FermiGrid

Keith Chadwick

Nov-2008

Work supported by the U.S. Department of Energy under contract No. DE-AC02-07CH11359.

FermiGrid - Strategy

In order to better serve the entire program of Fermilab, the Computing Division undertook the strategy of placing all of its production resources in a Grid "meta-facility" infrastructure called FermiGrid.

This strategy is designed to allow Fermilab:

- to insure that the large experiments who currently have dedicated resources to have first priority usage of those resources that are purchased on their behalf.
- to allow opportunistic use of these dedicated resources, as well as other shared Farm and Analysis resources, by various Virtual Organizations (VO's) that participate in the Fermilab experimental program and by certain VOs that use the Open Science Grid (OSG).
- to optimize the use of resources at Fermilab.
- to make a coherent way of putting Fermilab on the Open Science Grid.
- to save some effort and resources by implementing certain shared services and approaches.
- to work together more coherently to move all of our applications and services to run on the Grid.
- to better handle a transition from Run II to LHC in a time of shrinking budgets and possibly shrinking resources for Run II worldwide.
- to fully support Open Science Grid and the LHC Computing Grid and gain positive benefit from this emerging infrastructure in the US and Europe.

FermiGrid - Mission and Accomplishments

Mission:

- To establish and maintain Fermilab as a high performance, robust, highly available, expertly supported and documented Premier Grid Facility which supports the scientific program based on Computing Division and Laboratory priorities.

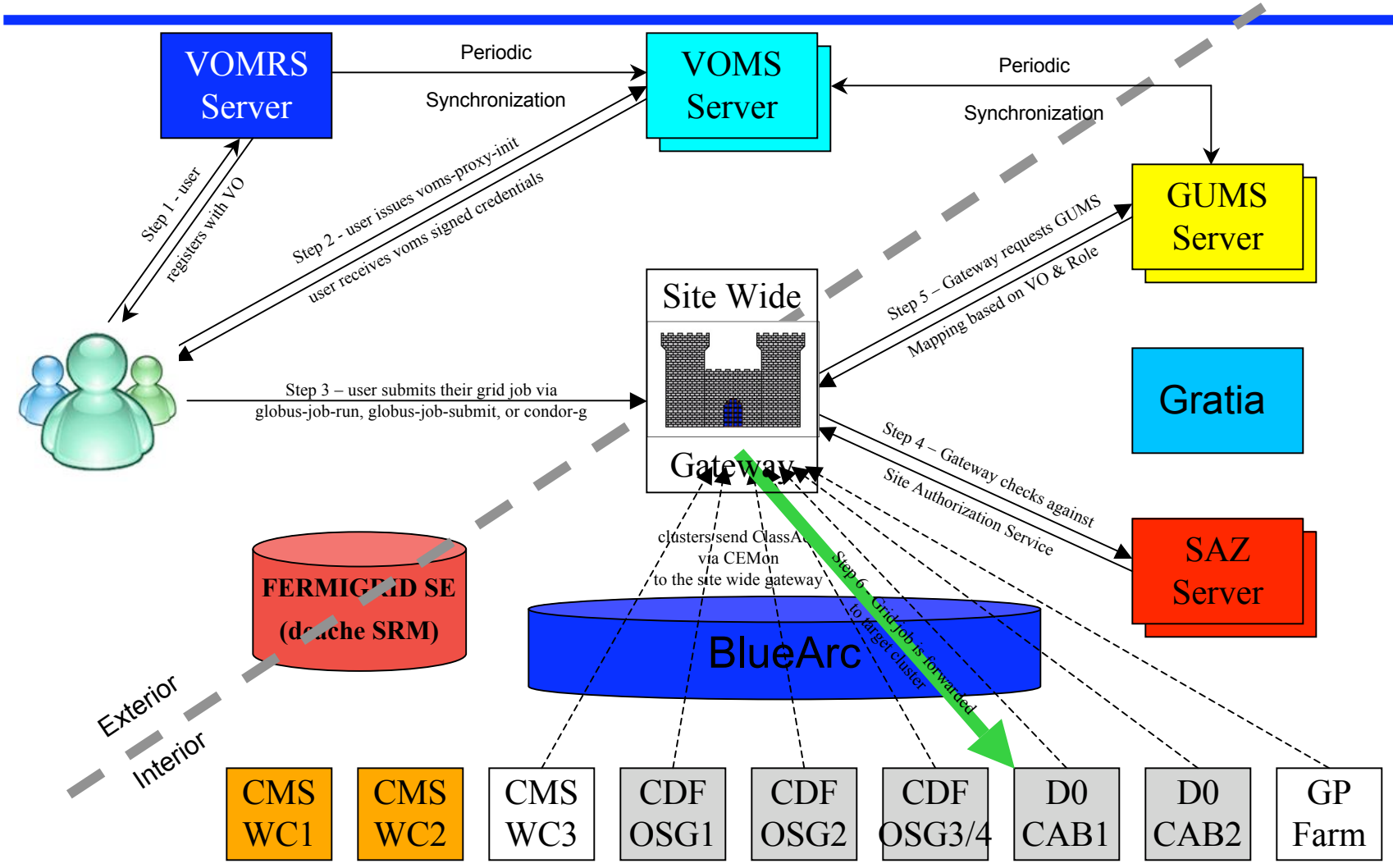
- The Fermilab Grid department provides production infrastructure and software, together with leading edge and innovative computing solutions, to meet the needs of the Fermilab Grid community and takes a strong leadership role in the development and operation of the global computing infrastructure of the Open Science Grid.

Accomplishments:

- FermiGrid is the coherent integration of multiple heterogeneous Fermilab resources into a common Campus Grid.

- The architecture of FermiGrid facilitates seamless interoperation of the Fermilab resources with the resources of the other regional, national and international Grids.

FermiGrid - Architecture



FermiGrid - Advantages

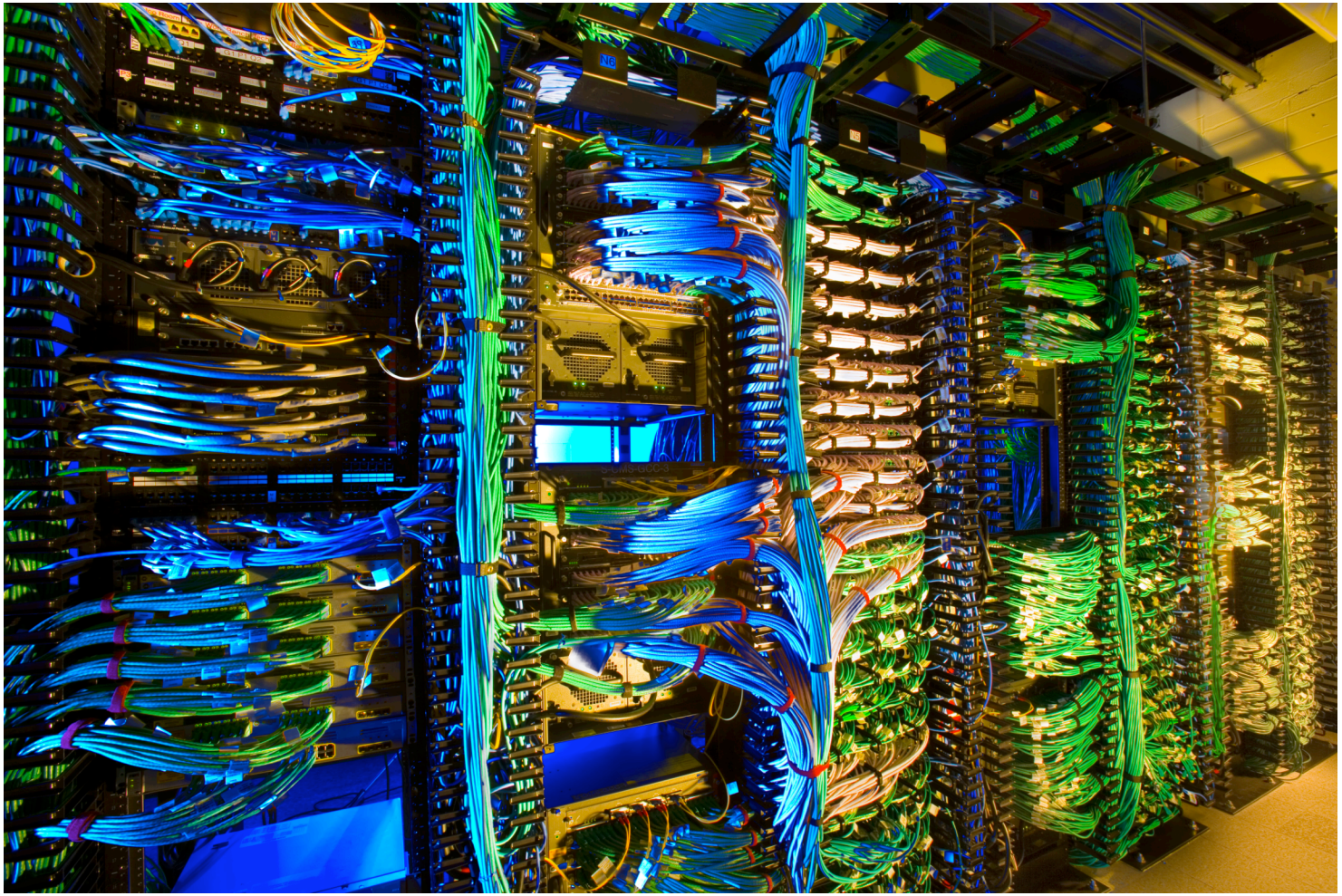
FermiGrid is the fulcrum by which Fermilab leverages it's contributions to the larger regional and national Grids.

Common site gateway for Fermilab Grid clusters.

- This single "portal" allows Fermilab to efficiently operate multiple Grid resources. Supports pre web services and web services GRAM.
- Performs matchmaking (resource selection) between the requirements of the users job and the available batch slots across multiple Fermilab Grid clusters.
- The users job is then transparently forwarded to the "best" Fermilab Grid cluster.

The FermiGrid shared common infrastructure allows for significant cost reductions and increased efficiency.

FermiGrid - Enabling Access to Computers



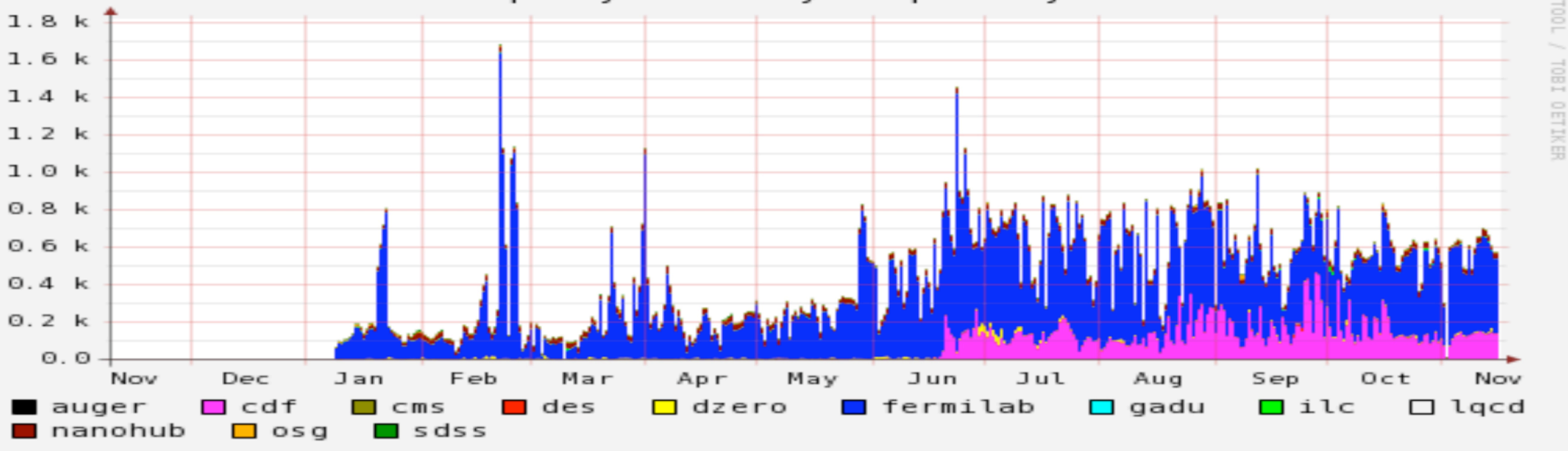
FermiGrid - Enabling Access to Storage

FermiGrid - Service Requests per Day

Service Name	Service Description	Average Requests/Day*	Maximum Requests/Day
VOMS	Virtual Organization Management Service	610	1,700
GUMS	Grid User Mapping Service	231,000	1,028,388
SAZ	Site AuthoriZation Service	176,000	438,193
Squid	Web Proxy Cache	1,975,000	5,625,194

* the average number of service requests/day for the past month.

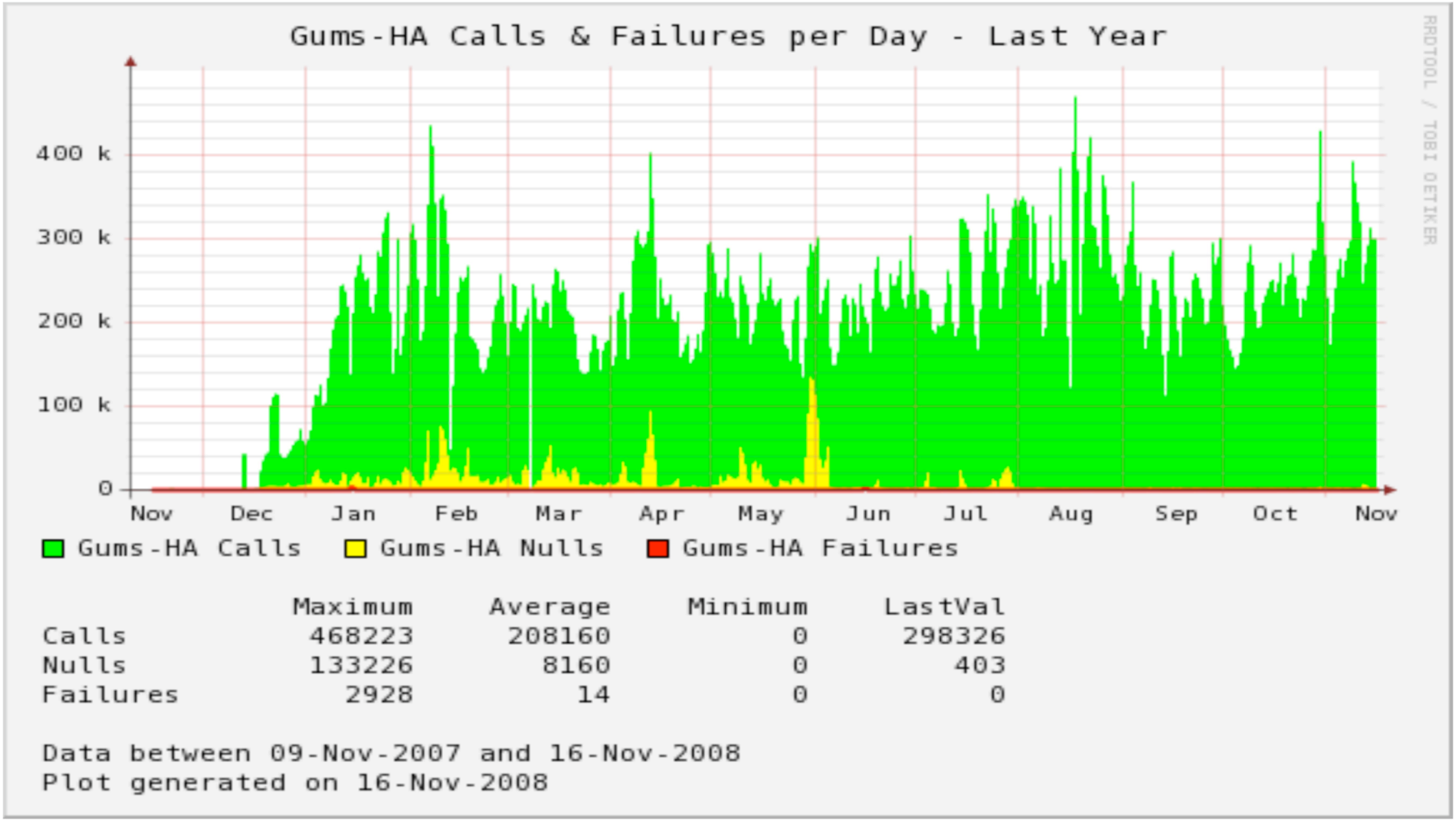
FermiGrid - Virtual Organization Management Service



	Maximum	Average	Minimum	LastVal
auger	0	0	0	0
cdf	463	60	0	137
cms	8	0	0	0
des	19	0	0	0
dzero	49	4	0	1
fermilab	1634	289	0	399
gadu	0	0	0	0
ilc	24	1	0	0
lqcd	0	0	0	0
nanohub	41	19	0	27
osg	26	2	0	2
sdss	7	0	0	0

Data between 09-Nov-2007 and 16-Nov-2008
Plot generated on 16-Nov-2008

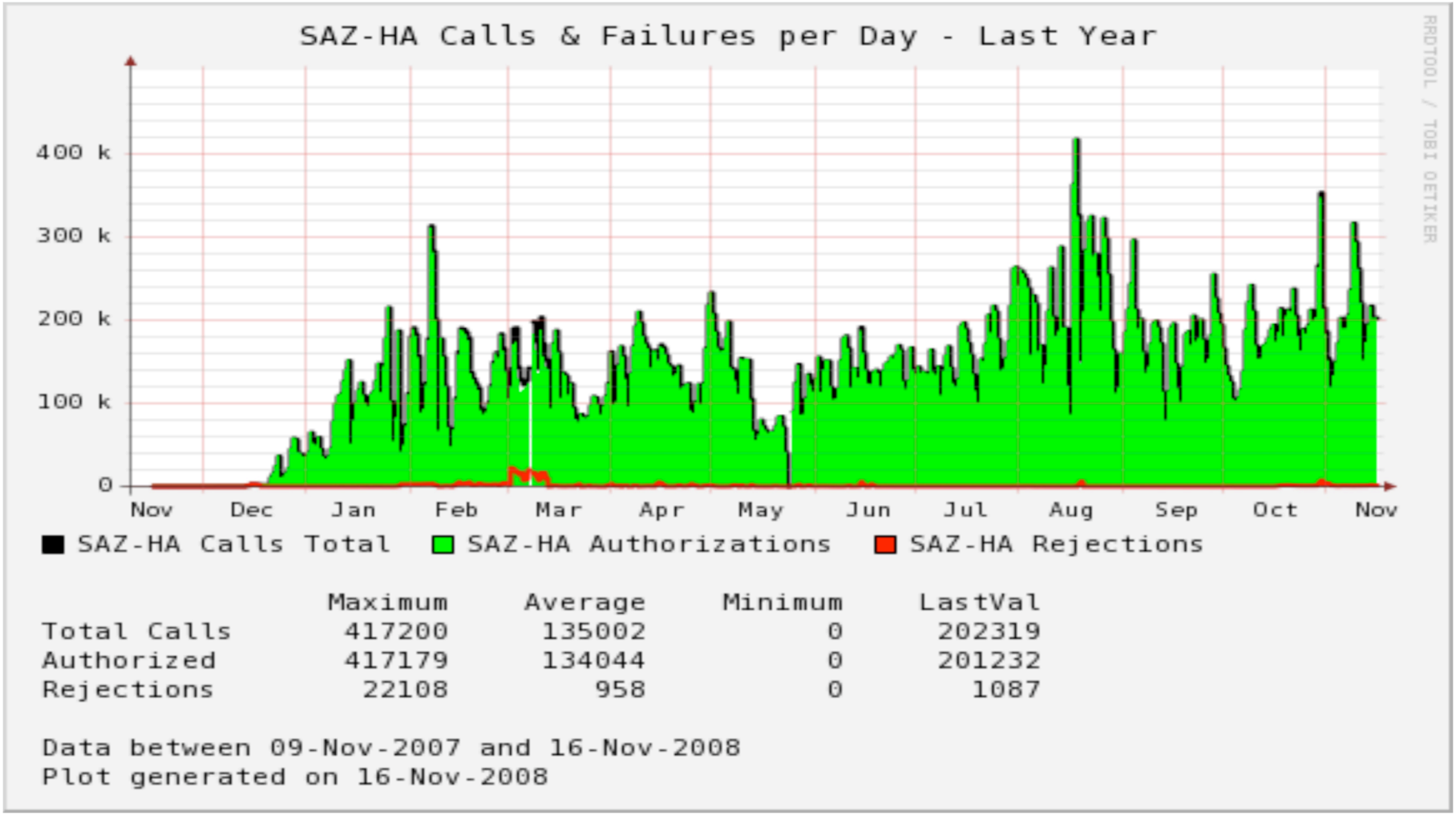
FermiGrid - Grid User Mapping Service (GUMS)



	Maximum	Average	Minimum	LastVal
Calls	468223	208160	0	298326
Nulls	133226	8160	0	403
Failures	2928	14	0	0

Data between 09-Nov-2007 and 16-Nov-2008
Plot generated on 16-Nov-2008

FermiGrid - Site AuthoriZation (SAZ) Service

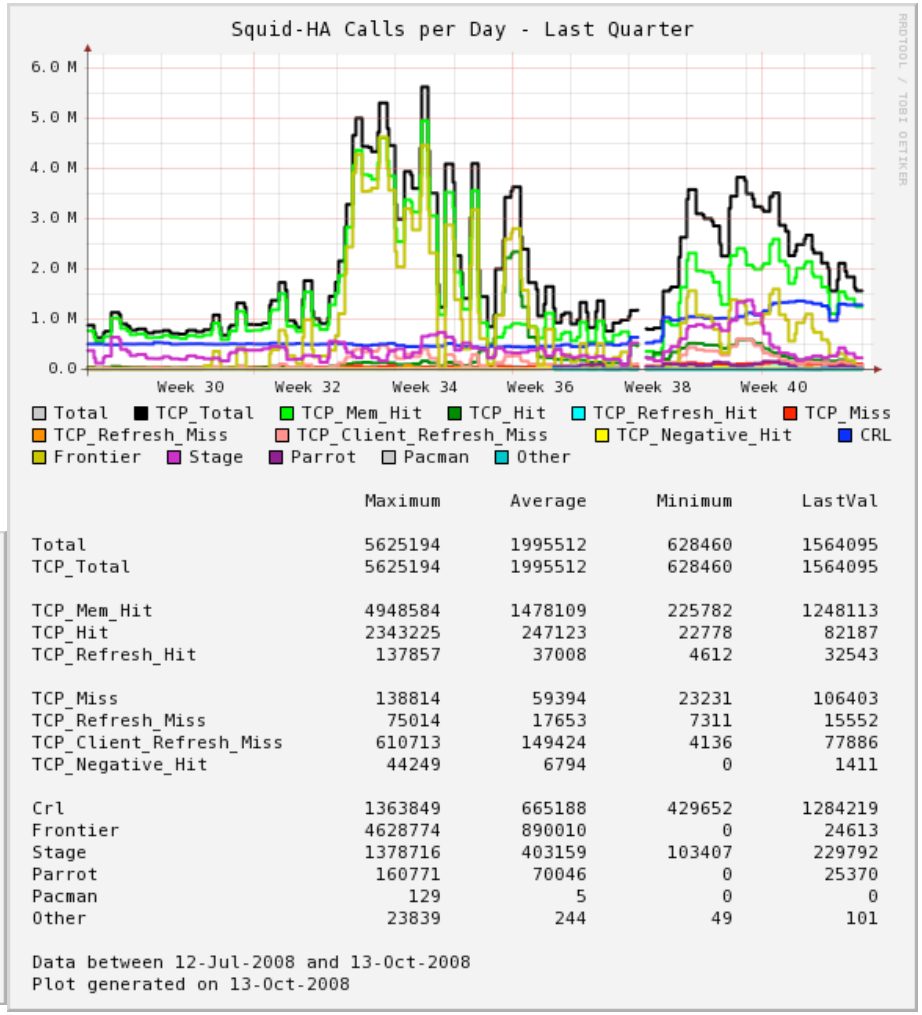


	Maximum	Average	Minimum	LastVal
Total Calls	417200	135002	0	202319
Authorized	417179	134044	0	201232
Rejections	22108	958	0	1087

Data between 09-Nov-2007 and 16-Nov-2008
Plot generated on 16-Nov-2008

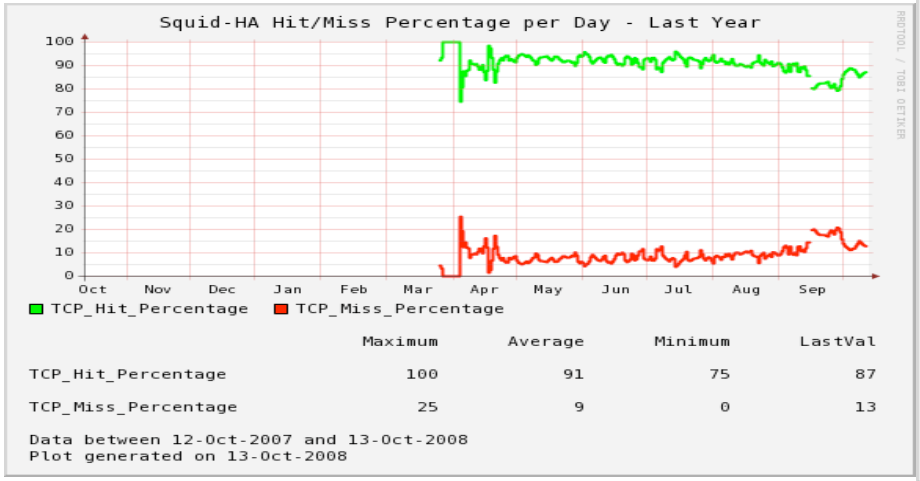
FermiGrid - Squid Web Cache

http://www.squid-cache.org/



	Maximum	Average	Minimum	LastVal
Total	5625194	1995512	628460	1564095
TCP_Total	5625194	1995512	628460	1564095
TCP_Mem_Hit	4948584	1478109	225782	1248113
TCP_Hit	2343225	247123	22778	82187
TCP_Refresh_Hit	137857	37008	4612	32543
TCP_Miss	138814	59394	23231	106403
TCP_Refresh_Miss	75014	17653	7311	15552
TCP_Client_Refresh_Miss	610713	149424	4136	77886
TCP_Negative_Hit	44249	6794	0	1411
Crl	1363849	665188	429652	1284219
Frontier	4628774	890010	0	24613
Stage	1378716	403159	103407	229792
Parrot	160771	70046	0	25370
Pacman	129	5	0	0
Other	23839	244	49	101

Data between 12-Jul-2008 and 13-Oct-2008
Plot generated on 13-Oct-2008



	Maximum	Average	Minimum	LastVal
TCP_Hit_Percentage	100	91	75	87
TCP_Miss_Percentage	25	9	0	13

Data between 12-Oct-2007 and 13-Oct-2008
Plot generated on 13-Oct-2008

FermiGrid-HA - Highly Available Grid Services

The majority of the services listed in the service catalog are deployed in high availability (HA) configuration that is collectively know as “FermiGrid-HA”.

- The goal for FermiGrid-HA is > 99.999% service availability.
- We achieved 99.9969% for the period of 01-Dec-2007 through 30-Jun-2008.

FermiGrid-HA utilizes three key technologies:

- Linux Virtual Server (LVS).
- Xen Hypervisor.
- MySQL Circular Replication.

We are actively working on the necessary configuration research that will allow us to deploy **ALL** services in the service catalog in a HA configuration:

- Gatekeeper-HA
- MyProxy-HA
- Gratia-HA/HP [OSG Accounting Service]
- ReSS-HA [OSG Resource Selection Service]

FermiGrid - Strategies to Deploy HA Services

FermiGrid employs several strategies to deploy HA services:

- Trivial monitoring or information services (such as Ganglia and Zabbix) are deployed on at least two independent virtual machines.
- Services that natively support HA operation (Condor Information Gatherer, FermiGrid internal ReSS deployment) are deployed in the standard service HA configuration on at least two independent virtual machines.
- Services that maintain intermediate routing information (Linux Virtual Server) are deployed in an active/passive configuration on at least two independent virtual machines. A periodic heartbeat process is used to perform any necessary service failover.
- Services that do not maintain intermediate context (i.e. are pure request/response services such as GUMS and SAZ) are deployed using a Linux Virtual Server (LVS) front end to active/active servers on at least two independent virtual machines.
- Services that support active-active database functions (circularly replicating MySQL servers) are deployed on at least two independent virtual machines.

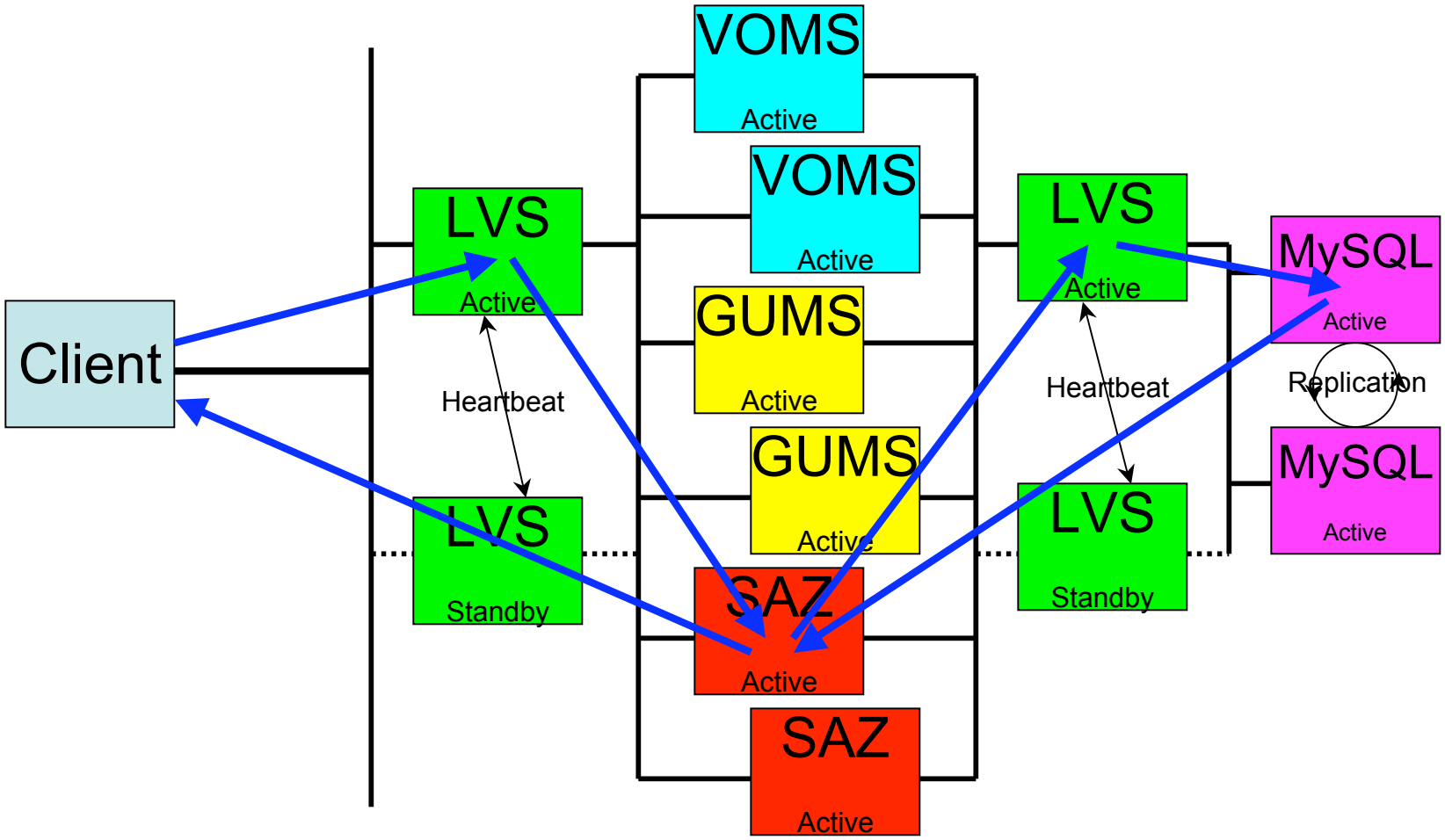
FermiGrid - Non-HA Services

The following services are not currently implemented as HA services:

- Globus gatekeeper services (such as the CDF and D0 experiment globus gatekeeper services) are deployed in segmented pools.
  - Loss of any single pool will reduce the available resources by approximately 50%.
- MyProxy
- Certain Open Science Grid (OSG) services (such as the OSG accounting repository [Gratia] and the OSG Resource Selection Service [ReSS]) are not currently implemented as HA services.
  - If the service fails, then the service will not be available until appropriate manual intervention is performed to restart the service.

These services will be addressed as part of the FermiGrid FY2009 activities.

FermiGrid - Logical View of Services



FermiGrid - Physical View of Services

fermigrid0 Monitoring host	
fg0x0	SL4+32
fg0x1	SL4+32
fg0x2	SL5+32 Future VOMRS
fg0x3	SL5+32 Future RSV
fg0x4	SL5+32
fg0x5	SL4+32
fg0x6	SL5+64
fg0x7	SL4+64

fermigrid1 Web Server	
fg1x0	OSG ST+E Tool ste.opensciencegrid.org (future lvs)
fg1x1	site gateway (primary) fermigridosg1
fg1x2	fginfo (standby)
fg1x3	condor master fermigridcm1
fg1x4	osg-tg gateway (standby)

fermigrid2	
fg2x0	unused (future lvs)
fg2x1	vomrs vomrs.fnal.gov vomrs.opensciencegrid.org future syslog-ng
fg2x2	ganglia zabbix
fg2x3	squid
fg2x4	myproxy (new)

fermigrid3	
fg3x0	unused (future lvs)
fg3x1	syslog-ng
fg3x2	ganglia zabbix
fg3x3	squid
fg3x4	myproxy (new)

fermigrid4	
fg4x0	myproxy (old) (future lvs)
fg4x1	site gateway (standby)
fg4x2	fginfo (primary)
fg4x3	condor master
fg4x4	osg-tg gateway (primary) tg.opensciencegrid.org

fermigrid5	
fg5x0	lvs
fg5x1	voms
fg5x2	gums
fg5x3	saz
fg5x4	mysql

fermigrid6	
fg6x0	lvs
fg6x1	voms
fg6x2	gums
fg6x3	saz
fg6x4	mysql

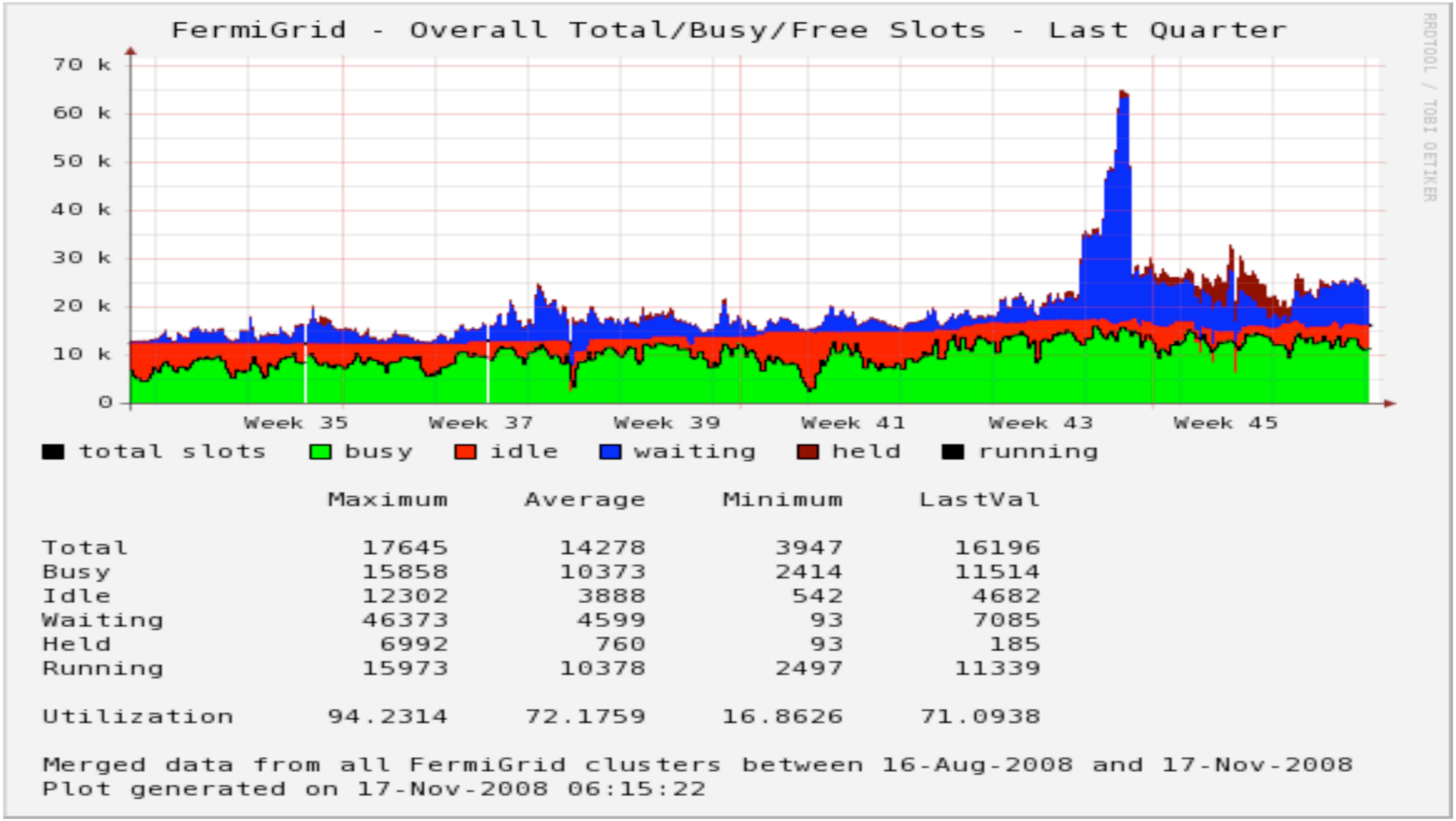
http://fermigrid.fnal.gov/fermigrid-organization.html

FermiGrid - Service Availability

	Last Week	Last Month	Last Quarter	Since 10-Jul-2008
Hardware	100.000%	100.000%	100.000%	100.000%
Core Services	100.000%	100.000%	99.970%	99.970%
Gatekeepers	100.000%	98.899%	98.843%	99.039%
Batch Services	99.953%	99.979%	99.916%	99.906%
ReSS	100.000%	100.000%	99.943%	99.597%
Gratia	99.688%	99.379%	99.720%	99.756%

The goal for FermiGrid-HA is > 99.999% service availability.

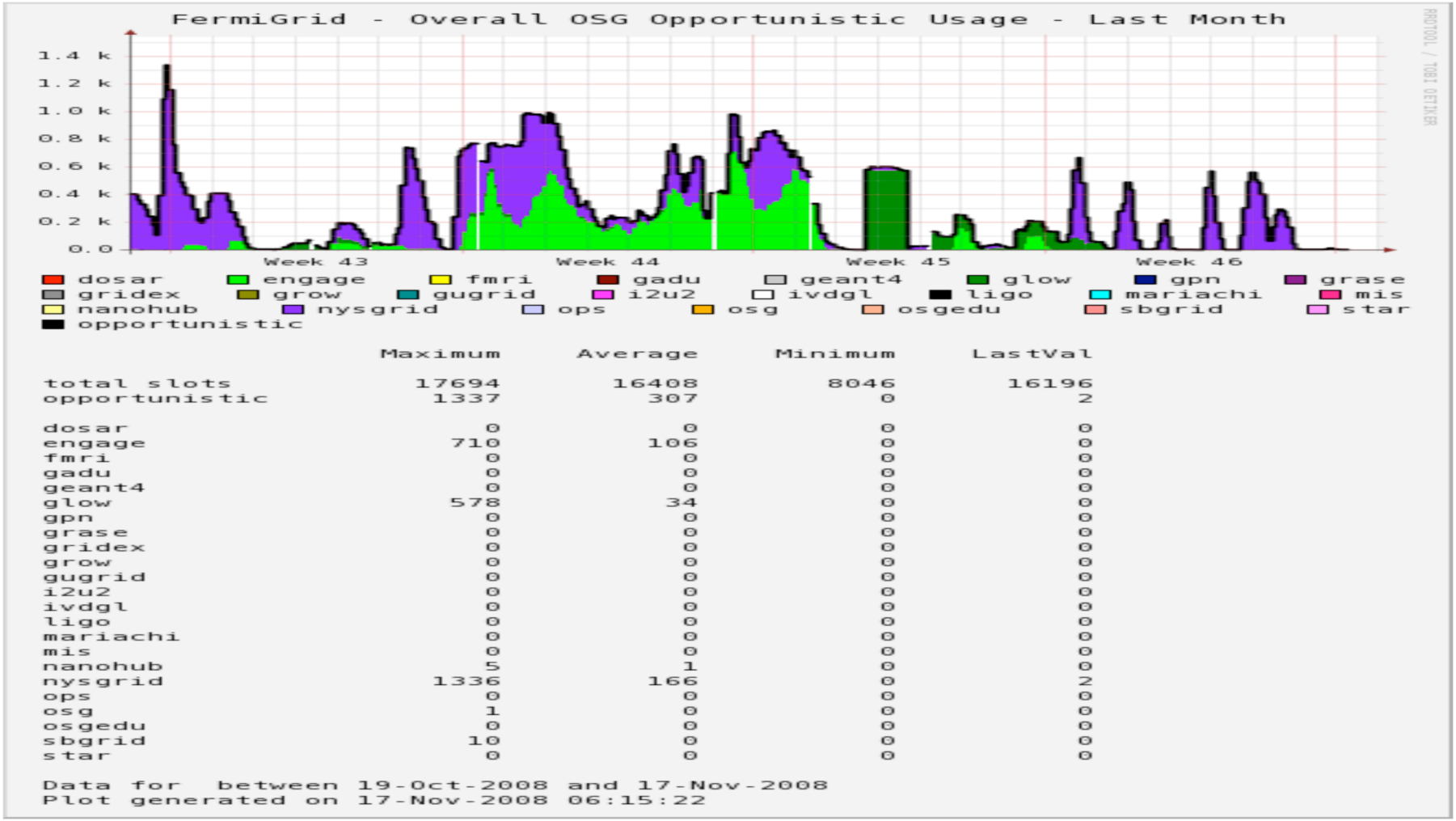
FermiGrid - Overall Utilization - Last Quarter



	Maximum	Average	Minimum	LastVal
Total	17645	14278	3947	16196
Busy	15858	10373	2414	11514
Idle	12302	3888	542	4682
Waiting	46373	4599	93	7085
Held	6992	760	93	185
Running	15973	10378	2497	11339
Utilization	94.2314	72.1759	16.8626	71.0938

Merged data from all FermiGrid clusters between 16-Aug-2008 and 17-Nov-2008
Plot generated on 17-Nov-2008 06:15:22

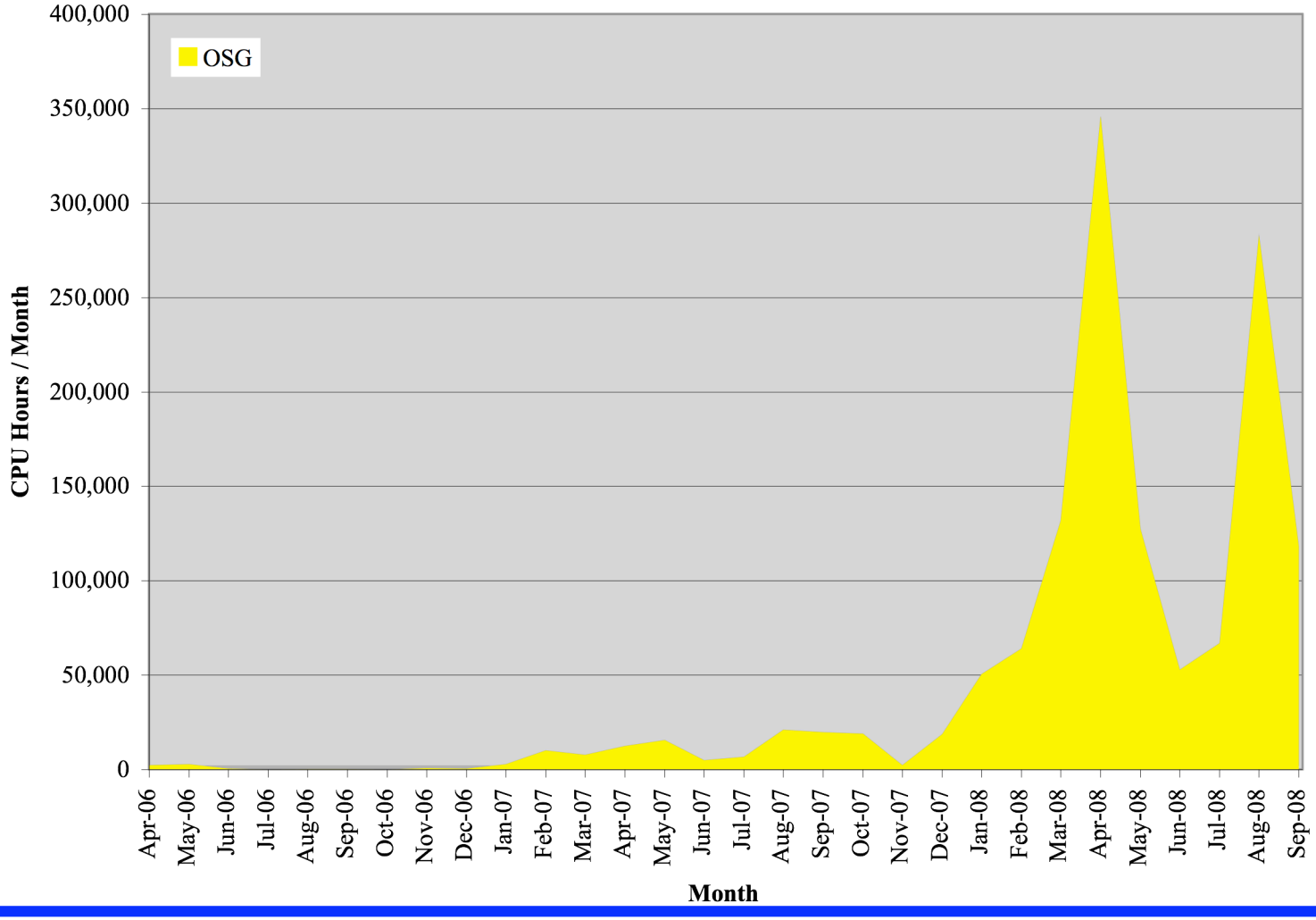
FermiGrid - OSG Opportunistic Usage by VO



	Maximum	Average	Minimum	LastVal
total slots	17694	16408	8046	16196
opportunistic	1337	307	0	2
dosar	0	0	0	0
engage	710	106	0	0
fmri	0	0	0	0
gadu	0	0	0	0
geant4	0	0	0	0
glow	578	34	0	0
gpn	0	0	0	0
grase	0	0	0	0
gridex	0	0	0	0
grow	0	0	0	0
gugrid	0	0	0	0
i2u2	0	0	0	0
ivdgl	0	0	0	0
ligo	0	0	0	0
mariachi	0	0	0	0
mis	0	0	0	0
nanohub	5	1	0	0
nysgrid	1336	166	0	2
ops	0	0	0	0
osg	1	0	0	0
osgedu	0	0	0	0
sbgrid	10	0	0	0
star	0	0	0	0

Data for between 19-Oct-2008 and 17-Nov-2008
Plot generated on 17-Nov-2008 06:15:22

FermiGrid - OSG Opportunistic Usage Summary



FermiGrid - CDF, D0, CMS, GP Cluster Utilization

	CDF	D0	CMS	GP	Total
# of Clusters	3	2	1	1	7
# of Gateway Systems	4	2	4	2	12
# of Batch Head Nodes	3	2	1	1	7
# of Worker Nodes	718	1,148	1,130	204	3,200
Total job slots	4,785	5,056	6,800	1,072	17,390
Maximum running jobs	4,128	5,054	6,158	960	15,022
Average running jobs	2,777	3,711	2,751	440	10,438
Maximum queued jobs	1,905	5,537	45,872	1,822	46,854
Average queued jobs	582	1,418	1,169	213	4208

FermiGrid - Plans for the Future

Ongoing software upgrades.

Routine expansion.

Extend HA service deployment.

Work with OSG and Globus to extend the FermiGrid SAZ service into a Grid wide banning tool.

Transition from HA to RS in FY2010.

Investigation of a Grid MPI capability.

Investigation of “Cloud Computing”.

FermiGrid - Summary

FermiGrid-HA is in operation offering high availability services.

- LVS, Xen and MySQL circular replication have given us significant cost savings.

The ~17.5K Batch slots across FermiGrid have an average utilization of ~72% with peaks of >94% utilization.

- Handle >46K queued batch jobs without incident.

FermiGrid regularly processes >2M Grid service requests per week.