IT-Symposium 2005

3E03: Layer 2 Redundanz im LAN mit Spanning Tree

Robert Krause
Bkom Business Kommunikationssysteme GmbH
Kirchheim bei München

Übersicht

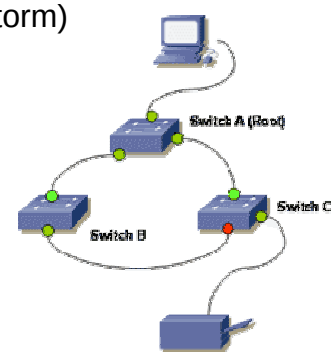
- Warum Spanning Tree?
- Die Verfahren:
  - IEEE Spanning Tree (IEEE 802.1D)
  - Per-VLAN Spanning Tree (PVST)
  - Rapid Spanning Tree (IEEE 802.1w)
  - Multi Instance Spanning Tree (IEEE 802.1s)
  - Ring Spanning Tree
- Interoperabilität - Mischen (Im)possible
- Layer 2 Redundanz-Konfigurationen
- Sicherheit
- Mögliche Konflikte mit Layer 3 Redundanz
- Fehlersuche

Layer 2 Redundanz

- Verbindung von Ethernet LAN Segmenten mit transparenten Bridges
  - Lernt, welche MAC Adresse an welchem Port
  - Bekannte MAC nur an zugehörigen Port
  - Broadcast, Multicast und unbekannte MAC ➔ an alle übrigen Ports (Flooding)
- Kein Routing (Layer 3), trotzdem gefordert:
  - Fehlertoleranz bei Ausfall von Verbindung (Link), Anschluss (Port), Netzwerkgerät (Bridge/Switch)
  - Mehrere mögliche Wege
  - Automatische Konfiguration

Loop

- PC schickt Broadcast Frame (z.B. ARP)
- Switch A dupliziert Broadcast an alle übrigen Ports
- Switch B und C tun das ebenfalls
- Jeder Switch empfängt die Broadcasts vom anderen...
- Maximale Auslastung (Broadcast Storm)
- Nichts geht mehr!



Spanning Tree

- Algorithmus Anfang der 80er Jahre bei DEC entwickelt von Radia Perlman
- Loop-freie, optimale Verbindung aller Segmente
- Automatische Konfiguration und Fehlertoleranz
- 1990: Standard IEEE 802.1D
  - Ersetzt DEC Spanning Tree
  - Gleicher Algorithmus
  - Andere BPDU
- Revisionen 1998 und 2004

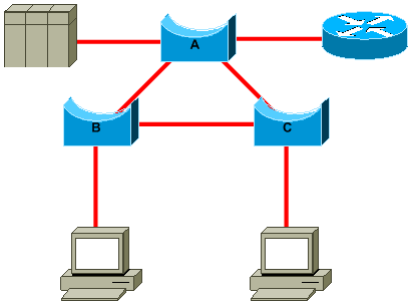
BPDU

- Bridge Protocol Data Unit
- Gesendet an Multicast Adresse 01-80-C2-00-00-00
- Topology Change Notification (TCN) nur 4 Bytes

```
protocol id:   0000 IEEE 802.1d
version id:    00
bpdu type:     00 config, 80 tcn
bit field:       1 byte
  1   : topology change flag
  2-7 : unused   0
  8   : topology change ack
root priority    2 bytes
root id:         6 bytes
root path cost:  4 bytes
bridge priority: 2 bytes
bridge id:       6 bytes
port id:         2 bytes
message age:     2 bytes in 1/256 secs
max age:         2 bytes in 1/256 secs
hello time:      2 bytes in 1/256 secs
forward delay:   2 bytes in 1/256 secs
```

Root

- Bridge Priority und Bridge ID (MAC) zusammen
- **Niedrigster** Wert wird Root
- Default Priority ist meist 32768 (0x8000) – aber nicht immer: aufpassen!
- Manche Hersteller geben z.B. (Priority / 4096) an !



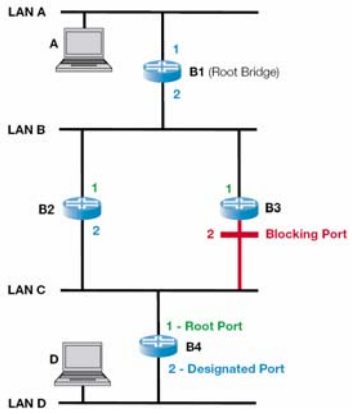
Path Cost

- Abhängig von Link Bandbreite
- Nach 802.1D: 1000/BW[Mbps]
- Heute laut nebenstehender Tabelle
- Default Werte nicht einheitlich!

Bandbreite	STP Cost
4 Mbps	250
10 Mbps	100
16 Mbps	62
45 Mbps	39
100 Mbps	19
155 Mbps	14
622 Mbps	6
1 Gbps	4
10 Gbps	2

Port Role - Port State

- Port Role:
  - Root Port ➔ Root Bridge
  - Designated Port ➔ Folgendes LAN Segment
- Port State:
  - Disabled
  - Blocking
  - Listening (BPDU)
  - Learning (MAC)
  - Forwarding



Rapid Spanning Tree

- 2001: IEEE 802.1w
- Schnelles Umschalten durch Vorbereitung
- Handshake Protokoll
- Zusätzliche Port Roles:
  - Alternate Port – Alternative zum Root Port
  - Backup Port – Alternative zum Designated Port
- Port State: Disabled, Discarding, Learning, Forwarding
- Link Type: Point-to-Point, Edge, Shared
- Konvergiert in 2-3 Sekunden bei Point-to-Point Links
- Kein Vorteil bei Shared LAN

VLAN Problematik

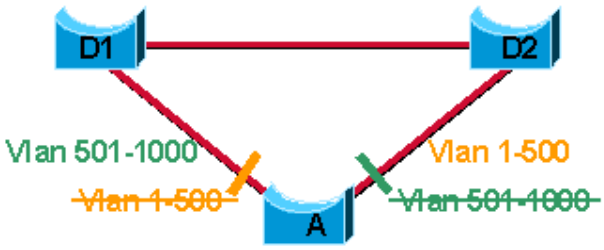
- Frame Tagging nach IEEE 802.1Q
- Mehrere VLANs auf einem physikalischen Link
- Spanning Tree ist Link basiert
- Kein Tagging der BPDUs
- Spanning Tree „sieht“ nicht, dass Ports in unterschiedlichen VLANs sind
- Folge: unerwünschtes Blocken von VLANs

Per-VLAN Spanning Tree (PVST)

- Von Cisco eingeführtes, proprietäres Verfahren
- Separater Spanning Tree Prozess pro VLAN
- Ermöglicht gleichzeitige Nutzung von redundanten Wegen durch unterschiedliche VLANs
- Neuerdings auch Rapid PVST
- Nachteile bei großer Zahl von VLANs:
  - BPDU Traffic pro VLAN
  - Hohe CPU Last im Switch
  - Verschärft bei Ausfall eines VLAN Trunks

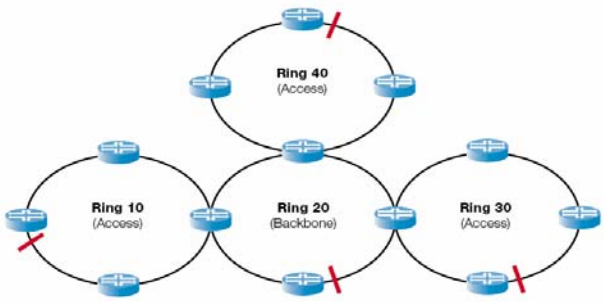
Multiple Spanning Tree (MST)

- 2002: IEEE 802.1s (als Zusatz zu 802.1Q)
- Zuordnung von VLANs zu Spanning Tree Prozessen (Gruppenbildung)
- Erweiterte BPDU (Region Name, Revision Number, Element Table)



Ring Spanning Tree

- Abwandlung von Rapid Spanning Tree
- Separater STP Prozess pro Ring (Ring ID)
- Nur Point-to-Point Links
- Kein Standard

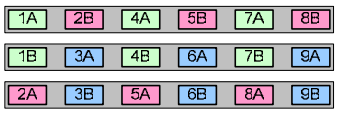


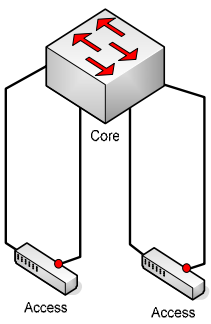
Mischen (Im)possible

- DEC Spanning Tree: nicht kompatibel
- 802.1w RSTP: Wenn 802.1D BPDU erkannt wird, schaltet Port in 1D Modus
  → interoperabel, aber Verlust der Vorteile
- PVST: Herstellerspezifisch - 1D und 1w werden nur in VLAN 1 (untagged) ausgeführt
  → Vorsicht bei Verbindungen zwischen Herstellern
- 802.1s MSTP: Zusammenspiel von MST und SST Regions im Standard geregelt
- Ring STP: Herstellerspezifisch

Redundante Uplinks

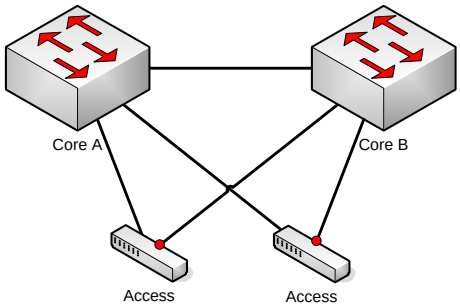
- Getrennte Kabelwege
- Core Switch redundant ausgelegt
- Uplinks auf verschiedene Linecards verteilt
- Im Beispiel sind 1,3,4,6,7,9 A aber 2, 5, 8 B aktiv (niedrigere Port ID)





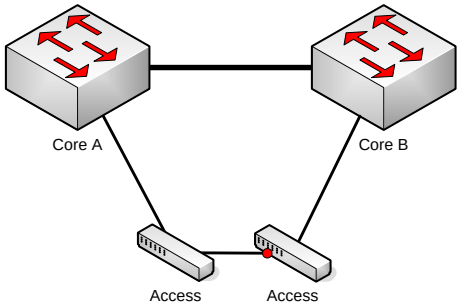
Redundante Core Switches

- Getrennte Kabelwege
- Querverbindung der Core Switches sinnvoll (sonst bei bestimmten Fehlern Traffic über Access Switch)



Dreiecks-Schaltung

- Spart Ports im Core bei voller Redundanz
- Geringere Performance im Fehlerfall
- Traffic auf Querverbindung beachten
- Geht natürlich auch mit nur einem Core Switch



Endgeräte und Spanning Tree

- Grundsätzlich ist alles Endgerät, was nicht auf Layer 2 forwarded
- Klassische Situation:
  - PC wird angesteckt → Link am Switch
  - Switch beginnt mit ST (...Listening, Learning...)
  - PC fährt hoch, schickt DHCP Request
  - Switch ist noch nicht in Forwarding → Mülleimer
  - PC hat keine Netzwerkverbindung
- Abhilfe: Portfast, Fast Learning, Edge Port (bei RSTP)
- Port geht sofort in Forwarding
- Gefahr von temporären Loops

Sicherheit

- Spanning Tree hat keine Sicherheitsmechanismen
- Jeder kann „günstige“ BPDUs schicken
  - Denial of Service
  - Man in the Middle
- BPDU Filter → Gefahr von Loops
- BPDU Guard:
  Port wird deaktiviert, sobald eine BPDU empfangen wird
- Root Guard:
  Port kann niemals Root Port werden
- Port Security (IEEE 802.1X)

ST und Layer 3 Redundanz

- HSRP, VRRP, XRRP
- Active/Standby über Hello Messages
- Übernahme von IP- **und** MAC-Adresse
- Wenn sich die Router nicht „sehen“ (Hello Timeout) werden beide Active → Chaos
- Auf Timeout Werte achten
- Wenn möglich RSTP verwenden
- Router sind **Endgeräte** aus Layer 2 Sicht
  - Also: Edge Port, Portfast, Fast Learning, etc.

Fehlersuche

- Bei Loop (Broadcast Storm) hilft nur systematisches Auftrennen der Links
- Unidirectional Link (gebrochene Faser)?
- Full-/Half-Duplex Mismatch?
- Wer ist Root? (Root ID)
- Verfolgung des Root Ports von Switch zu Switch
- Spanning Tree Port Status (Blocking)?
- In welchem VLAN?

Links

- http://www.cisco.com
  Technical Support & Documentation → Technology Support → LAN Switching → Spanning Tree Protocol
- http://www.riverstonenet.com/solutions/spanning_tree.shtml
  Spanning Tree: Death is Not an Option
- http://funkschau.de/heftarchiv/pdf/2003/fs1603/fs0316055.pdf
  Funkschau Heft 15-16/2003 Grundlagen: Spanning Tree
- http://www.dista.de/netstp.htm
  U.a.: „Algorhyme“ und Interview mit Radia Perlman

Fragen ???

Bkom Business Kommunikationssysteme GmbH
Schwabener Weg 1
85551 Kirchheim b. München

http://www.bkom-muc.de
mailto:robert.krause@bkom-muc.de
Telefon: (089) 159 23 65-0