UNIVERSITY OF CAMBRIDGE

CUDN PoP Switch Changes 2018

New hardware and configuration changes

UIS Networks - Systems

Agenda

- New PoP switch choices
  - Port assignments
- Recommendations on connecting
- Spanning Tree — now and changes
  - How these will interact with your network
- DHCP Snooping & ARP Inspection — now and changes

PoP switch choices

Catalyst 3850-24P-L — “1G Option”
20 of 24x 10/100/1000M copper, 435W PoE+
4x 1G SFP inc. 2x 1G up-/downlink
No upgrade charge
£2,688 /year

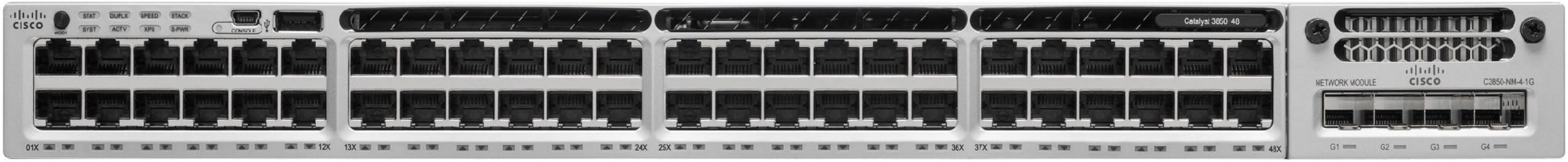
Catalyst 3850-48P-L — “10G Option 1”
44 of 48x 10/100/1000M copper, 800W PoE+
4x 1/10G SFP+ inc. 2x 10G up-/downlink
£5,270 upgrade now • £16,434 thereafter
£5,737 /year

Catalyst 3850-12XS-S — “10G Option 2”
12x 1/10G SFP+ inc. 2x 10G up-/downlink
£3,495 upgrade now • £14,660 thereafter
£5,484 /year

Other models *may* be available on request!

+ VAT, if applicable

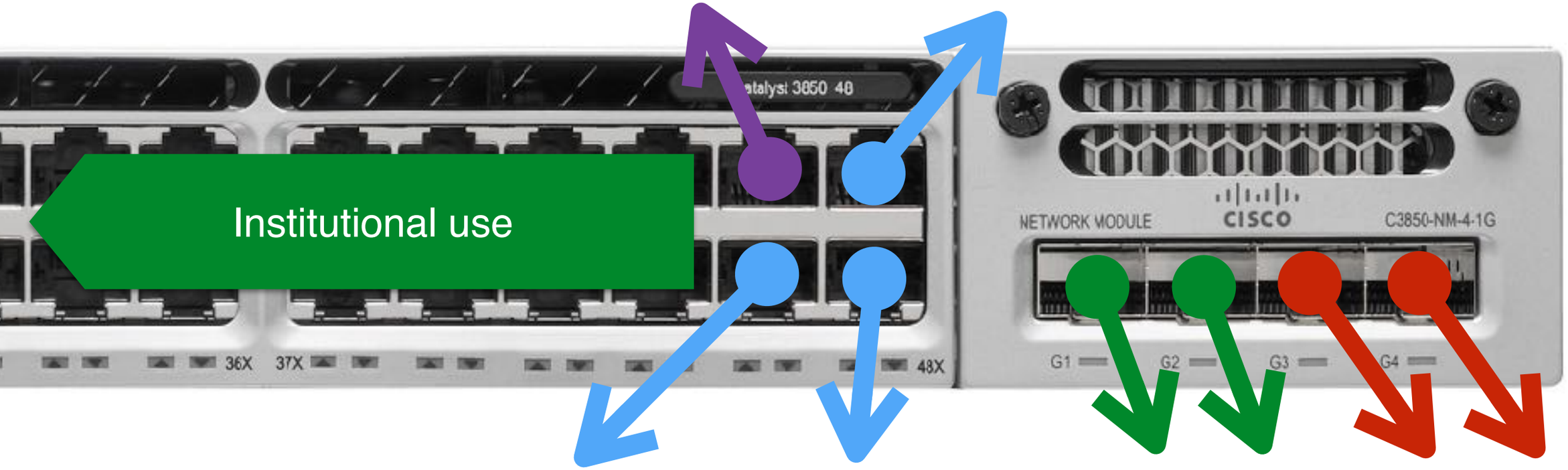
Copper PoP port assignments



↑ Module slot

Institution diagnostic port (voice+data)

UIS Networks reserved



Institutional use

UIS Networks monitoring

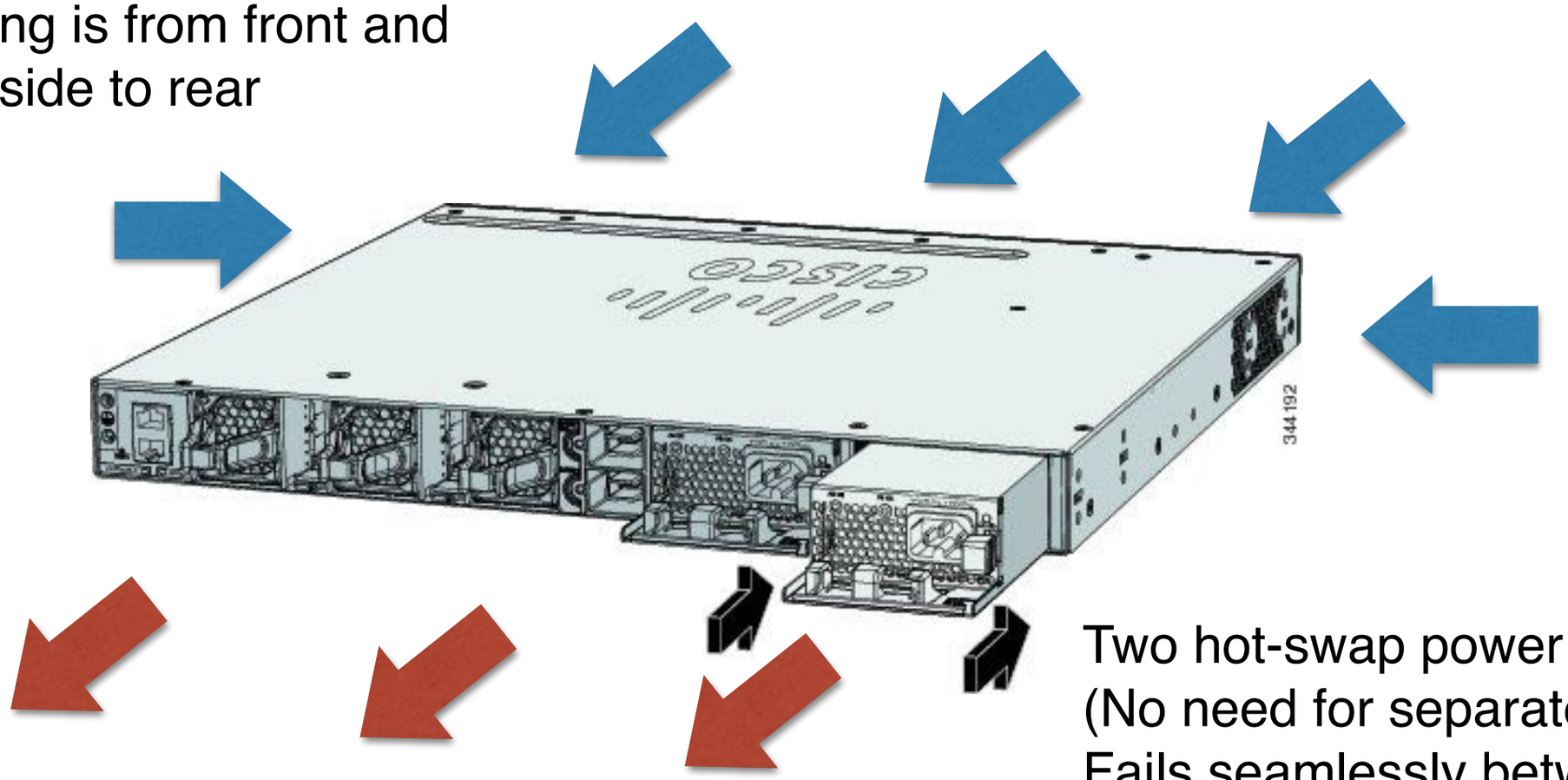
UPS or reserved

Institutional use

Uplinks to CUDN

Catalyst 3850 power & cooling

Cooling is from front and front/side to rear



Two hot-swap power supplies
(No need for separate RPS unit)
Fails seamlessly between units

UPS is now NOT supplied as standard: £275 /year (+ VAT, if applicable)
Cost covers replacing of batteries every 3 years t monitoring
(We'll be taking yours away, if you no longer want it.)

BGP choices

1G BGP Connection	2x 1GE up-/downlinks	No upgrade charge £2,198 [+VAT] /year
10G BGP Connection	2x 10GE up-/downlinks	No upgrade charge £3,386 [+VAT] /year

BGP requires you to:

- Route all VLANs with ACLs, DHCP relaying, etc.
- Handle multicast (for the voice network)
- Potentially handle VRFs for MPLS VPNs
  - ... else we may need to install separate switches

Recommendations connecting to a PoP

- Due to ECMP (Equal Cost MultiPath), the CUDN *down*links can deliver 1-2Gbit/s down to a 1G PoP, or 10-20Gbit/s to a 10G PoP
  - Later on we may upgrade *up*links similarly (GLBP?)
- This can overwhelm a single 1Gbit/s link
- The UIS Managed Firewall Service is connected "on a stick" and typically uses 2x links for redundancy and performance
  - Data VLANs fed through Managed Firewalls get 2x 1/10Gbit/s up and down now due to ECMP routing to/from them
  - Voice and wireless traffic do NOT go through it
- **Consider an LACP port-channel/trunk to your main, top level switch(es)**

Spanning Tree Protocol (STP) changes

Current STP configuration (2008)

- CUDN runs **Cisco Rapid PVST+** ("Per-VLAN Spanning Tree Plus") internally
- At the time, the CUDN used to extend VLANs across sites and use Spanning Tree to provide redundancy
- BPDUs (Bridge Protocol Data Units — Spanning Tree information frames) filtered on ports into institution, and "`portfast [trunk]`" enabled, blocking interaction with institutional Spanning Trees, for two reasons:

1. To avoid institutional spanning trees upsetting the CUDN backbone
2. Interoperability between different protocols/vendors

STP situation changes

- The CUDN no longer feeds VLANs between sites
  - The backbone is entirely routed
  - Best practice is now to use Spanning Tree only to detect *fault* loops, not to build redundant configurations, where possible
- On the CUDN, Spanning Tree only used between site routers and institutional PoP
- Loops within institutions are still a common problem and can upset a backbone router

New STP configuration (2018)

- The CUDN will continue to run **Rapid PVST+**
  - (Ideally we'd like to use a standard, but there isn't a practical one: MSTP [Multi STP] is the only one which supports VLANs, but it's a horrible mess, and doesn't work well across administrative boundaries.)
- **BPDUs will cease to be filtered** on the ports into an institution from the PoP and **portfast disabled** on trunk ports
- The CUDN will run a **root bridge** with a priority not lower than 16,384 (on the PoP)
- Institutions are free to join the Spanning Tree, if they wish
  - You can even take over the root bridge (priority <16,384)

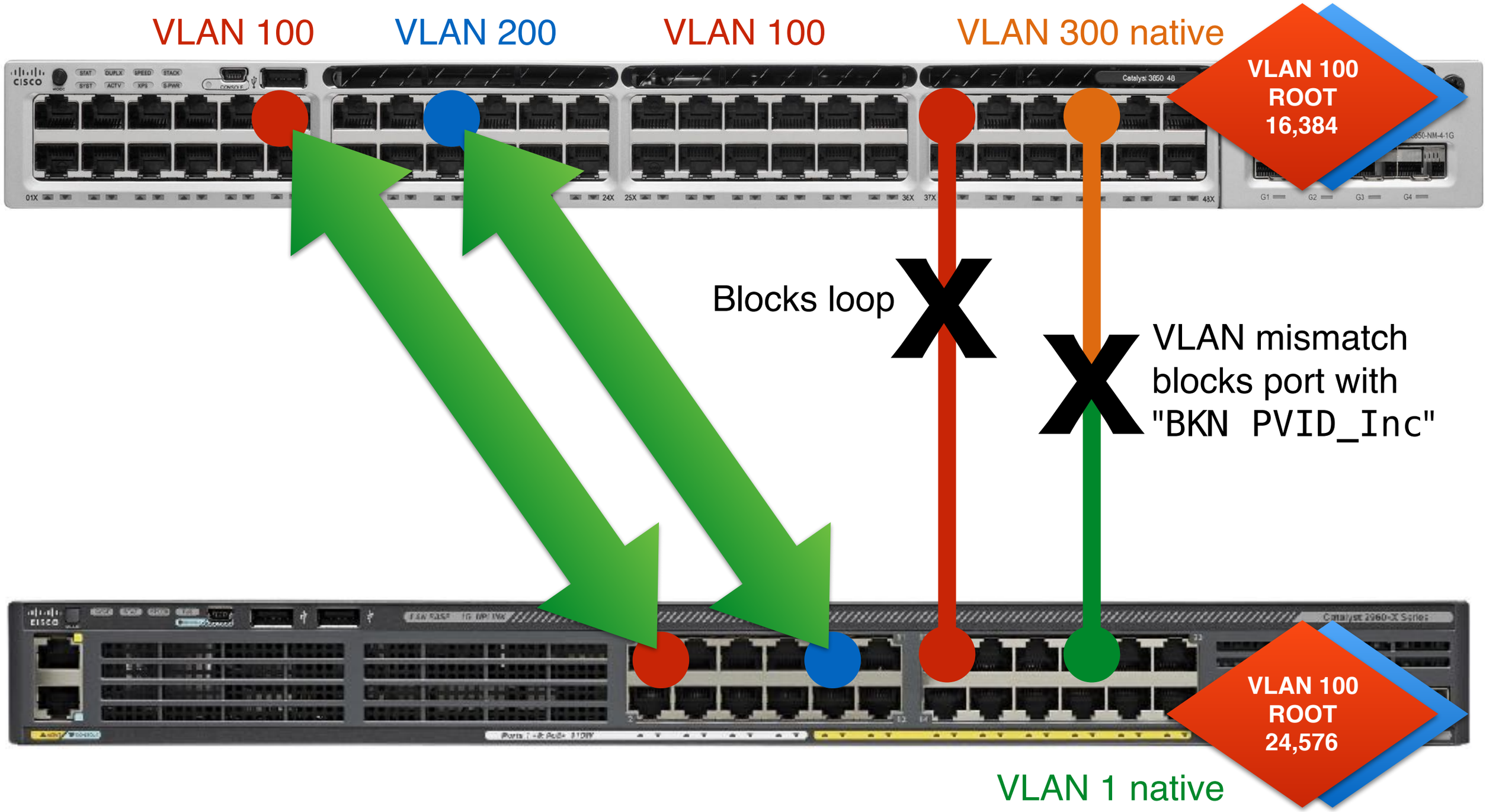
Spanning Tree Protocol (STP) changes

(What does that actually mean to me?)

Scenario 1: PVST+ ↔ PVST+

- **Cisco switches default to [non-rapid] PVST+**
  - Other manufacturers can often be put into this mode
- Institutional & PoP switches should now discover each other w.r.t. Spanning Tree
  - Links will begin forwarding immediately (no 30s delay) if Rapid
  - One root bridge, determined by priority
  - Loops should be detected and blocked, as appropriate
- **Things to beware of**:
  - IDs of VLANs on untagged/native ports MUST match else "BKN_PVID_Inc" port error and will block
  - If you're not using the CUDN VLAN ID, you should filter BPDUs ("`spanning-tree bpdufilter enable`")

PVST+ ↔ PVST+ — PoP links



PVST+ ↔ PVST+ — edge loop

VLAN 100

VLAN 100

VLAN 100

Looped VLANs blocked at edge, isolating problem and removing effect on rest of network

PVST+ ↔ PVST+ — edge VLAN mismatch

VLAN 100

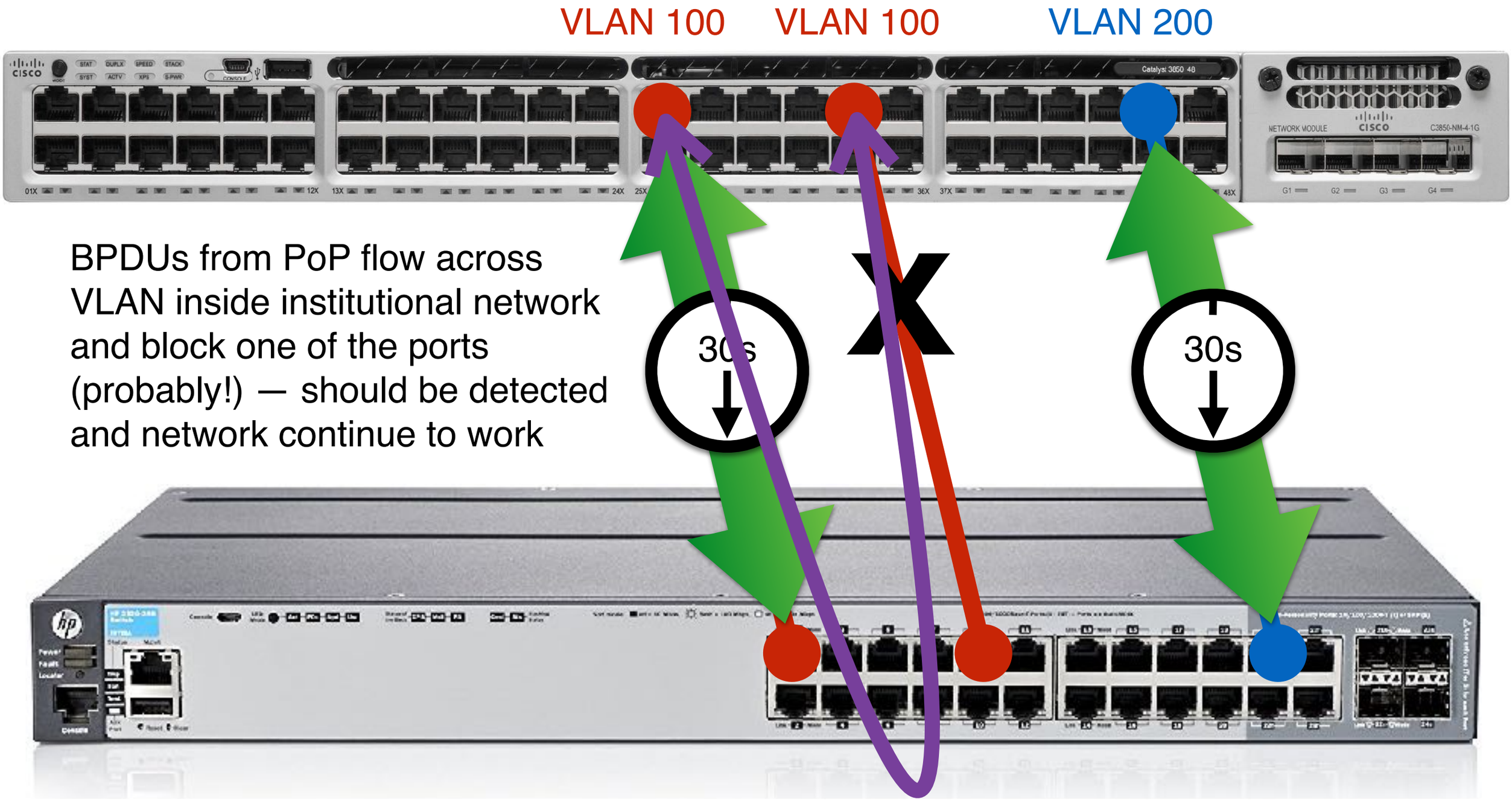
VLAN 100 VLAN 200

PVST+ detects mismatched VLAN and blocks port in "BKN PVID_Inc" state

Scenario 2: PVST+ ↔ IEEE STP

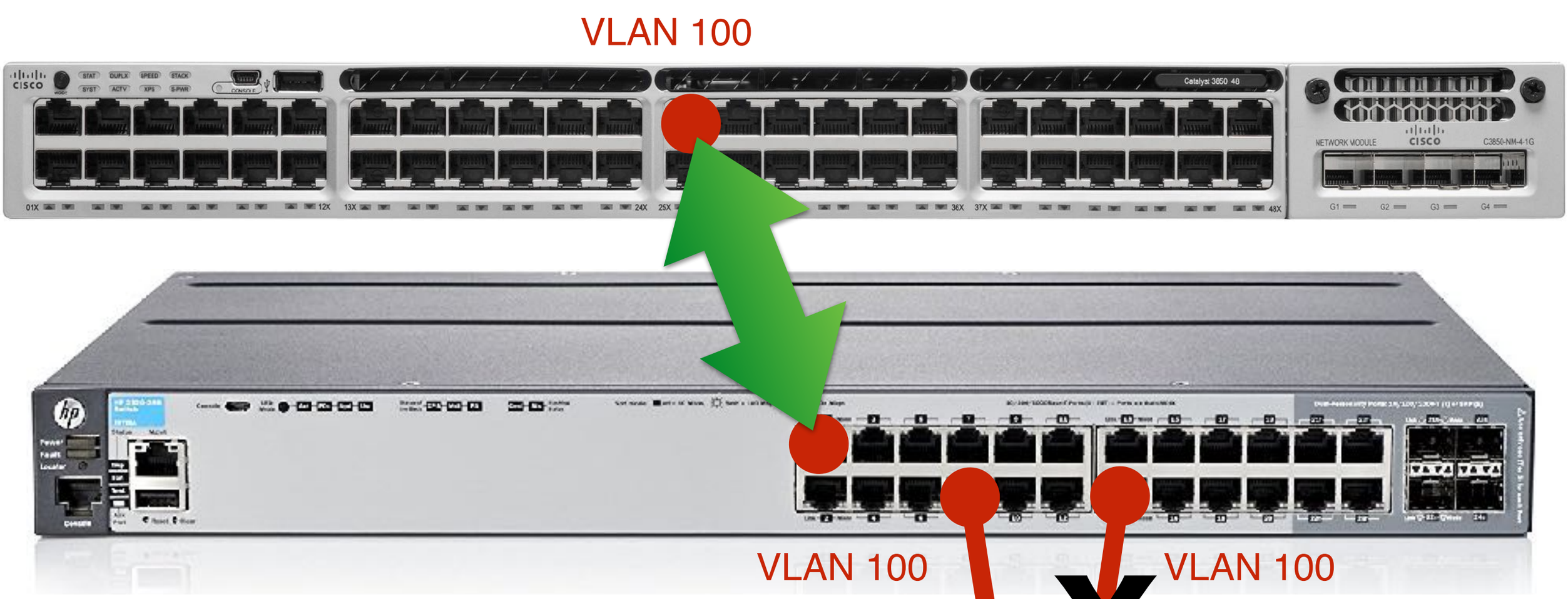
- IEEE STP is **RSTP (Rapid STP)** or **MSTP (Multi STP)**
  - **HP Comware (MST)** & most (other than Cisco) vendors default to these
- PVST+ and IEEE STP do NOT interact (except with VLAN 1 on PVST+)
  - BPDUs sent and received but are ignored by switches running the other protocol
  - Separate root bridges for PVST+ and IEEE STP
  - However, PVST+ BPDUs will *usually* flow through VLANs on IEEE STP switches and come back to the PoP
- **Things to beware of:**
  - Ports will take 30s to begin forwarding traffic
  - Making a loop on a VLAN will *likely* be detected by the PoP and one port will block — DO NOT build redundant topologies using this!
  - Joining two different VLANs will block the ports with "`BKN PVID_Inc`" on the PVST+ side on BOTH ports

PVST+ ↔ IEEE STP — PoP links



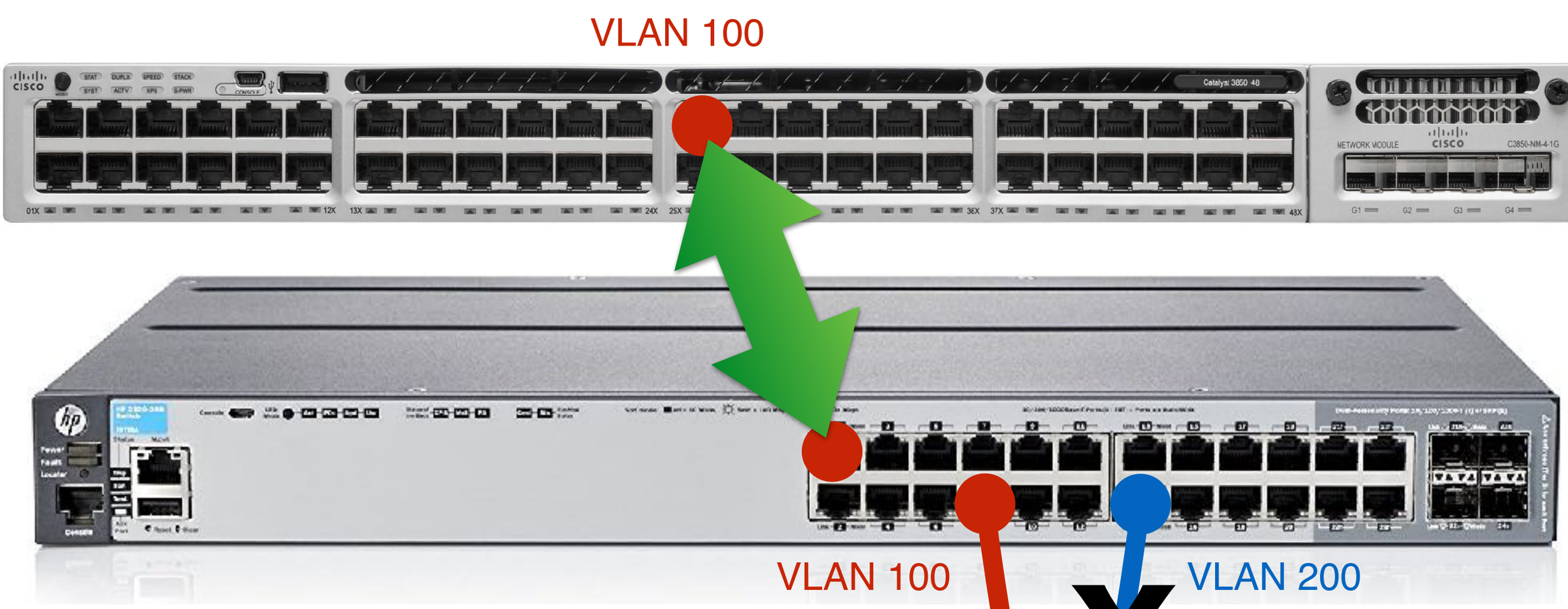
BPDUs from PoP flow across VLAN inside institutional network and block one of the ports (probably!) — should be detected and network continue to work

PVST+ ↔ IEEE STP — edge loop



Institutional IEEE STP will detect loop at the edge and block within institution, isolating problem and removing effect on rest of network

PVST+ ↔ IEEE STP — edge VLAN mismatch



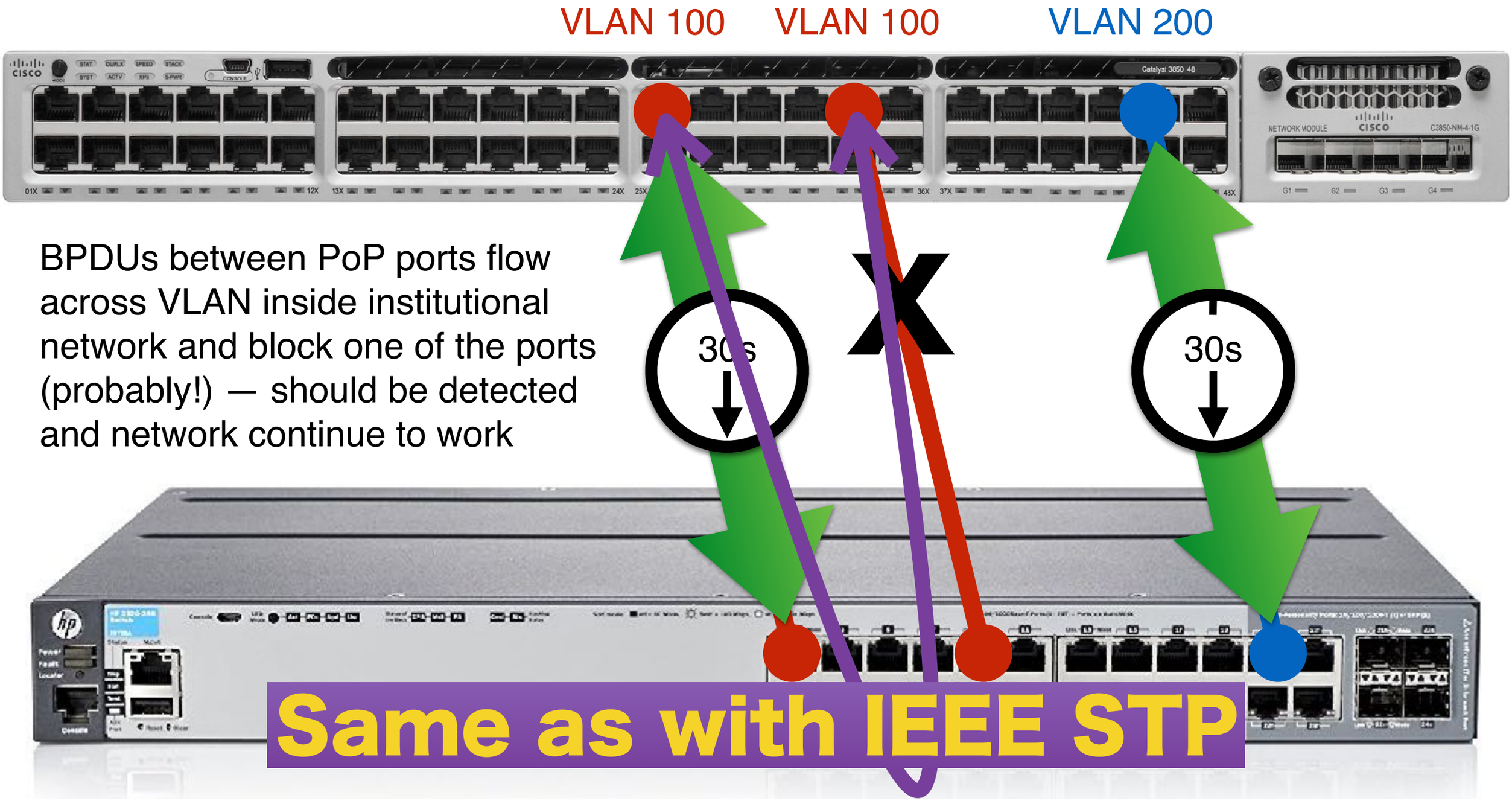
IEEE STP will detect looped port (it doesn't care about VLANs) and blocks the port at one end of the link

Also doesn't matter if VLAN fed from PoP or local

Scenario 3: PVST+ ↔ no STP

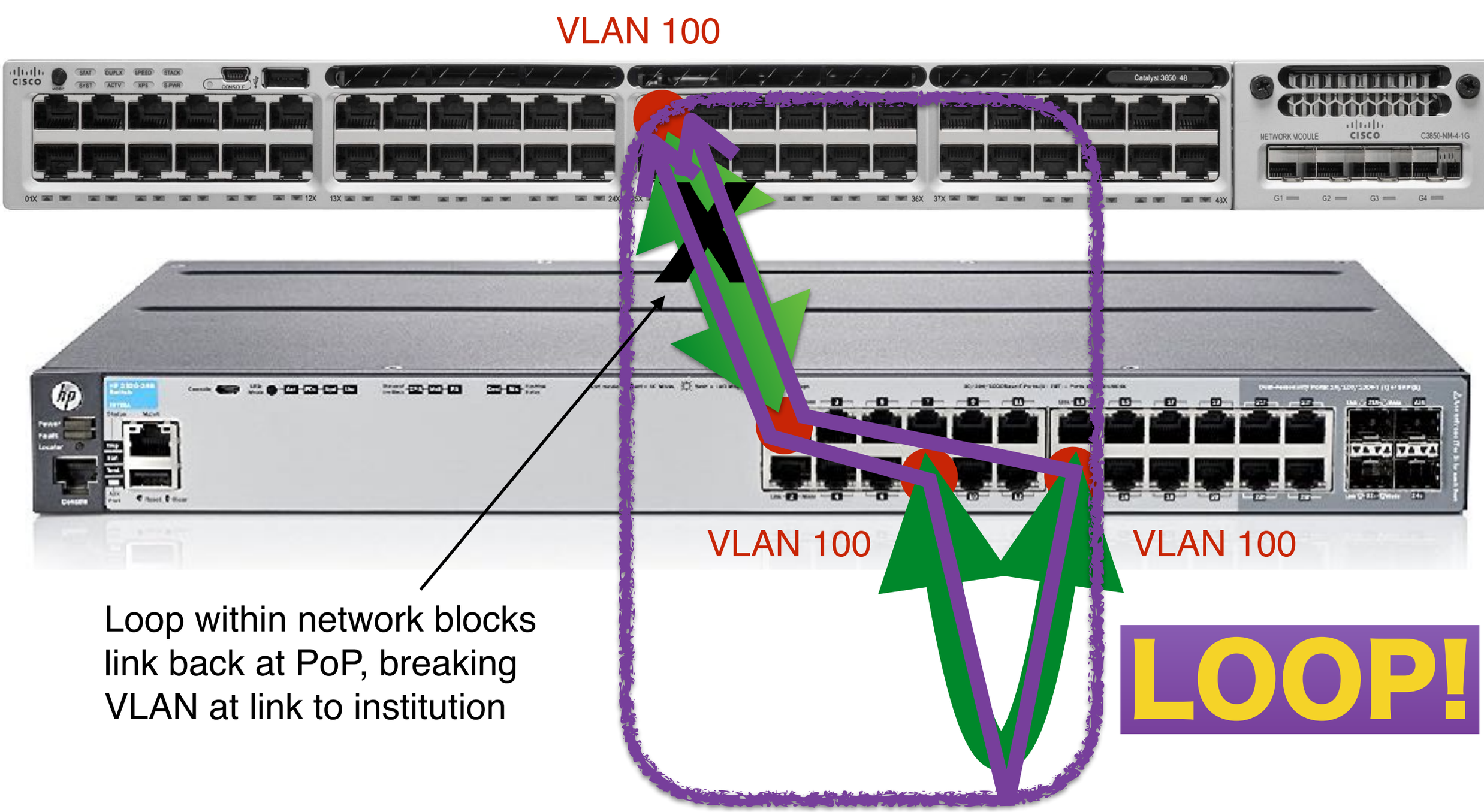
- **HP ProCurves default to Spanning Tree disabled** ("`no spanning-tree`")
- Effect is similar to when running an IEEE STP: the packets flow through the HP ProCurve and make their way back to the PoP
- **Things to beware of:**
  - Ports will take 30s to begin forwarding traffic
  - Institutional network will not detect loops or mismatched VLANs and block them itself
  - In the absence of this the PoP will likely do so, at a more institutional level!

PVST+ ↔ no STP — PoP links



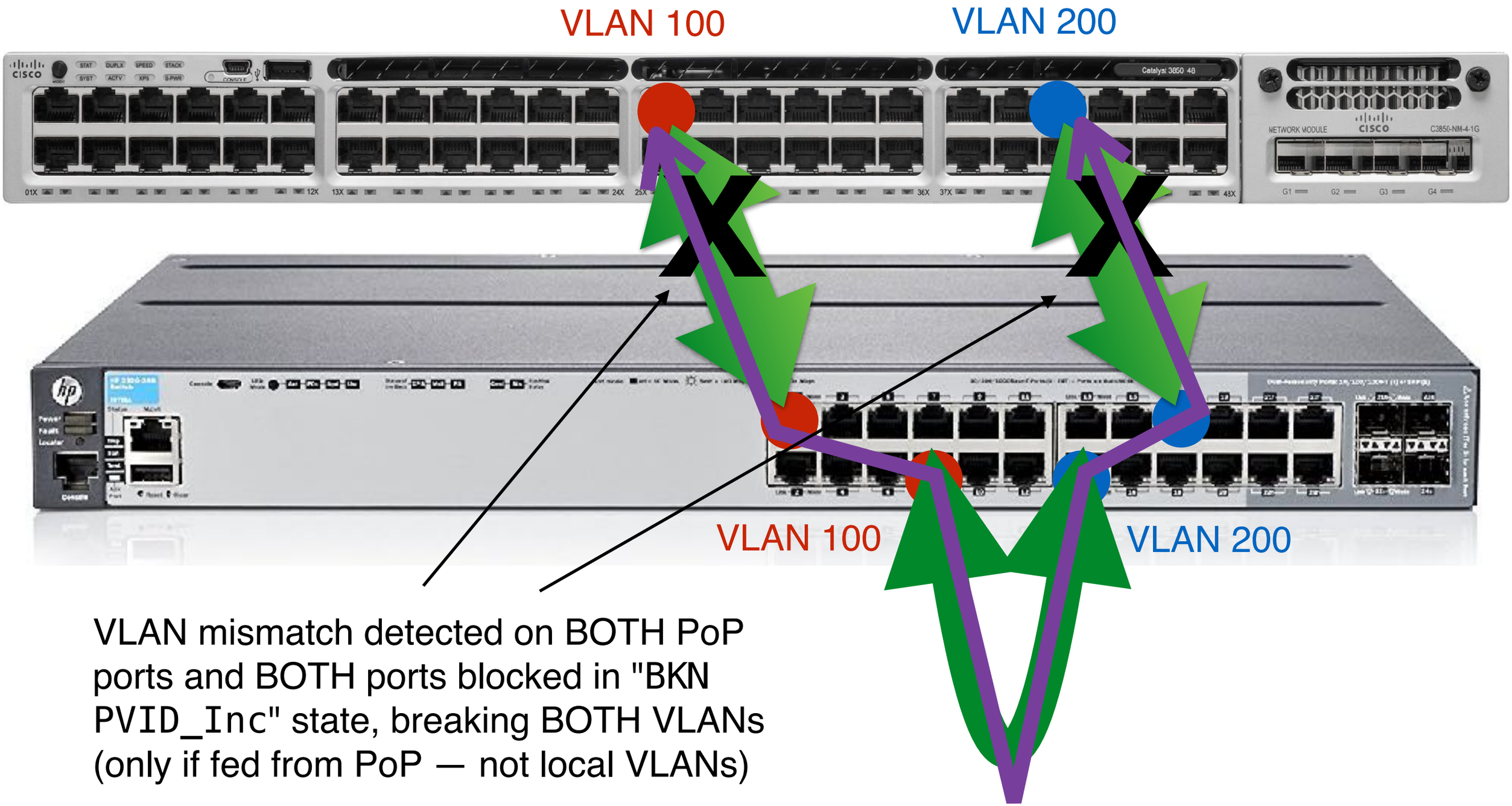
BPDUs between PoP ports flow across VLAN inside institutional network and block one of the ports (probably!) — should be detected and network continue to work

PVST+ ↔ no STP — edge loop



Loop within network blocks link back at PoP, breaking VLAN at link to institution

PVST+ ↔ no STP — edge VLAN mismatch

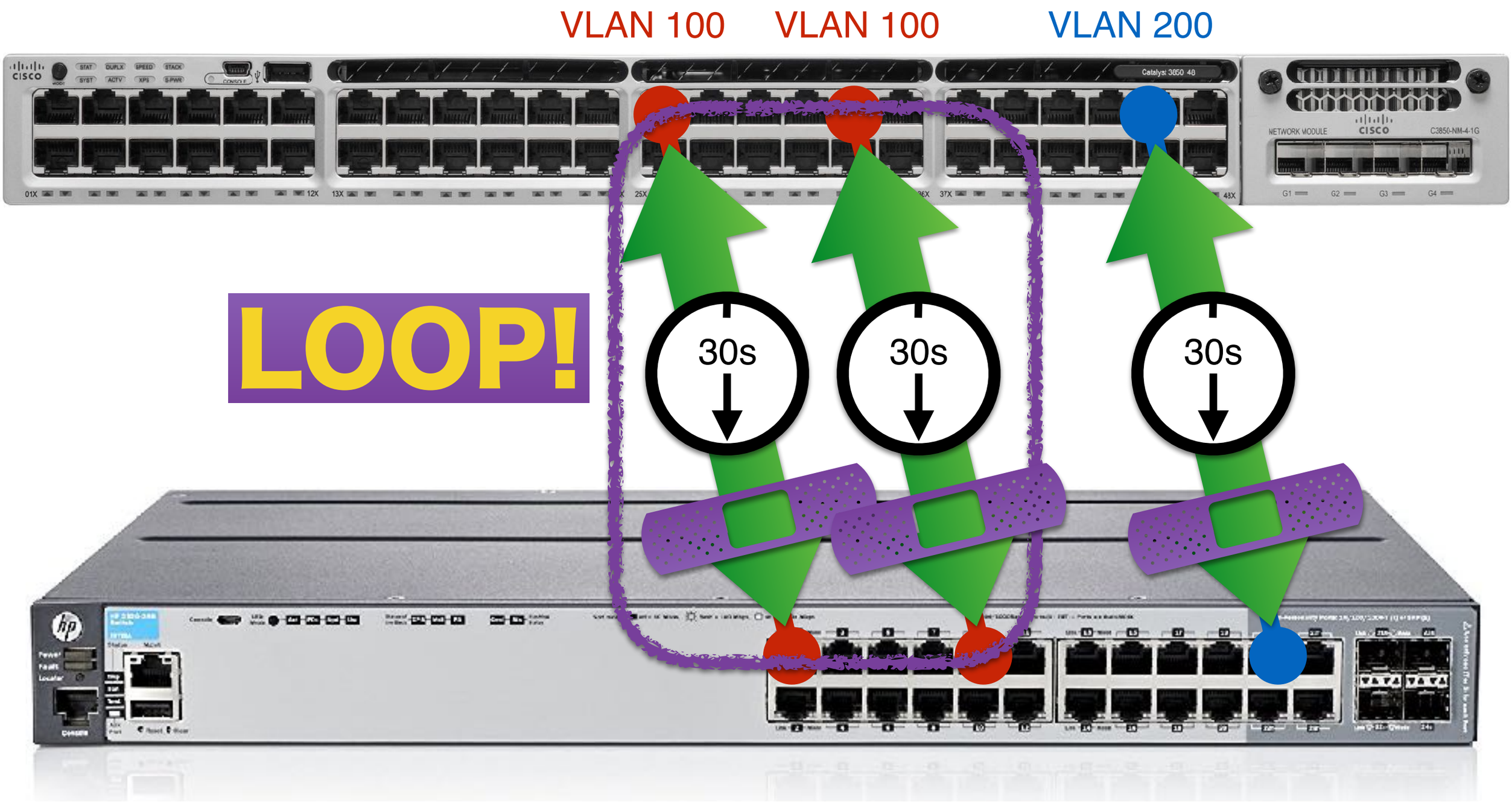


VLAN mismatch detected on BOTH PoP ports and BOTH ports blocked in "BKN PVID_Inc" state, breaking BOTH VLANs (only if fed from PoP — not local VLANs)

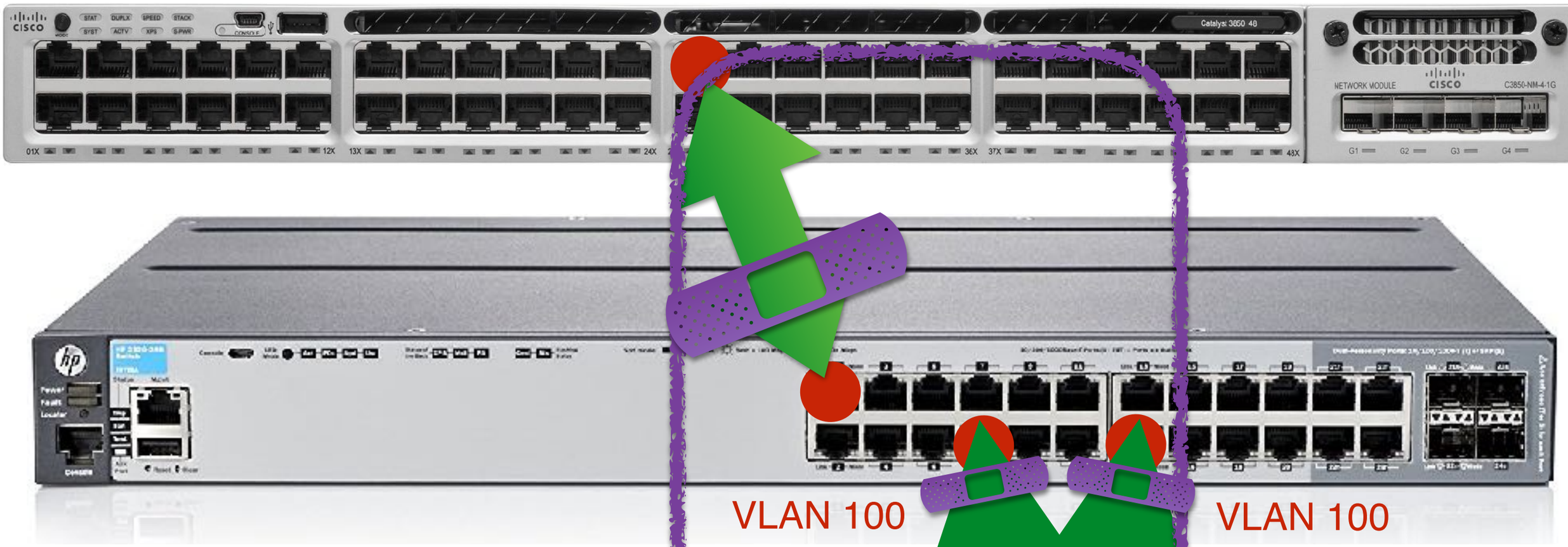
Scenario 4: PVST+ ↔ filtered BPDUs

- **Nothing does this by default** but you might be, given the previous CUDN configuration
- Institution & PoP switches will not discover each other w.r.t. Spanning Tree (as before)
- **Things to beware of**:
  - Ports will take 30s to begin forwarding traffic
  - Problems will not be discovered and issues may result in catastrophic failures!

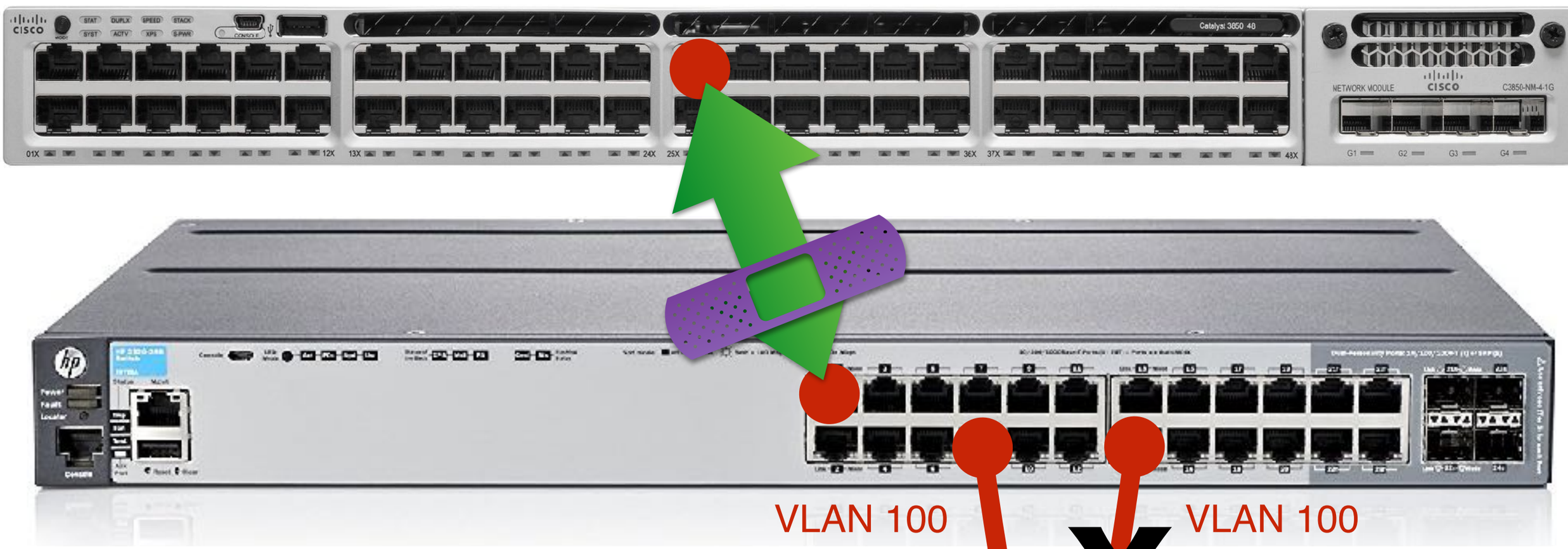
PVST+ ↔ filtered BPDUs — PoP links



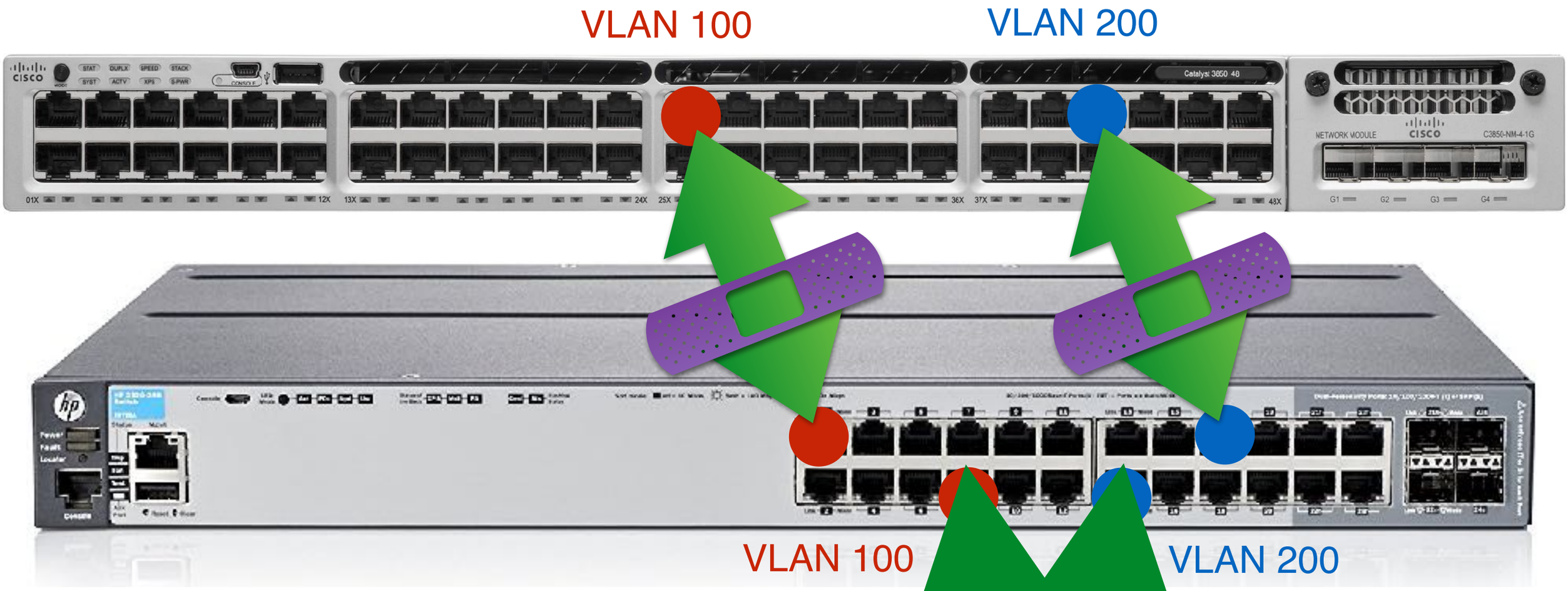
PVST+ ↔ superfiltered BPDUs — edge loop



PVST+ ↔ filtered BPDUs — edge loop



PVST+ ↔ filtered BPDUs — edge VLAN mismatch



Interconnected VLANs just join traffic together: will break things such as DHCP and a second such mismatch will cause loop

Why you shouldn't filter BPDUs

- Filtering prevents the loop detection working
- Keeps links up but can have serious impacts on the upstream network, beyond your institution
- With Spanning Tree, the port will move back into forwarding in 30s, once the fault is resolved
- In the absence of Spanning Tree, the PoP switches do other things to detect loops which are more serious and block ports for minutes on end
- In serious cases, we may manually disable your ports until the problem is resolved

Spanning Tree recommendations

- **Enable Spanning Tree**
- You can use either Rapid PVST+ or IEEE STP (RSTP or MSTP)
  - Rapid PVST+ may be preferable due to quicker convergence with the PoP (but beware VLAN mismatch)
- Set a priorities to locate the root bridge centrally
- Enable “root guard” or “BPDU guard” on edge ports
- Enable "portfast" on edge ports:
  - On Cisco use “`spanning-tree portfast default`”
  - On HP ProCurve use “auto-edge” mode (default)
- Use VLAN IDs matching those on the CUDN, or local IDs for internal VLANs

DHCP Snooping & ARP Inspection

DHCP Snooping

- The switch intercepts **DHCP packets** passing through it to process them
- Permit or deny the packet to flow through the switch and control which ports it goes to
  - Client→Server packets don't go to untrusted ports
  - Server→Client packets blocked from untrusted ports
- Rate limit packets
- Add "Option 82" information about the edge port
- Builds "binding table" to learn about assignments

ARP Inspection

- The switch intercepts ARP packets passing through it to process them
- Blocks ARP Replies from untrusted ports where the IP address wasn't seen being assigned via DHCP first
  - Uses the DHCP Snooping binding table
- Rate limit packets

Current PoP situation

- Some PoP switches use DHCP Snooping and ARP Inspection on non-data VLANs (e.g. voice, APs, etc.) to avoid address spoofing
- Some rate limits applied on institutional downlinks to filter out storms following loops
  - ... loops which Spanning Tree should have caught
- All a bit ad hoc
- Doesn't protect against issues on the data VLAN (at least unless we impose "special measures")

New PoP configuration

- DHCP Snooping and ARP Inspection enabled on ALL VLANs (inc. main data VLAN)
- **Trunk ports** will be set to:
  - Be trusted for DHCP Snooping and ARP Inspection
  - Will apply a [high] rate limit on both to filter out storms
- **Access/edge ports** will be set to:
  - Be *un*trusted for DHCP Snooping
  - Be trusted for ARP Inspection
  - Will apply a [low] rate limit on both to filter out storms

What will this mean?

- Probably nothing, especially if you have Spanning Tree turned on internally
- If you don't, DHCP and ARP upstream will be unreliable as random packets get dropped
- The PoP may become unresponsive (as DHCP and ARP packets are punted to the CPU to be processed, rather than switched in hardware)
- You can't run a DHCP Server on a PoP edge port
- We'll monitor the situation
- You should run these internally

Thank you — any questions?