JAIST

Demands for Storage Systems from a Customer Viewpoint in Japan

Assistant Prof. Satoshi UDA <zin@jaist.ac.jp>
Research Center for Advanced Computing Infrastructure,
Japan Advanced Institute of Science and Technology (JAIST).

About Me

- Satoshi UDA
  - Assistant Prof. at JAIST (2004-)
    - Research: Internet Architecture
    - Design & operate campus computing infrastructure
  - Research fellow at NICT (2005-)
    - JGN-X: Network testbed
    - StarBED: System testbed system
  - Board member of WIDE Project (2006-)
    - Researchers consortium mainly target to researching on the Internet technology.
  - NOC member of INTEROP Tokyo ShowNet (2001-)
    - NOC generalist (co-chair) of NOC Team (2010-)

JAIST

- Japan Advanced Institute of Science and Technology
  - Japanese national university w/ school of IS, MS, KS.
  - http://www.jaist.ac.jp/
  - Locate at Nomi city, Ishikawa.
- Members:
  - faculty/staff : 300+
  - grad student: 900+
    (incl. foreign students: 300+)

Research Center for Advanced Computing Infrastructure

- Mission
  - Research on computing infrastructure
  - Design, Procure, Deploy and Operate campus computing infrastructure
    - Desktop environment
    - Campus networking systems
    - Storage systems
    - Supercomputers
    - etc.

Research Center for Advanced Computing Infrastructure

The Research Center for Advanced Computing Infrastructure supports users' world-class research and educational environment by providing a high-speed advanced information environment. Based on the FRONTIER Project, a high-speed, high-availability network provides the foundation for the high performance file servers, massively parallel computers, and various servers that have enabled JAIST since its foundation to continuously provide users a convenient information environment in the form of FRONTNET.

- FRONTIER: Our campus computing environment (FRONT + Information EnviRonment)
  - Central operation for whole computers in JAIST
  - 24h/7d service

Services

High-performance Computers

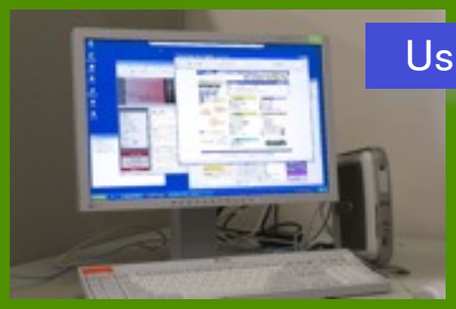
File storages

Campus Network

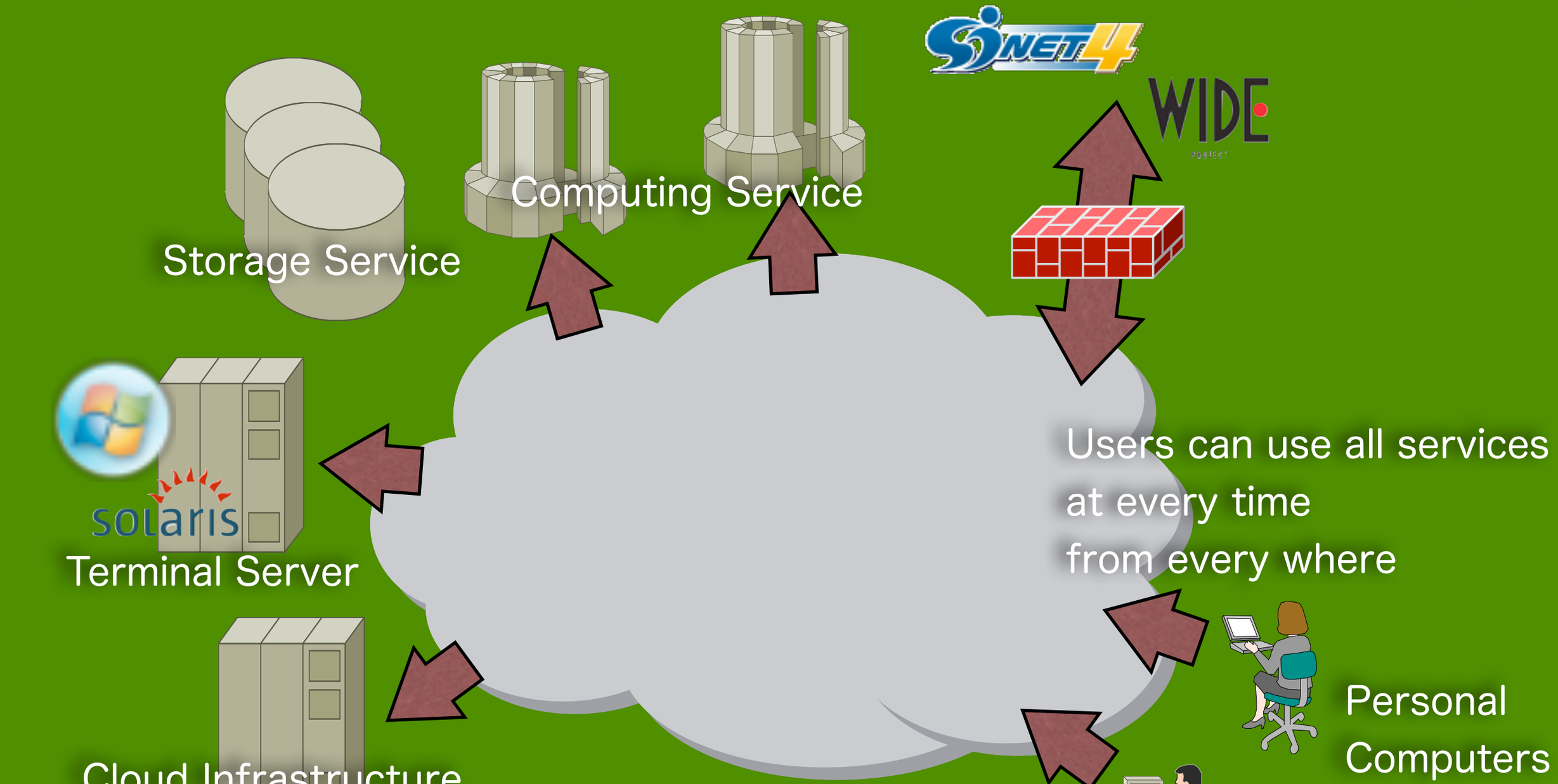
10Gbit/s Backbone (2006-)

User Terminal

Thin Client (2009-)



Overview of our services



Policy of our design

- Purpose:
  - Aspire to become a frontier
- Policy of our design:
  - Try hard with new/innovative technologies / products
  - But, don’t place a burden on users

  → We need a good sense of balance
- We are installing innovative systems on technologies 2-3 years out.
  - Our systems are setting up an example to another Universities and Colleges in Japan.

Products in JAIST

- A10 Networks
- Alaxala Networks
- Alied Telesis
- Brocade Networks
- Cisco Systems
- Cray
- DELL
- D-Link
- EMC
- Extreme Networks
- F5 Networks
- Fujitsu
- Juniper Networks
- Hitachi
- NetApp
- Oracle
- SGI
- VMware
- etc...

Campus Network

- Wide-bandwidth
  - Almost all inter-floor links are 10GbE.
    - 2 * 10G uplinks (WIDE and SINET)
- IPv4/v6 dual-stack
  - Users are using IPv6 w/o any intentional
- Redundancy
  - Almost all floor networks have redundant uplink.
    And we maintain primary/backup routers.

Super Computers

CRAY XC30

NEC SX-9

SGI Altix UV1000

Appro gB222X/1143H Cluster

SGI Altix XE

IBM Cell QS22

Storage services at JAIST

- Service Model
  - Many “data” in JAIST
    - documents owned by students, faculties and staffs
    - experimental data record,
    - etc.
  - Centralize data storing
    - hold down operation cost (incl. human resource)
    - highly-available enterprise systems
- Application
  - End-users’ home directory
    - They can access one home directory from several systems (UNIXes, Windows, Supercomputers, etc.)
  - Share store for project teams, sections, etc.

Requirements at JAIST

- Stability, Availability and Security
  - “Data” is one of most important property at Institutes
    - 24h/364d continuous service
    - secure access control
- Scale
  - parallel access from 1,500 users
- And...
  - multi protocol (NFS, CIFS) service providing for supporting multiple Systems / OSes
  - hold down running and operational costs
  - energy saving, ...

History of our storage systems

- 1990s
  - [NFS] NEWS-OS base systems
  - [NFS] Solaris base systems
  - etc.
- 2005.3 ～ 2009.2
  - [NFS/CIFS] Fujitsu/NetApp NR1000F/F540+R200 (60TB)
- 2007.3 ～ 2011.2
  - [CIFS] Hitachi SunRize AMS500
- 2008.3 ～ 2012.2
  - [NFS/CIFS] SGI IS4500F + Onstor Bobcat (100TB)
  - [iSCSI] EqualLogic PS400E (50TB) iSCSI
- 2009.3 ～ 2013.2
  - [NFS/CIFS] DELL EqualLogic PS5000X + PS5000E + Solaris (150TB)
  - [NFS] DDN S2A9900 + S2A6620 (1PB)

Operating Storage Systems

- 2011.3 - 2015.2
  - [CIFS] EMC Celerra NS-480 + NS-120 (100TB)
    - featured: remote replication for DR, stability
  - [iSCSI] DELL EqualLogic PS6510E + PS6010S
    - featured: Scale-out technology
- 2012.3 - 2016.2
  - [NFS/CIFS] Fujitsu/NetApp FAS3270 (250TB)
    - featured: secure NFS, ACL mapping
- 2013.3 - 2017.2
  - [NFS/CIFS] DELL Compellent w/ zNAS (1.3PB)
    - featured: hierarchical architecture, flash device

“IPv6”
that is great solution,
but has not have reality yet

IPv4 address has been depleted

- IPv4 addresses has been depleted,
  it'a serious condition in asia pacific region
  - some SPs looking for buyable IPv4 addresses
    around the world.
  - now, IPv4 address is a cost

- We need save using IPv4 address
  - use snippets...
  - cannot reserve for future use...

Scale-out storages use many IP addresses

- Scale-out storages use many IP addresses
  - Typically require one or more IPs per node
  - Total number will increase with adding nodes
- VM HVs require many IP addresses, too
  - Require on IP per I/F in multi-path case
  - Total number will increase with adding nodes

Renumber, Renumber, ...

- Storage Network subnet for cloud infrastructure need many IPs, and the number will increase
  - ex) we faced need to expand prefix-length
    - /25 -> /24 in 2011, /24 -> /23 in 2013

- Should we reserve addresses for future use?
  - No! we cannot keep for future use.
- Should we use private address?
- Should we use non-IP technology e.g. FCoE?
  - No! we want to keep global IP address to keep versatility

IPv6 is comming

- Wow! IPv6 release our mind from the number of IPs on each subnet. we can use many and many addresses on each subnet!!
- We've tried to switch to IPv6
  - Our storages is working fine with IPv6 !!
  - VM HVs are basically working fine with IPv6, but some important features not work :(
  - Some add-on softwares and management softwares have not support IPv6 yet.
- We strongly request vendors that storage and HV softwares fully support IPv6 soon!!

“Virtualize and Scale-out Technology”
that add flexibility to storage operation,
but raise complexity

Virtualize technology open up storage administration to site-admins

- On legacy system
  - It is very difficult to re-construct storage system configuration e.g. changing volume structure, share structure, because we need care from disk layer, storage network configuration, etc.
    - We cannot administrate storage systems by our self because almost all sites don't have storage specialists.
    - When we want to reconstruct or change configuration, we need order this to SIer and call specialists.
    - And also we do this under deep design consideration.
- Systems based on virtualize technology open up storage administrator to site-admins who has no storage speciality
  - because of hiding low layers
  - we can add/remove volumes/shares every time w/o special knowledge and deep consideration.

virtualize on vertualize on virtualize on ...

- Today's systems are very complex
  - Sometimes systems are on multiple virtualize technology.
    - e.g. ZFS filesystems on Scale-out iSCSI storage
  - Many virtualize technologies are vender proprietary, s.t. black-box
    - We are more than scary when we imagine that it breaks down by worst fortune

- However we adopt systems using virtualized technology because of flexibility merit
  - with believing it will not break down
  - with faith in developer and vendor

Measurement is important to operate systems

- on virtualized system
  - easy to operate system
  - difficult to measure the system load
    - Is this system busy or not??
    - how much additional users can handle?

- It is important for us to monitor performance, but we feel that there are many systems which this function is not enough.
  - Please think about these systems often operated by non-specialist site-admins.

Open Architecture v.s. Proprietary System
Think about Cost, Stability and Risk

Cost, Stability and Risk
Popular perception in Japan

- Open Architecture
  - Low Cost,
  - Low Stability,
  - and High Risk

- Proprietary System
  - High Cost,
  - High Stability,
  - and Low Risk

Difficulty on choosing opensource

- In many cases in Japan,
  the new systems are going to install
  under the leadership of SIer.
  - many Japanese users (companies) don't want to bear a risk by oneself
  - almost all major SIers cannot support opensource systems

- However some ventures which provide support service for opensource software growing up

Cost, Stability and Risk
Fledgling new perception

- Open Architecture
  - High Cost,
  - Middle Stability,
  - and Low Risk
- Proprietary System
  - Low Cost,
  - Middle Stability,
  - and High Risk

Long Long path from customer to developper

- Escalation path from Japanese customer
  - Local branch of SIers
  - Japan HQ of SIers
  - (OEM company)
  - Japan Branch of product vendor
  - HQ of product vendor
- We need long RTT in communicate with system developer / architect.

- We want to request developers to make a chance to talk direct with your customers.

Evaluation and Horner from public (vendor neutral) place is important

INTEROP Tokyo:
The Largest ICT Event in Japan

The biggest Live network demonstration with over 100 companies and 80 engineers. Challenges to provide stable network services with cutting edge technology every year.

Around 200 ICT leaders gives over 70 educational program during the show about network infrastructure, cloud and virtualization etc.

INTEROP

Launched in 1994. 2013 was the 20th show in Japan.

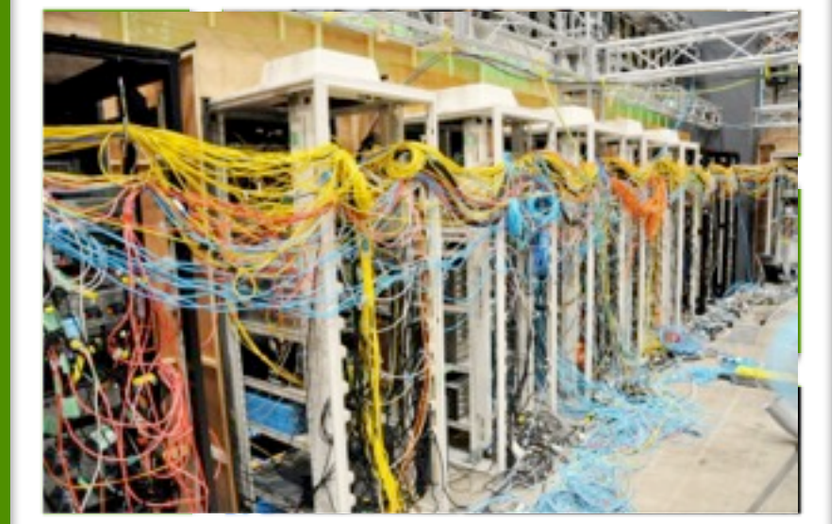
Around 130,000 business and technology professionals gather and meet the latest technology, products and services.

INTEROP Tokyo: “ShowNet”

- The largest Live Demonstration network
- The network is constructed by
  - Contributed Products,
  - Contributed Technologies, and
  - Contributed Prototype Products.
- The largest TEST and PRODUCT network.
  - Providing connectivity to participated companies and visitors.
  - However, it is a TEST network of new technologies.
- The network shows the near future of the Internet.
  - We aim to show the network of 3 to 5 years future.

INTEROP Tokyo:
Schedules of Constructing ShowNet

- From September 2012
  - Design and Planning of Network was started.
  - 28 NOC members had a meeting per month
- HotStage (May 30 - June 11)
  - Constructing Networks and Testing Interoperability
  - Over 100 people works all day and all night during the 13 days
- Main Show (June 12 - June 14)
  - Exhibition of 3 days



INTEROP Tokyo:
Challenges of ShowNet 2013

- Interoperability of Backbone / Core Network Technologies
- Verification New Datacenter Technology
- Introducing Software Defined Networking for Providing Customer Networks
- Verification IPv6 Migration Technologies
- Providing Comfortable Wireless Environment and WiFi Location Services

INTEROP Tokyo: ShowNet Contributors

SPONSER

AlaxalA NEC NTT Communications CISCO JUNIPER NETWORKS 東陽テクニカ "Your Best Partner" NISSHO ELECTRONICS JNETWORK INITIATIVE

CO-SPONSER

isa IXIA vmware A10 Customer Driven Innovation ATEN SII NTTAT AllOt PeerApp

FXC Schneider Electric NICT SEIKO DELL IBM hp HitachiCable

HUAWEI FireEye FORTINET FUJITSU FUJITSU Plat'Home FLUKE networks FURUNO SYSTEMS

BROCADE MKI YAMAHA

SUPPORTER

IT Frontier Infoblox Web Service Development NCLC ORIZON クオリティネット株式会社 JUNIPER NETWORKS soltec NCR AFFILIATE

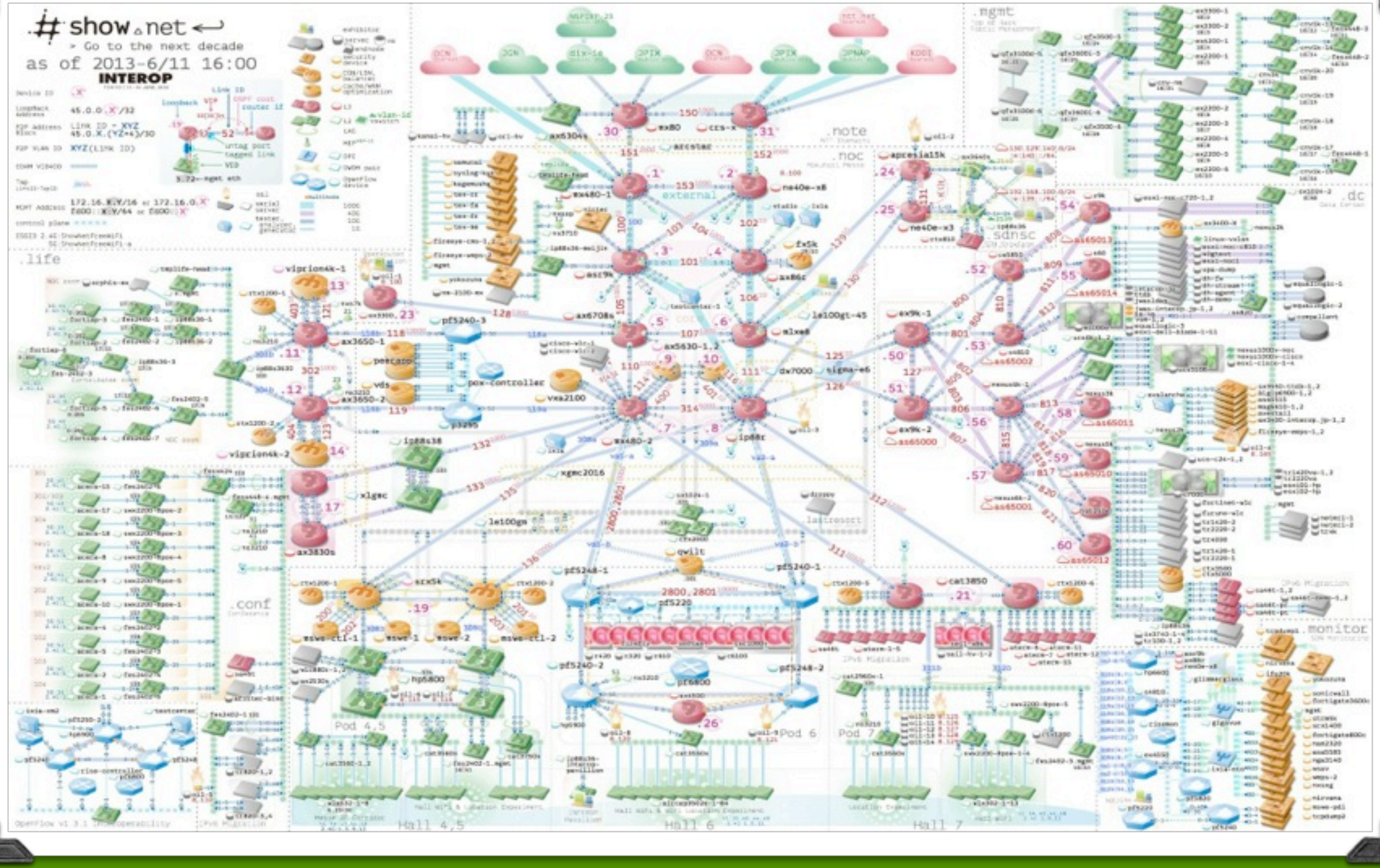
DYDEN comwic DATAHOTEL 東京大学 NTT西日本 jpix Bitrieve MAIN TECHNOLOGY Mellanox ALTIMA

RISO LogicVein

特別協力企業・団体

(株)アールシーエー
(株)インターネットイニシアティブ
インターネットマルチフィード(株)
金数芸術科学大学
慶應義塾大学
KDDI(株)
国立情報学研究所
さくらインターネット(株)
(株)Xenlon
センターピア(株)
ドコモ・システムズ(株)
トランスコスモス(株)
(株)ナビタイムジャパン
奈良先端科学技術大学院大学
日本テレガートナー(株)
古河ネットワークソリューション(株)
北陸先端科学技術大学院大学
WIDEプロジェクト

INTEROP Tokyo: ShowNet 2013



We are not end-user,
we have real end-users behind us

Service should be simple, easy to understand

- We are providing storage services by NFS & CIFS in JAIST
  - Users can access same contents from both protocol

- Topic in last 2-3 years is “ACL mapping”.
  - Users don’t care protocol who are using.
  - So we start to try enable mapping feature between NFSv4 ACL and CIFS ACL.
    - Write this feature on our procure specification.
    - But.... not working well on some systems. :(

End user make an unexpected request

- We are providing snapshot by automatic mount
  - Users can access their snapshop (old) files every time, it's useful...
- A day a user say...
  - "How can I delete files under snapshot directory?"
  - He was deleting files which had beed used in his joint research project.
  - The contract for the joint project was require deleting all files after finishing project period.
  - And he found files under snapshot directory....
- This is one of real end-user's voice.
  - We need to think how to solve this problem.

EOF