Distributed Computing R&D

progress report

Fabio HERNANDEZ *on behalf of*

IHEP: CHEN Gang, LI Weidong, QI Fazhi, WANG Lu, ZHANG Xiaomei, CHEN Yaodong, YAN Tian, SHI Jingyan, DU Ran, ZENG Shan, WANG Cong
CC-IN2P3: Ghita RAHAL, Vanessa HAMAR, Laurent CAILLAT-VALLET, Frédéric SUTER, Nicolas FOURNIALS
CPPM: Andreï TSAREGORODTSEV

10th FCPPL Workshop — Beijing, March 30th, 2017

Background

- Guiding principle

  to explore technologies of potential interest for the data processing needs of HEP experiments

- Partners

  CPPM, IN2P3 computing center, IHEP computing center

- Funding

  IHEP and IN2P3 through FCPPL 2016 call

  CNRS-NSFC joint program for international collaboration

Topics

- Topics of interest

  DIRAC-based computing platform for IHEP experiments

  High-Performance computing platforms

  Building blocks for inter-site bulk data transfer

  Understanding I/O behaviour of applications

  Experimentation with software-defined networking

DIRAC-based platform

DIRAC-based platform

- DIRAC instance at IHEP in production since several years

  currently being used by BES III, CEPC and JUNO

  single instance for various experiments: makes easier to add other experiments to the distributed infrastructure, lowers maintenance effort

- Single, multi-experiment storage element

  Lustre-managed local storage also exposed via grid protocols by StoRM

  access control and quotas based on grid certificates and VOMS attributes

- Example of recent usage

  CEPC simulation: 9.6 M events, 170 TB of data

DIRAC-based platform (cont.)

- Redesign and reimplementation of support of virtual machines by DIRAC: VMDIRAC 2.0

  contribution of IHEP team, now merged into official DIRAC source code

- Benefits of VMDIRAC 2.0

  scheduling of tasks for execution on (OpenStack) virtual machines: more reliable scheduling policy

  virtual machine started and stopped as a function of workload

  transparent to end-user

  greatly simplifies the integration of new cloud sites

  support of execution of multi-core applications in virtual machines

- In production at IHEP

  to be recommissioned at CC-IN2P3 after changes to OpenStack interfaces

High-performance platforms

High-performance computing platforms

- Both IHEP and CC-IN2P3 deployed pilot high-performance computing platforms

  IHEP (target): 1000 CPU cores, 150 GPUs, 50 Intel Xeon PHIs, Infiniband interconnection

  CC-IN2P3: 160 CPU cores and 40 NVIDIA TESLA K80 GPUs in 10 hosts, Infiniband interconnection

  programming environment: CUDA, OpenCL, OpenMP, MPI

- Early adopters getting familiar with those platforms

  the platforms are integrated to the workload management systems of both sites

- Sites understanding how to operate them and attracting users to these new facilities

  experience to be shared among sites operators

File metadata storage

File metadata storage

- Exploratory work being conducted at IHEP

- Goal: to explore suitable alternative storage platforms for handling file metadata

  motivating use case: number of objects managed by file systems rapidly increasing — metadata management becoming bottleneck in some cases

- RAMCloud (Stanford)

  aggregation of RAM of several machines in a cluster with fast interconnection

  key-value programming model on top of which indices, tables and transactions are built

  strong consistency guarantees, low latency, persistence

File metadata storage (cont.)

- Testbed deployed at IHEP

  3 servers, one client

  both Ethernet and Infiniband interconnection

- Performance measurements

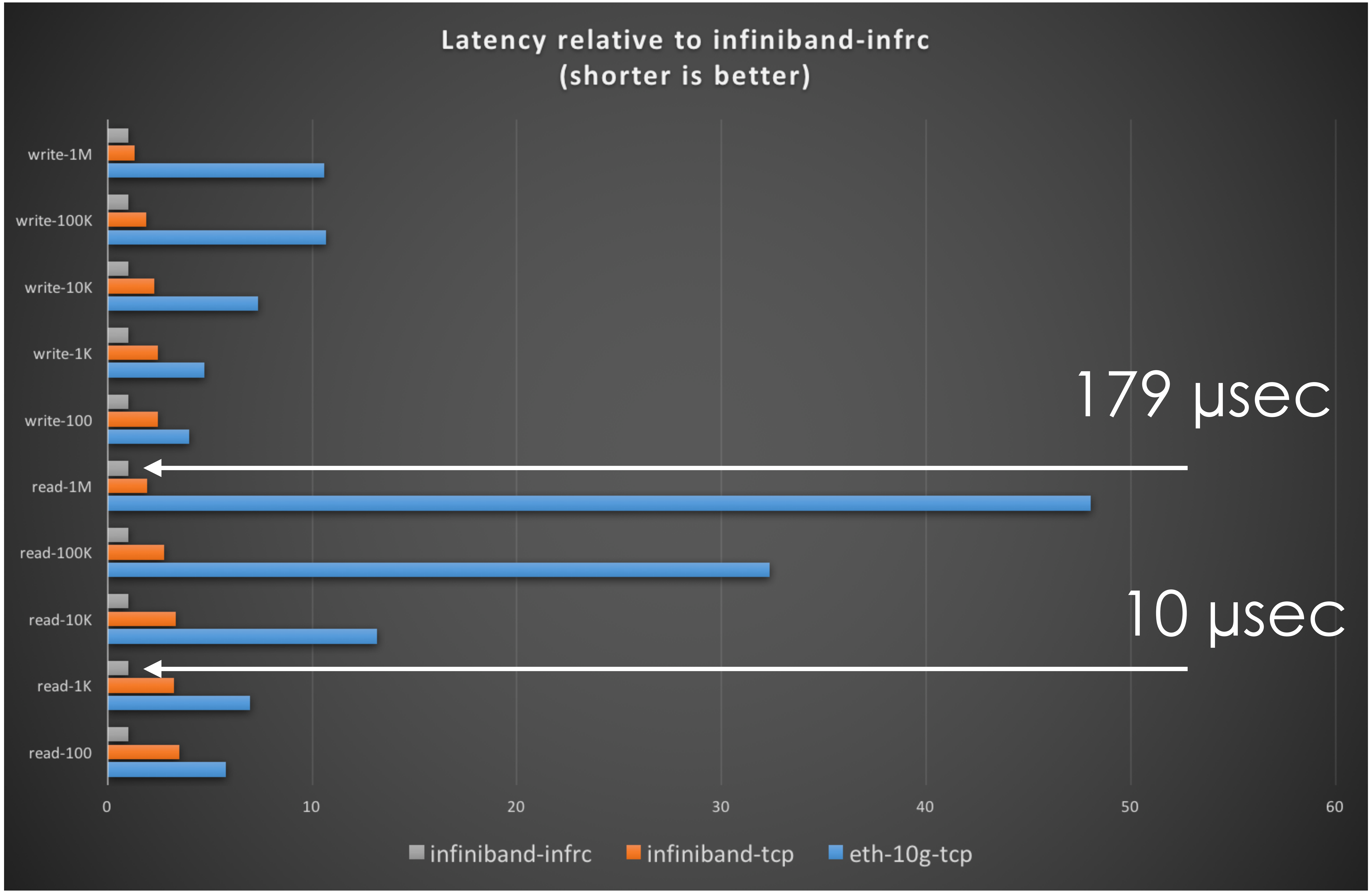
  using TCP over both Ethernet and Infiniband

  using proprietary infrc protocol over Infiniband

- Next steps

  to adapt metadata management component of IndexFS to exploit RAMCloud for storage

File metadata storage (cont.)

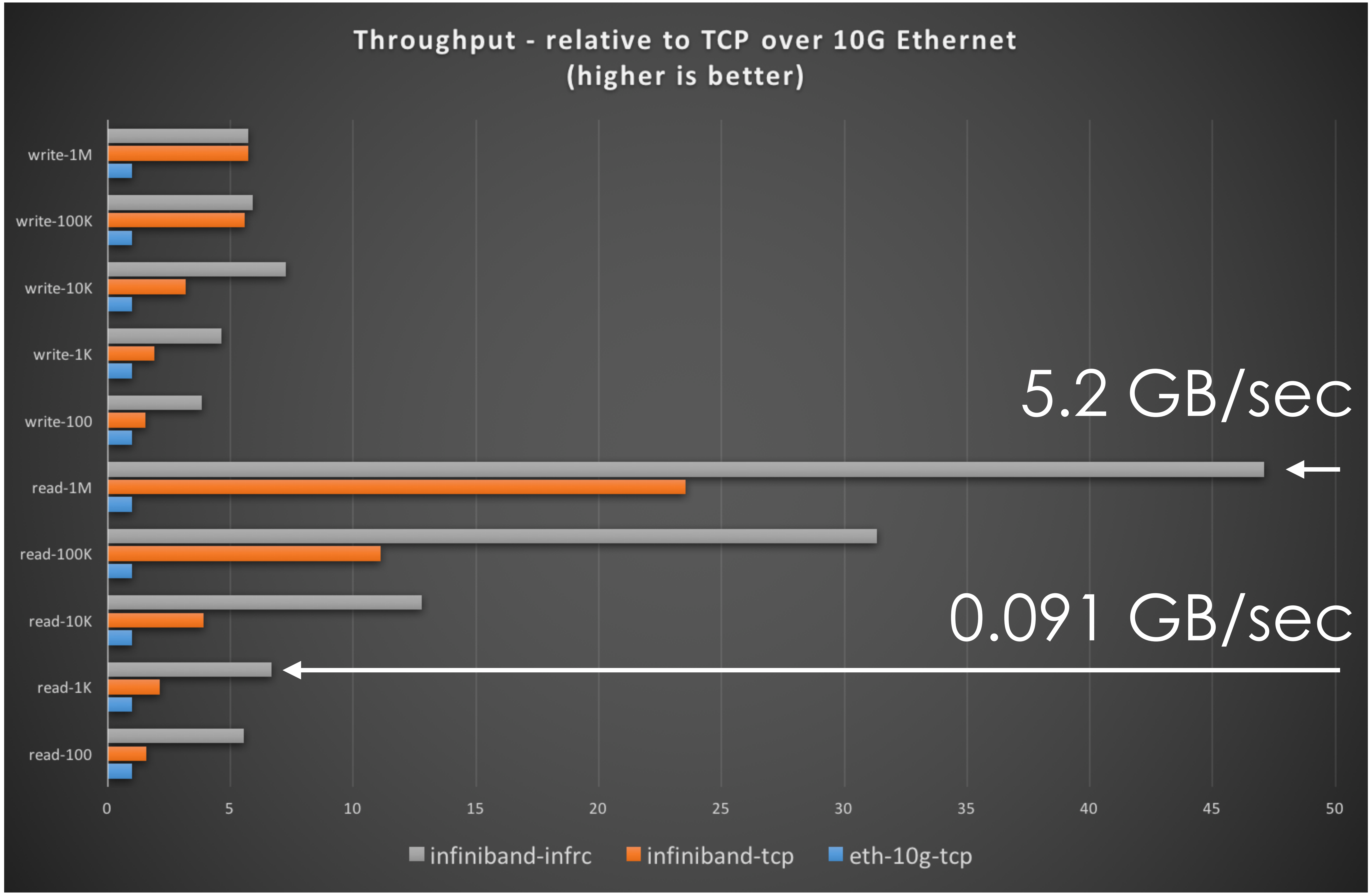


179 µsec

10 µsec

Source: WANG Lu

File metadata storage (cont.)



5.2 GB/sec

0.091 GB/sec

Source: WANG Lu

Software-defined networking

Software-defined networking

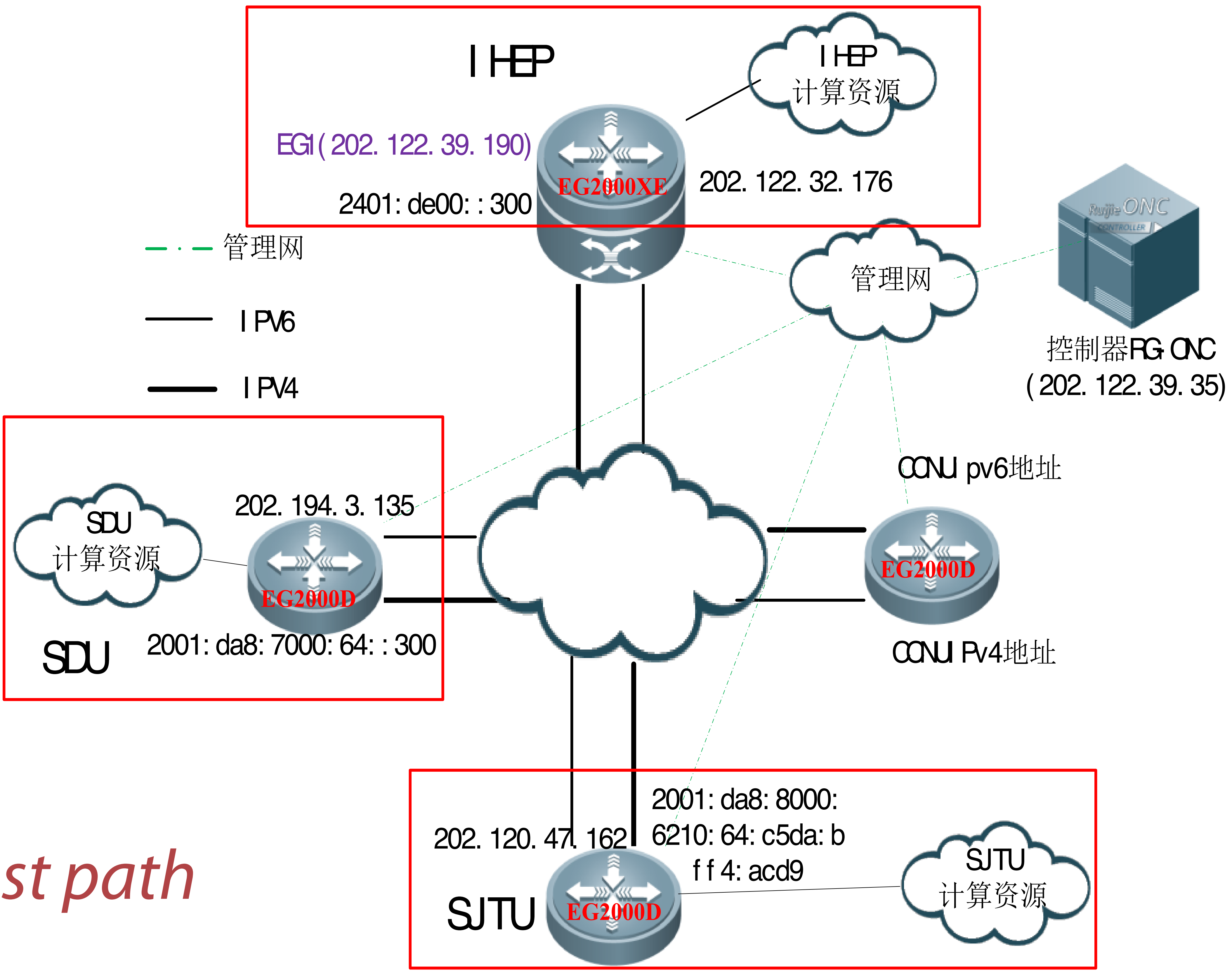
- Continuation of work presented last year
- Advanced experimentation with SDN technologies for 3 use cases

  performed by IHEP experts

- Use-case #1

  based on observed status of network links of 4 sites, dynamically reprogram the network for choosing the best path (IPv4 or IPv6 or aggregation of both)

  4 participating sites: HEP, Shanghai Jiaotong Univ., Shandong Univ., China Central Normal Univ.



Source: QI Fazhi

Software-defined networking (cont.)

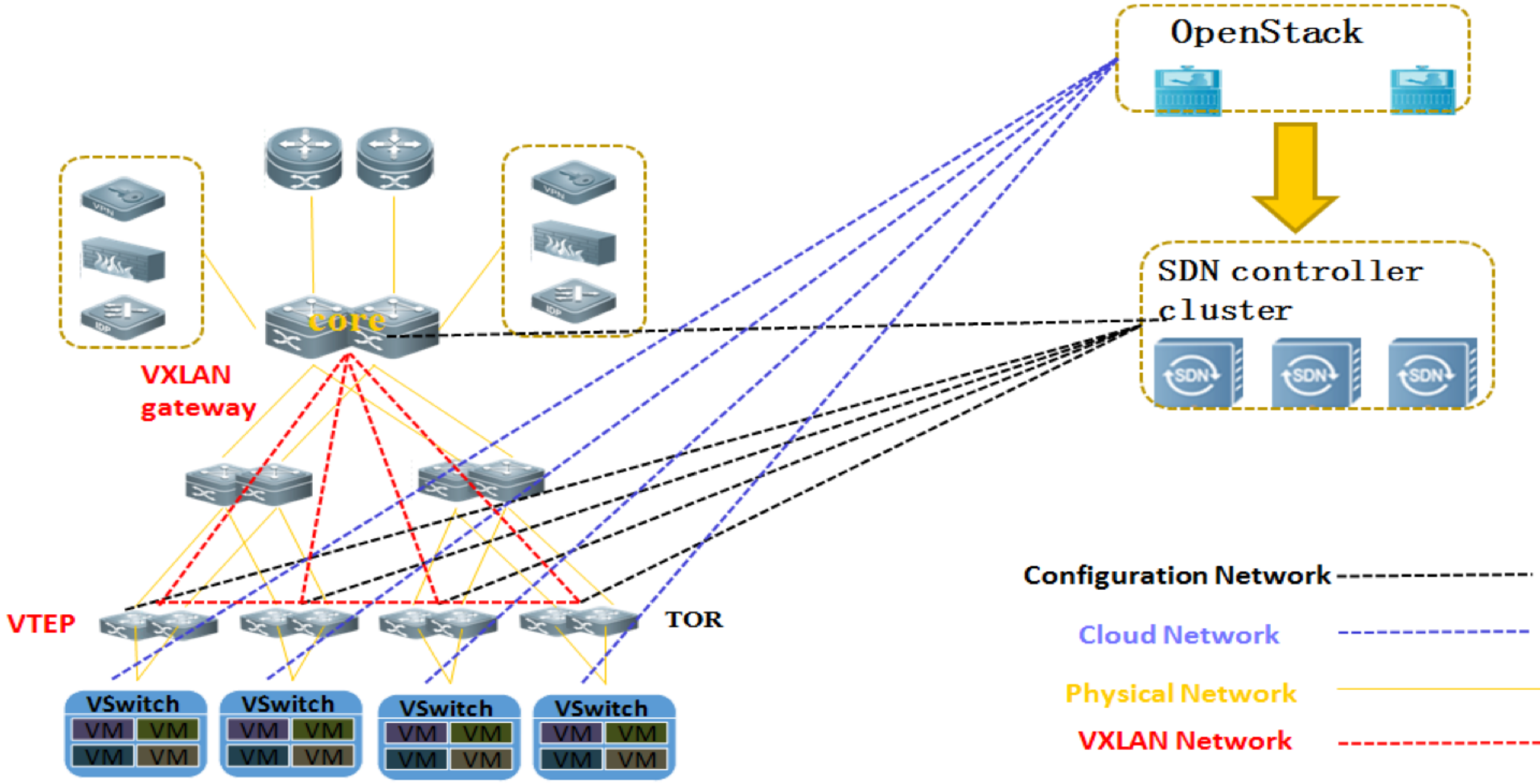
- Use case #2

  to provide flexibility in the network configuration of OpenStack virtual machines

  tenant can specify the bandwidth required on a per-VM basis

  underlying network is programmed according to available physical bandwidth and VM requirements

  experimentation ongoing with IHEP OpenStack installation



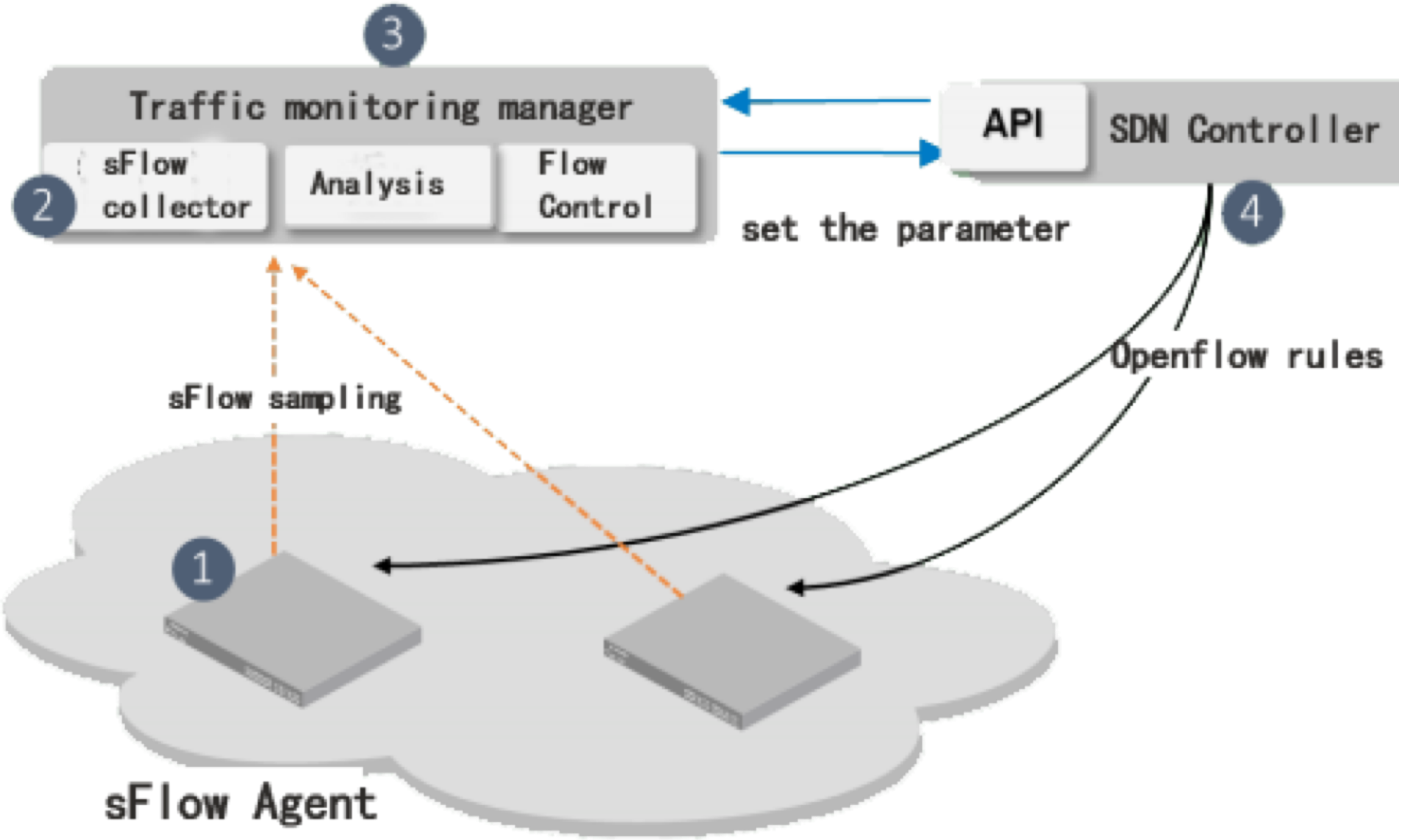
Source: QI Fazhi

Software-defined networking (cont.)

- Use case #3

  to dynamically program the the control path of network routers (via OpenFlow) based on observed behaviour capturing samples of packets (via SFlow)

  goal is to detect abnormal behaviour (e.g. network attacks) and react promptly by reprogramming the border routers



- Fazhi QI visited CC-IN2P3 in April 2016 and shared IHEP experience using SDN technology

  https://indico.in2p3.fr/event/13089

Source: QI Fazhi

Intersite data transfer

Intersite data transfer

- Goal: is HTTP suitable for bulk data transfer over high latency network?

  optimise for throughput, not latency

- Why HTTP?

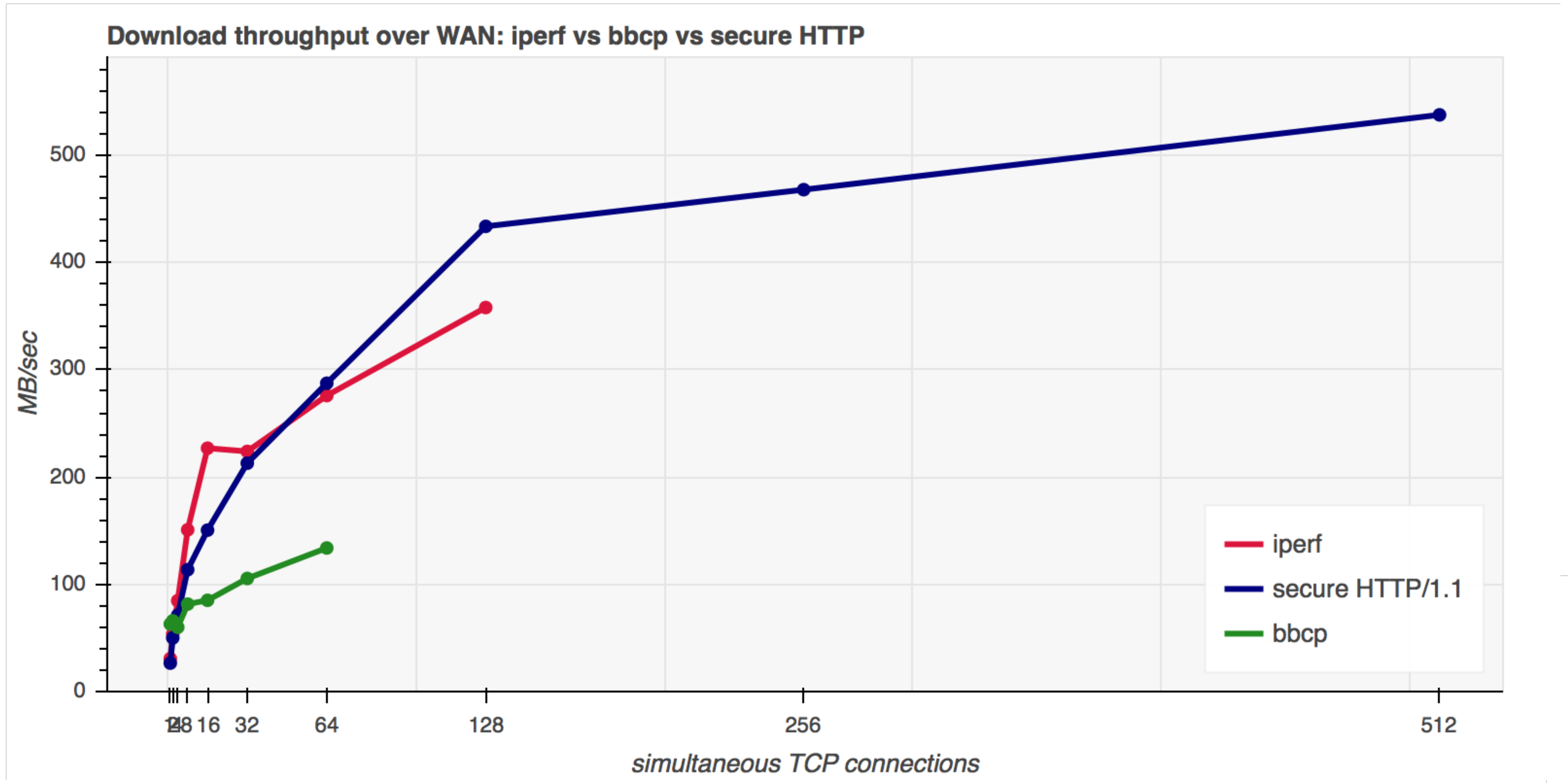
  standard, programmability of both client and servers in any relevant programming language, future-proof, ubiquitous, customisable semantics, …

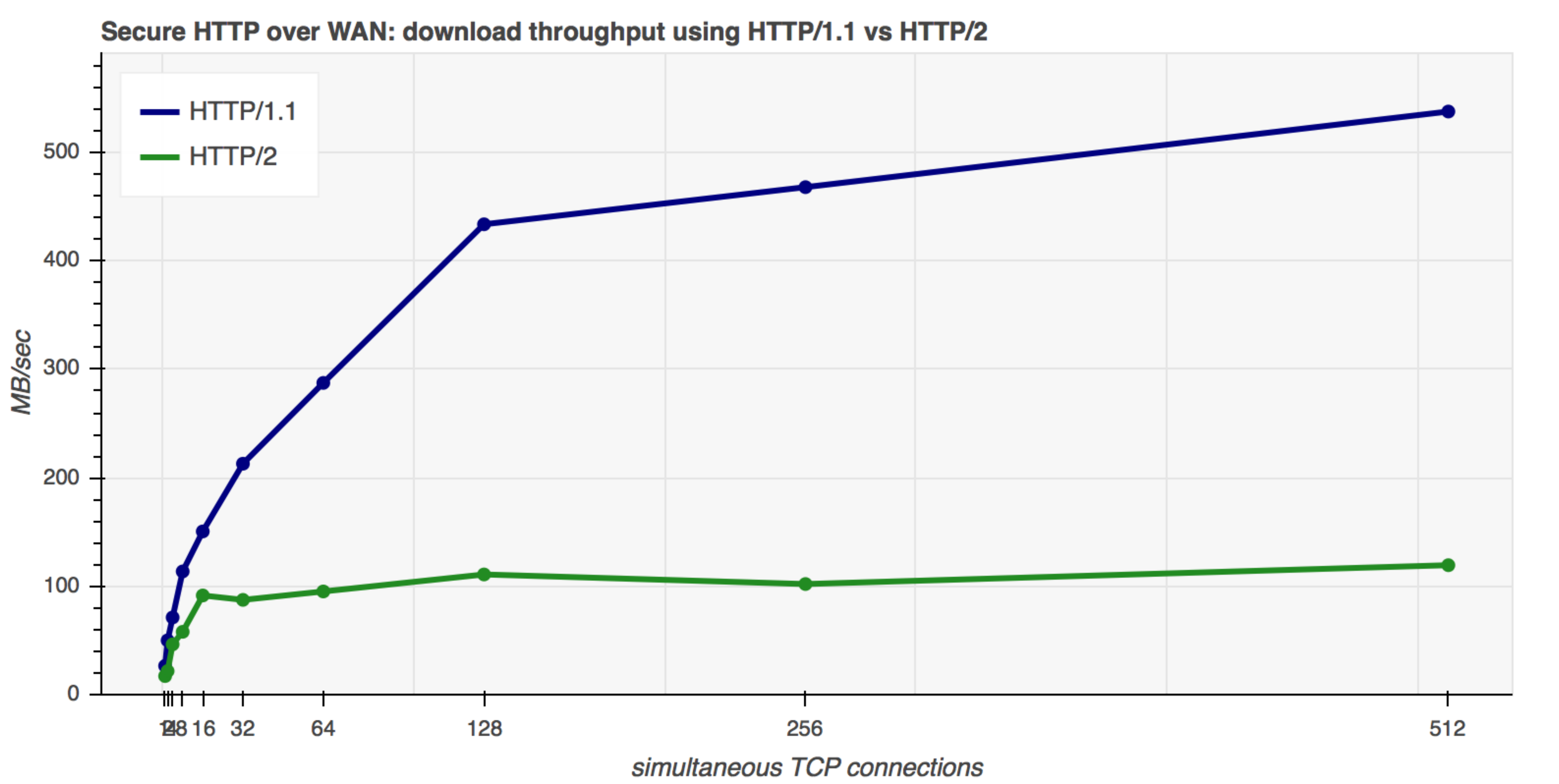
  confidentiality and data integrity ensured using standard TLS

- Developed building blocks for measuring memory-to-memory performance of data transport over HTTP

  over transatlantic network, 10 Gbps, 110 ms RTT

Intersite data transfer (cont.)





Analysis of I/O patterns

Analysis of I/O patterns

- Work performed by WANG Cong during her stay at CC-IN2P3
- Goal: understand the I/O behaviour of applications with the goal of simulate it
- Very preliminary results presented at CHEP 2016 (poster)

  http://indico.cern.ch/event/505613/contributions/2230984
- Additional areas of work identified, but this subject is on stand by because lack of manpower

People

People

- Fabio HERNANDEZ visited IHEP, January 2016
- Visit of QI Fazhi, LI Haibo and SUN Zhuhui to CC-IN2P3, April 2016

  http://indico.in2p3.fr/e/sdn
- ZHANG Xiaomei attended 6th DIRAC Users Workshop in France and visited CC-IN2P3, May 2016

  https://indico.cern.ch/event/477578/
- Fabio HERNANDEZ gave an invited talk at the 3rd Data Science conference in Shanghai, August 2016

  http://dc2016.codata.cn/
- CHEN Gang and LI Jianhui (CNIC) visited CC-IN2P3, March 2017

  https://indico.in2p3.fr/event/14339/
- Andrei TSAREGORODTSEV to visit IHEP, March 2017

Perspectives

Perspectives

- Project submitted to FCPPL 2017 call
- Topics

  DIRAC: improved integration of cloud and high performance computing resources

  hybrid computing platforms (CPU, GPU, accelerators)

  inter-site bulk data transfers

  RAM-based data stores for metadata management

  indexation of contents of ROOT files in external database for rapid location of events

  cyber security federation for HEP