

Measuring WLCG data streams at batch job level

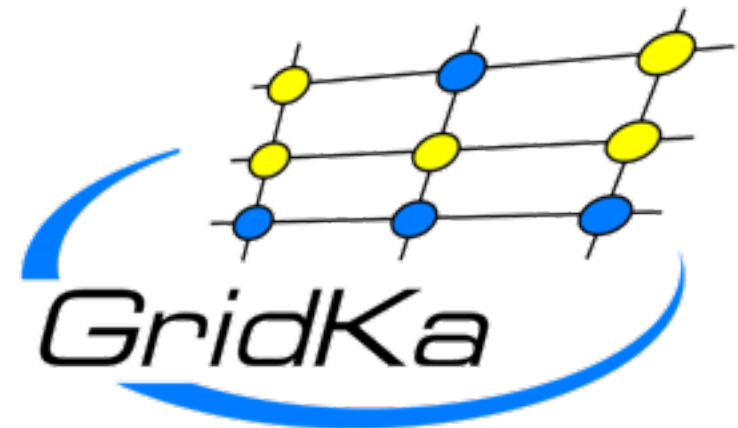
Eileen Kuehn

Steinbuch Centre for Computing, GridKa



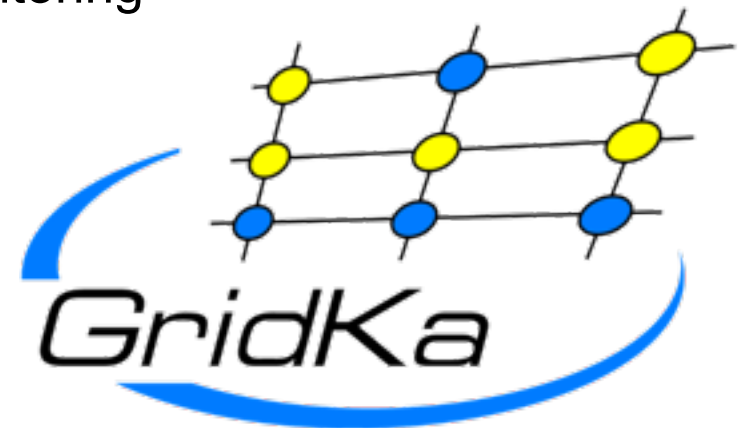
Context

- PhD student in computer science at KIT
 - Identification of data stream patterns of WLCG jobs
 - Anomaly detection
- Assigned to GridKa Tier1
- Cooperation with (local) CMS group



Motivation

- Introduction of remote IO (AAA, FAX, ...) to WLCG workflows
- Data can be read from anywhere
 - We'd like to know... [Who? What? How?]
- No monitoring of network resource usage via batch system
 - Monitoring for batch jobs desired
 - Network hardware does not support required details
 - Differentiation between internal / external traffic
- Technological foundation for pilot monitoring
 - Pilot vs. job traffic



„Wishlist“

- Network monitoring per process
- Grouping of processes by batch jobs
- Integration into batch system

Implementation

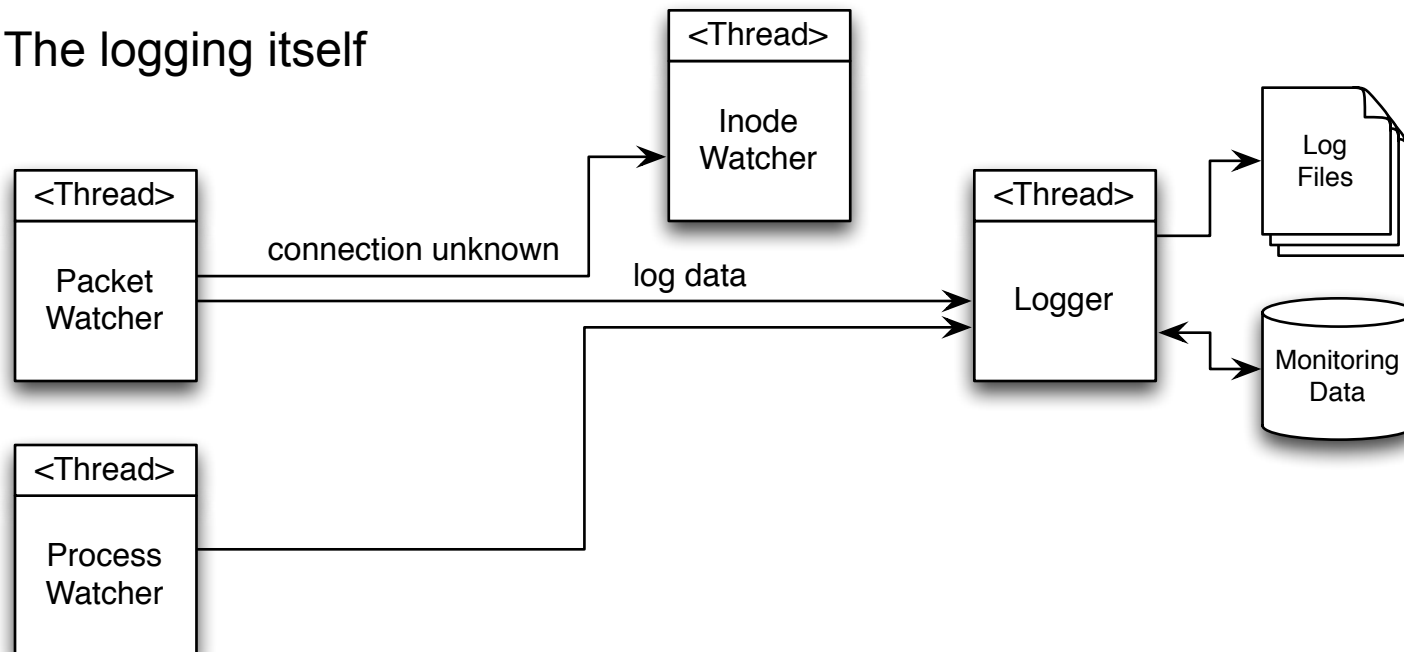
- Network traffic monitoring tool
 - Logging of network traffic
 - Logging of network related processes of jobs
 - Matching of traffic to processes
 - Matching processes to parent batch system process
 - Categorization into internal and external network traffic
 - Grouping of traffic by connection

```
[gridka]
intern=^10\.([0-9]{1,3})\.([0-9]{1,3})\.([0-9]{1,3})
extern=.*
splitinternextern=true
format=csv
skipotherpids=false
reducedoutput=true
loginterval=20
detail=connection
```

Example configuration

Implementation

- Tool is based on different modules
 - Packet watcher
 - Inode watcher
 - Process watcher
 - The logging itself



Implementation: Inode Watcher

- Repeated search of inodes in
 - /proc/net/tcp
 - /proc/net/tcp6
- Matching of sockets to inodes

```

eileen -- admin@carl:~ -- admin@carl:~ -- ssh -- 156x46
sl local_address rem_address st tx_queue rx_queue tr tm->when retrnsmt uid timeout inode
0: 00000000:BFEB 00000000:0000 8A 00000000:00000000 00:00000000 00000000 29 0 13501 1 ffff8806742bd400 99 0 0 10 -1
1: 00000000:1FAA 00000000:0000 8A 00000000:00000000 00:00000000 00000000 18000 0 4118059402 1 ffff8802415952c0 99 0 0 10 -1
2: 00000000:D6AA 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14878 0 4114449833 1 ffff88032e9104c0 99 0 0 10 -1
3: 00000000:BD2B 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14878 0 4119843186 1 ffff8803d2450200 99 0 0 10 -1
4: 00000000:3AAD 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 414816 1 ffff8806742bc700 99 0 0 10 -1
5: 00000000:976F 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14878 0 4120015973 1 ffff8801a9b40e40 99 0 0 10 -1
6: 00000000:006F 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 13241 1 ffff880c73e3a6c0 99 0 0 10 -1
7: 00000000:DAD0 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14878 0 4119843246 1 ffff8803b96a52c0 99 0 0 10 -1
8: 00000000:81F2 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 4100753218 1 ffff88011ffe2300 99 0 0 10 -1
9: 0100007F:2495 00000000:0000 8A 00000000:00000000 00:00000000 00000000 11999 0 4120729379 1 ffff88032e910b40 99 0 0 10 -1
10: 00000000:BE56 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14878 0 4119952011 1 ffff880173f3ce8c0 99 0 0 10 -1
11: 00000000:0016 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 1962184037 1 ffff8802dbd0d00 99 0 0 10 -1
12: 00000000:0010 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 1962184037 1 ffff8802dbd0d00 99 0 0 10 -1
100: 6007010A:A4F4 8004618A:0C38 08 00000000:00000001 00:00000000 00000000 496 0 4120995300 1 ffff8803ed112900 20 3 10 10 -1
101: 6007010A:B3BE 7EBFE183:2813 01 00000000:00000000 02:00019F48 00000000 14878 0 4119843230 2 ffff8803b96a5940 34 3 16 10 -1
102: 6007010A:A1F3 4088E80:271E 01 00000000:00000000 02:0007778F 00000000 14878 0 4114449845 2 ffff8802995f9000 21 3 16 10 -1
103: 6007010A:BF5D 4088E80:2706 01 00000000:00000000 02:000AD6EE 00000000 14878 0 4121756537 2 ffff88024ea3ac00 21 3 0 10 -1
104: 6007010A:8E08 7EBFE183:2813 01 00000000:00000000 02:0001A1C6 00000000 14878 0 4119843693 2 ffff8803be28d240 33 3 1 10 -1
105: 6007010A:80C3 8004618A:0C38 08 00000000:00000001 00:00000000 00000000 496 0 3550324150 1 ffff880674b67600 20 3 10 10 -1
106: 6007010A:CC4E 7EBFE183:2813 01 00000000:00000000 02:00019FF6 00000000 14878 0 4119843345 2 ffff8802dbd0d000 34 3 12 10 -1
107: 6007010A:8F24 4088E80:271E 01 00000000:00000000 02:000766F0 00000000 14878 0 4120960067 2 ffff88032b902e00 21 0 0 5 8
108: 0100007F:94A8 0100007F:8822 01 00000000:00000000 02:000282C3 00000000 14878 0 4120015970 2 ffff8801a9b414c0 20 3 30 10 -1
109: 6007010A:C196 4088E80:2664 01 00000000:00000000 02:0000E27A 00000000 14878 0 4121357962 2 ffff8802e1583240 21 0 0 9 9
110: 6007010A:8F9D 4088E80:2706 01 00000000:00000000 02:000AD603 00000000 14878 0 4121750126 2 ffff88024ea3b200 21 3 0 10 -1
111: 6007010A:98AC 7EBFE183:2764 01 00000000:00000000 02:0004465A 00000000 14878 0 4114471765 2 ffff8802416573c0 34 3 1 10 -1
112: 0100007F:DA15 0100007F:85A9 01 00000000:00000000 02:0002273E 00000000 14878 0 4119952000 2 ffff880173f3cf5c0 20 3 30 10 -1
113: 6007010A:C68A 7EBFE183:2764 01 00000000:00000000 02:0007E33E 00000000 14878 0 4114449890 2 ffff88011ffe2900 33 3 12 10 -1
114: 6007010A:87AF A888E80:0FF0 01 00000000:00000000 02:000282C0 00000000 14878 0 4120015925 2 ffff880229a5e440 21 3 28 10 -1
115: 6007010A:803D 4088E80:25E8 01 00000000:00000000 02:000ADAEE 00000000 14878 0 4121761945 2 ffff8803f7ba8440 21 3 1 12 -1
116: 6007010A:96FE 7EBFE183:2840 01 00000000:00000000 02:000ADA31 00000000 14878 0 4121760473 2 ffff88012b33f140 34 3 1 10 -1
  
```

Implementation: Inode Watcher

- Repeated search of inodes in
 - /proc/net/tcp
 - /proc/net/tcp6
- Matching of sockets to inodes

```
eileen -- admin@carl:~ -- admin@carl:~ -- ssh -- 156x46
sl local_address rem_address st tx_queue rx_queue tr tm->when retrnsqt uid timeout inode
0: 00000000:BFEB 00000000:0000 8A 00000000:00000000 00:00000000 00000000 29 0 13501 1 ffff8806742bd400 99 0 0 10 -1
1: 00000000:1FAA 00000000:0000 8A 00000000:00000000 00:00000000 00000000 18000 0 4118059402 1 ffff8802415952c0 99 0 0 10 -1
2: 00000000:D6AA 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14870 0 4114449833 1 ffff88032e9104c0 99 0 0 10 -1
3: 00000000:8D2B 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14870 0 4119843186 1 ffff8803d2450200 99 0 0 10 -1
4: 00000000:3AAD 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 414816 1 ffff8806742bc700 99 0 0 10 -1
5: 00000000:976F 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14870 0 4120015973 1 ffff8801a9b40e40 99 0 0 10 -1
6: 00000000:006F 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 13241 1 ffff880c73e3a6c0 99 0 0 10 -1
7: 00000000:DAD0 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14870 0 4119843246 1 ffff8803b96a52c0 99 0 0 10 -1
8: 00000000:81F2 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 4100753218 1 ffff88011ffe2300 99 0 0 10 -1
9: 0100007F:2495 00000000:0000 8A 00000000:00000000 00:00000000 00000000 11999 0 4120729379 1 ffff88032e910b40 99 0 0 10 -1
10: 00000000:BE56 00000000:0000 8A 00000000:00000000 00:00000000 00000000 14870 0 4119952011 1 ffff880173fce8c0 99 0 0 10 -1
11: 00000000:0016 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 1962184037 1 ffff88020dbdad00 99 0 0 10 -1
12: 00000000:0010 00000000:0000 8A 00000000:00000000 00:00000000 00000000 0 0 1962184037 1 ffff88020dbdad00 99 0 0 10 -1
```

```
112: 0100007F:DA15 0100007F:85A9 01 00000000:00000000 02:0002273E 00000000 14870 0 4119952008 2 ffff880173fcf5c0 20 3 30 10 -1
```

Implementation: Process Watcher

- Monitoring creation of new processes
 - CONFIG_CONNECTOR and CONFIG_PROC_EVENTS of kernel
- Process tree with additional information about relevance for monitoring

```

pid: 18604 pid from process: 18604 ppid: 6263 Unimportant (gridka_monitor)-./gridka_monitor (0x2061e70)
pid: 18692 pid from process: 18692 ppid: 15857 Groupedprocess (python)-python (0x2061f40)
pid: 18792 pid from process: 18792 ppid: 18488 Groupedprocess (1023127)-/bin/bash (0x2062020)
pid: 18811 pid from process: 18811 ppid: 18792 Groupedprocess (CREAM940071419_)-/bin/sh (0x20620e0)
pid: 19016 pid from process: 19016 ppid: 1 Unimportant (cvmfs2)-/usr/bin/cvmfs2 (0x20621a0)
pid: 19026 pid from process: 19026 ppid: 1 Unimportant (cvmfs2)-/usr/bin/cvmfs2 (0x2062270)
pid: 19448 pid from process: 19448 ppid: 20415 Trafficgroupedprocess (sge_shepherd)-sge_shepherd-1054928 (0x2062330)
pid: 19458 pid from process: 19458 ppid: 20415 Trafficgroupedprocess (sge_shepherd)-sge_shepherd-1054933 (0x20623f0)
pid: 19506 pid from process: 19506 ppid: 19448 Groupedprocess (1054928)-/bin/bash (0x20624c0)
pid: 19507 pid from process: 19507 ppid: 19458 Groupedprocess (1054933)-/bin/bash (0x2062590)
pid: 19512 pid from process: 19512 ppid: 19506 Groupedprocess (CREAM846634783_)-/bin/sh (0x2062650)
pid: 19521 pid from process: 19521 ppid: 19507 Groupedprocess (CREAM617249015_)-/bin/sh (0x2062710)
pid: 19840 pid from process: 19840 ppid: 19512 Groupedprocess (perl)-perl (0x20627e0)
pid: 19841 pid from process: 19841 ppid: 19521 Groupedprocess (perl)-perl (0x20628b0)
pid: 19842 pid from process: 19842 ppid: 19840 Groupedprocess (sh)-sh (0x2062970)
pid: 19843 pid from process: 19843 ppid: 19840 Groupedprocess (perl)-perl (0x2062a30)
pid: 19845 pid from process: 19845 ppid: 19841 Groupedprocess (sh)-sh (0x2062af0)
pid: 19846 pid from process: 19846 ppid: 19841 Groupedprocess (perl)-perl (0x2062bb0)
pid: 19848 pid from process: 19848 ppid: 19842 Groupedprocess (glidein_startup)-/bin/bash (0x2062c70)
pid: 19851 pid from process: 19851 ppid: 19845 Groupedprocess (glidein_startup)-/bin/bash (0x2062d30)
pid: 20187 pid from process: 20187 ppid: 18538 Groupedprocess (perl)-perl (0x2062e00)
pid: 20202 pid from process: 20202 ppid: 20187 Groupedprocess (sh)-sh (0x2062ed0)
pid: 20203 pid from process: 20203 ppid: 20187 Groupedprocess (perl)-perl (0x2062f90)
pid: 20205 pid from process: 20205 ppid: 20202 Groupedprocess (glidein_startup)-/bin/bash (0x2063050)
pid: 20415 pid from process: 20415 ppid: 1 Groupingfactor (sge_execd)-/opt/gridengine/bin/lx-and64/sge_execd (0x2063110)
  
```

Implementation: Logger

- Accumulation of data regarding log intervals
- Different output methods
 - UNIVA Grid Engine / Command
 - CSV, JSON

```

1394888071-process.log-20140503
twe,pid,ppid,uid,name,cmd,exit_code,state,gpid
1398993166,27783,18971,0,(curl),/usr/bin/curl,0,exit,0
1398993179,18975,18974,11999,(lcp-cp),/usr/bin/lcp-cp,0,...1
1398993186,18975,18974,11999,(lcp-cp),/usr/bin/lcp-cp,0,exit
1398993186,11767,11765,11999,(lcp-get-checksu),lcp-get-check
1398993186,11767,11765,11999,(lcp-get-checksu),lcp-get-check
1398993186,11784,11783,11999,(curl),curl,0,...13637
1398993186,11784,11783,11999,(curl),curl,0,exit,13637
1398993187,23348,18923,11999,(python),/usr/bin/python,0,exit
1398993214,17868,17867,11999,(lcp-cp),/usr/bin/lcp-cp,0,...1
1398993221,17868,17867,11999,(lcp-cp),/usr/bin/lcp-cp,0,exit
1398993222,18109,18108,11999,(lcp-get-checksu),lcp-get-check
1398993223,18109,18108,11999,(lcp-get-checksu),lcp-get-check
1398993223,18286,18285,11999,(curl),curl,0,...13637
1398993224,18286,18285,11999,(curl),curl,0,exit,13637
1398993224,18398,18397,11999,(curl),curl,0,...13637
1398993225,18398,18397,11999,(curl),curl,0,exit,13637
1398993258,21799,14348,11999,(curl),curl,7168,...13637
1398993253,22378,12766,0,(sntp),sntp,0,...0
1398993270,21799,14348,11999,(curl),curl,7168,exit,13637
1398993271,24568,24565,11999,(globus-url-copy),globus-url-c
1398993272,24568,24565,11999,(globus-url-copy),globus-url-c
1398993273,24785,24783,11999,(globus-url-copy),globus-url-c
1398993273,24785,24783,11999,(globus-url-copy),globus-url-c
1398993274,24987,24973,11999,(ssh),/usr/bin/ssh,0,...13637
1398993275,24987,24973,11999,(ssh),/usr/bin/ssh,0,exit,1363
1398993275,25878,25873,11999,(ssh),/usr/bin/ssh,0,...13637
1398993275,25878,25873,11999,(ssh),/usr/bin/ssh,0,exit,1363
1398993275,13637,19721,0,(sge_shepherd),sge_shepherd-969939
1398993307,38193,19721,0,(sge_shepherd),sge_shepherd-971499
  
```

```

1394888071-traffic.log-20140503
twe,pid,ppid,uid,in_rate,out_rate,in_cnt,out_cnt,conn,gpid
1398993191,0,0,0,0.8347656,0.8408203,4,6,10.1.7.104:1018-10.65.165.122:2049,0
,15628,28271,0,0.80957831,0.80478516,2,1,10.1.7.104:22-10.33.1.115:44494,0
,28994,3158,0,0.8193359,0.417773,6,6,10.1.7.104:22-10.33.1.115:48189,0
,3849,3884,14878,0.8193359,0.396924,6,4,10.1.7.104:39710-131.225.191.126:9774,30493
,0,0,0,0.857373,0.8587482,6,4,10.1.7.104:40638-10.33.1.114:8084,0
,3849,3884,14878,0.8193359,0.396924,6,4,10.1.7.104:41502-128.142.187.77:9774,30493
,19721,1,0,0.8253418,0.4375,4,7,10.1.7.104:49223-10.97.13.140:15020,0
,6751,3852,14878,0.80541992,0.80322266,1,1,10.1.7.104:58593-128.142.187.77:9774,30493
,31796,31795,0,0.80322266,0.80322266,1,1,10.1.7.104:58733-10.97.4.128:3128,0
1398993211,0,0,0,0.102734,0.12168,12,18,10.1.7.104:1018-10.65.165.122:2049,0
,28994,3158,0,0.8225586,0.487482,7,7,10.1.7.104:22-10.33.1.115:48189,0
,6809,3158,0,0.80957831,0.80478516,2,1,10.1.7.104:22-10.33.1.115:48204,0
,18398,18397,11999,0.888811,0.735156,18,19,10.1.7.104:32813-188.184.140.25:25443,13637
,17868,17867,11999,0.980342,0.911377,33,29,10.1.7.104:43173-192.188.46.82:2811,13637
,24743,23868,14878,0.80322266,0.80322266,1,1,10.1.7.104:43998-131.225.191.126:9667,13756
,18286,18285,11999,0.755371,0.14043,15,15,10.1.7.104:46376-128.142.171.211:443,13637
,17868,17867,11999,0.789668,0.648828,28,19,10.1.7.104:58239-192.188.45.69:8443,13637
,17868,17867,11999,0.626562,0.639941,19,18,10.1.7.104:58248-192.188.45.69:8443,13637
,17868,17867,11999,0.695215,0.669922,26,19,10.1.7.104:58241-192.188.45.69:8443,13637
,17868,17867,11999,0.682324,0.63916,28,18,10.1.7.104:58242-192.188.45.69:8443,13637
,17868,17867,11999,0.642969,0.644336,19,17,10.1.7.104:58245-192.188.45.69:8443,13637
,18189,18188,11999,0.715137,0.649609,28,19,10.1.7.104:58246-192.188.45.69:8443,13637
,0,0,0,0.121094,9.4877,41,27,10.1.7.104:56876-10.65.70.113:24551,0
1398993231,28994,3158,0,0.828125,0.484375,8,8,10.1.7.104:22-10.33.1.115:48189,0
,6809,3158,0,0.80322266,0.80634766,1,1,10.1.7.104:22-10.33.1.115:48204,0
,13333,3158,0,0.80957831,0.80478516,2,1,10.1.7.104:22-10.33.1.115:48212,0
  
```

Batch System Integration

- UNIVA Grid Engine as batch system being used at GridKa
- Worker nodes act as submit hosts
 - submission of predefined variables via qalter

```
qalter -ac gridka_network_ext_rate_in=xx -ac  
gridka_network_ext_rate_out=xx -ac  
gridka_network_ext_volume_in=xx -ac  
gridka_network_ext_volume_out=xx -ac  
gridka_network_int_rate_in=xx -ac  
gridka_network_int_rate_out=xx -ac  
gridka_network_in_volume_in=xx -ac  
gridka_network_in_volume_out=xx JobID
```

Example call

Current State

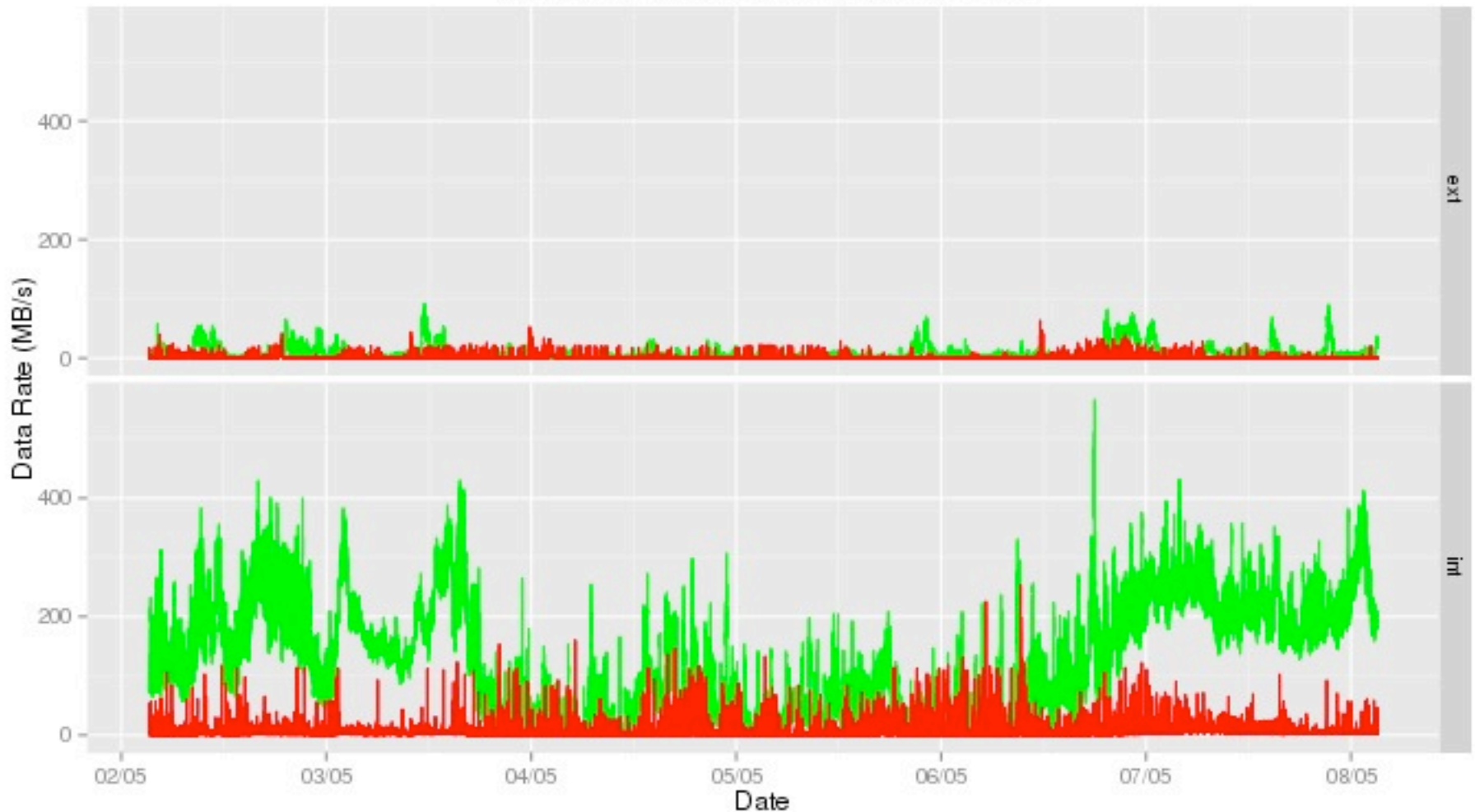
- Long-time measurements and evaluations
 - Running since mid of march
 - 32 worker nodes (1 rack)
 - 16 physical cores
 - 32 hyperthreaded cores
 - 24 job slots
 - 1 separate worker node (ALICE)

Current State

- Some facts
 - Variables being measured: 16
 - Approx. log file size / h with 20 seconds log interval at connection level
 - Batch jobs and background load: 302 kB
 - Batch jobs only: 208 kB
 - Toolbox
 - C/C++ implementation
 - Preprocessing: Python
 - Visualization: R

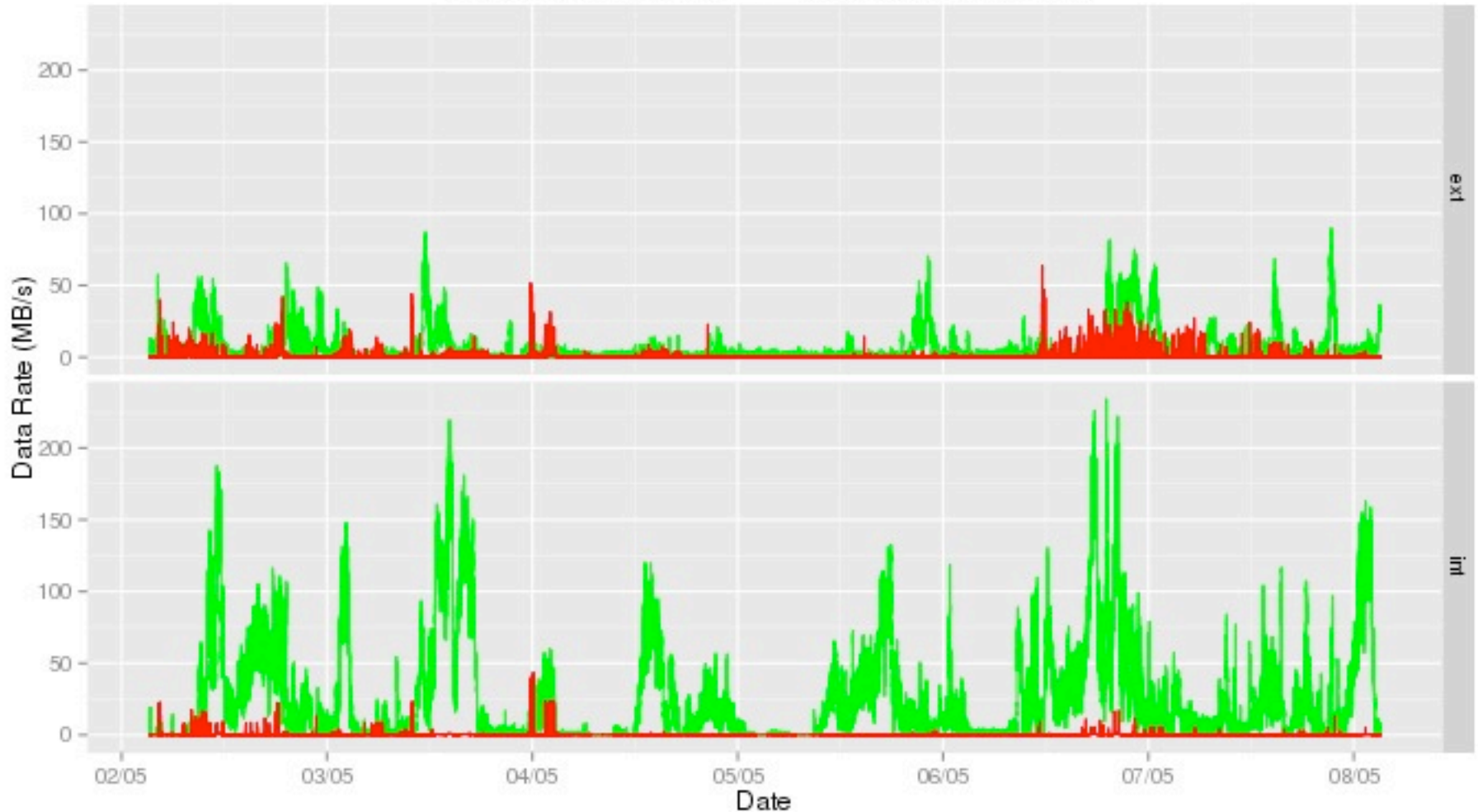
First Results

Total Data Rates of 21 Worker Nodes



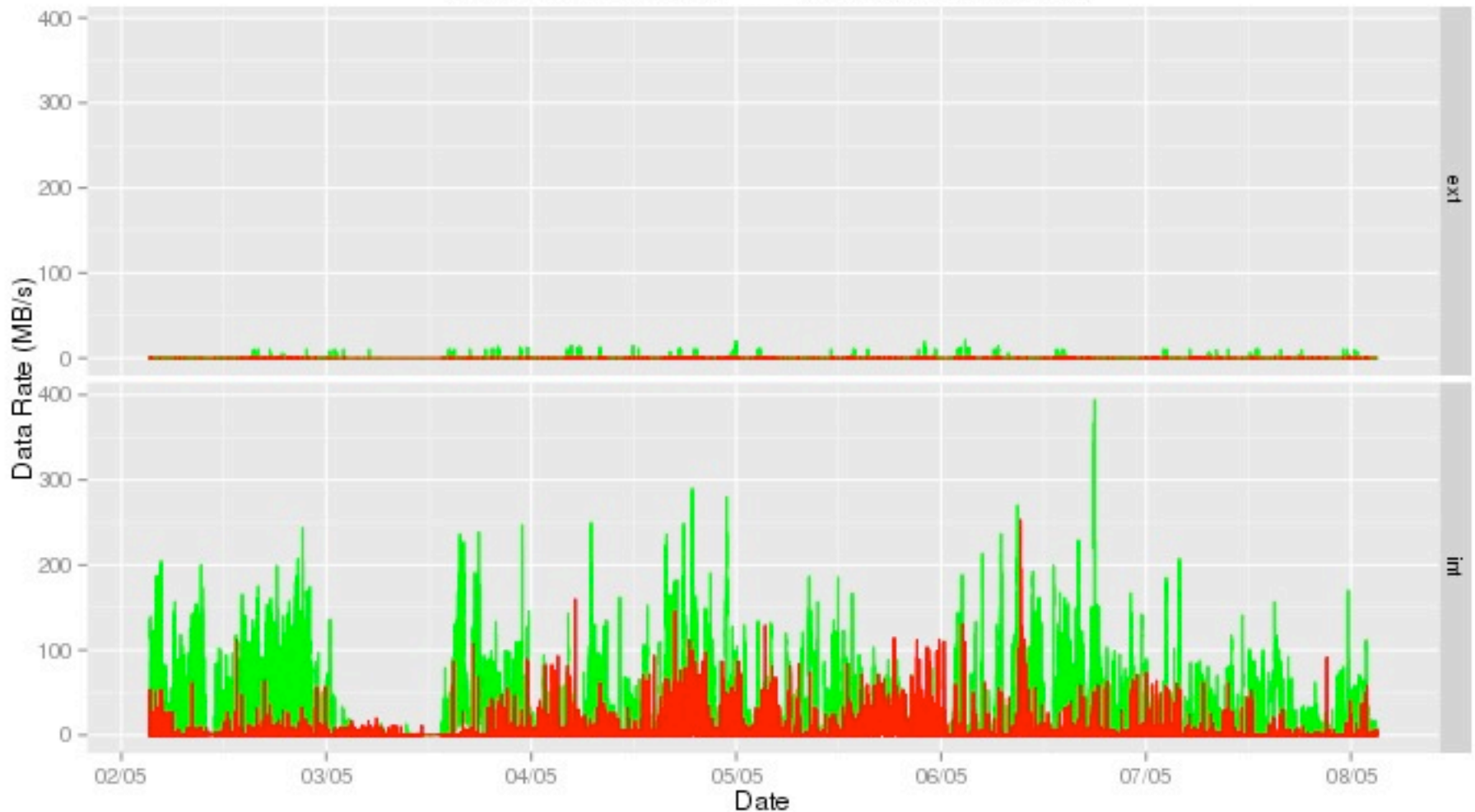
First Results: VO1

Total Data Rates for VO1 at 21 Worker Nodes



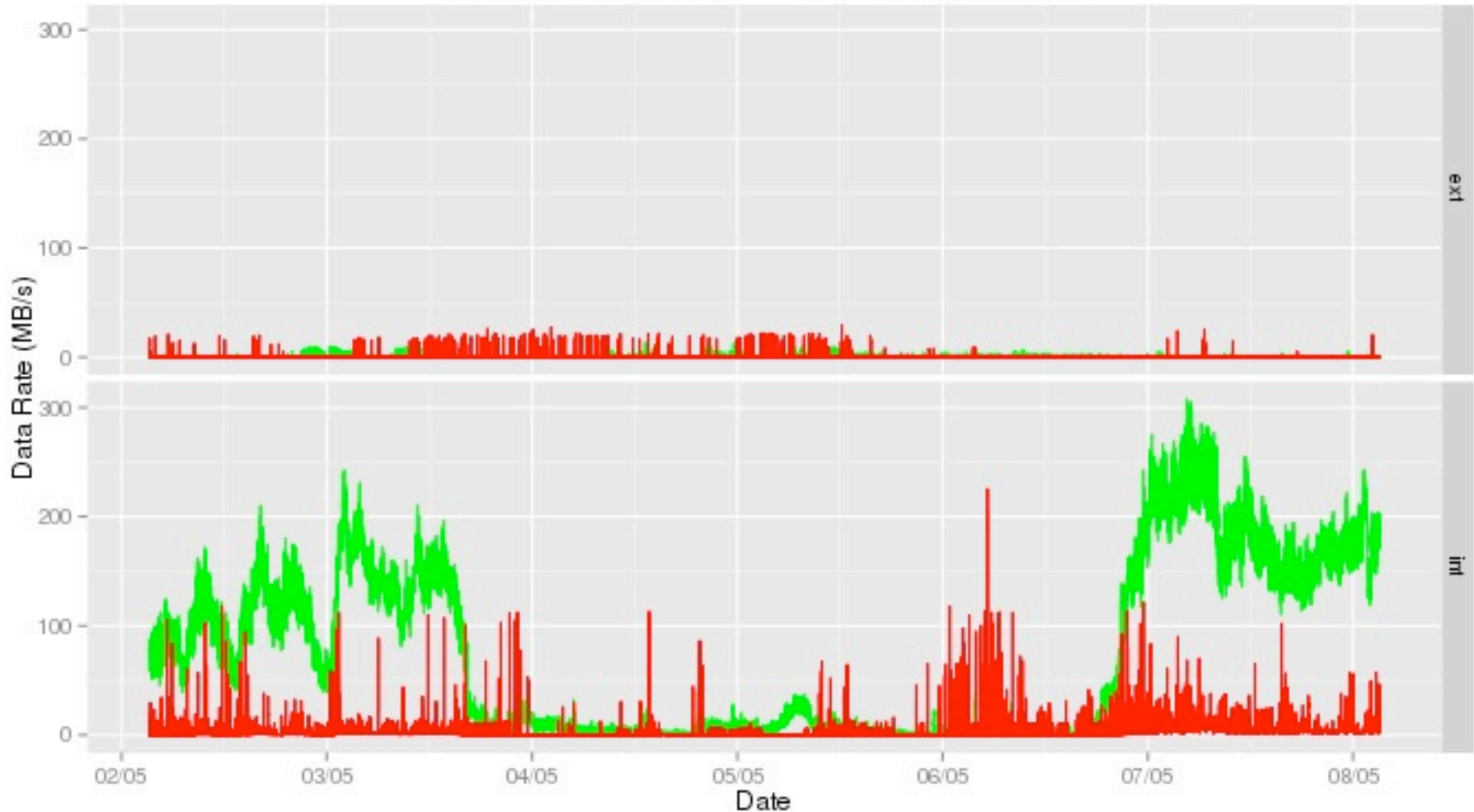
First Results: VO2

Total Data Rates for VO2 at 21 Worker Nodes



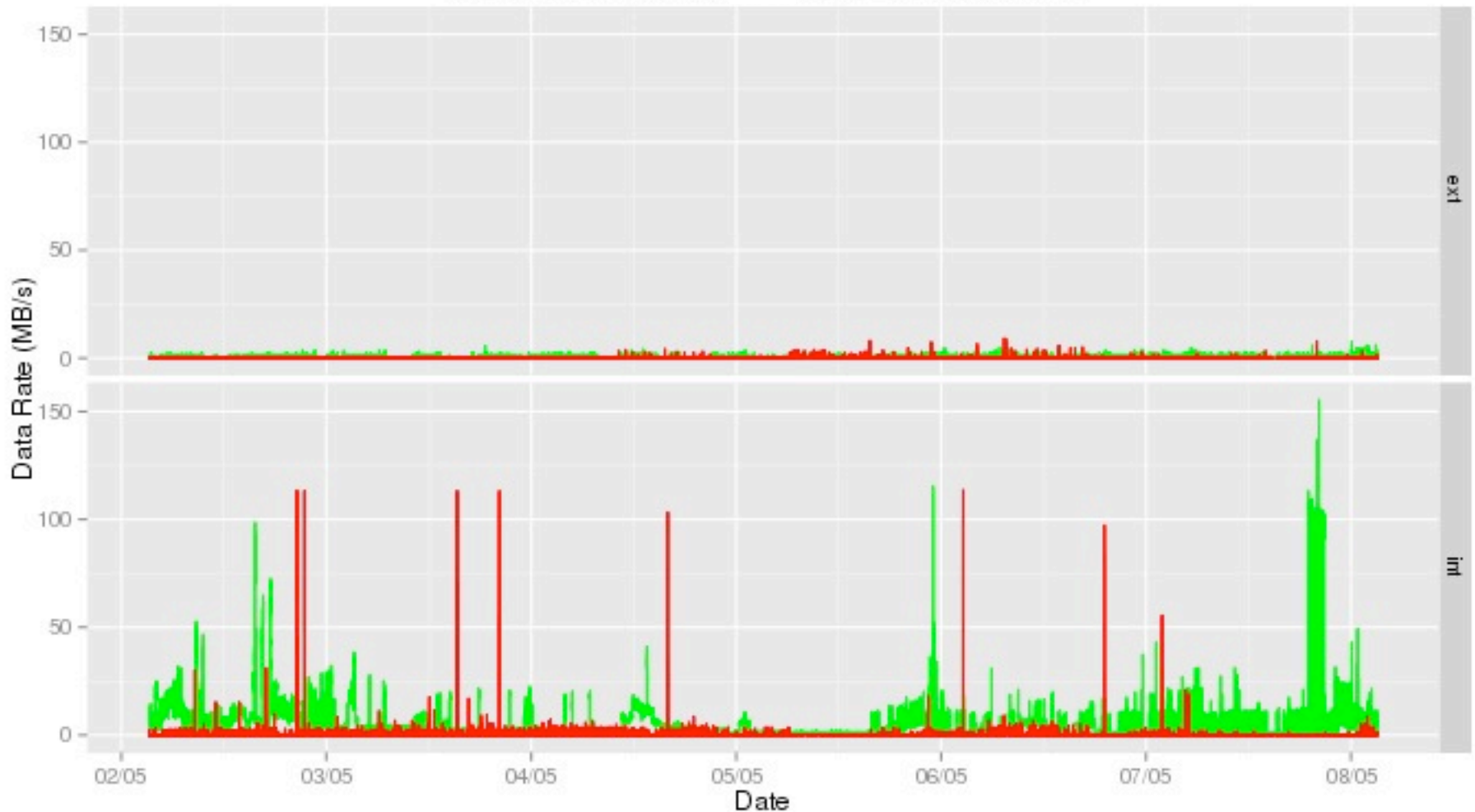
First Results: VO3

Total Data Rates for VO3 at 21 Worker Nodes



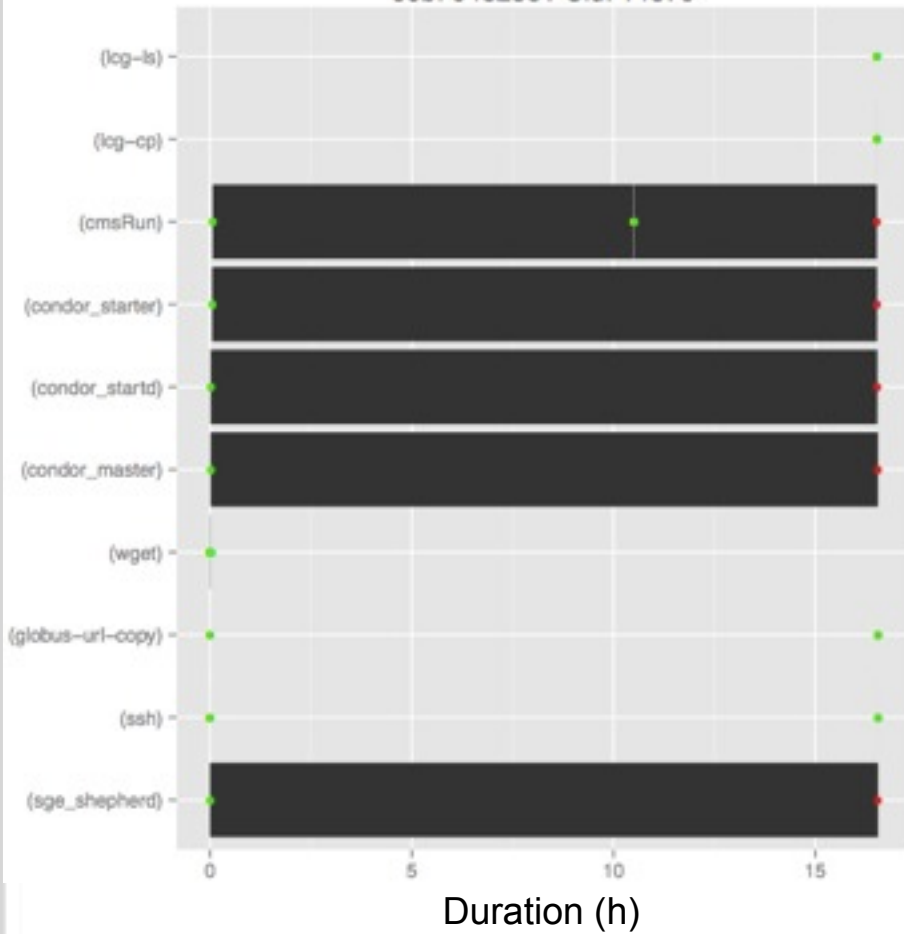
First Results: VO4

Total Data Rates for VO4 at 21 Worker Nodes

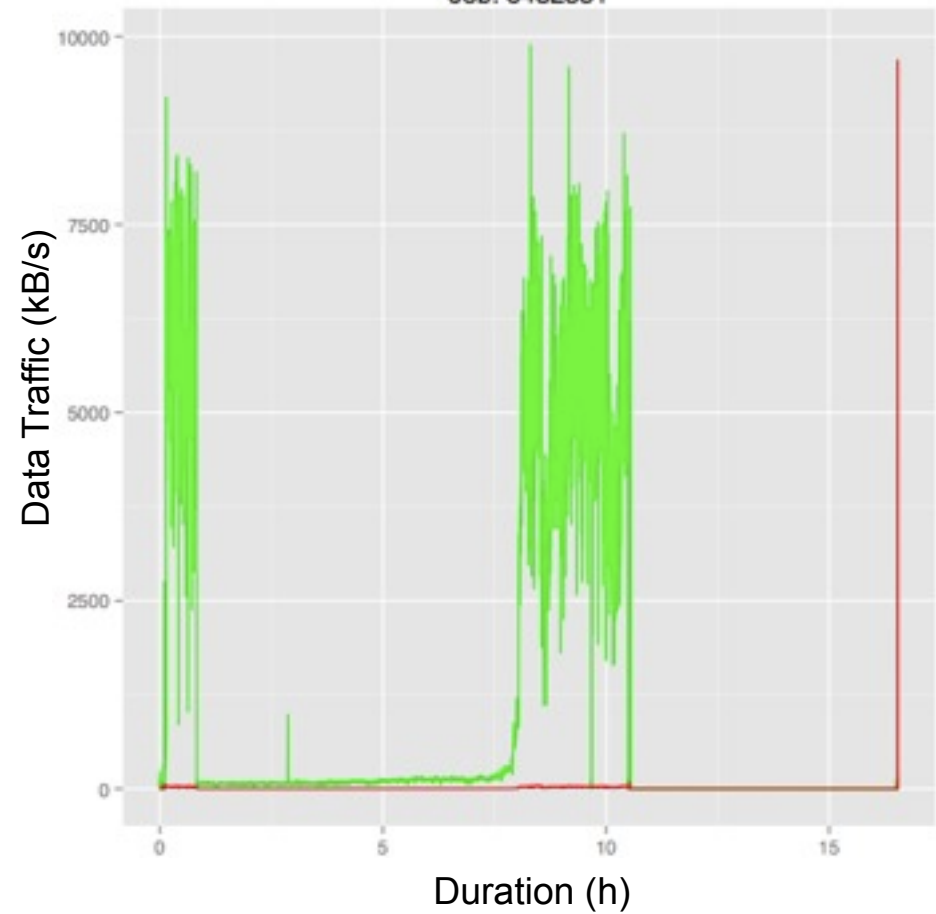


First Results

Job: 6482551 Uid: 14870



Job: 6482551



Future Improvements

- Further optimizations
- Two parallel developments
 - Tool for production use integrated into batch system
 - Tool focused on PhD work for more detailed information
- Monitoring cloud environments to predict application runtimes
 - Cooperation with University of Applied Sciences Zittau/Görlitz

Questions?

Steinbuch Centre for Computing, GridKa

