

UCSD in IRIS-HEP DOMA

Frank Würthwein

SDSC/UCSD

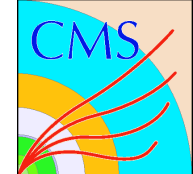
IRIS-HEP DOMA Meeting

June 7th 2018

Big Picture

- We have short term and long term objectives.
- Short term:
 - Work on issues related to globus replacement
 - E.g. benchmarking http transfers
 - Arriving at a satisfactory Xcache release for stable operations for ATLAS & CMS
 - E.g. benchmarking candidates in the Petabyte scale Caltech/UCSD caching prototype
- Long Term:
 - Understanding how to make the HL-LHC computing needs not blow up the budget
 - E.g. Data Analytics and modelling effort

People

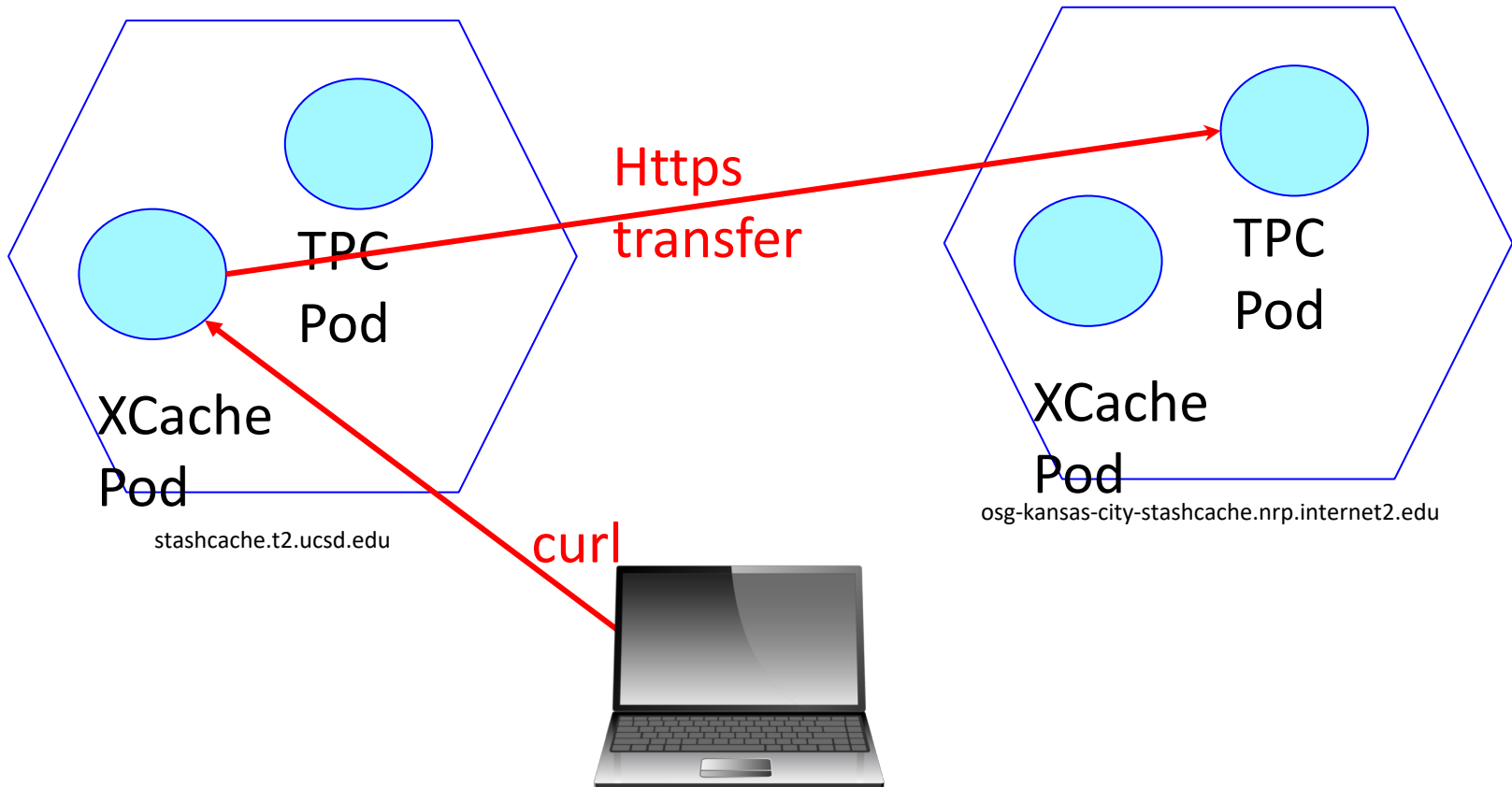


- Edgar/Diego/Igor
 - Computing professionals
 - We have 2/3 of an FTE in IRIS-HEP total.
 - Nominally, this is split between SSL and DOMA.
- Ilan Scheinkman
 - UCSD undergraduate that has a summer research fellowship to work full time with us.
 - He also has worked with us in his spare time throughout the year on benchmarking and other short term objectives.
- Jonathan Guiang
 - UCSD graduate student starting Fall 2019. He has a GSOC fellowship for the summer to work on data analytics.
- There are 4 other undergraduates involved at levels where it's as yet unclear whether or not it provides value to IRIS-HEP.

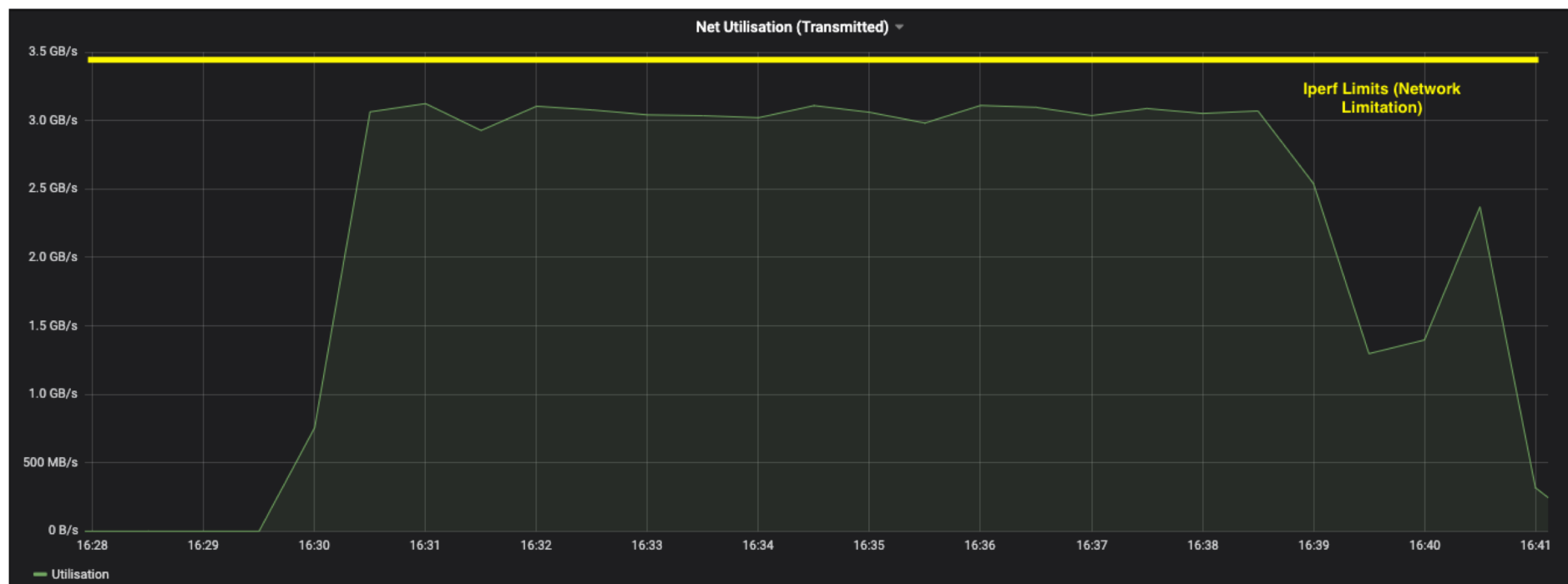
Benchmarking http transfers

- Edgar/Ilan doing the work.
- Edgar is presenting at Lyon workshop next week on this. I'm taking some info from his draft slides here:
- https://docs.google.com/presentation/d/1Z-NYvbaTF4Ni-AlmAr1zb0HgjnGse77C08LEnC7ItPE/edit?ts=5cf5a3bf#slide=id.g25f6af9dd6_0_0

Benchmarking



This is NVME to NVME on hosts with 100Gbps connectivity



The **yellow** line shows the limits of doing Iperf3. Between same hosts.
Different pods

Testing Xrootd HTTP TPC architecture

We do not understand (yet) why only ~25Gbps, even with Iperf.

Data Analytics (I)

- Working within WLCG DOMA Access working group as context.
- Working with global CMS data analytics working group.
- Diego presented proposal for making data public at last Tuesday DOMA Access meeting.
 - Various topics are being pursued, all ultimately surrounding DOMA Access issues towards an understanding of the utility of “Data Lake” concept.
- Fkw gave 2 presentations at DOMA Access so far.
 - On HL-LHC expected data use by CMS, and how that maps onto Data Lake concept.
 - On the benefits for a site like UCSD from the concepts discussed.

Data Analytics (II)

- Everything we do presently is looking at data organization and management, rather than access issues, i.e. is at the level of files, blocks of files, datasets, and annual campaigns.
- In the future (maybe in 2020?), we'd like to go back and work on lower level things.
 - Object reuse
 - Object use predictability
 - Intelligent storage

Activities in 2020 and beyond

- Once we have a stable XRootD release
 - Stable & performant
 - Includes monitoring dashboard that meets operational needs.
 - Includes logging of monitoring to support data analytics
 - Has the features necessary to replace gridftp for bulk transfers, and Xcache for data access.
- We are interested in pursuing other things, e.g.
 - Can a case be made for on the fly compression changes within context of iDDS, for example.