



東京大学
素粒子物理国際研究センター
International Center for Elementary Particle Physics
The University of Tokyo



ATLAS
EXPERIMENT

Status of ATLAS Tier 2 Centre at Tokyo/ICEPP

25th Sep. 2025

The 9th Asian Tier Center Forum (ATCF9)

Masahiko Saito, on behalf of the operation team
ICEPP, The University of Tokyo

International Center for Elementary Particle Physics (ICEPP)



ICEPP
The University of Tokyo

- Leading international collaborations in elementary particle physics.
- Our Mission: Unraveling the universe's fundamental laws.

Main projects



ATLAS Experiment



LHC, CERN

*Exploring new physics
at the energy frontier.*

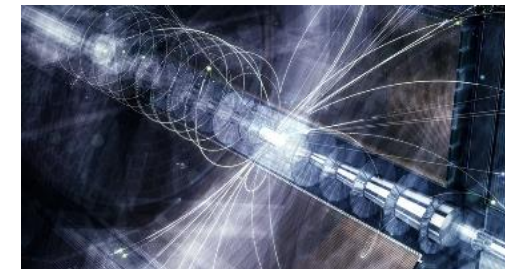


MEG II Experiment



PSI

*Searching for rare decays,
probing beyond the Standard Model.*



ILC Project



Future project

*Precision studies of the Higgs
boson with a lepton collider.*

Quantum AI Technology: *Innovating for future experiments.*

ICEPP's Key Contribution: Tokyo Regional Analysis Center

- ICEPP operates Tokyo Regional Analysis Center for ATLAS/ATLAS-Japan
 - Only computing center for ATLAS in Japan

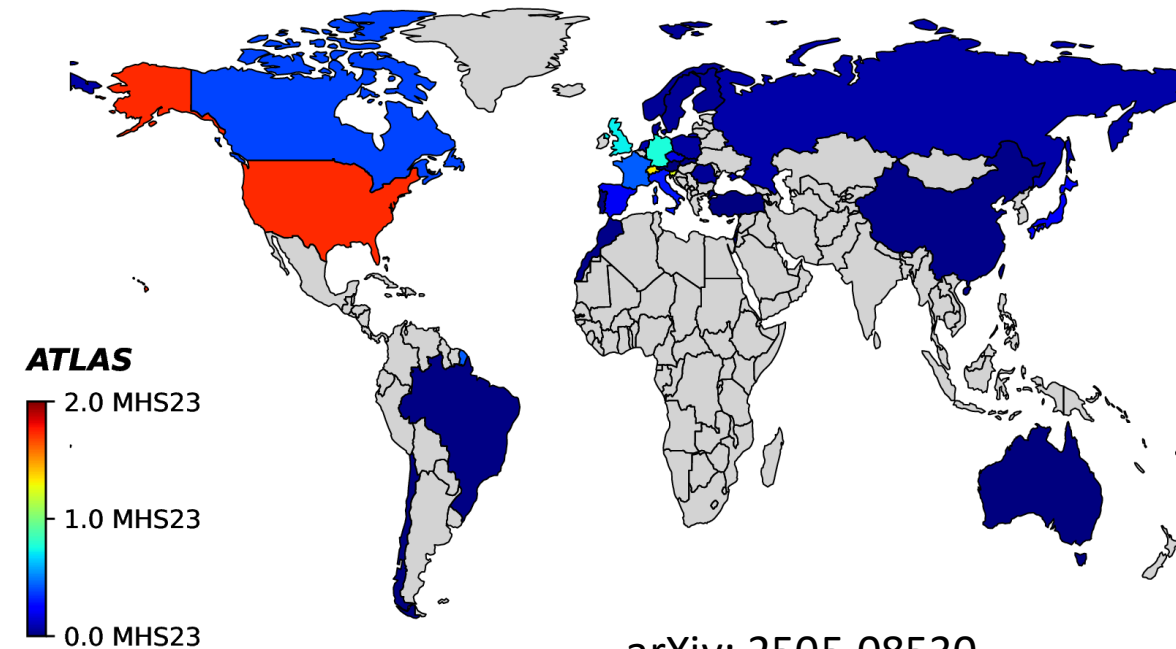
Tier2 (for WLCG)

- Worker nodes (ARC-CE/HTCondor): ~11k cores
 - ~4% of total ATLAS resources
- Storage (dCache): ~13 PB
 - ~3% of total ATLAS resources

Tier3 (for ATLAS-Japan)

- Interactive nodes: ~ 200 cores
- Worker nodes (HTCondor): ~ 1.7k cores
- Storage (GPFS): 3 PB
- GPU resources: 2 GPU servers with 10 GPUs

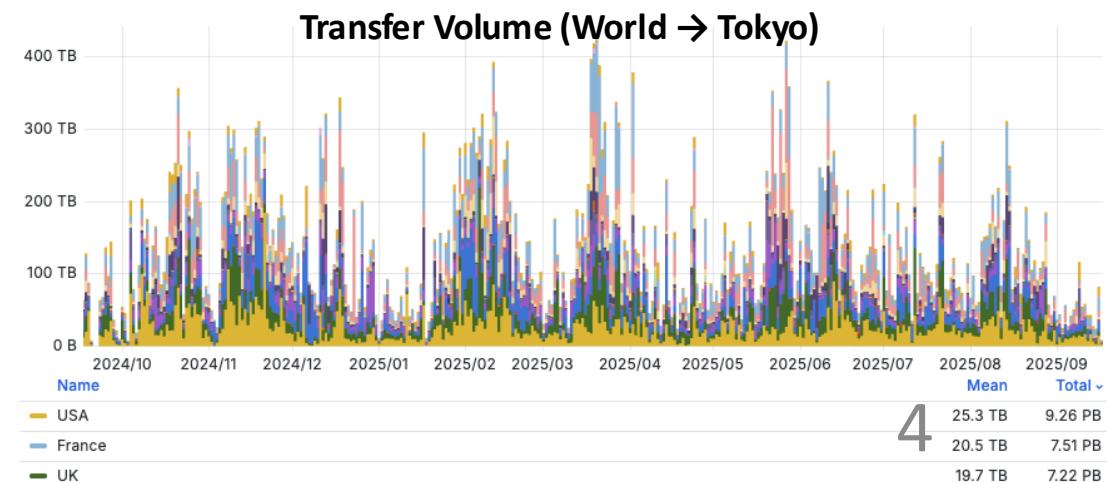
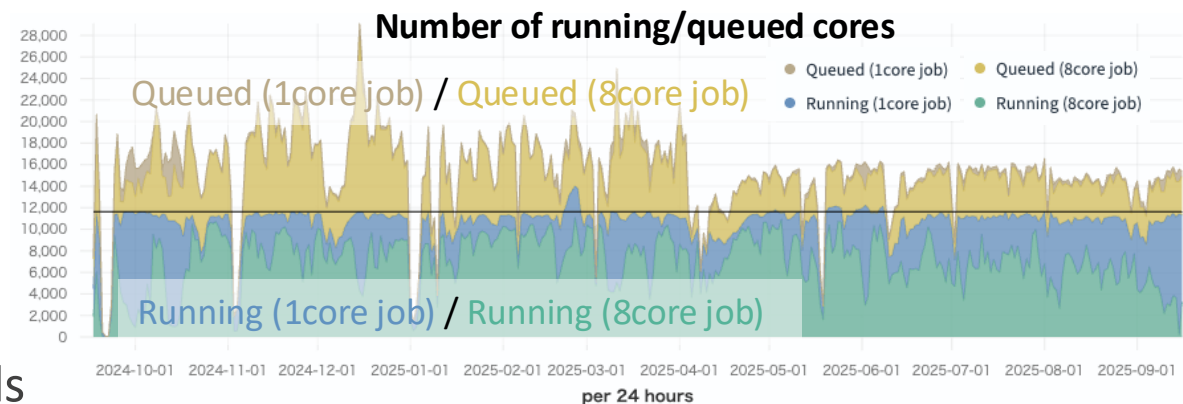
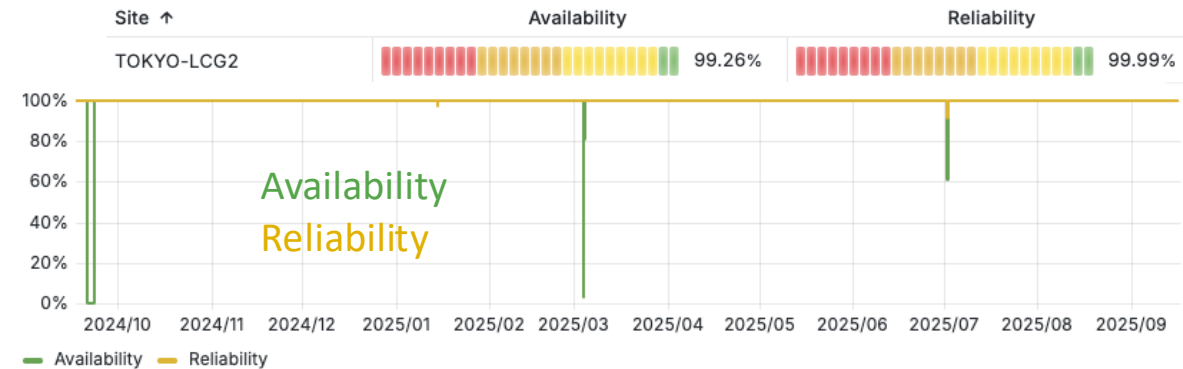
The worldwide distribution of ATLAS computing on average in 2023-2024



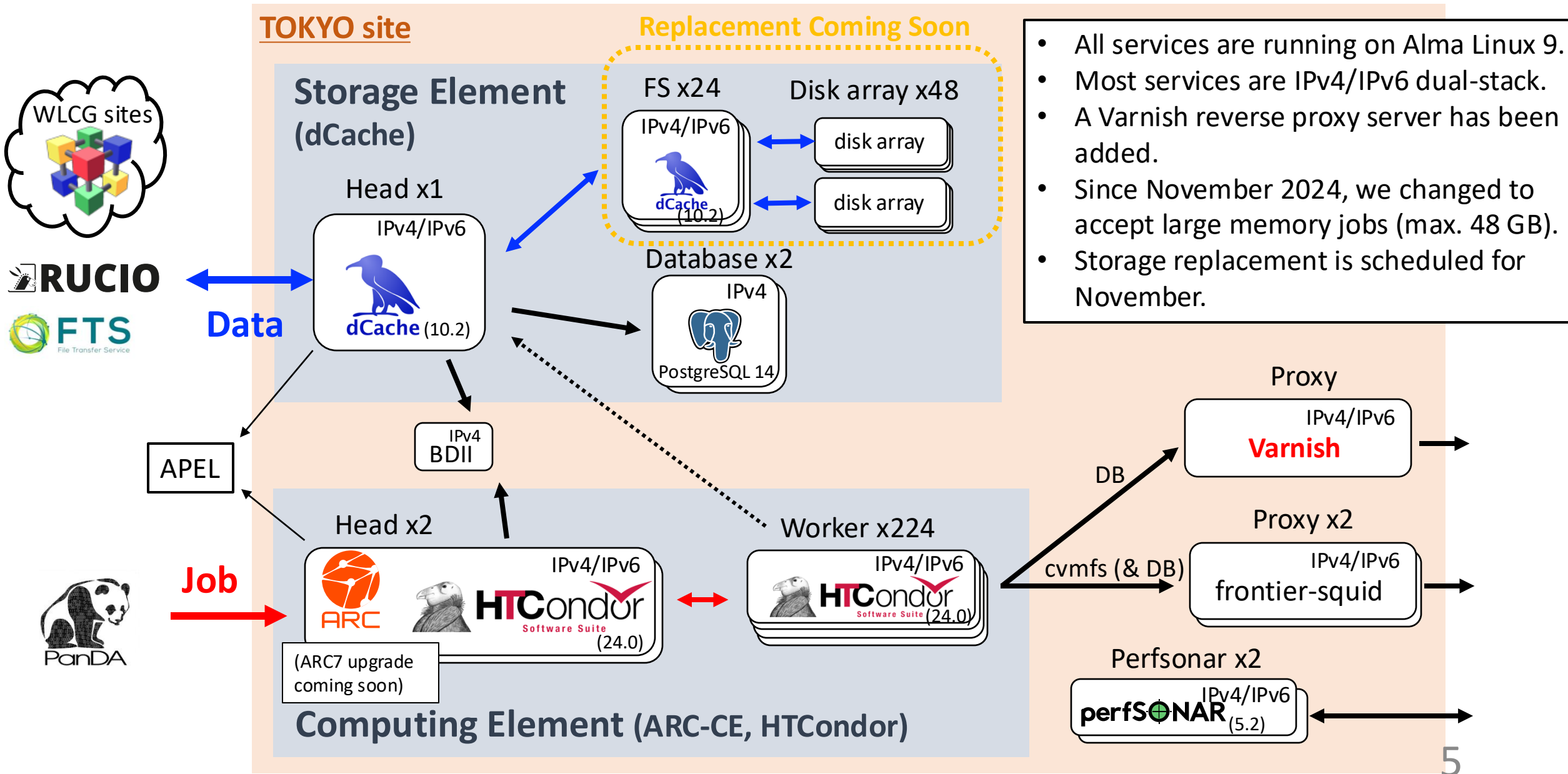
arXiv: 2505.08530

TOKYO Tier2: Status and Performance Summary

- High Availability and Reliability
 - 99.26% Availability / 99.99% Reliability
 - Typically 2-3 scheduled downtimes per year.
 - The current system (deployed in 2022) continues to run stably.
- Key metrics (Last 12 Months)
 - Completed jobs: 10.3 Million
 - Walltime (Successful Jobs): 6.36 Trillion HS23 seconds
 - Data transfer: 50 PB incoming / 35 PB for outgoing
 - Storage Capacity: A total of 11 PB for ATLAS Pledge, and up to 2 PB for national users (local group disk)



Grid Services: System Overview



- All services are running on Alma Linux 9.
- Most services are IPv4/IPv6 dual-stack.
- A Varnish reverse proxy server has been added.
- Since November 2024, we changed to accept large memory jobs (max. 48 GB).
- Storage replacement is scheduled for November.

Network Connectivity

Tokyo Tier2 Regional Center (RC) ↔ SINET6

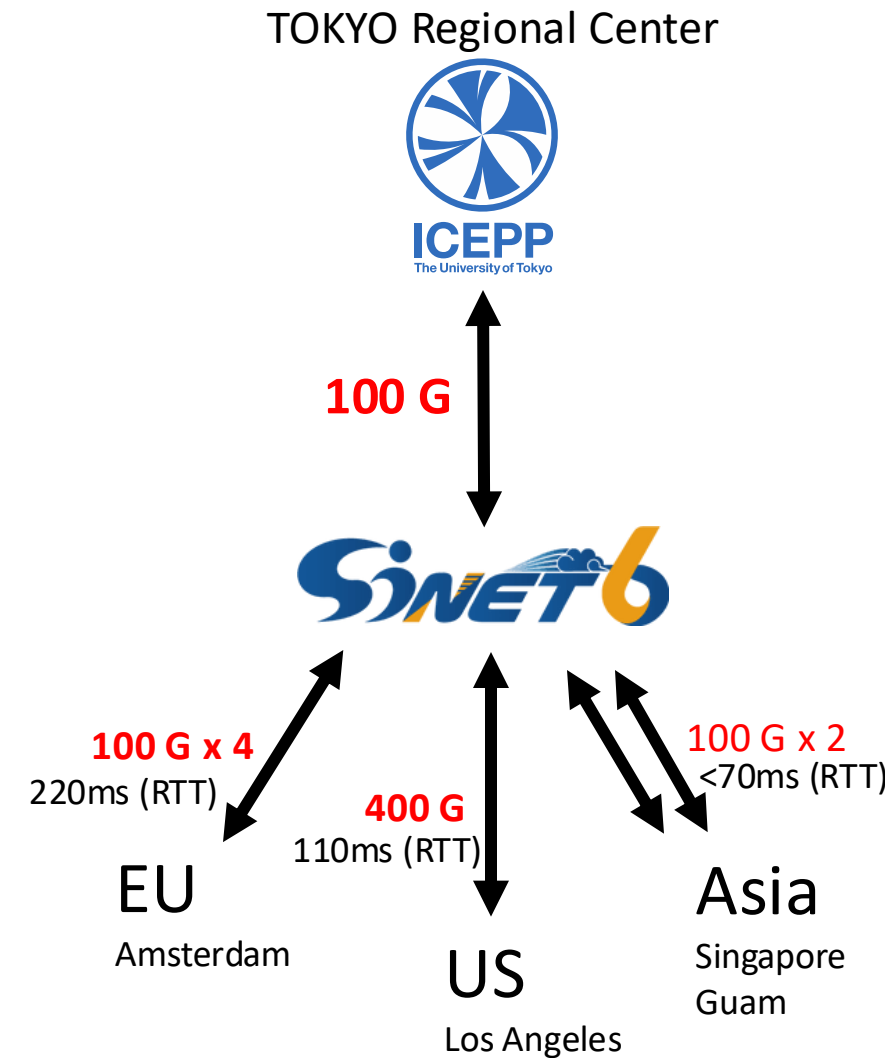
- Tokyo regional center is connected to SINET6.
- Bandwidth is 100 Gbps (since January 2024).

SINET international connections

- Connected to major global hubs via multiple 100+ Gbps
 - Amsterdam, Los Angeles, Singapore, Guam

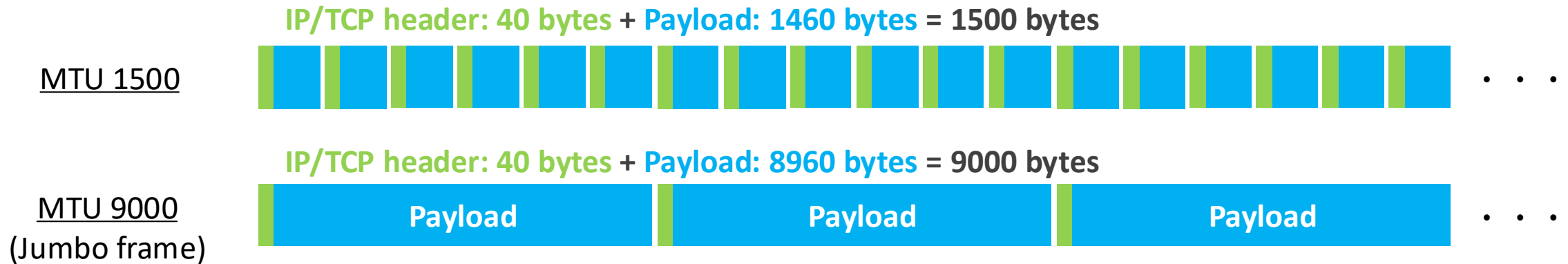
Record

- Data transfer volume:
 - 50 PB (inbound) + 32 PB (outbound) per year → **~220 TB / day**
- Dominant transfer region is Europe, followed by North America.



Network R&D: Jumbo Frame

- Jumbo frame: Frame with MTU > 1500 (typically 9000 bytes)

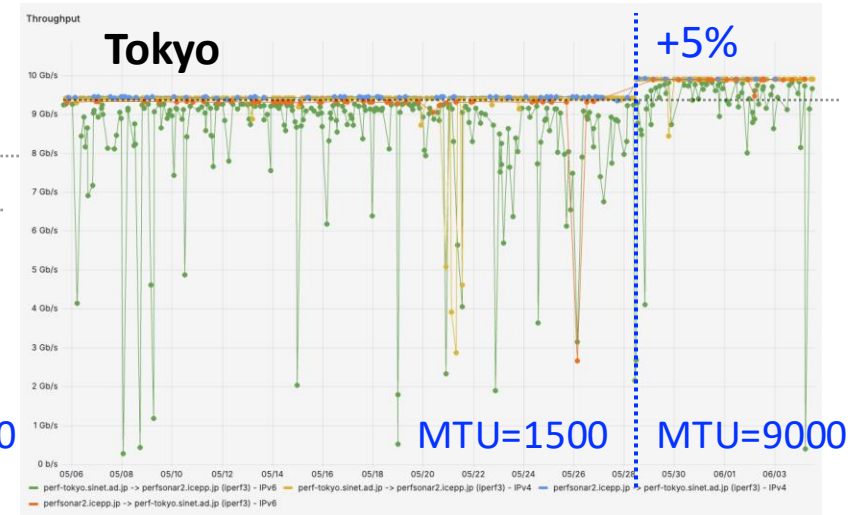
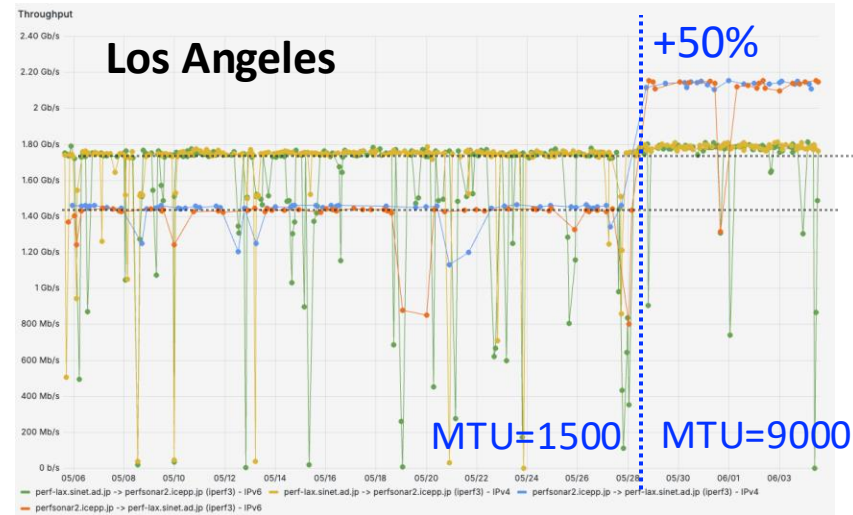
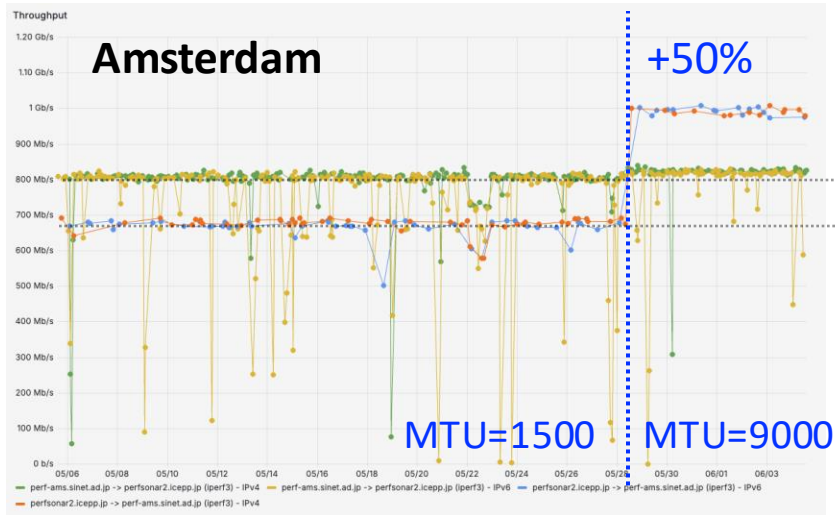


- Pros: Reduces CPU load.
 - More effective for long-distance transfers. Most Tokyo site traffic has >200ms RTT.
- Cons: Fails if any router on the path does not support Jumbo Frames
- Test ongoing using PerfSONAR

Network R&D: Jumbo Frame

TOKYO-LCG2 Perfsonar ↔ SINET Perfsonar

Tokyo → Others (IPv4) Others → Tokyo (IPv4)
Tokyo → Others (IPv6) Others → Tokyo (IPv6)



After changes MTU to 9000

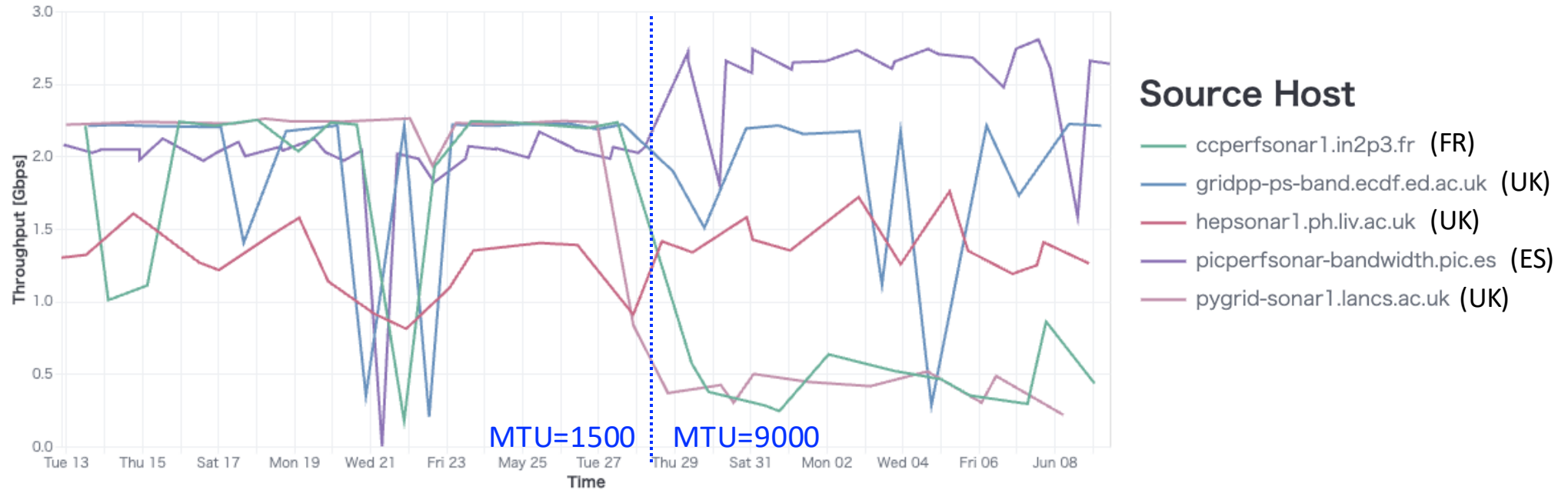
- Outbound(● ●): Significant improvement, especially to distant sites
- Inbound(● ●): little improvement → Under investigation
 - Likely due to TCP window auto-tuning issues

Dedicated test with Amsterdam PerfSONAR

- Fixing window size improved throughput by ~7-9% for both in/out-bound.

Network R&D: Jumbo Frame

TOKYO-LCG2 perfSONAR ↔ ATLAS site perfSONAR



- Improved at some sites, degraded at some sites.
- Needs further investigation and tuning.
- Plan to apply this to File servers after the investigation

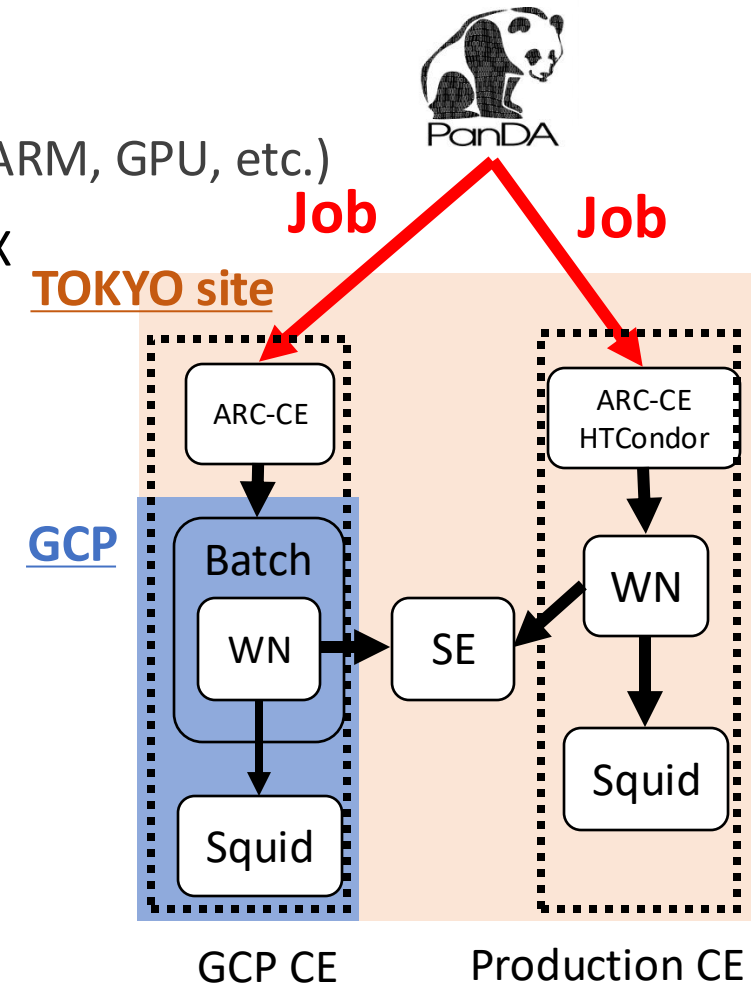
Usage of Cloud Resources as Grid Worker Nodes

Cloud resources

- Utilization of cloud resources as well as on-premise resources
- Ongoing study to integrate external resources as Grid resources
 - For on-demand use and temporary use of special hardware (high mem, ARM, GPU, etc.)
- Two cloud resources are tested: Google Cloud Platform (GCP) and MDX

GCP

- Implemented using ARC-CE + GCP Batch system
 - ARC-CE accepts jobs and submits to the GCP batch.
 - GCP batch manages the queue and WN assignments.
 - Site admins don't need to manage (spot-)WN instances directly.
- Record:
 - 4.4K completed jobs; 13.1 CPU years (success), 0.82 CPU years (failed)

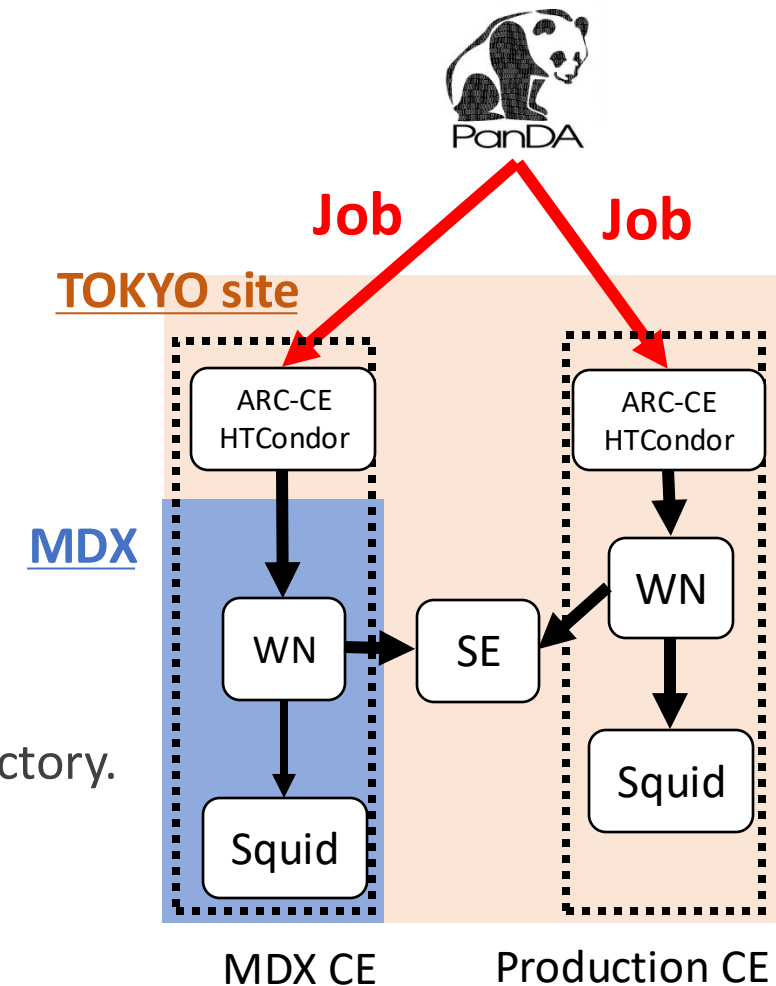


Usage of Cloud Resources as Grid Worker Nodes



Academic cloud platform for supporting data science and cross-disciplinary research collaborations

- Pros & Cons
 - Pricing is much lower than commercial cloud. And no network charge.
 - Connected to SINET. Transfers with TOKYO site are very fast.
 - Minimal functionality compared to commercial cloud. But it might be enough for our use-case.
- Implemented ARC-CE + HTCondor
 - CE (ARC-CE + HTCondor CM/AP) hosted in on-premise resources
 - WN and squid deployed on MDX
 - Local SSD resources are limited. Lustre volume is used for working directory.
- Record:
 - 30K completed jobs; 70 CPU years (success), 15 CPU years (failed)
- Considering future use as SE as well.



Next Hardware Procurement

- The Storage system will be replaced in November.
 - Overall Architecture will remain unchanged. (server count, disk connection, etc.)
 - Volume: Increasing from 22.2 PB to 37.3 PB
 - The HDD volume increase (from 14 TB to 22 TB per drive) is the main contributor to this capacity growth.
 - Data migration (13PB) is planned to be performed online utilizing dCache functionality.
- Delivery is currently underway

File server and Disk array

File server (Dell R660xs)

Disk array (Infortrend ESDS 3024)



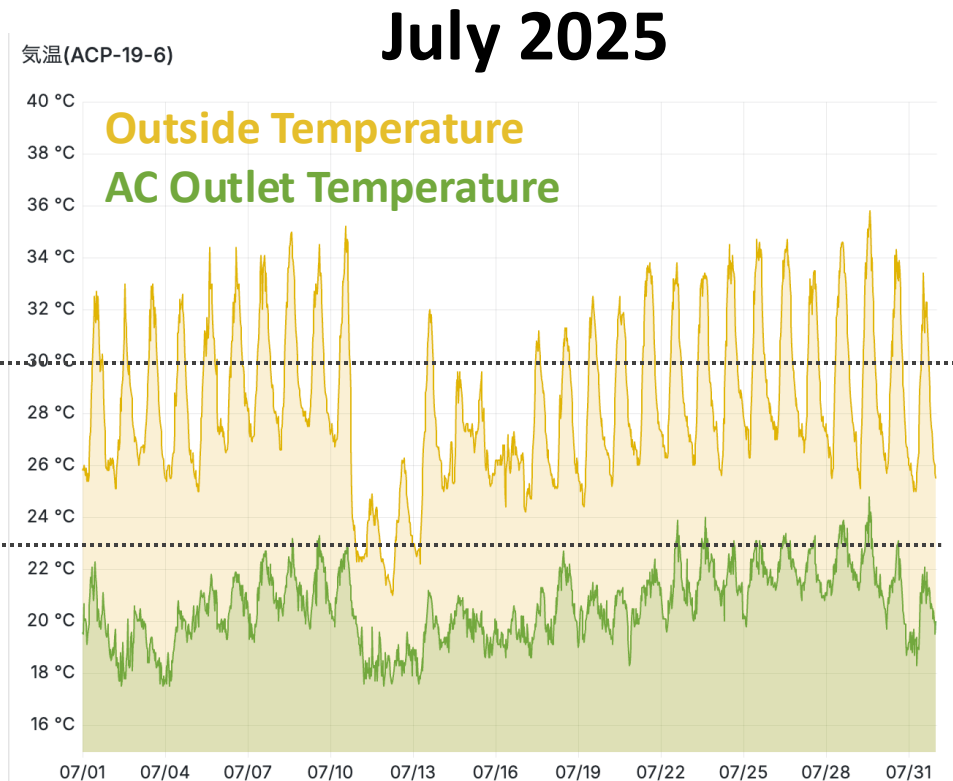
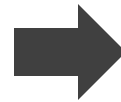
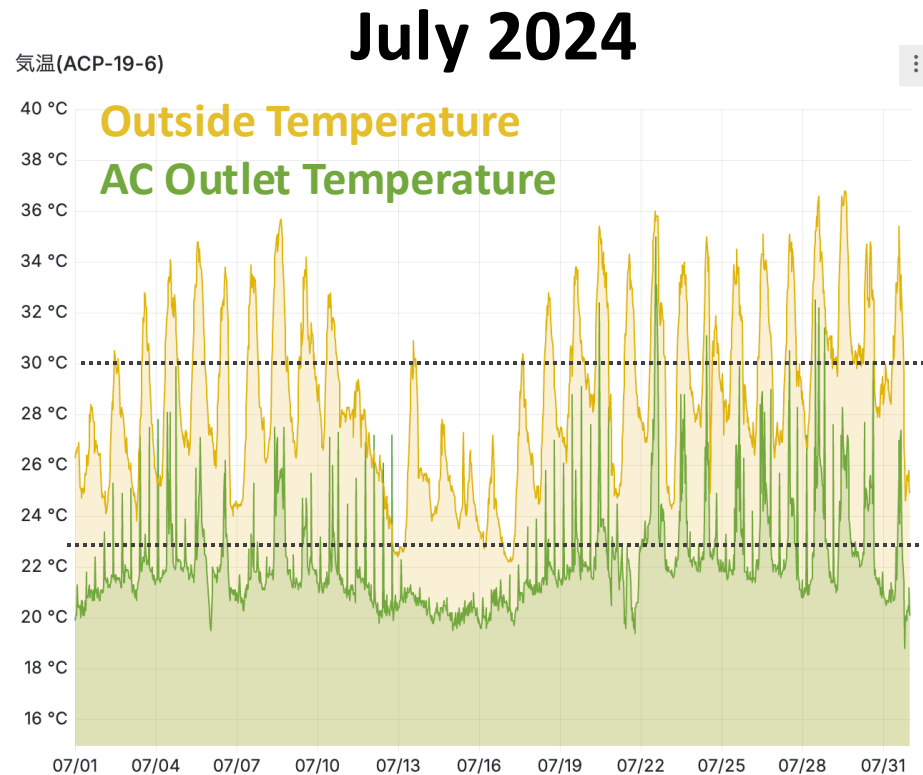
Server rack for storage



- CPUs will remain unchanged; however, more may be purchased depending on the budget.¹²

Server Room Cooling

- The temperature in Tokyo is getting higher, and our air conditioners are approaching their performance/lifespan limit. We are gradually phasing in replacements.
 - The air conditioner in the most severe location was replaced in March, which resulted in a significant improvement in cooling capacity.



Summary and plan

- ICEPP Regional Analysis Center Operation
 - Contributes to ~4% CPU and ~3% Disk of ATLAS Grid sites
 - All Tier2 services are smoothly operated.
 - Varnish (reverse proxy)
 - Large memory jobs
- R&D
 - Jumbo frame
 - Cloud resources usage: Academic cloud is promising.
- Next Procurement:
 - Storage is scheduled to be upgraded in Nov this year (22 PB → 37 PB).
 - Air conditioner replacement is gradually on-going.

Backup

The 6th system vs the 7th system

		Total	For Tier2
CPU	6 th system	304 nodes, 15808 cores (26 cores / CPU) Intel Xeon Gold 5320 2.2 GHz (Icelake) 337 kHS06 1.92 TB SSD / node	224 nodes, 11648 cores 21.34 HS06 / core 2.5 GB RAM / core
	7 th system		
Disk storage	6 th system	72 disk arrays, RAID6 22,176 TB (14 TB / HDD)	48 disk arrays, RAID6 14,784 TB (14 TB / HDD)
	7 th system	74 disk arrays, RAID6 35,816 TB (22 TB / HDD)	56 disk arrays, RAID6 27,104 TB (22 TB / HDD)

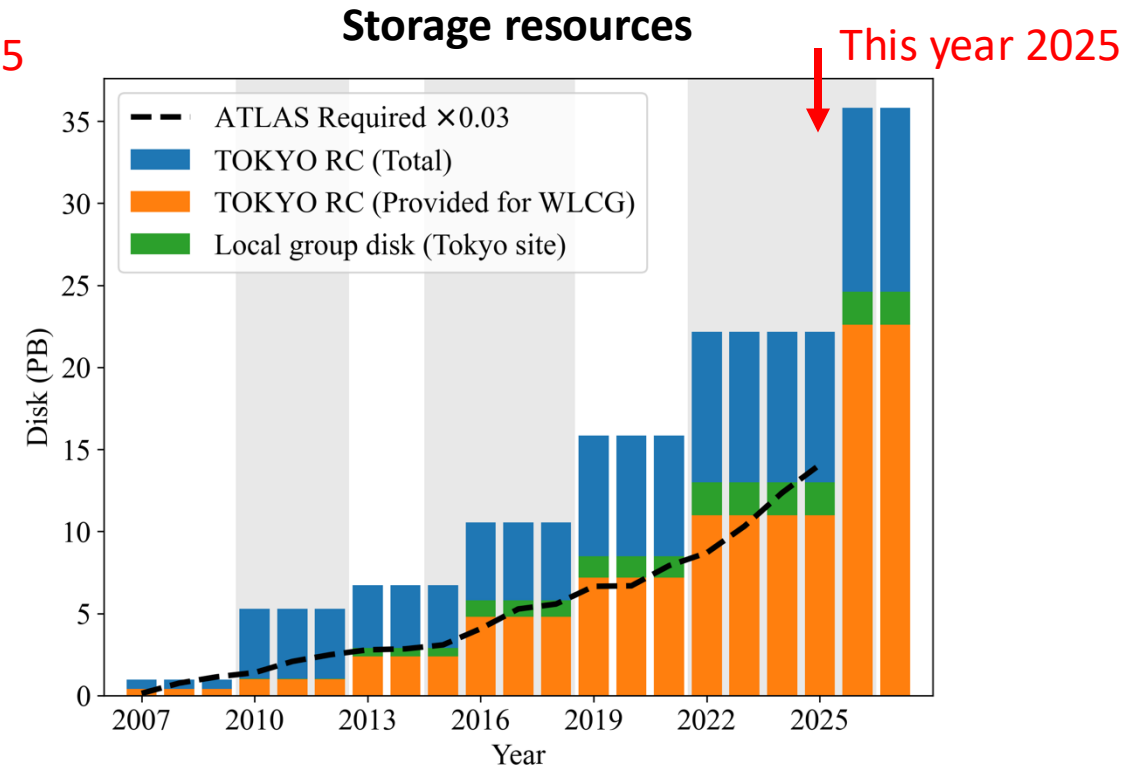
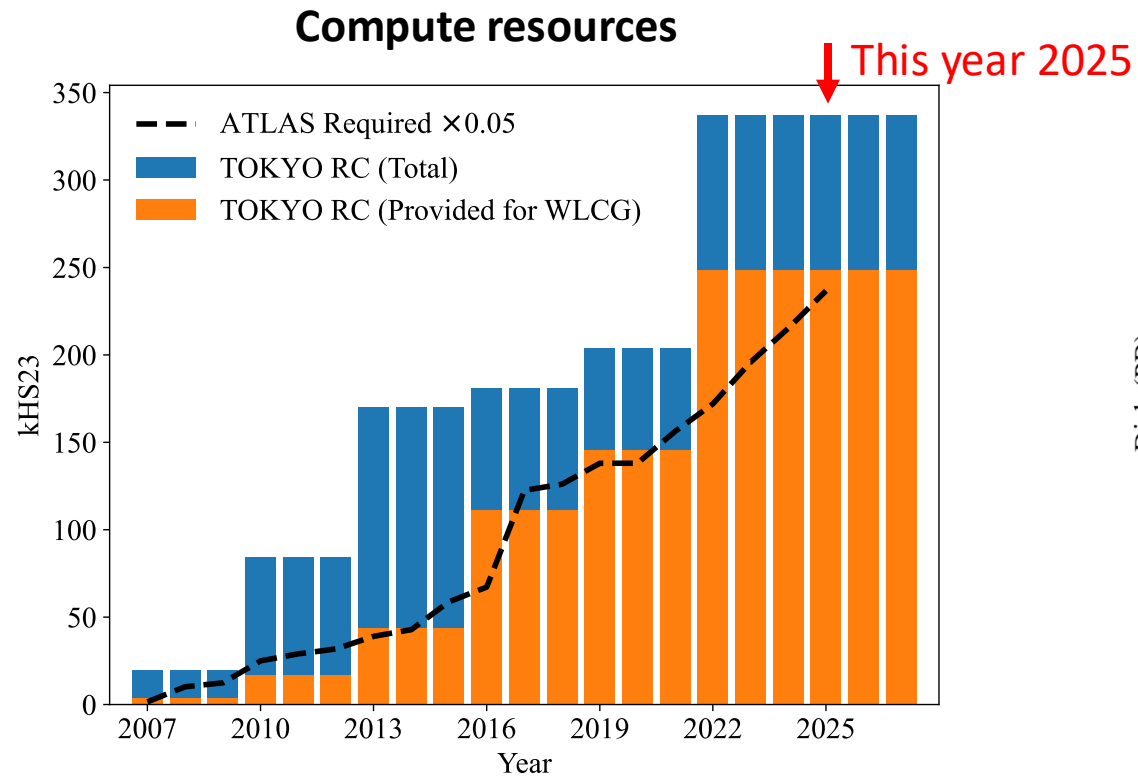
Perfsonar4 ↔ SINET Amsterdam (Throughput results)

Throughput: iperf3, 0-10s

Window size	parallel	Receiver Throughput (Mbps) (Amsterdam→Tokyo)		Gain (%)	Receiver Throughput (Mbps) (Tokyo→Amsterdam)		Gain (%)
		MTU=1500	MTU=8986		MTU=1500	MTU=8986	
auto	1	684	714	4.5	553	828	49.6
32 MB		467	498	6.6	463	494	6.7
64 MB		892	963	8.0	884	953	7.8
128 MB		1730	1850	6.9	1730	1850	6.9
256 MB		3340	3600	7.8	3350	3600	7.5
auto	4	2600	2830	8.8	2170	3240	49.3
64 MB		3470	3780	8.9	3540	3800	7.3
256 MB		5200	5240	0.8	6820	7780	14.1
auto	8	5320	5550	4.3	4310	6370	47.8
64 MB		5430	6050	11.4	6880	7580	10.2
256 MB		4590	6060	32.0	7030	7980	13.5

An overall improvement of approximately 7-9%, with up to 50% improvement in some configurations.

Next hardware procurement



- The Tokyo site's hardware has been upgraded every 3 years so far.
- Due to a challenging procurement schedule, we can no longer follow this policy.
- Only storage will be replaced in November this year.
- Unchanged CPUs; possibly buy more depending on budget.

Network Update: Tokyo-Amsterdam Line Rerouting Impact

Before Mar 2024



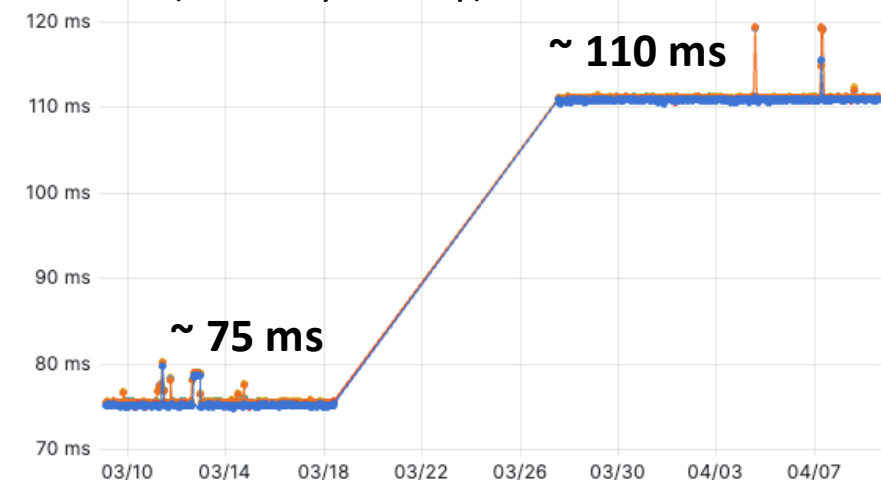
After Apr 2024



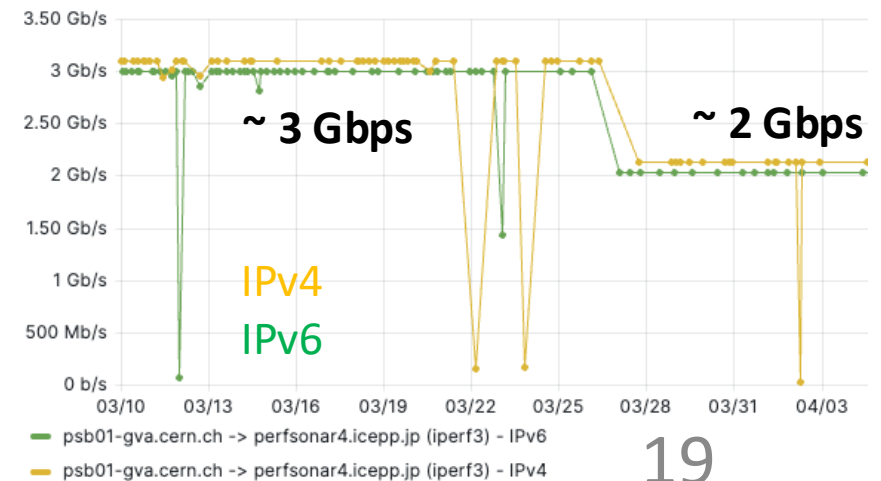
from [SINET6 webpage](#)

- Geographical cable routing has been changed
- **Bandwidth improved: 100G to 100G x 4**
- **Latency increased: 150ms to 220ms (RTT)**
- No major production impact with this change.

perfSONAR latency (Tokyo → Amsterdam)
(One-way latency)

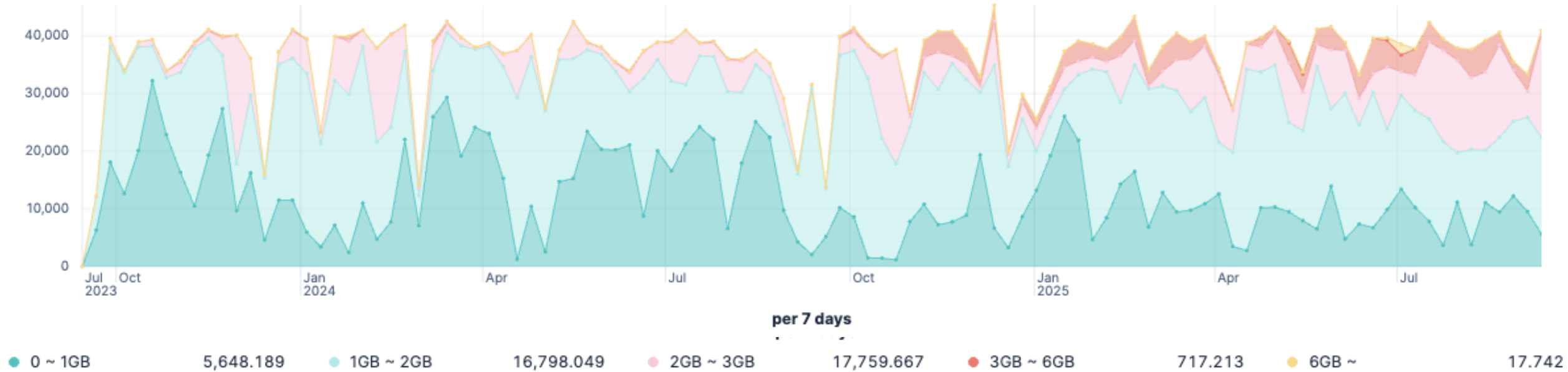


perfSONAR throughput (CERN → Tokyo)
(Single-stream throughput)



Flexible slots with memory

kHS23 hours



- Jobs that require more than 3 GB are coming since Nov 2024.