Simple Inter-AS CoS

draft-knoll-idr-qos-attribute
draft-knoll-idr-cos-interconnect

Thomas Martin Knoll
Chemnitz University of Technology
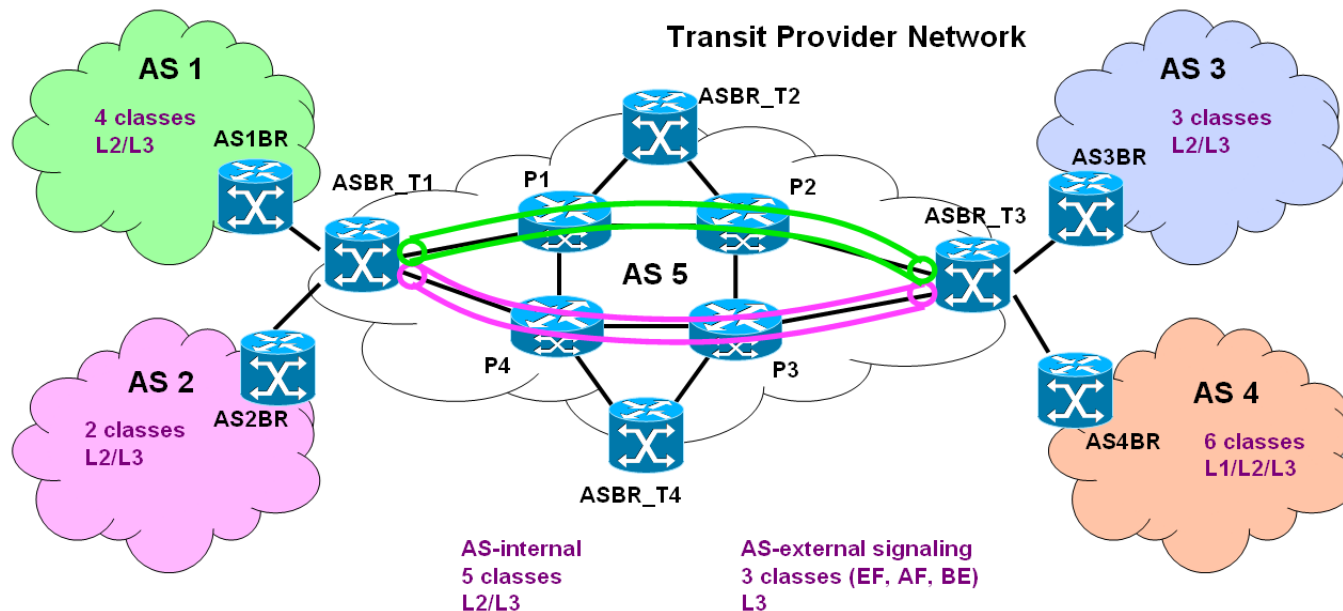**Communication Networks**
Phone   +49 (0)371 531 33246
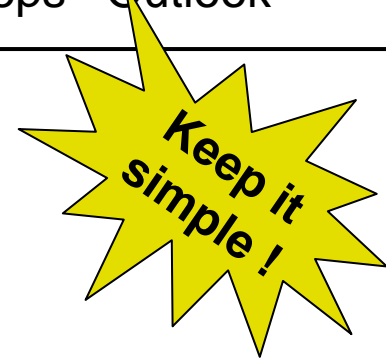**Email   knoll@etit.tu-chemnitz.de**

Recap

Traffic separation is key → AS local & Inter-AS

- BE Interconnection today:
  AS **local traffic separation** + **BE forwarding** + **costly multi-parameter ingress classification**.
- Such "**quality islands**" exist **independently**, **peer with BE** traffic, run **uncoordinated QoS concepts** and might **not** even be **known globally**.

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Recap

Keep it simple !

draft-knoll-idr-qos-attribute

http://tools.ietf.org/html/draft-knoll-idr-qos-attribute-03

- **Provides knowledge** about the available **traffic separations** and their **encoding at** the prefix **origin and locally** optionally with **Cross-layer mapping + encoding**
- **AS local traffic separation + CoS inter-AS forwarding - costly multi-parameter ingress classification.**

draft-knoll-idr-cos-interconnect

http://tools.ietf.org/html/draft-knoll-idr-cos-interconnect-01
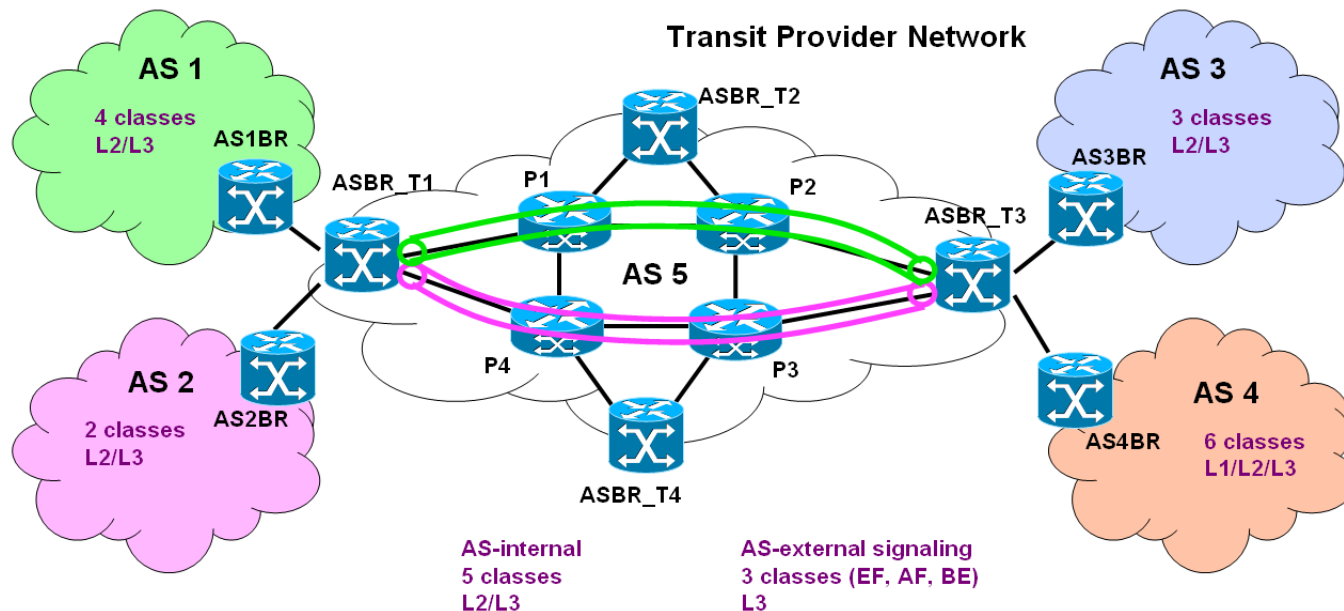
- **Prevent Class-Overload with Ingress Filters**
  → signalling for fair and square (predictable) operation

Separate drafts:
       complementary, but not depending → free to join concepts
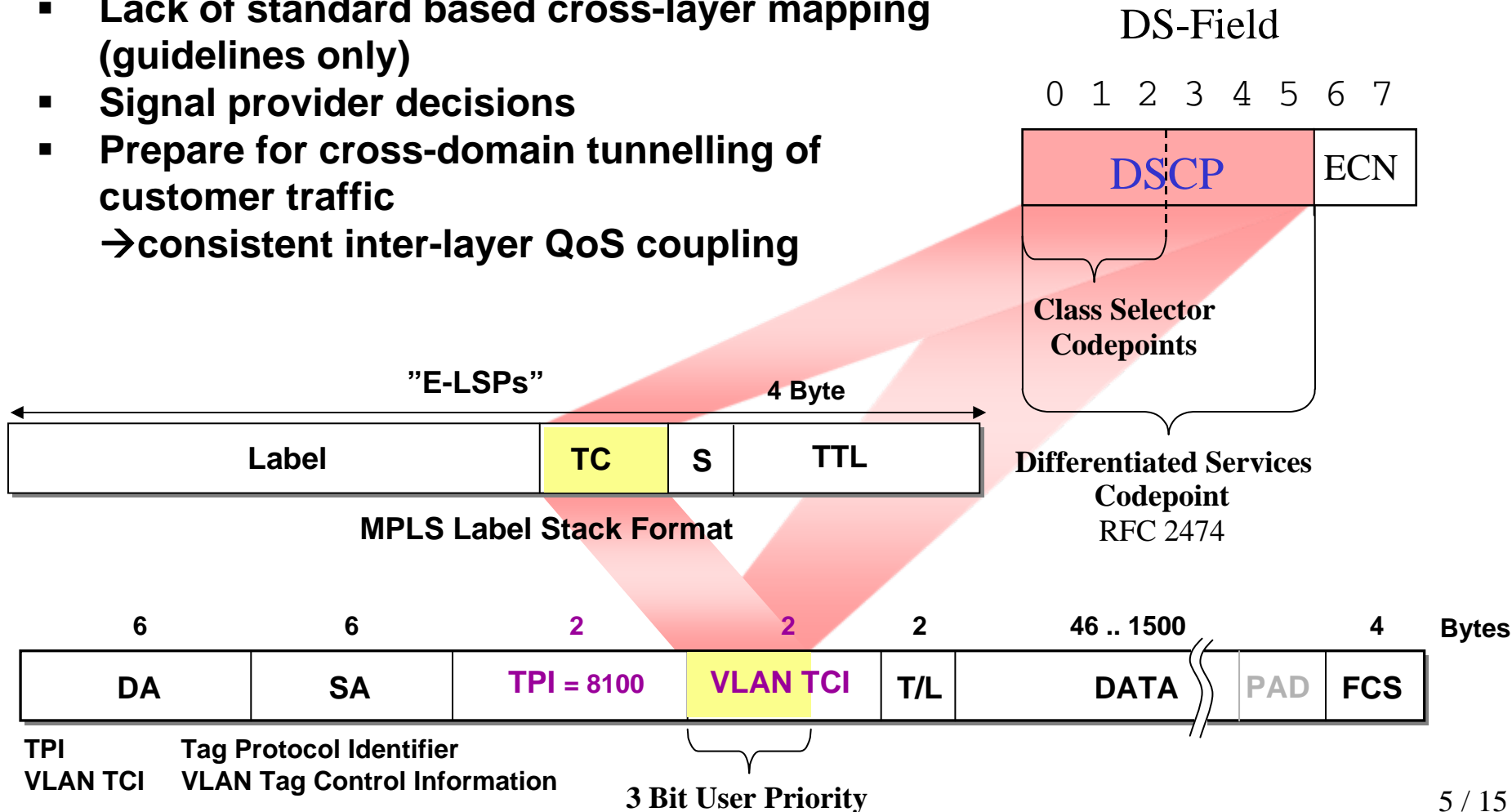   **topical interest on the marking**

3 / 15

Cross-Domain QoS signalling

- **Originating AS advertises** its **prefixes and** the supported **QoS Class Set** on several layers.
- **Relaying ASes signal support or ignore** status and their local encoding
- BGP free cores either agree on L2 Class Set or signal "ignore" status.
- **Tunnelling** of customer traffic is **preferred** for transparent transport.
- Differentiation between **internal or transit QoS Class Set**

Transit Provider Network

AS 1
4 classes
L2/L3
AS1BR

ASBR_T1   P1   P2   ASBR_T3   AS 3
3 classes
L2/L3
AS3BR

ASBR_T2

AS 5

AS 2
2 classes
L2/L3
AS2BR

P4   P3

ASBR_T4

AS 4
6 classes
L1/L2/L3
AS4BR

AS-internal
5 classes
L2/L3

AS-external signaling
3 classes (EF, AF, BE)
L3

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll
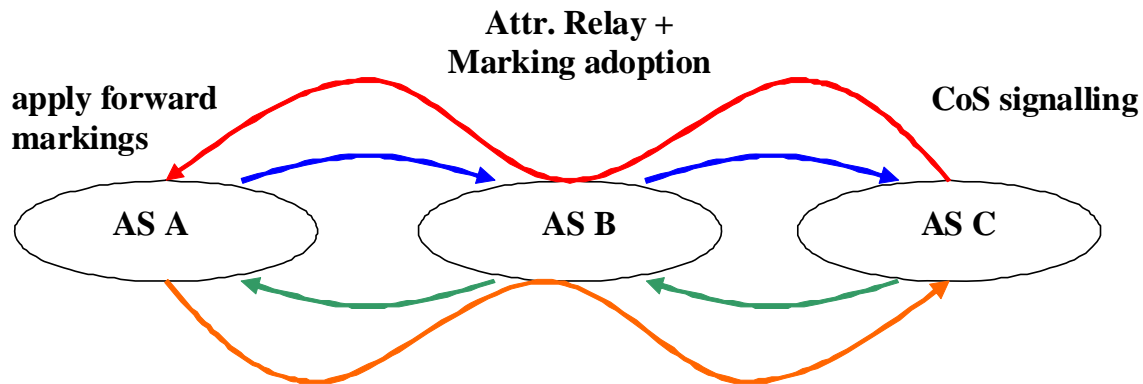
Cross-Layer QoS mapping

- **Lack of standard based cross-layer mapping (guidelines only)**
- **Signal provider decisions**
- **Prepare for cross-domain tunnelling of customer traffic**
  →consistent inter-layer QoS coupling

DS-Field

0 1 2 3 4 5 6 7

DSCP | ECN

Class Selector Codepoints

Differentiated Services Codepoint
RFC 2474

"E-LSPs"   **4 Byte**

Label	TC	S	TTL

MPLS Label Stack Format

6	6	2	2	2	46 .. 1500	4	**Bytes**
DA	SA	TPI = 8100	VLAN TCI	T/L	DATA	PAD	FCS

TPI   **Tag Protocol Identifier**
VLAN TCI   **VLAN Tag Control Information**

3 Bit User Priority

5 / 15

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Status & Changes

draft-knoll-idr-qos-attribute

Keep it simple !

Attr. Relay +
Marking adoption

apply forward
markings

CoS signalling

AS A   AS B   AS C
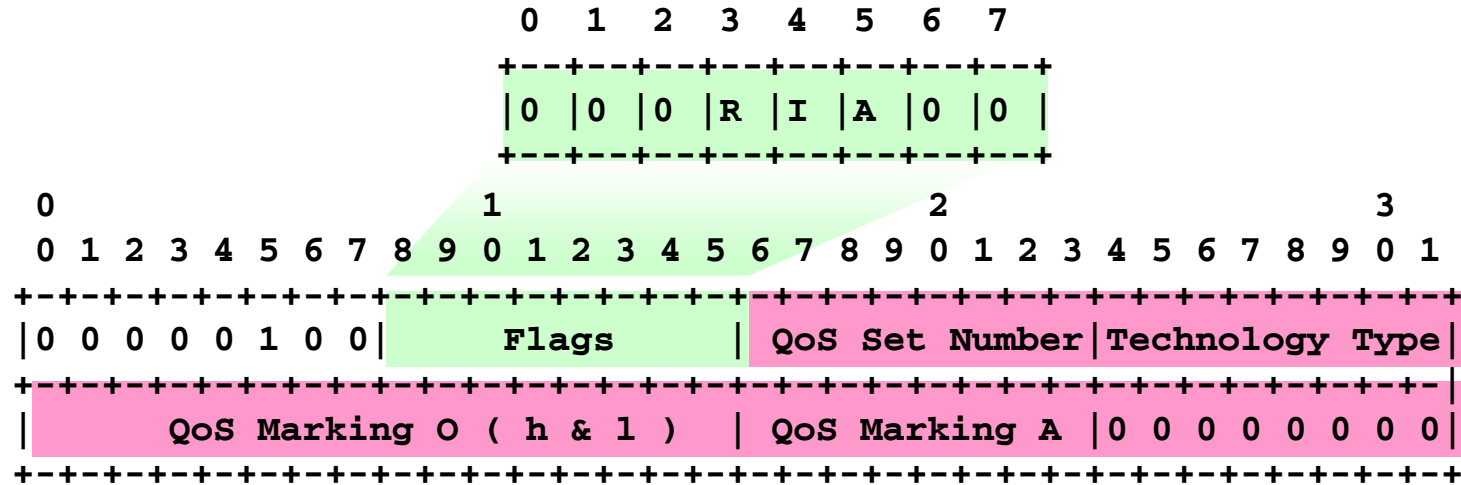
- **Supports IPv4 / IPv6** -> no change, but should be pointed out
- **New number assignment by IANA** -> 0x04, 0x44, 0x40
  http://www.iana.org/assignments/bgp-extended-communities
- Now **transitive + non-transitive** extended communities
- **Processing Count %**
  -> Confederation remove AS Path parts on egress -> falsify CoS P. count
  -> Attribute storage increases with differing P. Counts
- **Copy** of marking information **in Internet Routing Registry (IRR) for security** reasons

Status & Changes

draft-knoll-idr-qos-attribute

Keep it simple !

```
                        0   1   2   3   4   5   6   7
                       +--+--+--+--+--+--+--+--+
                       |0 |0 |0 |R |I |A |0 |0 |
                       +--+--+--+--+--+--+--+--+

  0                             1                     2                             3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |0 0 0 0 0 1 0 0|      Flags      | QoS Set Number|Technology Type|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-|
 |      QoS Marking O ( h & l )    | QoS Marking A |0 0 0 0 0 0 0 0|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Status & Changes

draft-knoll-idr-cos-interconnect

Keep it simple !
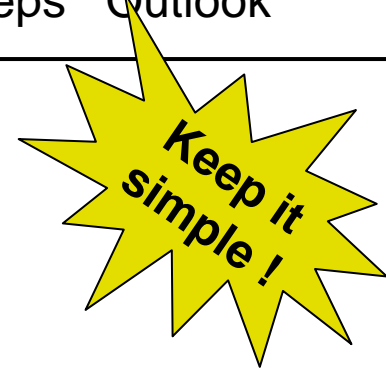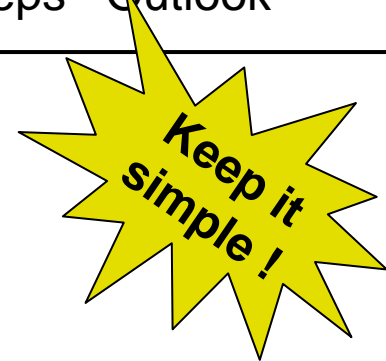
Bit	Flag	Encoding
0	BE	Default to '1' to signal general "Best Effort" PHB support
1	EF	'1' ... "Expedited Forwarding" PHB support [10]
2	AF	'1' ... "Assured Forwarding" PHB group support [11]
3	LE	'1' ... "Lower Effort" PHB support [5]
4 .. 7	unused	Default to '0'

Class Set selection
→ 2 classes – BE + LE ●
→ 3 classes – BE + EF + AF
→ 4 classes – BE + LE + EF + AF

1 Octet	1 Octet	2 Octet

Type								CoS Flags								
0	1	0	0	0	0	0	0	BE	EF	AF	LE	0	0	0	0	

Currently Unused - default to '0'

8 / 15

Status & Changes

draft-knoll-idr-cos-interconnect

Keep it simple !
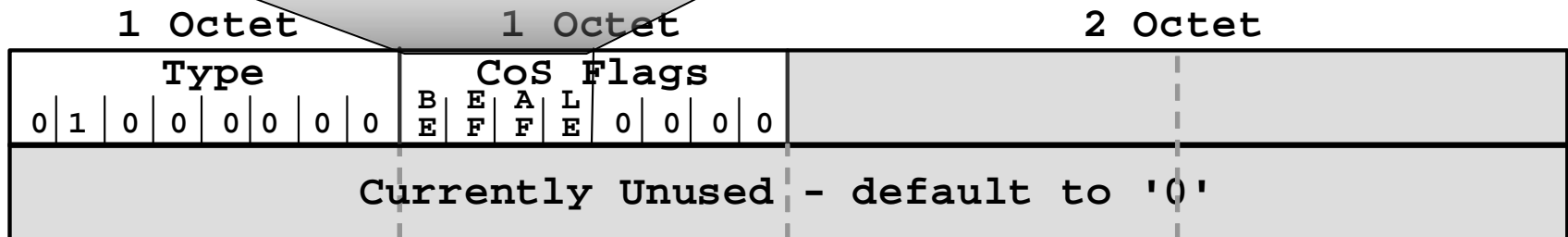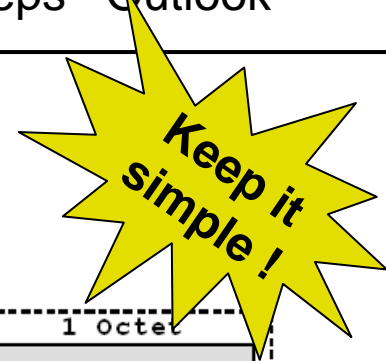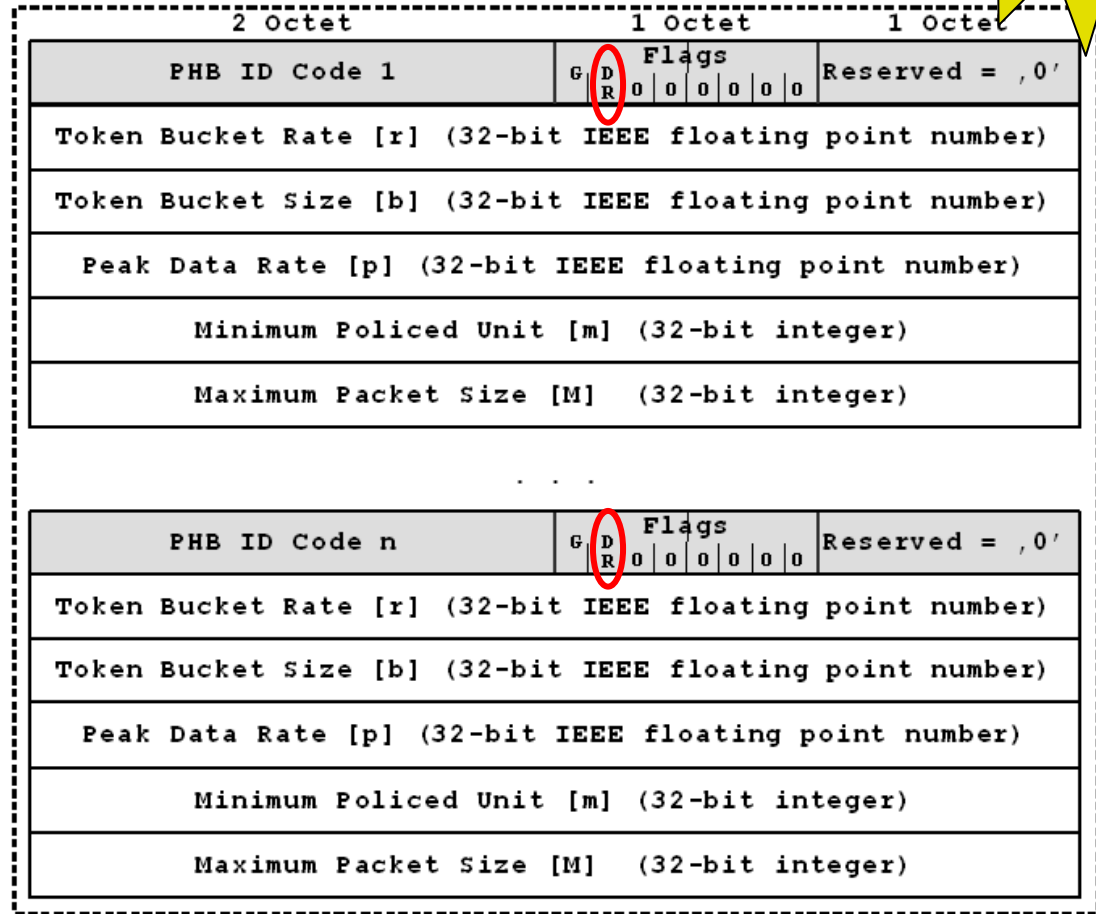
The new CoS Parameter Attribute is a **variable length non-transitive attribute**, which is not readily available as yet.

2 Octet		1 Octet	1 Octet
PHB ID Code 1	G D R Flags 0 0 0 0 0 0	Reserved = ,0'
Token Bucket Rate [r] (32-bit IEEE floating point number)			
Token Bucket Size [b] (32-bit IEEE floating point number)			
Peak Data Rate [p] (32-bit IEEE floating point number)			
Minimum Policed Unit [m] (32-bit integer)			
Maximum Packet Size [M] (32-bit integer)			

. . .

PHB ID Code n	G D R Flags 0 0 0 0 0 0	Reserved = ,0'
Token Bucket Rate [r] (32-bit IEEE floating point number)		
Token Bucket Size [b] (32-bit IEEE floating point number)		
Peak Data Rate [p] (32-bit IEEE floating point number)		
Minimum Policed Unit [m] (32-bit integer)		
Maximum Packet Size [M] (32-bit integer)		

G flag … globally or NLRI local

DR flag… drop / remarking

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll
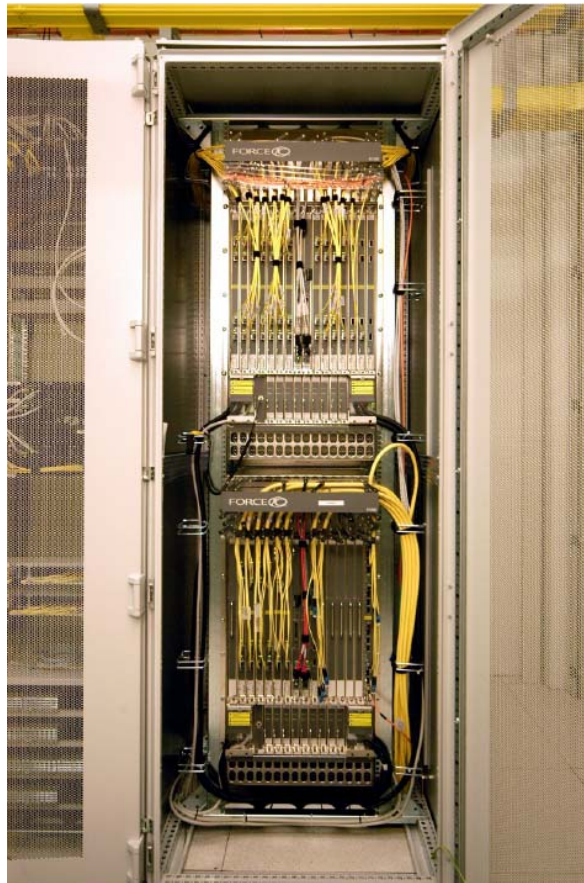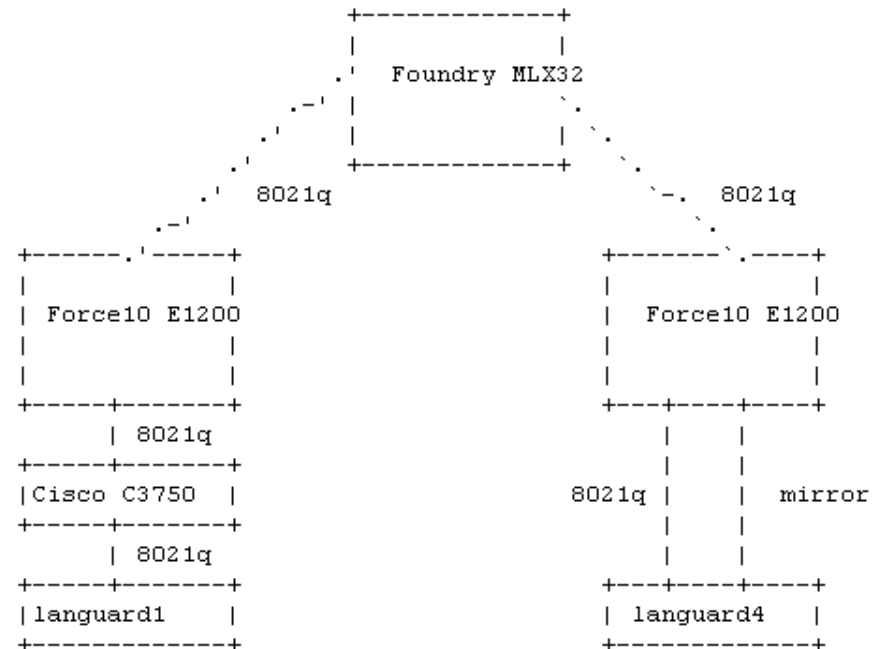
Implementation / Practical usage

- **Linux BGP routing (Quagga)** has been augmented with the selective Cross-domain & Cross-Layer CoS Marking scheme.

- The **Network Analyzer "Wireshark"** has been officially extended to detect and interpret the new BGP extended community attributes.
  http://www.wireshark.org/download.html

- Quagga + Wireshark -> regular type (8bit) introduced

- Lab tests with Cisco routers have been performed, which enabled feasibility testing, statistical calculations on real world resource usage and revealed the output of the new attribute information within Debug logs.

- An **online decoding** service for **debug logs** is available at:
  http://www.bgp-qos.org/draft-knoll/decode_attributes.php *(code available upon request)*

Implementation / Practical usage

Tests with **Internet Exchange Points** have been performed in order to document their **CoS support (IEEE 802.1p)** on the switching platform.

[DE-CIX]

```
Test Setup @ DE-CIX 2009-03-11

                         +------------+
                         |            |
                       .'|  Foundry MLX32
                     .-' |            |    `.
                   .'    |            |      | `.
                 .'    +------------+        |   `.
               .'  8021q                       `-.  8021q
             .-'                                    `.
+------.'-----+                           +--------`.----+
|            |                           |            |
| Force10 E1200                          | Force10 E1200
|            |                           |            |
|            |                           |            |
+-----+------+                           +---+----+----+
      | 8021q                                |    |
+-----+------+                               |    |
|Cisco C3750 |                          8021q |    |  mirror
+-----+------+                               |    |
      | 8021q                                |    |
+-----+------+                           +---+----+----+
|languard1   |                           | languard4  |
+------------+                           +------------+

                                    arnold.nipper@de-cix.net
```

11 / 15

Implementation / Practical usage

- http://www.bgp-qos.org/qos-ixp/

Details of Internet Exchange Point: **DE-CIX**

Name	DE-CIX
URL	http://www.de-cix.net/
Country	DE
First Name	Arnold
Surname	Nipper
E-mail	arnold.nipper@de-cix.net
Phone	+49 69 1730 902 -0
Fax	+49 69 4056 2716
URL	http://www.de-cix.net/
QoS VLAN	X
QoS VLAN priority	X
IEEE 802.1p	802.1p support on request + non-blocking line-rate switching
Nr. of QoS queues	8

Details of Internet Exchange Point: **AMS-IX**

Name	AMS-IX
URL	http://www.ams-ix.net/
Country	NL
First Name	AMS-IX NOC
Surname	AMS-IX NOC
E-mail	noc@ams-ix.net
Phone	+31 (20) 305 89 99
Fax	+31 (20) 305 89 90
URL	http://www.ams-ix.net/contact/
QoS VLAN	X
QoS VLAN priority	X
IEEE 802.1p	802.1p support + non-blocking line-rate switching
Nr. of QoS queues	4 (7 later in 2009)

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Partial extended communities

Partial bit next
to community
type ?

```
☐ Border Gateway Protocol
  ☐ UPDATE Message
       Marker: 16 bytes
       Length: 92 bytes
       Type: UPDATE Message (2)
       Unfeasible routes length: 0 bytes
       Total path attribute length: 65 bytes
  ☐ Path attributes
    ⊞ ORIGIN: IGP (4 bytes)
    ⊞ AS_PATH: 65001 65000 35    87   339    (19 bytes)
    ⊞ NEXT_HOP: 192.168.202.1 (7 bytes)
    ☐ EXTENDED_COMMUNITIES: (35 bytes)
      ⊞ Flags: 0xc0 (Optional, Transitive, Complete)
         Type code: EXTENDED_COMMUNITIES (16)
         Length: 32 bytes
      ☐ Carried Extended communities
       •☐ QoS Marking – transitive
            Type: 0x04
          ⊞ Flags: 0x00
            QoS Set Number: 0x2e
            Technology Type: 0x02 (MPLS using E-LSP)
            QoS Marking O (16 bit): 1011 0100 0000 0010 = 0xb402
            QoS Marking A  (8 bit): 0010 1101 = 0x2d (decimal 45)
            Defaults to zero: 0x00
       •⊞ QoS Marking – transitive
       •⊞ QoS Marking – transitive
       •⊞ QoS Marking – transitive
    ⊞ Network layer reachability information: 4 bytes
```

13 / 15

Next Steps

- **Consider a trial with the free code** (email request)

- **Acceptance as wg document – no hurry**

- **Thanks for hints and feedback so far**

- **Request for comments (on/off list)**

- **Clarification on partial community handling**
  suggest partial bit next to type (at least for regular types)

Outlook

- **Detailed analysis of resource usage**
  max. 8 classes
  max. 3 techn. (IP (v4+v6 independ), Ethernet, E-LSP)
  max. 15 different Marking A
  max. 30 (15 & group bit) different Marking O
  ➔ < 0.8 MB additional routing table memory

- **TC** (traffic control) **& BGPd interaction**

- **Lab (field?) tests**
  (commercial routers with Quagga attribute source – works, because of transitive type)

Separation & Simplicity is key

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Backup slides

CoS configuration Example

```
router bgp 9000
 bgp router-id 10.3.0.1
 network 10.4.0.0/14
 network 10.8.0.0/14
 network 10.16.0.0/14
 neighbor 10.0.0.1 remote-as 9000
 neighbor 10.0.0.1 route-map rttag out
!
qos-marking Set_A
 set qos-set BE 0
 set qos-marking transitive 0 0 0
 set qos-set LE 1
 set qos-marking transitive 1 0 8192
qos-marking Set_A2
 set qos-set BE 0
 set qos-marking transitive 0 0 0
 set qos-marking transitive 0 2 0
 set qos-set LE 1
 set qos-marking transitive 1 0 8192
 set qos-marking transitive 1 2 0
qos-marking Set_B
 set qos-set BE 0
 set qos-marking transitive 0 0 0
 set qos-set LE 1
 set qos-marking transitive 1 0 8192
 set qos-set EF 2
 set qos-marking transitive 2 0 47104
 set qos-set AF 3
 set qos-marking transitive 3 0 10240
```

```
!
ip prefix-list netA seq 5 permit 10.4.0.0/14
ip prefix-list netA seq 10 permit 10.8.0.0/14
ip prefix-list netB seq 5 deny 10.4.0.0/14
ip prefix-list netB seq 10 deny 10.8.0.0/14
ip prefix-list netB seq 15 permit 0.0.0.0/0 le 32
!
route-map rttag permit 5
 match ip address prefix-list netA
 set extcommunity qos-marking Set_A
!
route-map rttag permit 10
 match ip address prefix-list netB
 set extcommunity qos-marking Set_A2
!
```

17 / 15

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

CoS configuration Example

```
bgpd# show qos-marking
QoS-Marking Table of local AS
 QoS-Table Set_A:
  QoS-Set: BE (0x00)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x0000               0x00               yes
  QoS-Set: LE (0x01)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x2000               0x08               yes

 QoS-Table Set_A2:
  QoS-Set: BE (0x00)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x0000               0x00               yes
            0x02               0x0000               0x00               yes
  QoS-Set: LE (0x01)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x2000               0x08               yes
            0x02               0x0000               0x00               yes

 QoS-Table Set_B:
  QoS-Set: BE (0x00)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x0000               0x00               yes
  QoS-Set: LE (0x01)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x2000               0x08               yes
  QoS-Set: EF (0x02)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0xb800               0x2e               yes
  QoS-Set: AF (0x03)
    Technology Type     Original Marking     Active Marking     transitive
            0x00               0x2800               0x0a               yes
```
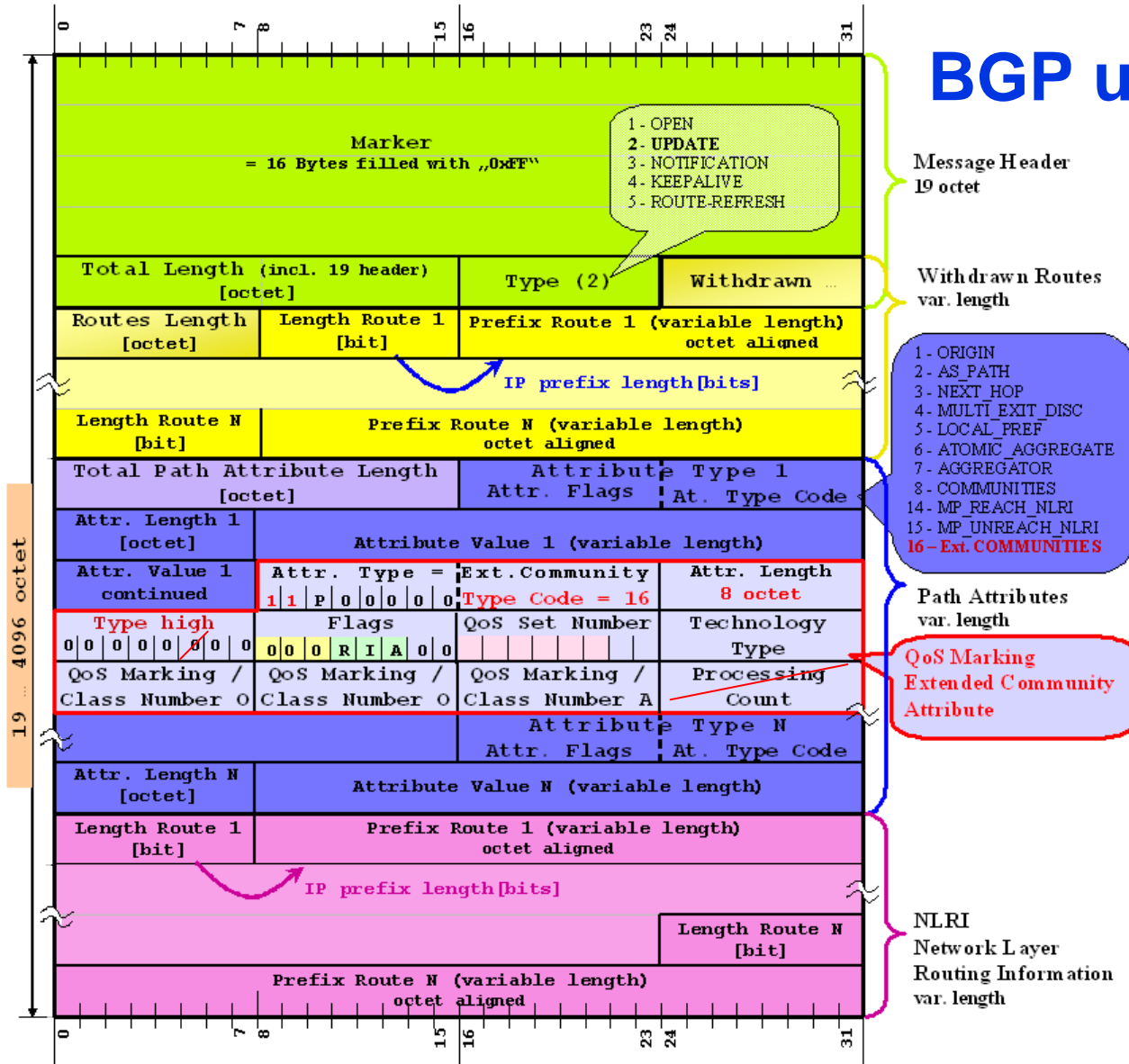
IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Communities as Signature

- **[Googlev6] page 20: "IPv6 Trusted Tester program"** → **signup using community 15169:6666**

- **QoS Marking signalling as signup** for the respective CoS support for the stated Prefixes ?

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Technology Type

Own list ← lack of one consistent one & simplification

```
+-------+----------------------------------------------------+
| Value | Technology Type                                    |
+-------+----------------------------------------------------+
| 0x00  | DiffServ enabled IP (DSCP encoding)                |
| 0x01  | Ethernet using 802.1q priority tag                 |
| 0x02  | MPLS using E-LSP                                    |
| 0x03  | Virtual Channel (VC) encoding using separate channels for |
|       | QoS forwarding / one channel per class (e.g. ATM VCs, FR |
|       | VCs, MPLS L-LSPs)                                   |
| 0x04  | GMPLS - time slot encoding                         |
| 0x05  | GMPLS - lambda encoding                            |
| 0x06  | GMPLS - fibre encoding                             |
+-------+----------------------------------------------------+
```

BGP update message

IETF 74 in San Francisco - Inter-Domain Routing WG (IDR) - Thomas M. Knoll

Related Work

[I-D.jacquenet-bgp-qos] -> QOS_NLRI attribute

[I-D.boucadair-qos-bgp-spec] -> based on QOS_NLRI + combined QoS parameter sets

[I-D.liang-bgp-qos] -> extending AS_PATH

[I-D.zhang-idr-bgp-extcommunity-qos] -> TOS value signalling

[MIT_CFP] "Inter-provider Quality of Service - White paper draft 1.1" -> 2 class strict QoS
<http://cfp.mit.edu/docs/interprovider-qos-nov2006.pdf>

Sources

[IANA_EC]   IANA, „Border Gateway Protocol (BGP) Data Collection Standard
            Communities", online available
            http://www.iana.org/assignments/bgp-extended-communities
[I-D.knoll-idr-qos-attribute]
            Knoll, T., "BGP Extended Community Attribute for QoS Marking", draft-knoll-
            idr-qos-attribute-00 (work in progress), June 2008.
[Googlev6] Colitti, L., "A strategy for IPv6 adoption", RIPE 57, October 2008,
            http://www.ripe.net/ripe/meetings/ripe-57/presentations/Colitti-
            A_strategy_for_IPv6_adoption.Z8ri.pdf
[MIT_CFP]   Amante, S, Bitar, N., Bjorkman, N., and others, "Inter-provider Quality of
            Service - White paper draft 1.1",November 2006,
            http://cfp.mit.edu/docs/interprovider-qos-nov2006.pdf
[RFC4271]   Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)",
            RFC 4271, January 2006.
[RFC4360]   Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities
            Attribute", RFC 4360, February 2006.
[RFC4594]   Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ
            Service Classes", RFC 4594, August 2006.
[Y.1541]   ITU-T, "Network performance objectives for IP-based services", Y.1541,
            February 2006