Cybermedia Center

Osaka University, Japan

SC17 BOOTH 1943

About Us：Cybermedia Center, Osaka University

As a resource provider of knowledge and technology derived from advanced researches conducted in Osaka University, the Cybermedia Center (CMC) offers support in the areas of large-scale computation, information communication, multimedia content and education. The center also works closely with educational and research organizations within Osaka University, as well as with industries and institutes outside the University. By sharing its resources and encouraging local communities to use its facilities for public lectures and other events, CMC has helped to create a more internationally-oriented IT society for the region.

Location Map

Location

University-Wide Services

Large-Scale Computer System, we provide a high-performance computing environment, consisting of the NEC SX-ACE supercomputer and PC clusters, to both the academic and industrial communities. Part of the overall computer system is provided, as a computational resource, to the national High-Performance Computing Infrastructure (HPCI).

Information Media Education Multimedia Language Education, we have implemented a consistent curriculum, from the basics of computer utilization to advanced subject matter, while the Computer Assisted Language Learning System supports foreign language learning and cross-cultural understanding in accordance with each individual's language-proficiency level.

Cybermedia Commons is an active learning space for students, exploiting a wide variety of the Cybermedia Center's information technology, to support student's active learning and research activities.

Cybermedia Commons

Digital Library provides academic information databases and remote access to electronic journals. It is equipped with multimedia terminals and public network jacks with an authentication system.

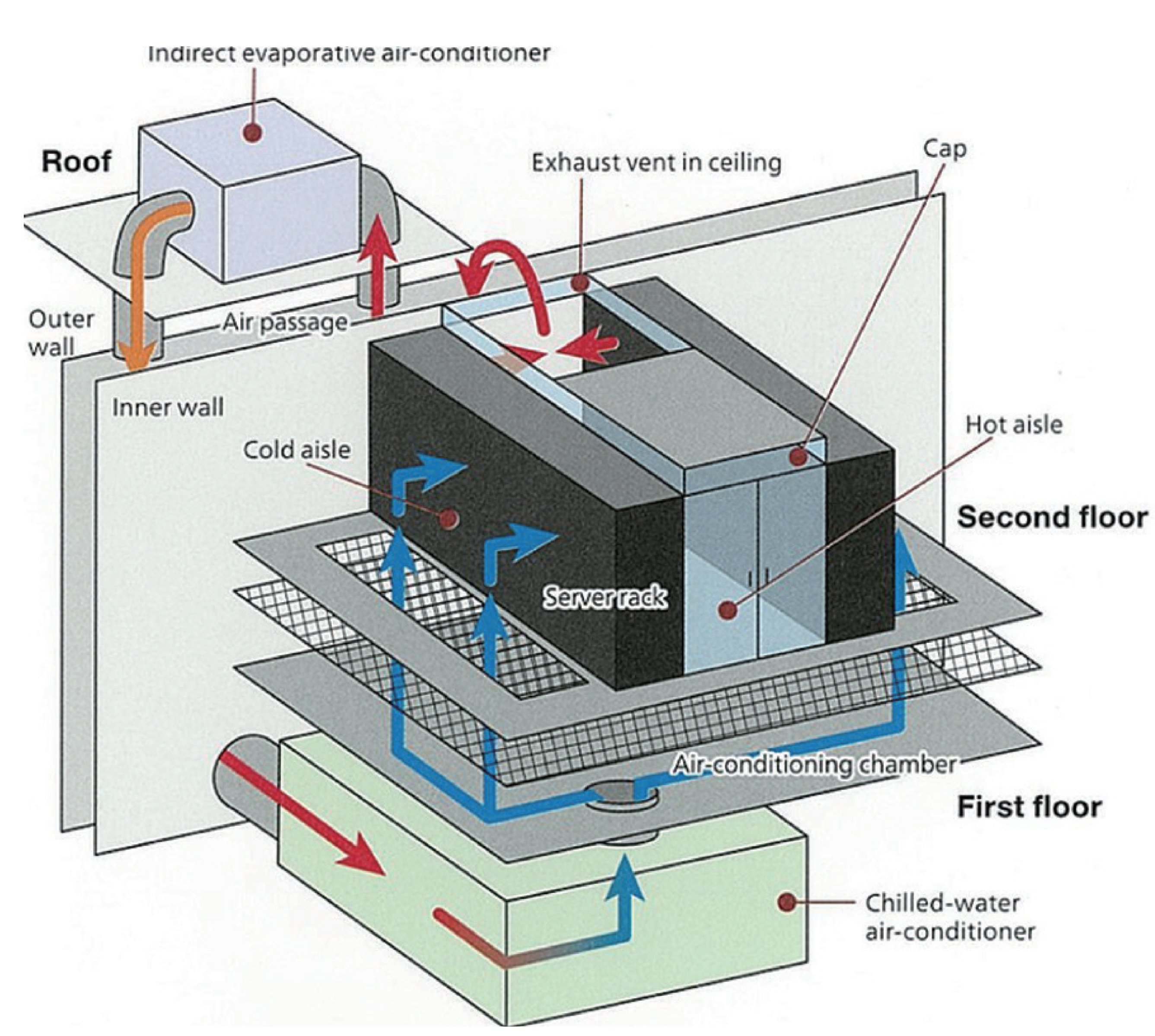
Repair and Maintenance of the Information Network, a high-speed, stable and reliable campus-wide network environment, as well as wireless access networks, as information infrastructure for supporting the educational, research, and social contribution activities of Osaka University.

Visualization Services, we maintain two types of high-resolution stereo visualization systems, as primary visualization facilities. The systems can be used for scientific visualization, information visualization, visual analytics, and other research activities.

Visualization Services

Academic Cloud improves the integration of computing resources scattered across the university. The objectives of the system are to optimize administrative operations, enhance security, and reduce costs.

IT Core Annex is a two-story steel-frame data center housing large-scale computers. The perimeter wall is designed with gently curving surface and light-permeable metal panels, to harmonize with the surrounding environment.



Air-conditioning mechanism in IT Core Annex

Contact : system@cmc.osaka-u.ac.jp　Visit us at www.cmc.osaka-u.ac.jp

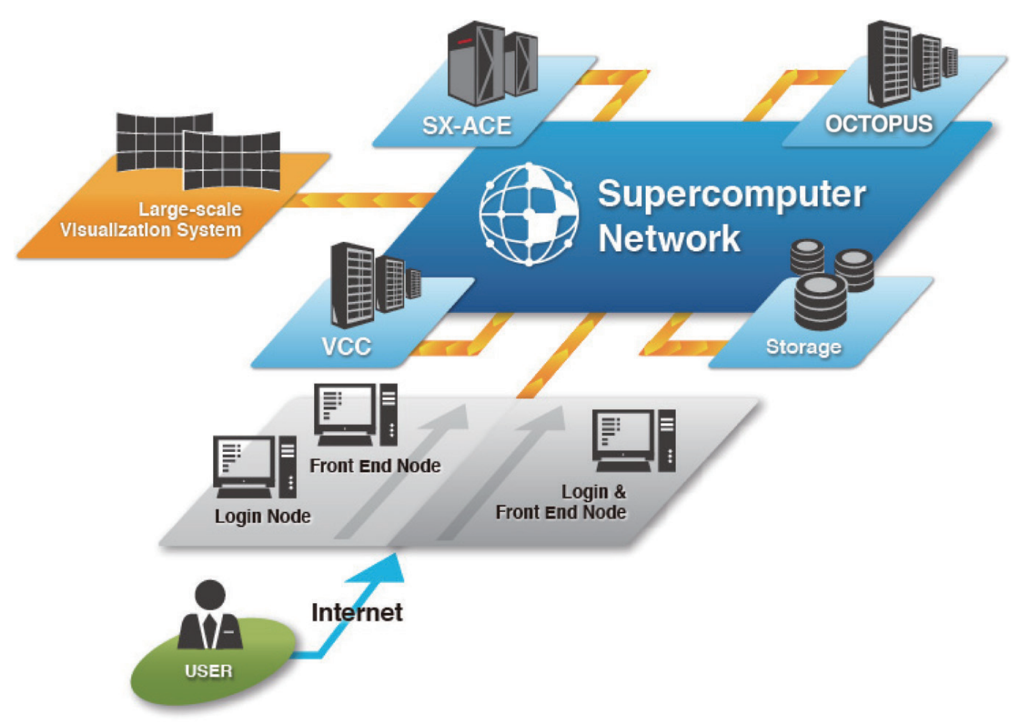
Cybermedia Center

Cybermedia Center
Osaka University

Osaka University, Japan

SC17 BOOTH 1943

Large-scale Computing and Visualization Systems at the Cybermedia Center



Overview of high-performance computing environment at the CMC

Large-scale computing systems (SX-ACE, VCC, and OCTOPUS), and large-scale visualization systems are deployed on CMC-Supercomputer network, a.k.a CMC-SCinet, a low-latency and wide-bandwidth network. This architectural design allows users to access to large-scale storage systems, perform large-scale high-performance computation and analysis on our large-scale computing systems, and then visualize its computation and analysis results without losing any important information on our large-scale visualization systems.

Large-scale Computing System

The large-scale computing systems at the CMC are classified into (1) Vector-typed Supercomputer and (2) Scalar-typed Supercomputer.

SX-ACE

Type: Vector
OS: Super UX
# of nodes: 1536
# of cores: 6144
Peak performance: 423 TFlops

SX-ACE is a "clusterized" vector-typed supercomputer, composed of 3 cluster, each of which is composed of 512 nodes. Each node equips 4-core multi-core CPU and a 64 GB main memory. These 512 nodes are interconnected on a dedicated and specialized network switch, called IXS (Internode Crossbar Switch) and form a cluster. Note that IXS interconnects 512 nodes with a single lane of 2-layer fat-tree structure and as a result exhibits 4 GB/s for each direction of input and output between nodes.

Library

MathKeisan(BLAS, LAPACK, etc)
ASL, ASLSTAT, ASLQUAD
MPI/SX
HPF/SX
XMP

OCTOPUS

Type: Scalar
OS: Linux
# of nodes: 319
Peak performance: 1.463 Pflops
Interconnect: InfiniBand EDR

OCTOPUS means Osaka university Cybermedia cenTer Over-Petascale Universal Supercomputer. OCTOPUS is a new cluster system supposed to start its operation in December 2017. this system is composed of different types of 4 cluster, General purpose CPU nodes, Xeon Phi nodes, GPU nodes and Large-scale shared-memory nodes, total 319 nodes. These nodes and large-scale storage "EXAScaler" are interconnected on InfiniBand EDR and form a cluster.

Library

Intel MKL(BLAS, LAPACK, etc)	IntelMPI, OpenMPI, MVAPICH2
ASL, ASLSTAT, ASLQUAD	XMP

General purpose CPU node × 236

CPU: Intel Xeon Gold 6126 × 2 (2.6 GHz, 12 cores)
Memory: 192 GB
Performance: 1.996 TFlops

Xeon Phi node × 44

CPU: Intel Xeon Phi 7210 (1.3 GHz, 64 cores)
Memory: 192 GB
Performance: 2.662 TFlops

GPU node × 37

CPU: Intel Xeon Gold 6126 × 2 (2.6 GHz, 12 cores)
Memory: 192 GB
Accelerator: NVIDIA Tesla P100×4
Performance: 23.196 TFlops

Large-scale shared-memory node × 2

CPU: Intel Xeon Platinum 8153 × 8 (2.0 GHz, 16 cores)
Memory: 6 TB
Performance: 8.192 TFlops

Sstorage

File system: DDN EXAScaler (Lustre)
capasity: 3.1PB

Application

AVS/Express	TensorFlow	OpenFOAM
NICE DCV	Torch	GAMESS
VisIt	Caffe	ABINIT-MP
Gaussian09	Theano	Octave
IDL	Chainer	Relion
LAMMPS	Quantum Espresso	GROMACS

VCC (PC Cluster for large-scale visualization)

Type: Scalar
OS: Linux
# of nodes: 69
Peak performance: 31.104 TFlops
Accelerator: NVIDIA Tesla K20×59

Library

Intel MKL(BLAS, LAPACK, etc)
Intel MPI, Open MPI

VCC is a cluster system composed of 69 nodes. These nodes are interconnected on InfiniBand FDR and form a cluster. Also, this system has introduced ExpEther, a system hardware virtualization technology. Each node can be connected with extension I/O nodes with which GPU resource, and SSD on 20Gbps ExpEther network. A major characteristic is that this cluster system is reconfigured based on user's usage and purpose by changing the combination of node and extension I/O node.

Large-scale Visualization System

The large-scale visualization systems at the CMC are set up on Campus and on CMC's Umekita Office. Large-scale and interactive visualization processing becomes possible through the dedicated use of PC cluster for large-scale visualization (VCC) on these systems. The visualization system in Campus is composed of 24 50-inch Full HD (1920x1080) stereo projection module (Barco OLS-521). Also, OptiTrackFlex13, a motion capturing system has been introduced in this visualization system. By making use of the software corresponding to the motion capturing system, interactive visualization leveraging Virtual Reality (VR) becomes possible.

Contact : system@cmc.osaka-u.ac.jp Visit us at www.hpc.cmc.osaka-u.ac.jp

Cybermedia Center

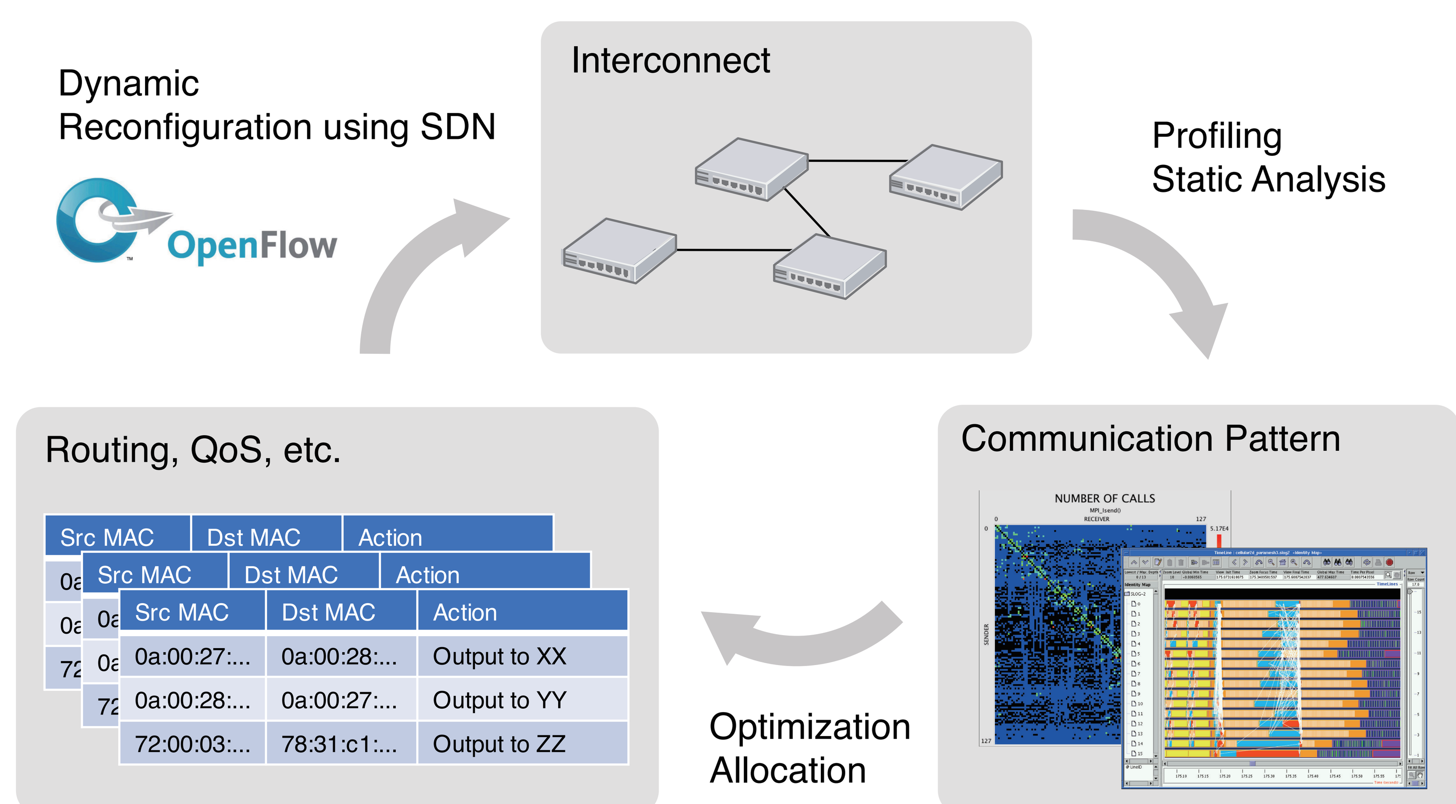
Osaka University, Japan

SC17 BOOTH 1943

Dynamically Optimized Interconnect Architecture Based on SDN

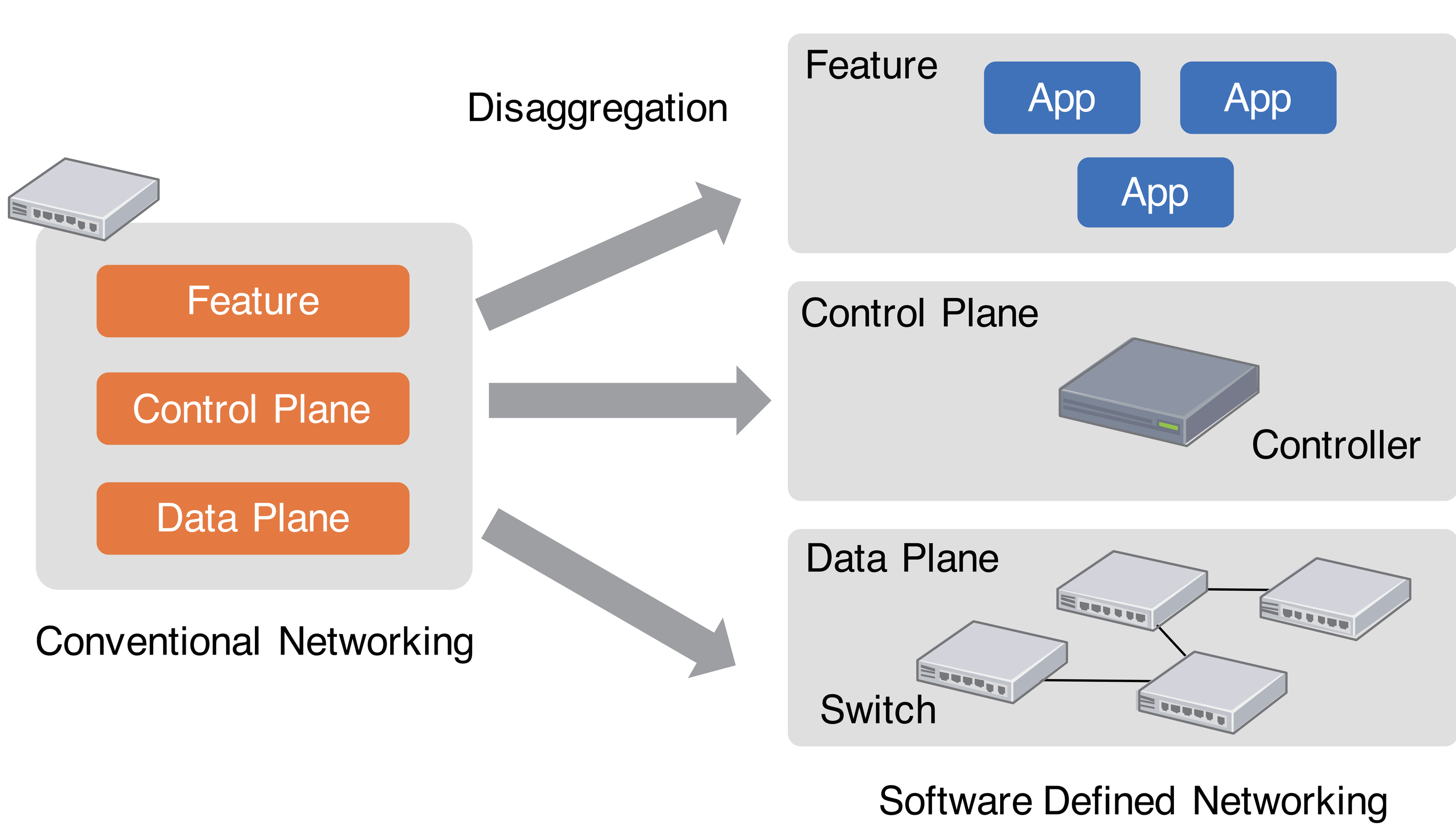
Fundamental Idea of SDN-enhanced MPI

Can we accelerate MPI communication and improve the utilization of interconnect by leveraging the network programmability of SDN?

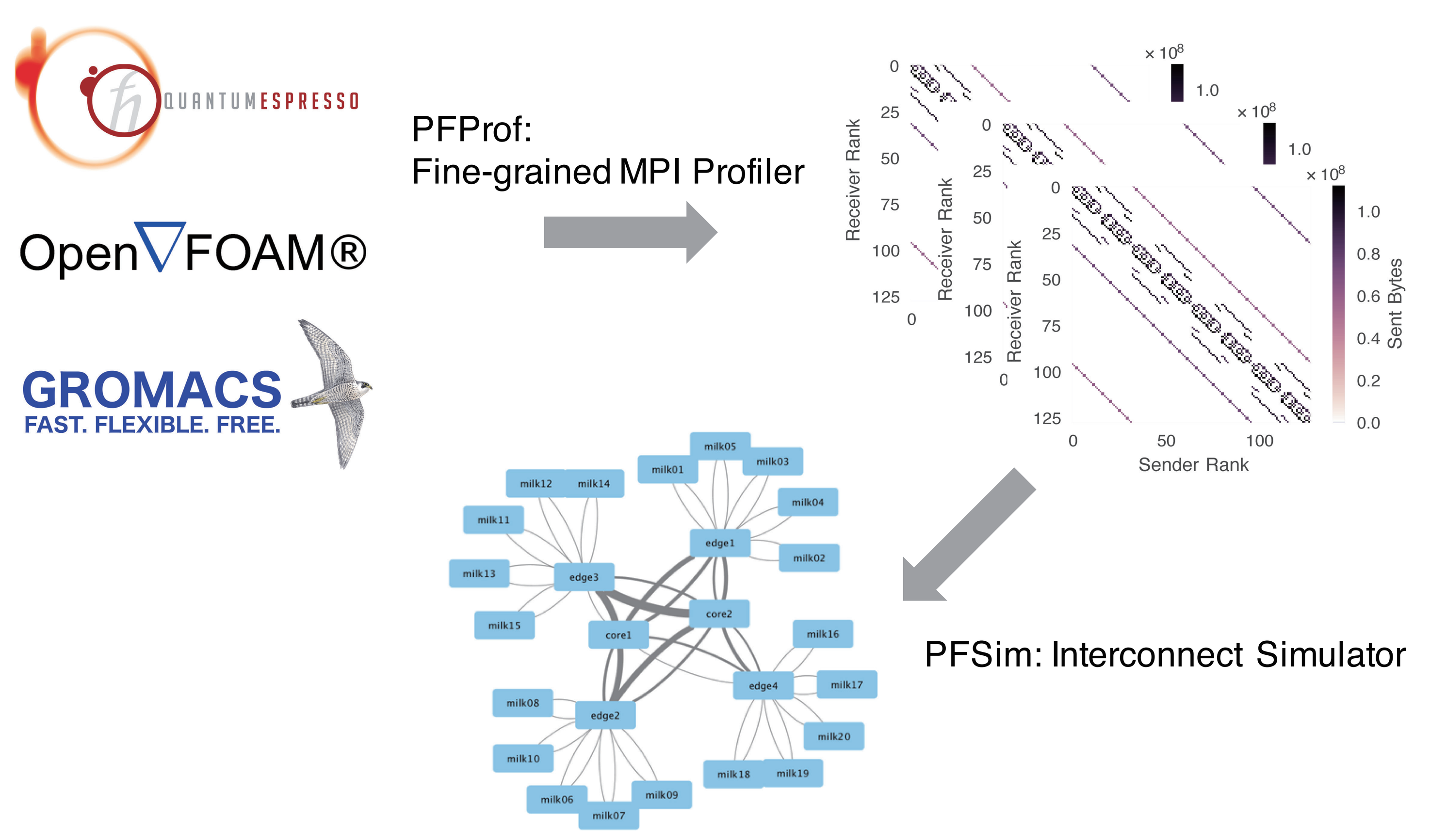


What is Software-Defined Networking (SDN)?

Software-Defined Networking (SDN) is a novel network architecture that decouples conventional networking function into a programmable control plane (responsible for deciding how to control the packets) and a data plane (responsible for the actual packet delivery).

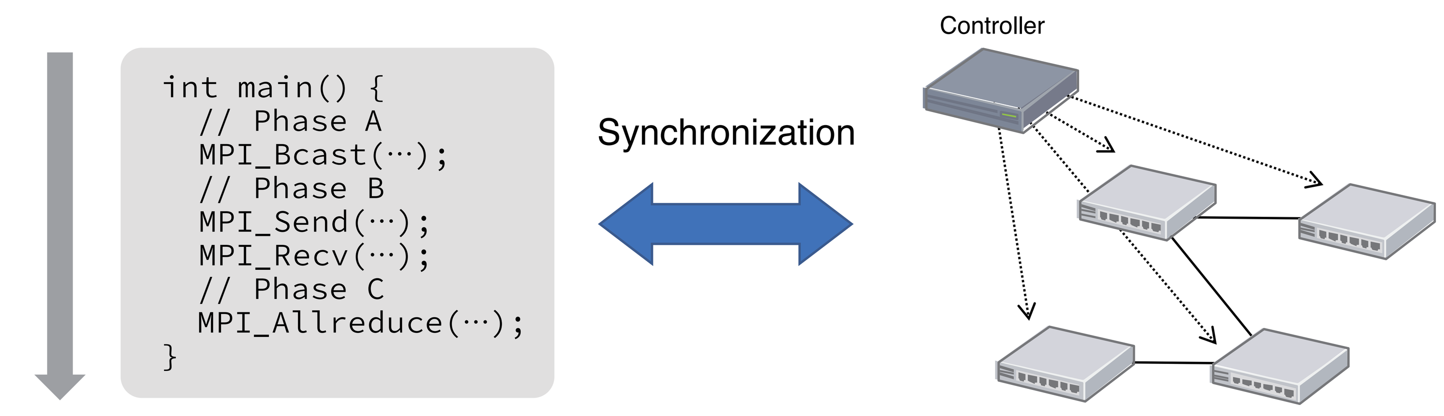


Toolset for Analyzing Application-aware Dynamic Interconnects

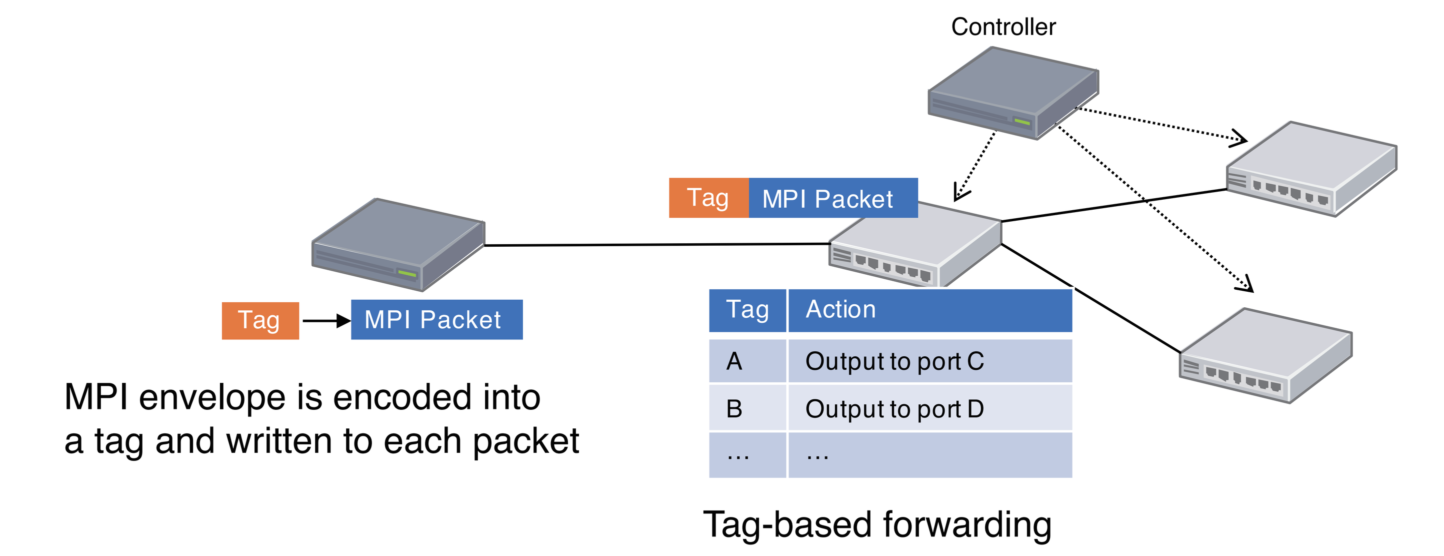


Coordination Mechanism of Computation and Communication

How do we reconfigure the interconnect in accordance with the execution of application?

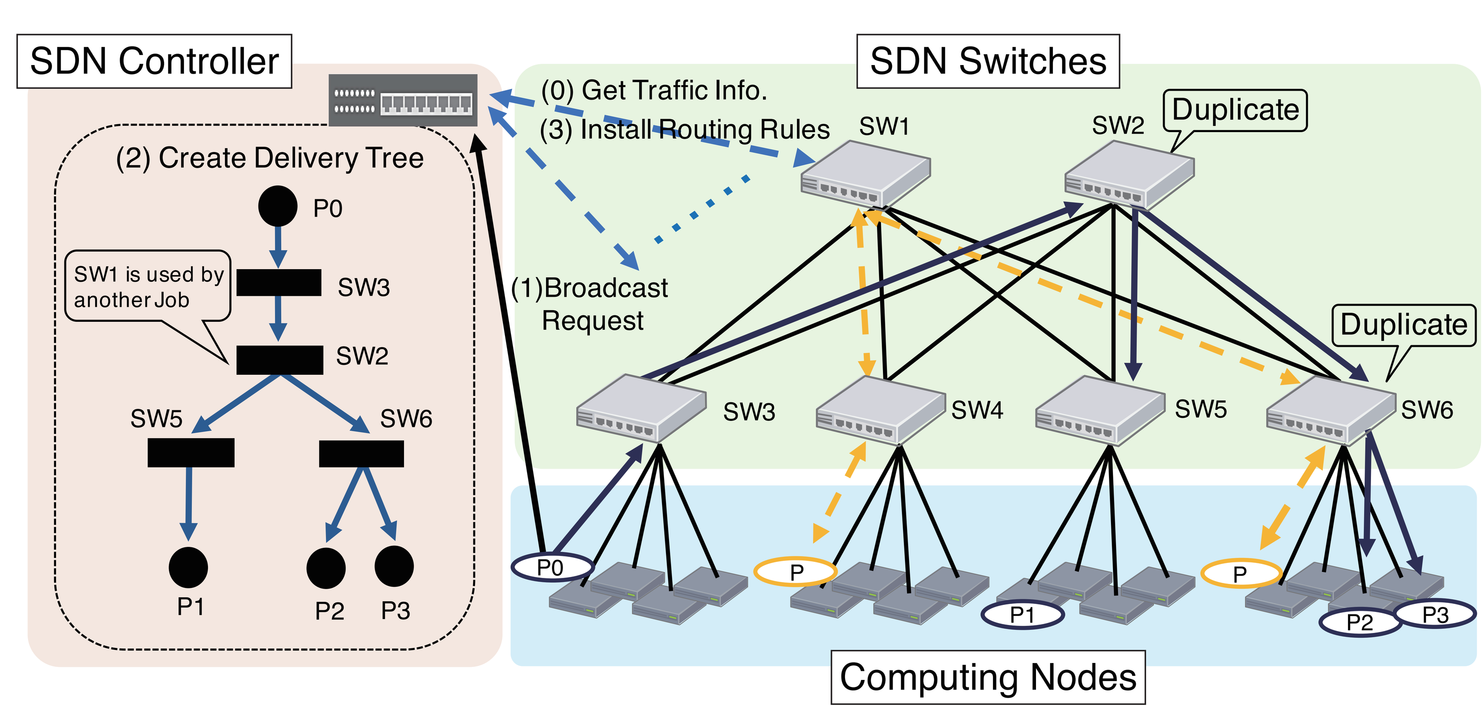


Encode the MPI envelope into a tag and embed into each packet in the kernel



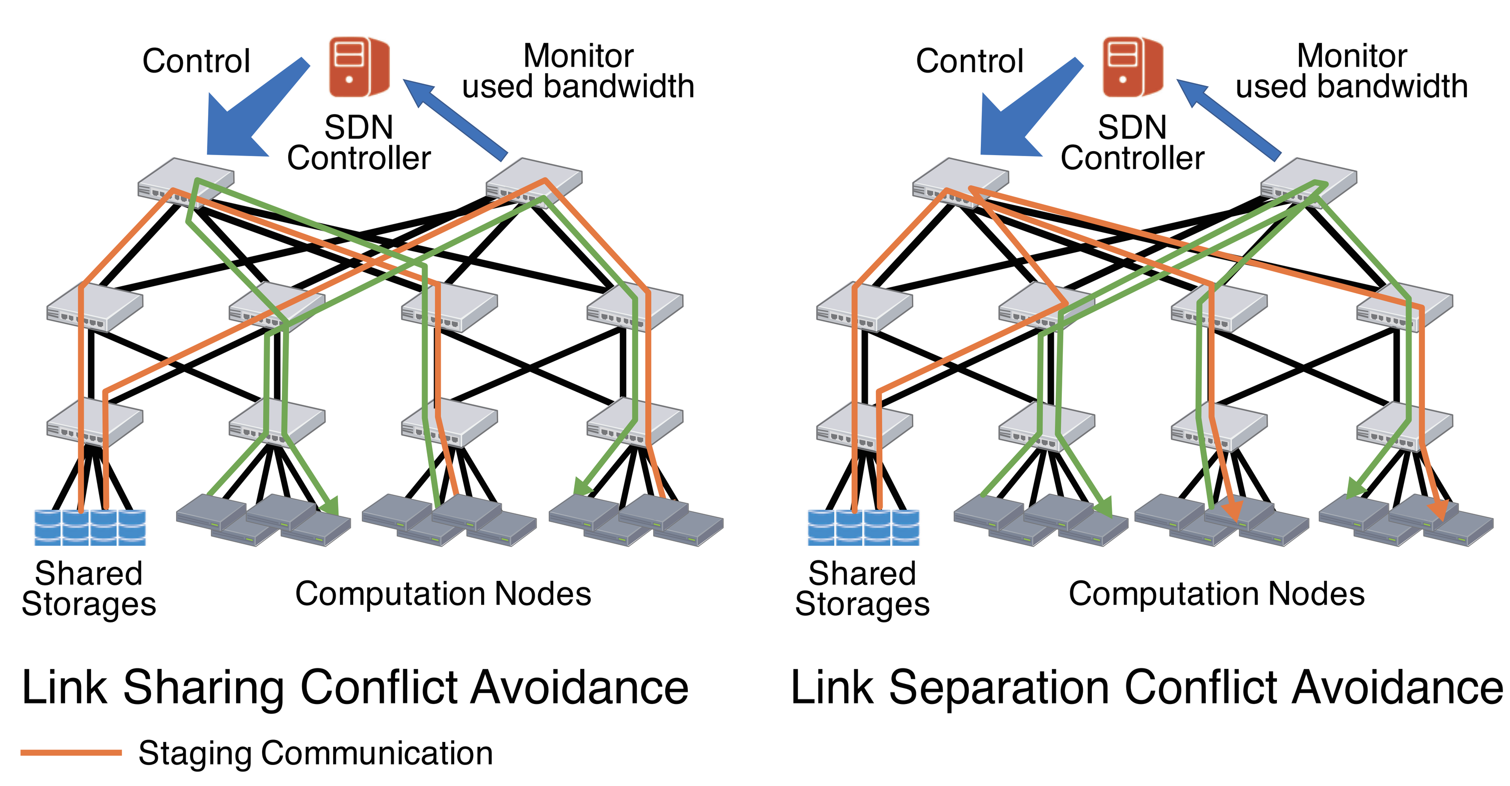
SDN-enhanced MPI Broadcast

We propose an accelerated implementation of MPI broadcast using SDN (SDN-MPI_Bcast). In this implementation, the SDN controller dynamically installs broadcasting rules to SDN switches based on the topology of the interconnect and other jobs running on the same cluster.



Contention Avoidance Between Staging Communication and Inter-process Communication

We propose two conflict avoidance methods to investigate whether the conflict between the two types of communication has influence on the performance of applications.



Keichi Takahashi: takahashi.keichi@ais.cmc.osaka-u.ac.jp
Arata Endou: endou.arata@ais.cmc.osaka-u.ac.jp
Hiroaki Morimoto: morimoto.hiroki@ais.cmc.osaka-u.ac.jp

Cybermedia Center

Osaka University, Japan

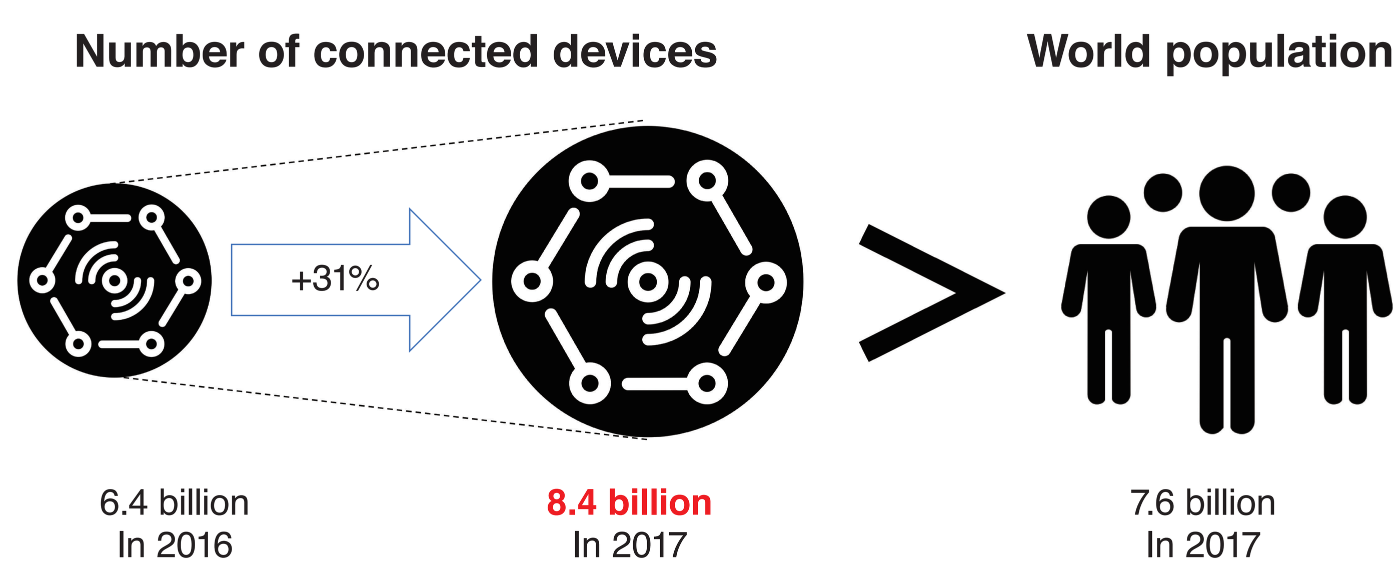
SC17 BOOTH 1943

Access control mechanism & resource sharing platform for IoT Era

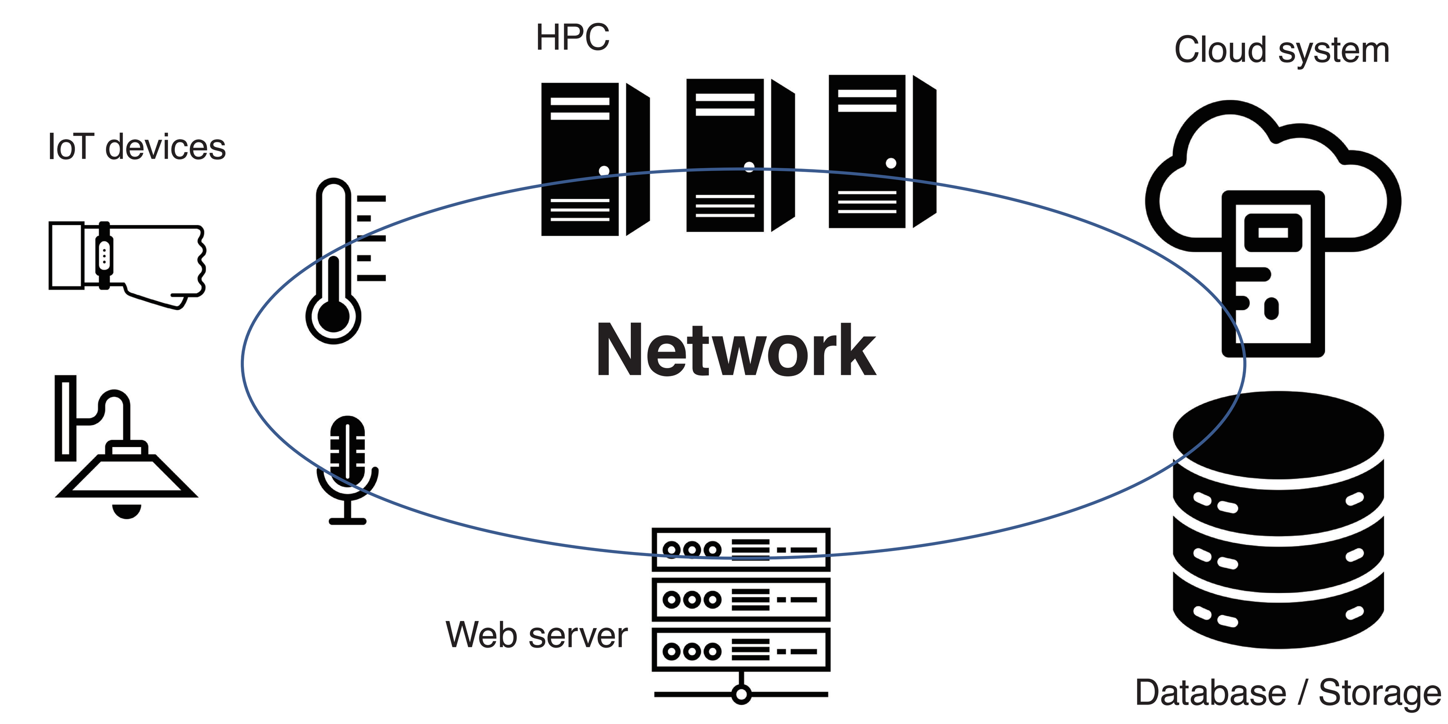
Motivation

Today, there is ***a large number of IoT devices*** distributed widely throughout world. According to the Gartner, Inc. 's forecast, "***8.4 billion connected things*** will be in use worldwide in 2017" [1].

[1] Gartner, Inc., "Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016," http://www.gartner.com/newsroom/id/3598917



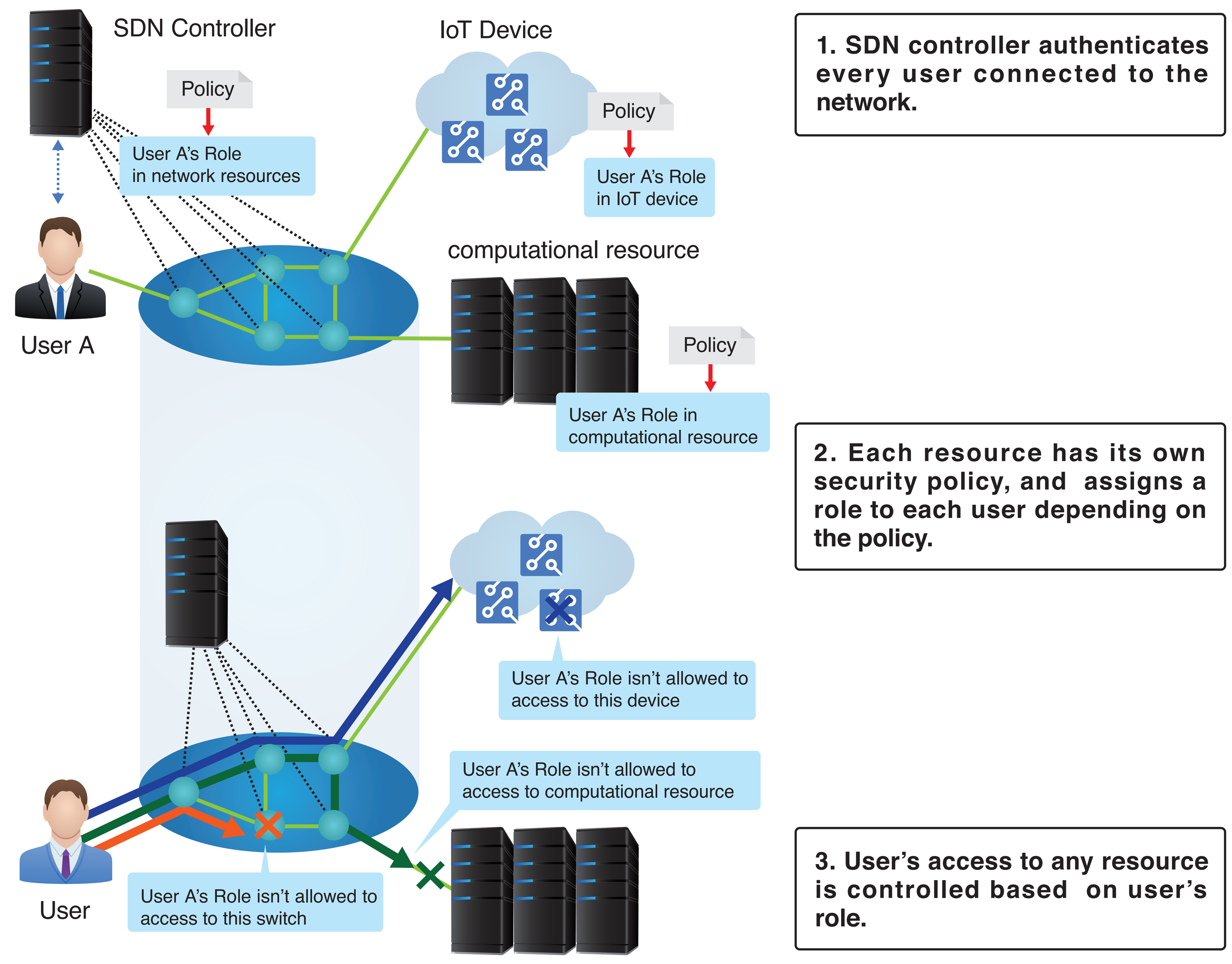
In general, in an IoT system, many types of & a large number of resouces (IoT devices, computational resources, Storage, etc.) are connected by network and cooperated with each other.



Therefore, it is imporant that ***how to manage / connect a large number of resources distributed widely throughout world.***
In thus background, we have been tackling two research themes;
1. FlowSieve: Access control mechanism that dynamically organizes a flexible and secure network linking devices.
2. Sharing Economy of Things Platform: Platform for sharing IoT resources (sensors, computational resources, actuators, etc.)

FlowSieve

To date, various security technologies have been proposed and implemented. However, these technologies have targeted only computational resources.
We propose an access control mechanism ***that targets network resources as access-controllable resources***. We have adopted **RBAC** (Roll Based Access Control) and **SDN** to develop the mechanism. The mechanism works as shown below, and ***provides user-dedicated infrastructure.***



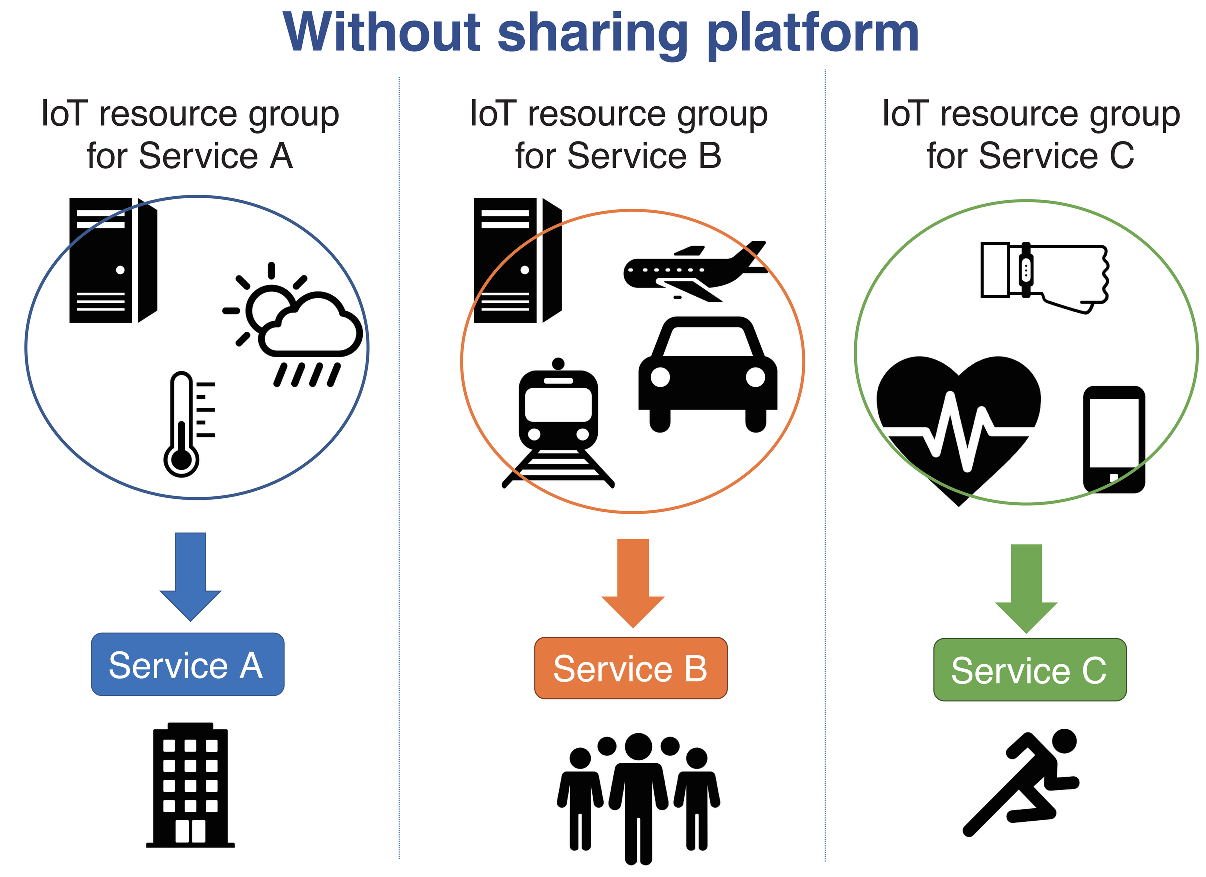
FlowSleve is a preliminary implementation of the access control mechanism. FlowSieve is implemented as an OpenFlow controller program, and can ***control access to network resouces***. However, under the current version of FlowSieve, all devices and network resources are controlled under only one security policy.
In the next plan, we extend FlowSieve so that each resource can have its own security policy towards multi-site network.

GitHub: https://github.com/shimojo-lab/flowsieve

Sharing Economy of Things Platform

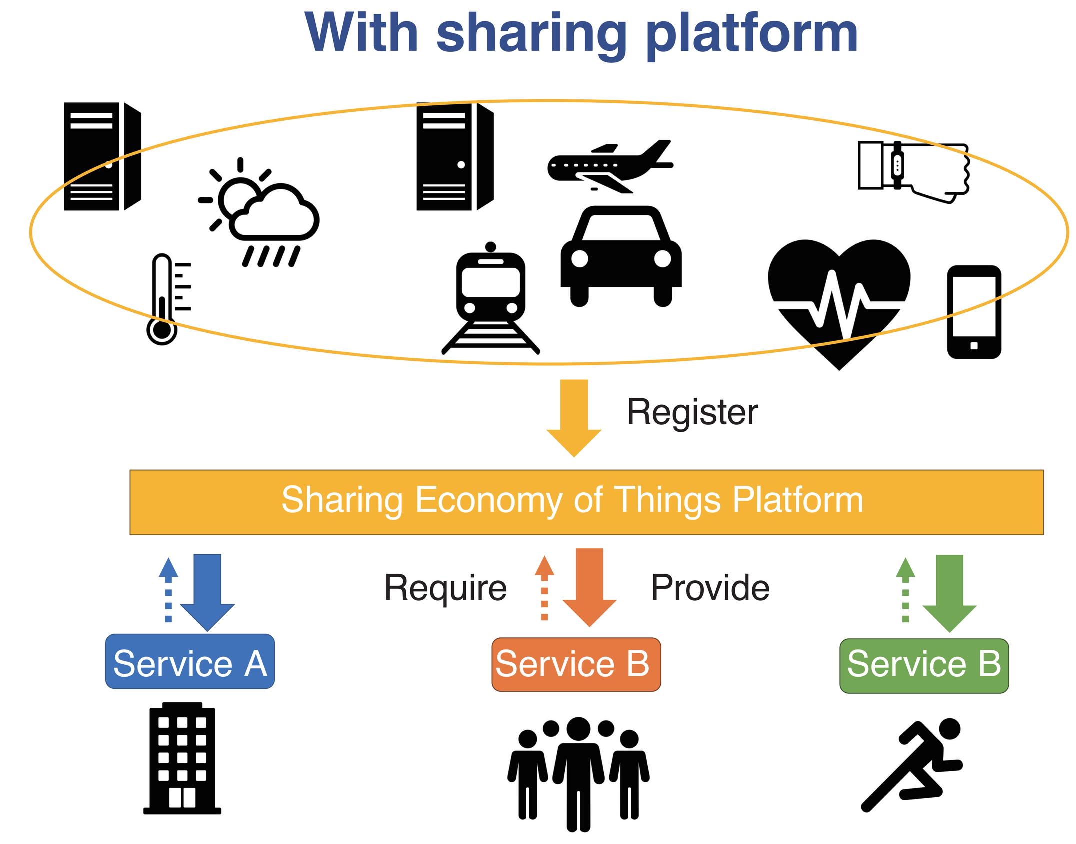
With the arrival of Internet of Things (IoT) era, diversity of IoT devices have been connected to the Internet. However, IoT services have never been popular because of ***the lack of a platform for sharing IoT devices.*** We tackle this problem by proposing a platform for sharing IoT devices based on ***the idea of Sharing Economy.***
Without sharing platform, Each service developer prepares resource group for building each service.

Without sharing platform



With sharing platform, resources are register with the pltaform, and service developers can build services by using registered resources.

With sharing platform



Contact: Takuya Yamada yamada.takuya@ais.cmc.osaka-u.ac.jp

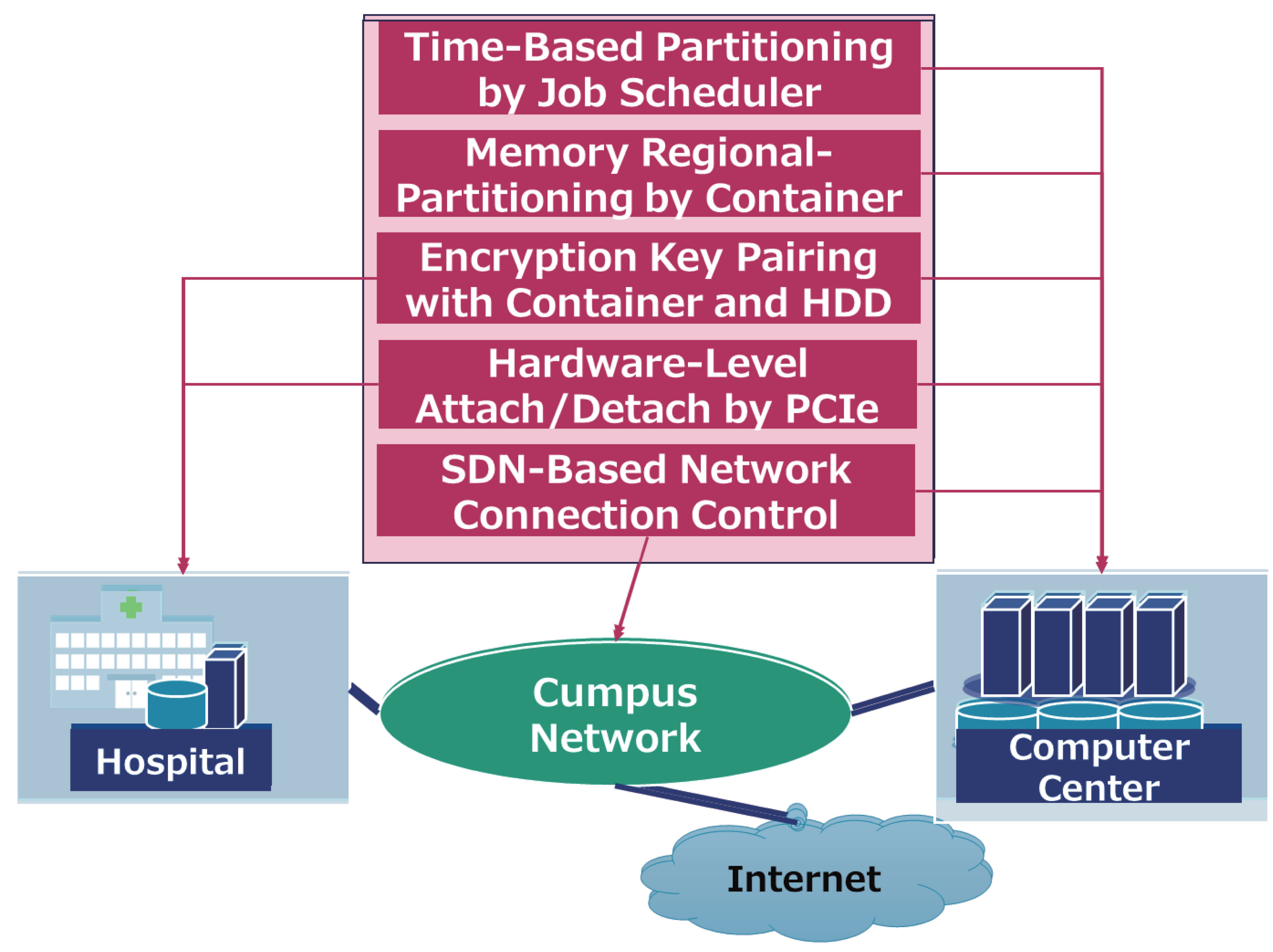
Dynamic and Secure Staging for Medical Data to Be Processed in The Computer Center

Dynamic and Secure Staging for Medical Data to Be Processed in The Computer Center

Dynamic and secure staging mechanism is developed to process medical data by using high performance computers in the computer center. The system is composed of five secure partitioning technologies including time-based partitioning by a job scheduler, memory regional-partitioning by container, hardware-level attach/detach by PCI Express, encryption key pairing with container and hardware device, and SDN-based network connection control. With combination of those technologies, this system provides various enhanced security level appropriate for the medical data concealment of each data to realize approval of medical data usage out of the hospital location.

Partitioning Mechanism and Intensity

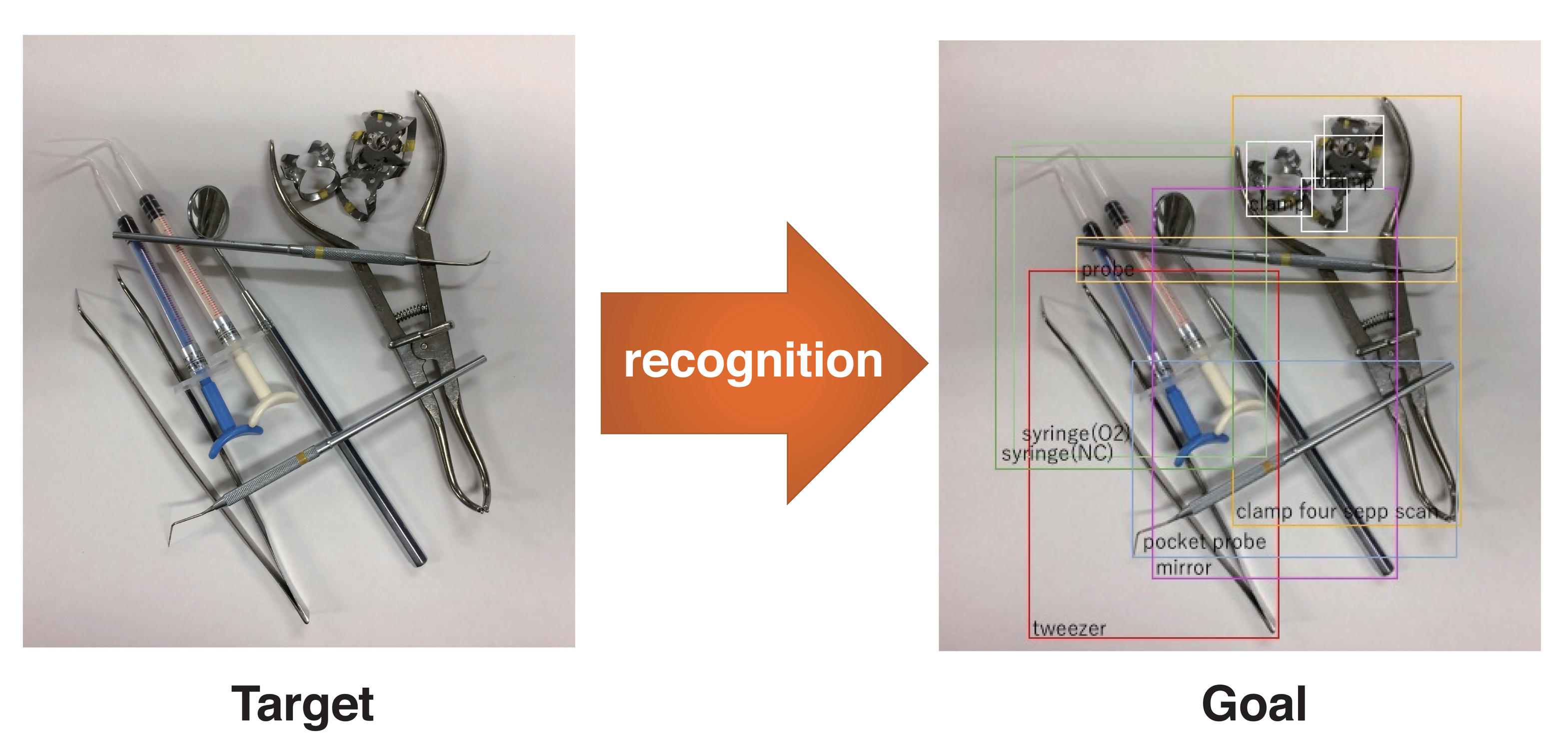
Mechanisum	Description	Partitioning Intensity	
Scheduler	Time-based separation	weak	data load only during job execution
Container	Erase Container after Job execution	weak	clear memory data w / container
EE DISC	Atacch / Detach HDD by ExpEther	strong	equivalent to device detachment
EE NIC	Atacch / Detach NIC by ExpEther	moderate	equivalent to cable disconnect
SDN EE	ExpEther Path Separation by SDN	moderate	partitioning at IP unreachable layer
SDN IP	IP path Separation by SDN	moderate	partitioning at IP unreachable layer
Key Pairing	Encryption Key Pairing w/Disk	moderate	pairing spacific container and dvice



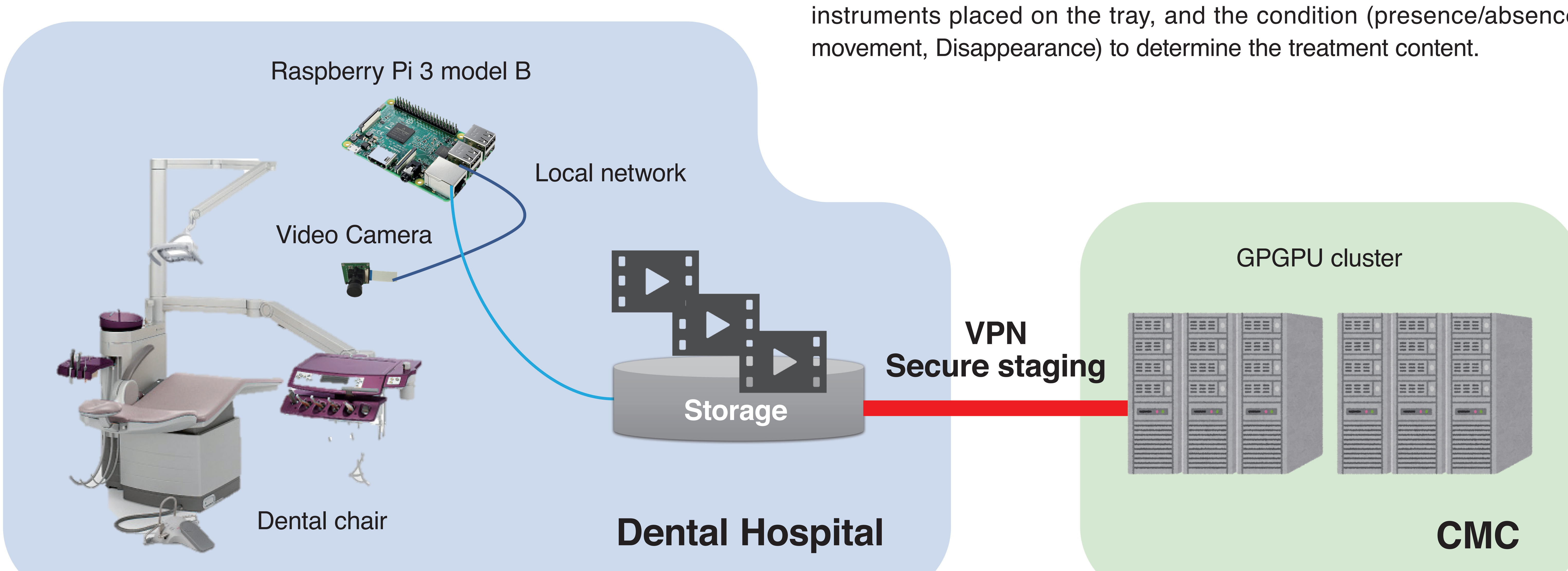
Secure Staging System Diagram and Dynamic Secure Partitioning Mechanism

myDentalAI

myDentalAI is a machine that performs stochastic reasoning. In probabilistic reasoning, it is possible to estimate the state most likely to occur from the combination by considering what kind of probability distribution various events occur. That is, when inputting dental treatment or disease state into a machine, it outputs a remarkable treatment result.Furthermore, myDentalAI gets the behavior history and condition of the dentist in the treatment room. It is necessary for the behavior history to include information that can infer the treatment content. Furthermore, Panorama / Dental images before and after treatment are recognized by AI, and changes in images due to treatment contents are learned. If a consent is obtained from a patient for photographing his/her oral cavity, more accurate suggestion will be available.



Movie recognition: firstly, a tray (table) attached to the dental chair is photographed with a video camera, AI is made to recognize therapeutic instruments placed on the tray, and the condition (presence/absence, movement, Disappearance) to determine the treatment content.



Contact: Oka S, Nozaki K, Hayashi M Osaka University Dental Hospital