Silicon Hill

Vysoká dostupnost

Bronislav Robenek, 4. Prosinec 2014

Osnova - Co vás čeká

- Úvod do HA a redundantních systémů
- Základy redundance na L2 (STP, PC, VPC)
- Základy redundance pro FH (VRRP/HSRP)
- Základy redundance na L3 (OSPF)
- Redundance na internetu (BGP)
- Vysoká dostupnost aplikací/serverů

Vysoká dostupnost

- 99.999% -> 5.26 min/rok downtime
- Uptime vs. Reachability vs. “funguje to”
- SLA na Uptime, RTT, …
- Redundance
  - Active/Passive (Master/Backup)
  - Active/Active (Master/Hot Standby)
  - Active/Active + Load balancing

Vysoká dostupnost v IT

Uživatel

Application (Software)

Platform (OS, virtualizace)

Infrastructure

Compute
(CPU, RAM)

Storage
(LS, SAN)

**Networking
(LAN, WAN)**

Hardware (servery, síťové prvky, disky, elektřina …)

Síťové vrstvy a HA

#	ISO/OSI	TCP/IP	Jednotka	
7	Application	Application	Data	DNS, DDNS
6	Presentation			
5	Session			
4	Transport	Transport	Segment	TCP, MTCP
3	Network	Network	Packet	FHRP, IGP, BGP
2	Data link	Link	Frame	STP, PC, VPC, FabricPath
1	Physical		Bit	APS (SONET)

Redundance na L2

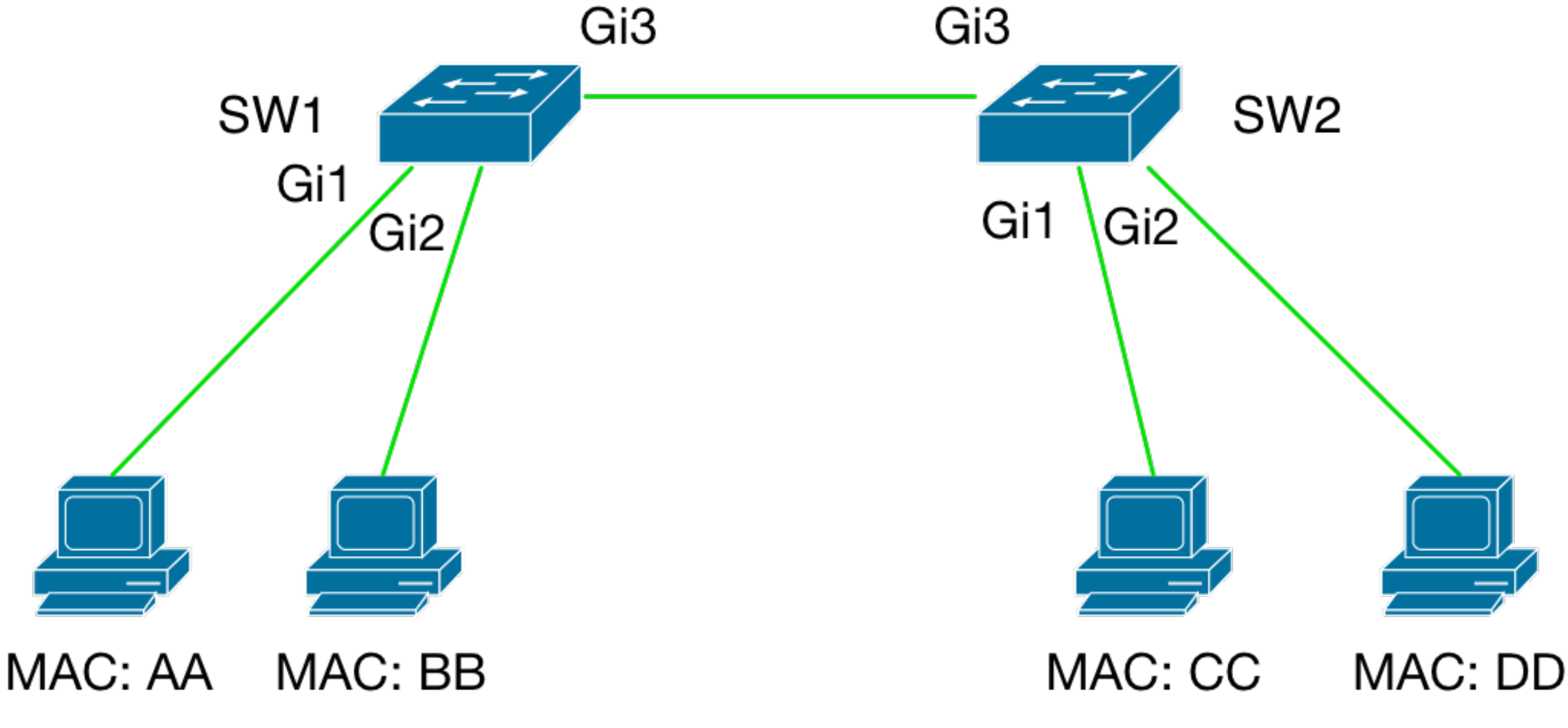
Active/Backup - STP

SW1 L2 FIB

PORT	MAC
Gi1	AA
Gi2	BB
Gi3	CC DD

SW2 L2 FIB

PORT	MAC
Gi1	CC
Gi2	DD
Gi3	AA BB

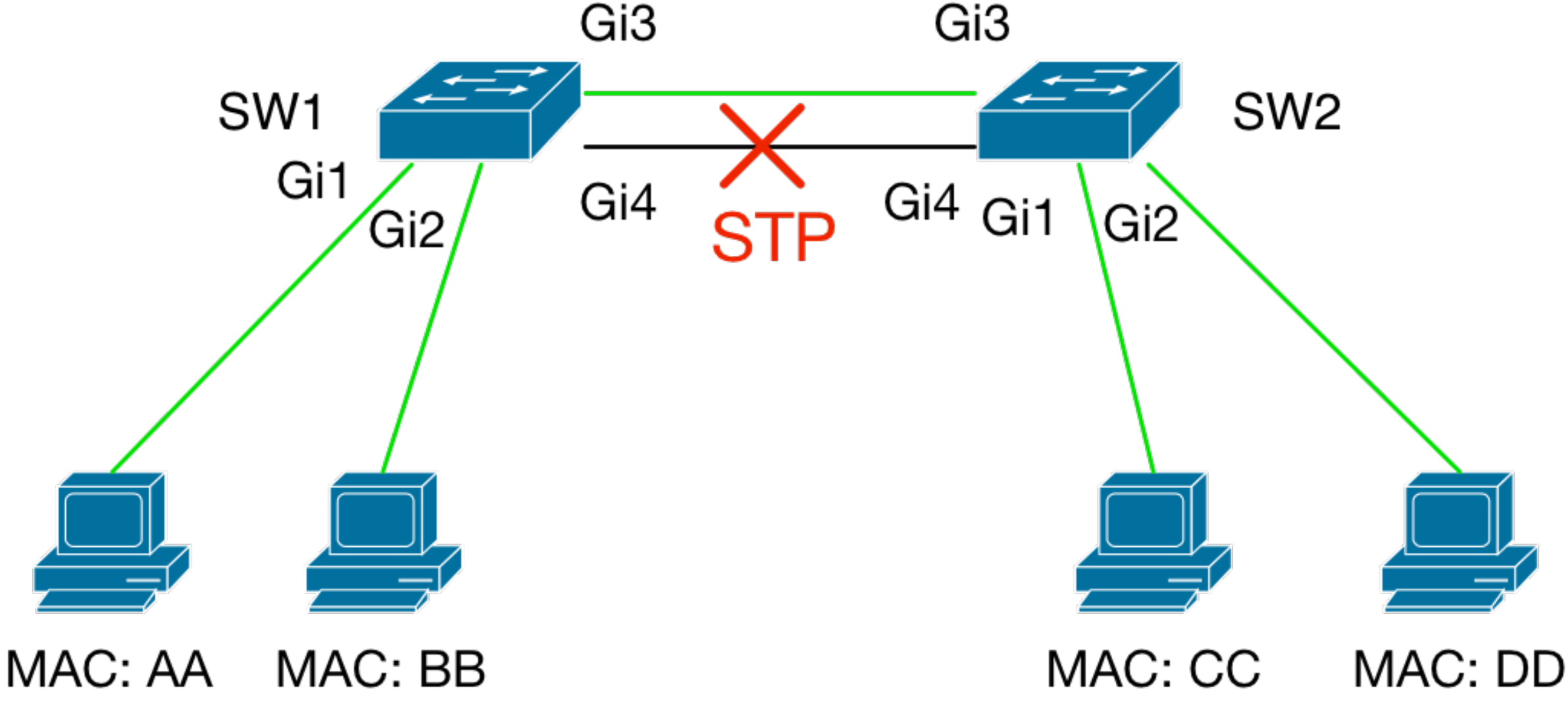


SW1 L2 FIB

PORT	MAC
Gi1	AA
Gi2	BB
Gi3	CC DD

SW2 L2 FIB

PORT	MAC
Gi1	CC
Gi2	DD
Gi3	AA BB

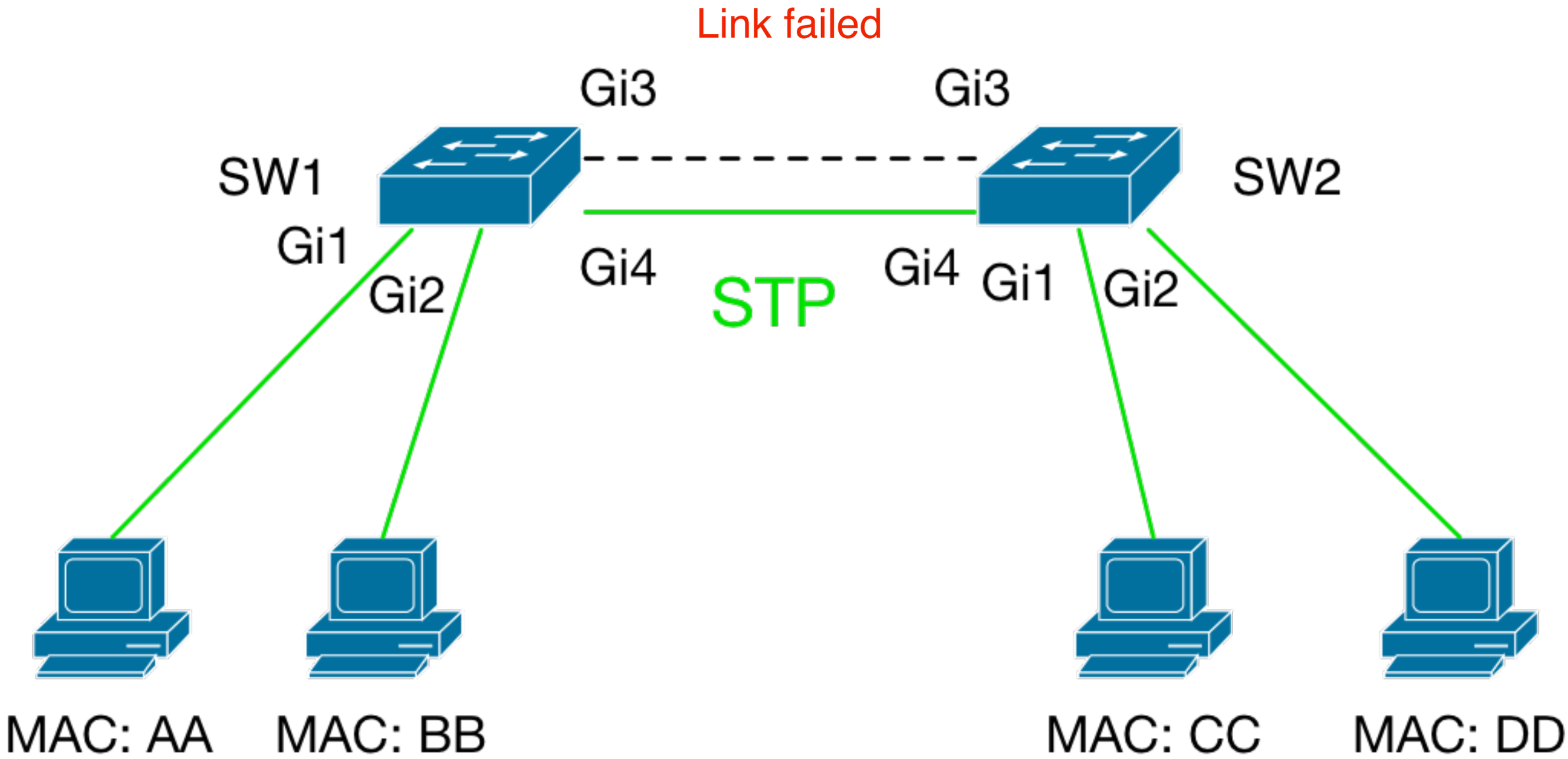


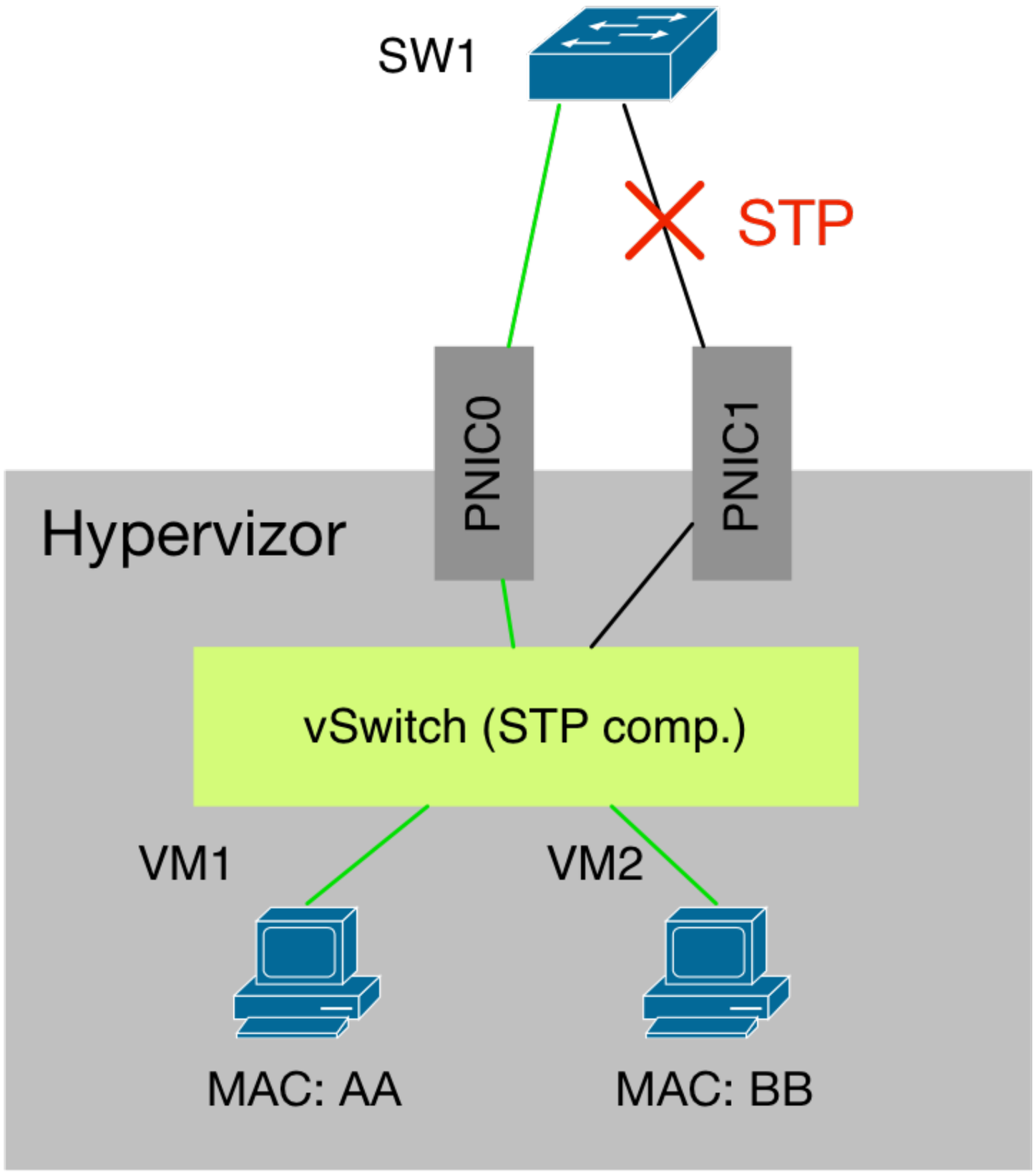
SW1 L2 FIB

PORT	MAC
Gi1	AA
Gi2	BB
Gi4	CC DD

SW2 L2 FIB

PORT	MAC
Gi1	CC
Gi2	DD
Gi4	AA BB




SW1
STP
PNIC0
PNIC1
Hypervizor
vSwitch (STP comp.)
VM1
VM2
MAC: AA
MAC: BB

Redundance na L2

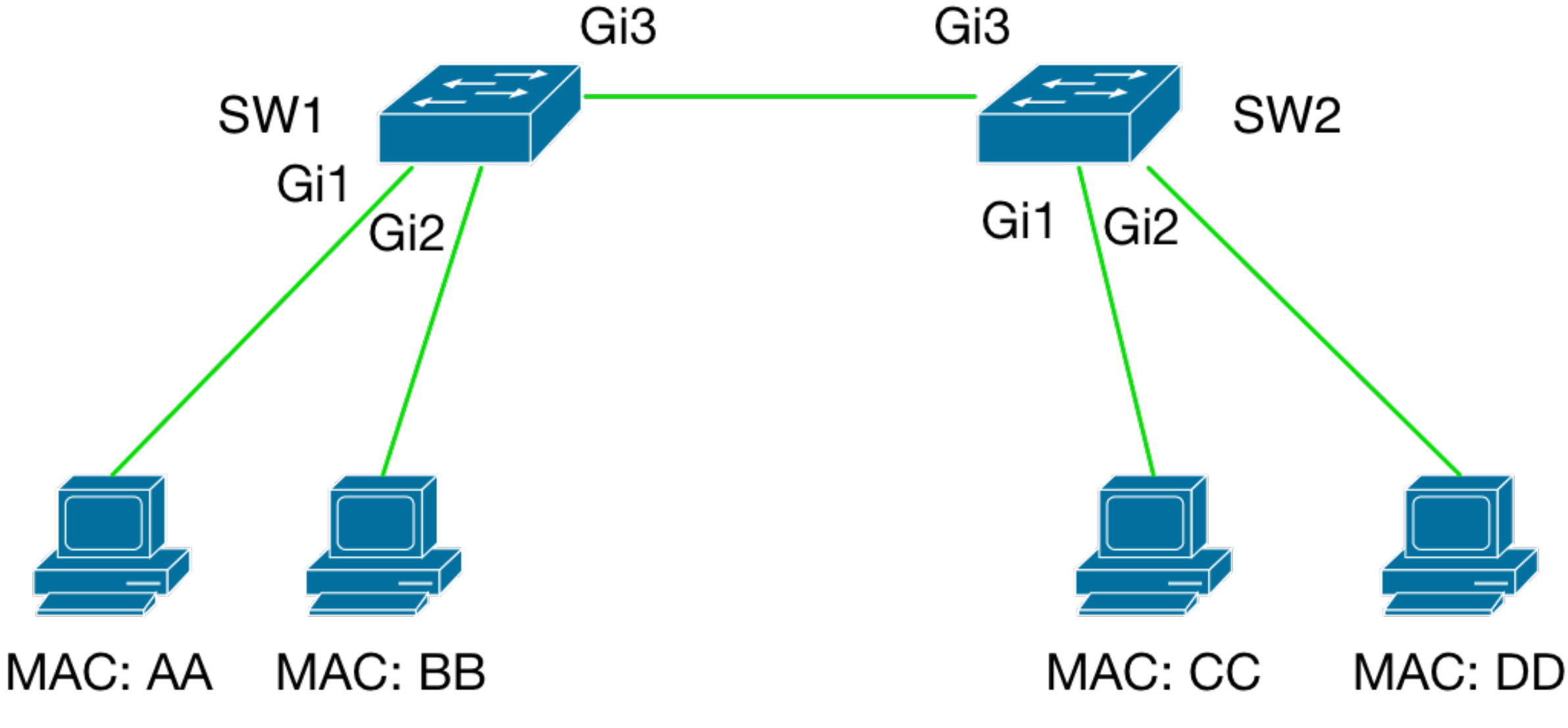
Active/Active - Port-channel (LACP)

SW1 L2 FIB

PORT	MAC
Gi1	AA
Gi2	BB
Gi3	CC DD

SW2 L2 FIB

PORT	MAC
Gi1	CC
Gi2	DD
Gi3	AA BB



SW1#

```
interface Gi 3
  switchport access vlan 1
  channel-group 1 mode active

interface Gi 4
  switchport access vlan 1
  channel-group 1 mode active

interface Po 1
  switchport access vlan 1
```

SW2#

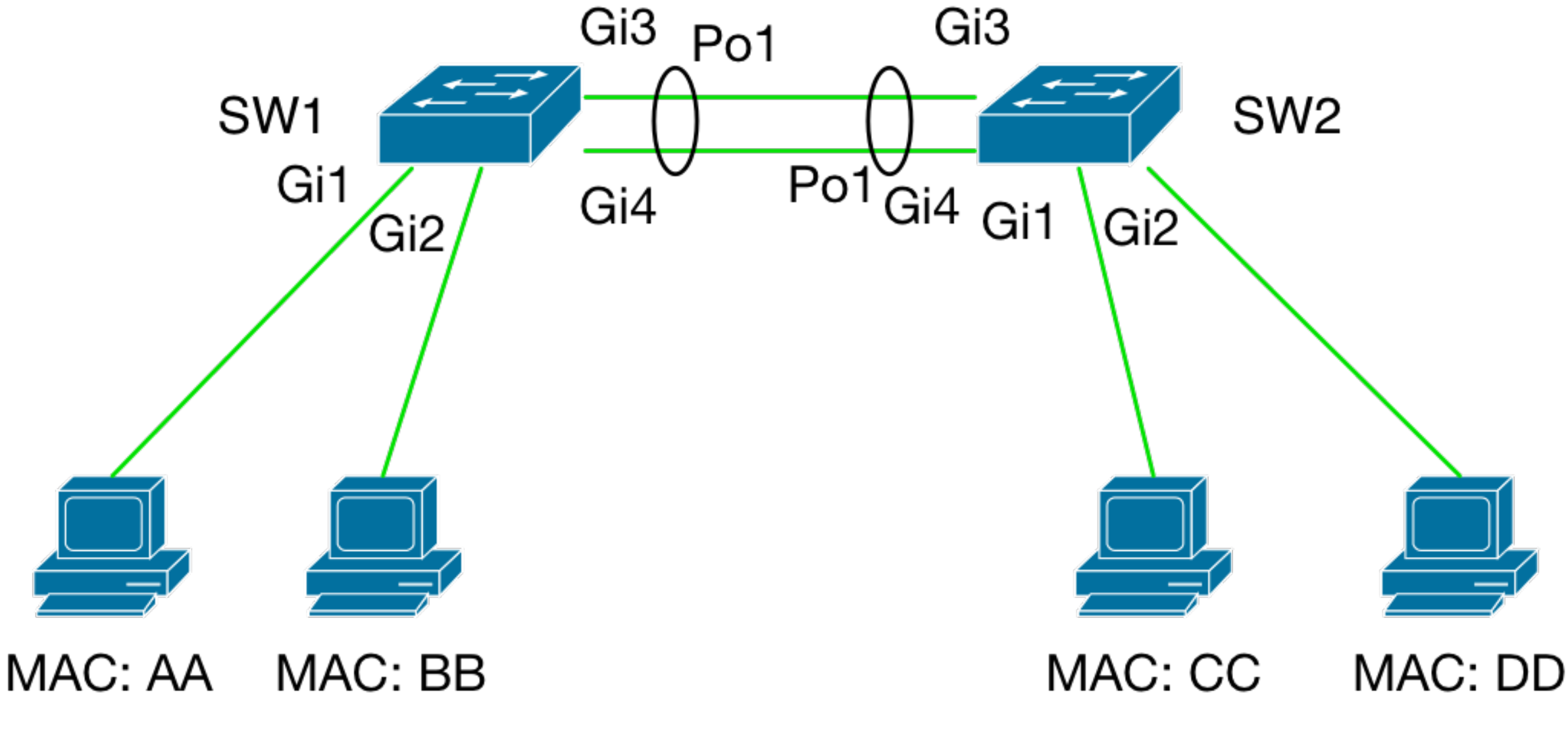
```
interface Gi 3
  switchport access vlan 1
  channel-group 1 mode active

interface Gi 4
  switchport access vlan 1
  channel-group 1 mode active

interface Po 1
  switchport access vlan 1
```

SW1 L2 FIB

PORT	MAC
Gi1	AA
Gi2	BB
Po1	CC DD

SW2 L2 FIB

PORT	MAC
Gi1	CC
Gi2	DD
Po1	AA BB



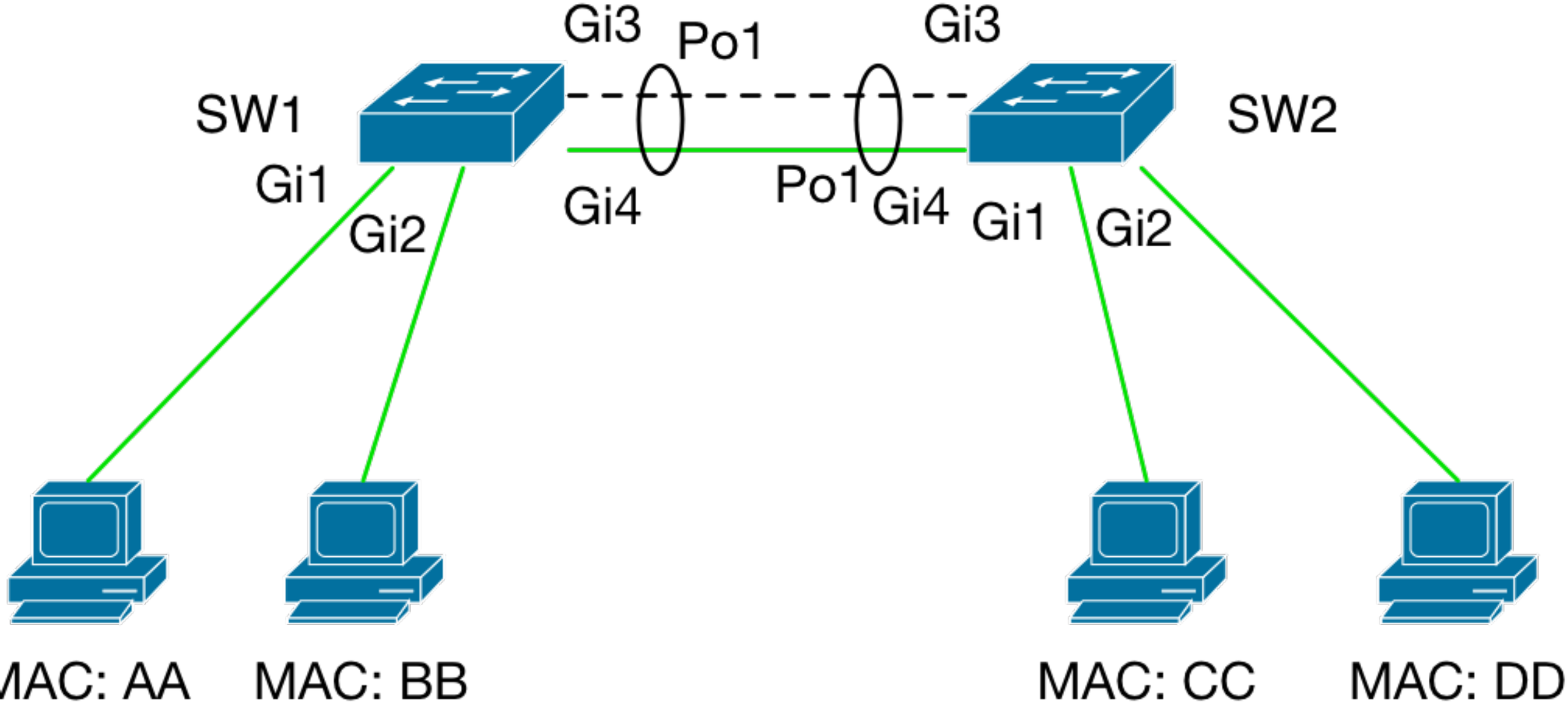
SW1 L2 FIB

PORT	MAC
Gi1	AA
Gi2	BB
Po1	CC DD

SW2 L2 FIB

PORT	MAC
Gi1	CC
Gi2	DD
Po1	AA BB

Member port failed -> OK



Redundance na L2

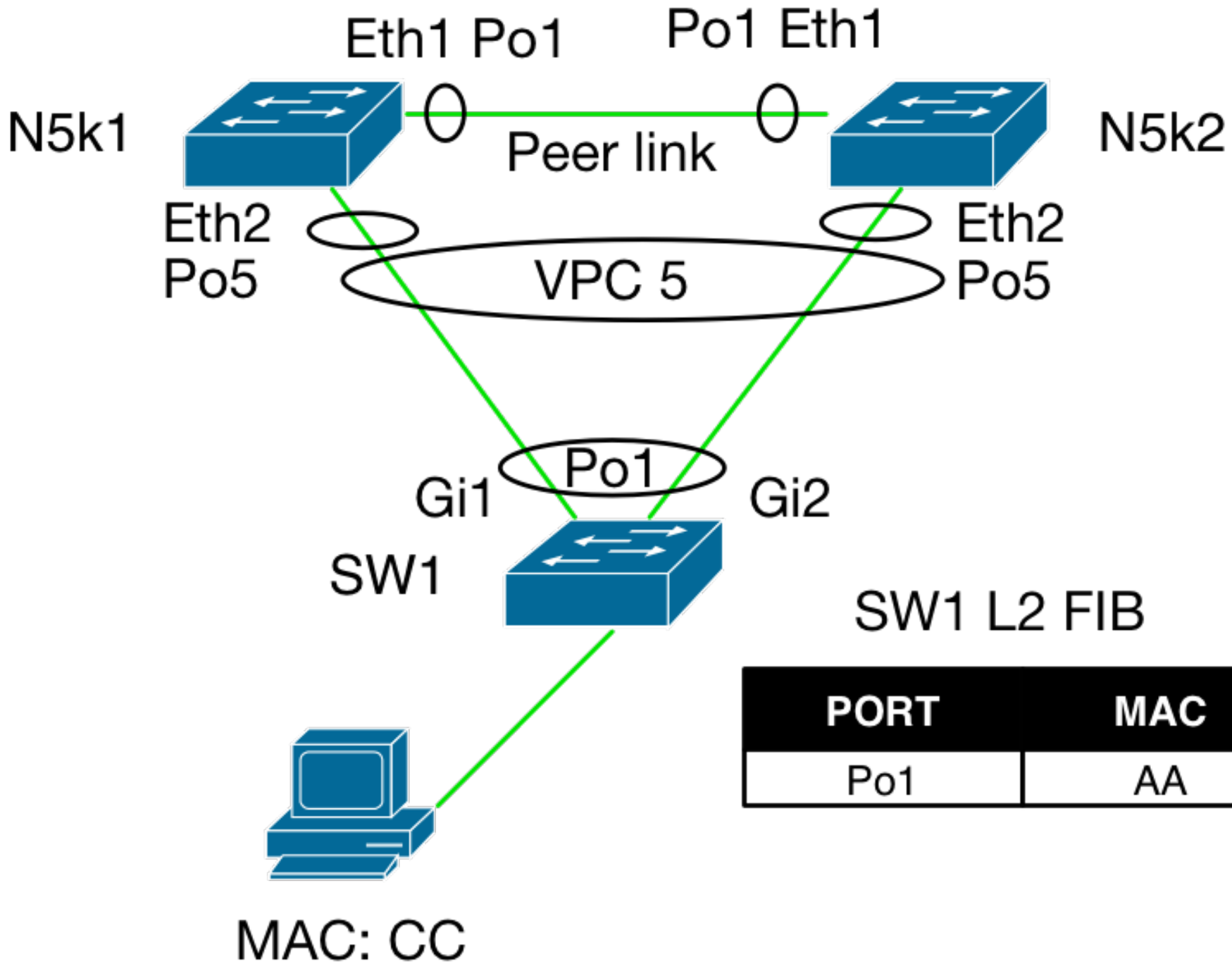
Active/Active - Virtual Port-channel (Multichassis port channel)

N5k1 FIB

PORT	MAC
Po5	CC

N5k2 FIB

PORT	MAC
Po5	CC



SW1 L2 FIB

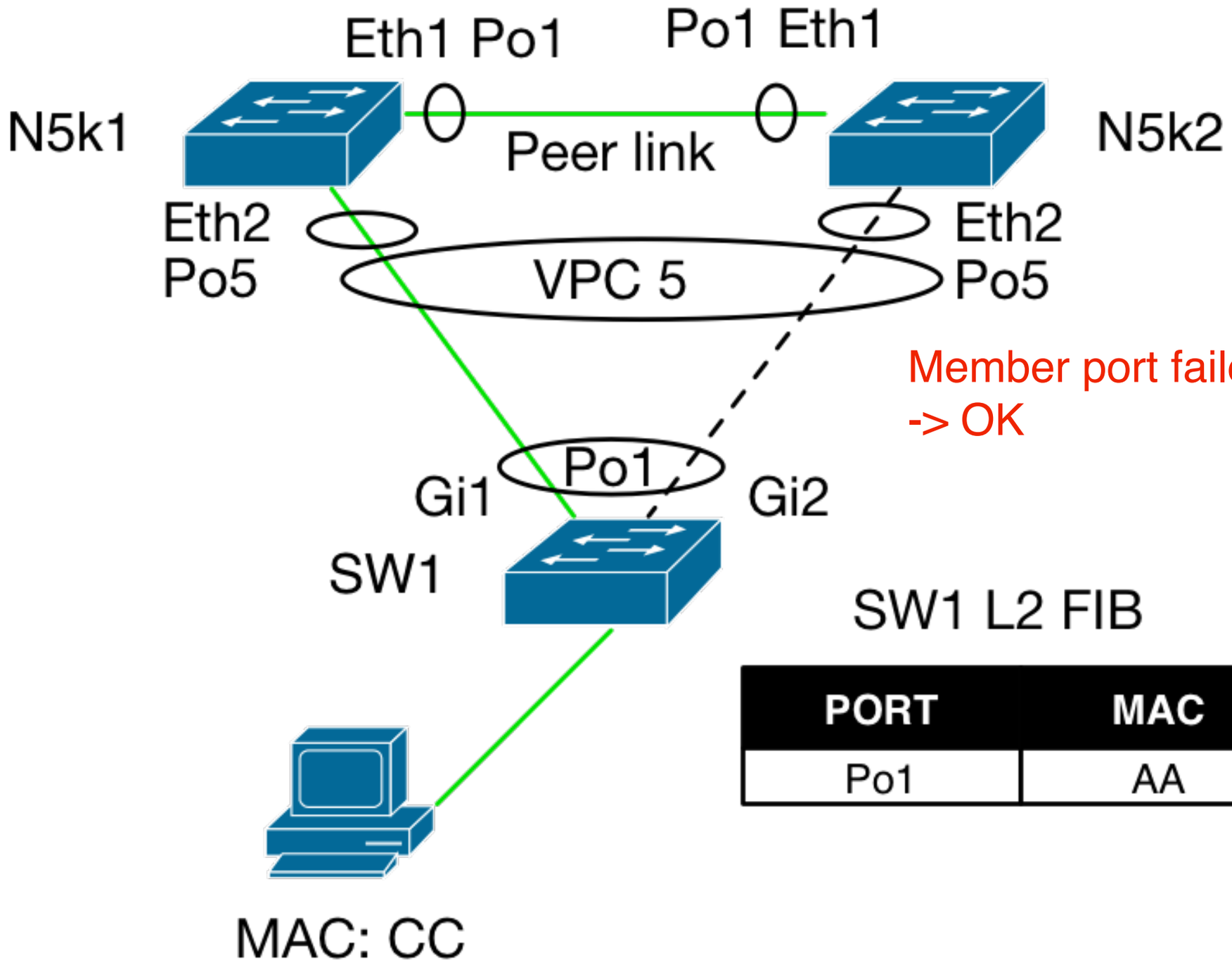
PORT	MAC
Po1	AA

N5k1 FIB

PORT	MAC
Po5	CC

N5k2 FIB

PORT	MAC
Po1	CC



N5k1 FIB

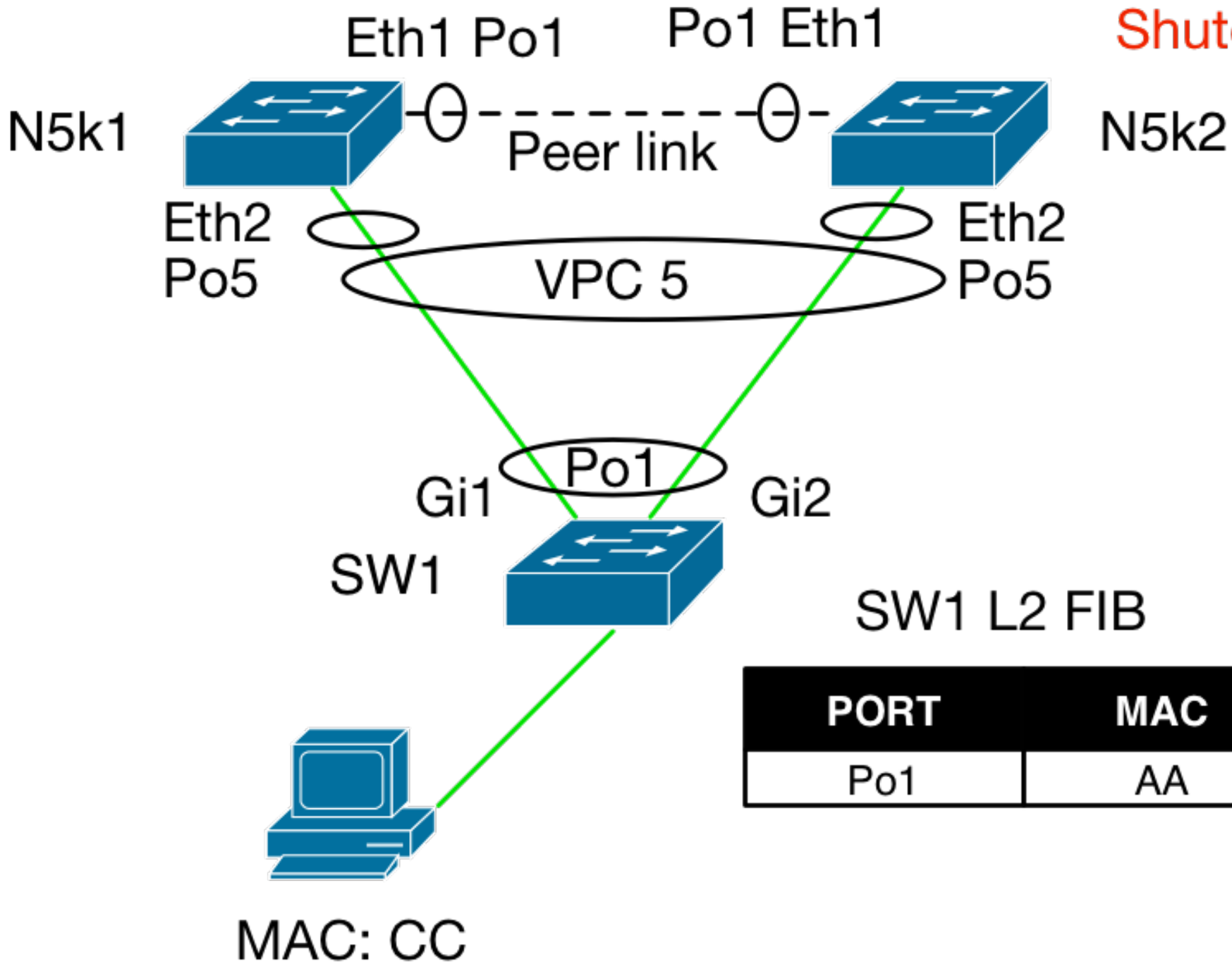
PORT	MAC
Po5	CC

N5k2 FIB

PORT	MAC
Po5	CC

Peer link failed ->

VPC Secondary Shutdown!



SW1 L2 FIB

PORT	MAC
Po1	AA

N5k1 FIB

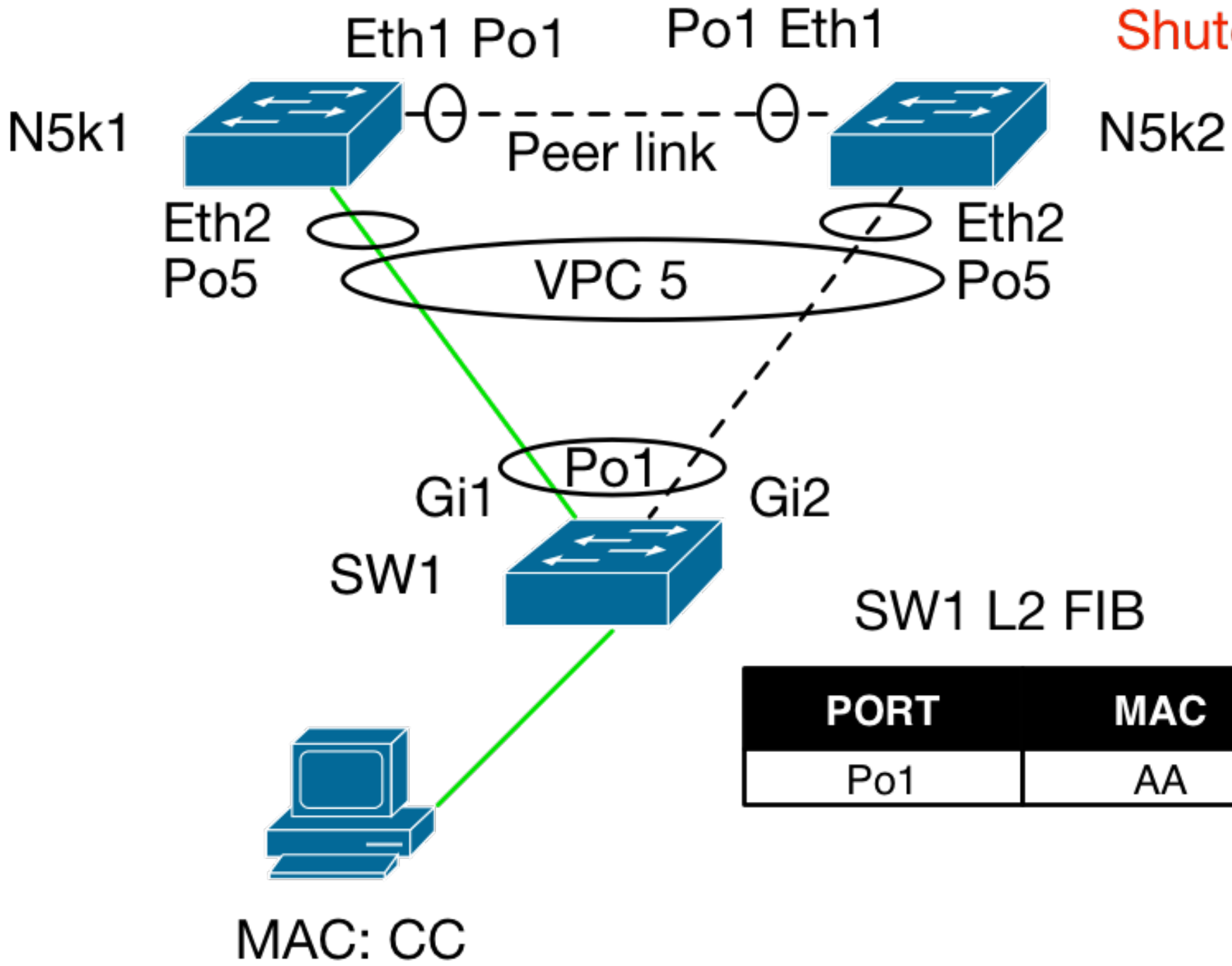
PORT	MAC
Po5	CC

N5k2 FIB

PORT	MAC
N/A	N/A

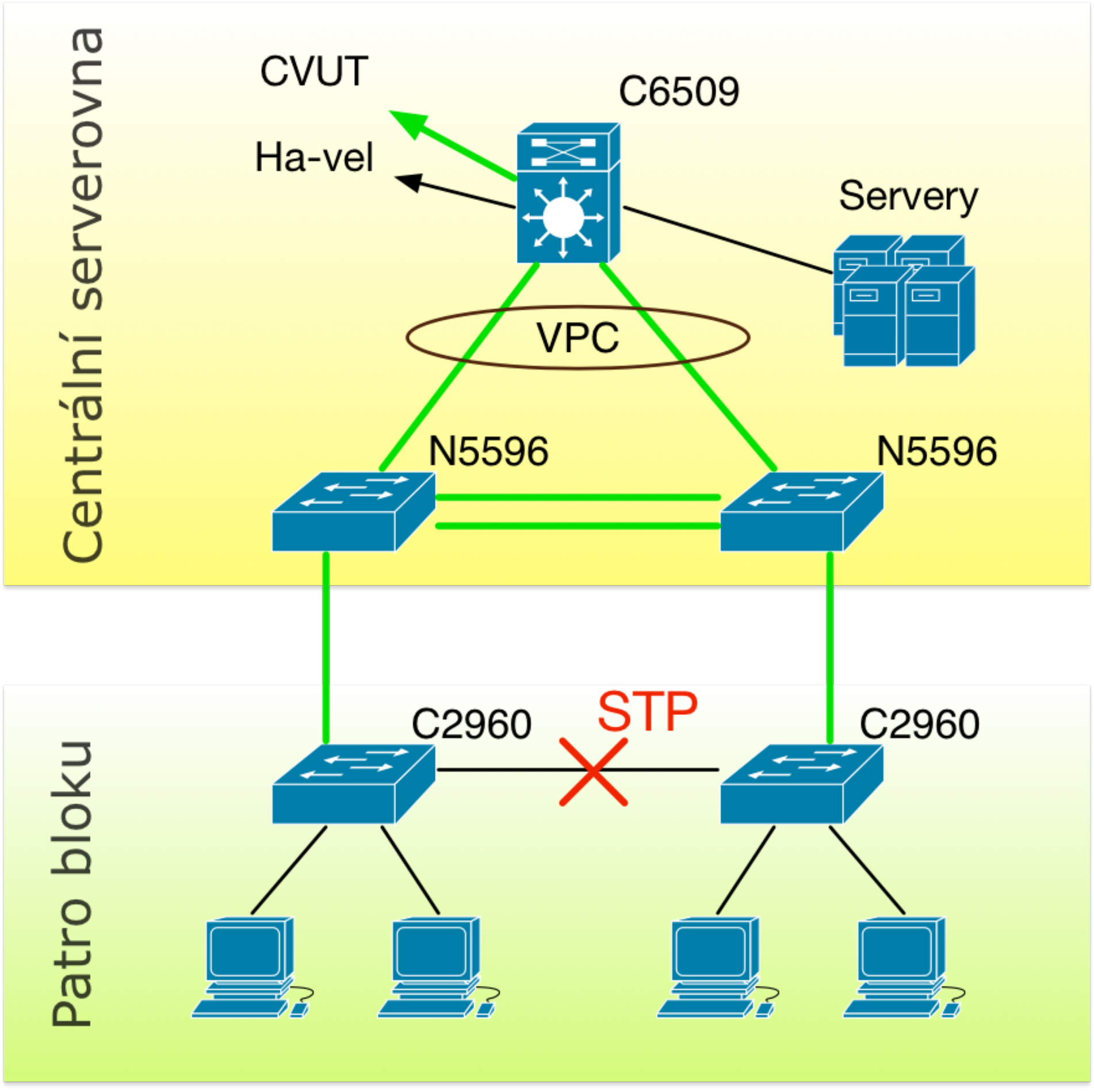
Peer link failed ->

VPC Secondary Shutdown!

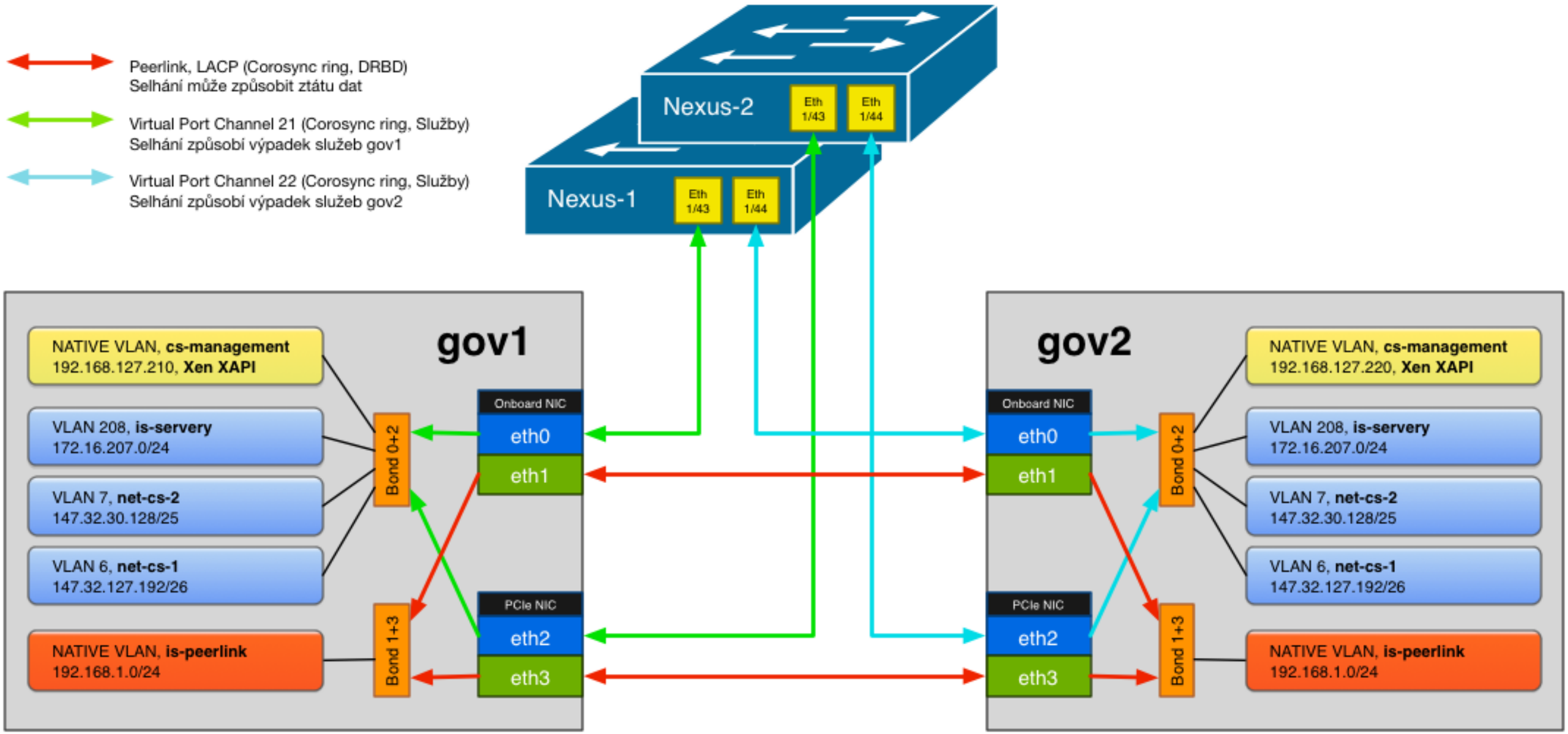


SW1 L2 FIB

PORT	MAC
Po1	AA


Centrální serverovna
CVUT
Ha-vel
C6509
Servery
VPC
N5596
N5596
Patro bloku
C2960
STP
C2960

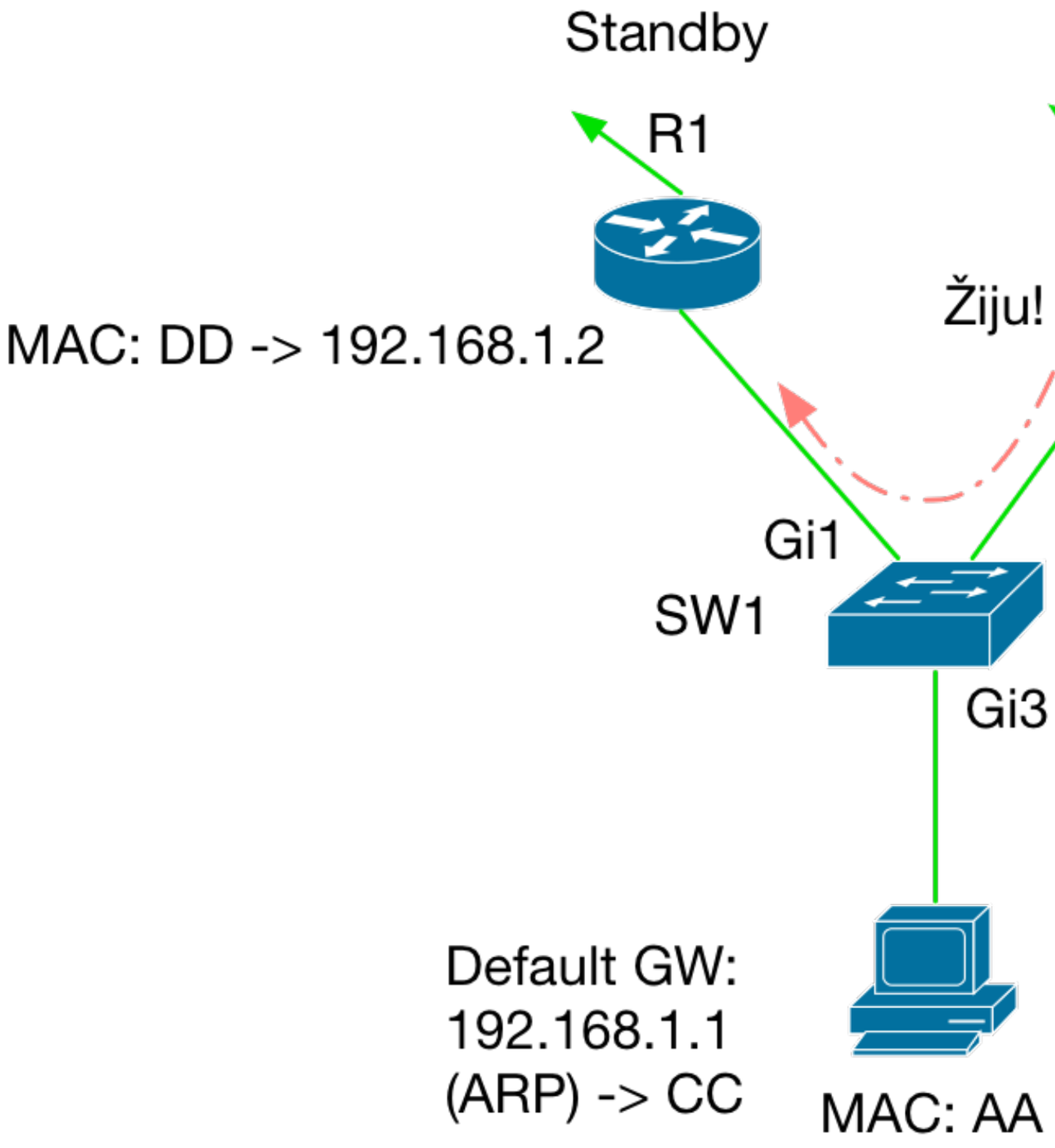
Zapojení serverů Informačního systému klubu Silicon Hill
Live configuration, 14.3.2014



Redundance pro First Hop L2

Virtual Router Redundancy Protocol (VRRP)

Hot Standby Router Protocol (HSRP)



SW1 L2 FIB

PORT	MAC
Gi1	DD
Gi2	EE CC
Gi3	AA

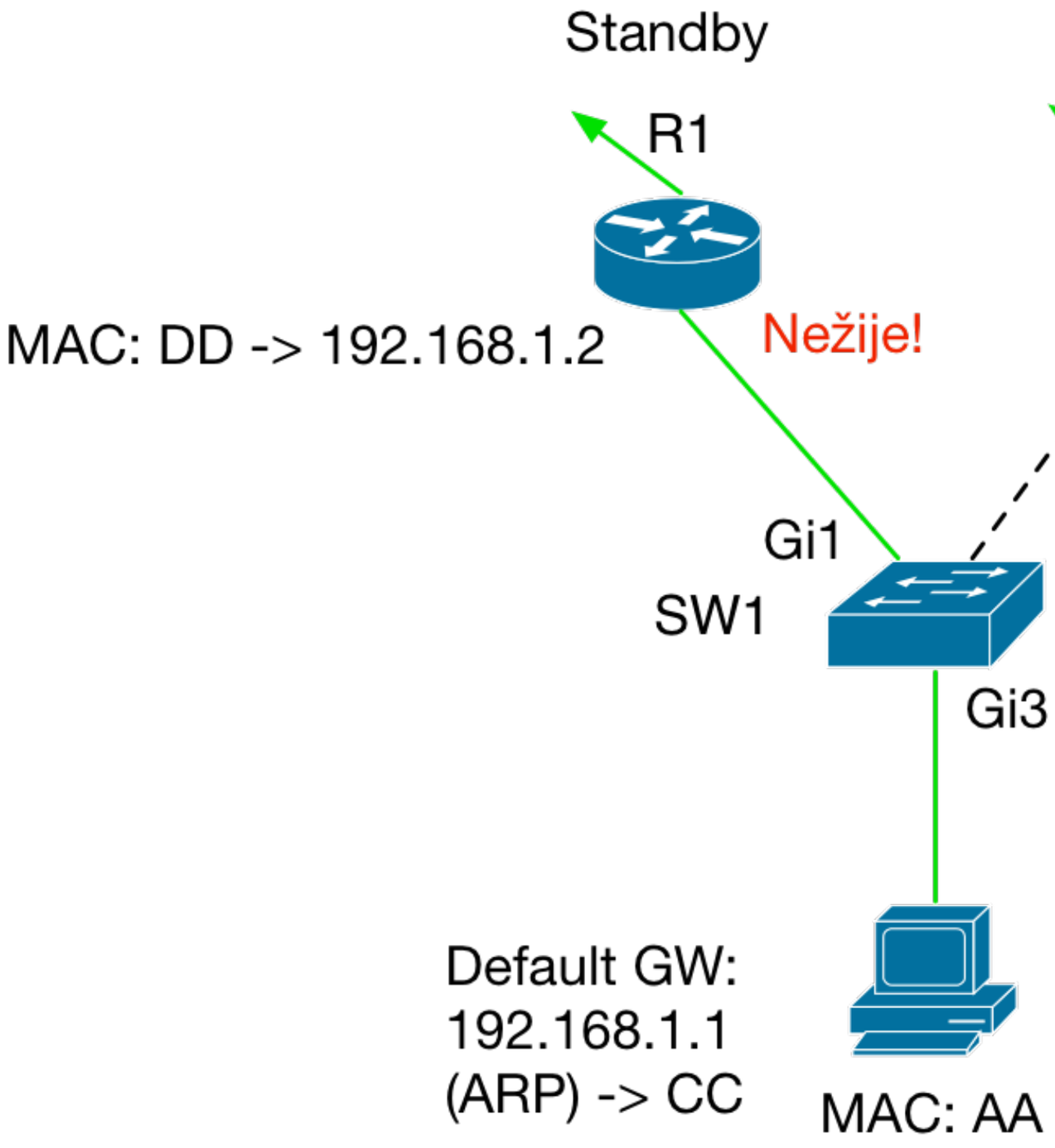
R1#

```
interface Gi 1
  ip address 192.168.1.2/24
  standby 2 priority 100
  standby 2 preempt
  standby 2 ip 192.168.1.1
```

R2#

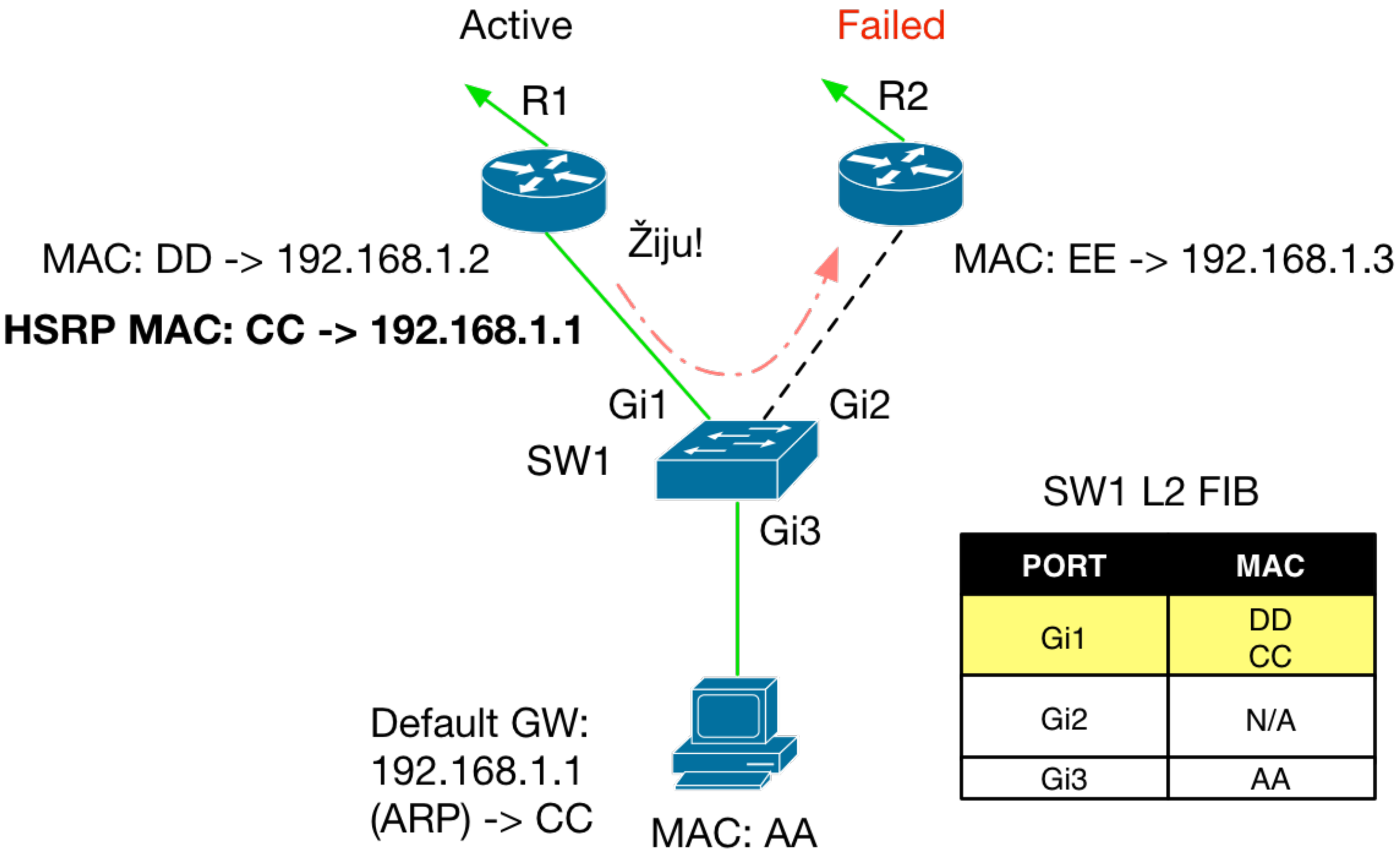
```
interface Gi 1
  ip address 192.168.1.3/24
  standby 2 priority 110
  standby 2 preempt
  standby 2 ip 192.168.1.1
```

HSRP/VRRP MAC = f(GROUP-ID) -> Kolize na stejné VLAN !!
Vždy zapnout ověření (standby GROUP-ID auth …) !!



SW1 L2 FIB

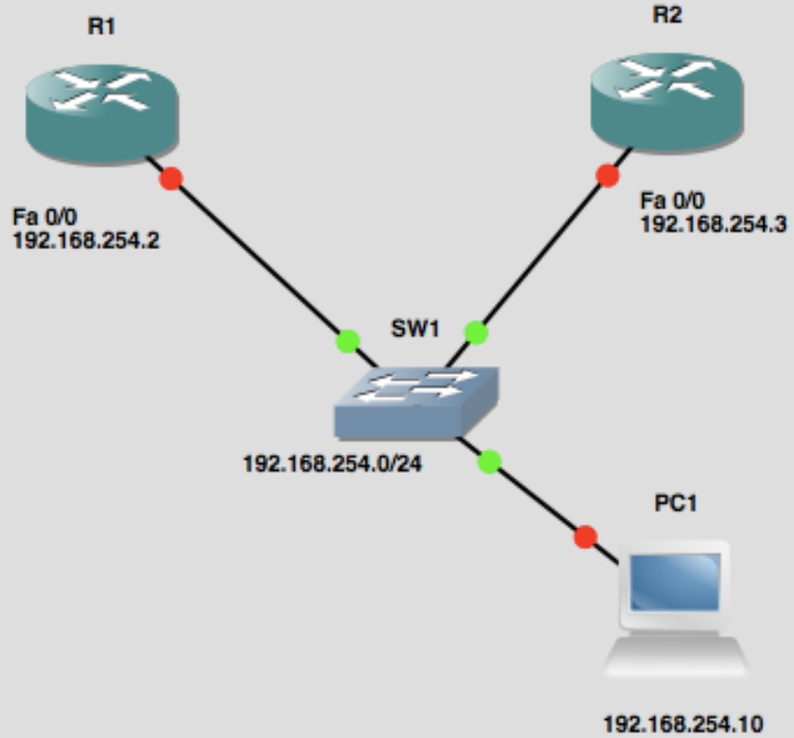
PORT	MAC
Gi1	DD
Gi2	N/A
Gi3	AA



SW1 L2 FIB

PORT	MAC
Gi1	DD CC
Gi2	N/A
Gi3	AA

FHRP.gns3



Topology Summary

- SW1
- R1
- R2
- PC1

Jungle Newsfeed

GNS3 Jungle

WANT TO TRY IOU IN GNS3?

Grab your copy of IOU and follow these instructions to have it up and running in no time.

Integrate IOU

Console

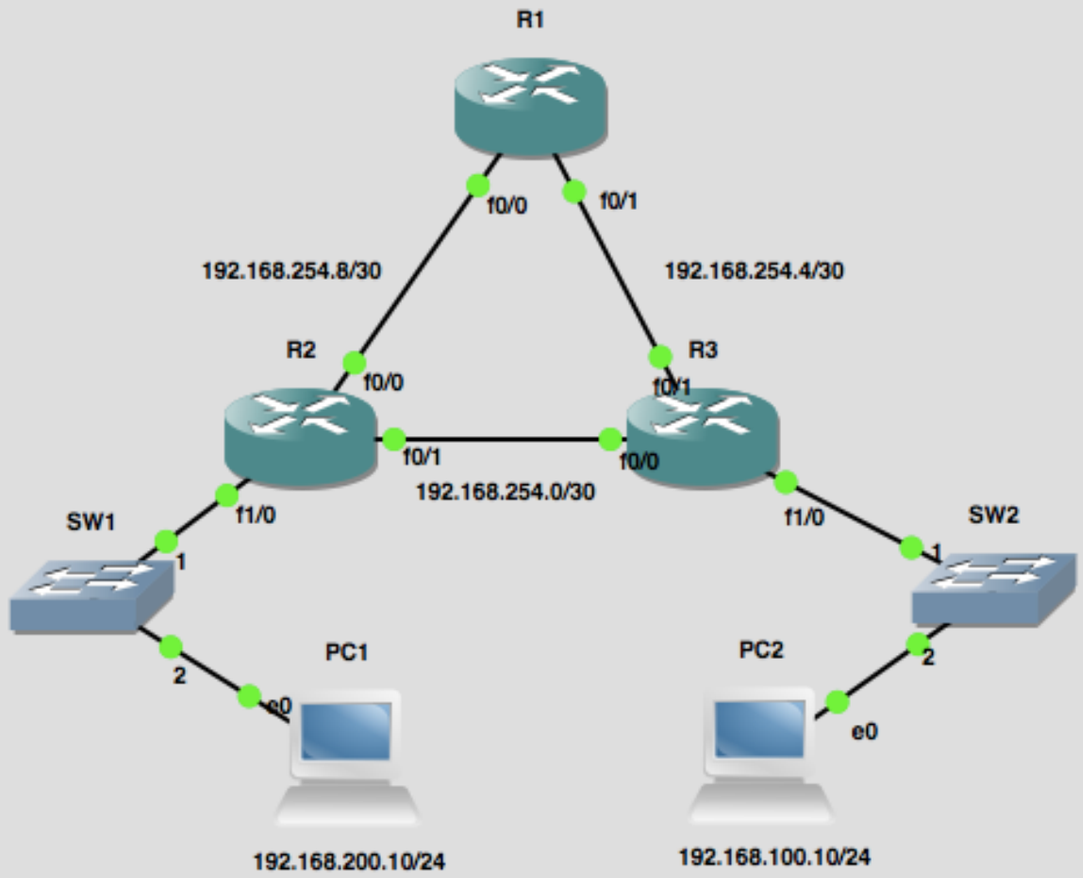
```
Documented commands (type help <topic>):
========================================
console  debug  help  reload  show  start  stop  suspend  version

=>
```

X: 45.0 Y: -97.0 Z: 2.0

Redundance na L3

Dynamický routovací protokol (OSPF)

OSPF.gns3



Topology Summary

- R1
- R2
- R3
- SW1
- SW2
- PC1
- PC2

Jungle Newsfeed

GNS3 Jungle

SOLARWINDS NCM IN GNS3!

Complete guide to setting up Solar-Winds NCM directly in GNS3. **Comes with a Free Download of NCM**

DOWNLOAD NCM

Console

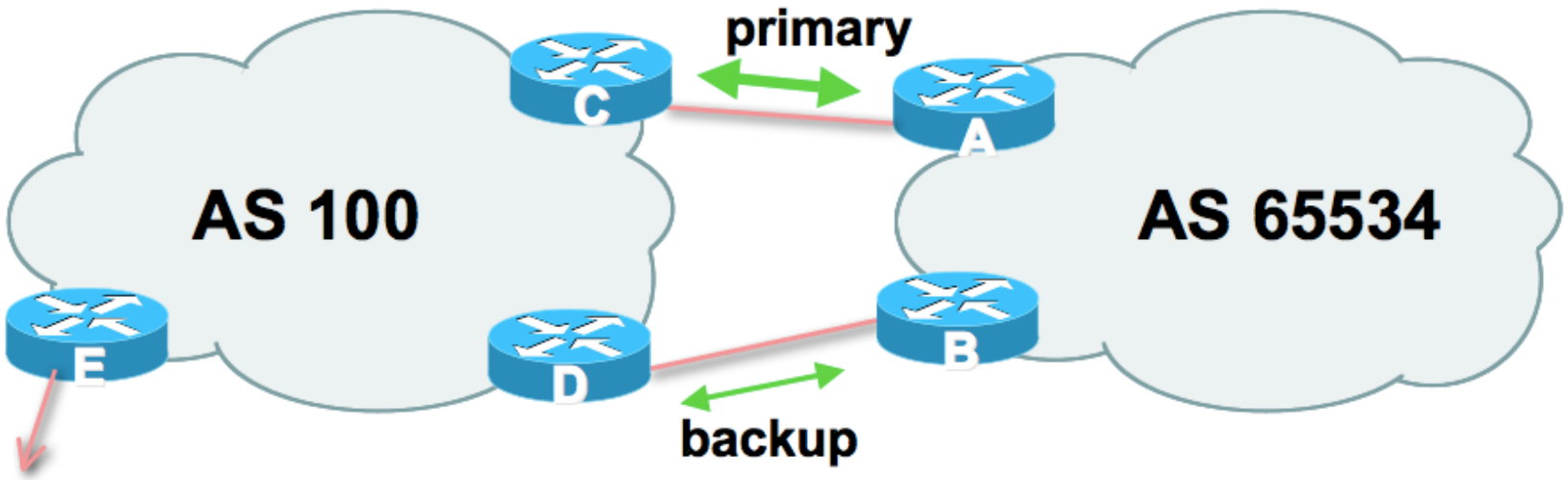
```
Documented commands (type help <topic>):
========================================
console  debug  help  reload  show  start  stop  suspend  version

=>
```

X: -327.5 Y: 107.0 Z: 1.0

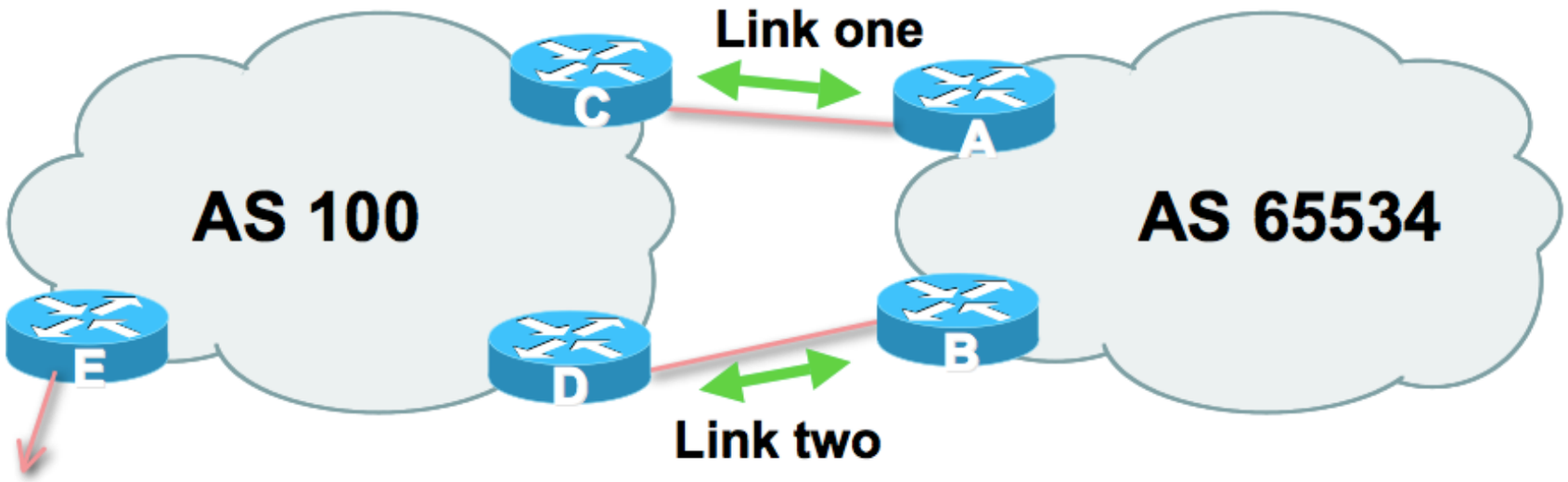
Redundance na internetu

Border Gateway Protocol (Základní informace)

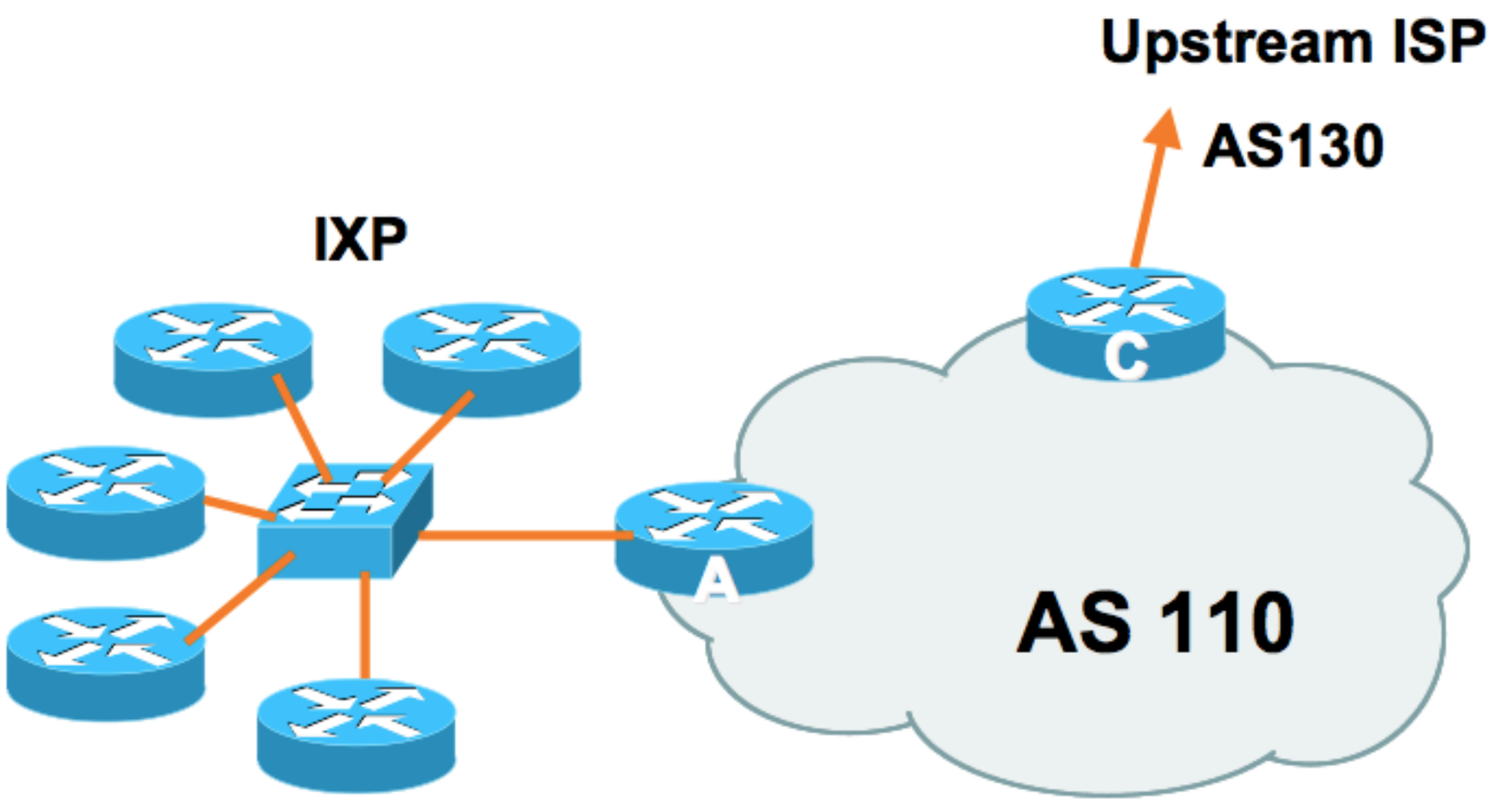
Two links to the same ISP (one as backup only)



Loadsharing to the same ISP

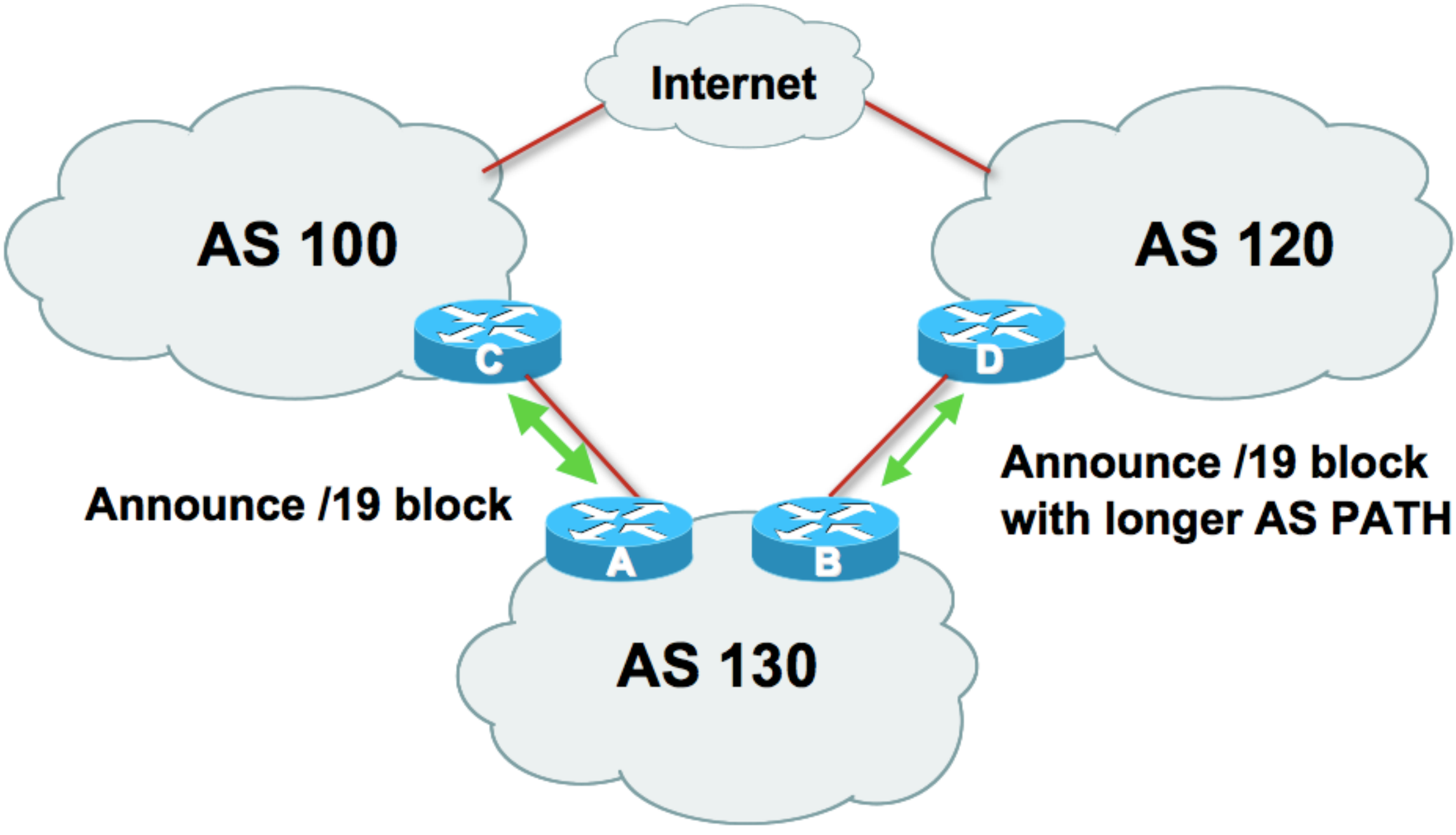


One Upstream, Local Exchange Point

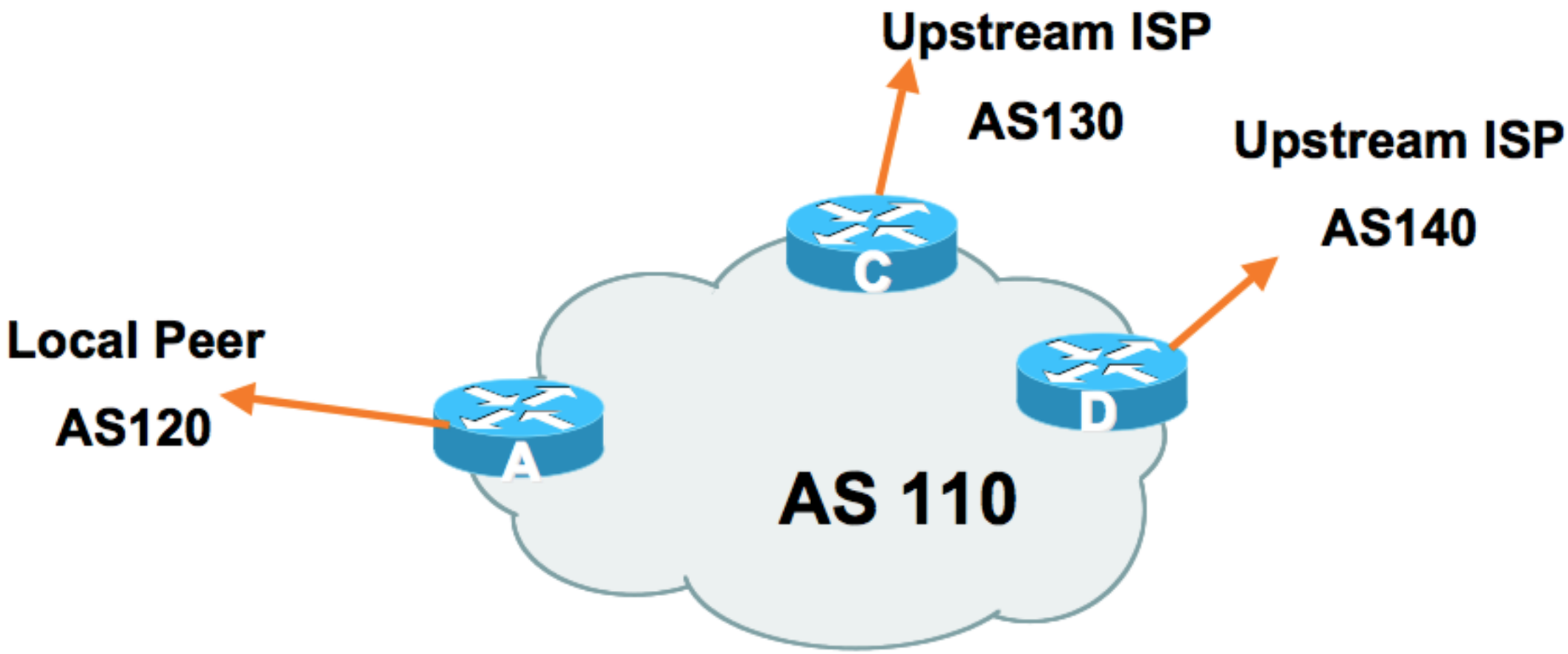


https://www.nanog.org/meetings/nanog41/presentations/BGPMultihoming.pdf

Two links to different ISPs (one as backup only)

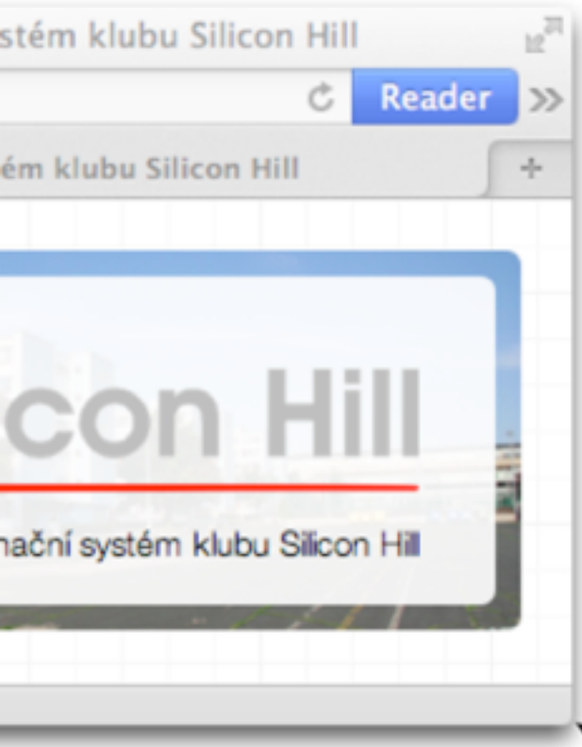


Two Upstreams, One Local Peer



Vysoká dostupnost aplikací

Informační systém klubu Silicon Hill



PostgreSQL

Materialized views

Read Only SQL

POWERDNS

DNS Server/Recursor/Cache

ISC DHCP + Patch

pgboucner

POSTFIX

SMTP Server

R/W SQL

HTTPS

SSH

Síťové prvky

Model

View/Controller

Workery
(Delayed Job)

IS Applikace

R/W LDAP

OpenDJ

LDAP Server

RO LDAP

freeRADIUS

RADIUS Server

GOV 1

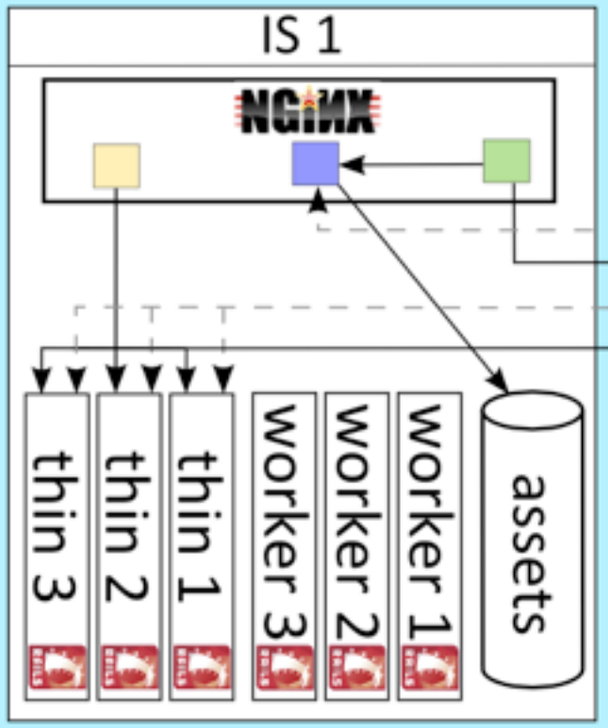
GOV 2

NGINX load balancer
is.sh.cvut.cz
+ ssl offload

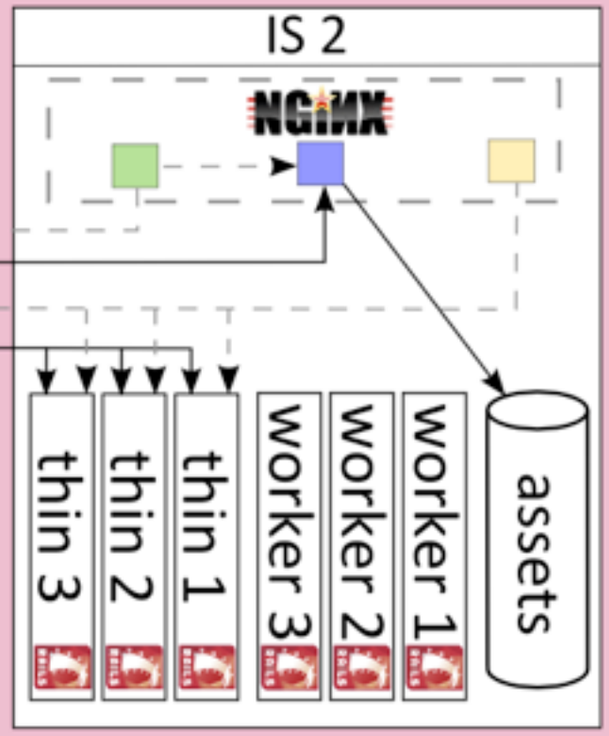
NGINX load balancer
static.is.sh.cvut.cz

NGINX pro assets

IS 1



IS 2



LDAP 1

OpenDJ
Congregate - Locate

RADIUS

LDAP 2

OpenDJ
Congregate - Locate

RADIUS

SVC 1

PowerDNS

ISC DHCPD
+ SH DBI Patch

SMTP (Postfix)

SVC 2

PowerDNS

ISC DHCPD
+ SH DBI Patch

SMTP (Postfix)

PGSQL 1

PostgreSQL

DRBD

PGSQL 2

PostgreSQL

DRBD

VYSOKÁ DOSTUPNOST

- Řešíme SPOF (Single point of failure)
- Cílem je “aby se to samo nepokazilo” => **Prevence**
  => monitoring (trendy, ...) -> *Munin*
  => včasné varování (místo na disku, ...) -> *Nagios*
- Systém by si, ale “měl poradit sám” =>
  Redundance
  => fail over + [load balancing]
  => replikace dat
  => fencing

PACEMAKER

- *Pacemaker* = Cluster manager
  (http://clusterlabs.org , http://www.corosync.org)

- *Node* = Server, několik serverů je ve společném clusteru
  (rozhoduje zvolený master; komunikace Corosync)

- *CIB* = Cluster Information Base
  (konvergovaný stav - všechny nody ví všechno = konfigurace a stav)

- *Resource* = “služba”, například IP adresa, nebo proces

- *Resource Agent* (RA) = “wrapper” kolem konkrétní služby
  (parametrizovaný SH skript - start, stop, ...)

PACEMAKER - KONFIGURACE

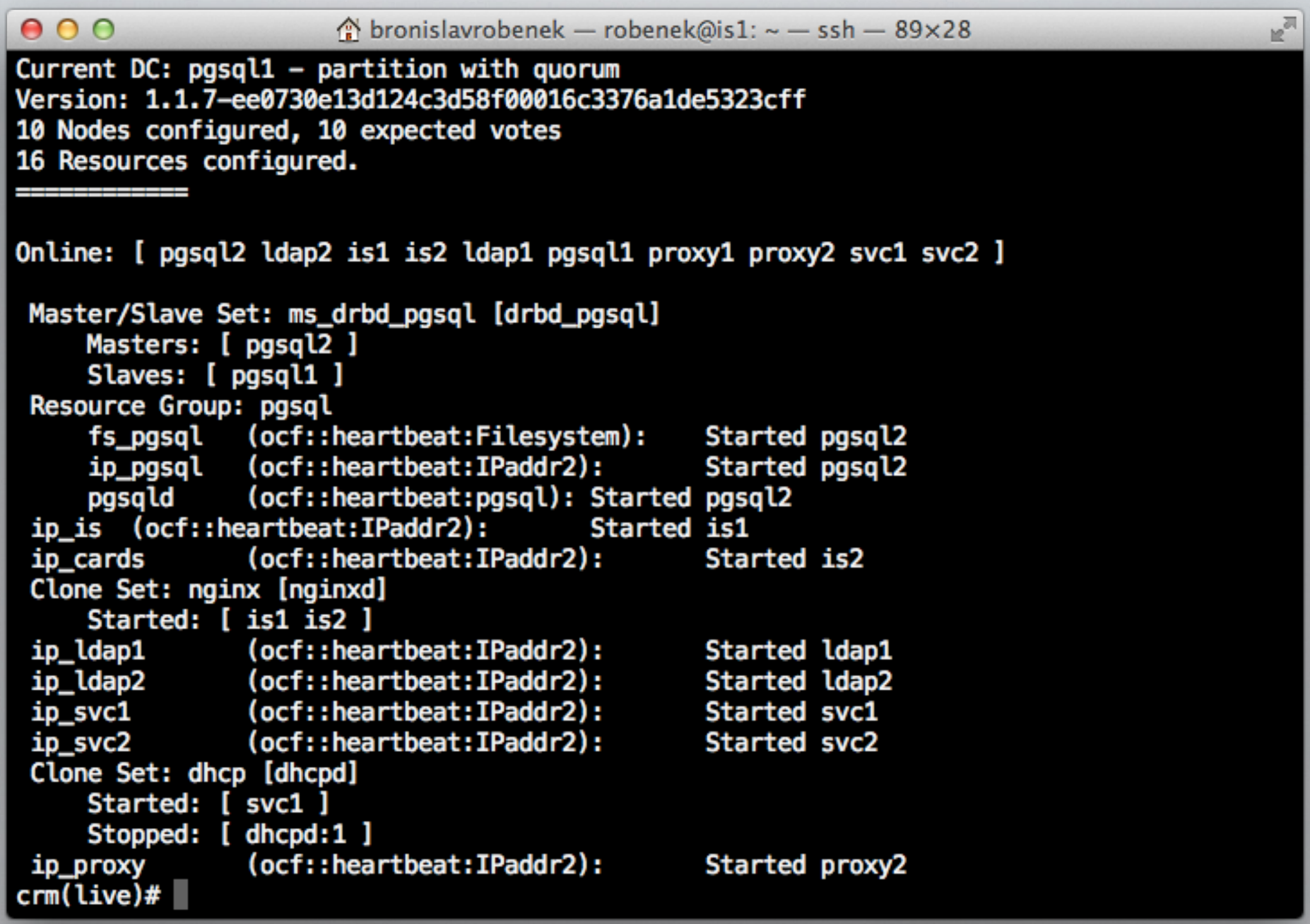
- Definice služeb (RA, parametry)
  `primitive ip1 ocf:heartbeat:IPaddr2 params ip="1.2.3.4" cidr_netmask="24"`

- Kde může která služba běžet (primitve, score, node)
  `location loc_ip1_node1 ip1 100: node1`

- Kolikrát má služba běžet v clusteru (primtive, počet)
  `clone nginx nginxd meta clone-max="2"`

- Logické uspořádání služeb
  `group pgsql fs_pgsql ip_pgsql pgsqld`
  `colocation ip1_on_nginx inf: ip_1 nginx`

- Pořadí spouštění služeb
  `order nginx_after_ip inf: ip1 nginx`

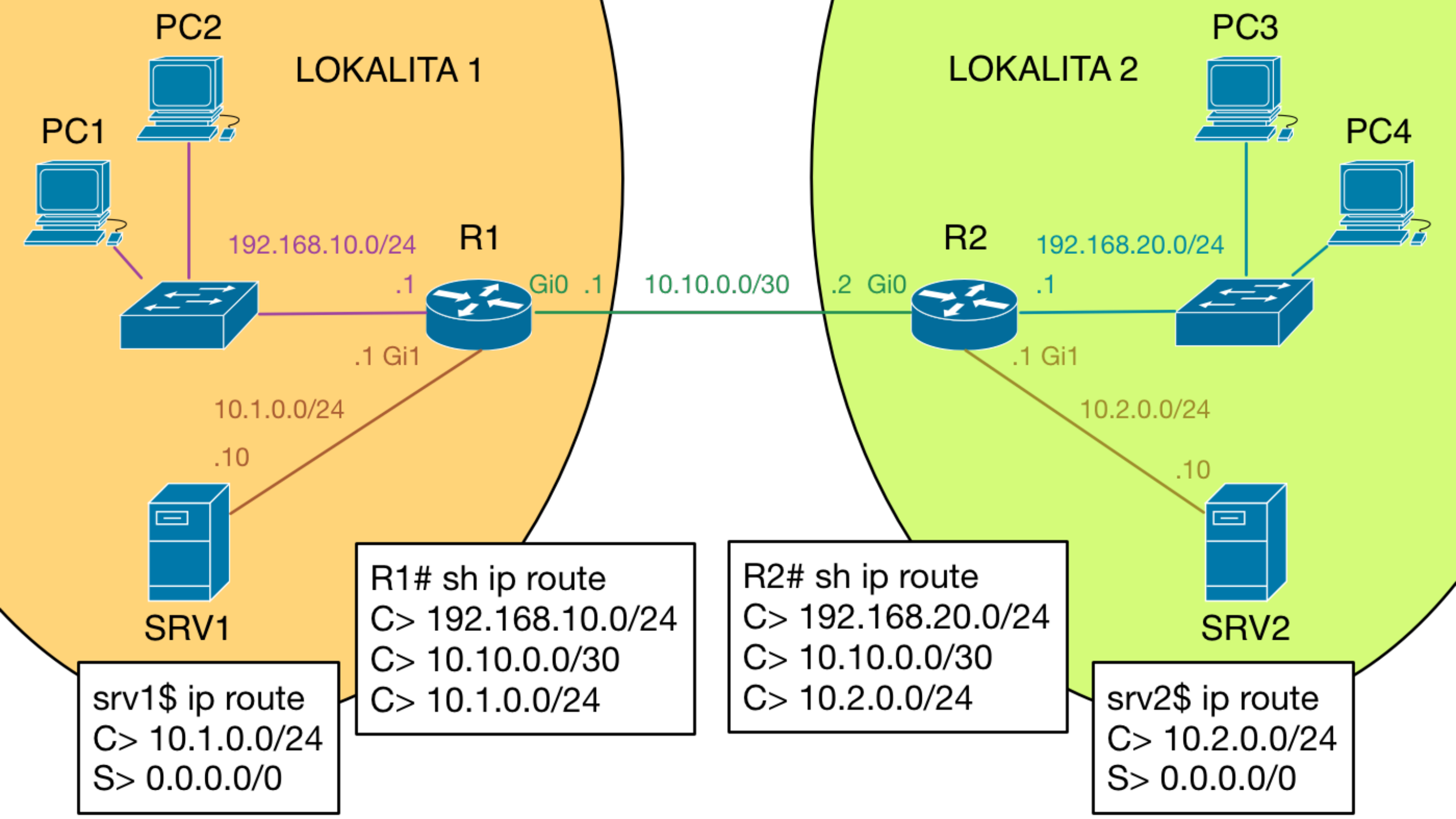

bronislavrobenek — robenek@is1: ~ — ssh — 89×28
Current DC: pgsql1 - partition with quorum
Version: 1.1.7-ee0730e13d124c3d58f00016c3376a1de5323cff
10 Nodes configured, 10 expected votes
16 Resources configured.
Online: [pgsql2 ldap2 is1 is2 ldap1 pgsql1 proxy1 proxy2 svc1 svc2]
Master/Slave Set: ms_drbd_pgsql [drbd_pgsql]
Masters: [pgsql2]
Slaves: [pgsql1]
Resource Group: pgsql
fs_pgsql (ocf::heartbeat:Filesystem): Started pgsql2
ip_pgsql (ocf::heartbeat:IPaddr2): Started pgsql2
pgsqld (ocf::heartbeat:pgsql): Started pgsql2
ip_is (ocf::heartbeat:IPaddr2): Started is1
ip_cards (ocf::heartbeat:IPaddr2): Started is2
Clone Set: nginx [nginxd]
Started: [is1 is2]
ip_ldap1 (ocf::heartbeat:IPaddr2): Started ldap1
ip_ldap2 (ocf::heartbeat:IPaddr2): Started ldap2
ip_svc1 (ocf::heartbeat:IPaddr2): Started svc1
ip_svc2 (ocf::heartbeat:IPaddr2): Started svc2
Clone Set: dhcp [dhcpd]
Started: [svc1]
Stopped: [dhcpd:1]
ip_proxy (ocf::heartbeat:IPaddr2): Started proxy2
crm(live)#

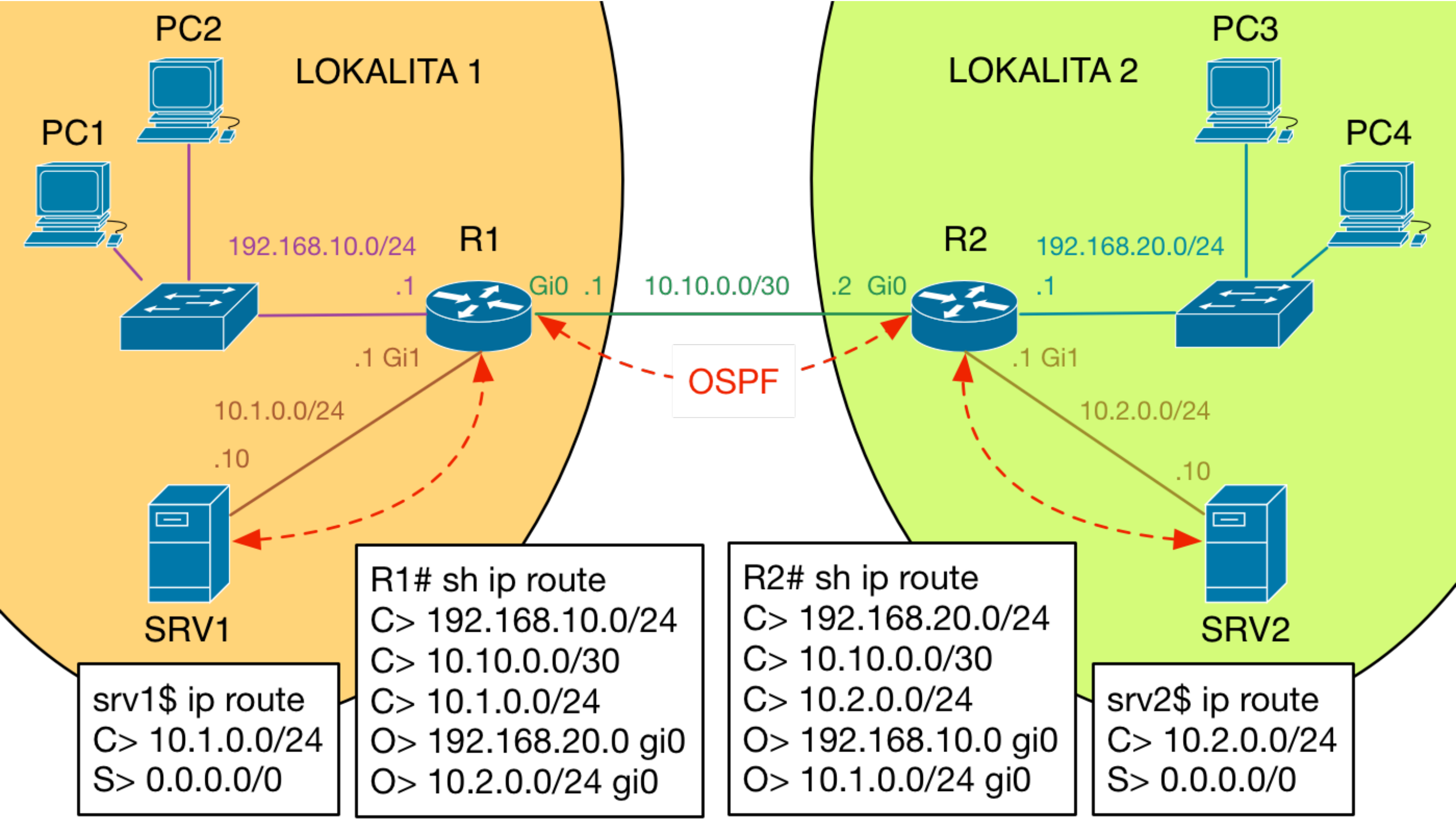
Vysoká dostupnost aplikací

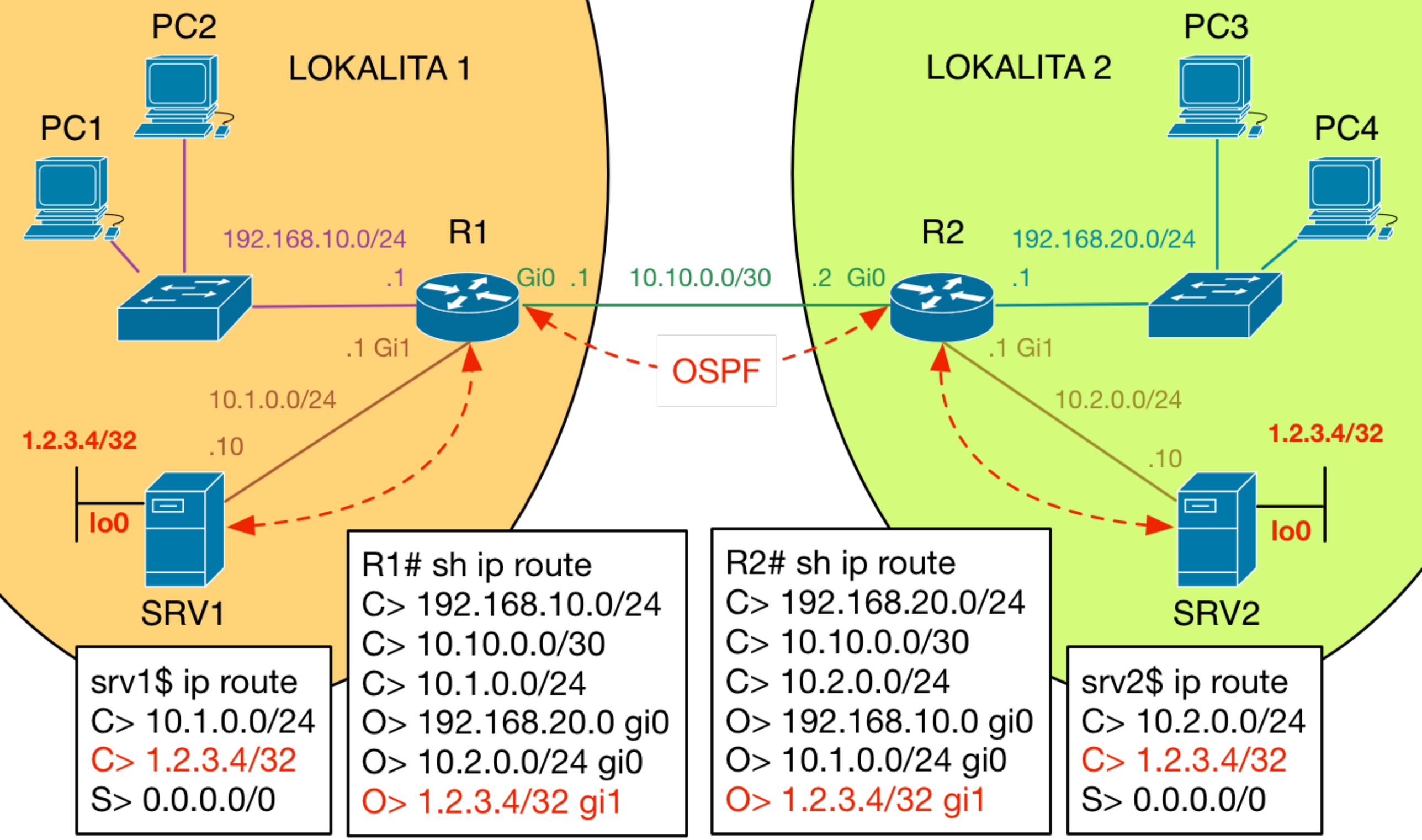
Aplikace OSPF místo fail-overu IP adres

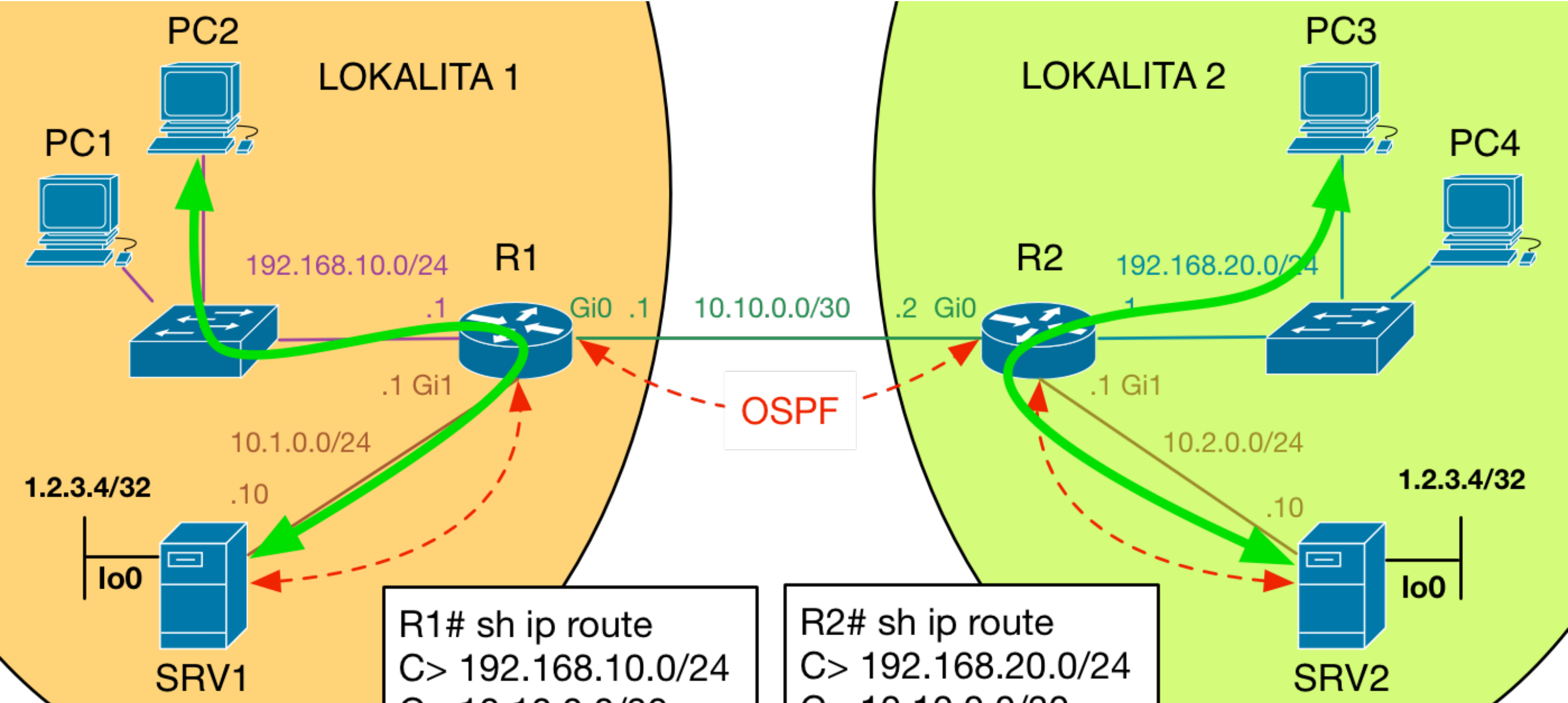
HA služba @ 2 lokality

- 2 lokality spojené spojovací sítí
- V každé lokalitě je umístěn server který realizuje službu **http://1.2.3.4**
- V standardním režimu místní server odbavuje místní klienty
- V případě, že server přestane fungovat měl by jej nahradit server v druhé lokalitě
- Na dalších slide v routovacích tabulkách: “gi0” a “gi1” uvedeno místo IP adres


PC2
LOKALITA 1
PC3
LOKALITA 2
PC1
PC4
192.168.10.0/24
R1
R2
192.168.20.0/24
.1
Gi0 .1
10.10.0.0/30
.2 Gi0
.1
.1 Gi1
.1 Gi1
10.1.0.0/24
10.2.0.0/24
.10
.10
SRV1
SRV2
R1# sh ip route
C> 192.168.10.0/24
C> 10.10.0.0/30
C> 10.1.0.0/24
R2# sh ip route
C> 192.168.20.0/24
C> 10.10.0.0/30
C> 10.2.0.0/24
srv1$ ip route
C> 10.1.0.0/24
S> 0.0.0.0/0
srv2$ ip route
C> 10.2.0.0/24
S> 0.0.0.0/0


PC2
LOKALITA 1
PC3
LOKALITA 2
PC1
PC4
192.168.10.0/24
R1
R2
192.168.20.0/24
.1
Gi0 .1
10.10.0.0/30
.2 Gi0
.1
.1 Gi1
OSPF
.1 Gi1
10.1.0.0/24
10.2.0.0/24
.10
.10
SRV1
SRV2
srv1$ ip route
C> 10.1.0.0/24
S> 0.0.0.0/0
R1# sh ip route
C> 192.168.10.0/24
C> 10.10.0.0/30
C> 10.1.0.0/24
O> 192.168.20.0 gi0
O> 10.2.0.0/24 gi0
R2# sh ip route
C> 192.168.20.0/24
C> 10.10.0.0/30
C> 10.2.0.0/24
O> 192.168.10.0 gi0
O> 10.1.0.0/24 gi0
srv2$ ip route
C> 10.2.0.0/24
S> 0.0.0.0/0


PC2
LOKALITA 1
PC3
LOKALITA 2
PC1
PC4
192.168.10.0/24
R1
R2
192.168.20.0/24
.1
Gi0 .1
10.10.0.0/30
.2 Gi0
.1
.1 Gi1
OSPF
.1 Gi1
10.1.0.0/24
10.2.0.0/24
1.2.3.4/32
.10
.10
1.2.3.4/32
lo0
lo0
SRV1
SRV2
srv1$ ip route
C> 10.1.0.0/24
C> 1.2.3.4/32
S> 0.0.0.0/0
R1# sh ip route
C> 192.168.10.0/24
C> 10.10.0.0/30
C> 10.1.0.0/24
O> 192.168.20.0 gi0
O> 10.2.0.0/24 gi0
O> 1.2.3.4/32 gi1
R2# sh ip route
C> 192.168.20.0/24
C> 10.10.0.0/30
C> 10.2.0.0/24
O> 192.168.10.0 gi0
O> 10.1.0.0/24 gi0
O> 1.2.3.4/32 gi1
srv2$ ip route
C> 10.2.0.0/24
C> 1.2.3.4/32
S> 0.0.0.0/0

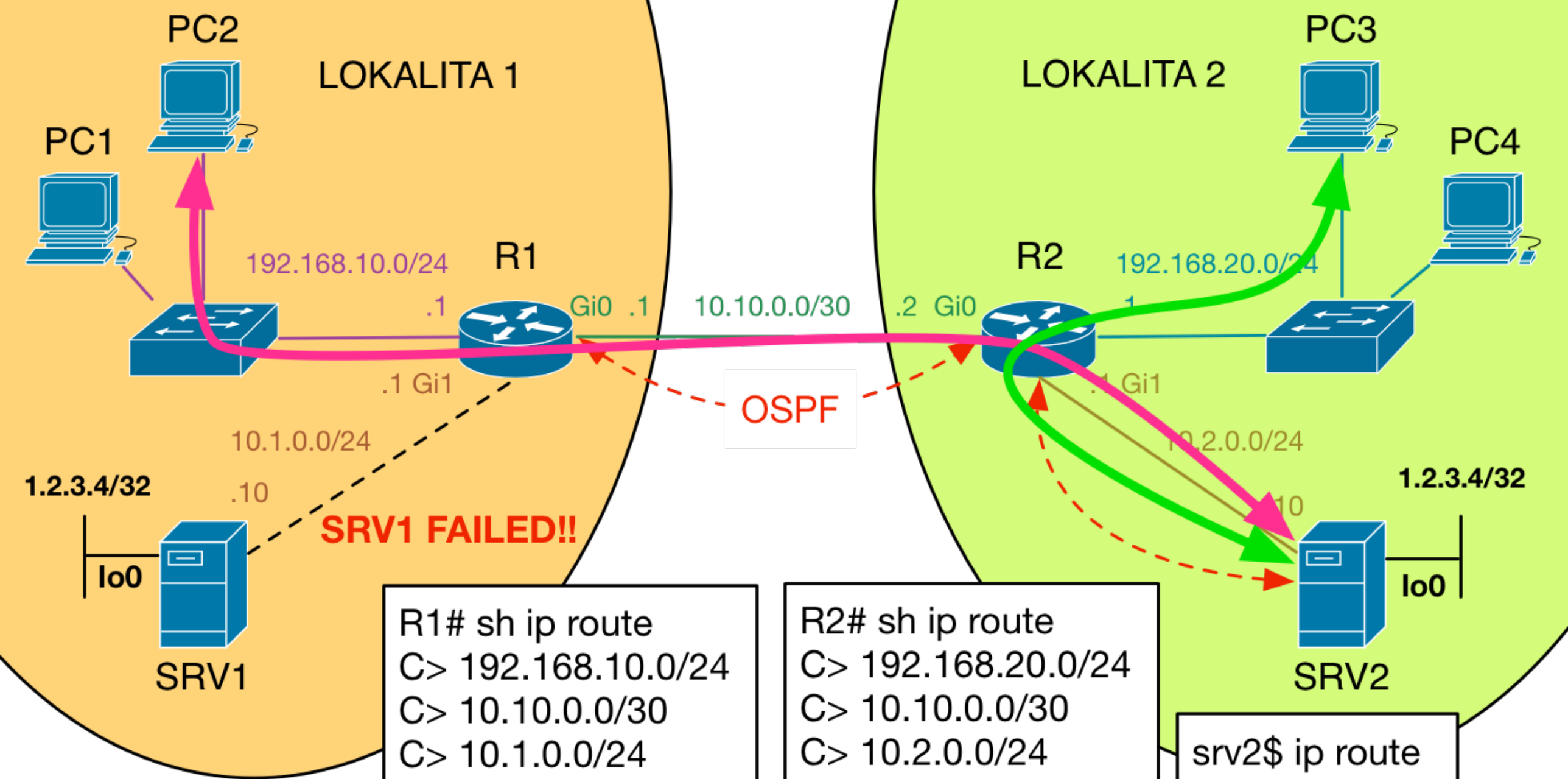


srv1$ ip route
C> 10.1.0.0/24
C> 1.2.3.4/32
S> 0.0.0.0/0

R1# sh ip route
C> 192.168.10.0/24
C> 10.10.0.0/30
C> 10.1.0.0/24
O> 192.168.20.0 gi0
O> 10.2.0.0/24 gi0
O> 1.2.3.4/32 gi1

R2# sh ip route
C> 192.168.20.0/24
C> 10.10.0.0/30
C> 10.2.0.0/24
O> 192.168.10.0 gi0
O> 10.1.0.0/24 gi0
O> 1.2.3.4/32 gi1

srv2$ ip route
C> 10.2.0.0/24
C> 1.2.3.4/32
S> 0.0.0.0/0



R1# sh ip route
C> 192.168.10.0/24
C> 10.10.0.0/30
C> 10.1.0.0/24
O> 192.168.20.0 gi0
O> 10.2.0.0/24 gi0
O> 1.2.3.4/32 **gi0**

R2# sh ip route
C> 192.168.20.0/24
C> 10.10.0.0/30
C> 10.2.0.0/24
O> 192.168.10.0 gi0
O> 10.1.0.0/24 gi0
O> 1.2.3.4/32 gi1

srv2$ ip route
C> 10.2.0.0/24
C> 1.2.3.4/32
S> 0.0.0.0/0

Co dál?

- Virtual Switching System (VSS)
- FabricPath / TRILL
- QoS, FRR
- Disaster Recovery
- Synchronizace databází
- …

Dotazy?