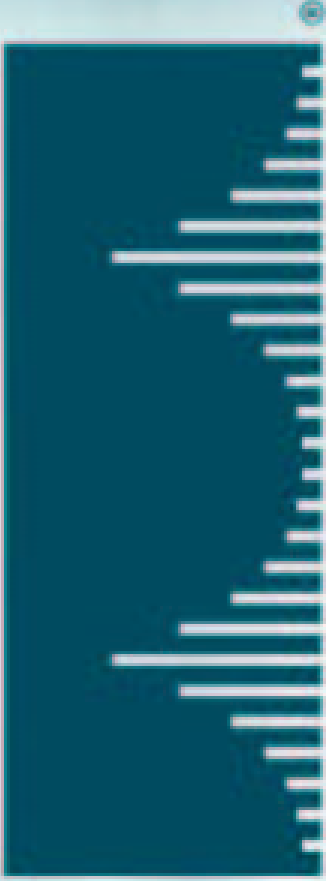


CISCO SYSTEMS



IEEE Spanning Tree 802.1w und 802.1s

Frank Bartel
Internetworking Consultant
fbartel@cisco.com

Agenda

- **IEEE Spanning Tree Erweiterungen**
 - 802.1w - Rapid Spanning Tree
 - 802.1s - Multiple Spanning Tree
- **Implementation in der Cisco Catalyst Familie**
- **Weiter Informationen zu den Themen**
- **Q & A**

IEEE 802.1w Rapid Spanning Tree

Was ist IEEE 802.1w

- **802.1w - Rapid Spanning-Tree beschreibt eine Erweiterung des bestehenden 802.1D SPT.**
- **802.1 w ist mehr eine Evolution des SPT als eine Revolution.**
- **Ist ähnlich zu den bekannten UplinkFast und BackboneFast Erweiterungen von Cisco.**
- **802.1w verbessert die Konvergenzzeiten von STP.**
- **Cisco empfiehlt bei Neu-Installationen (Green Fields) durchgehend 802.1s/w.**
- **Ciscos Implementierung von 802.1s und 802.1w erlaubt z.T. keine Trennung zwischen beiden Standards. Details: siehe weiter unten**

RSTP: Neue Konzepte

- **Neue Port Rolle und Status**
- **Modifizierte BPDU**
- **“Proposal/Agreement” Nachrichten zwischen den Bridges**
- **BPDU Behandlung**
- **Neuer “Topology Change” Mechanismus**
- **PVST+/802.1D Kompatibilität**

Port Status: IEEE & Cisco Terminologie

Cisco.com

- **Discarding** == **Blocking** == **Listening**
- **Learning** == **Learning**
- **Forwarding** == **Forwarding**

RSTP Rollen und Aufgaben (1)

RSTP definiert 4 Port-Rollen:

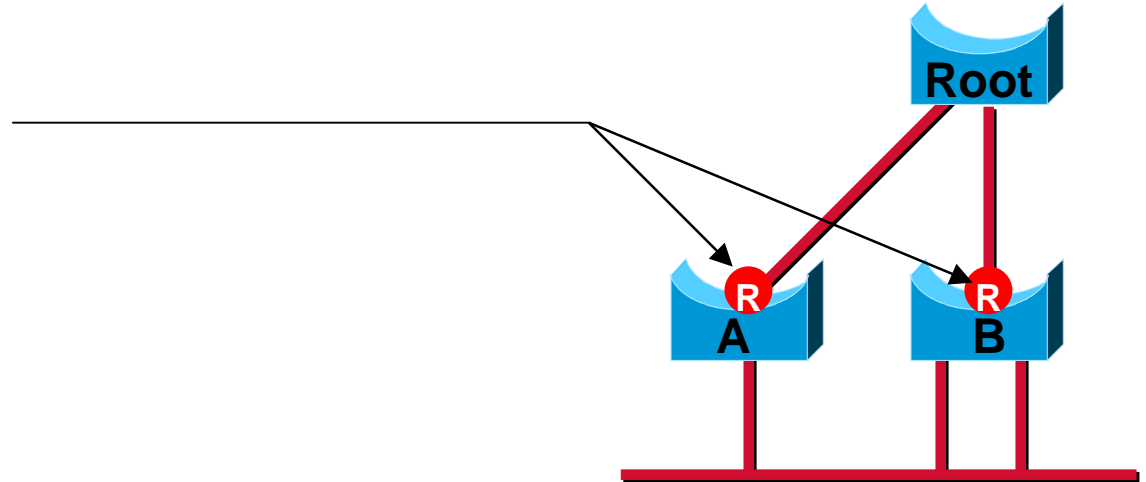
- Root port
 - Designated port
 - Alternate port
 - Backup port
- } Beide im blocking modus

**Port Status (blocking, forwarding, learning)
ist unabhängig von der Port Rolle**

RSTP Rollen (2)

Cisco.com

R Root Port



Root Port:

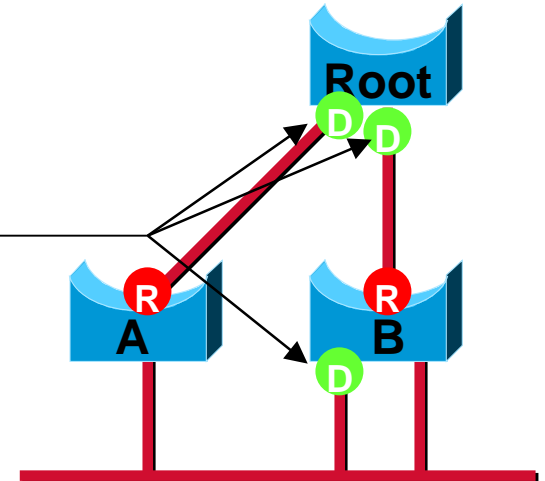
Port, der die besten BPDUs für diese Bridge empfängt.

Kürzester Pfad zur Root unter dem Gesichtspunkt path cost.

RSTP Rollen (3)

Cisco.com

- R** Root Port
- D** Designated Port



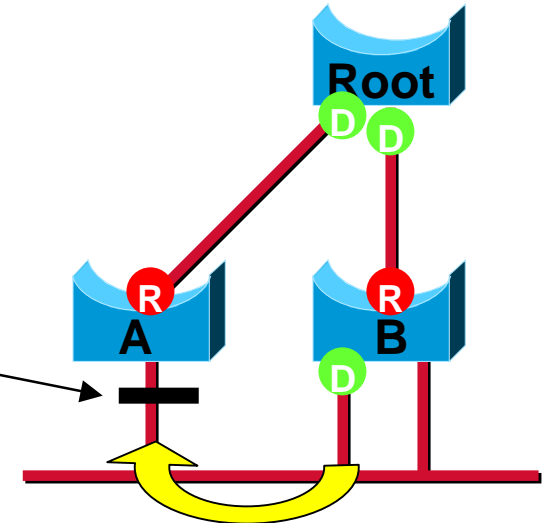
Designated Port:

Port, der die besten BPDUs auf ein Segment sendet.

RSTP Rollen (4)

Cisco.com

- Root Port
- Designated Port
- Alternate Port



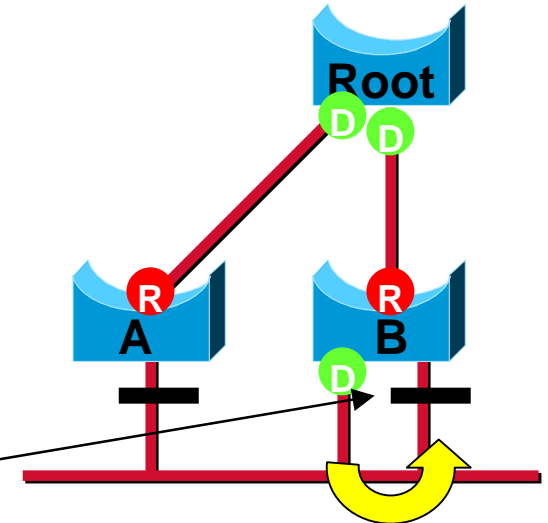
Alternate Port:

Port, der durch BPDUs einer anderen Bridge geblockt ist - Redundanter Pfad zur Root.

RSTP Rollen (5)

Cisco.com

- **R** Root Port
- **D** Designated Port
- Alternate Port
- Backup Port



Backup Port:

Port, der durch BPDUs der gleichen Bridge geblockt ist
Redundanter Pfad zu einem Segment.

Modifizierte BPDU

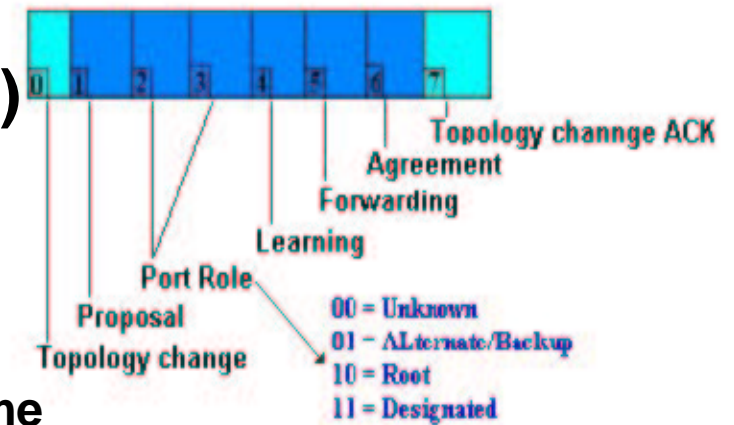
- Protokoll Version ist jetzt 2 (zuvor 0)
- Kein dedizierten TCN BPDU mehr
- BPDUs agieren als keepalives:

Alle Bridges senden BPDUs zu jeder hello time

Port Information ist ungültig nach max. 3 x hello time

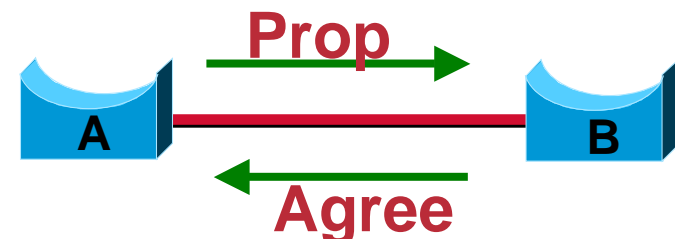
↔ 3 BPDUs gehen verloren

- 802.1D Bridges ignorieren RSTP BPDUs
- **Resultat: Schnellere Erkennung von Fehlerzuständen**

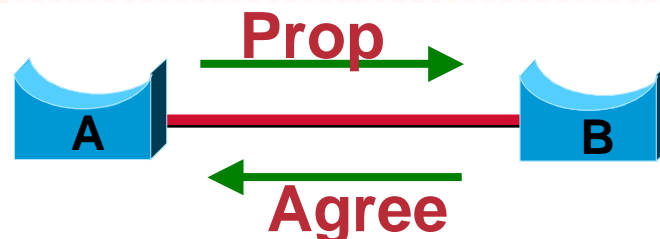


Agreement/Proposal

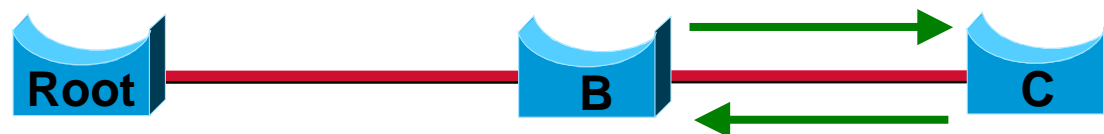
- Expliziter Handshake Mechanismus zwischen Bridges
- Wenn ein link up entdeckt wird, sendet die Bridge ein Proposal, um für dieses Segment die “designated Bridge” zu werden
- Die Antwort ist ein “**agreement**”, wenn die remote Bridge den Port auf dem sie das Proposal empfangen hat, als root Port erkennt.
- Sobald das “**agreement**“ empfangen wurde, geht der Port in den forwarding Modus



Agreement/Proposal



- A hat eine bessere Priorität als B
- A sendet ein Proposal zu B um “designated” zu werden
- B vergleicht die empfangene Priorität und antwortet mit einem “agreement” – B’s Port wird “Root Port”



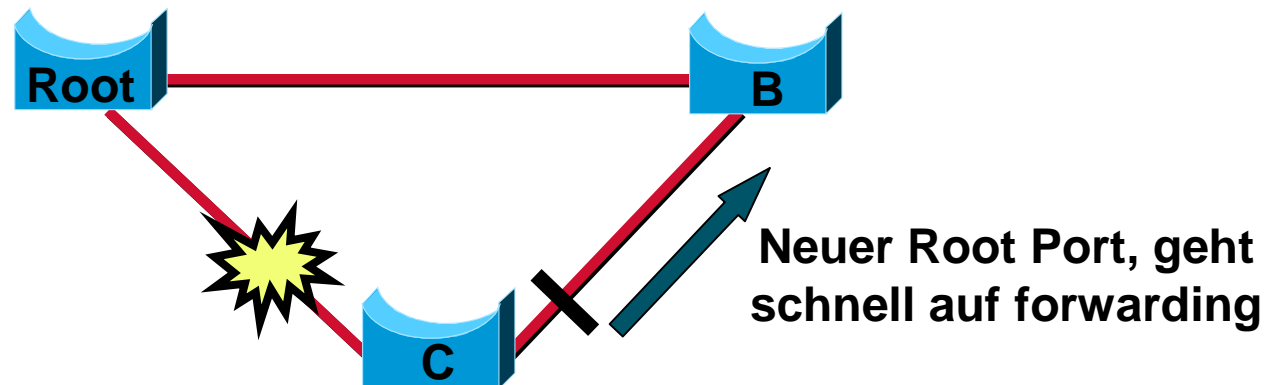
- Der gleiche Prozeß wird wiederholt wenn eine neue Bridge dazukommt.

BPDU Proposal/Agreement Merkmale

- **Proposal** kann nur von einem „**designated non-forwarding**“ Port geschickt werden.
- **Agreement** kann nur von einem „**root non-forwarding**“ port gesendet werden
- Auf PVST+ Ports werden PVST+ anstelle von RSTP BPDU gesendet, ebenso für TCN
- BPDUs werden auf jeden Link in beide Richtungen geschickt, im Gegensatz zu 802.1D

Integriertes UplinkFast

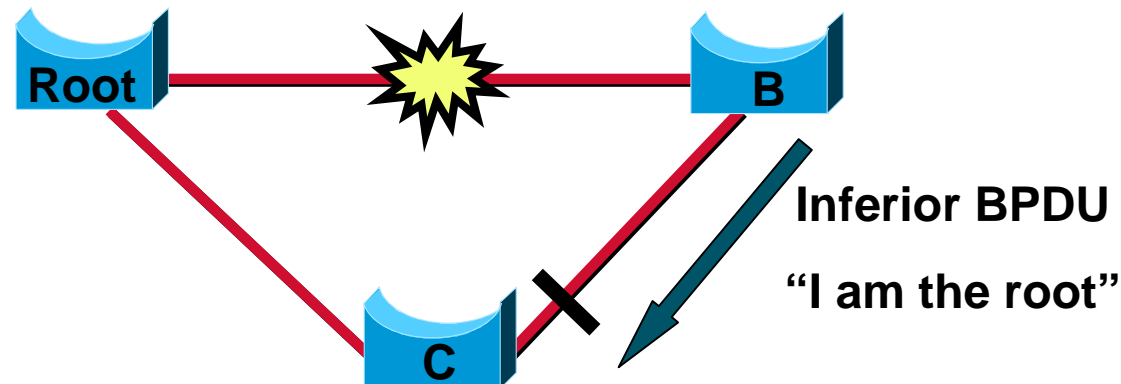
Cisco.com



- Wenn eine Unterbrechung auf dem Root Link entdeckt wird, schaltet C sofort auf den neuen Root Port um.
- Kein “Dummy Multicasts” werden ausgesandt, um die CAM Table von B zu erneuern – C generiert ein TCN, die die CAM Table von B löscht.
- Sehr ähnlich zu Ciscos UplinkFast Mechanismus.

Integriertes BackboneFast

Cisco.com



- **B verliert seinen Root Port und sendet ein BPDU zu C und gibt an, daß er Root ist.**
- **C geht sofort in “designated blocking” Status und sendet ein Proposal zu B um schnellstens wieder in “forwarding” zu kommen.**
- **B antwortet mit einem “Agreement” und C setzt seinen Port in “forwarding”.**

Schnelle Umschaltung

- Um eine schnelle Umschaltung zu gewährleisten gilt es, folgendes zu identifizieren:
 - Edge ports (**aktiviert durch Portfast**)
 - Point-to-point links (**abgeleitet vom duplex Modus - default**)
- **Achtung: wenn der Duplex Modus nicht korrekt gesetzt ist, gibt es keinen “Proposal/Agreement” Mechanismus zwischen den Bridges !**
- **Interoperabilität zwischen 802.1D und 802.1w ermöglicht nur eine sehr langsame Konvergenz – wenn möglich dringend vermeiden.**

802.1D Topology Change

- Wenn ein port in “forwarding oder blocking” geht, initiiert diese Bridge eine TCN BPDU zur Root.
- Die TCN BPDU werden bestätigt (ACKed) bis sie zur Root gelangen.
- Root setzt dann das TC flag in den gesendeten BPDUs für die Zeit (Fwd_Delay + Max_Age).
- Cam_Aging Timer auf den Bridges wird auf Fwd_Delay geändert.

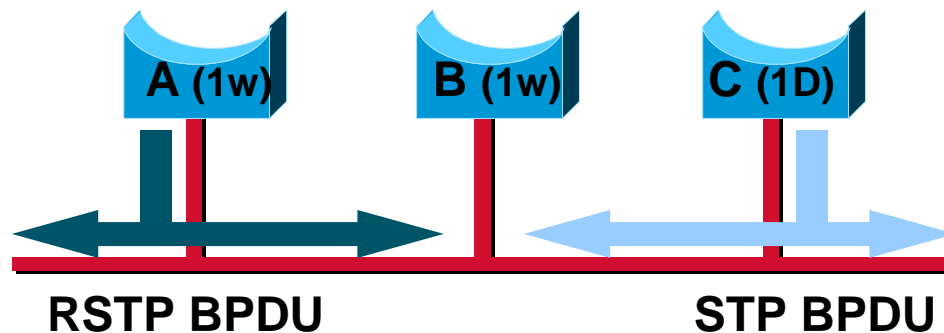
RSTP Topology Change

- TC wird nur auf non-edge Ports, die auf “forwarding” gehen, ausgesendet.
- Werden vom “Initiator” ausgeschickt (nicht von der Root) und von den Nachbarn weitergeleitet.
- TC bit wird für 2 x hello time gesetzt (TC While timer), es gibt kein ACK mehr ...
- Die CAM Table wird sofort gelöscht, außer für:
 - Port auf dem die TCN empfangen wurde
 - Edge ports
 - > Flooding, aber die Verbindungen werden sofort aufgebaut
 - > Keine Notwendigkeit für “Dummy Multicast”

Protokoll Migration (1)

Mix von 802.1D und 802.1w Bridges:

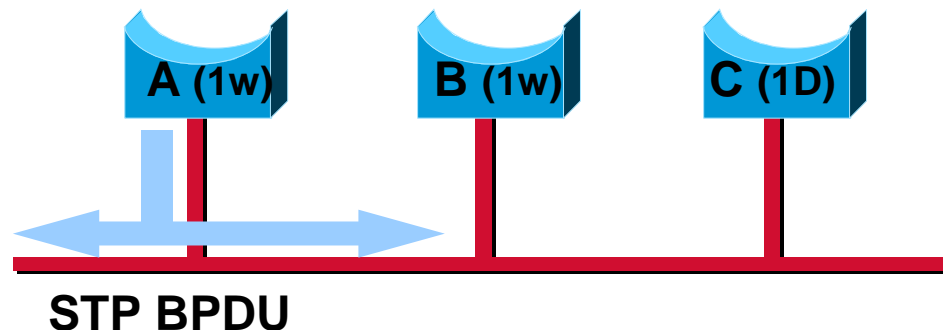
- 1D Bridges ignorieren 1w BPDUs (unterschiedliche BPDU Versionen) ⇒ 1D Bridges senden also immer 1D BPDUs



- Nach der “Migration Delay” Zeit (2 x Hello Time), sendet die RSTP Bridge 802.1D BPDU aus, wenn sie auf diesem Segment eine Legacy Bridge entdeckt (je Port).

Protokoll Migration(2)

- RSTP Bridge erkennt nicht, wenn die 802.1D Bridge entfernt wird.



- Wenn C entfernt wird, muß manuell eingegriffen werden, um die RSTP Vorteile zu nutzen (`set spanntree mst x/y redetect-neighbor`)
- 1D-1w Interoperabilität = Langsame Konvergenz:
⇒ wenn immer möglich, vermeiden.

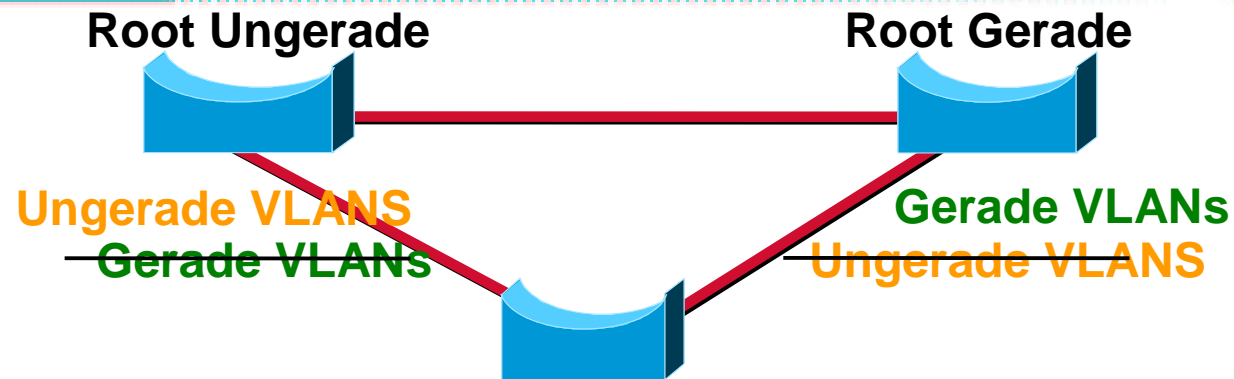
IEEE 802.1s Multiple Spanning Tree

Was ist IEEE 802.1s

- **802.1s - Multiple Spanning-Tree Spezifikation beschreibt eine Erweiterung des bestehenden 802.1D SPT**
- **Ist ähnlich zu Multiple Instance Spanning Tree (MISTP) Erweiterungen von Cisco Systems**
- **802.1s ist immer noch ein IEEE Draft**
 - z. Zt. in der Version 13 vom Juni 2002
- **802.1s verbessert die Skalierbarkeit von STP**
- **Ciscos Implementierung von 802.1s und 802.1w erlaubt z.T. keine Trennung zwischen den beiden Standards. Details: siehe weiter unten.**

Warum MST ? Vergleich zu PVST+

Cisco.com



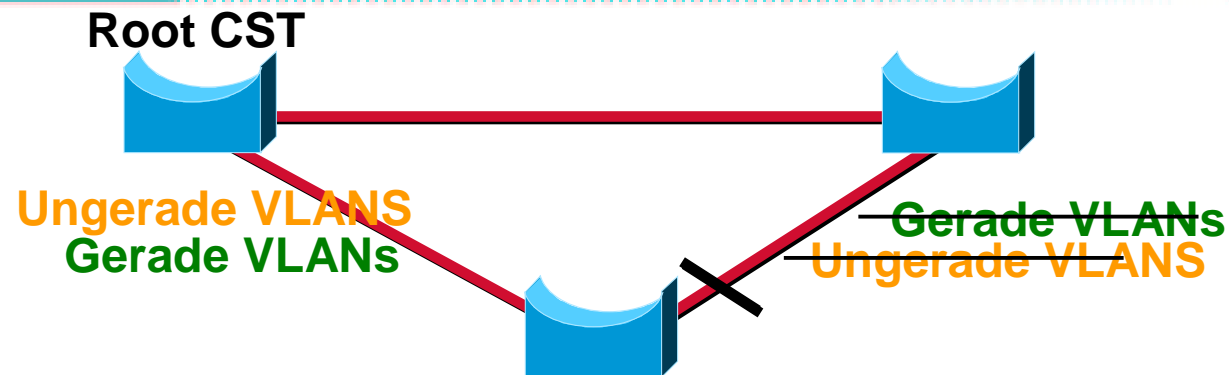
PVST+ = 1 STP pro VLAN

😊 **Flexibles load-balancing**

☹️ **CPU muß bis zu 4000 Instanzen des SPT für nur zwei verschiedene logische Topologien verwalten**

Warum MST ? Vergleich zu 802.1Q

Cisco.com

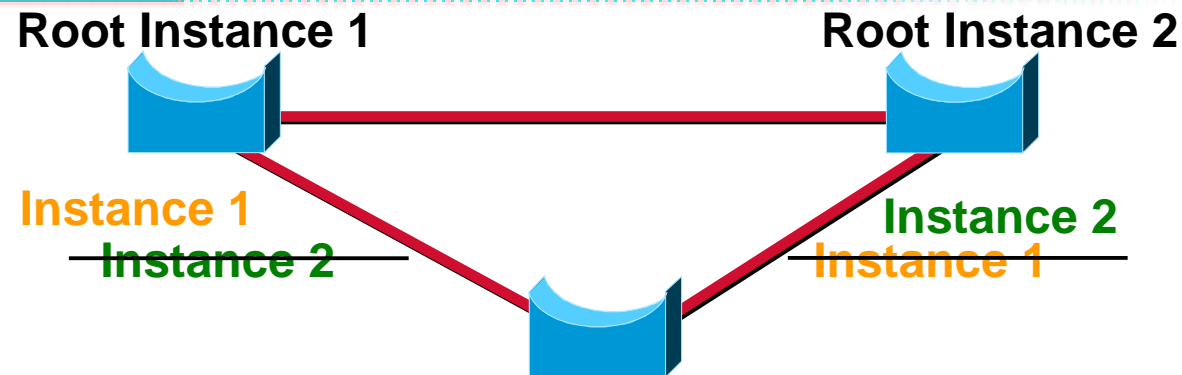


Alle Vlans laufen unter einer Spanning Tree Instanz

- ☺ **Kein Overhead für die CPU: nur eine Instanz**
- ☹ **Eine einzige logische Topologie für alle VLANs, geringe Flexibilität**

Der MST Ansatz

Cisco.com



- 😊 **Flexibles load-balancing**
- 😊 **Die CPU muß nur zwei Spanning Tree Instanzen für zwei Topologien verwalten**
- ☹️ **Neue und komplexe Interoperabilität mit dem klassischen 802.1D Protokol**

MST Regionen

- Eine MST Region umfaßt eine Gruppe von Switches mit der gleichen Konfiguration
- MST trennt VLANs und STP Instanzen (Instanzen ohne zugeordnete VLANs sind möglich)
- Bridges innerhalb einer Region müssen konsistente VLAN zu Instanz Zuordnungen haben

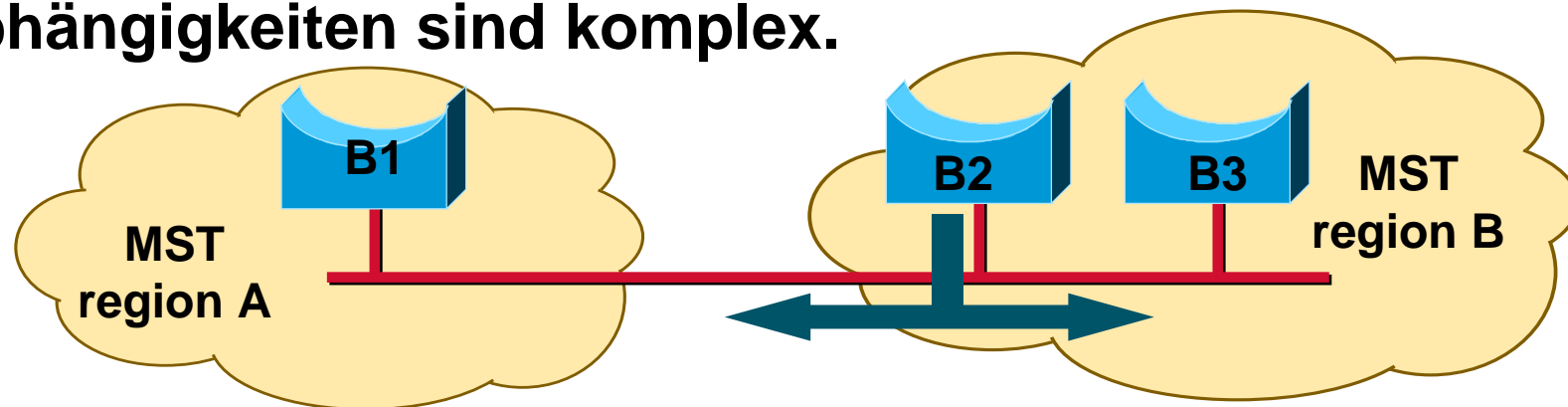
Die MST Konfiguration beinhaltet Werte für

- Einen Namen (32 bytes)
- Eine Revisions Nummer (16 bits)
- Eine Vlan zu Instanz Zuordnungstabelle

Und **muß** auf allen Switches gleich sein

MST Regions Grenzen

- Eine Zusammenfassung der Konfiguration wird innerhalb der MST BPDUs versandt – sind diese verschieden, bedeutet dies verschiedene Regionen !
- Ein Port ist an einer Regionsgrenze, wenn sich die designated Bridge auf seinem Segment in einer anderen Region befindet, oder wenn sich eine 802.1D Bridge auf dem Segment befindet.
- Abhängigkeiten sind komplex.



B2 designated => B1 boundary, B2 & B3 internal

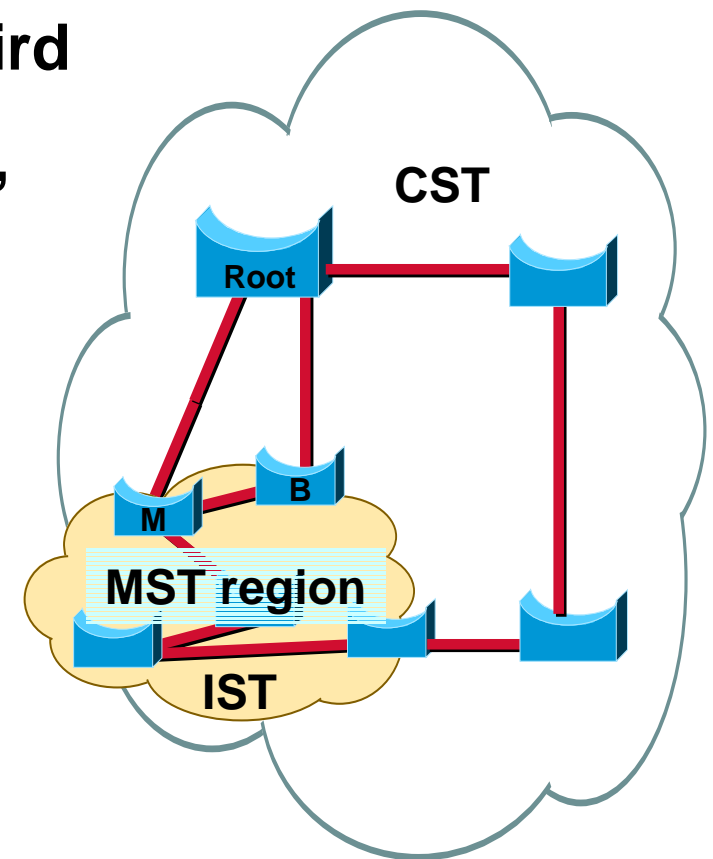
MST Instanzen

MST Bridges müssen mindestens zwei Instanzen verwalten:

- **Eine IST – der Internal Spanning Tree (Instanz 0) – existiert immer auf allen Ports**
- **Wenigstens eine MSTI - Multiple Spanning Tree Instance**
- **Cisco Implementation: 16 Instanzen (0-15)**
- **Die Terminologie entwickelt sich weiter
⇒ pre-standard**

Die IST Instanz

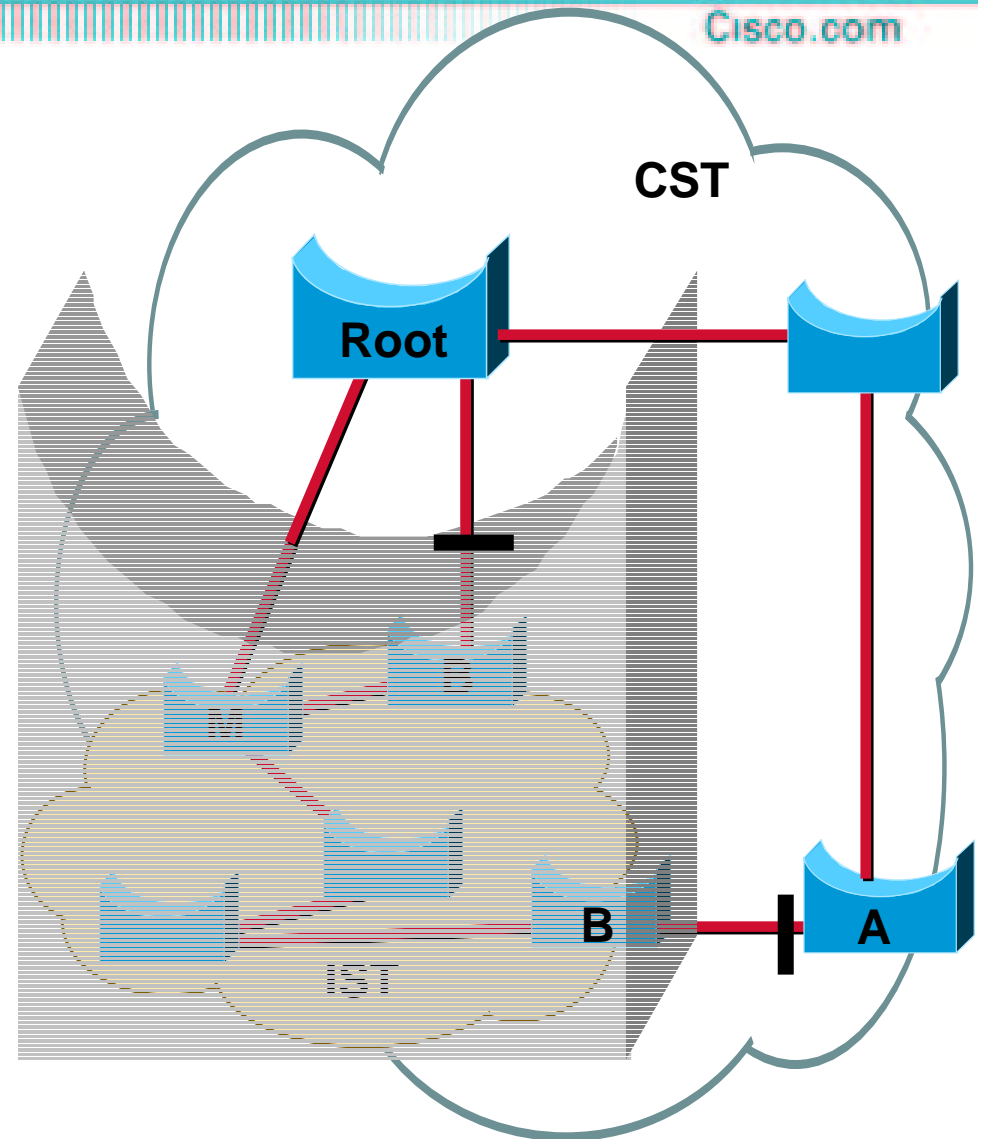
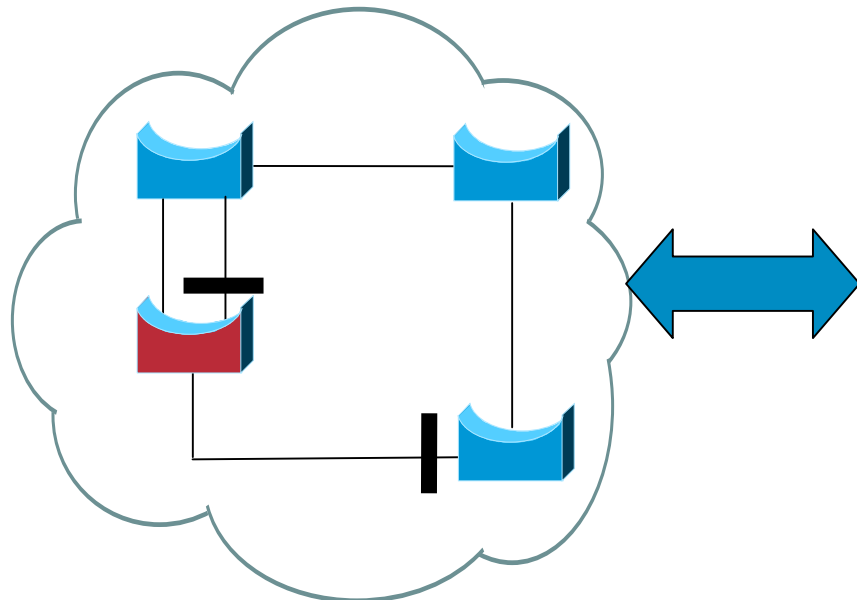
- In einem Standard 802.1Q Netzwerk gibt es nur eine STP Instanz, die CST (Common Spanning Tree) genannt wird
- Die IST Instanz ist eine RSTP Instanz, die den CST in die MST Region reinträgt
- Die MST Region sieht den Rest der Welt nur durch den Informationsaustausch zwischen IST/CST



MST Region: eine virtuelle Bridge (1)

Cisco.com

Die gesamte MST Region erscheint nach außen als eine virtuelle CST bridge



MST Region: eine virtuelle Bridge (2)

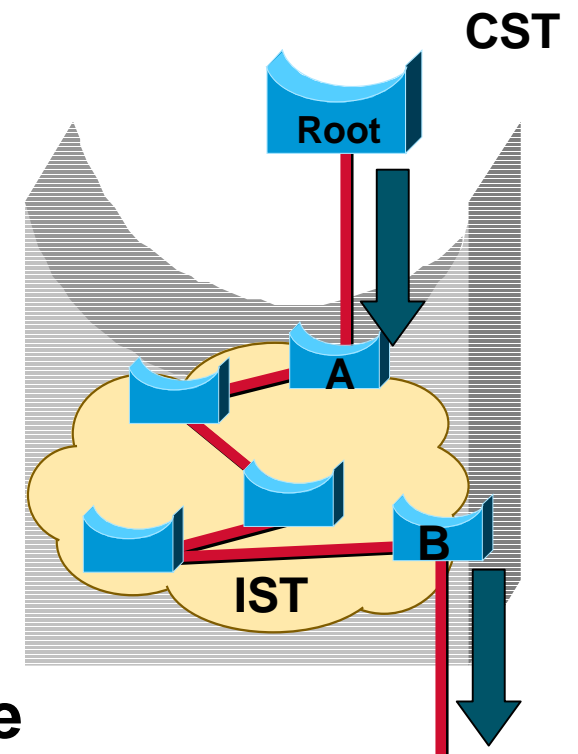
Cisco.com

Im BPDUs Frame von Bridge B werden:

- **Root Path Cost**
- und
- **Message Age**

so erhöht, als wenn nur eine Bridge passiert wurde.

Bridge B fügt seine Bridge ID in das Sender Bridge ID Feld ein (ähnlich wie tunneling)

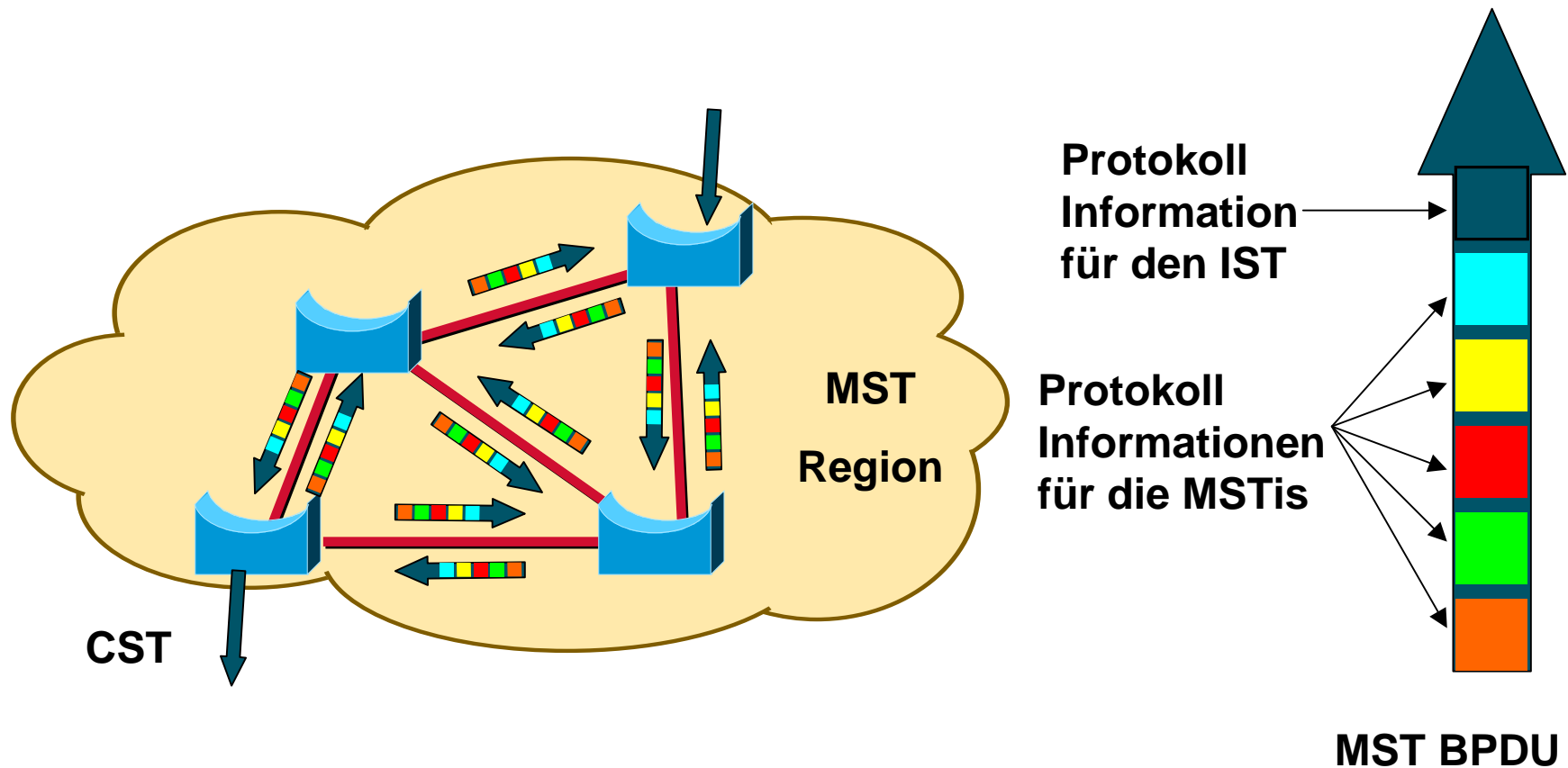


MSTIs (1)

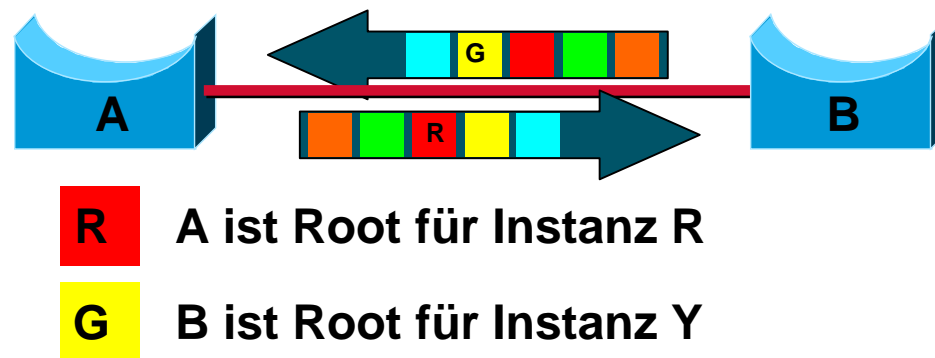
- **MSTIs sind RSTP Instanzen, die nur innerhalb der gleichen Region existieren.**
- **Sie kommunizieren nicht direkt mit der Welt außerhalb der Region.**
- **MST sendet nur eine BPDU für alle Instanzen – ein Record pro Instanz.**
- **Nur eine Instanz besitzt Timer relevante Parameter - die IST Instanz.**

MSTIs (2)

- MST BPDUs werden auf jedem Port gesendet



MSTIs (3)

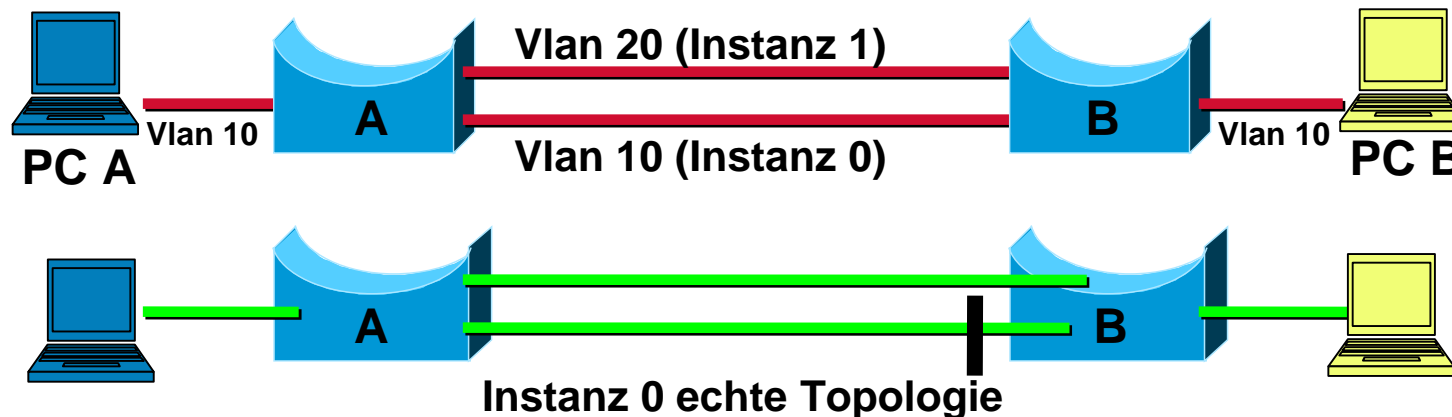


- BPDUs können auf einem Link in beide Richtungen gesendet werden – z.B.: B ist Root für Gelb aber A ist Root für Rot -> bi-dir BPDUs nötig.
- Ein Record für jede Instanz ist in den BPDUs enthalten.

MSTIs (4)

- Die IST Instanz existiert auf allen Ports, unabhängig vom Vlan mapping

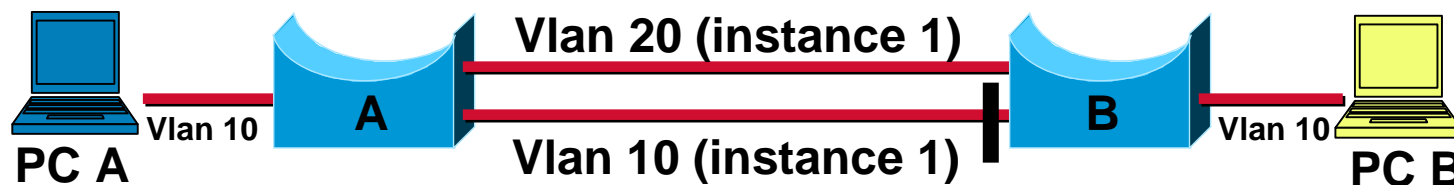
Keine Verbindung zwischen PC A & PC B, warum @#?!?



- Kein Mapping eines VLANs zur IST kann eine Lösung sein.

MSTIs (5)

- Das gleiche Problem kann auftreten mit VLANs, die zur gleichen Instanz gehören, die aber nicht auf dem gleichen Port erlaubt sind.

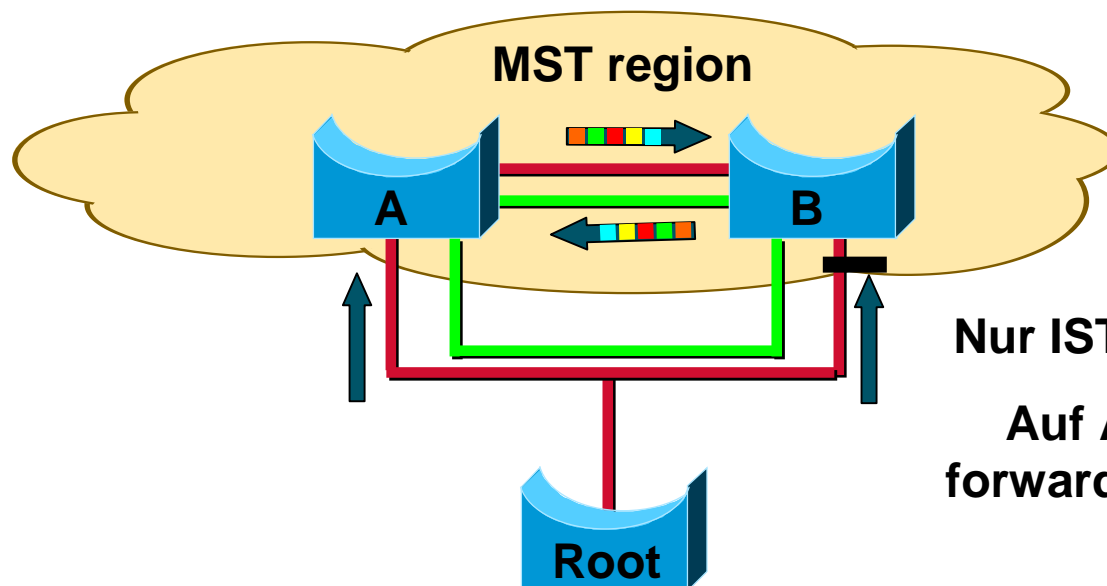


- Hier gilt: die VLANs nicht prunen.
- Oder: Alle VLANs in der gleichen Instanz sollten auf einem Trunk entweder alle disabled oder enabled sein.

MSTIs (6)

MSTIs an einer Boundary

- **MSTIs senden keine BPDUs auf Boundary Ports**
⇒ alle MSTIs sind “designated”
- **Nur die IST kommuniziert mit der CST**
- **Loop in der grünen Instanz ??**

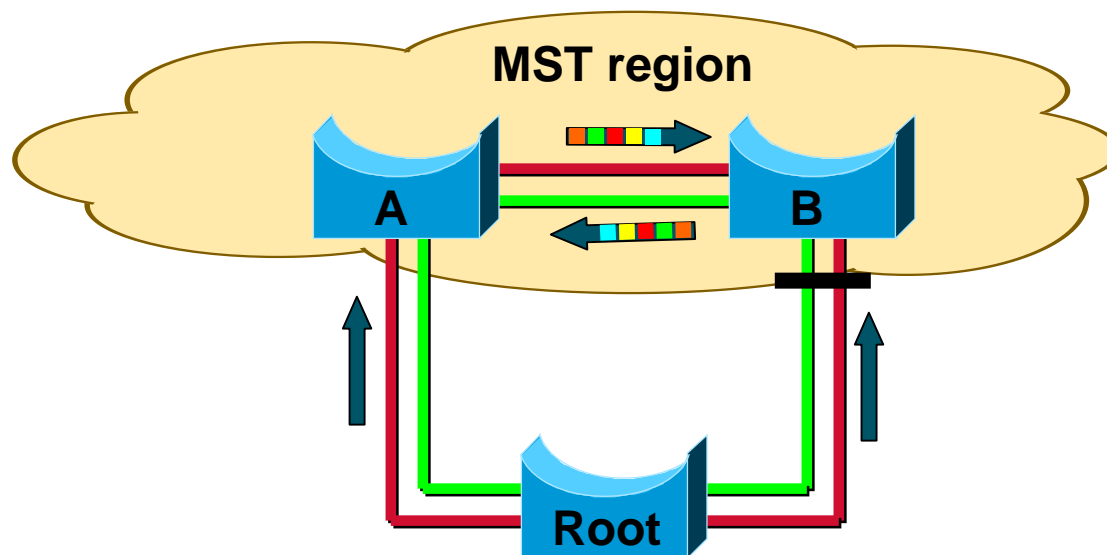


Nur IST kommuniziert mit der CST
Auf A&B ist die grüne Instanz forwarding (empfängt keine BPDUs)

MSTIs (7)

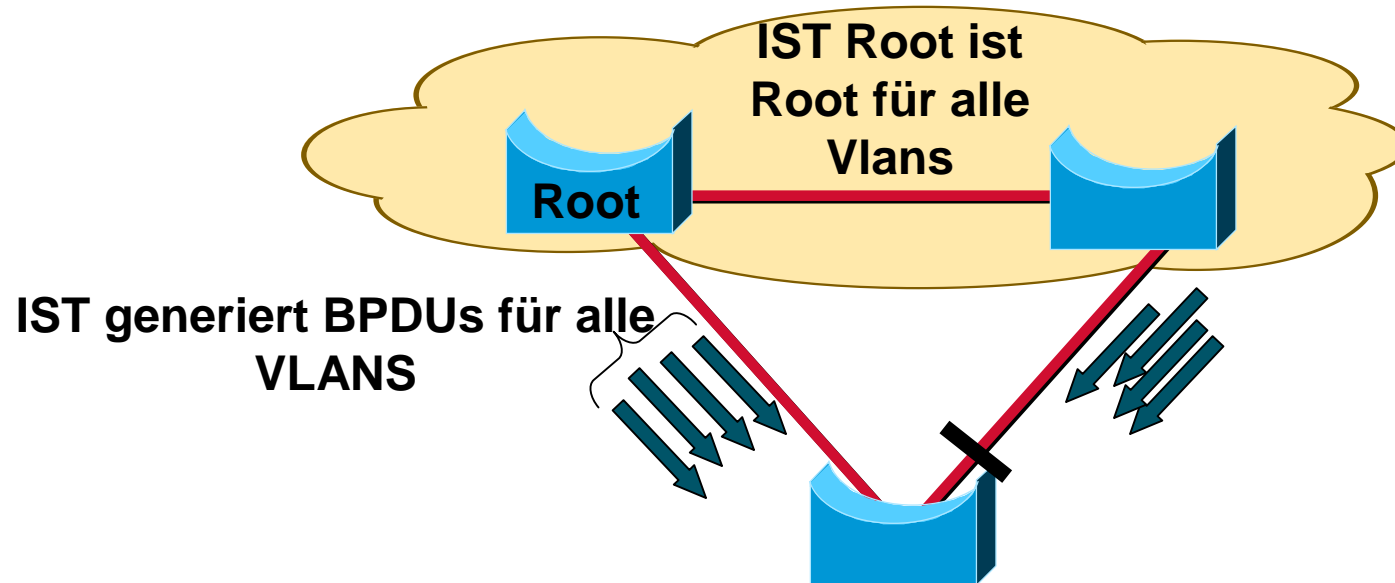
MSTIs an einer Boundary

- Ein Boundary Port einer MSTi folgt immer dem Status der IST.
- Resultat: auf B wird der Port der grünen Instanz auch geblockt, weil die IST blockt ⇒ der Loop wird verhindert.



Zusammenspiel mit PVST(+) Bridges

Cisco.com



- Duplizieren der IST BPDUs auf alle Vlans
- Empfohlene Konfiguration: die IST ist Root für alle PVST(+) Instanzen
- Load Balancing (Tuning der Kosten) und Uplinkfast ist weiterhin möglich

Implementation in der Cisco Catalyst Familie

Cisco Systems Catalyst Switch Serie

Cisco.com

Catalyst 6500



- Höchste Switching Kapazität
- HA— Stateful Failover
- 10 Gigabit Ethernet Interfaces
- Non-Blocking Gig Dichte: 142 Ports
- LAN/WAN Interfaces (bis OC 48)
 - L2-L7 Switching
 - Integrierte In-Line Power Option
 - Service Module
 - Red. Supervisor, Switch Fabric, PSU

Catalyst 4000/4500



- Hohe Switching Kapazität
- Höchste Portdichte für 10/100/1000
- Non-Blocking Gig Dichte: 32 Ports
- LAN/WAN Interfaces (T1/E1)
- L2-L4 Switching
- In-Line Power Option
- Hohe Dichte für 100FX
- Redundanz: PSU
- Kostengünstig und modular
- Flexible modulare Konfiguration

Catalyst 3550 / 2950



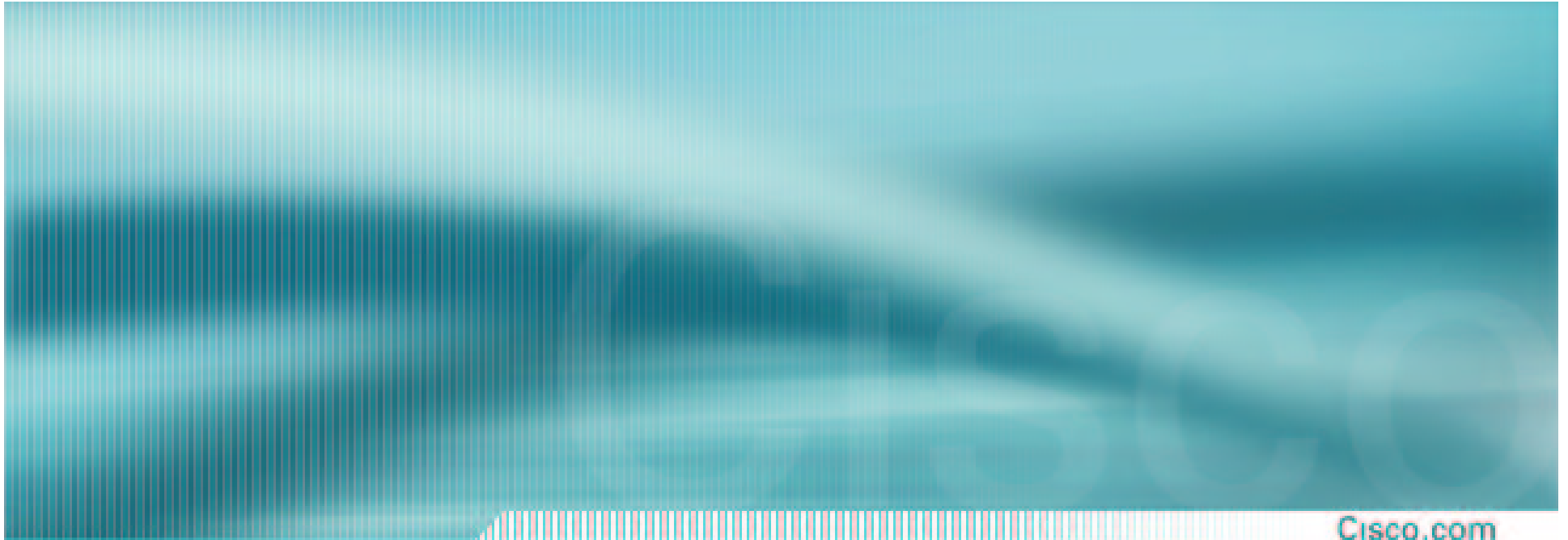
- Stackable
- Kompaktes Gehäuse
- Non-Blocking Gig Dichte: 12 Ports
- Einfaches Clustering Management
- Kostengünstiger Stack
- L2-L4 Switching

Preis / Leistung

Funktionalität

MST und RSTP Support Matrix

MST and RSTP Software availability table		Rapid-PVST+
Platform	Release	
Catalyst 6x00 SUP 1/2	Native IOS 12.1(11b)EX1 CatOS 7.1	Native IOS 12.1(13)E1 CatOS 7.5(1)
Catalyst 4000 SUP 3/4 Catalyst 4000 SUP 1/2	Native IOS 12.1(12c)EW CatOS 7.1	Ridgeway, Q2 CY03 CatOS 7.5(1)
Catalyst 3550 Catalyst 2950	12.1(9)EA1	12.1(13)EA1



Weiterführende Informationen zu den Themen

Weiterführende Informationen zu 802.1s/w

Cisco.com

Cisco TAC Tips und Beispiele

- <http://www.cisco.com/warp/public/473/123.html>
- <http://www.cisco.com/warp/customer/473/147.html>
- <http://www.cisco.com/warp/customer/473/146.html>

IEEE Informationen zu 802.1s/w

- <http://grouper.ieee.org/groups/802/1/pages/802.1s.html>
- <http://grouper.ieee.org/groups/802/1/pages/802.1w.html>



CISCO SYSTEMS



EMPOWERING THE
INTERNET GENERATIONSM