



Traffic Engineering für neue Qualitätsinfrastrukturen: Qualitätszusicherung in IP-Netzen



Dipl.-Ing. Kai-Oliver Detken
Senior IT Consultant, <http://www.detken.net>
DECOIT e.K., <http://www.decoit.de>
Düsseldorf, Januar 2002

Inhalte

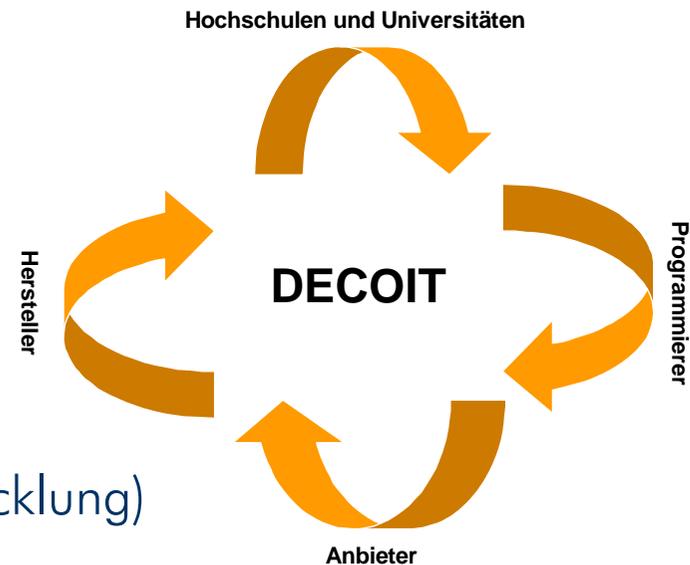
- ◆ Schaffung einer Dienstqualität mittels IP (IntServ, DiffServ)
- ◆ Einsatz von Traffic Management (TE) für Provider-Netze mittels MPLS
- ◆ Lösungen und Einsatzgebiete
- ◆ Messungen
- ◆ Ergebnis und Fazit

Kurzvorstellung

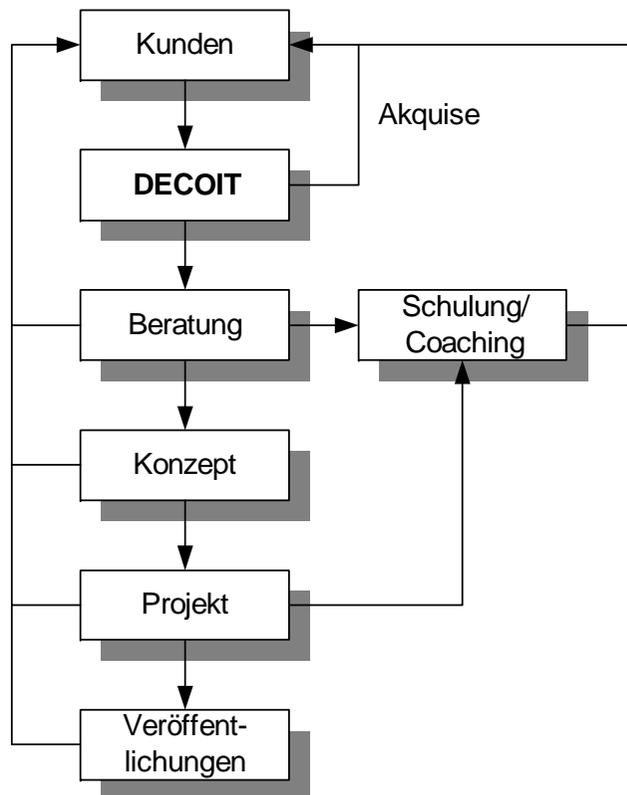


Dienstleistungsportfolio der DECOIT

- ◆ Consulting
 - Netzkonzepte
 - Sicherheitskonzepte
 - Sicherheitsüberprüfung
 - Strategiekonzepte
 - Softwarekonzepte
- ◆ Produkt- und Herstellertests
- ◆ Förderprojekte (Forschung und Entwicklung)
- ◆ Schulungen und Workshops
- ◆ Software-Entwicklung für intelligente IP-Lösungen
- ◆ Qualitätssicherung
- ◆ Projektmanagement



Zusammenspiel



- ◆ Bündeln der Informationen
- ◆ Wissensvorsprung
- ◆ Integration neuer Technologien
- ◆ Kosten-/Nutzenanalyse
- ◆ Erkennen von Visionen, Trends und Prognosen
- ◆ Abschätzung und Evaluierung der neuesten Technologien
- ◆ Zusammenarbeit mit den führenden Herstellern
- ◆ Erkennen von Stärken/Schwächen
- ◆ Ganzheitliche Lösungen

Schaffung einer Dienstqualität



Dienstgüte im Internet (1)

- ◆ Wenn das Internet nicht als einzelnes Netz, sondern aus einem Verbund gleichberechtigter Netze betrachtet wird, müssen nicht nur technische Vorkehrungen beim Übergang in andere Teilnetze getroffen werden, sondern auch gegenseitige Abkommen bezüglich der Einhaltung der Dienstgüte im eigenen Netz
- ◆ Hinzu kommt, dass die Tarifierung der angebotenen QoS-Dienste und die Authentifizierung übergreifend sichergestellt werden muss
- ◆ Eine Authentifizierung muss im allgemeinen sowohl am Netzeingang als auch an den Netzübergängen stattfinden
- ◆ Dies ist aber im Internet ein offenes Thema!

Dienstgüte im Internet (2)

- ◆ Als wichtige Vermittlungseinheiten im Internet agieren Router
- ◆ Verkehrsströme einzelner Rechner (Hosts) kommen am Eingang des Routers in eine Warteschlange
- ◆ Der Router analysiert den Paketkopf und entscheidet anhand dessen, auf welchen Ausgang das Paket weitergereicht wird
- ◆ In der Regel werden am Eingangspunkt des Netzes mehrere kleinere Datenströme zu einem Großen zusammengefasst
- ◆ Da die einzelnen Hosts im allgemeinen unabhängig voneinander das Netz benutzen, wird auch die Auslastung der Ressourcen statistisch verteilt
- ◆ Folgende Parameter können die Verkehrscharakteristik beeinflussen:
 - Die (momentane) Menge der ankommenden Datenpakete
 - Verarbeitungsgeschwindigkeit des Routers
 - Die verfügbare Bandbreite zwischen den Routern im Backbone

Dienstgüte im Internet (3)

- ◆ IP ist ein Protokoll der Vermittlungsschicht (Schicht 3) beschreibt nicht den Zugriff auf ein physikalisches Medium
- ◆ Dies ermöglicht zwar, dass das Internet Protocol (IP) unabhängig von der Übertragungstechnologie arbeiten kann, Mechanismen zwischen den Layer-2-Protokollen und IP müssen aber vorhanden sein, um beispielsweise QoS, Priorisierung oder Ressourcenreservierung ausnutzen zu können
- ◆ Dafür sind aber nicht alle Layer-2-Protokolle unbedingt geeignet

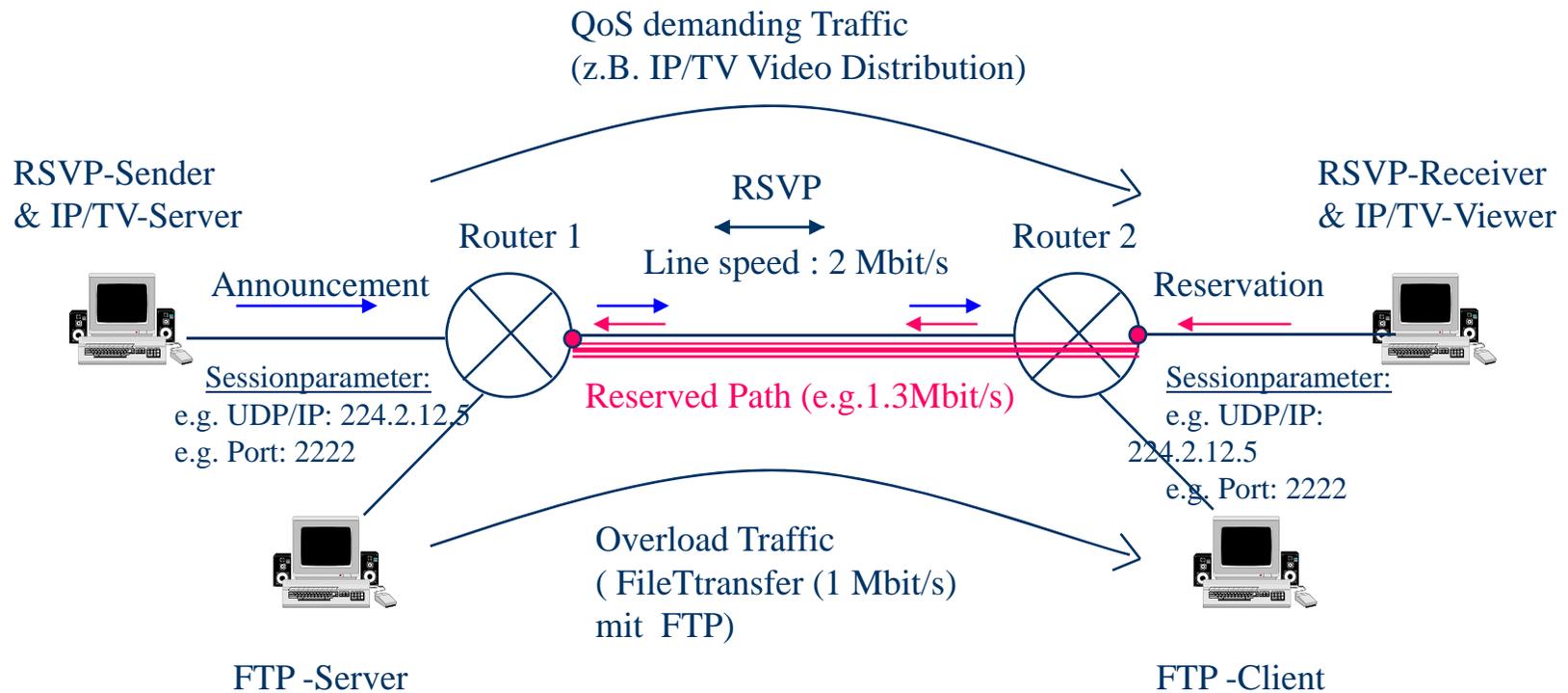
Integrated Service

- ◆ Best-Effort Service
- ◆ Controlled Load Service
 - Performance so hoch wie möglich
 - keine qualitativen Zusicherungen
- ◆ Guaranteed Service
 - Durchsatz und Verzögerung
 - Jitter oder durchschnittliche Verzögerung werden nicht berücksichtigt
 - Jedes Element in einem Übertragungspfad muss Delay begrenzen können
 - Ist nicht in jeder Umgebung einsetzbar (z.B. Ethernet als Shared Media)
- ◆ RSVP wird als Signalisierungsprotokoll verwendet (Skalierungsprobleme)

RSVP-Funktionalität bei IntServ

- ◆ **Merging:** Der abschnittsweise festgelegte QoS ermöglicht Verschmelzungen von Datenströmen mehrerer Sender
- ◆ **Unidirektionale Ressourcenetablierung:** Ressourcen werden unidirektional angefordert; Up- und Downstream sind bezüglich ihrer Dienstgüte voneinander entkoppelt
- ◆ **Regelmäßige Bestätigung der Ressourcen:** Soft-State muss regelmäßig bestätigt werden
- ◆ **Empfänger initiiert die Ressourcen:** die Reservierung der Ressourcen wird durch den Empfänger eingeleitet
- ◆ **Transparenz:** Strecken, die nicht RSVP-fähig sind, können aufgrund der Transparenz des Protokolls einbezogen werden
- ◆ **Routing:** Ausnutzung vorhandener Routing-Mechanismen (z.B. OSPF, BGP)
- ◆ **Protokollunterstützung:** IPv4 und IPv6 sowie Multicast-Pakete

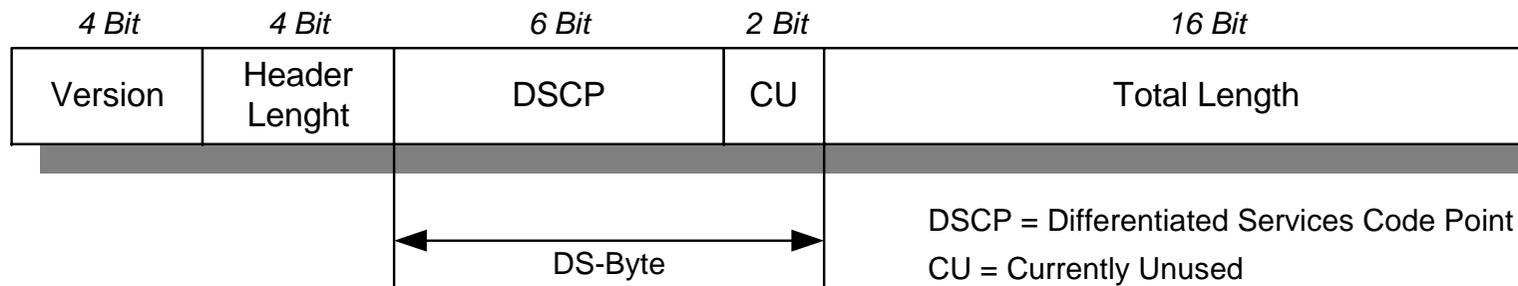
RSVP-Szenario



Differentiated Services (DiffServ)

- ◆ Endgeräte-bezogene Session-Behandlung ist zu aufwendig
- ◆ Gleichartige Sessions werden zusammengefaßt (Skalierbarkeit)
- ◆ Definition von drei Servicelevels (RFC-2475/-2474/-2597)
 - Expedited Forwarding (EF): Premium-Klasse; max. Datenrate und geringe Verzögerung (DSCP=101100)
 - Assured Forwarding (AF): Unterteilung in weitere vier Klassen, die wiederum in drei Verlustraten unterscheiden: erwartete Datenrate, keine Garantien
 - Best-effort: keine Zusicherung (DSCP = 000000)
- ◆ TOS-Feld des IP-Headers enthält DS-Byte
- ◆ Wird die zugesicherte Qualität durch eine Anwendung verletzt, so wird der Mehranteil über der PCR nicht mehr garantiert

Differentiated Services



- ◆ Andere Einteilung des TOS-Feldes im IP-Header
- ◆ Per-Hop Behaviour (PHB):
 - Drop Threshold
 - Pufferzuordnung
 - Service Priorität
 - Service Rate
- ◆ Mehrere DSCP können auf ein PHB abgebildet wird
- ◆ PHB + Netzdimensionierung = Qualität

IP-QoS durch DiffServ

- ◆ Nutzung des TOS-Feldes des IP-Headers
- ◆ Differentierte Behandlung von Datenströmen nach Klassifizierung an Edge-Routern (Ingress und Egress Router)
- ◆ Dienstgüteklassen: Premium bis Olympic
- ◆ Einhaltung von Service Level Agreements möglich
- ◆ Einsatz auch in Kombination mit IntServ/RSVP
- ◆ Weiterführung des Ansatzes von IPv6 (Traffic Class Byte)
- ◆ Durch DSCP-Feld (6 Bit) sind theoretisch 64 Klassen ermöglicht worden (!)

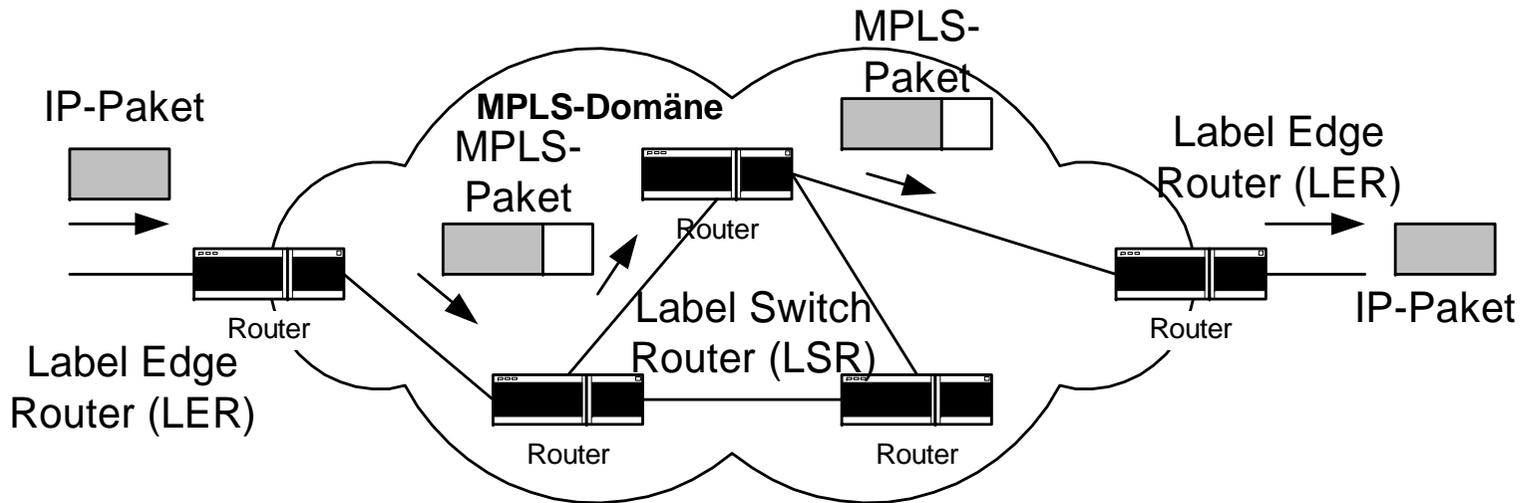
Einsatz von Traffic Management



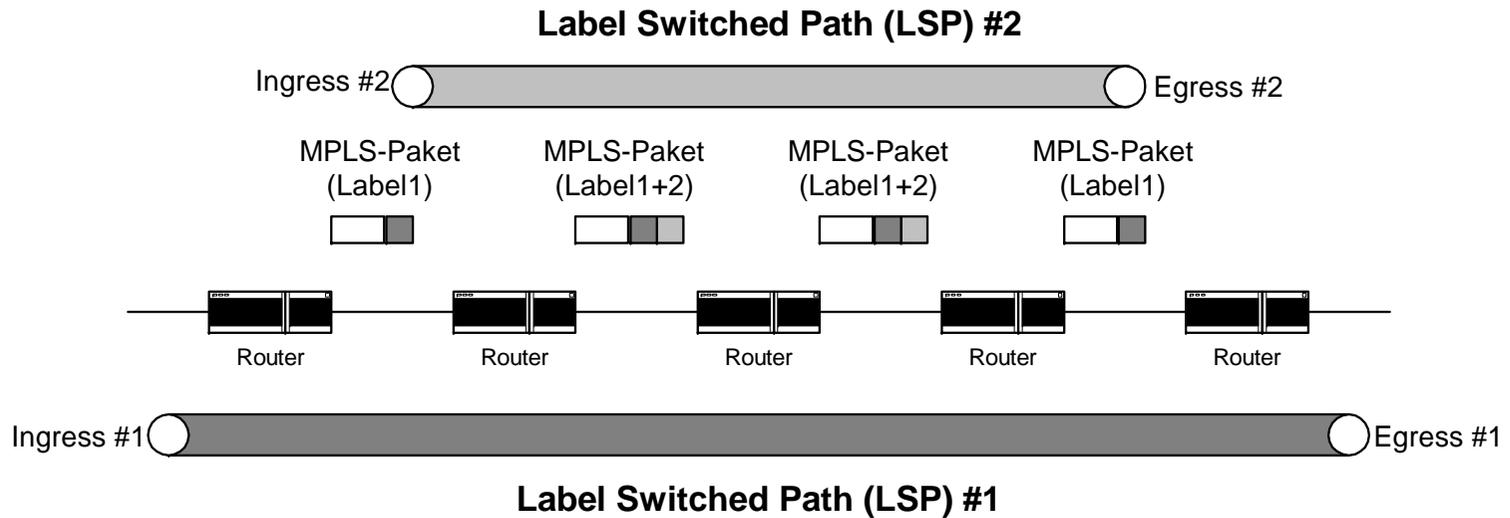
Label Switching Ansatz (1)

1. **Layer-3-Switching:** bessere Performance durch Switching, nachdem Routing-Funktionen übertragen wurden (Label Switching)
2. **Interworking:** Durch die Koexistenz von IP und ATM im Internet ist ein Interworking notwendig. Die komplexen Mechanismen, die entwickelt wurden, um IP und ATM anzupassen bzw. zu integrieren (z.B. CLIP), weisen alle Schwachpunkte auf
3. **Skalierbarkeit:** Die Skalierbarkeit beschreibt das Aufbrauchen von Ressourcen, wenn ein Netzwerk wächst. Da durch steigende Teilnehmerzahlen die Netze immer größer werden, ist es daher unbedingt erforderlich, die Skalierbarkeit zu optimieren
4. **Traffic Engineering:** In konventionellen Routern wird ausschließlich Destination-Based Routing betrieben. Zusätzliche Routing-Funktionen, wie das Explicit Routing mit dem Ziel des Traffic Engineering müssen eingeführt werden

Label Switching Ansatz (2)

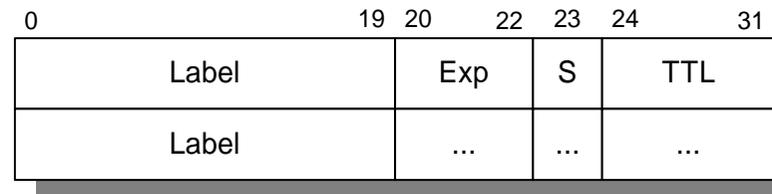


Label Switched Path (LSP)



Generic MPLS Encapsulation

- ◆ **Label-Feld:** trägt das eigentliche Label für die Pakete einer bestimmten FEC (Forwarding Equivalence Class)
- ◆ **Exp-Feld:** Dieses Feld beinhaltet drei Bits und ist reserviert.
- ◆ **S-Feld:** Dieses Feld ist 1 Bit groß und kennzeichnet für den Wert 1 den „Bottom-of-Stack“
- ◆ **TTL-Feld:** Das Time-to-Live-Feld besteht aus 8 Bits und wird zur Verhinderung von Schleifenbildungen eingesetzt



Exp = Experimental Use
 S = Bottom-of-Stack (=1 im letzten Eintrag)
 TTL = Time-to-Live

Labeltransport über undefinierte Schicht-2-Felder (ATM – AAL5)

- ◆ **SVP Encoding:** schreibt das Top-Label des Label-Stacks in das VPI-Feld. Falls im Labelstack weitere Labels vorhanden sind, wird der zweite Label-Wert in das VCI-Feld geschrieben
- ◆ **SVP Multipoint Encoding:** schreibt das Top-Label des Label-Stacks in das VPI-Feld. Falls im Label-Stack weitere Labels vorhanden sind, wird der zweite Label-Wert in einen Teil des VCI-Felds geschrieben. Der andere Teil des VCI-Felds wird für die Kennzeichnung des Ingress für diesen LSP verwendet. Auf diese Weise können Zellen einem Flow zugeordnet werden, wodurch es möglich ist, Label Merging zu betreiben

Lösungen und Einsatzgebiete



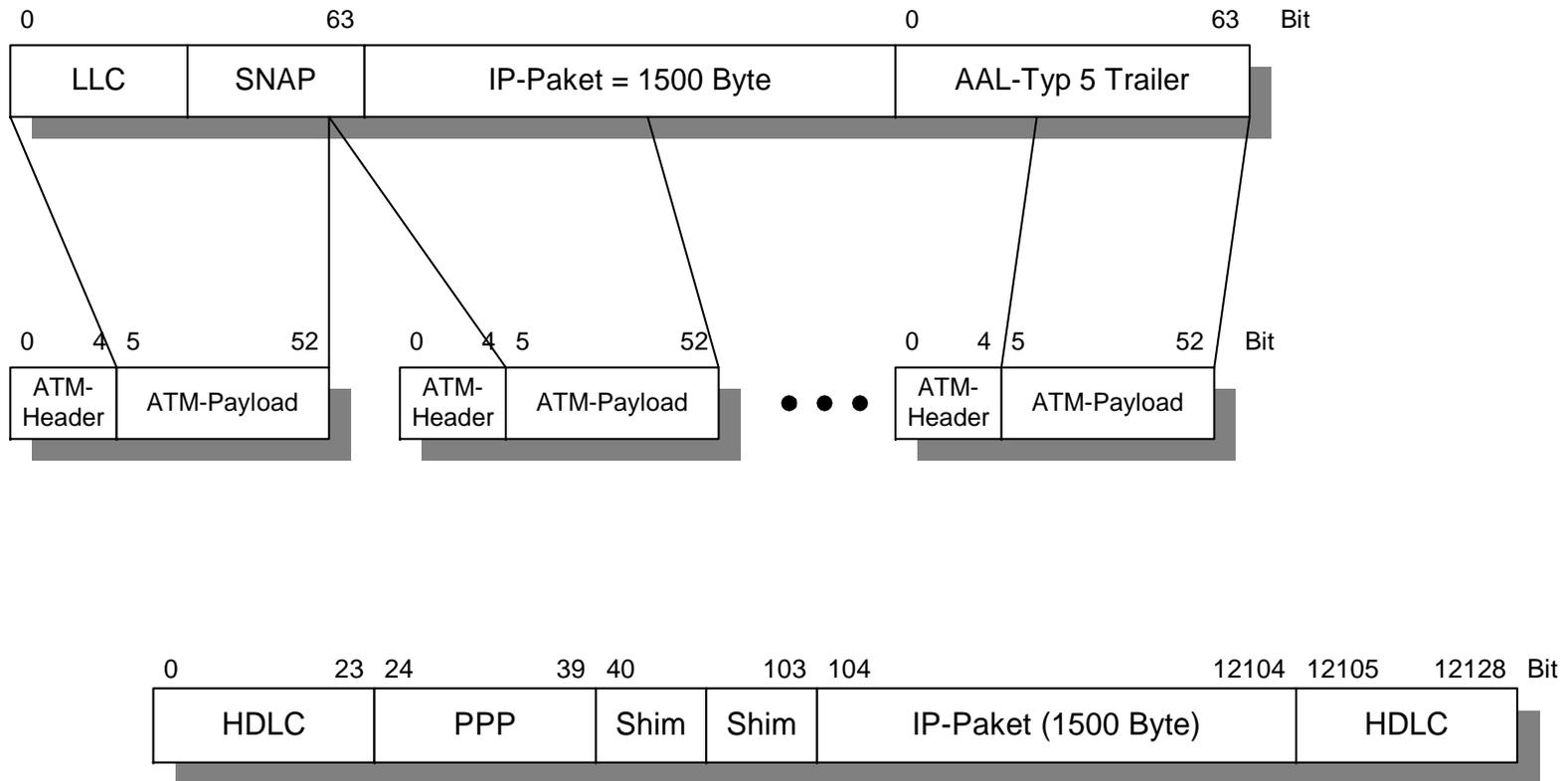
MPLS-Eigenschaften (1)

- ◆ **Virtual Private Network (VPN):** Durch das Label-Konzept ist MPLS sehr gut dazu geeignet, VPNs zu realisieren
- ◆ **Multi-Protocol- und Multi-Link-Unterstützung:** Obwohl hauptsächlich für IP Version 4 und Version 6 genutzt, bietet MPLS auch die Möglichkeit, andere Protokolle wie IPX zu übertragen. Weiterhin ermöglicht es die Kommunikation und Ressourcenreservierung über verschiedene Techniken, wie zum Beispiel Ethernet, POS, ATM und WDM
- ◆ **Quality-of-Service:** Mit MPLS ist es möglich, Ressourcenreservierung im Netz vorzunehmen. Hierfür gibt es speziell angepasste Verbindungsaufbauprotokolle, die diese Reservierungen durchführen.

MPLS-Eigenschaften (2)

- ◆ **Verschiedene Service-Klassen:** MPLS ermöglicht dem Service Provider, seinen Kunden verschiedene Dienstklassen anzubieten. Der Kunde profitiert von dieser Wahlmöglichkeit. Er wählt den Dienst nach seiner Qualität und den damit verbundenen Kosten.
- ◆ **Einsatz mit ATM:** Kombiniert mit ATM bietet MPLS auch harte Dienstgüte wie Bandbreitengarantien, Zusicherung von Verzögerungszeiten und Laufzeiten der Pakete. Vorhandene ATM-Switches können in ein MPLS-Netz integriert und weitergenutzt werden, auch wenn sie selbst kein MPLS unterstützen.
- ◆ **Packet-over-SONET (PoS):** PPP zur Einkapselung der Daten. HDLC-Rahmen zur Weitervermittlung und sicheren Übertragung. SDH-Rahmen auf der Layer 2 für das SDH-Framing. Keine direkte Anpassung auf Layer-3-Protokolle notwendig, da POS-Ansatz keine eigene Intelligenz besitzt!

MPLS-Overhead bei ATM und PoS



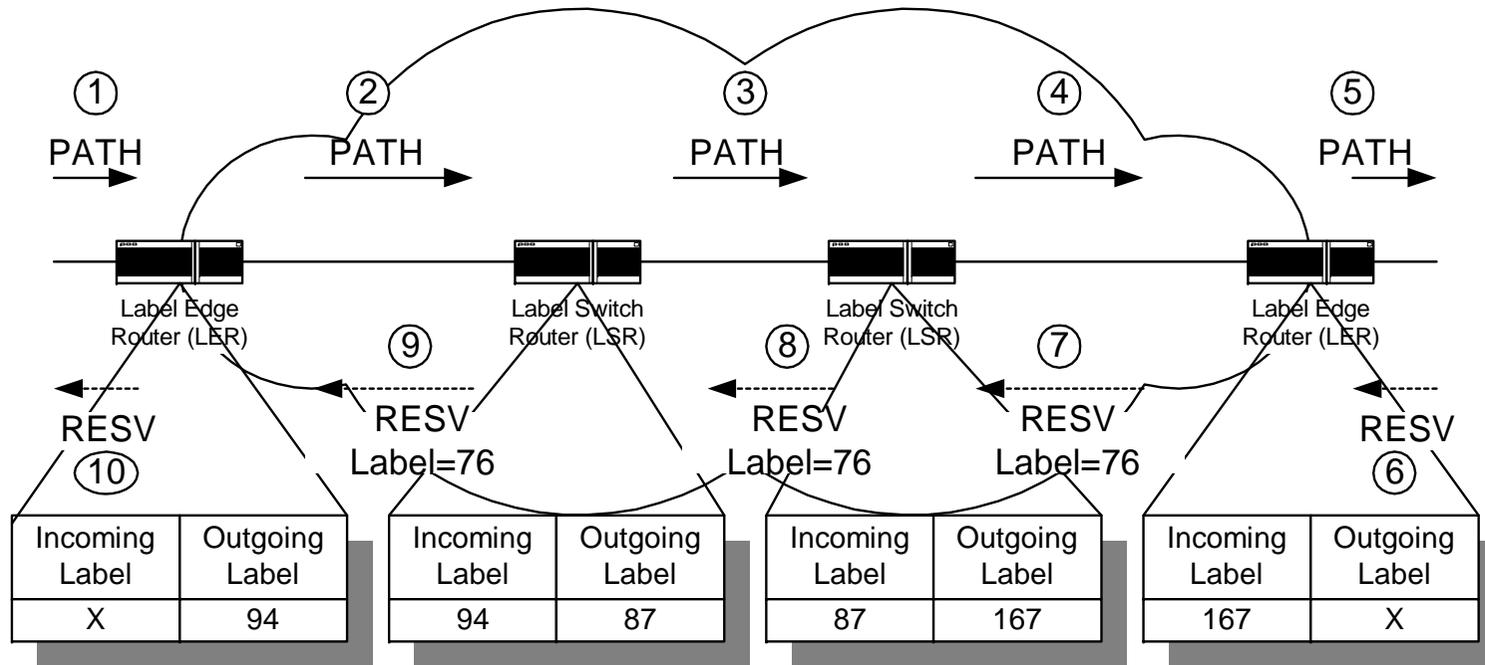
Bewertung ATM

- ◆ Anpassung/Integration relativ komplex, aber unterschiedliche Verfahren (CLIP, LANE, MPOA, MPLS) sind verfügbar
- ◆ Benötigt höheren Overhead (10,5% bei 1500 Byte Payload) als POS
- ◆ Hohe Performance durch Switching-Technologie
- ◆ Ermöglicht Traffic Engineering bereits auf Layer 2
- ◆ QoS ist auf Layer-2-Ebene als Hard-State durchführbar
- ◆ MPLS macht das ATM-Konzept weniger komplex (ATM-Funktionen wie u.a. Routing werden vermieden)

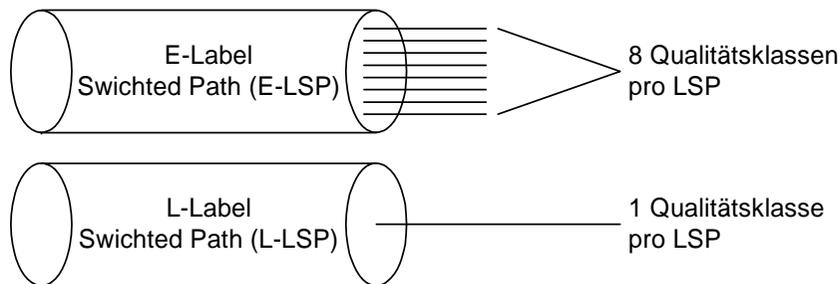
Bewertung POS

- ◆ IP-Integration wird durch einfache PPP Encapsulation vorgenommen
- ◆ Overhead (0,5-1 % bei 1500 Payload) ist deutlich geringer als bei ATM
- ◆ Performance ist hoch, da keine Intelligenz vorhanden
- ◆ Traffic Engineering wird erst durch MPLS ermöglicht (erzwingen von Übertragungspfaden)
- ◆ Weicher QoS kann mittels DiffServ/IntServ realisiert werden
- ◆ Realisierung ist effizienter/einfacher gegenüber ATM

Unterstützung von QoS: IntServ



Unterstützung von QoS: DiffServ

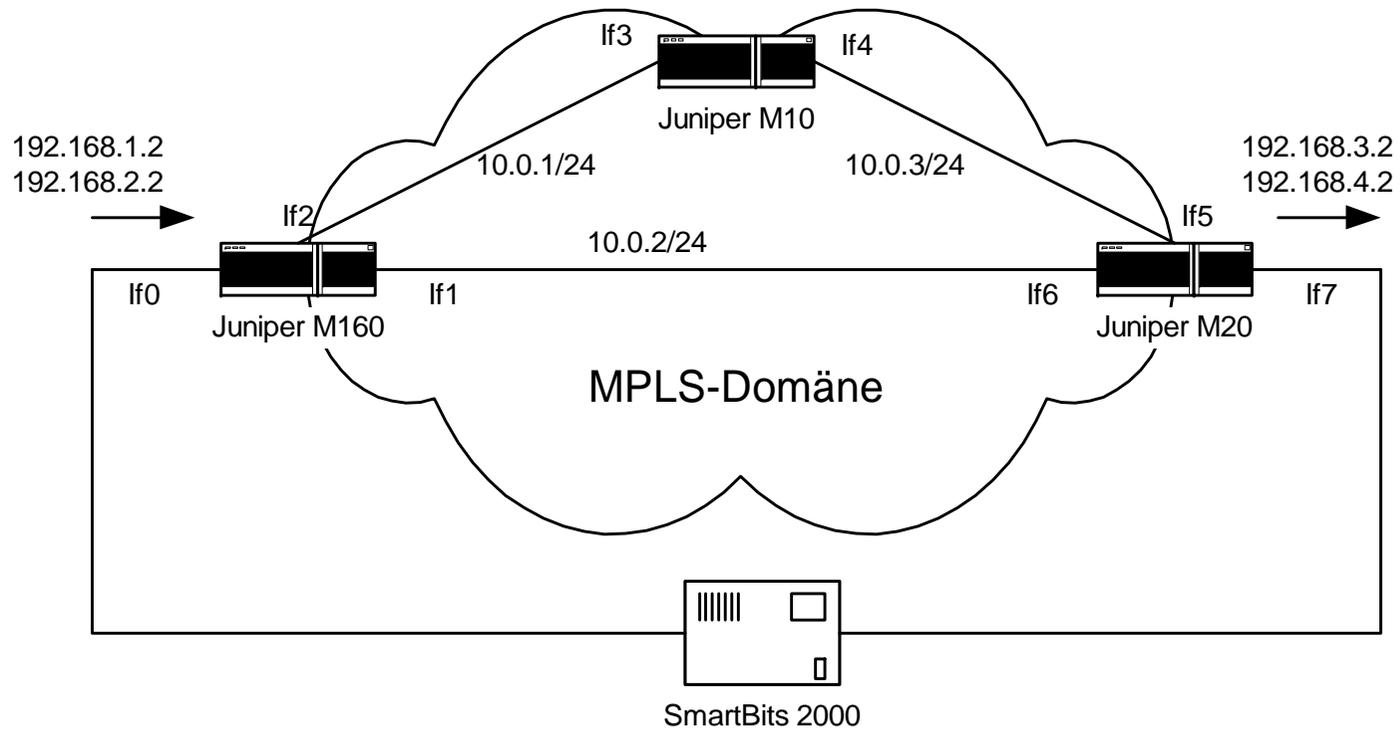


- ◆ Im MPLS werden Schicht-3-Header in den LSR nicht ausgewertet; andere Verfahren für die Bestimmung des PHB eines Pakets notwendig
 - Shim-Header (Exp-Bits): Codierung von 8 QoS-Klassen (E-LSP)
 - Labelwert: Codierung von beliebig vielen QoS-Klassen über direkte Labelzuordnung (L-LSP)

Messungen



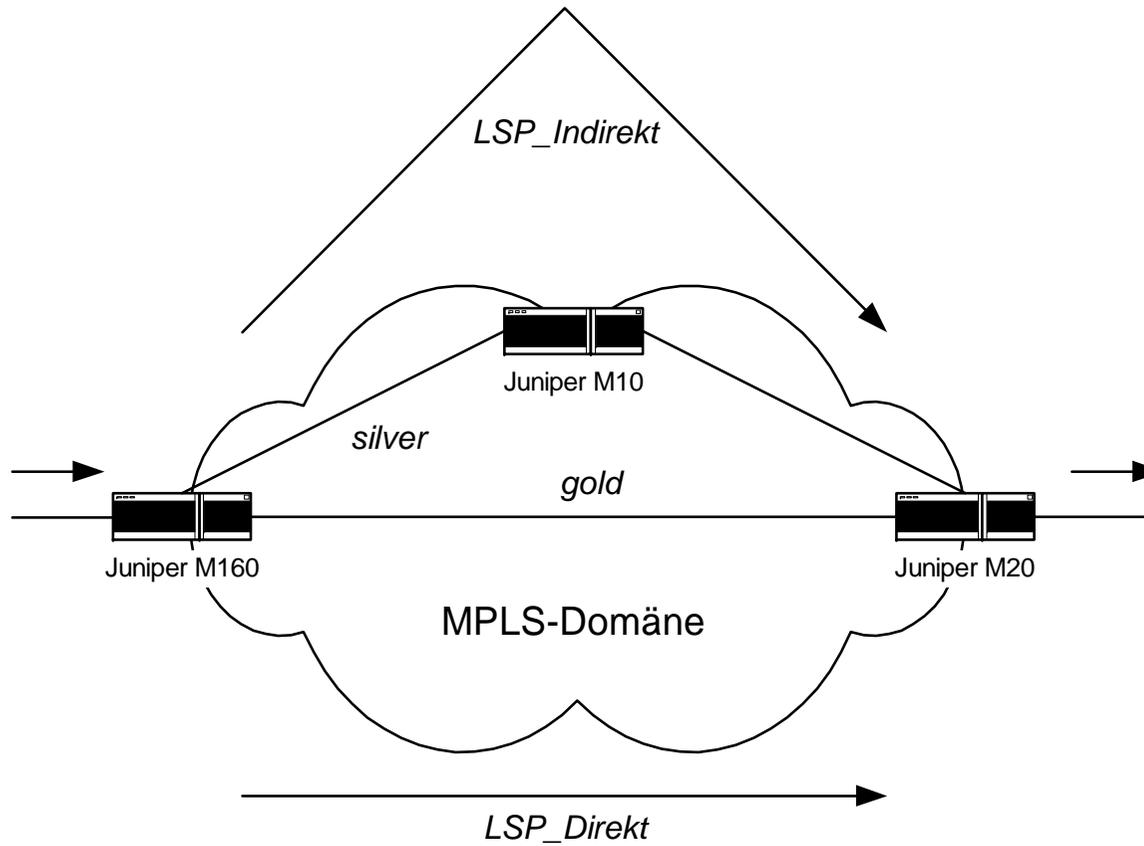
Messaufbau



Testszenario

- ◆ Das Testszenario soll diese Funktion für den Aufbau von LSPs in der praktischen Anwendung zeigen
- ◆ Zu diesem Zweck wird ein LSP als Constraint-based Routing LSP konfiguriert
- ◆ Der zweite LSP wird zum Vergleich als konventionell gerouteter LSP aufgesetzt
- ◆ Es wird hierbei der Aufbau der beiden LSPs bei folgenden Vorgaben untersucht:
 - intakten Verbindungen
 - Trennung der Verbindung M160-M20
 - Trennung der Verbindung M160-M10

LSP-Aufbau bei intakter Verbindung



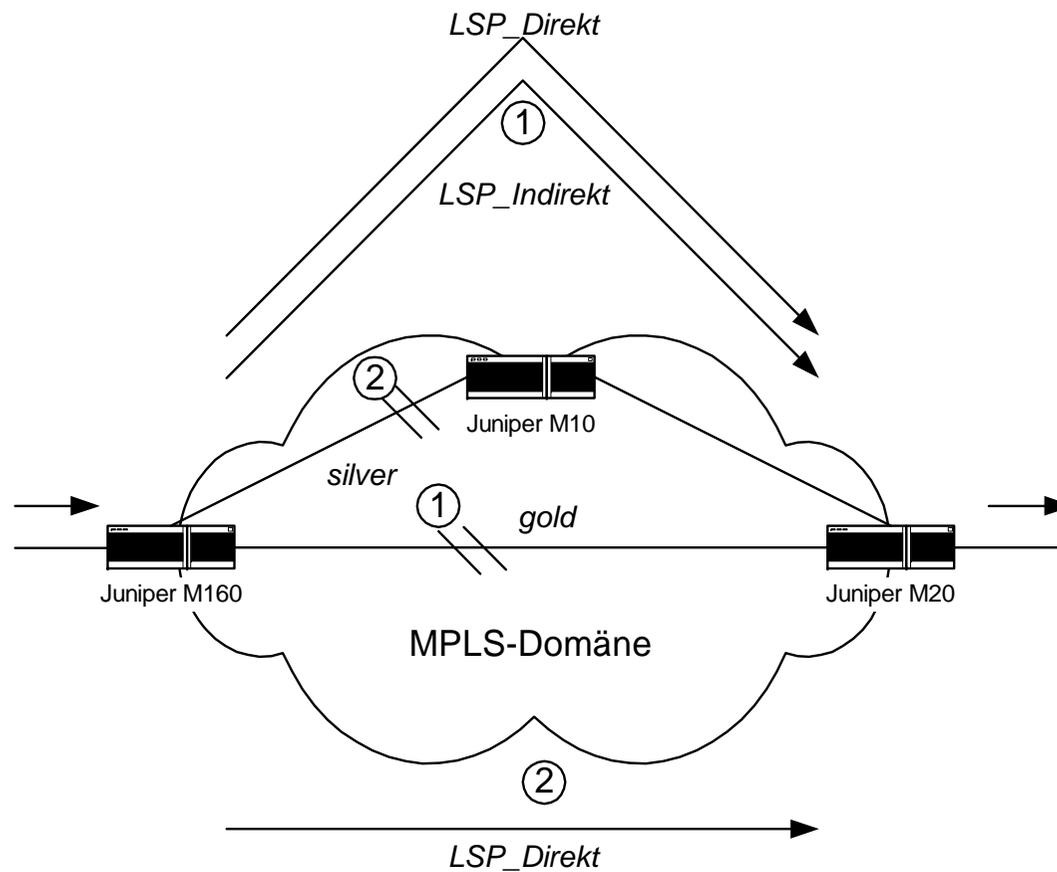
Funktionsweise (1)

- ◆ Die Bestimmung des Pfades, über den ein LSP aufgebaut wird, erfolgt durch eine Berechnung des Algorithmus Constraint Shortest Path First (CSPF)
- ◆ Dieser Algorithmus basiert auf dem Algorithmus Shortest Path First, welcher beispielsweise von OSPF verwendet wird
- ◆ Das CSPF ist aber zusätzlich in der Lage, weitere Parameter für die Routing-Entscheidung zu berücksichtigen
- ◆ Ziel des CSPF ist es ein Explicit Route Object (ERO) zu bestimmen. Das ERO ist eine Adressenliste und steuert den Aufbau der LSPs
- ◆ Für den Aufbau des LSP_Direkt bei intakten Verbindungen sollte der CSPF-Algorithmus die Verbindung mit der geringsten Anzahl von Hops (M160-M20) wählen

Funktionsweise (2)

- ◆ Für den LSP_Indirekt dürfte diese Verbindung nicht verwendet werden, da seine Konfiguration die Nutzung von Links der Übertragungsklasse *gold* ausschließt. Dies sollte in den Berechnungen des CSPF-Algorithmus berücksichtigt werden
- ◆ Der Aufbau des LSP_Indirekt sollte nur über den Pfad M160-M10-M20 möglich sein.
- ◆ Die Trennung der Verbindung M160-M20 ist automatisch mit dem Abbau des LSP_Direkt verbunden
- ◆ Über konventionelles Routing sollte für diesen LSP der Alternativpfad über die Verbindung M160-M10-M20 gewählt werden

LSP Re-routing



CSPF-Berechnung (Auszug)

```

Jun 5 12:51:39 mpls lsp LSPIndirekt primary Down
Jun 5 12:51:39 mpls lsp LSPDirekt primary No Route
Jun 5 12:51:39 mpls lsp LSPIndirekt primary No Route
Jun 5 12:51:39 mpls lsp LSPDirekt primary Deselected as active
Jun 5 12:51:39 MPLS lsp LSPDirekt down on primary()
Jun 5 12:51:39 mpls lsp LSPIndirekt primary Deselected as active
Jun 5 12:51:39 MPLS lsp LSPIndirekt down on primary()
Jun 5 12:51:41 TED free LINK j2-re0.00(172.168.1.160)->Transit.00(172.168.1.10)
Jun 5 12:51:41 TED_2_CSPF Start
Jun 5 12:51:41 mpls lsp LSPDirekt primary CSPF: link down/deleted
Jun 5 12:51:41 CSPF adding path LSPDirekt(primary ) to CSPF queue 1
Jun 5 12:51:41 CSPF creating CSPF job
Jun 5 12:51:41 mpls lsp LSPIndirekt primary CSPF: link down/deleted
Jun 5 12:51:41 CSPF adding path LSPIndirekt(primary ) to CSPF queue 1
Jun 5 12:51:41 TED_2_CSPF end elapsed time 0.000159s
Jun 5 12:51:41 CSPF job starting
Jun 5 12:51:41 CSPF for path LSPDirekt(primary ), starting at j2-re0.00
Jun 5 12:51:41 CSPF final destination 172.168.1.20
Jun 5 12:51:41 CSPF starting from j2-re0.00 (172.168.1.160) to 172.168.1.20, hoplimit 254
Jun 5 12:51:41 CSPF Reached target
Jun 5 12:51:41 CSPF completed in 0.000083s
Jun 5 12:51:41 CSPF ERO for LSPDirekt(primary ) (1 hops)
Jun 5 12:51:41      node 10.0.2.2/32
Jun 5 12:51:41 mpls lsp LSPDirekt primary CSPF: computation result accepted
Jun 5 12:51:41 mpls lsp LSPDirekt primary Clear Call
Jun 5 12:51:41 CSPF job starting
Jun 5 12:51:41 CSPF for path LSPIndirekt(primary ), starting at j2-re0.00
Jun 5 12:51:41      path exclude: 0x00000004
Jun 5 12:51:41 CSPF final destination 172.168.1.20
Jun 5 12:51:41 CSPF starting from j2-re0.00 (172.168.1.160) to 172.168.1.20, hoplimit 254
Jun 5 12:51:41      constrains exclude 0x00000004
Jun 5 12:51:41 CSPF completed in 0.000077s
Jun 5 12:51:41 CSPF couldn't find a route to 172.168.1.20
Jun 5 12:51:41 mpls lsp LSPIndirekt primary CSPF failed: no route toward 172.168.1.20
Jun 5 12:51:41 CSPF job done!
Jun 5 12:51:41 mpls lsp LSPDirekt primary Up
Jun 5 12:51:41 mpls lsp LSPDirekt primary Record Route: 10.0.2.2 S
Jun 5 12:51:41 mpls lsp LSPDirekt primary Selected as active path
Jun 5 12:51:41 MPLS lsp LSPDirekt up on primary() Route 10.0.2.2 S
Jun 5 12:51:41 TED free LINK Transit.00(172.168.1.10)->j2-re0.00(172.168.1.160)
Jun 5 12:51:44 mpls lsp LSPIndirekt primary 10.0.2.2: Explicit Route: wrong delivery
Jun 5 12:51:53 mpls lsp LSPIndirekt primary No Route

```

Neuer Pfad für *LSPDirekt*

Kein alternativer Pfad für *LSPIndirekt* gefunden

RSVP-Bandbreitenreservierung

RSVP interface: 2 active							
Interface	Active State	Subscr- resv	Static tion	Static BW	Available BW	Reserved BW	Highwater mark
so-0/1/1.0	Up	1	100%	155Mbps	55Mbps	100Mbps	100Mbps
so-0/1/3.0	Up	1	100%	155Mbps	55Mbps	100Mbps	100Mbps

Verfügbare Bandbreite

Reservierte Bandbreite

- ◆ Der Aufbau dieses LSP darf dabei nur dann erfolgen, wenn es dem CSPF möglich ist, einen Pfad zu bestimmen, welcher diese Bandbreitenanforderung auf allen Verbindungen erfüllt.
- ◆ Mit dem Aufbau des LSP ist ebenfalls die Reservierung der geforderten Bandbreite verbunden
- ◆ Zusätzlich ist es bei Constraint-based Routing möglich, einem LSP eine Pre-emption zuzuordnen. Die Pre-emption stellt eine Priorisierung des LSP dar und gliedert sich in Setup- und Holding-Priorität. Mögliche Werte der Pre-emption bewegen sich zwischen 0 und 7, wobei 0 den höchstmöglichen Wert darstellt

Ergebnis

- ◆ Für das Re-routing des LSP_Direkt musste der CSPF-Algorithmus einen alternativen Pfad zum M20 berechnen, welcher eine verfügbare Bandbreite von 100 MBit/s aufwies
- ◆ Die einzig mögliche Verbindung bestand im Link M160-M10-M20
- ◆ Aufgrund der Reservierungen des LSP_Indirekt standen hier aber nur 55 MBit/s zur Verfügung
- ◆ Da die Setup-Priorität des LSP_Direkt aber höher als die Holding-Priorität des LSP_Indirekt ausgelegt war, musste der LSP_Indirekt abgebaut und seine Ressourcen für den Aufbau des LSP_Direkt zur Verfügung gestellt werden
- ◆ Das darauffolgende Re-routing des LSP_Indirekt war nun aufgrund der verfügbaren Ressourcen und der Pre-emption-Werte nicht mehr möglich, was ebenfalls die Messungen bestätigten

Zusammenfassung



Fazit

- ◆ Label Switching wird momentan stark in der IETF vorangetrieben
- ◆ MPLS ist im Bereich der Signalisierung nicht fertig spezifiziert
- ◆ CR ist gegenüber RSVP-TE Signalisierung vorzuziehen bei MPLS
- ◆ Interoperabilität zwischen verschiedenen Herstellerlösungen kann momentan nicht sichergestellt werden!
- ◆ Einführung von QoS nach IntServ- oder DiffServ-Modell für Provider möglich
- ◆ Traffic Engineering ermöglicht bessere Auslastung des Netzes
- ◆ VPN-Funktionalität kann ebenfalls eingesetzt werden
- ◆ Komplexität von IP-Netzen wird allerdings um ein Vielfaches angehoben

Ende des Vortrags

Vielen Dank für ihre
Aufmerksamkeit

DECOIT e.K.
Zu den Stauwiesen 18
D-28879 Grasberg/Bremen
E-Mail: detken@decoit.de
Tel.: 04208/8945-48
Fax: 04208/8945-47