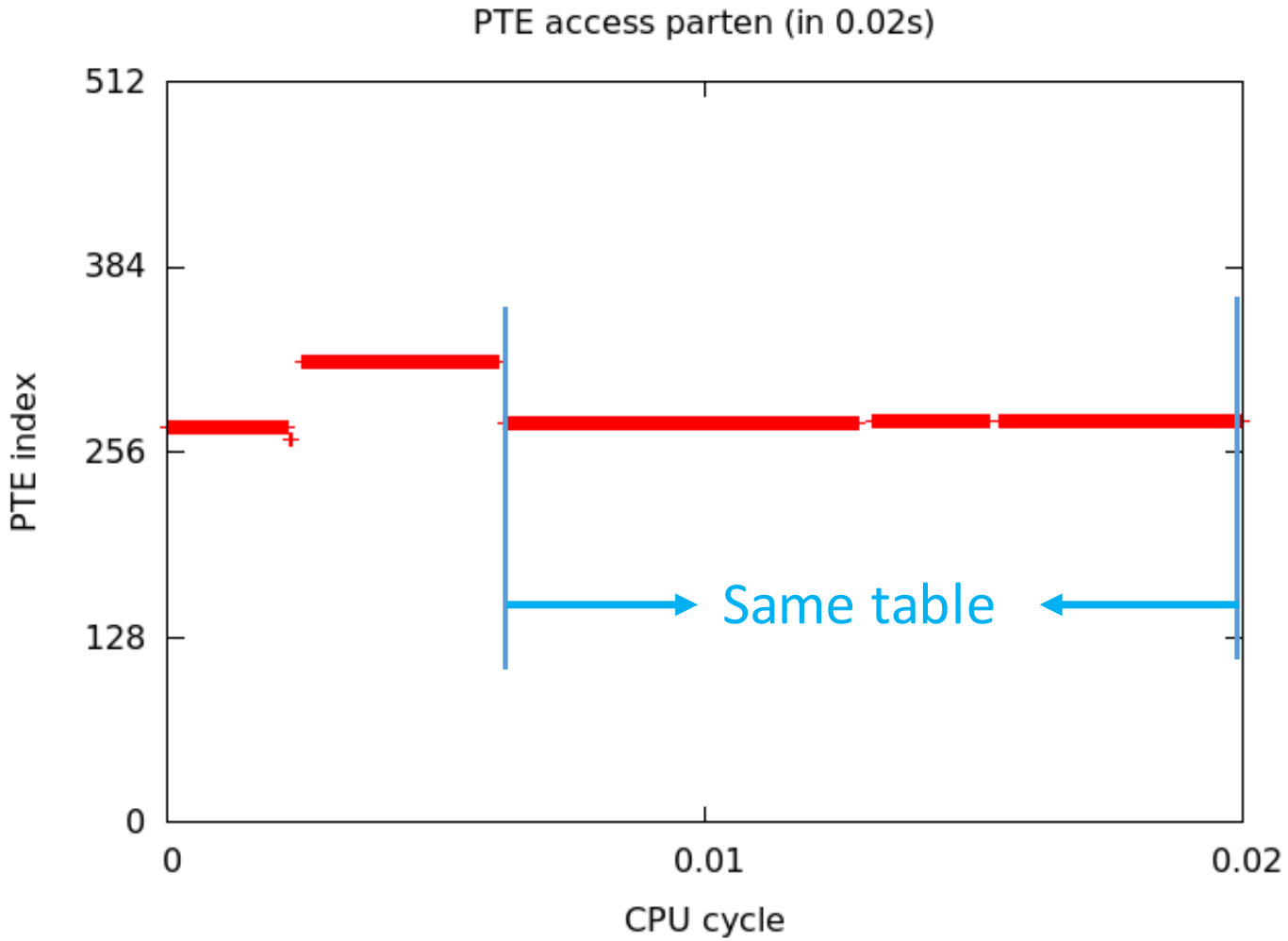


Up to 7.5k times

Hot page tables



Continuous updates



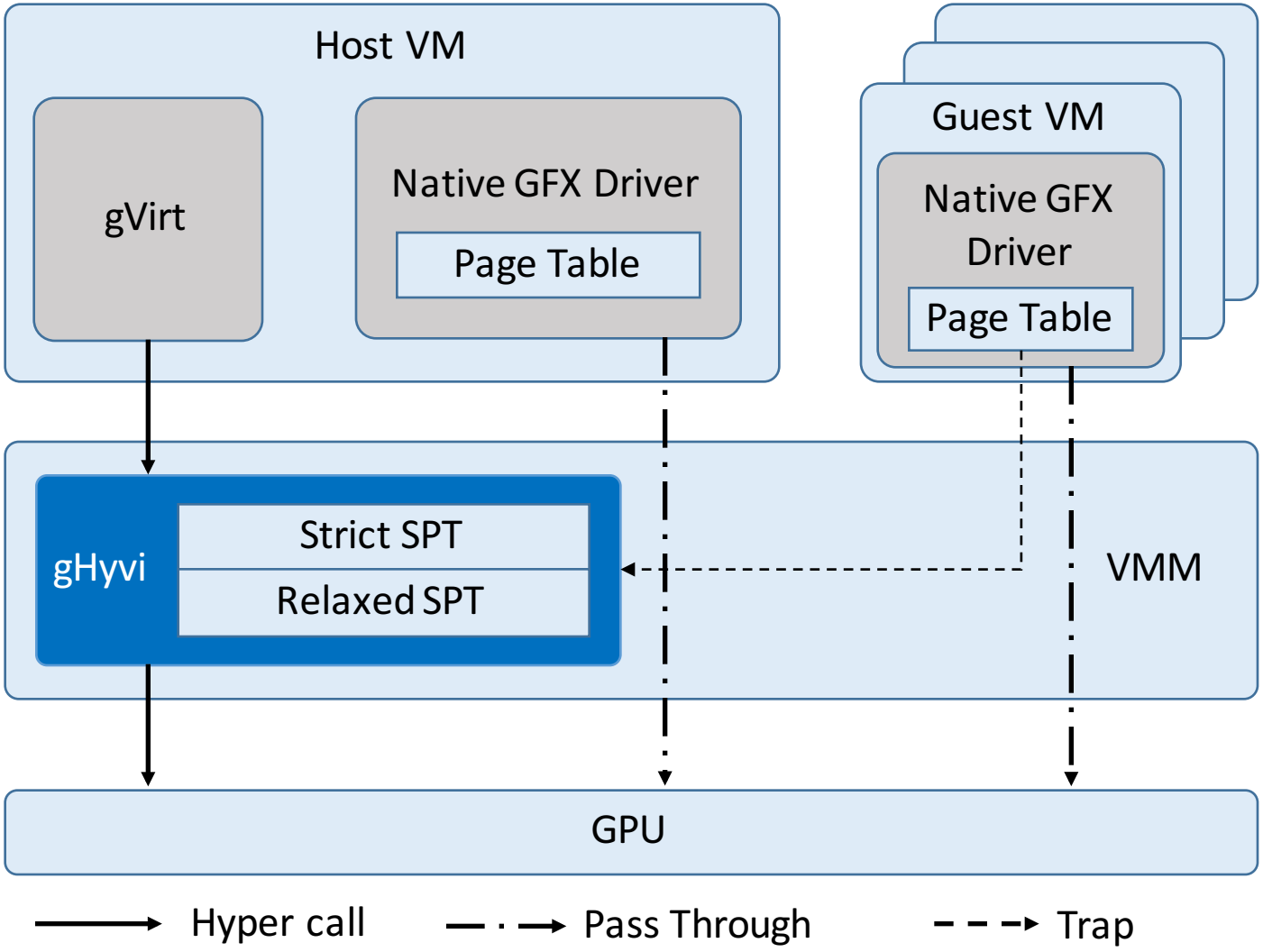
Conclusion of massive update issue

VM is frequently swapping graphic memory pages.
It modifies the entries of page table massively.

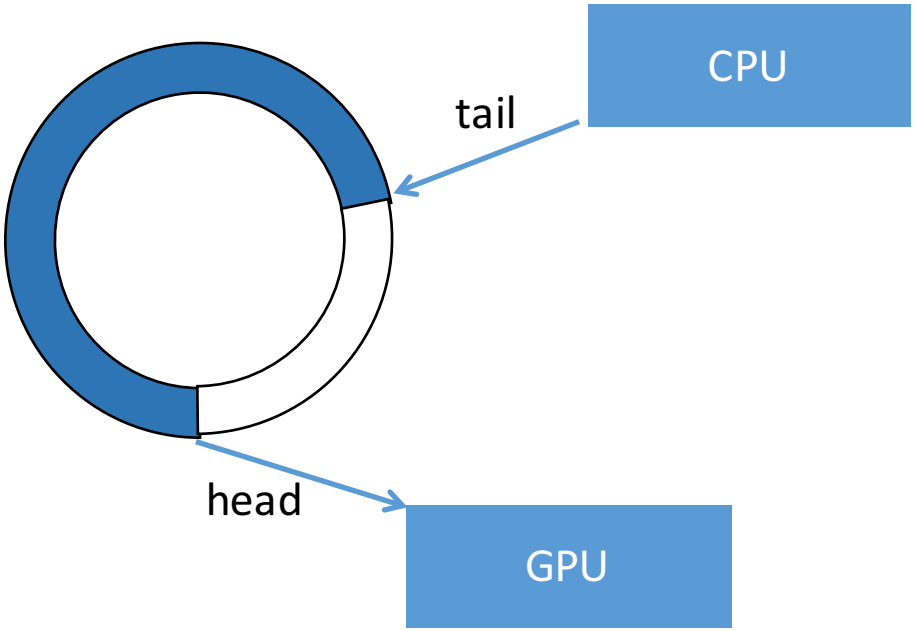
- Large amount of updates(7.5k in 10s)
- Updates focus on certain pages (hot pages)
- Updates are continuous (on the same page)

Modifications lead to busy trap-and-emulations.
Eventually, overhead happens. (Up to 95%)

gHyvi architecture

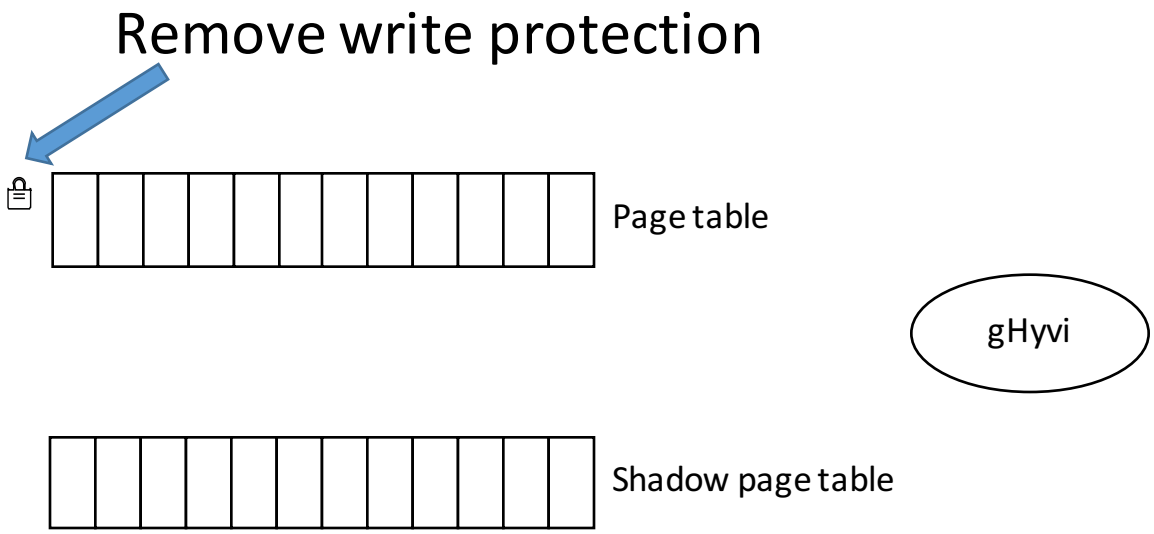


GPU programming model

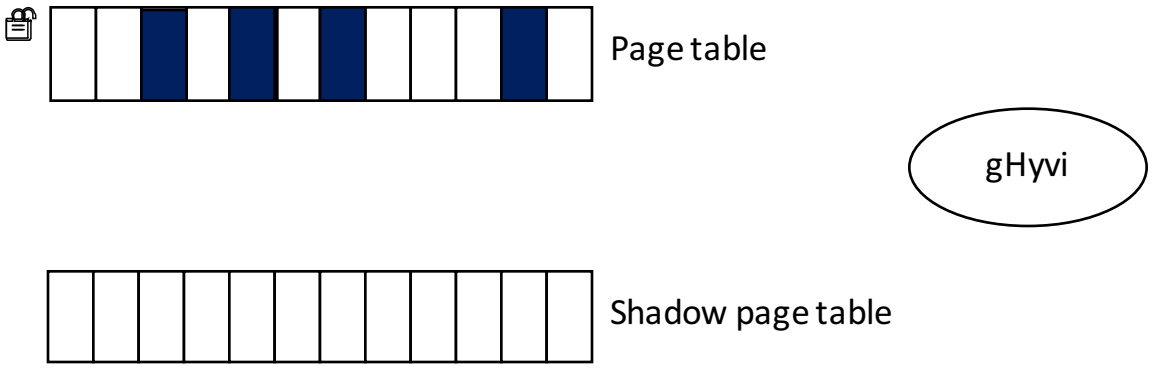


The commands fed by CPU won't take effect until they are fetched by the GPU.

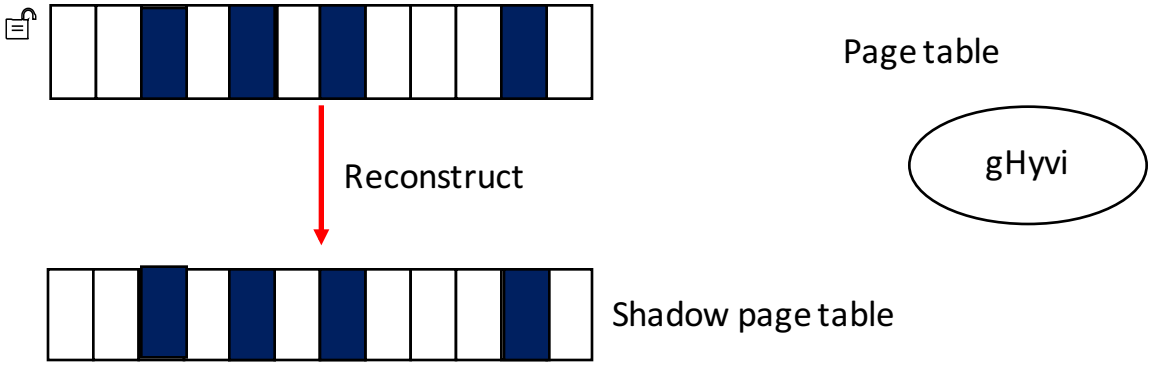
Relaxed shadow page table



Relaxed shadow page table

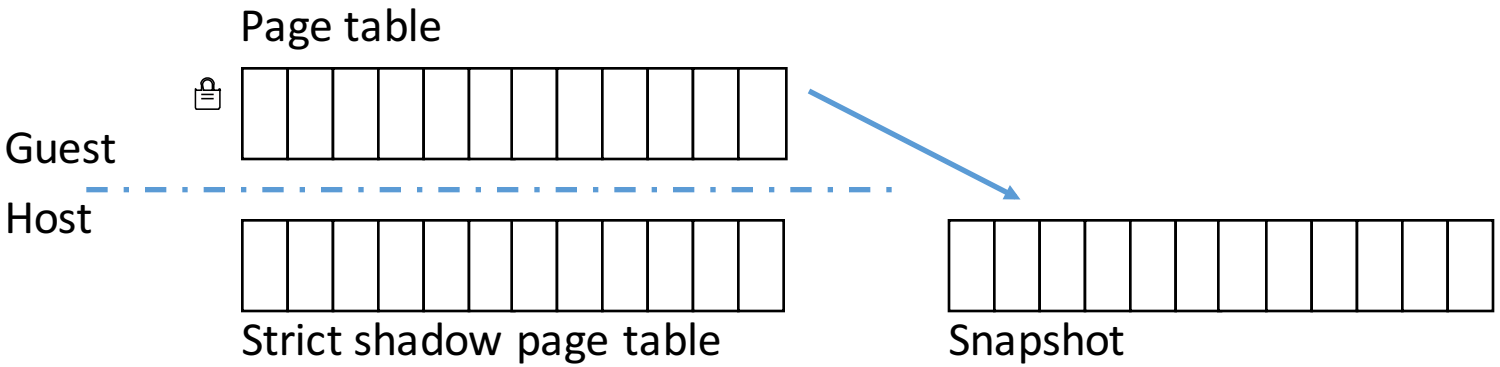


Relaxed shadow page table



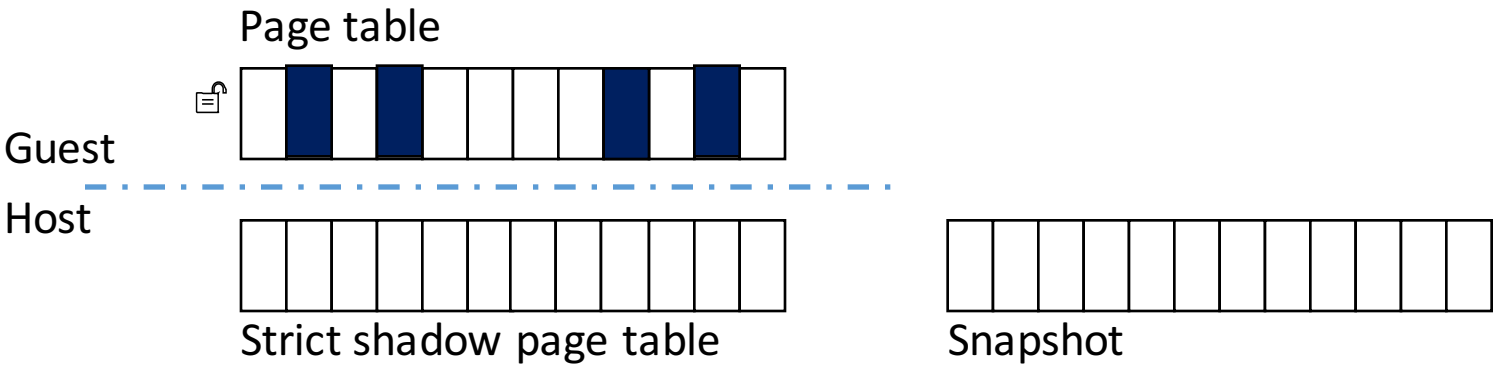
Page table reconstruction

Step1: Take a snapshot of guest page table



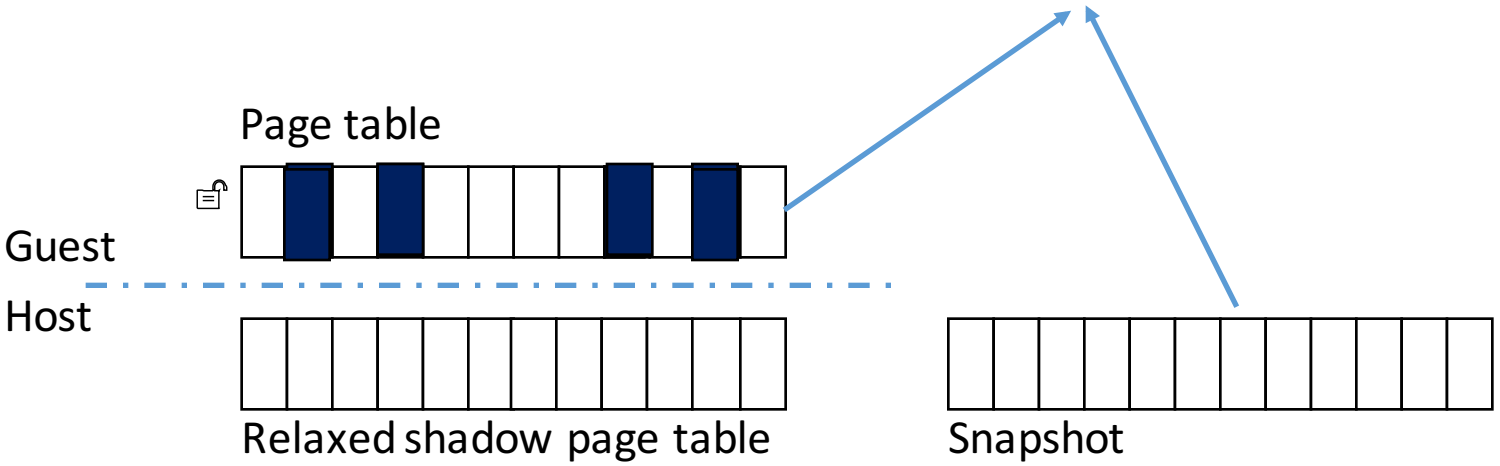
Page table reconstruction

Step2: Massive update



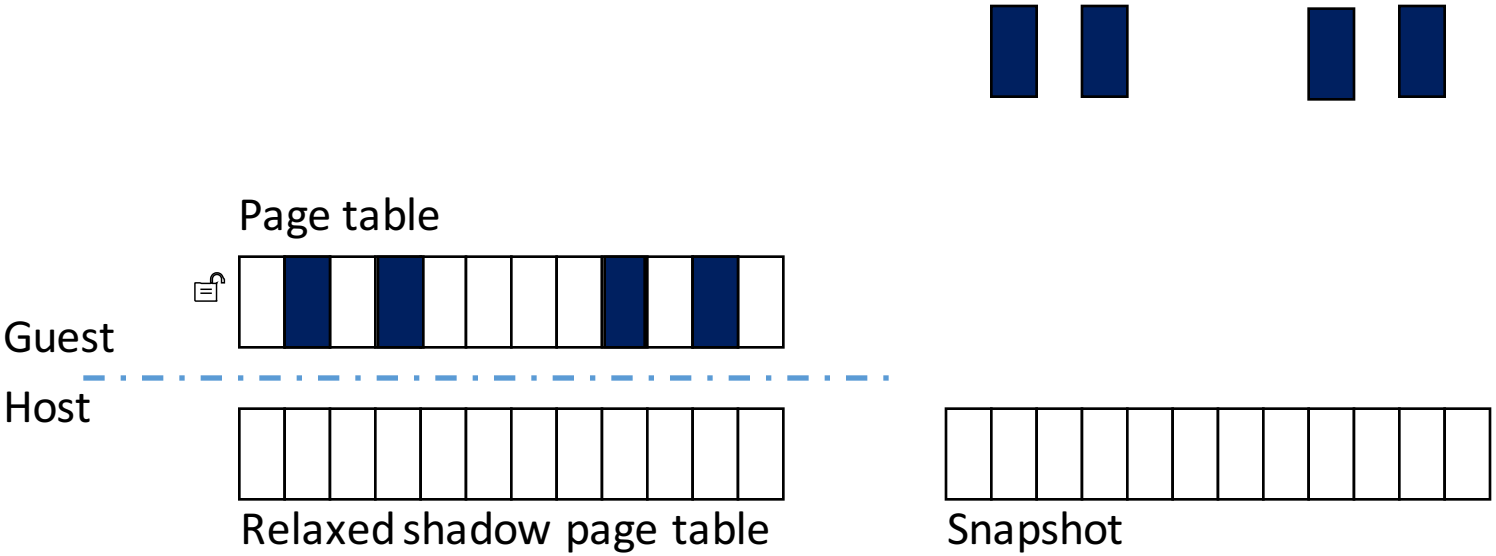
Page table reconstruction

Step3: Compare with snapshot

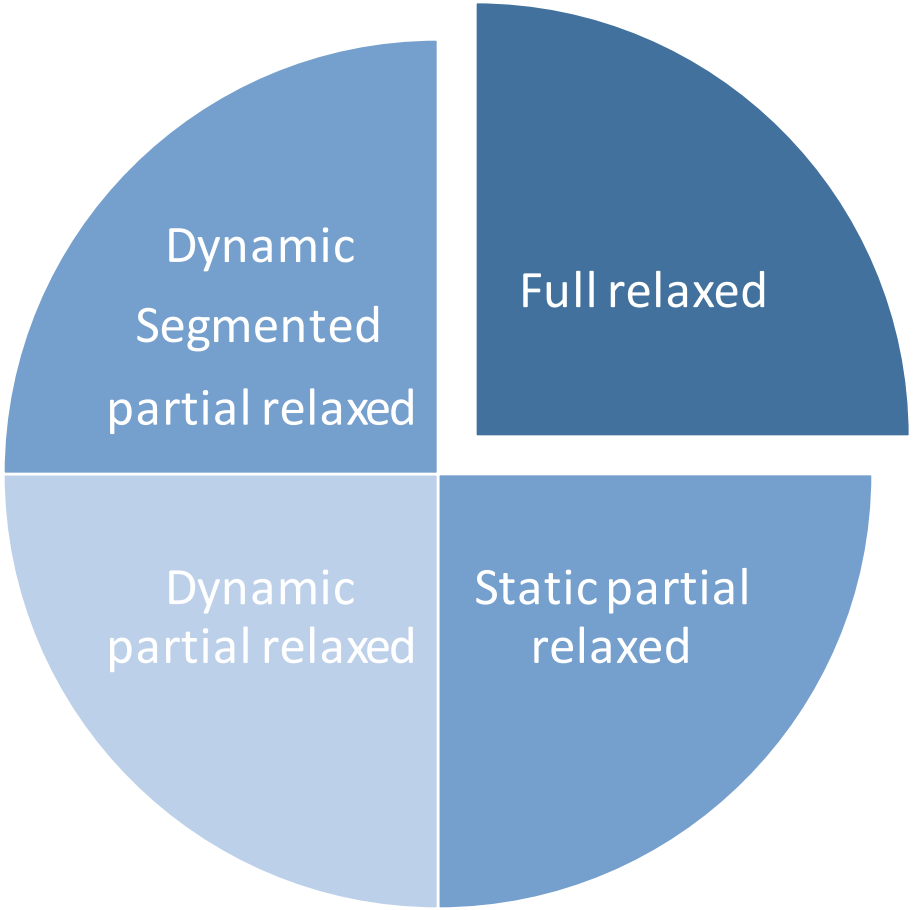


Page table reconstruction

Step 4: Reconstruct the different part



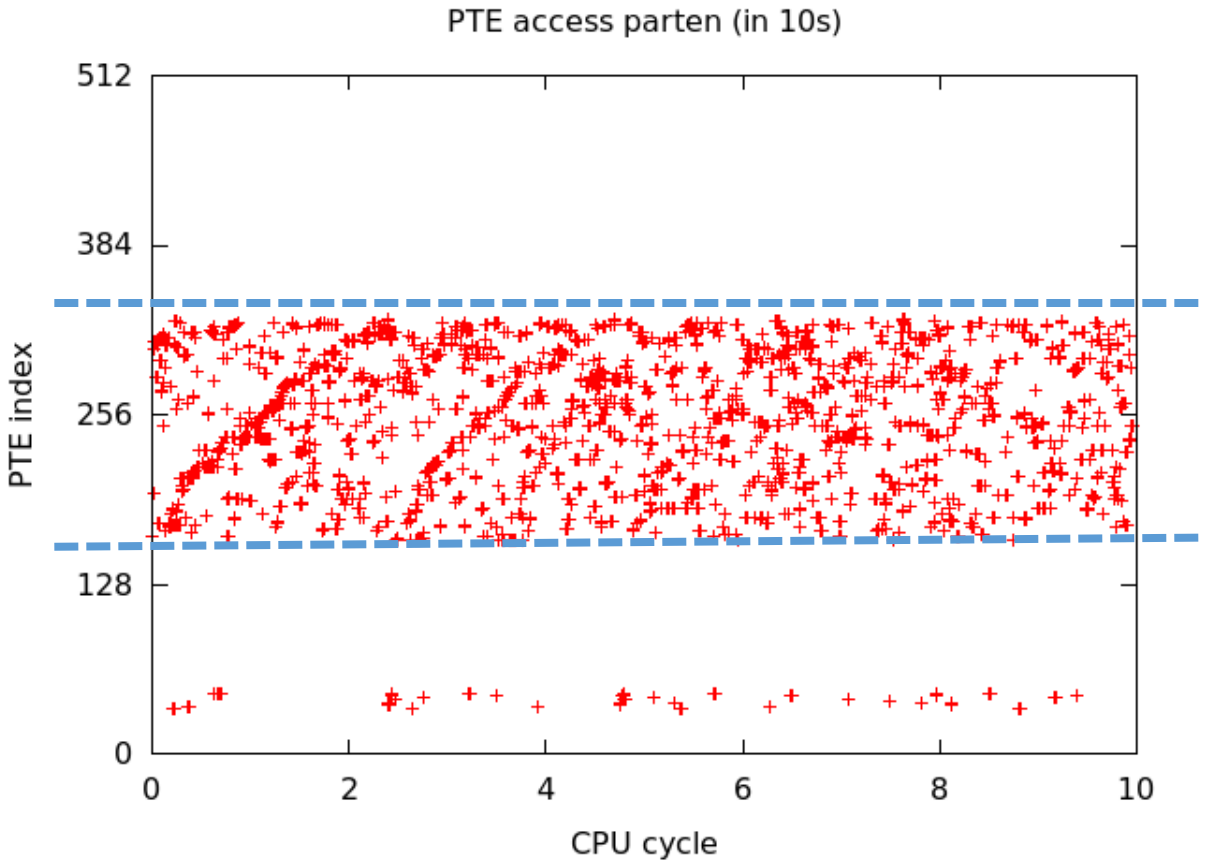
Hybrid page table shadowing



4 Policies for Performance Tuning

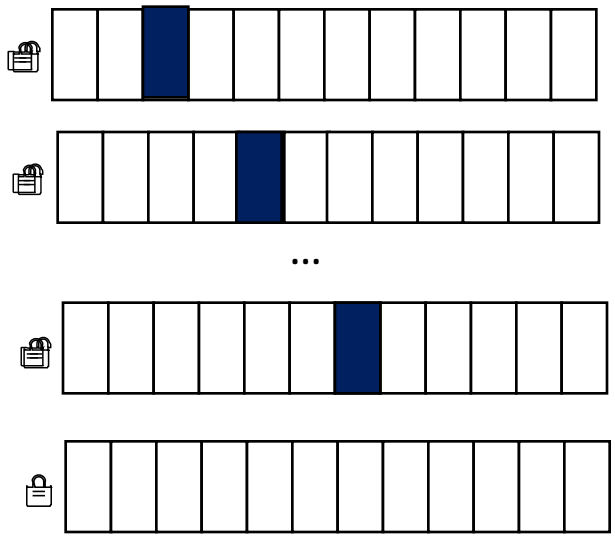
Static partial relaxed

Select 50, 100, 200, 300 hot pages
Switch them into relaxed mode



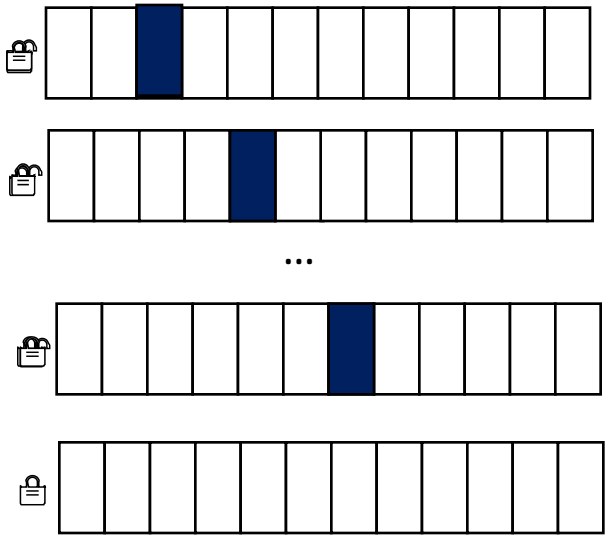
Dynamic partial relaxed

- All the pages are in strict mode at first
- Switch to relaxed mode once it's touched



Dynamic segmented partial relaxed

- All the pages are in strict mode at first
- Switch to relaxed mode once it's touched
- Reset the relaxed pages after reconstruction



Evaluation

4 policies of page table shadowing

- Full reconstruction
- Static partial reconstruction
- Dynamic partial reconstruction
- Dynamic segmented partial reconstruction

Linux 2D/3D performance

Windows 2D/3D performance

Configuration

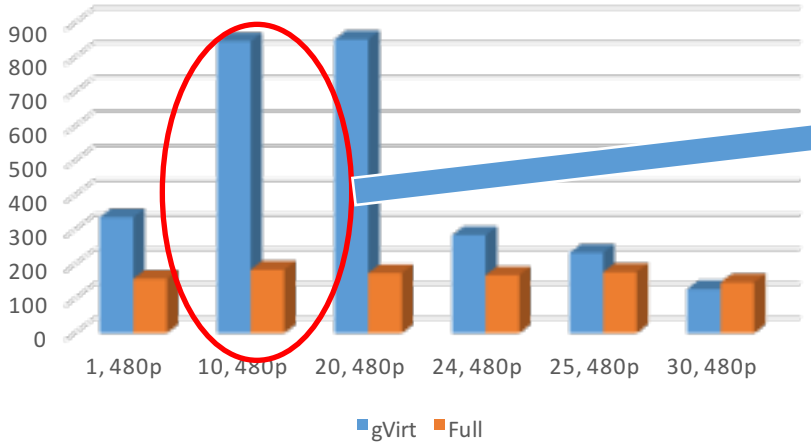
Hardware

CPU	4th Intel Haswell i5 (2.4Ghz)
RAM	8GB
HDD	Seagate 500GB
GPU	Intel Processor Graphics With 2GB global graphics memory

Software

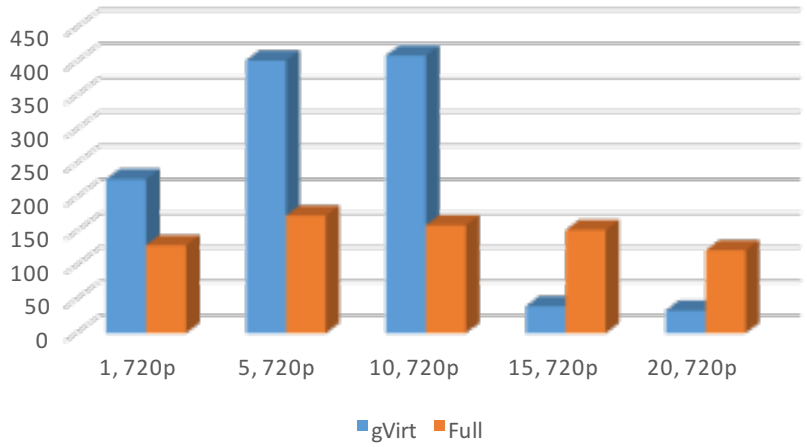
Linux VM	64bit Ubuntu 12.04
Windows VM	64bit Windows 7
Xen	4.3
VM configuration	4 VCPUS and 2GB system memory

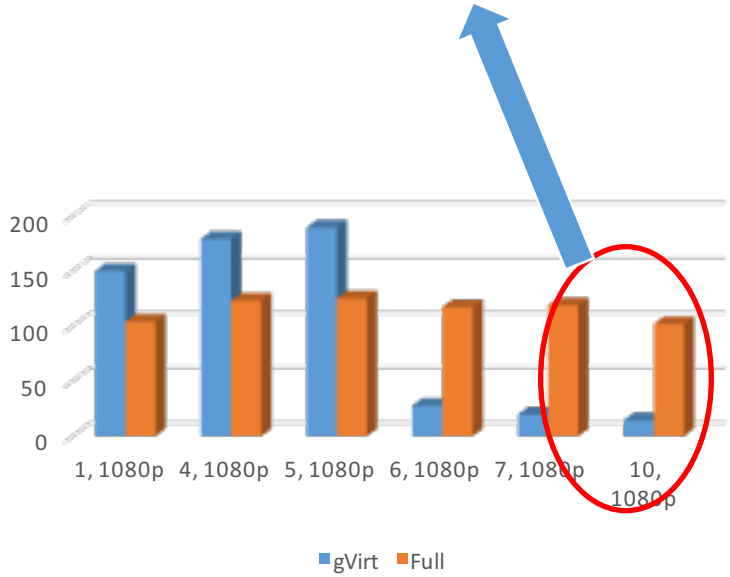
Evaluation: full relaxed



The performance of normal cases are degraded.

The performance of issue cases are improved.

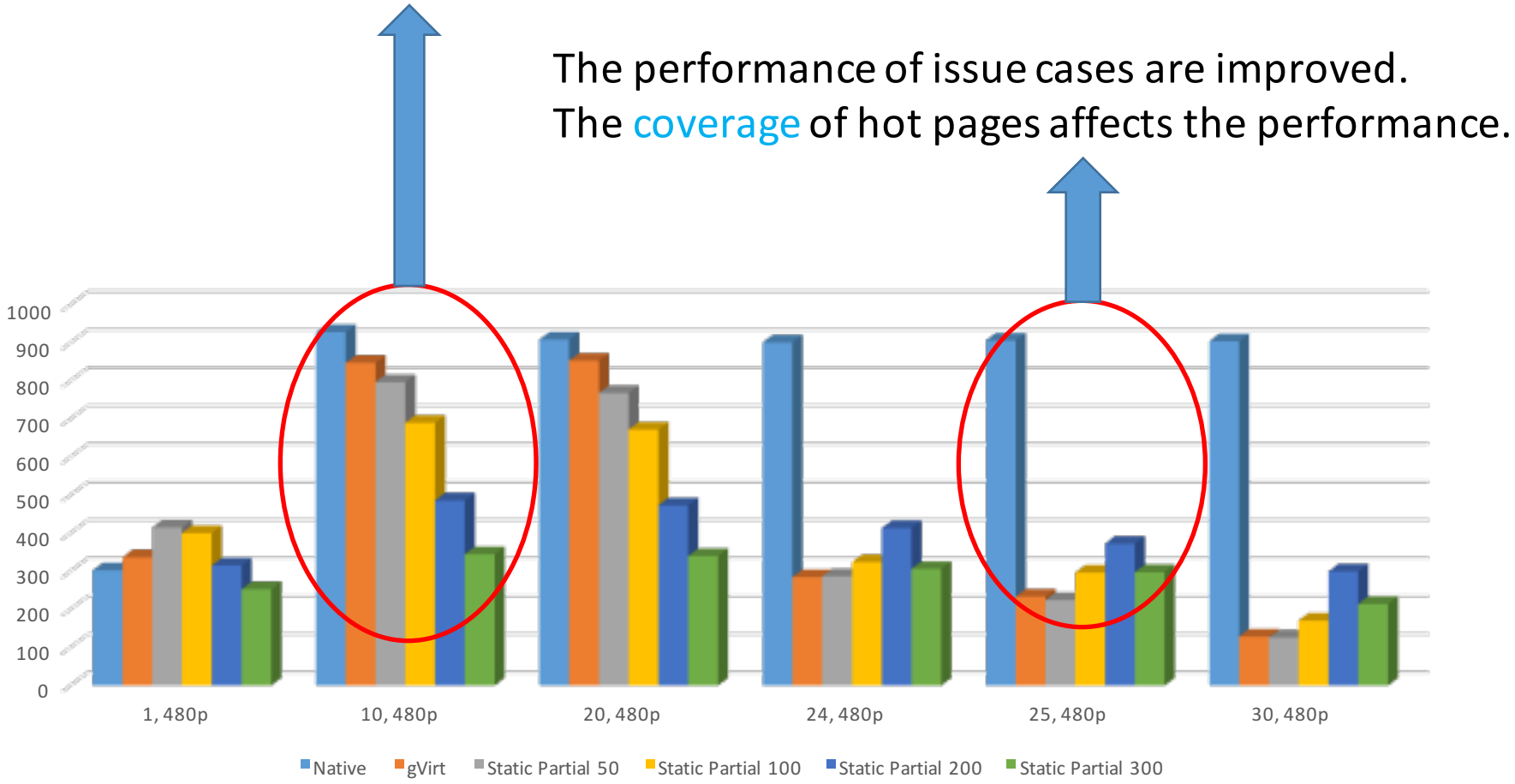




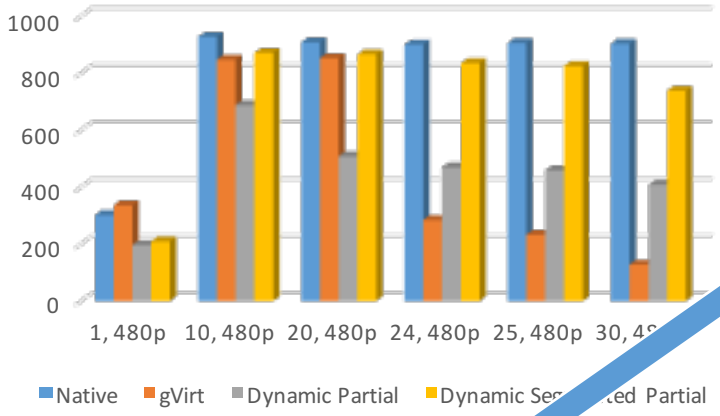
Evaluation: static partial relaxed

The performance of normal cases
becomes worse with more relaxed pages.

The performance of issue cases are improved.
The coverage of hot pages affects the performance.

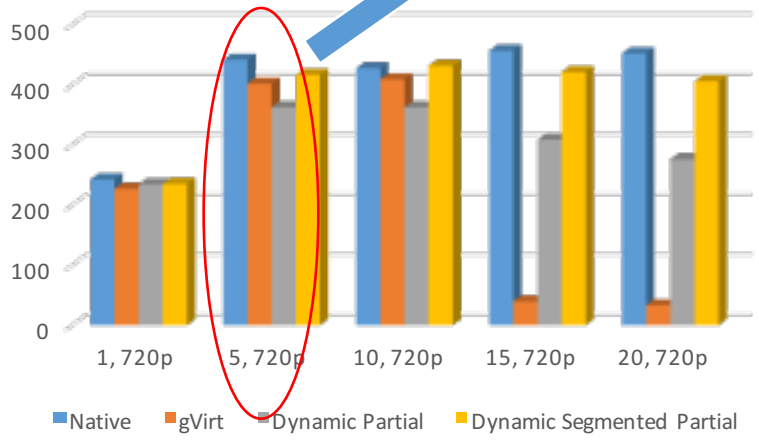


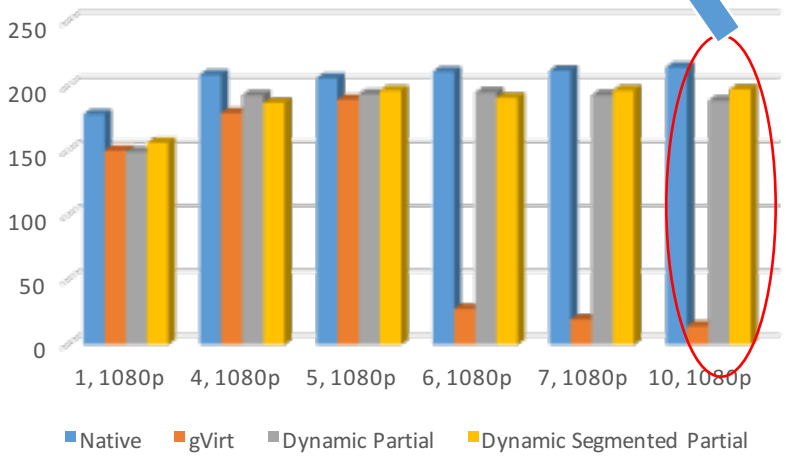
Evaluation: dynamic relaxed



Dynamic segmented partial works fine on normal cases.

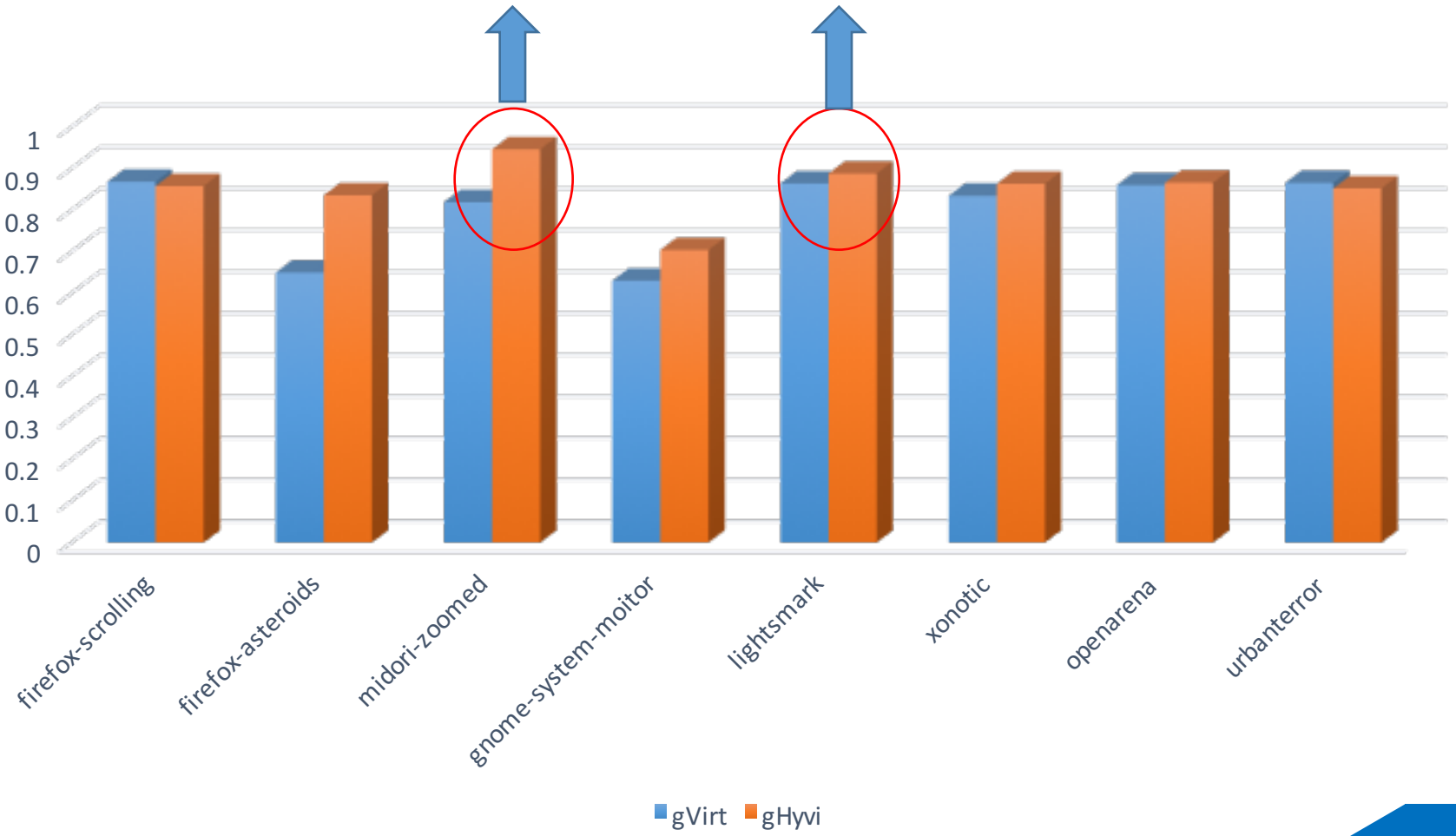
Up to 13x of gVirt
85% of native



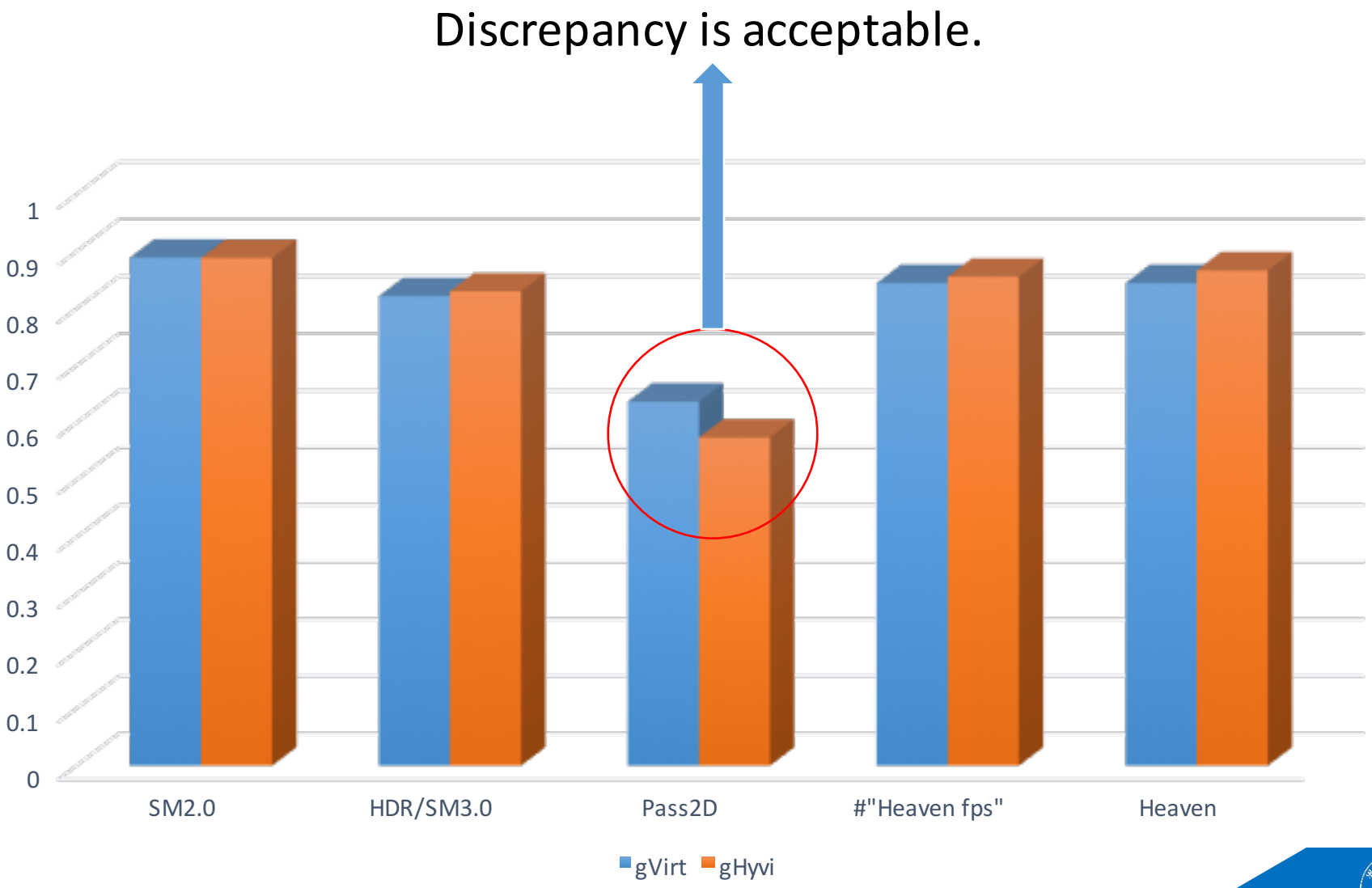


Linux 2D/3D performance

Slightly better than gVirt.



Windows 2D/3D performance



Summary

An optimized GPU virtualization scheme based on gVirt.

- New shadow page table: relaxed shadow page table
- Adaptive hybrid page table shadowing policies

Up to 13x performance improvement.

Source code is available

https://01.org/igvt-g

Q&A

Boosting GPU Virtualization Performance with Hybrid Shadow Page Tables