



# **HEP Application Behavioral Modeling under Multidimensional Resource Restrictions**

*System Performance Modeling WG Meeting*

Riccardo Maganza

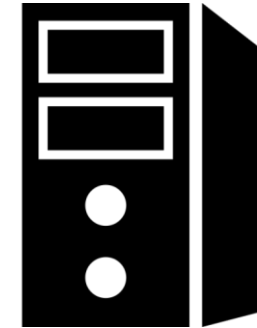
07/08/2019

# CONTEXT

- **Corentin Bugnot's 2018 studies on reference HEP application's behavior under emulated resource conditions**
  - **Focus on overall performance (runtime)**
- **PRmon detailed behavior of core metrics over time**
- **Now: Combining the two approaches to get input for modelling the behavior of individual metrics under resource restrictions**
- **In addition parametrization of PRmon timeseries as series of linear approximations**

# TEST SETUP

- **64 GB DDR4 Memory**
- **Dual Socket, 8 Core, 2HT**
- **Total: 32 logical cpus**
- **Intel PCIe SSDs for local disk IO**



## Reference workloads:

- **ATLAS: Single job, 16 Processes**
  1. **Monte Carlo Generation : Memory**
  2. **DigiReco : Memory, Bandwidth, Latency**
  3. **Deriv + Prod : Memory, Bandwidth, Latency**
- **CMS: Single job, 16 Threads**
  1. **Monte Carlo Generation: Latency**
  2. **Digi + Trigg + PileUp Sim: Memory, Latency**
  3. **Reco + Analysis: Memory, Latency**
- **Resource restriction range**
  - **Network bandwidth (1 to 1250 MBps)**
  - **Network latency (0.1 to 64 ms)**
  - **Memory (4 to 64 GB)**

# OUTPUT DESCRIPTION

- **Each plot represents a job restricted in one of the mentioned resources**
- **Each color represents a different resource limit**
- **Different resource measurements:**
  - **Memory**
    - **PSS**
    - **RSS**
    - **Swap**
    - **Vmem**
  - **Application**
    - **rchar**
    - **wchar**
  - **Network**
    - **rx\_bytes**
    - **tx\_bytes**
    - **rx\_packets**
    - **tx\_packets**
  - **Disk**
    - **read\_bytes**
    - **write\_bytes**

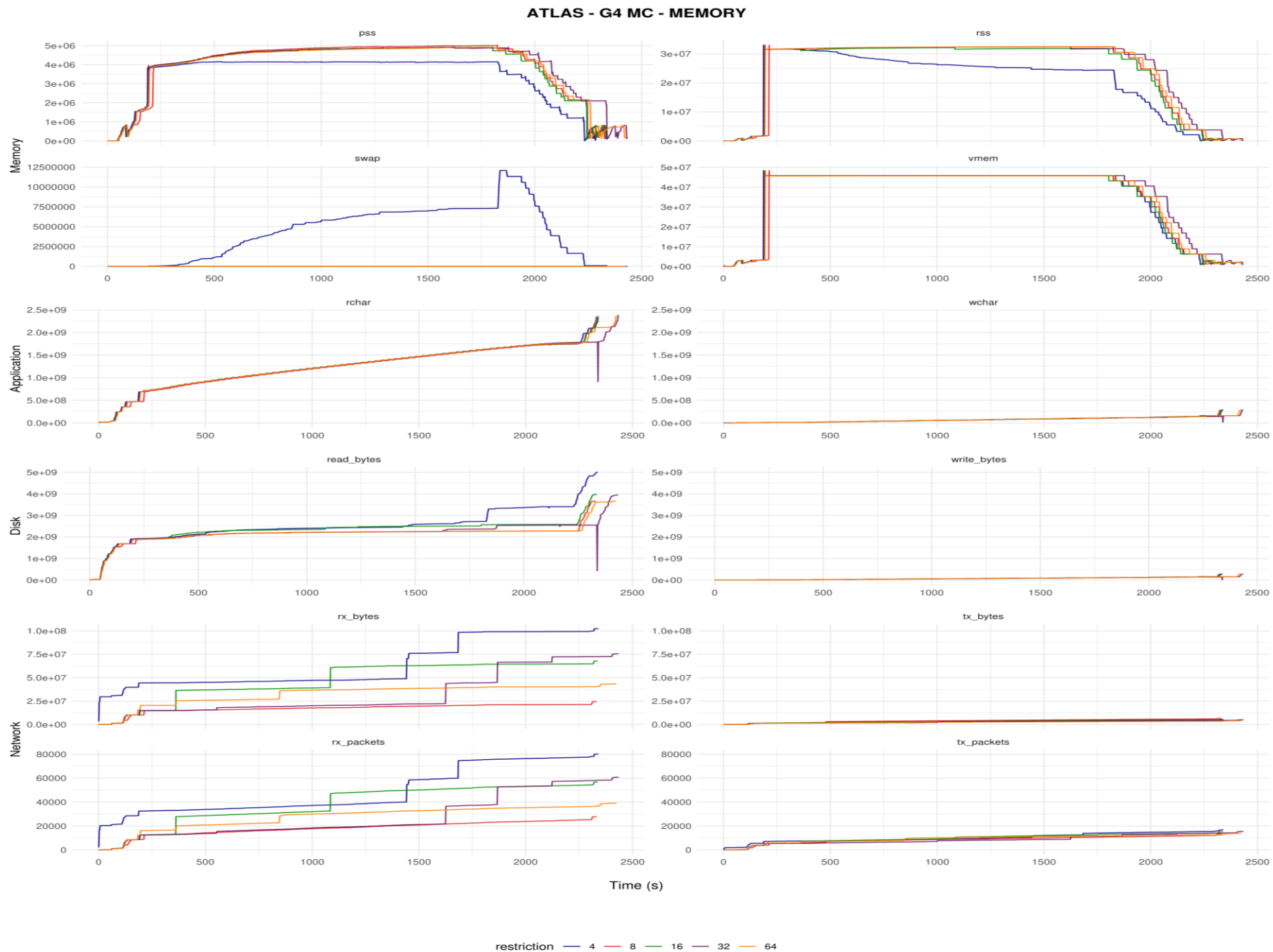


# ATLAS WORKLOADS



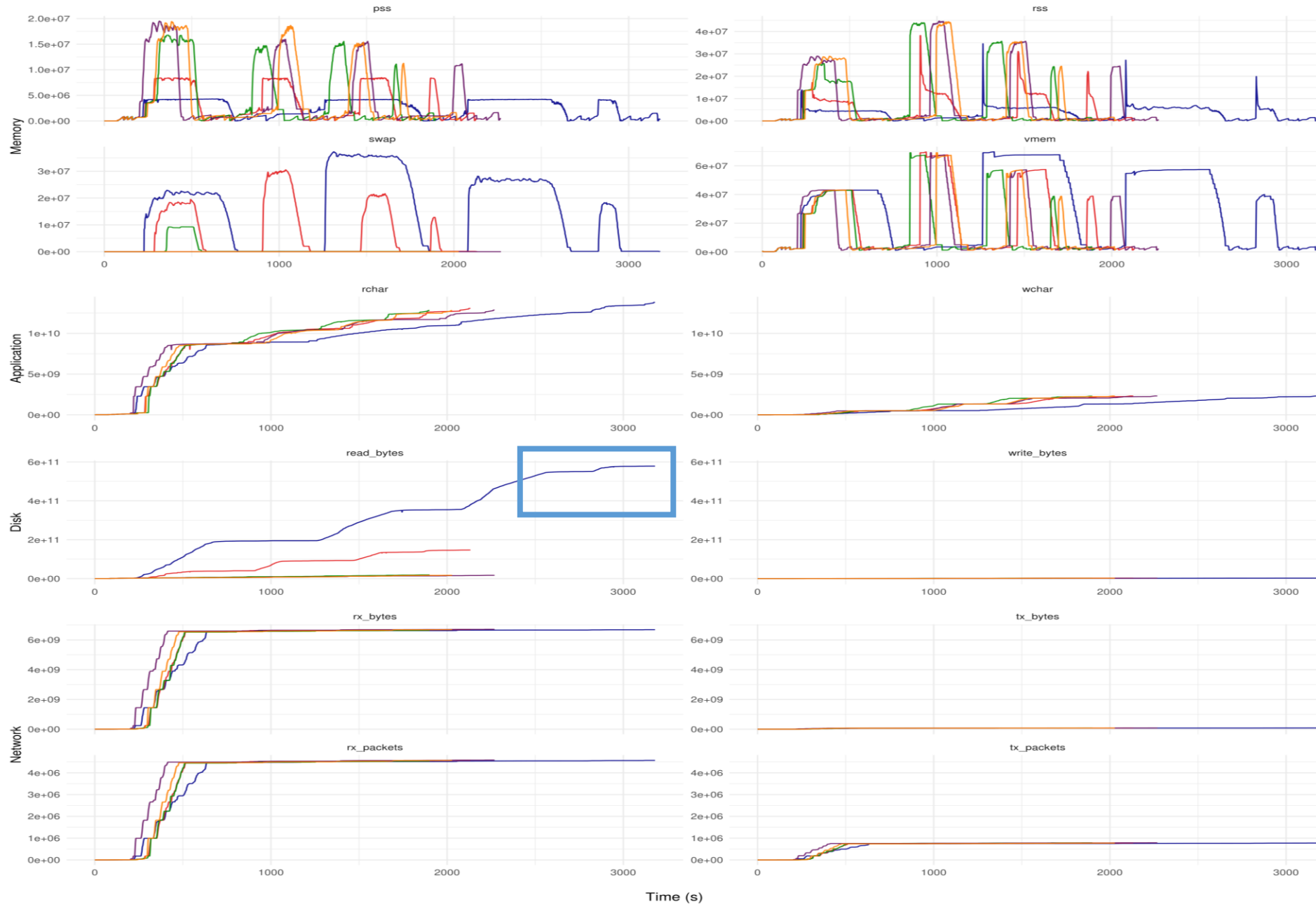
# ATLAS G4 MC: Memory

- Only after decreasing memory to 4 GB, swap is utilized
  - Job insensitive to memory limitation
- Disk and network usage increase slightly when restricting memory
  - 4GB limit causes 1GB more disk reads



# ATLAS DIGIRECO: Memory

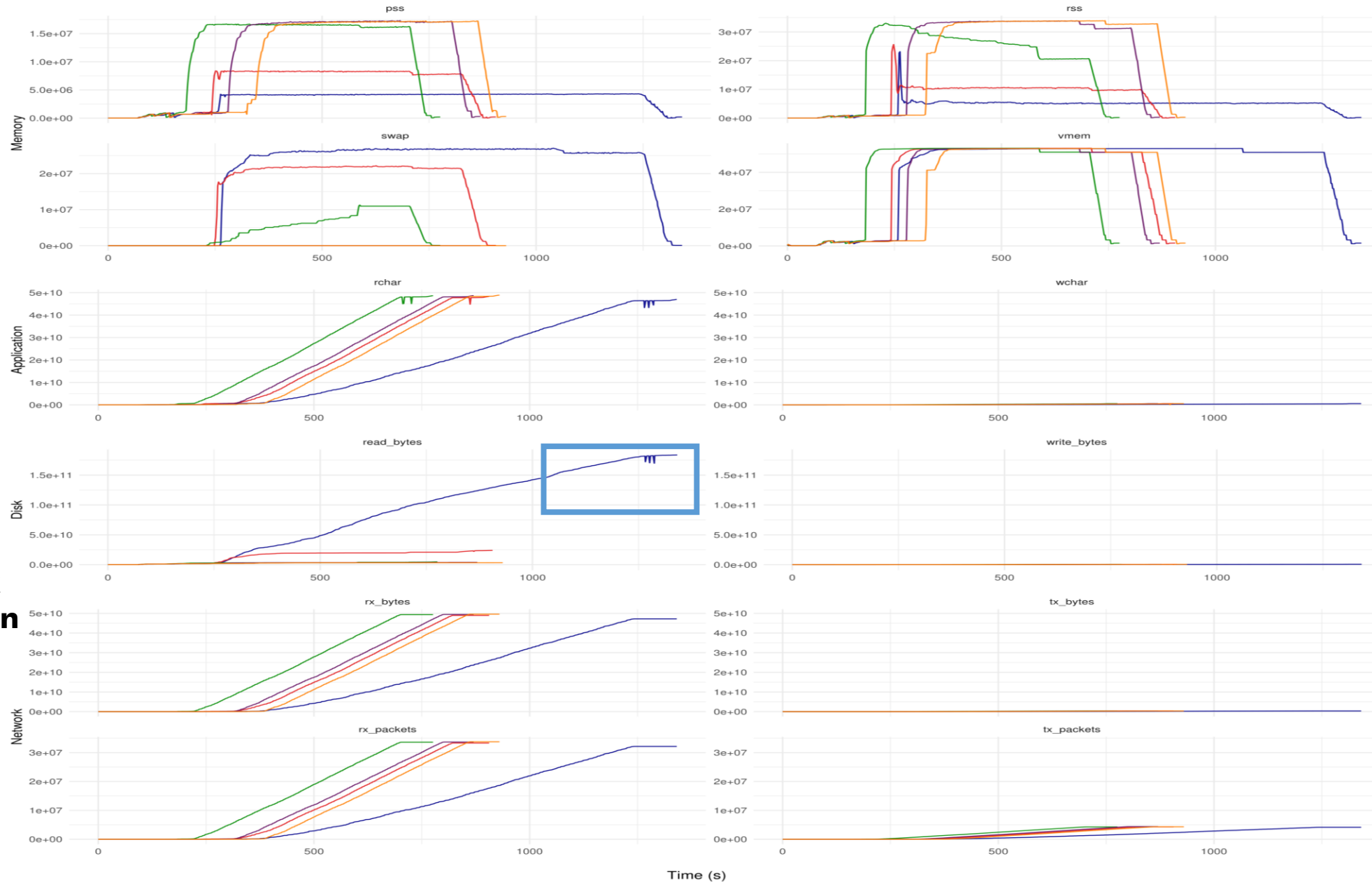
- **Swap activity starts from 16GB**
- **In the case of the 4GB restriction, the application reads 600GB of data from disk.**
- **There are very high differences with respect to disk reading between the various restrictions, while the amount of data requested from the application is roughly the same**



# ATLAS DerivProd: MEMORY

- **Swapping starts from the 16GB restriction**
- **4GB and 8GB have very similar swap behaviors, yet very different disk usage, with 4GB reading 100GB more from disk than the 8GB restriction**

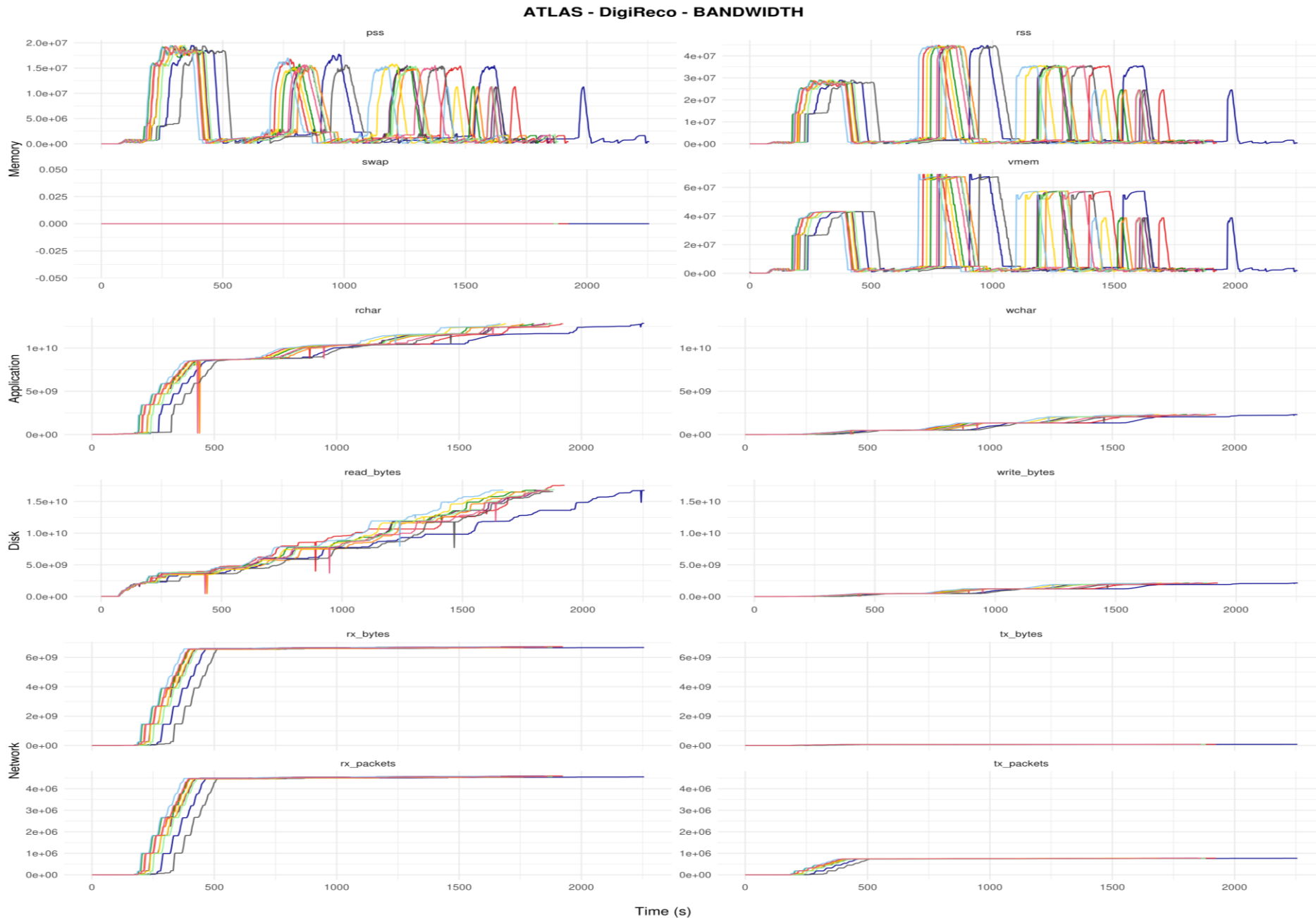
ATLAS - DerivProd - MEMORY





# ATLAS DigiReco: Bandwidth

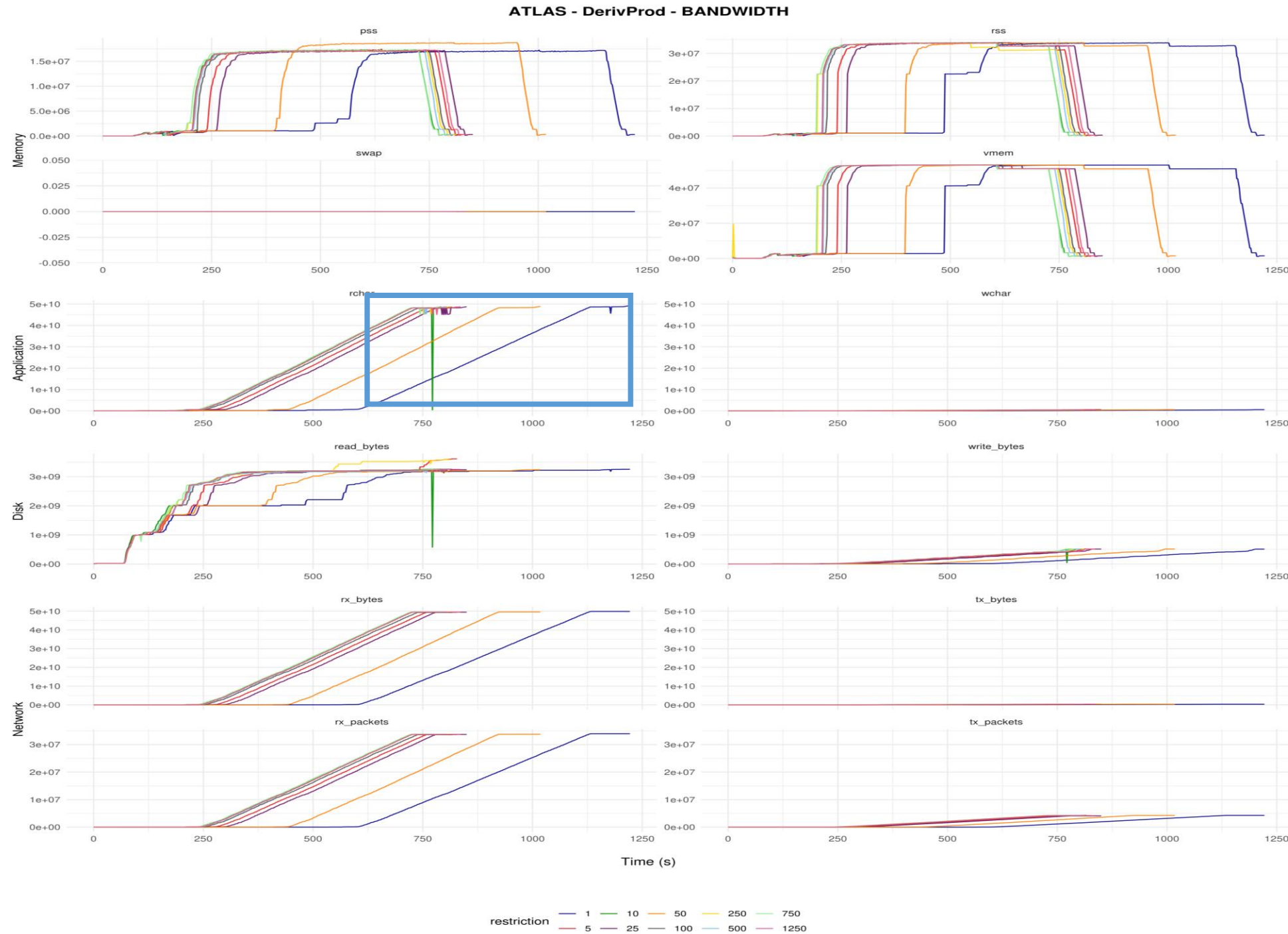
- **Plots show shift in time series as we increase bandwidth limit**
- **The workload reads from disk more than what the application asks for. e.g for 4 and 8 GB restrictions the application requests for ~12.5 GB but the job reads ~17 GB.**



# ATLAS

## DerivProd: Bandwidth

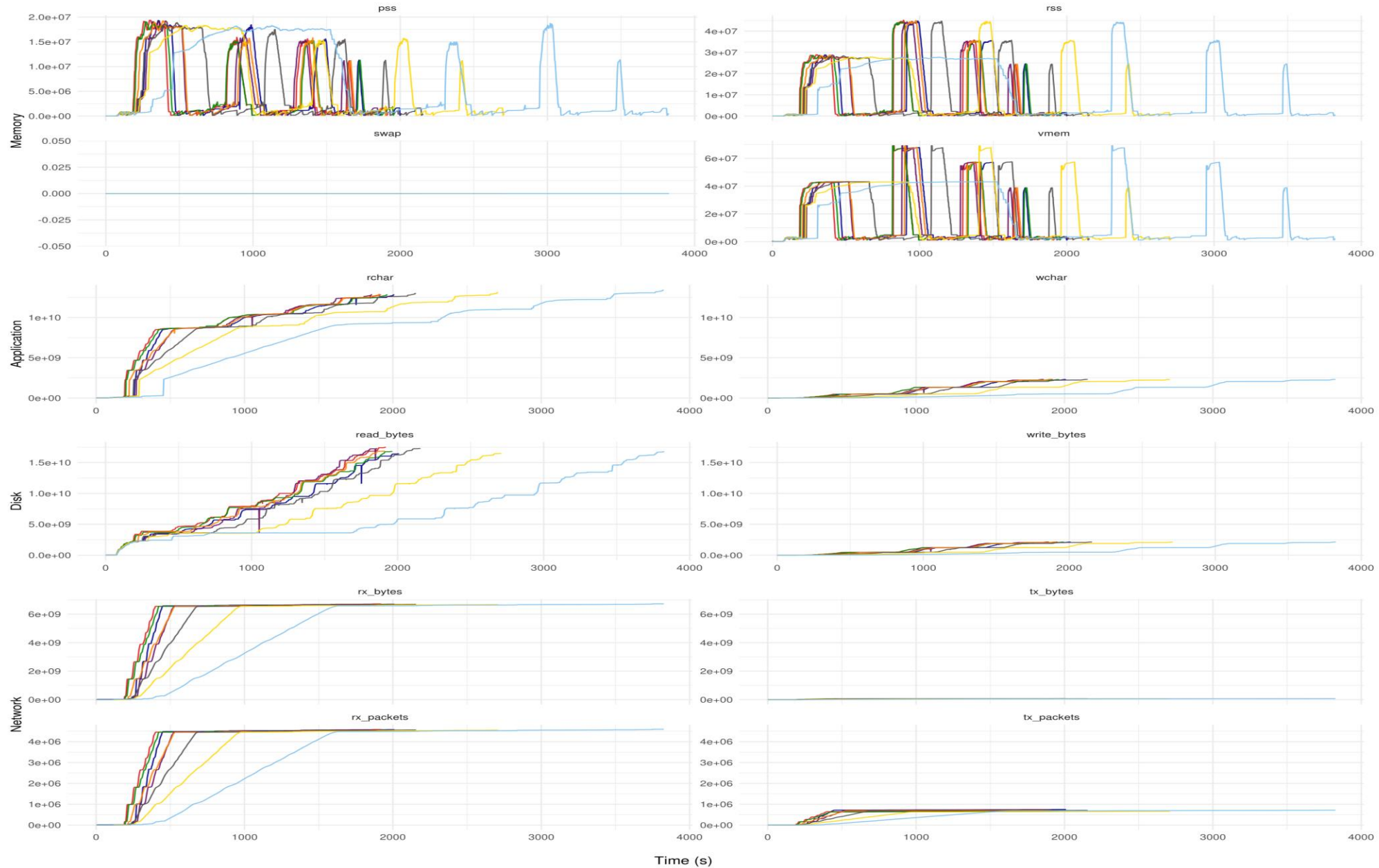
- The slopes of the curves are similar even with varying limits, while we expect lower gradients with less bandwidth
- The only effect is that IO activity seems delayed
- The time shift causes the worst case scenario to take about twice as long than the fastest run



# ATLAS DigiReco: Latency

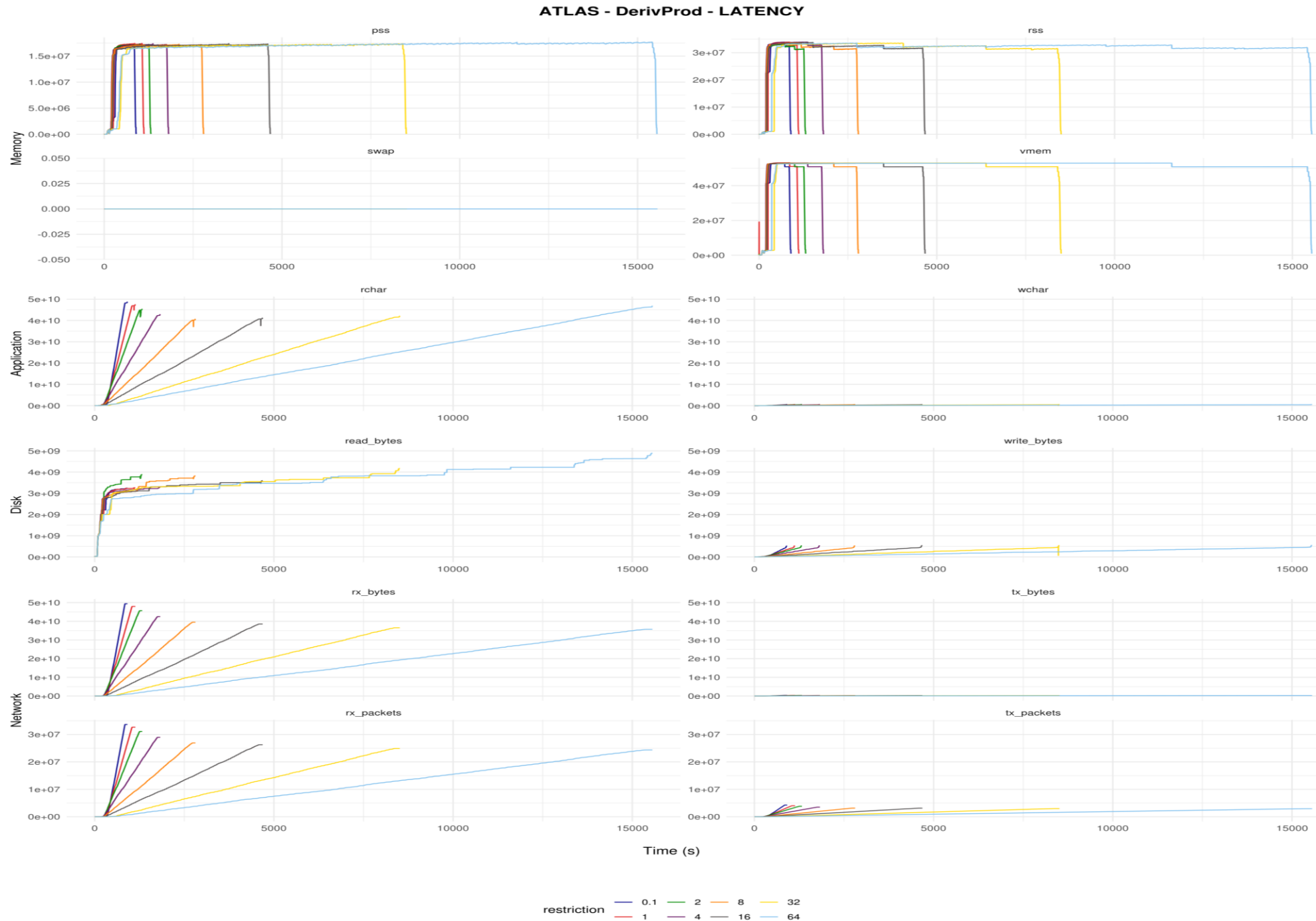
- The plots show a time shift as well, but slopes change.
- Time dilation is more evident: the best case scenario takes ~1800 s, the worst case takes ~3700s

ATLAS - DigiReco - LATENCY



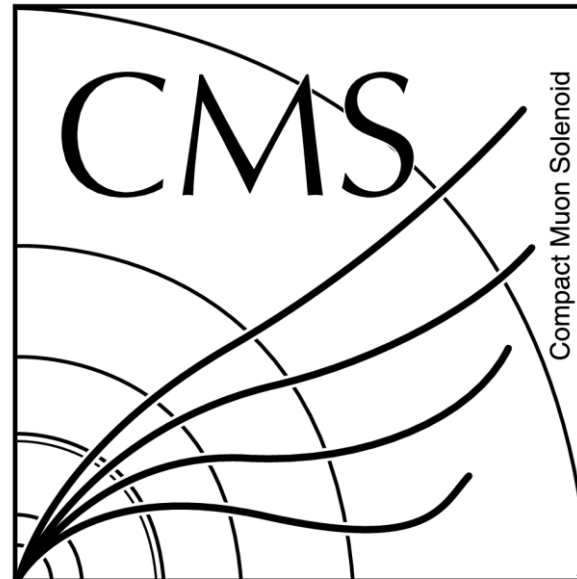
# ATLAS DerivProd: Latency

- There are noticeable changes in the slopes of the curves.
- The worst case scenario takes more than 10 times to complete with respect to the best case scenario.





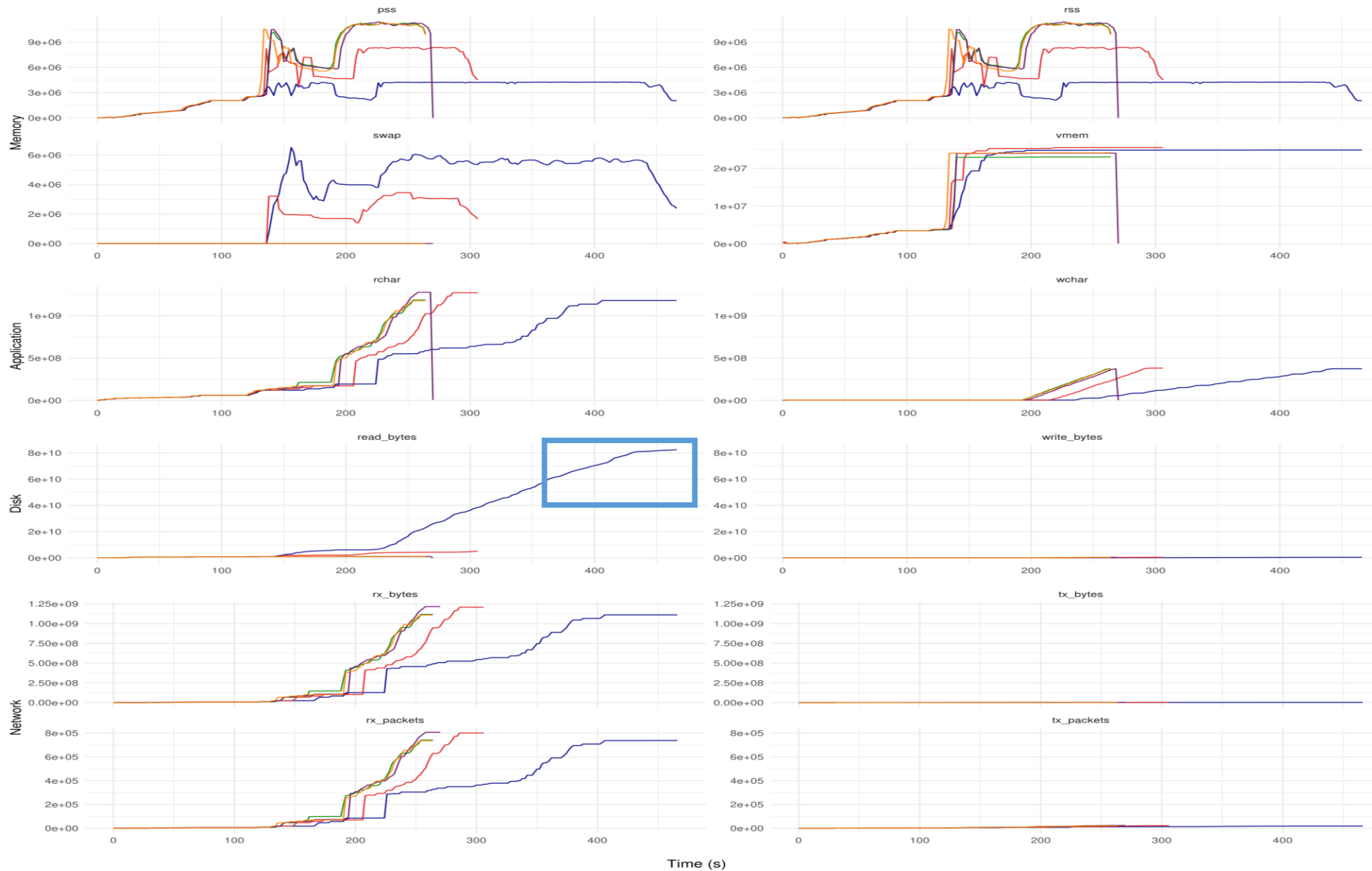
# CMS WORKLOADS



Riccardo Maganza

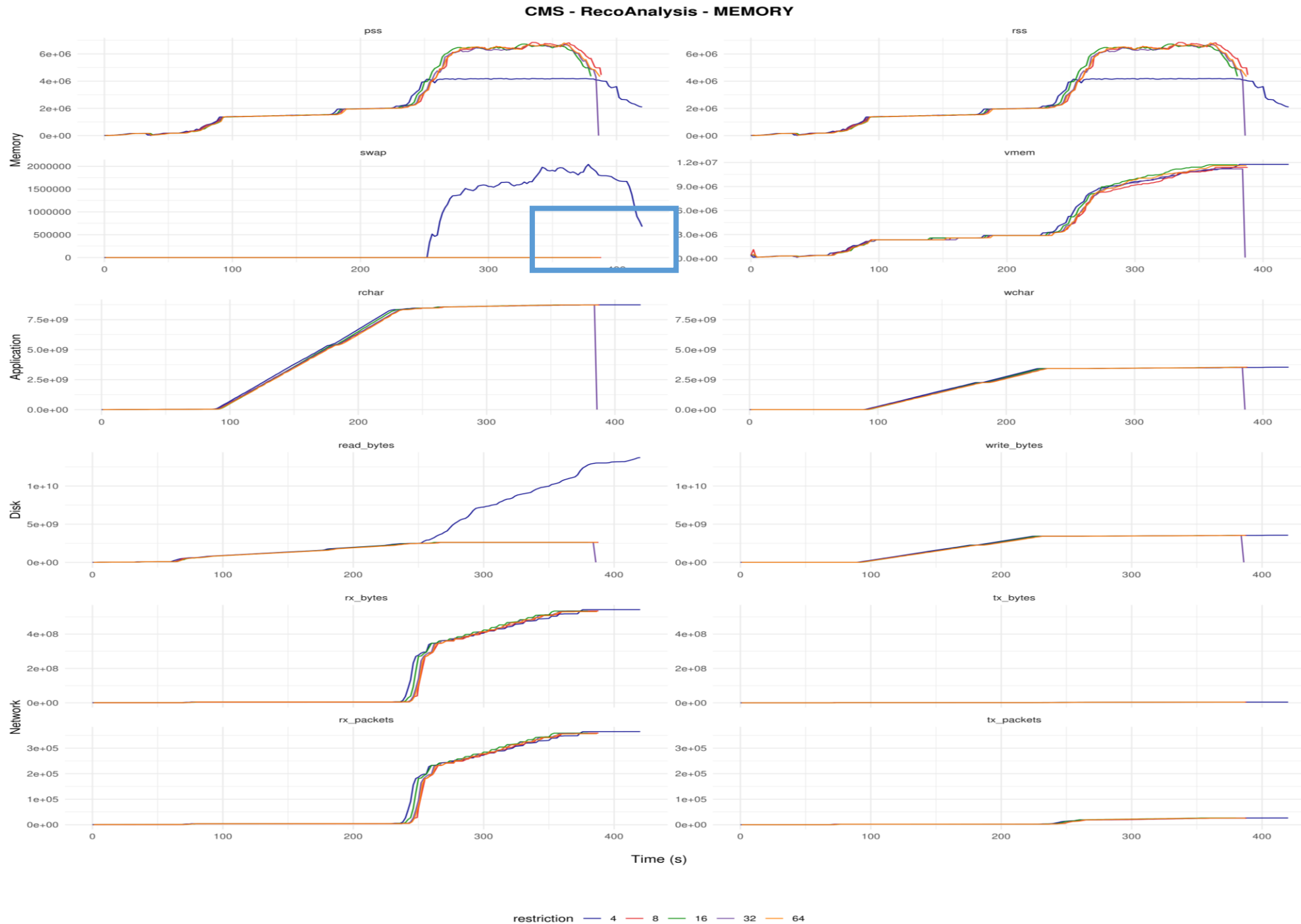
# CMS DigiTriggPile UpSim: Memory

- **4GB and 8GB restrictions show very similar swap patterns**
- **However the 4GB restriction reads 80GB of data from disk**



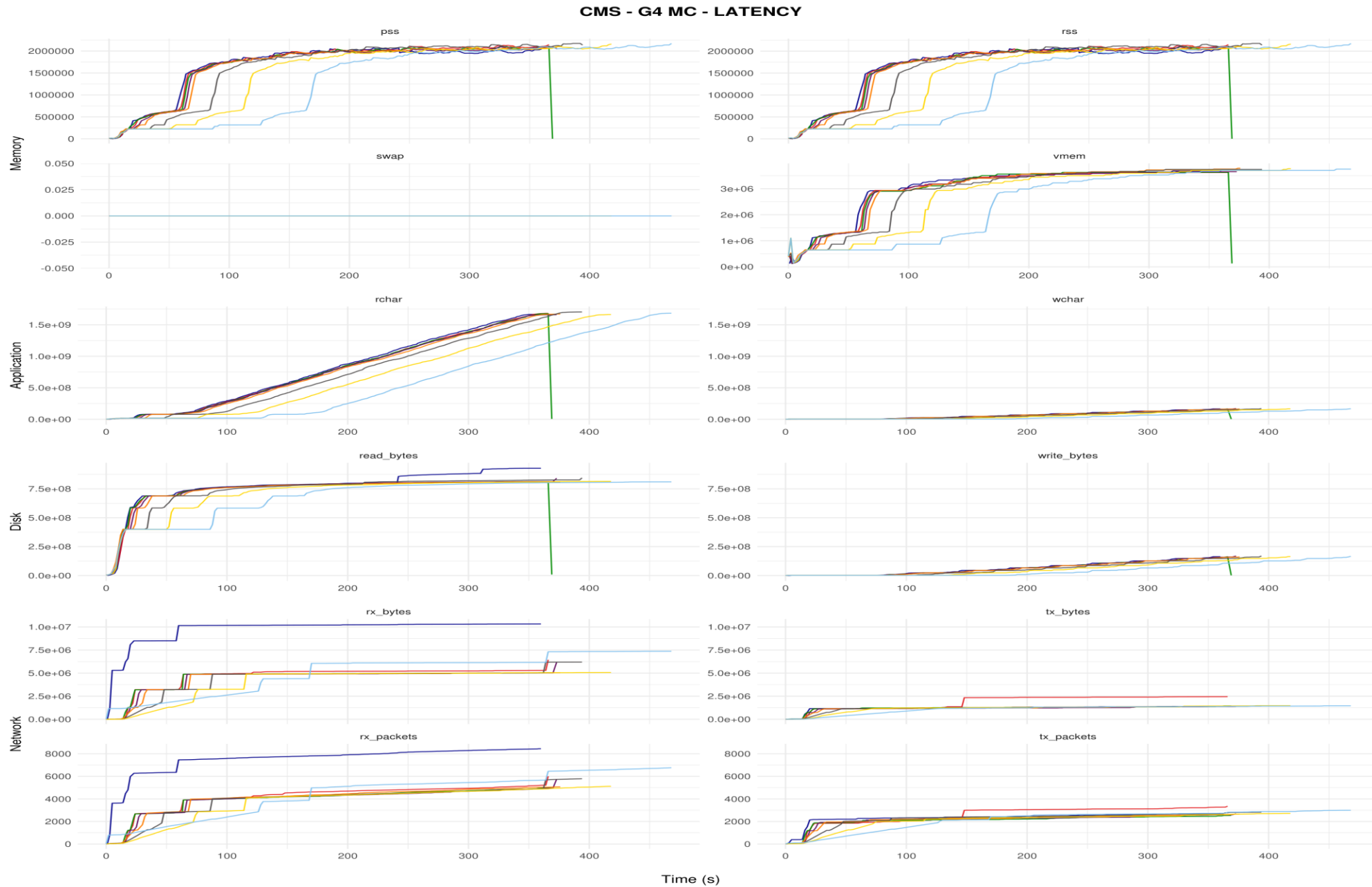
# CMS RecoAnalysis: Memory

- **Swaps only starts at 4GB**
- **There is only a ~50 second (~15%) difference between the lowest limitation and the highest one.**



# CMS G4 MC: Latency

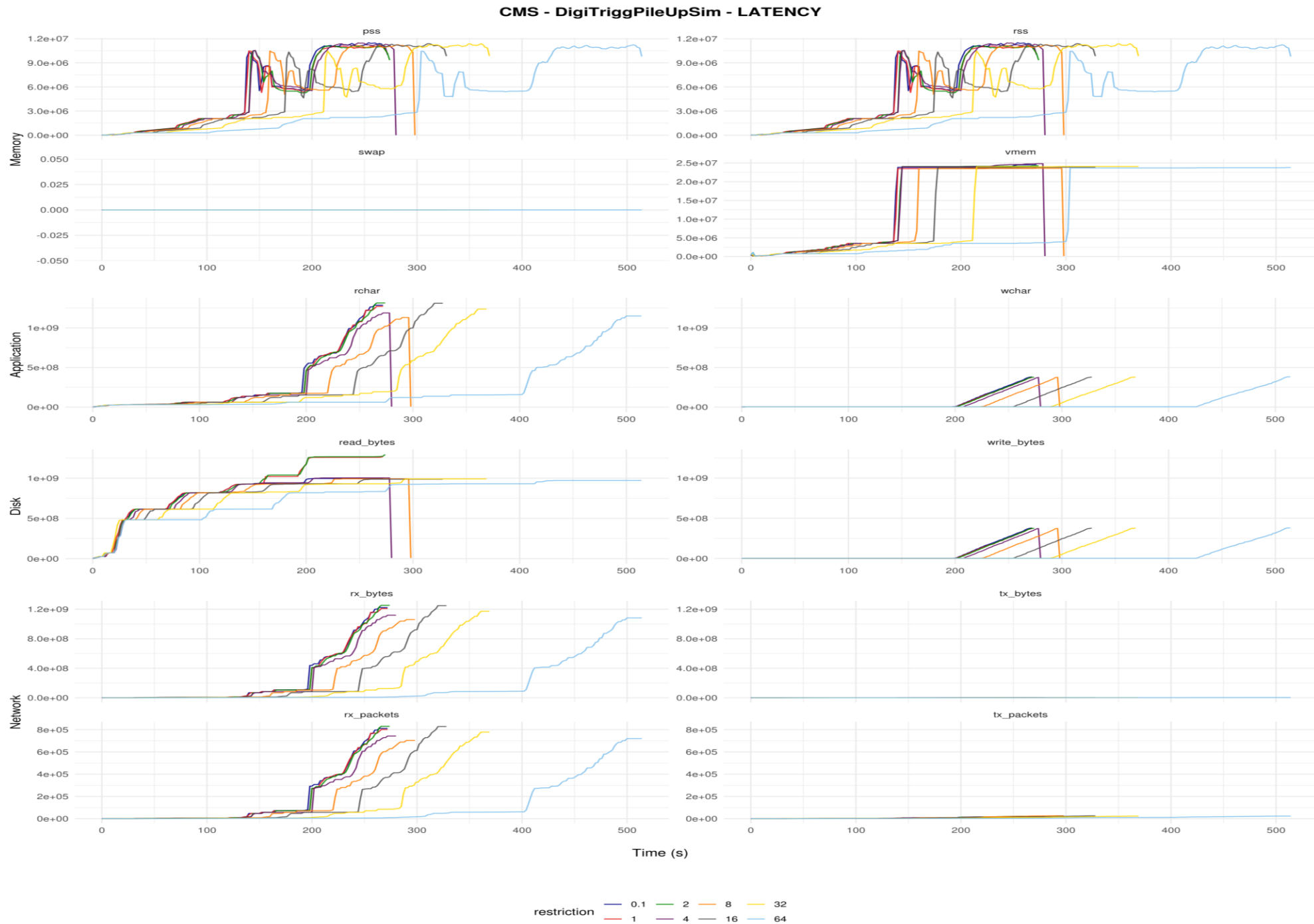
- The workload shows a ~30% time increase between best and worst case scenario.





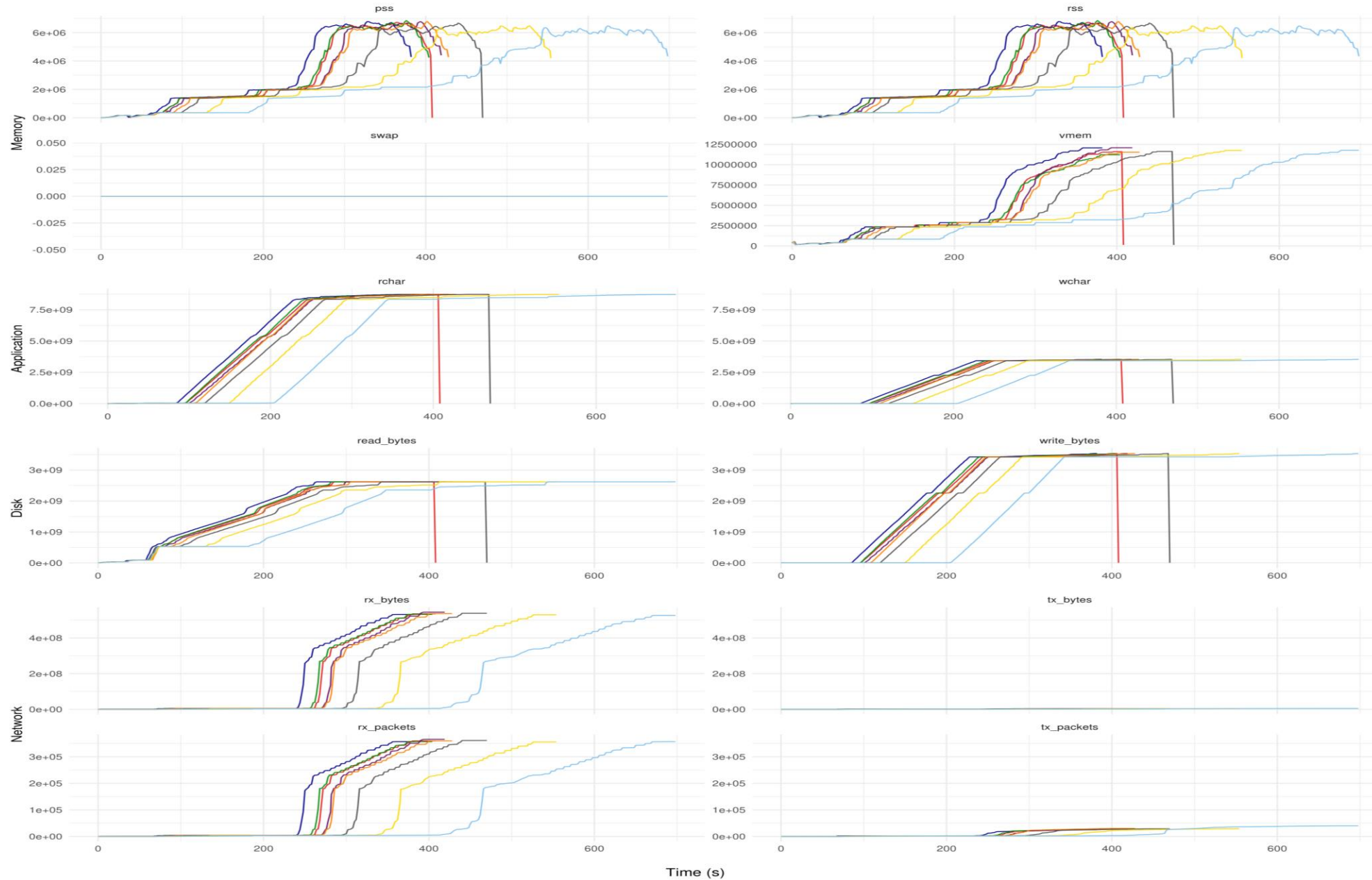
# CMS DigiTriggPile UpSim: Latency

- Curves show very slight slope changes, and mostly time shifted behaviours



# CMS RecoAnalysis: Latency

CMS - RecoAnalysis - LATENCY



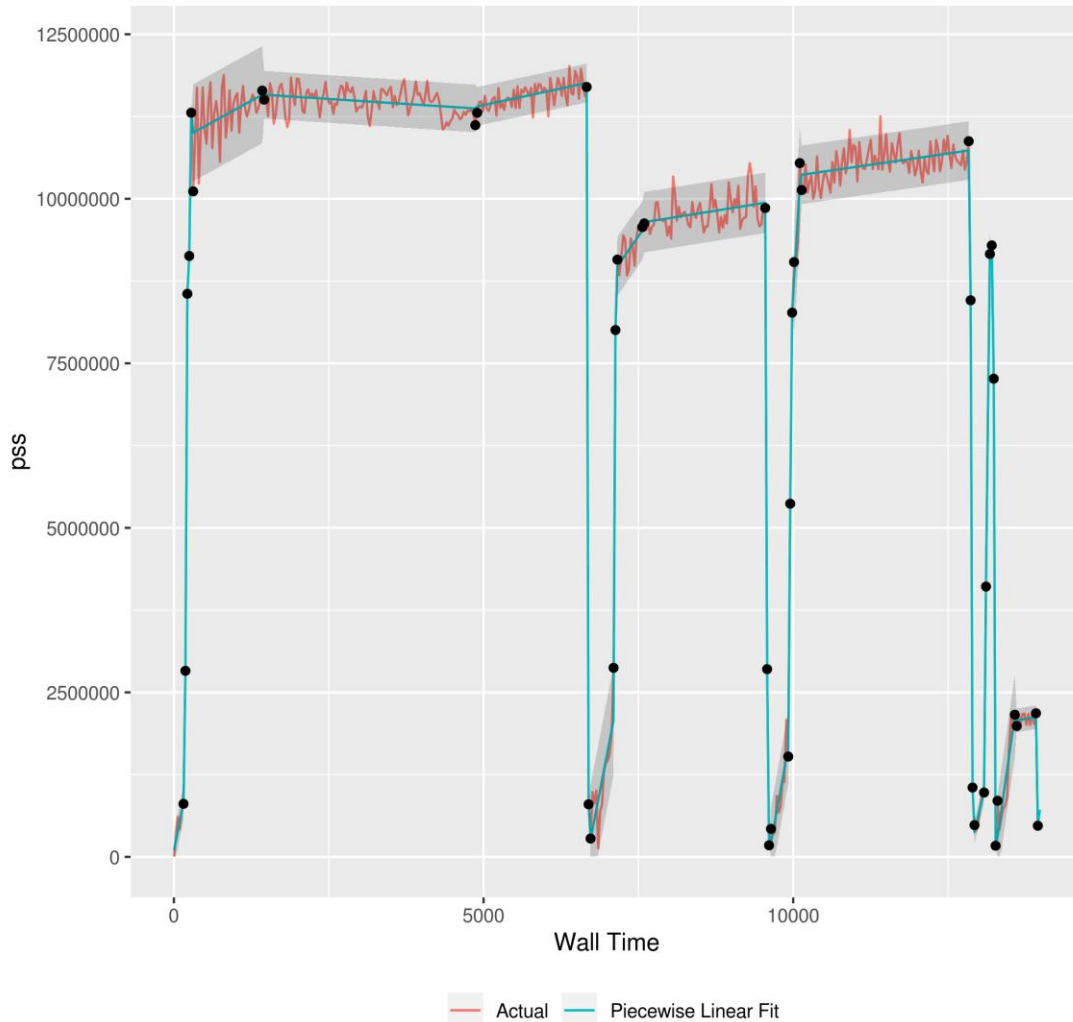
- The curves again only show time shifting of similar patterns. In this case there is a 2x difference between worst and best case scenarios

# NEXT STEP

- **Apply parametrization to all time series**
- **Include impact of restrictions**

# PRELIMINARY RESULT

atlas-atlas-digireco-190302-pss

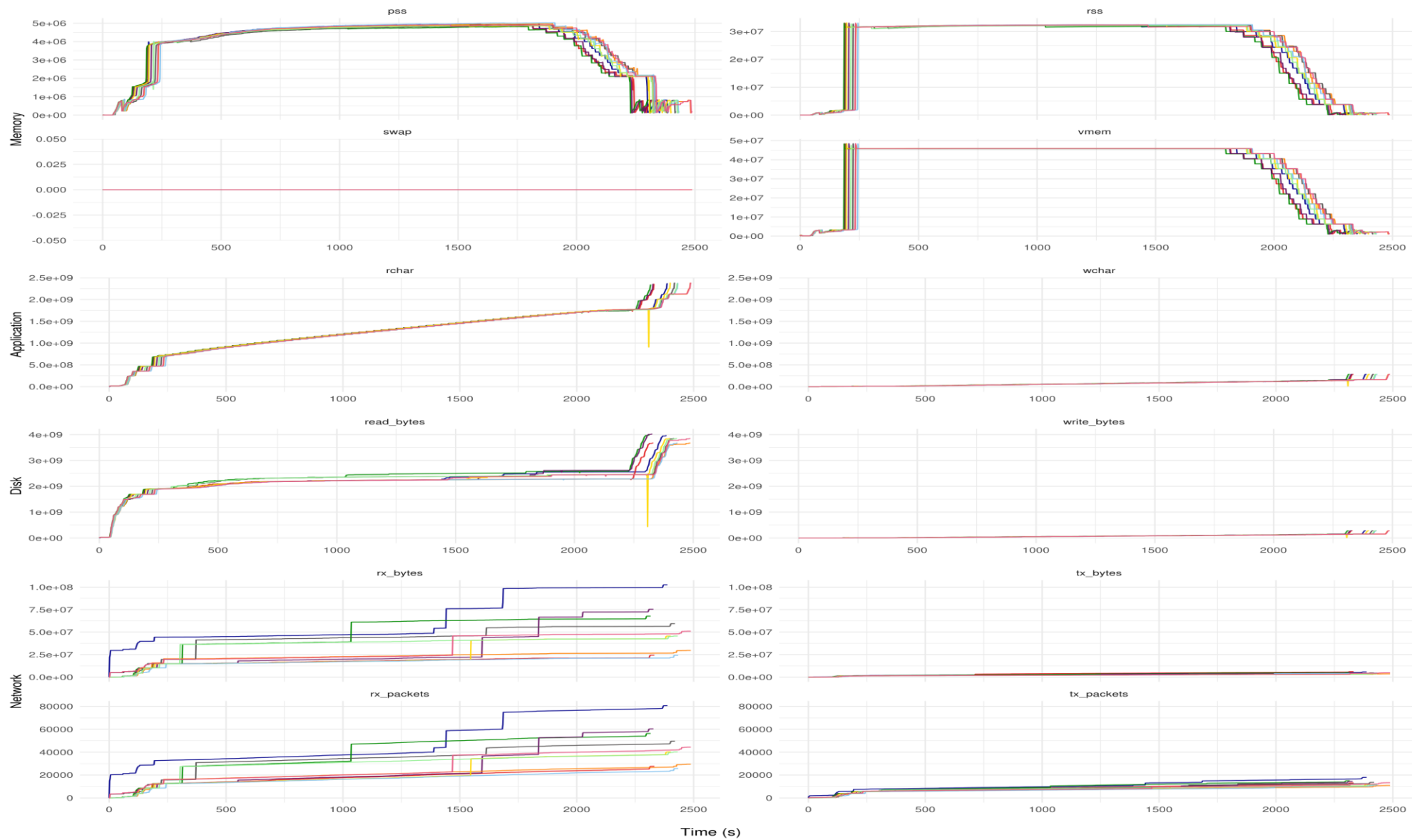


```
1 start,end,b0,b1,MSE
2 0,155,99611.80826145464,4717.217061011314,135538.10772202484
3 186,186,2827225,0,0
4 217,217,8556247,0,0
5 248,279,-8279601.000000007,70205.80645161294,0
6 311,1426,10841393.154296715,518.9666674248671,376204.6478722952
7 1457,4867,11670426.98306852,-61.20637229629499,184584.25990544527
8 4898,6665,10416023.393337343,201.61652845633668,150465.334534868
9 6696,6696,799041,0,0
10 6728,7099,-32562733.02887043,4879.34136364999,423704.58494714374
11 7130,7130,8005469,0,0
12 7161,7564,-626709.2880632706,1340.8171017654977,228311.56636695808
13 7595,9548,8497873.86795191,151.0505634495127,233892.96012561486
14 9579,9579,2852760,0,0
15 9610,9610,176160,0,0
16 9641,9920,-50017299.80760436,5209.6769940089935,281956.93215174146
17 9952,9983,-926672650.451533,93653.45161289521,0
18 10014,10106,-147930177.4346719,15648.139037716834,438831.0652264979
19 10137,12834,8977720.071649626,136.83416229427934,227117.91335645493
20 12865,12865,8457589,0,0
21 12896,12896,1053922,0,0
22 12927,13082,-51122484.997611344,3983.788345025733,92060.96727472759
23 13113,13176,-1047325022.3364912,80186.47883776212,126224.43604014115
24 13206,13238,845047689.3749259,-63286.06249999439,0
25 13268,13268,169912,0,0
26 13299,13578,-86777317.19521393,6548.306385725739,319591.5005350885
27 13609,13920,-103425.49716278126,159.9884579791461,91369.81797705288
28 13950,13981,-108612280.99999064,7819.741935483203,0
```

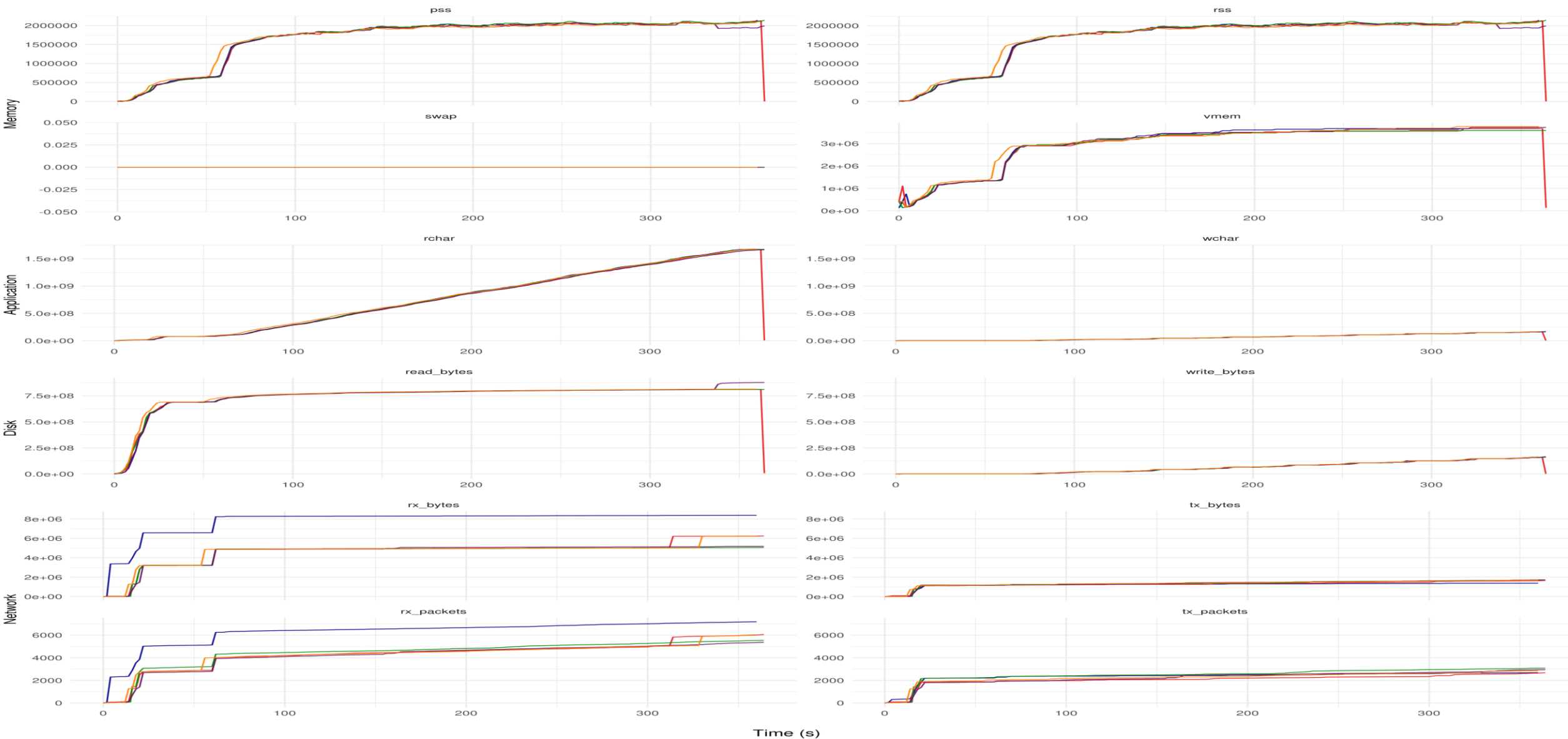


# BACKUP SLIDES

# ATLAS - G4 MC - BANDWIDTH

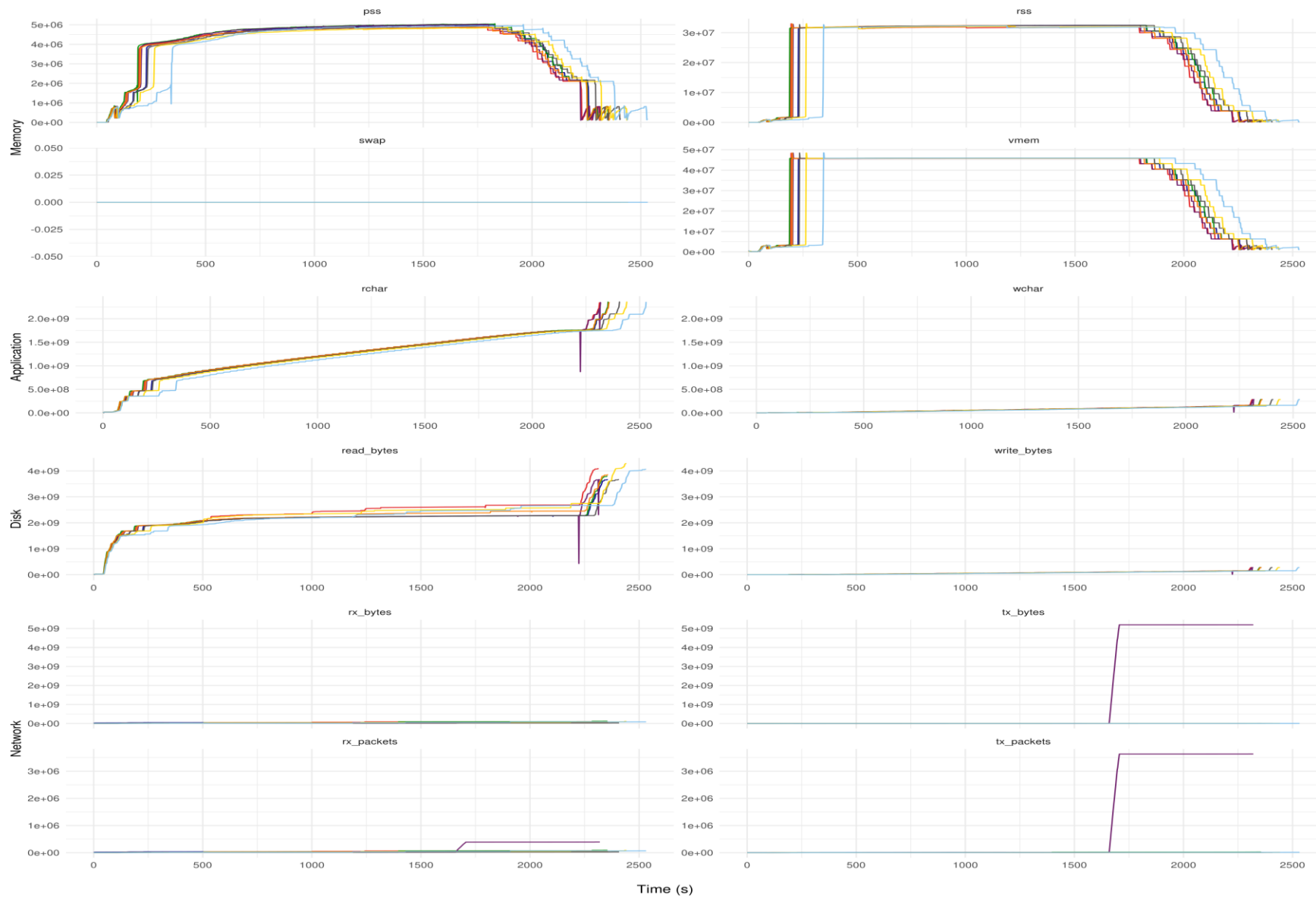


restriction — 1 — 10 — 50 — 250 — 750  
 — 5 — 25 — 100 — 500 — 1250



restriction — 4 — 8 — 16 — 32 — 64

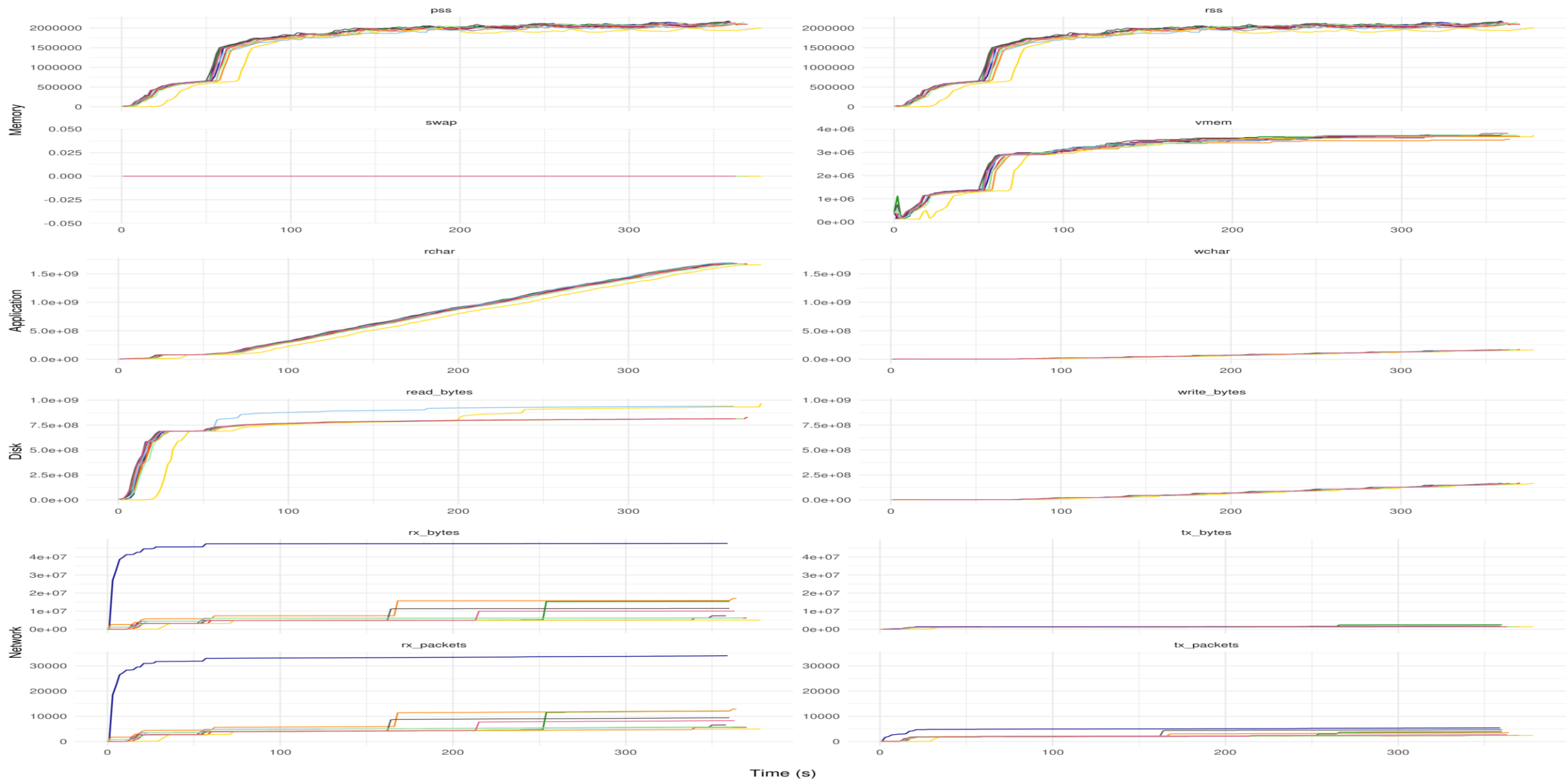
# ATLAS - G4 MC - LATENCY



restriction 0.1 2 8 32  
1 4 16 64

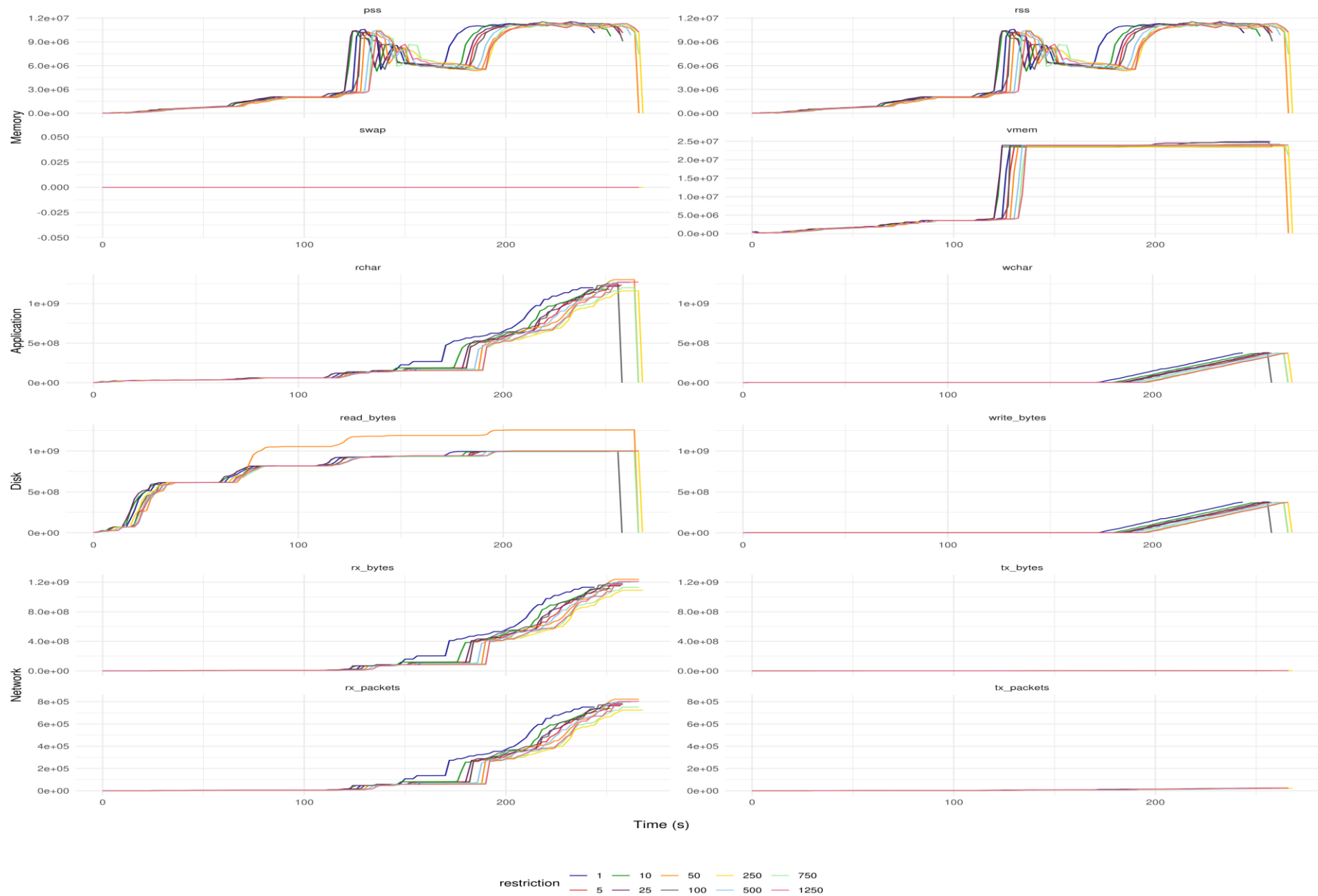


# CMS - G4 MC - BANDWIDTH

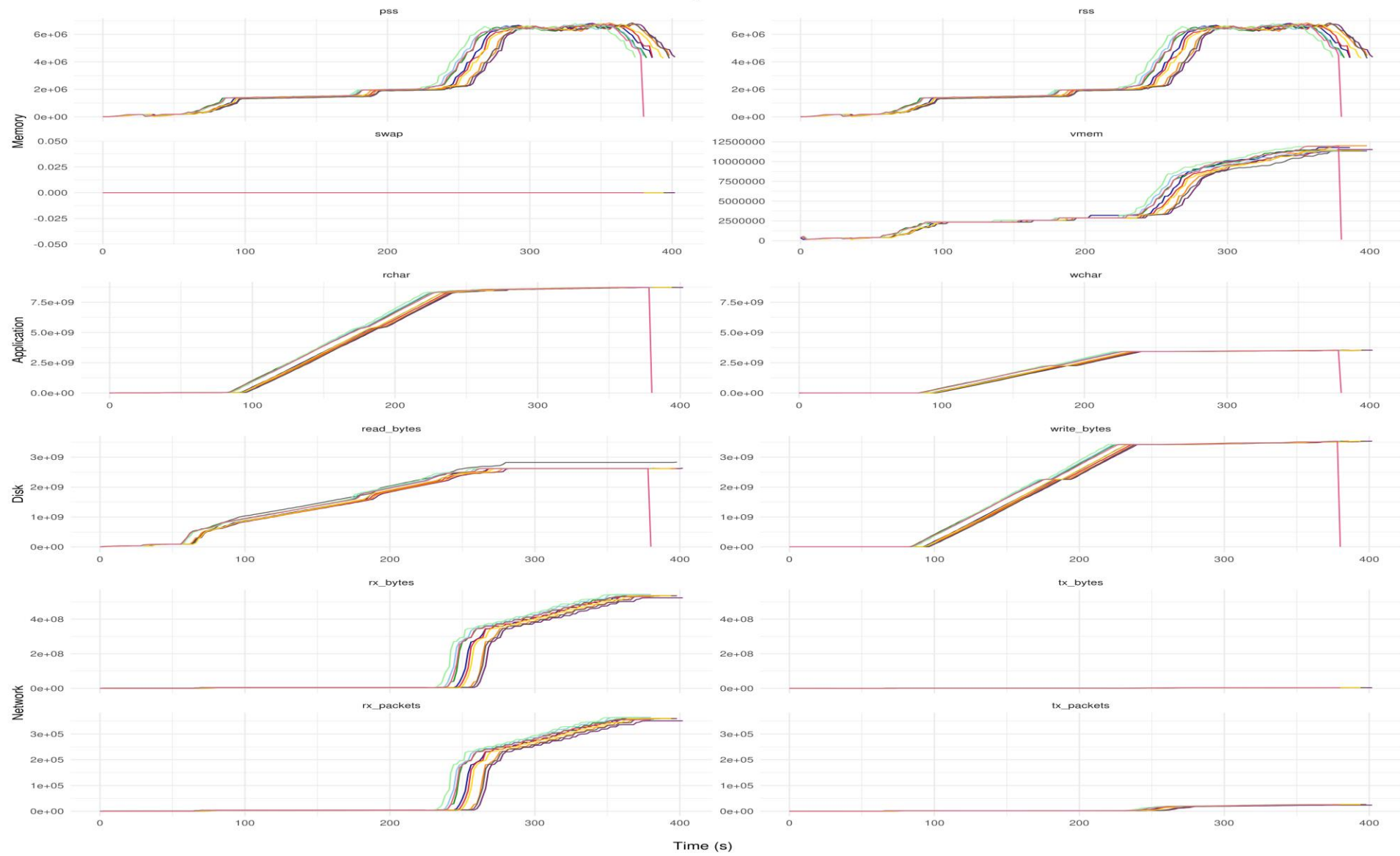


restriction — 1 — 10 — 50 — 250 — 750  
 — 5 — 25 — 100 — 500 — 1250

CMS - DigiTriggPileUpSim - BANDWIDTH



CMS - RecoAnalysis - BANDWIDTH



restriction — 1 — 5 — 10 — 25 — 50 — 100 — 250 — 500 — 750 — 1250



# QUESTIONS?

*riccardo.maganza@cern.ch*

@RiccardoMaganza