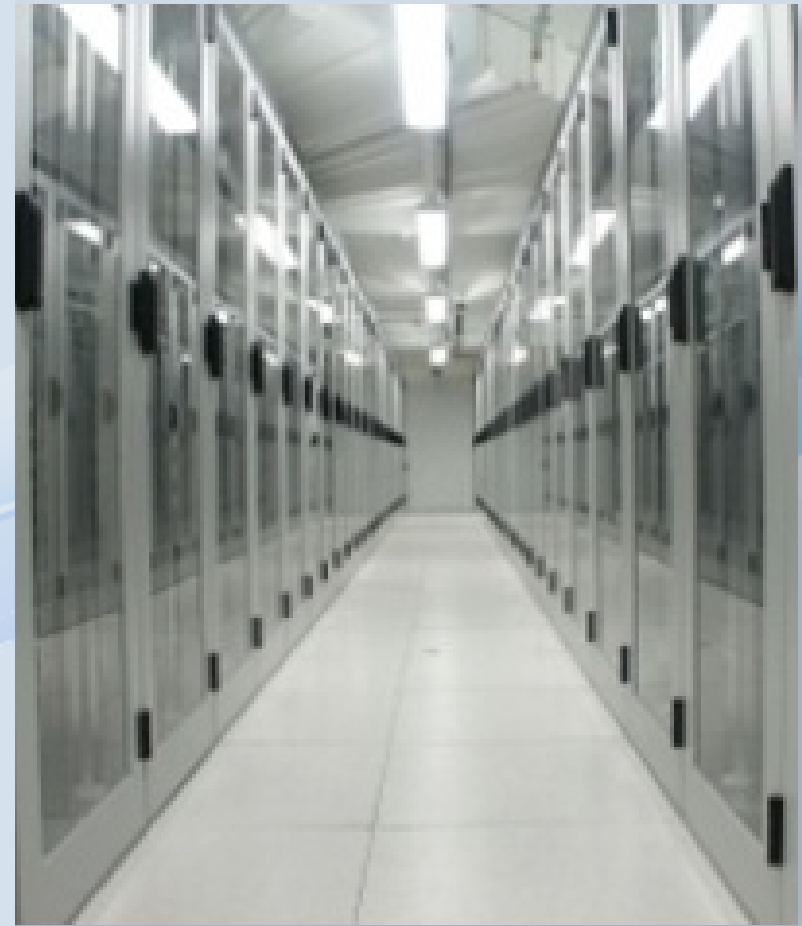


blkreplay: Experiences with Commercial vs OpenSource Storage Systems



LCA2013 presentation by Thomas Schöbel-Theuer

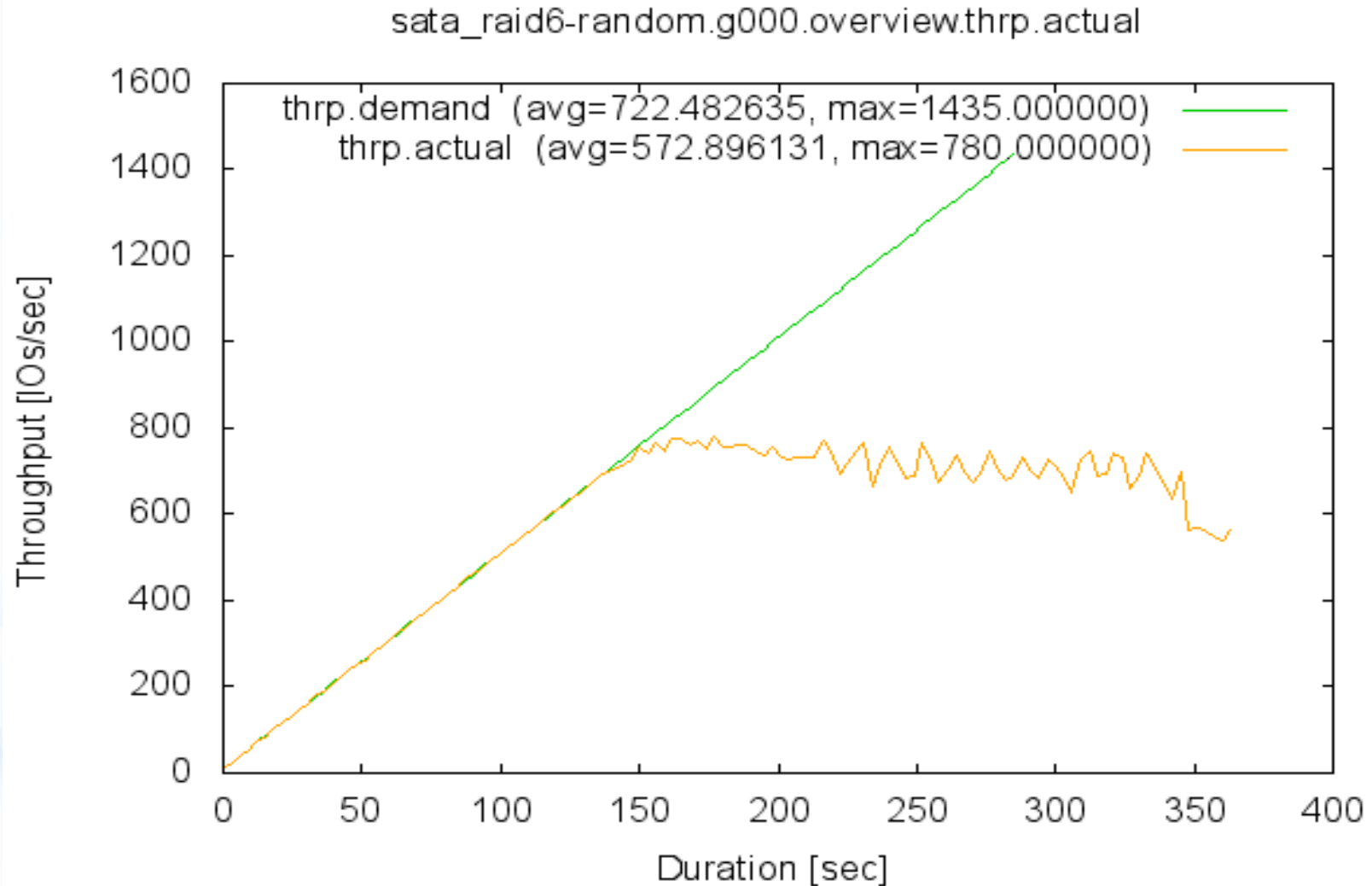
- `blkreplay` Features
- Why Artificial Benchmarks suck
 - Example: random-sweep comparison
- `blkreplay`: Real-Life Performance
 - Example continued
- Pitfall: EMPTY vs FILLED
- Chances for OSS



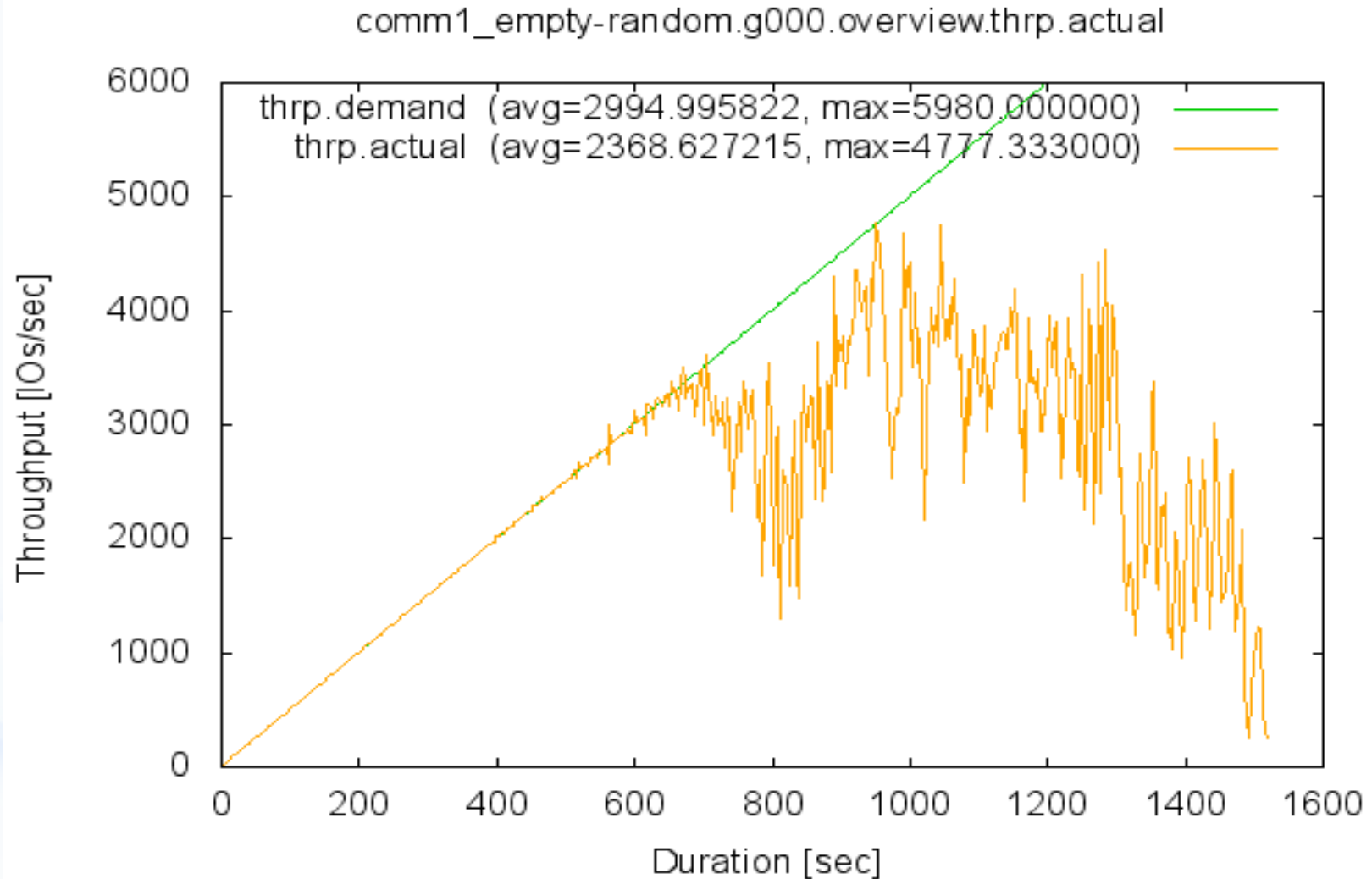
- Reproduction of both artificial and **natural loads** (block level)
 - Positionly behaviour
 - Timely behaviour
 - IO parallelism
- **Test suite** for **automation** of large benchmarking projects / stress-testing, etc
 - extensible with plugins
- Large **database** (>70GB) with natural loads from 1&1 datacenters on `blkreplay.org`
 - contributions welcome!



Example 1a: random sweep on Linux SATA RAID-6



Example 1b: random sweep on Commercial Box



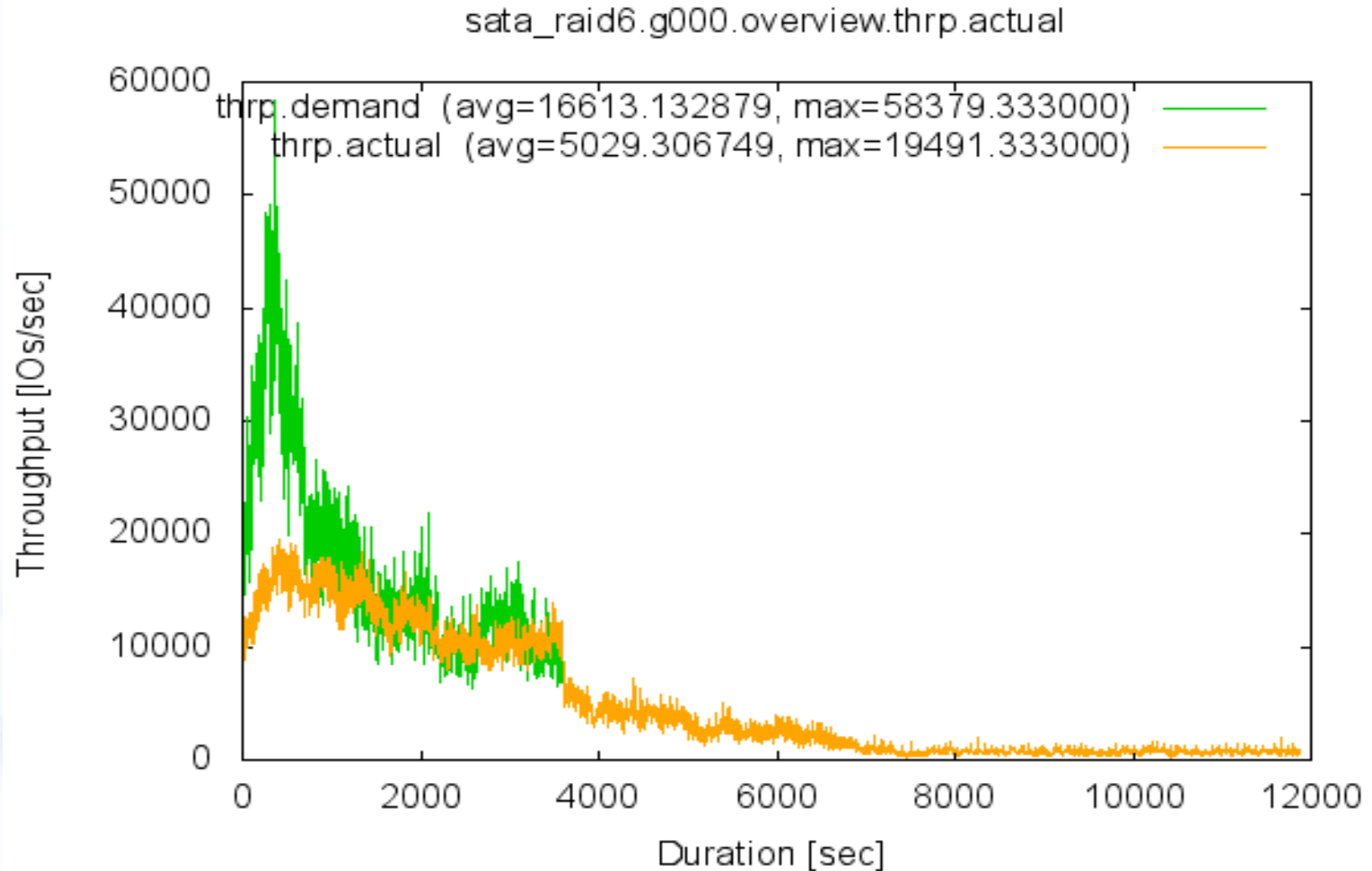
Who is *really* the winner?

- Artificial random IO can be **extremely different** from real life
- Alternative: use `blkreplay.org`
 - Record your **real** application behaviour with `blktrace`
 - Or, use a published real-life load from `blkreplay.org`
 - Exactly replay your original timely and positionly behaviour, degree of IO parallelism, etc
 - Don't use AIO [bottleneck, distortions from page cache]
 - Use processes / threads
- Okay, does it make a difference?
=> next slides

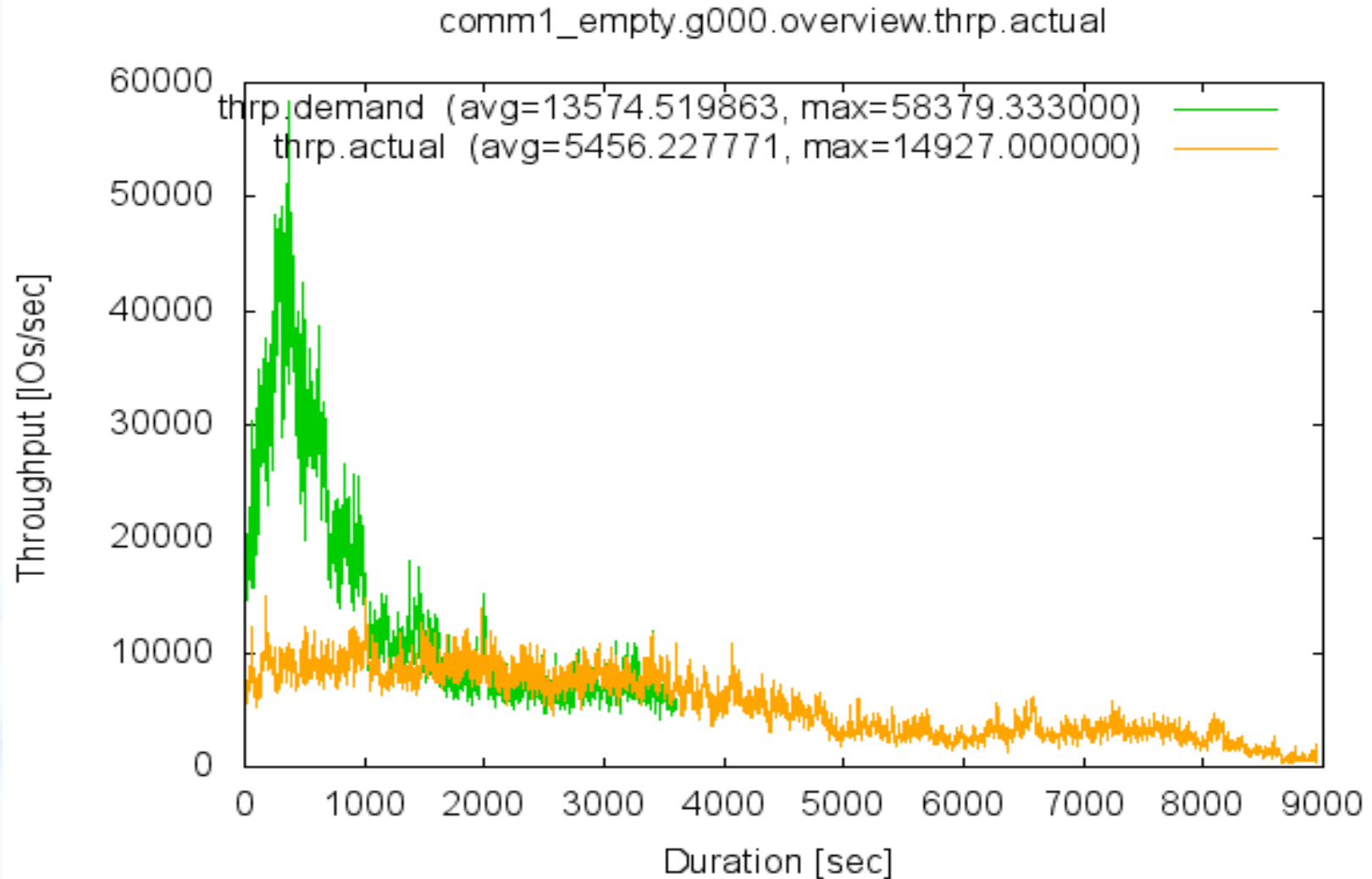


25 VMs (XenServer) in parallel, iSCSI over 10GbEth

Example 2a: real-life load on Linux SATA RAID-6

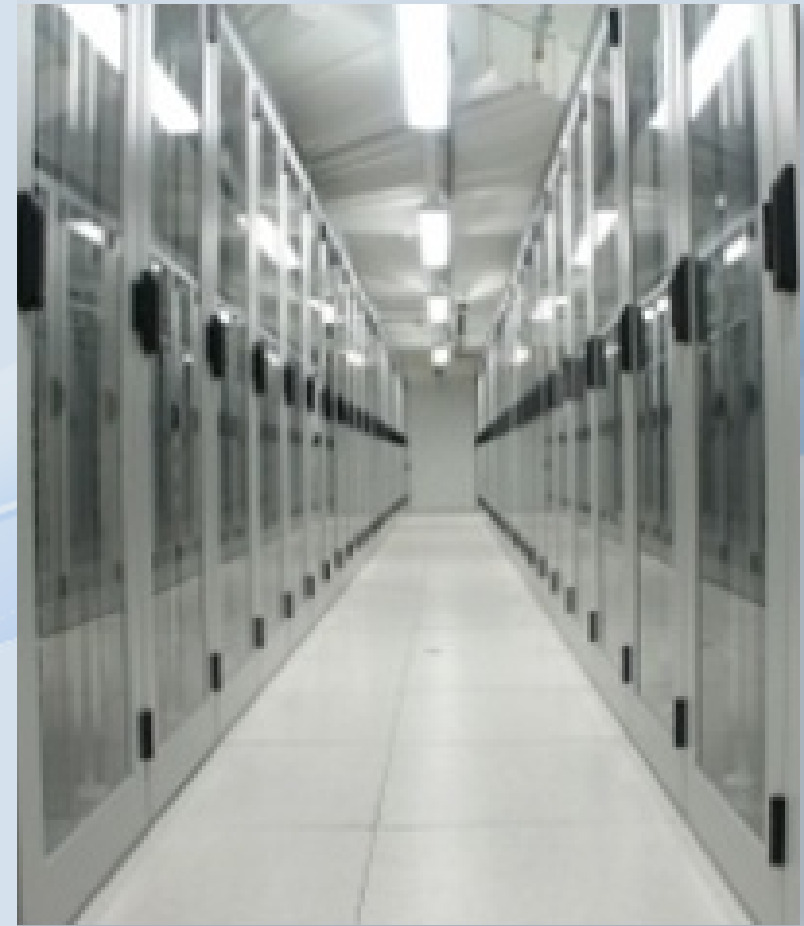
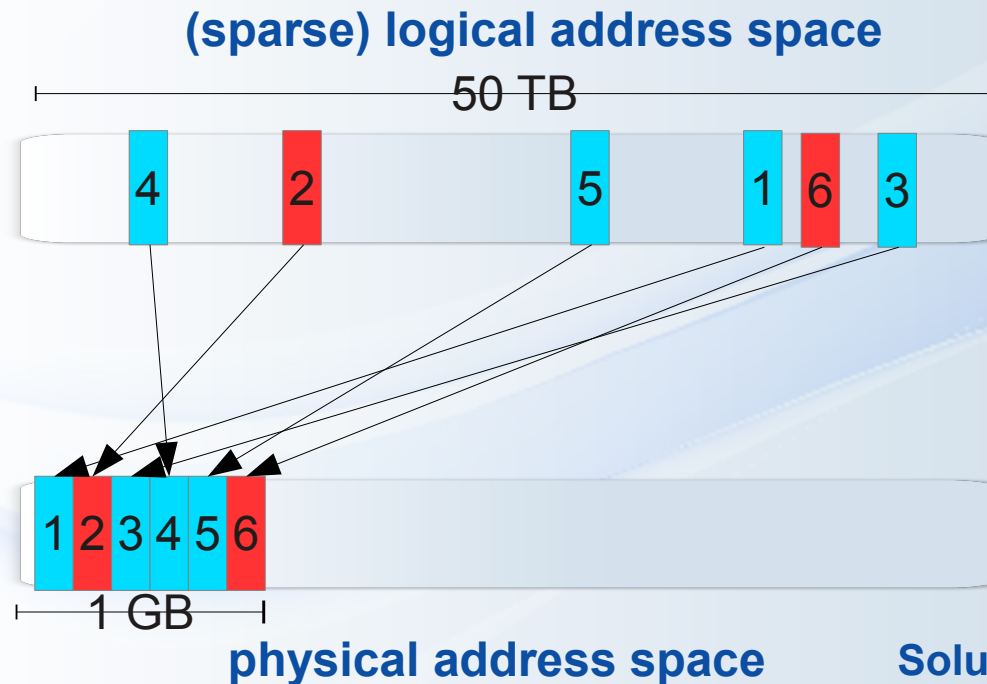


Example 2b: real-life load on EMPTY Commercial Box



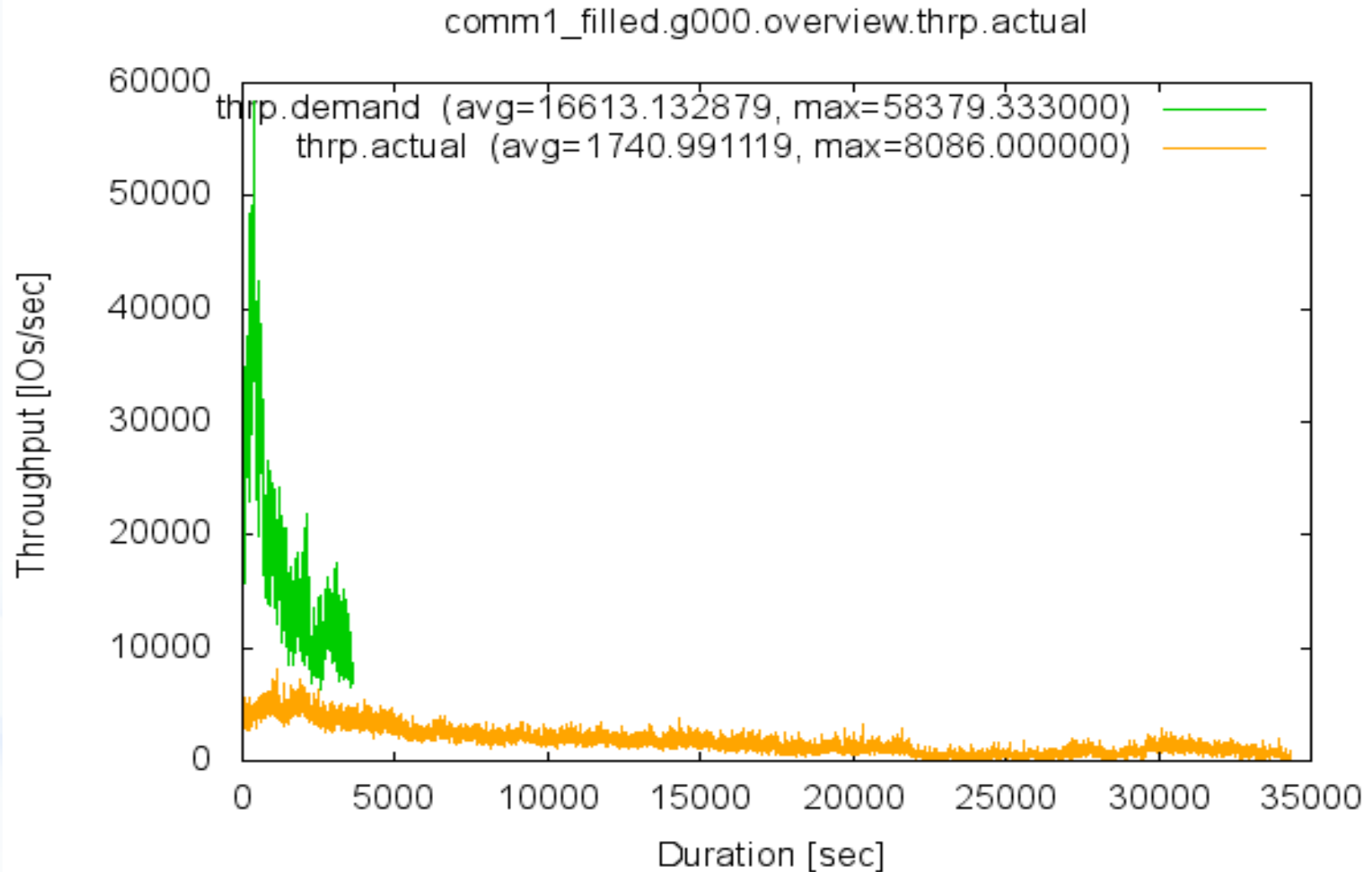
Pitfall: Filled vs Empty Logical Volumes

- Commercial black-boxes / SSDs / etc often implement **Storage Virtualization**
- Translation from **logical block addresses** to **physical block addresses**
- Problem: benchmarks touch only a **tiny fraction!**



Solution: pre-fill the whole LV with random data

Example 2c: real-life load on FILLED Commercial Box



- Mass Data: > 1 PB total
→ price/TB matters
- Admins know what they are doing
- Management often believes sales personnel from commercial storage vendors
→ find out the TRUTH
prejudices can be HARD
→ **evaluation projects**
 - ✓ **Automated** by the `blkreplay` test suite
- convince your management that OSS can do often better & cheaper



Conclusions

- Never trust *any* claim / benchmark from **sales!**
- Always check yourself, e.g. with natural loads from **blkreplay.org**
- OSS Performance often better
- OSS Price / Performance even more often better

