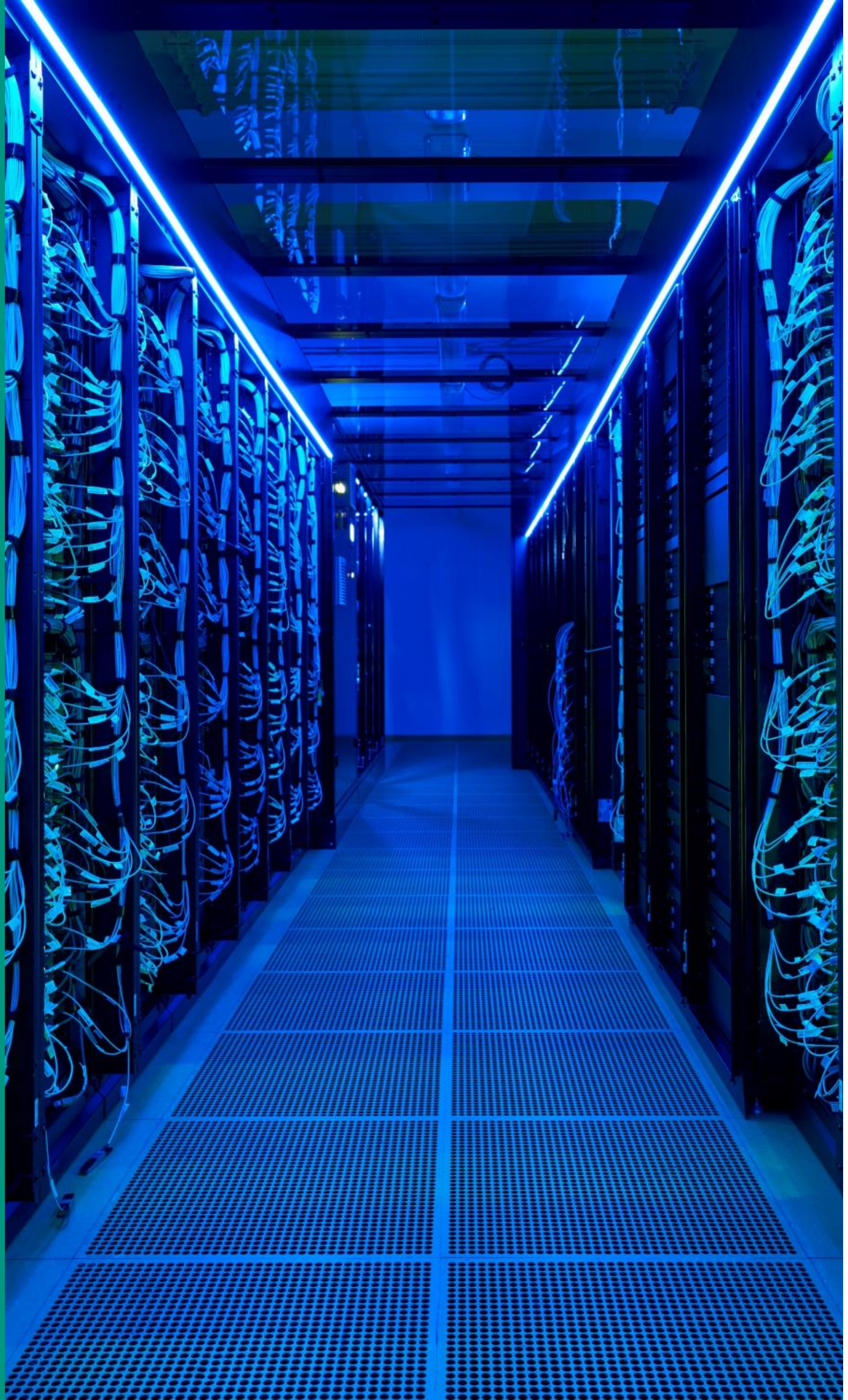
DE-KIT Network Update

LHCOPN-LHCONE meeting #55 – 07-09/10/2025

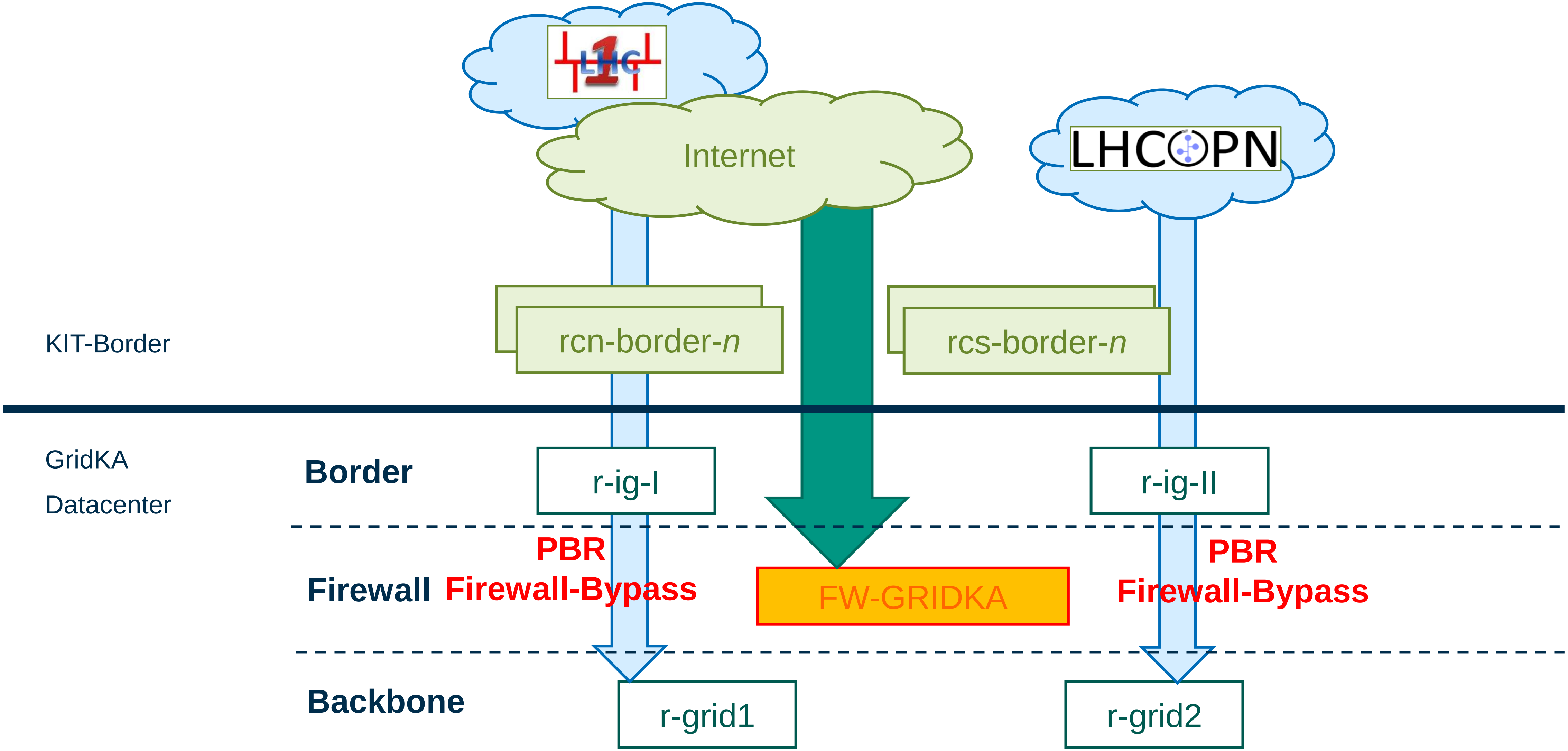
Aurélie Reymund, SCC/NET, 07.10.2025.



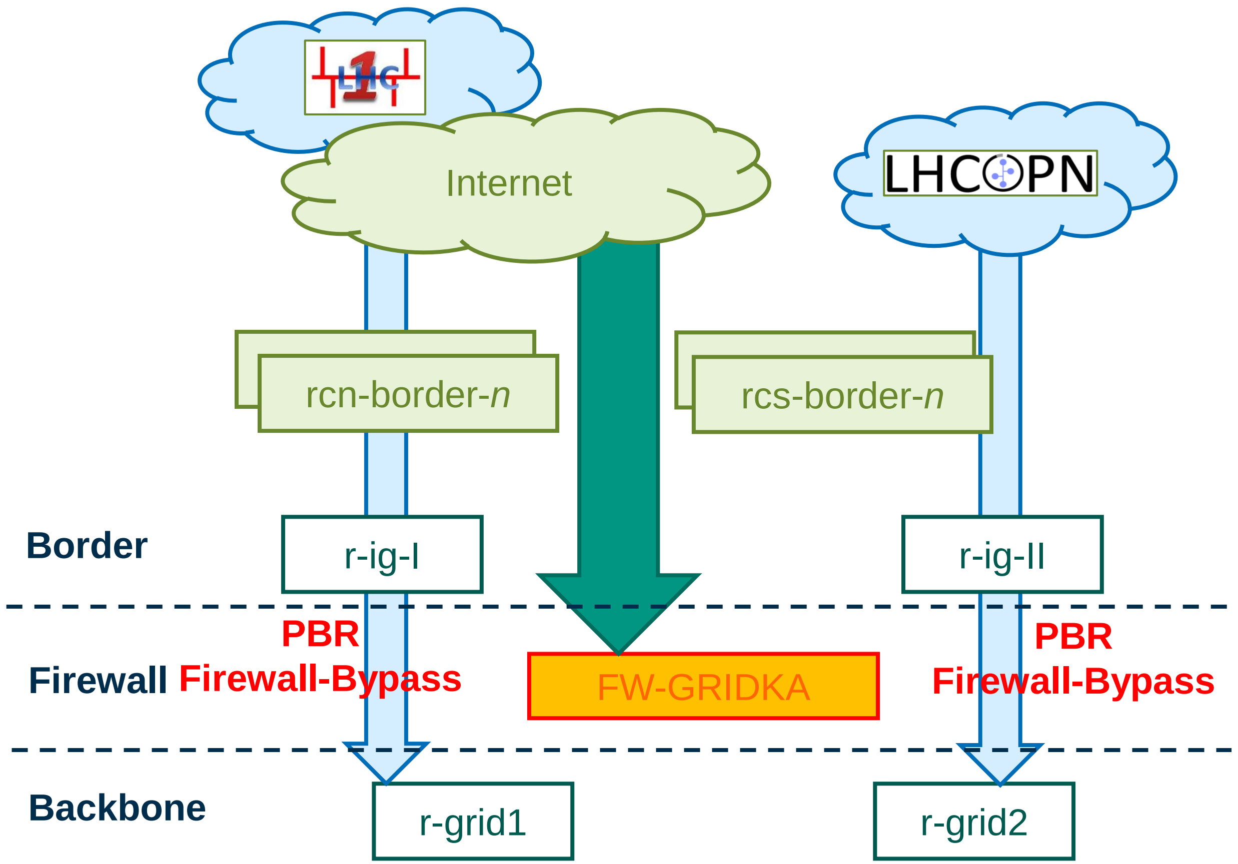
Content

1. **The network before update**
    - What we no longer want
    - The firewall problem
2. **Motivation**
3. **The new network**
    - Bandwidth requirements
    - IP addressing
    - Design concept
4. **Status of the migration**

The Network Before Update

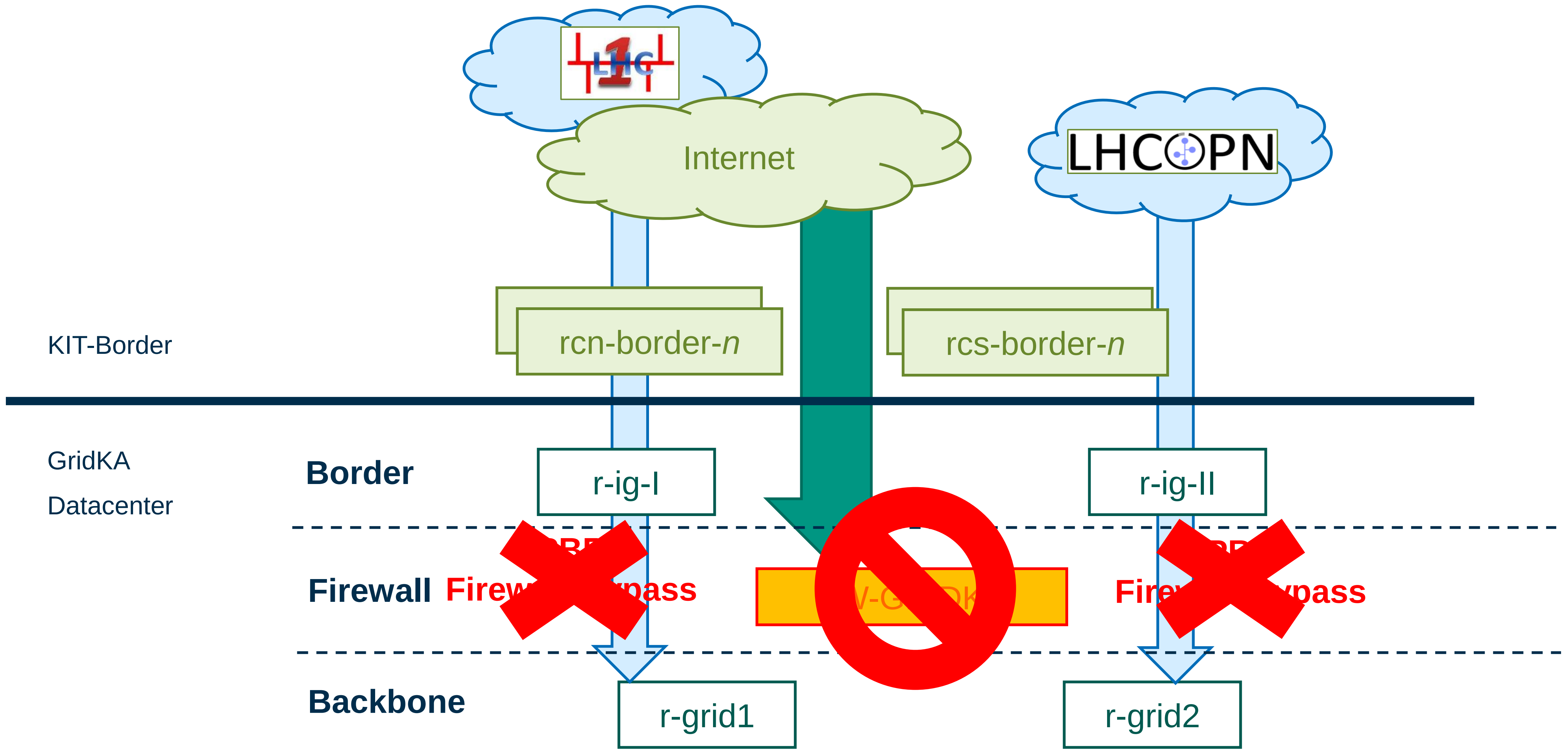


The Network Before Update

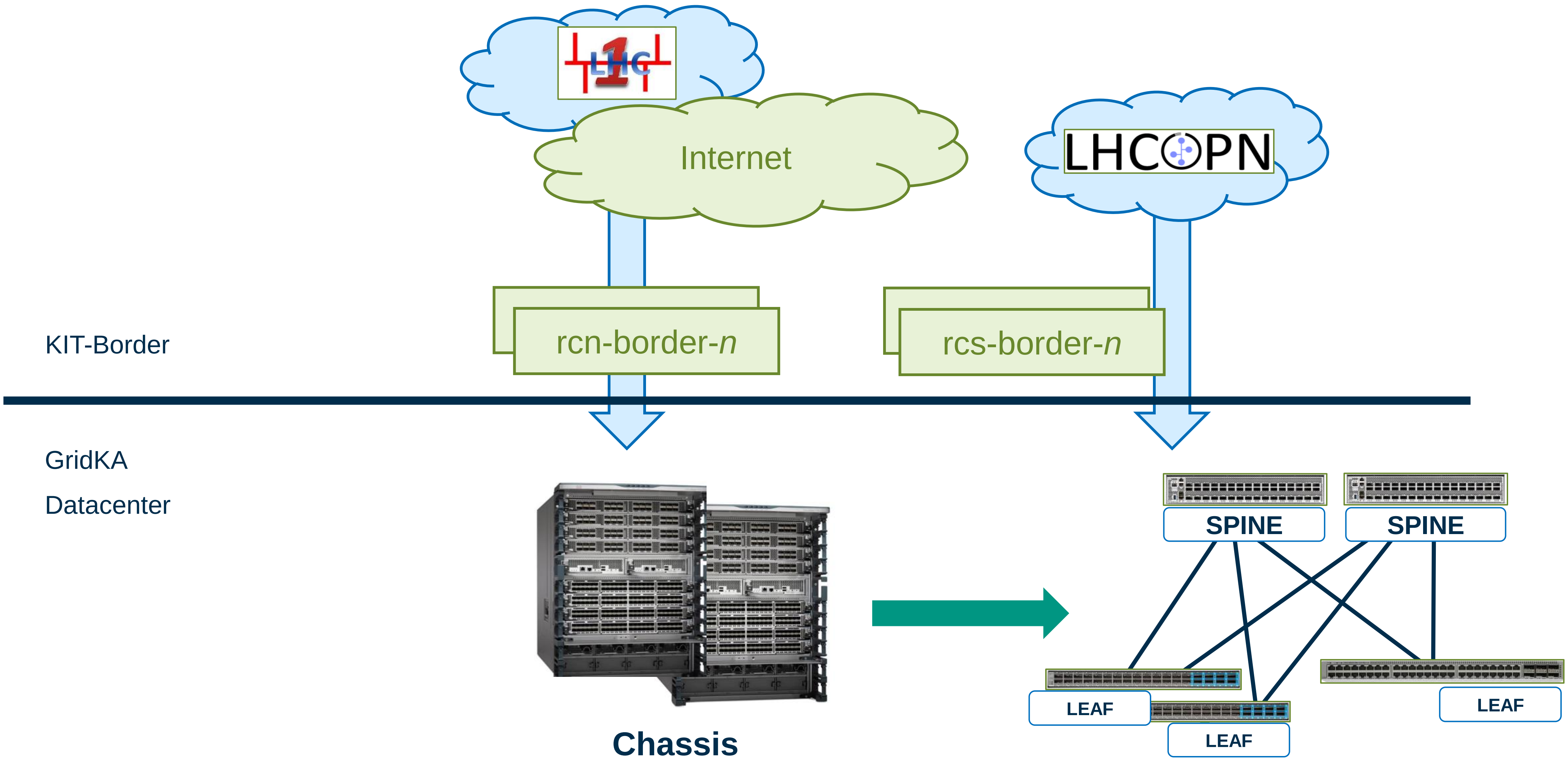


- Historical infrastructure:
  - Unchanged design since years
  - Complexity grew up with time
    - ex.: maintening of LHCOPN/ONE filters
    - WN had only private IPv4-addresses -> need NAT to communicate to outside
    - Filserver need to transfer a huge amount of data
    - Firewall as bottleneck -> need to build firewall-bypass
    - …
- Datacenter network topology different from KIT network topology:
  - Redesign of KIT-Core network in 2016
  - Since 2019 in production
  - From classical backbone ring to spine/leaf architecture
- GridKa „as an island“

DE-KIT Network Update – What we no longer want (1/2)



DE-KIT Network Update – What we no longer want (2/2)



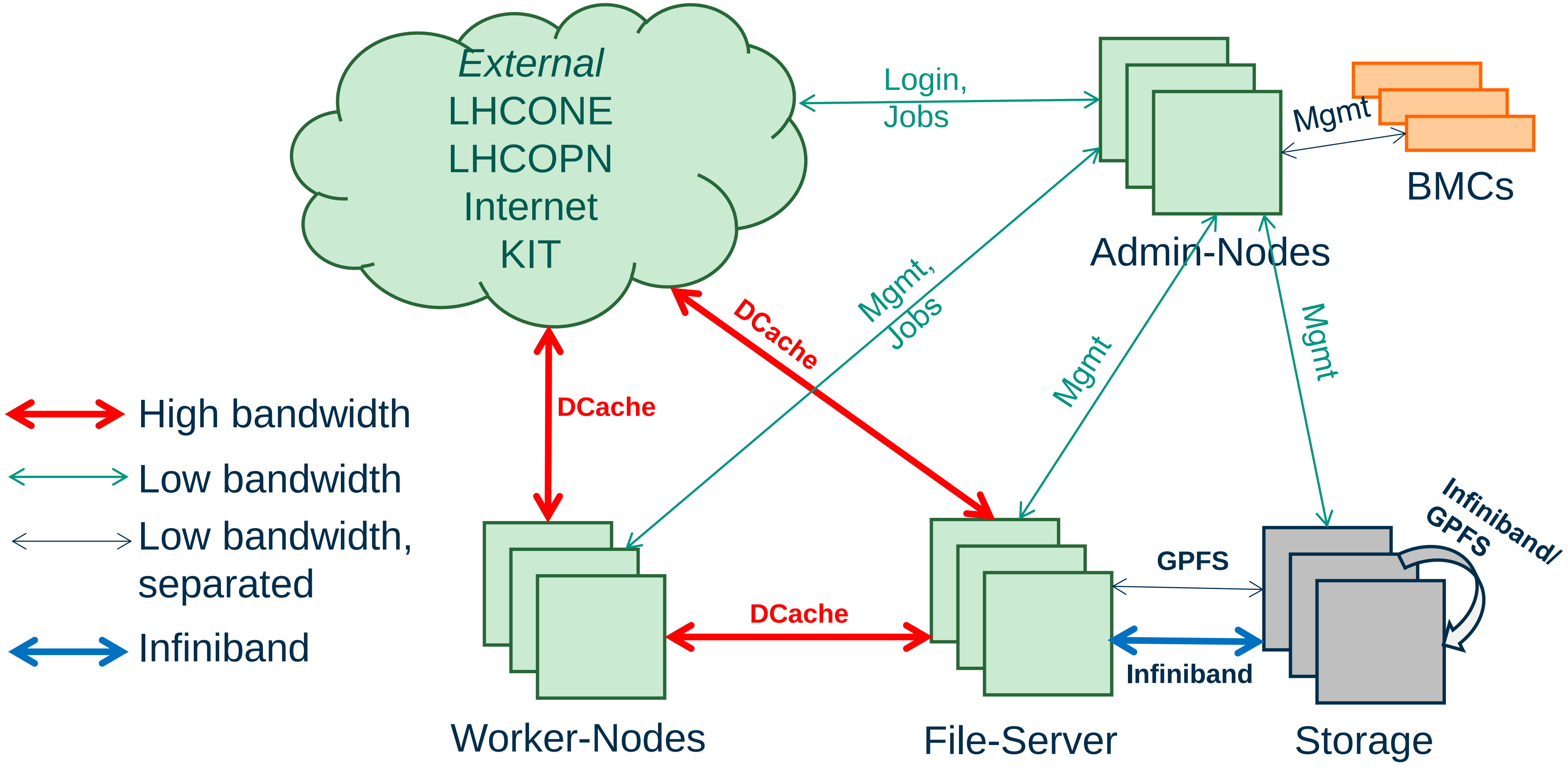
DE-KIT Network Update – The Firewall Problem

- Bottleneck
  - Need to build Firewall-bypass to increase bandwidth
  - Or buy a firewall with a high bandwidth (-> expensive)
- Firewall Bypass
  - PBR is confusing, hard to extend, hard to debug
  - Need to configure ACLs with thousands of entries (true!)
  - Very complex
- Current Firewall reaches end of lifetime

DE-KIT Network Update - Motivation

- **Simplicity**
  - ✓ Simple network setup, simple deployment of new services
  - ✓ Design integrated in the KIT-Core network design
- **Consistence**
  - ✓ Configuration (ex. IPv4/IPv6)
- **Flexibility and extensibility**
  - ✓ The network should be easily expandable for more bandwidth, new services
- **High availability**
  - ✓ Redundancy
  - ✓ Upgrade of network component should be transparent
- **Security**
  - ✓ Systems are protected via firewall
- **Integration of the GridKa network administration into the network department**
  - ✓ Thanks to consistent platform, tools and deployment scenarios

DE-KIT Network Update – Bandwidth Requirements



DE-KIT Network Update – IP addressing (1/2)

- Historical: every host has a private IPv4-address
  - Central firewall was used for outside communication
  - IPv4 NAT, shared public IPv4
  - Also used for provisioning
- Public IPv4 needed for FS and Admin-Nodes: separate vlan

➢ **What about IPv6? No need of a firewall to NAT**

  - Decentral Firewall: no IPv4 sharing anymore
  - Public IPv4 was bought from market (2021)
  - 157.180.228.0/22 and 157.180.232.0/22

✓ **Every Worker/Fileserver/Adminserver gets a public IPv4 address**
✓ **Private IPv4 address still needed for provisioning**

F01-120
- Vlan **xx** name f01-120
- 10.65.120.0/24
- 2a00:139c:2::/n

- Vlan **xy** name f01-120-e
- 192.108.46.x/y
- 2a00:139c:5::/n

A01-001
- Vlan **zz** name a01-001
- 10.33.1.0/24
- 2a00:139c:2::/n

- Vlan **yy** name a01-001-e
- 192.108.45.0/24
- 2a00:139c:3::/n

C01-120
- Vlan **yy** name c01-120
- 10.1.120.0/24
- <no_public_ipv4>
- 2a00:139c:4::/n

DE-KIT Network Update – IP addressing (2/2)

- 1 VLAN per Rack
  - Untagged connection
  - Private IPv4 and IPv6 on subnet -> *used for provisioning!*
  - Public IPv4 and IPv6 routed (BGP)

✓ Modern, simple design
✓ No additional VLAN on network interface
✓ No condition with asymmetrical routing
✓ Same network during lifecycle
✓ All public IPv4 are usable

New

F02-120
- Vlan **xx** name f01-120
- 10.65.120.0/24
- 2a00:139c:8:1::/n

loopback public
157.180.u.w/32
2a00:139c:8:0:1000::/128

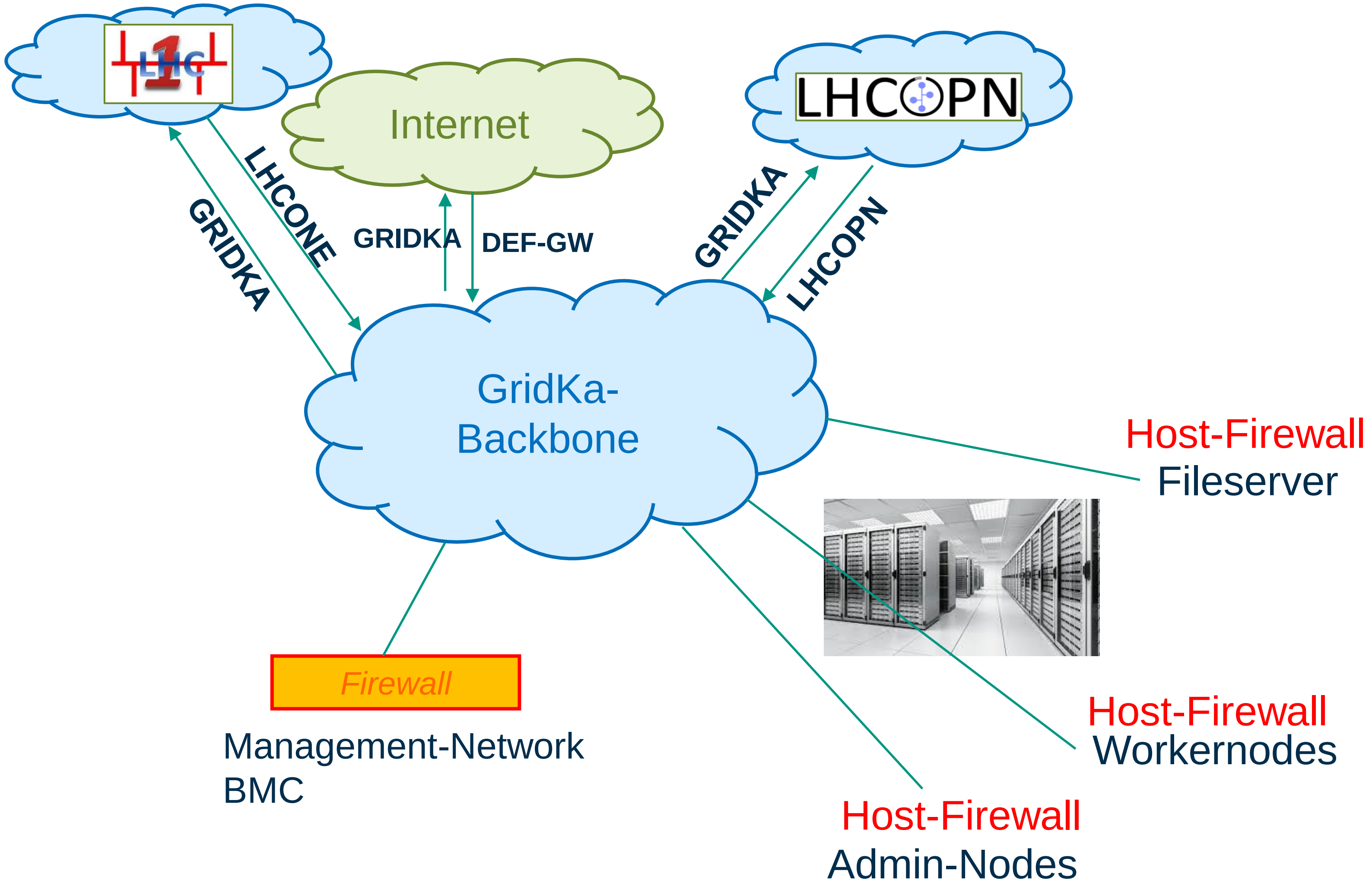
A02-001
- Vlan **zz** name a01-001
- 10.33.1.0/24
- 2a00:139c:a:1::/n

loopback public
157.180.u.w/32
2a00:139c:a:0:4000::/128

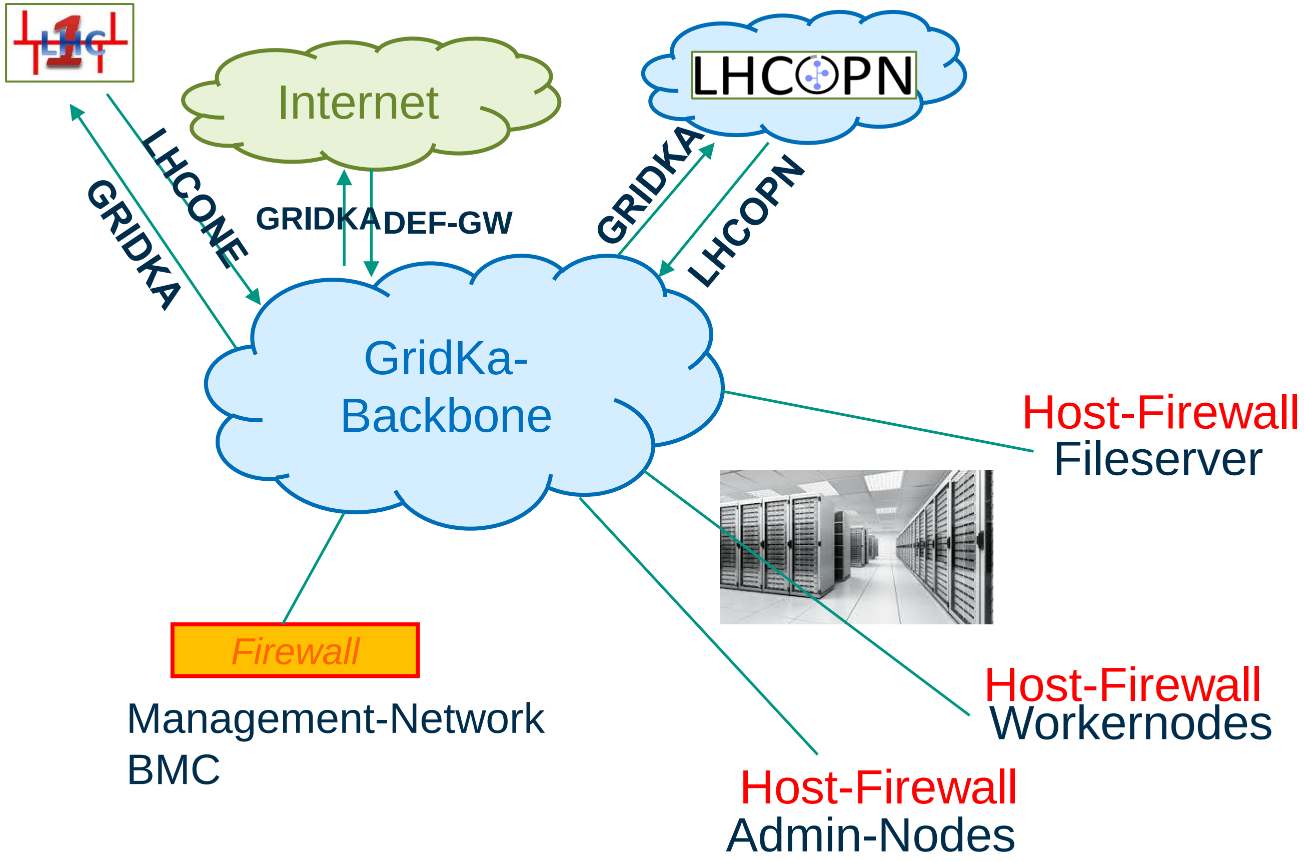
C02-120
- Vlan **yy** name c01-120
- 10.1.120.0/24
- 2a00:139c:9:1::/n

loopback public
157.180.u.w/32
2a00:139c:9::/128

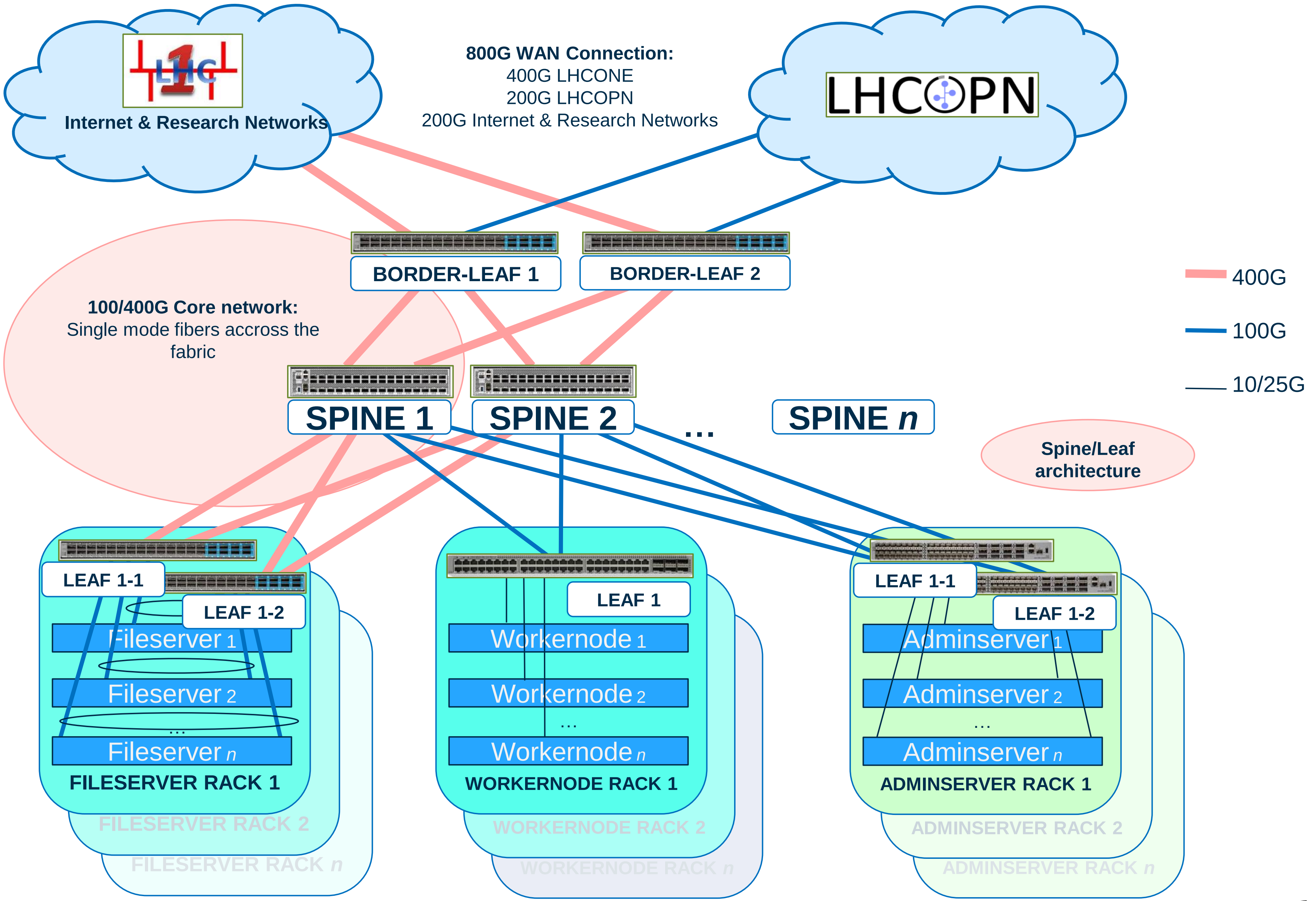
New Network Design Concept (1/2)



New Network Design Concept (2/2)



- High bandwidth for storage and worker nodes
- Decentral firewall (move to hosts)
  - Firewall bandwidth grows with number of hosts
- Dedicated firewall only for special hosts like appliances, BMCs, etc
  - Local firewall not possible / trusted / inflexible
  - Black boxes, unknown OS patch level
- PBR for firewall bypass is not needed anymore
- Align IPv4 and IPv6 network design
- Spine/leaf architecture -> reuse well known design patterns from KITnet


800G WAN Connection:
400G LHCONE
200G LHCOPN
200G Internet & Research Networks
Internet & Research Networks
LHCOPN
BORDER-LEAF 1
BORDER-LEAF 2
100/400G Core network:
Single mode fibers accross the fabric
SPINE 1
SPINE 2
...
SPINE n
400G
100G
10/25G
Spine/Leaf architecture
LEAF 1-1
LEAF 1-2
Fileserver 1
Fileserver 2
...
Fileserver n
FILESERVER RACK 1
FILESERVER RACK 2
FILESERVER RACK n
LEAF 1
Workernode 1
Workernode 2
...
Workernode n
WORKERNODE RACK 1
WORKERNODE RACK 2
WORKERNODE RACK n
LEAF 1-1
LEAF 1-2
Adminserver 1
Adminserver 2
...
Adminserver n
ADMINSERVER RACK 1
ADMINSERVER RACK 2
ADMINSERVER RACK n

New Network – Migration Status

- **Compute-Node-Racks**
  - 100% complete
    - ✓ Routed in the new network
    - ✓ 2 racks phys. connected to the new infrastructure
    - ✓ 13 racks phys. connected to the old infrastructure
    - ✓ 5 new racks upcoming

- **Fileserver-Racks**
  - Work in progress…
    - ✓ 1 rack phys. connected to the new infrastructure
    - ✓ ~ 12 racks to migrate

- **Adminserver-Racks**
  - Work in progress
    - ✓ Mainly virtual host, hardware remains for special purpose

The plan is: end of the migration in 2026

When all CN, FS, Adminserver have moved to the new network -> move LHCOPN/ONE links

KIT