Traffic Categorisation and Inter-AS Peering

Introducing generally available coarse grained Class of Service without guarantees

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

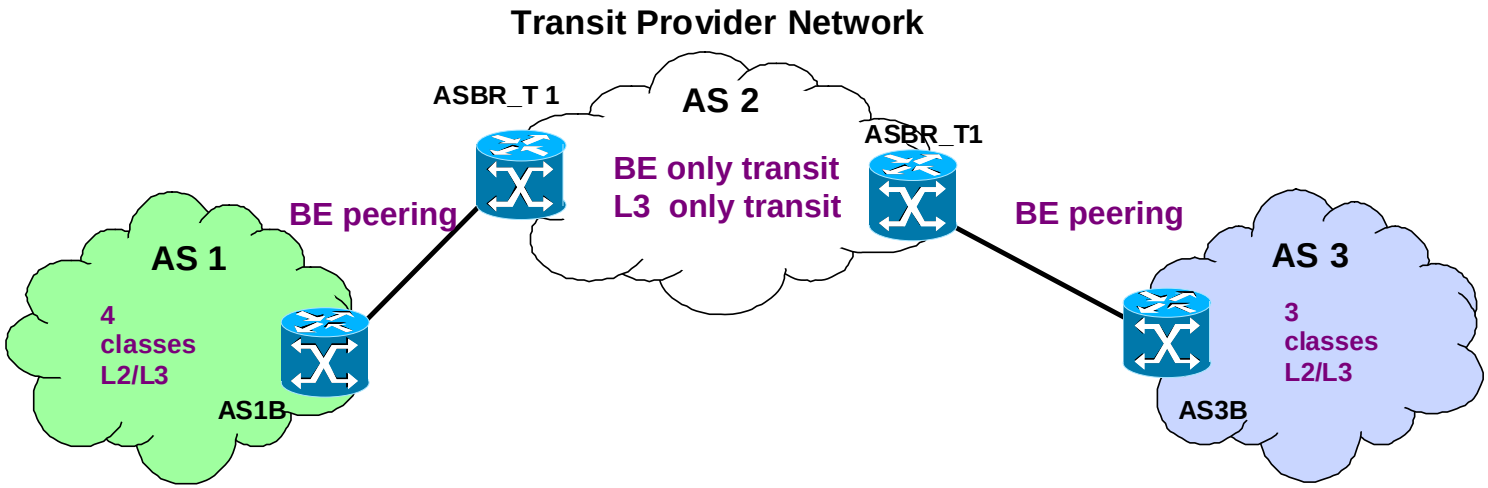
Outline

1. Motivation
2. Proposed Improvements and Focus
3. Addressed Issues
4. Usage of BGP for signalling
5. Definition of the new BGP attributes
6. Remarks
7. Summary

Motivation

Current QoS support in the Internet

- The current "**Best Effort**" packet transport in IP networks is currently being augmented by **locally applied traffic separation** with **prioritized forwarding** together with **costly multi-parameter ingress classification**.
  (markings are ignored, reset and re-classified)
- Such "**quality islands**" exist **independently**, **peer with BE** traffic, run **uncoordinated QoS concepts** and might **not** even be **known globally**.
- Complex approaches exist, which aim for **guaranteed (parameterized) QoS** support for future inter-AS peerings



Proposed Improvements / Focus

Proposed Improvements of the new Concept

- **Provides knowledge** about the available traffic separations and markings. **Cross-layer mapping** & **transitive Cross-domain signalling** is a novel feature.
- Enables **marking adoption** (and possibly route selection) **without guarantees**.
- Fair signalling of class overload limitations and excess traffic handling with local scope
- Greatly improves inter-AS packet forwarding.
- **Twofold "free to join" concept** (single or combined usage):
  1. global class set + cross-layer marking signalling (transitive attribute)
  2. local class set + rate limitation signalling (non-transitive attributes)

Traffic Separation is key:

QoS in this approach refers to primitive traffic separation into several classes, which will experience differently prioritized forwarding behaviour in relaying nodes. Enqueueing in separate queues is thereby aspired.

Proposed Improvements / Focus

Focus of the new Concept

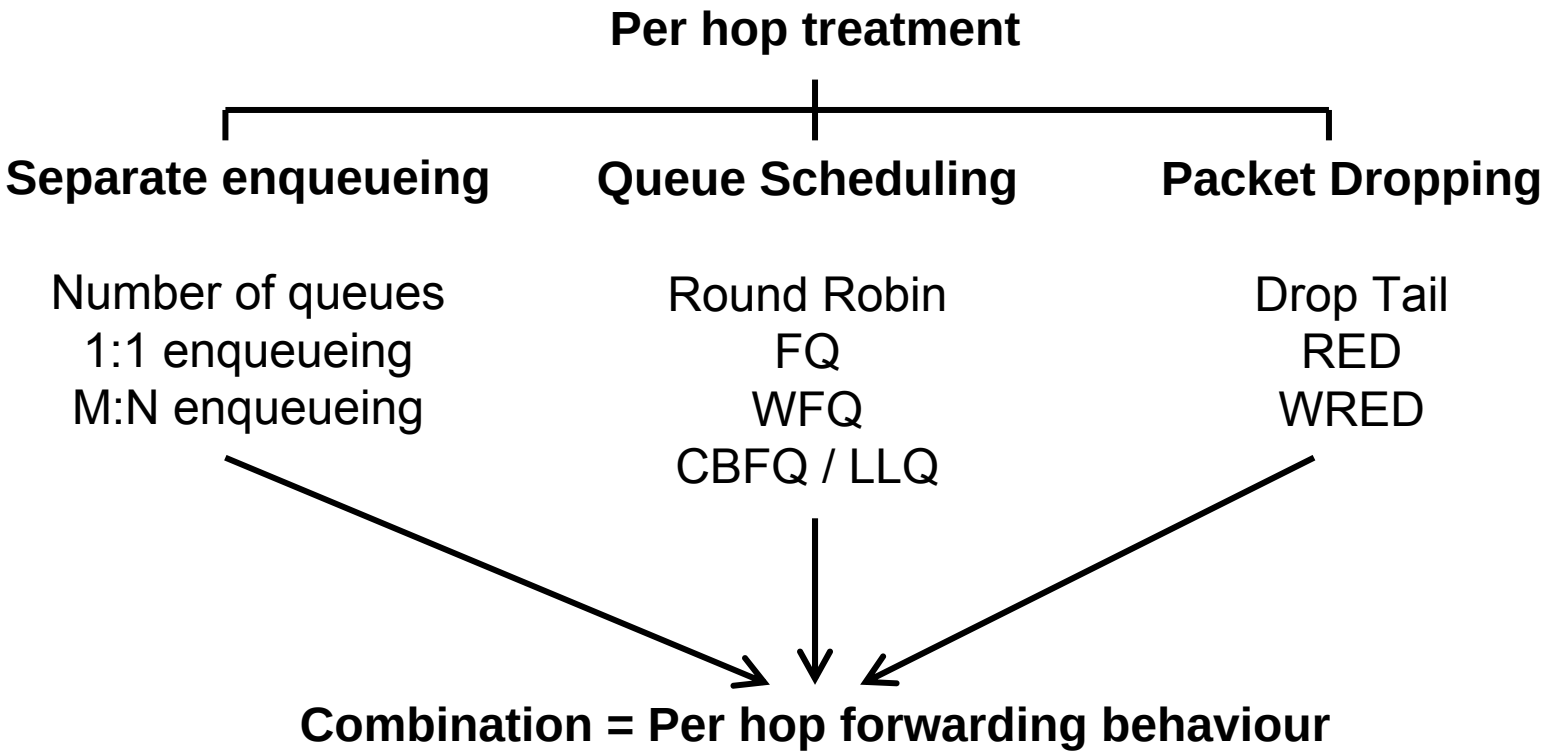
Destinction between AS level CoS:

1. **CoS based Forwarding → use case**
2. **CoS based Routing → possibly future use case**
3. **CoS based Tunnelling → use case**

Proposed Improvements / Focus

CoS based Forwarding → use case

- ”CoS based Forwarding”
  = traditional path selection + “in-path” CoS based per hop treatment

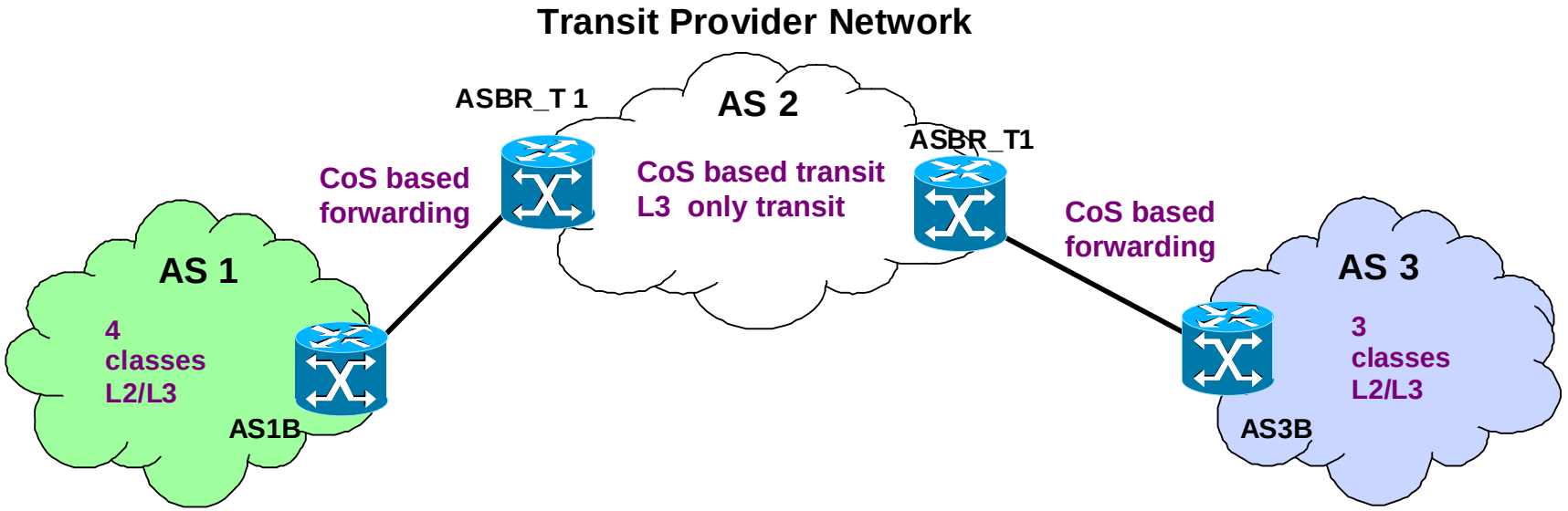


Proposed Improvements / Focus

CoS based Forwarding → use case

- ”CoS based Forwarding”

  = traditional path selection + “in-path” CoS based per hop treatment



Proposed Improvements / Focus

CoS based Routing → possibly future use case

- ”CoS based Routing”

  = CoS based path selection w/wo “in-path” CoS based Forwarding

CoS based path selection

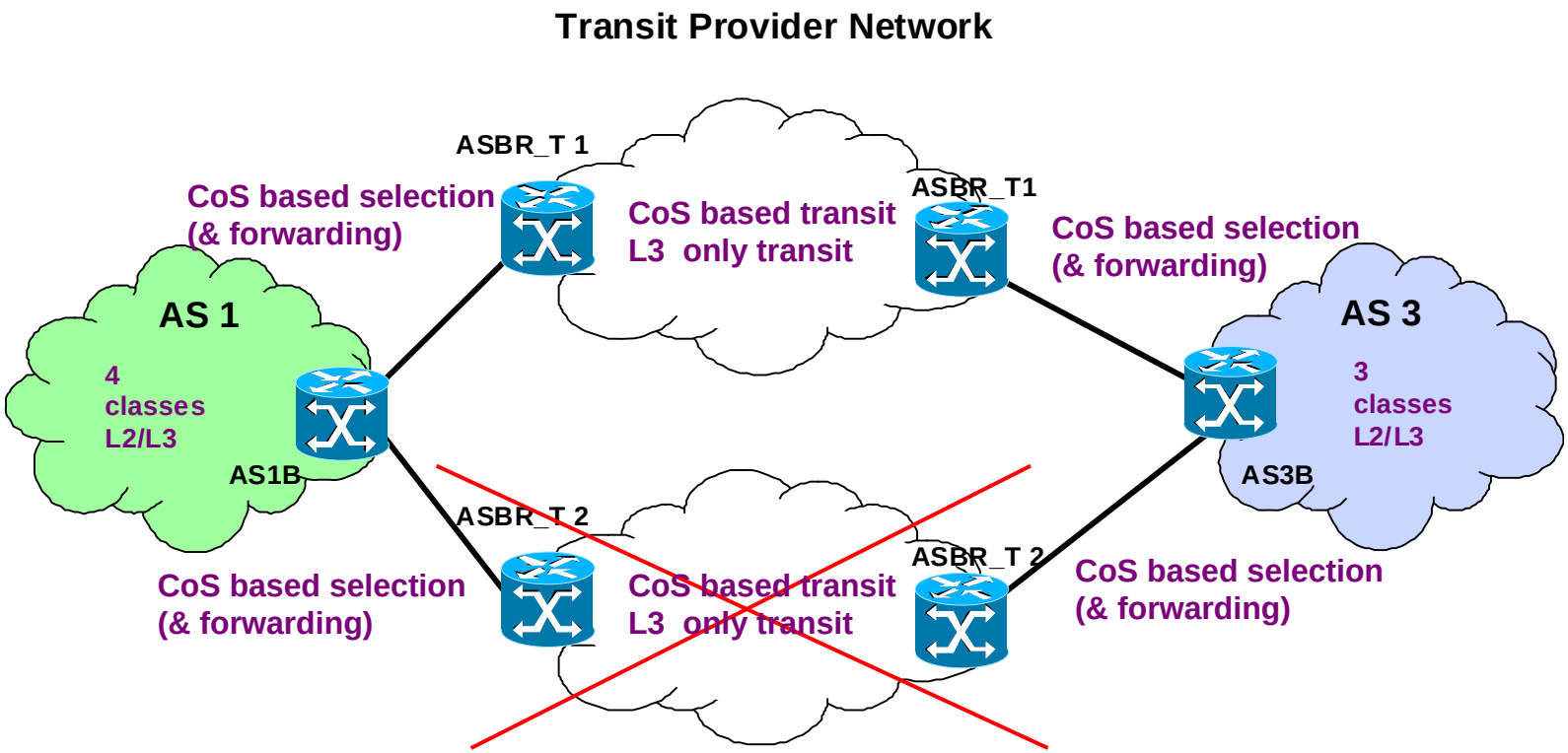
Best path selection process	**Multi-path traffic assignment “CoS based load balancing”**
Additional selection condition for (extent of) CoS support	Available in meshed setups, where **multiple** (instead of “best”) **paths** are selected **for a prefix** Requires FEC classification for next hop / interface selection
↓	↓
Best path selection modification !	**Multi-path support required**

Proposed Improvements / Focus

CoS based Routing → possibly future use case

- ”CoS based Routing”

  = CoS based path selection w/wo “in-path” CoS based Forwarding



Proposed Improvements / Focus

CoS based Tunnelling → use case

- ”CoS based Tunelling”
  = traditional or CoS based path selection + tunnelled forwarding

Tunnelling options

Intra-AS tunnelling

- E-LSPs, CE+UPC
- L-LSPs, VCs, λs, fibres etc.

↓

Tunnelled forwarding strongly recommended
Operator’s choice

At peering point

- Inter-AS E-LSPs …
- “tunnelling” through L2 marking (VLAN, UPC)
- Unlikely: layered peering

↓

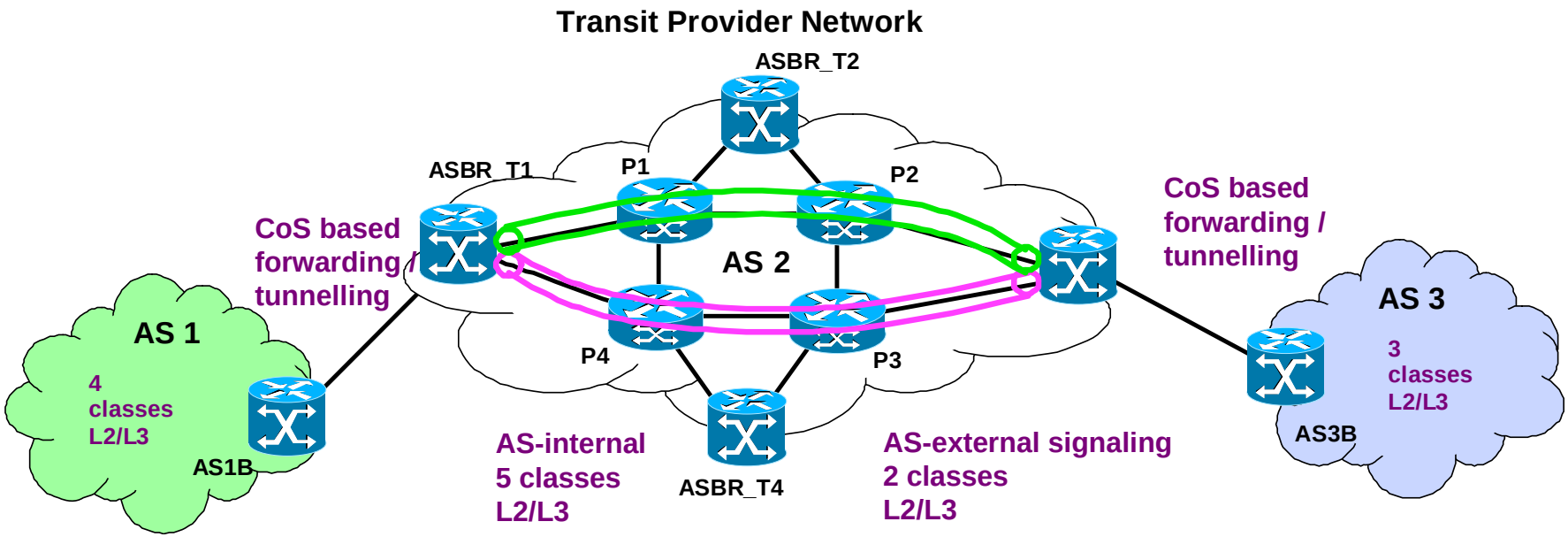
L2 marking most likely
Mutual agreement

Proposed Improvements / Focus

CoS based Tunnelling → use case

- ”CoS based Tunelling”

  = traditional or CoS based path selection + tunnelled forwarding



Addressed Issues

Cross-Layer QoS mapping

Network type	Supported QoS classes
IP supporting DiffServ	64 (currently 21 defined)
IP supporting ITU Y.1541	6
Ethernet (IEEE 802.3)	8 (802.1p priority tag)
MPLS	8 (E-LSP) or 2^{20} (L-LSP)
ATM	4 major QoS categories
UMTS	4 major QoS categories

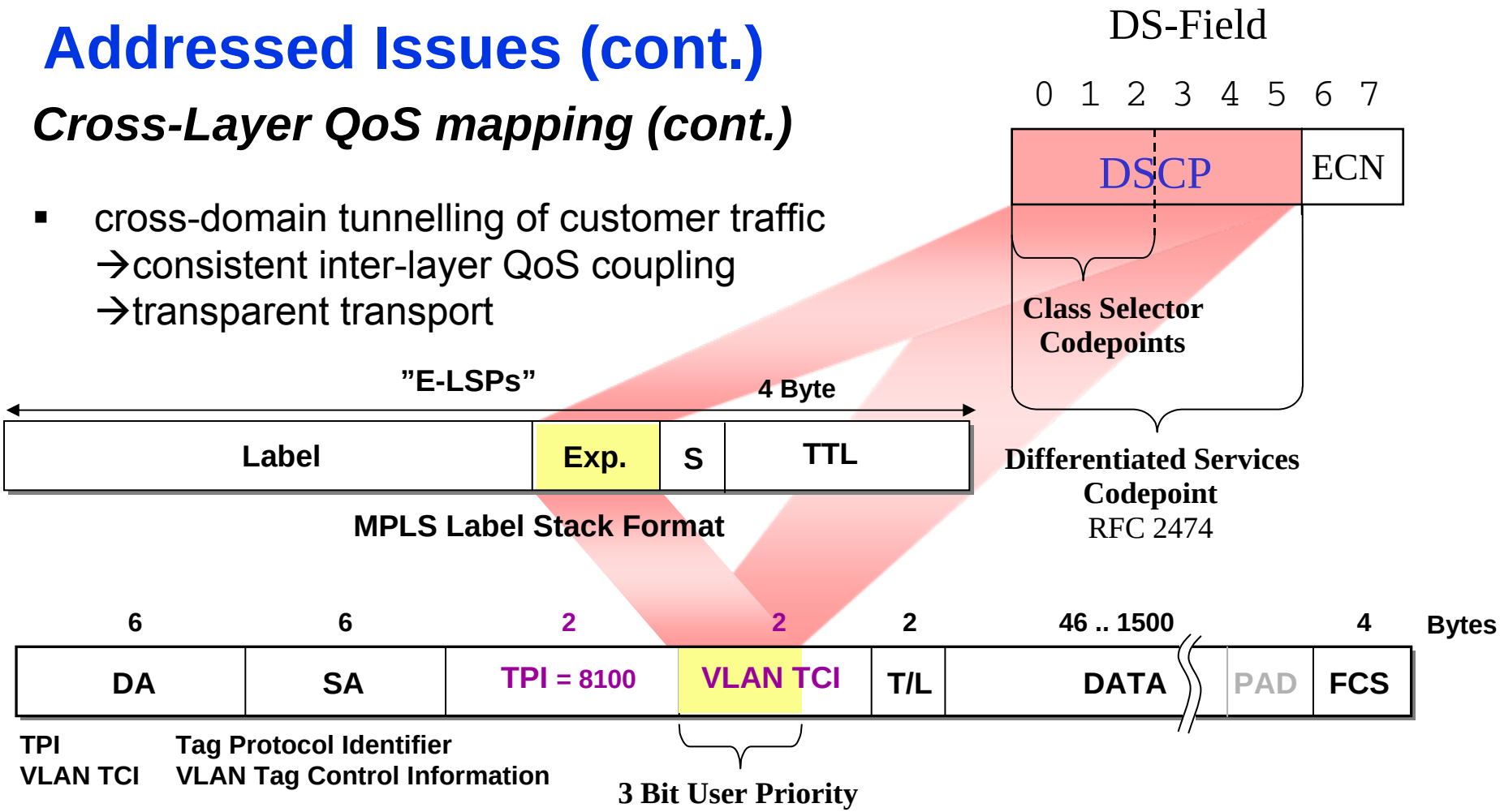
- IP as layer 3 and most layer 2 mechanisms support traffic class differentiation
- The number of classes and their encoding and mapping can freely be chosen by network providers.
- Diverse usage and internal QoS strategies are not necessarily visible outside a network domain
- Internal BGP (iBGP) is one choice for domain-internal QoS policy propagation.
- Increased usage of **tunnelling mechanisms** (MPLS(-TP)+COS, PBT+UPC, OTH channels, GRE etc.) put even more pressure on **consistent inter-layer CoS coupling** – especially for **BGP free cores**
- **Tunnels** (virtual channels) allow for QoS-based traffic engineering, which will be **regarded as Layer 1 class differentiation** in possible future uses.

The aim is consistent classification and a consistent class-based forwarding behaviour on all layers of a transit traffic path.

Addressed Issues (cont.)

Cross-Layer QoS mapping (cont.)

- cross-domain tunnelling of customer traffic
  →consistent inter-layer QoS coupling
  →transparent transport

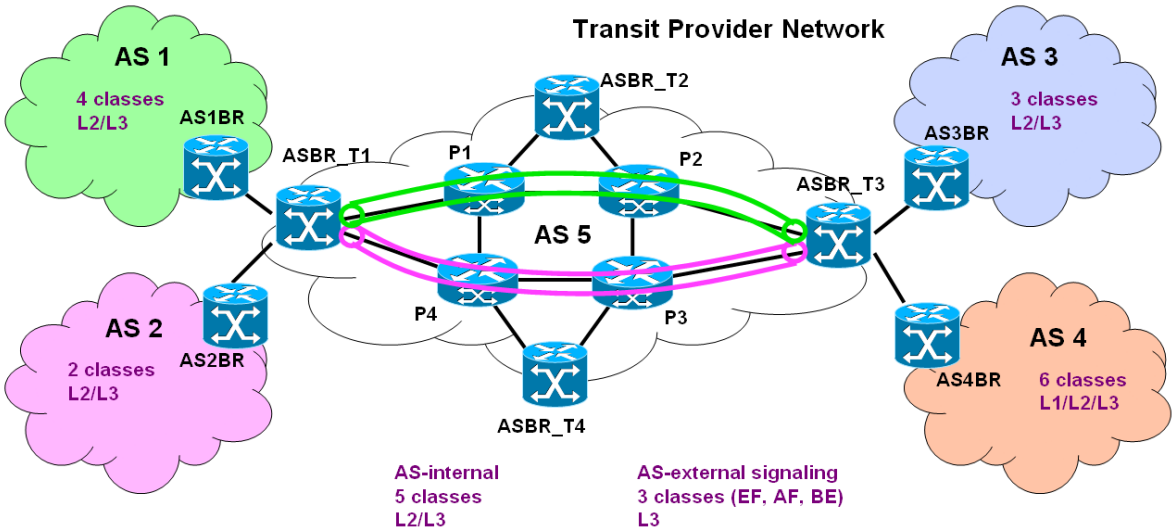


The aim is consistent classification and a consistent class-based forwarding behaviour on all layers of a transit traffic path.

Addressed Issues (cont.)

Cross-Domain QoS signalling

- Current Practice: **Best Effort only IP** traffic **peering** between ASes
- Individual agreements on class support between neighbouring ASes
- Diverse usage and internal QoS strategies are **not visible outside an AS**
- External BGP (**eBGP**) is used for **Inter-Domain signalling**



The aim is consistent classification and a consistent class-based forwarding behaviour on all layers of a transit traffic path.

Addressed Issues (cont.)

CoS – Class Overload prevention

- Traffic separation and enqueueing into separate – prioritised – queues tempts users to **overload the higher priority classes**.
- **Limitation and punishment concept** on the following slides

Usage of BGP

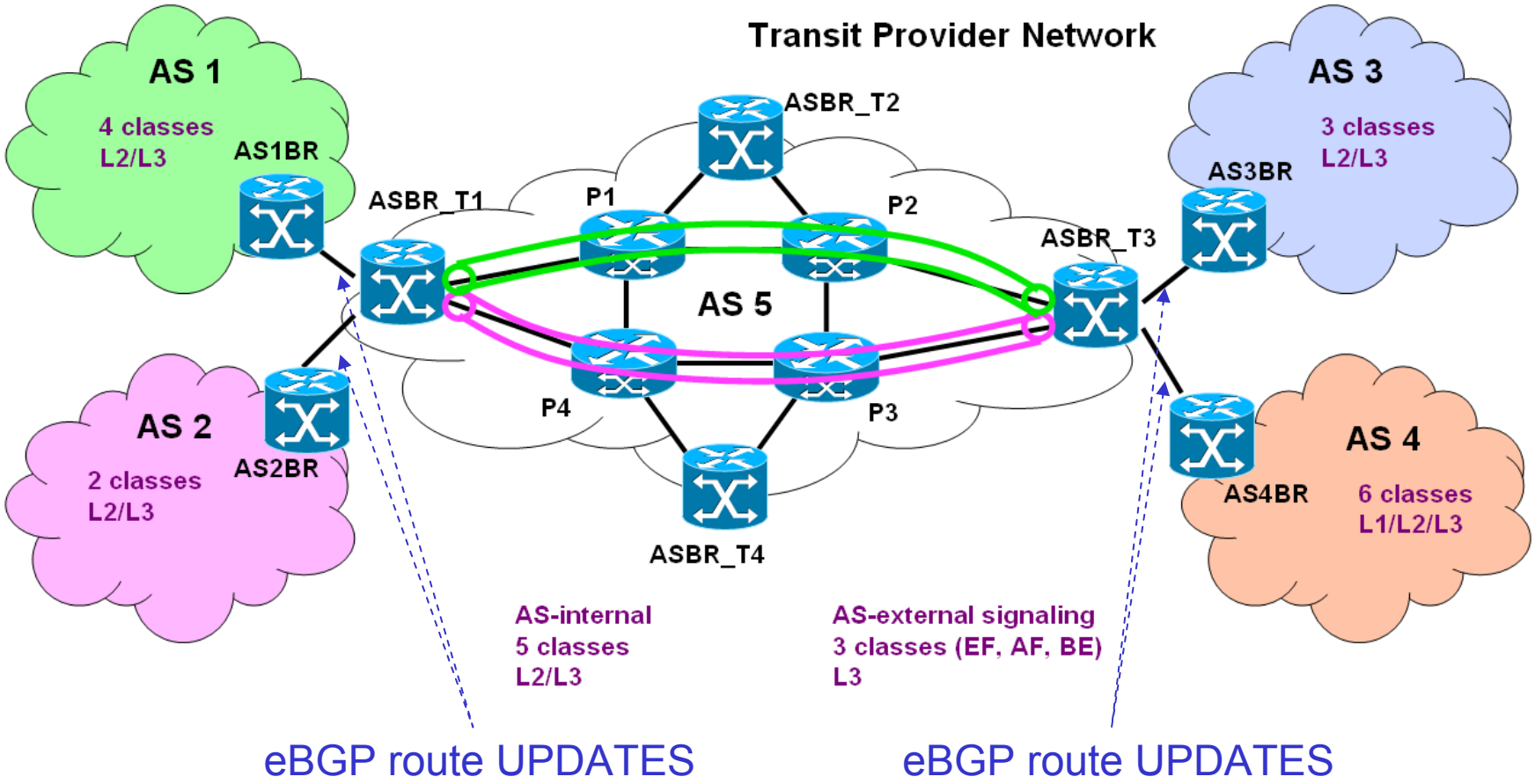
Why to use BGP for signalling

- BGP is the de-facto peering protocol
- Globally accepted / globally available
- Well designed flexible protocol that allows for such extensions
- BGP exchanges reachability information and can tag this information with route related attributes
  ➔ use BGP to signal available traffic separation tags along the announced routes ➔ no multiple advertisement of the same prefixes

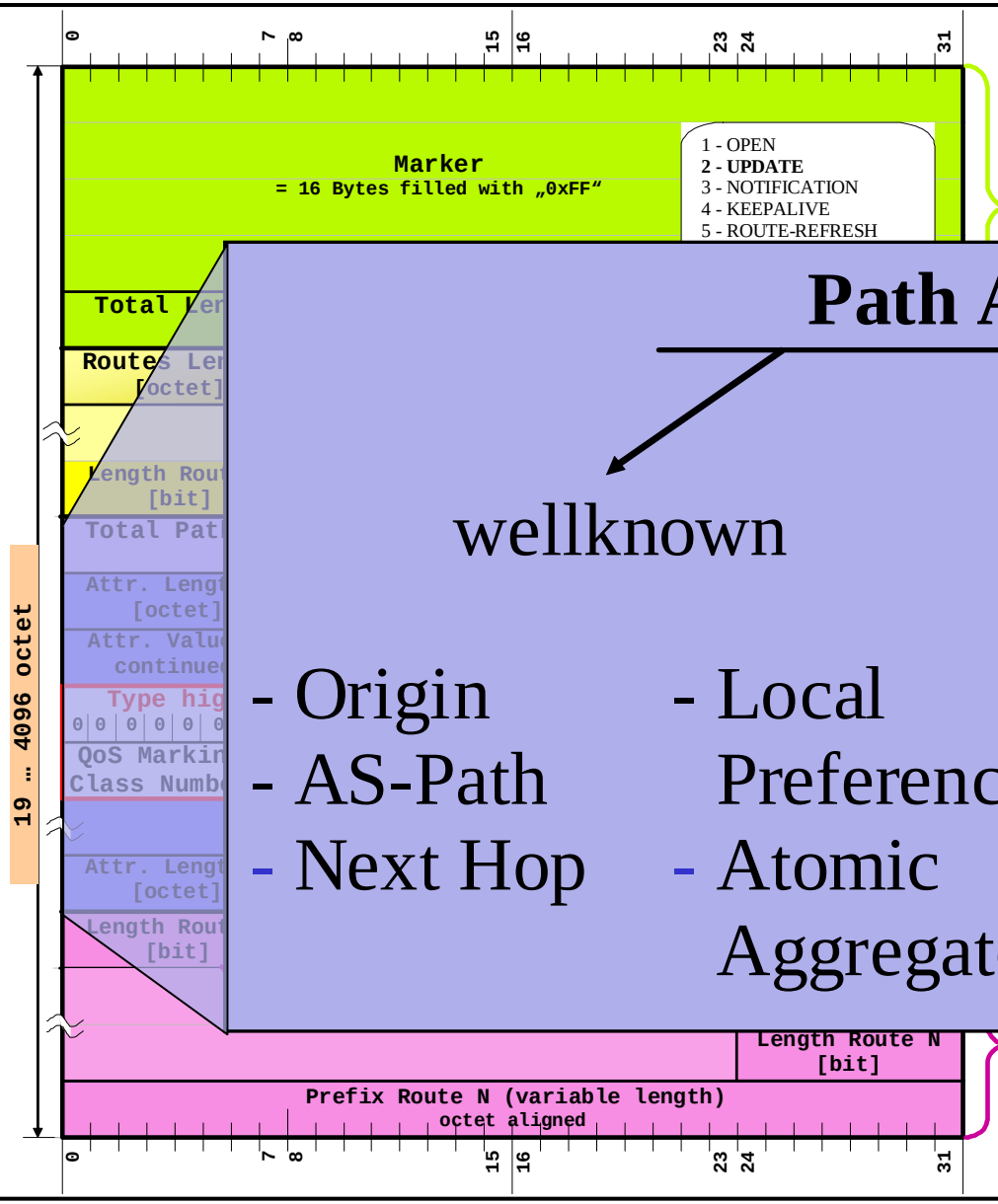
Why not to use BGP for signalling

- BGP's stability is achieved through dampened UPDATE rates and the concept of failure confinement within routing areas or confederations
  → avoid any fast changing information to be convey in UPDATE messages
  → hourly or slower changes are acceptable
- Avoid BGP signalling, if globally accepted standard definitions exist
- Avoid BGP signalling, if the size or number of UPDATE messages becomes largely increased

eBGP peering between neighbouring ASes



BGP update message



Path Attributes

wellknown		optional
- Origin	- Local Preference	- Aggregator
- AS-Path	- Atomic Aggregate	- MED
- Next Hop		- Community
		- Ext. Comm.

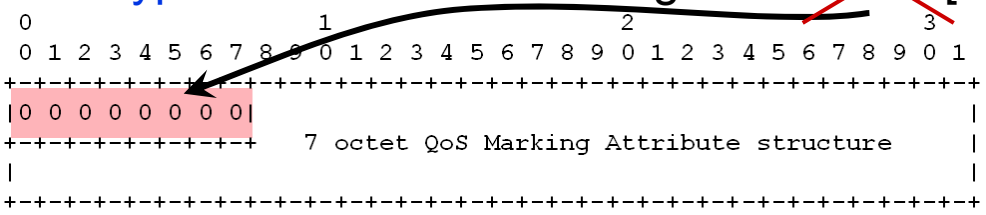
Definition of the QoS Marking Attribute

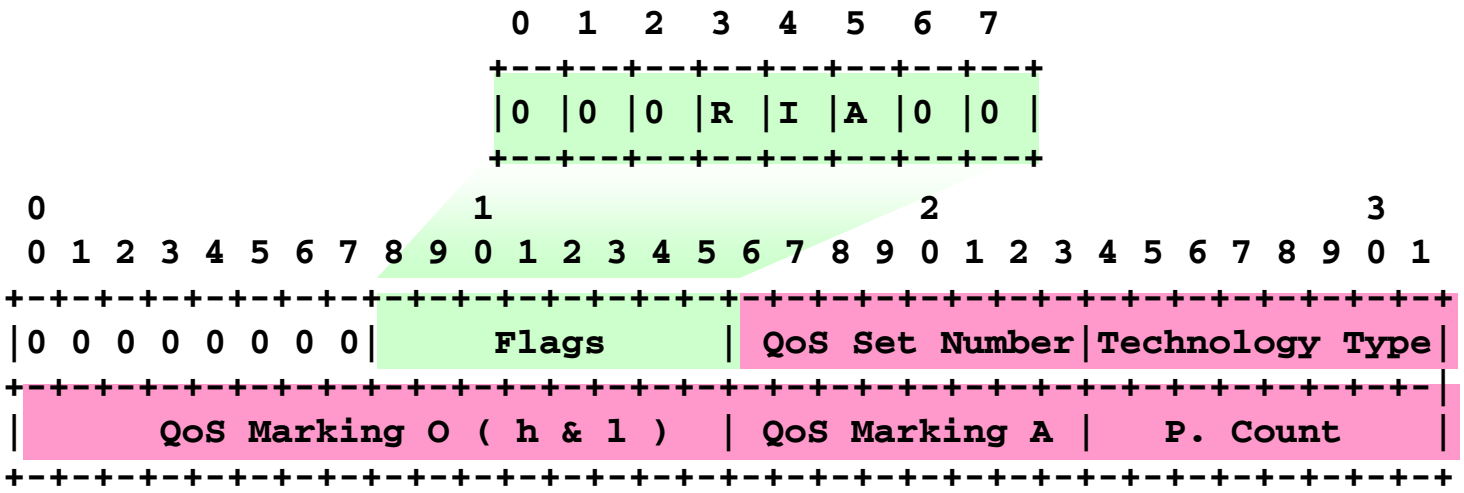
Ext. Community Attribute

http://tools.ietf.org/html/draft-knoll-idr-qos-attribute-02

The new QoS Marking Attribute is encoded as a **BGP Extended Community Attribute** [RFC4360]. It is therefore a **transitive optional** BGP attribute with **Type Code 16**.

The Type Value has been assigned to ~~0x00~~ possibly 0x04 [IANA_EC].





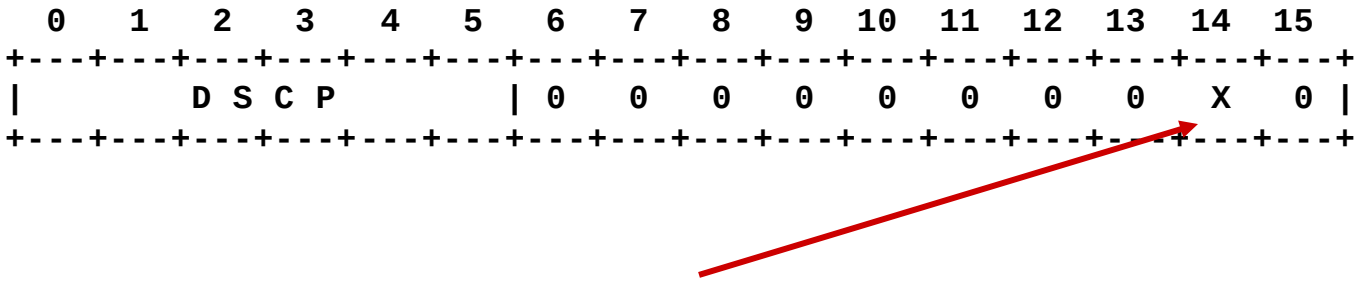
Definition of the QoS Marking Attribute

QoS Marking O as PHB ID code

Inter-domain signalling of PHBs is to be done using the

PHB ID code format - RFC 3140

Focus here on standards track PHBs with assigned values:



Differentiation between

single PHBs (‚0‘) and **PHB groups** (‚1‘)

➔ Allows for **cross-layer marking of PHB groups**

Definition of the QoS Marking Attribute

Optional transitive Attribute

- Smooth integration and transparent transport across ignoring ASes
- Fixed fields guarantee unchanged values / other fields for local adaptation

QoS Set – Concept of "linked" together attributes

- Several QoS Attributes will be included, which are virtually grouped together
- Grouping not fixed to technology or DSCP etc.

Technology Type

- Lack of common enumeration of different layer technologies
  → own enumeration list

Processing Count

- Detection of non-cooperative ASes (Count vs. diff. AS numbers in AS_PATH)
- Route selection based on 'I' flag and P. Count possible
- Additional usage of the attribute's 'P'-flag

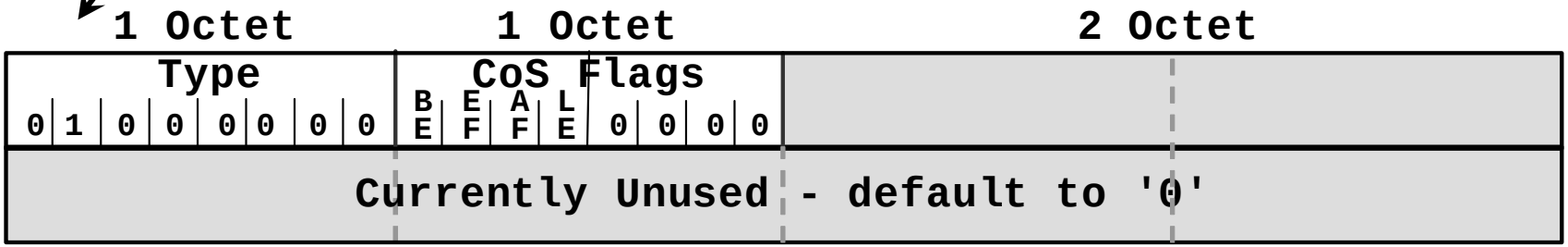
Definition of the CoS Capability Attribute

Ext. Community Attribute http://tools.ietf.org/html/draft-knoll-idr-cos-interconnect-00

The new CoS Capability Attribute is encoded as a **BGP Extended Community Attribute** [RFC4360]. It is therefore a **transitive optional** BGP attribute with **Type Code 16**. Inside the extended community, a further distinction of transitive and non-transitive is made.

The regular Type Value of this non-transitive ext. community has been assigned to 0x40 [IANA_EC].

possibly 0x44

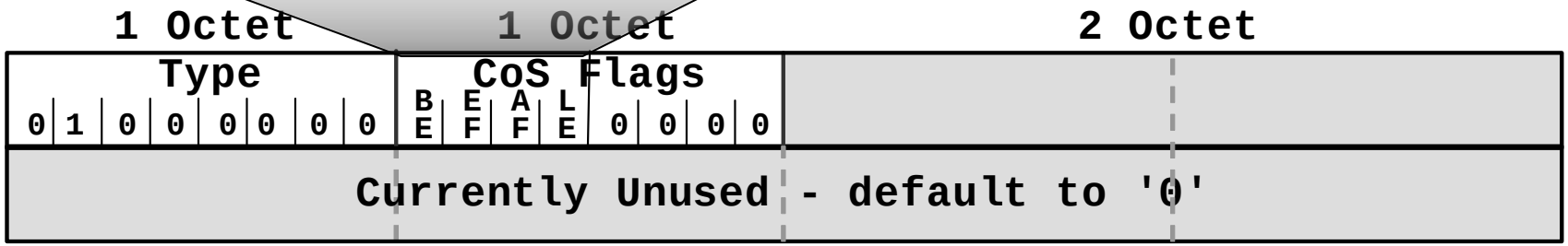


Definition of the CoS Capability Attribute

Ext. Community Attribute http://tools.ietf.org/html/draft-knoll-idr-cos-interconnect-00

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



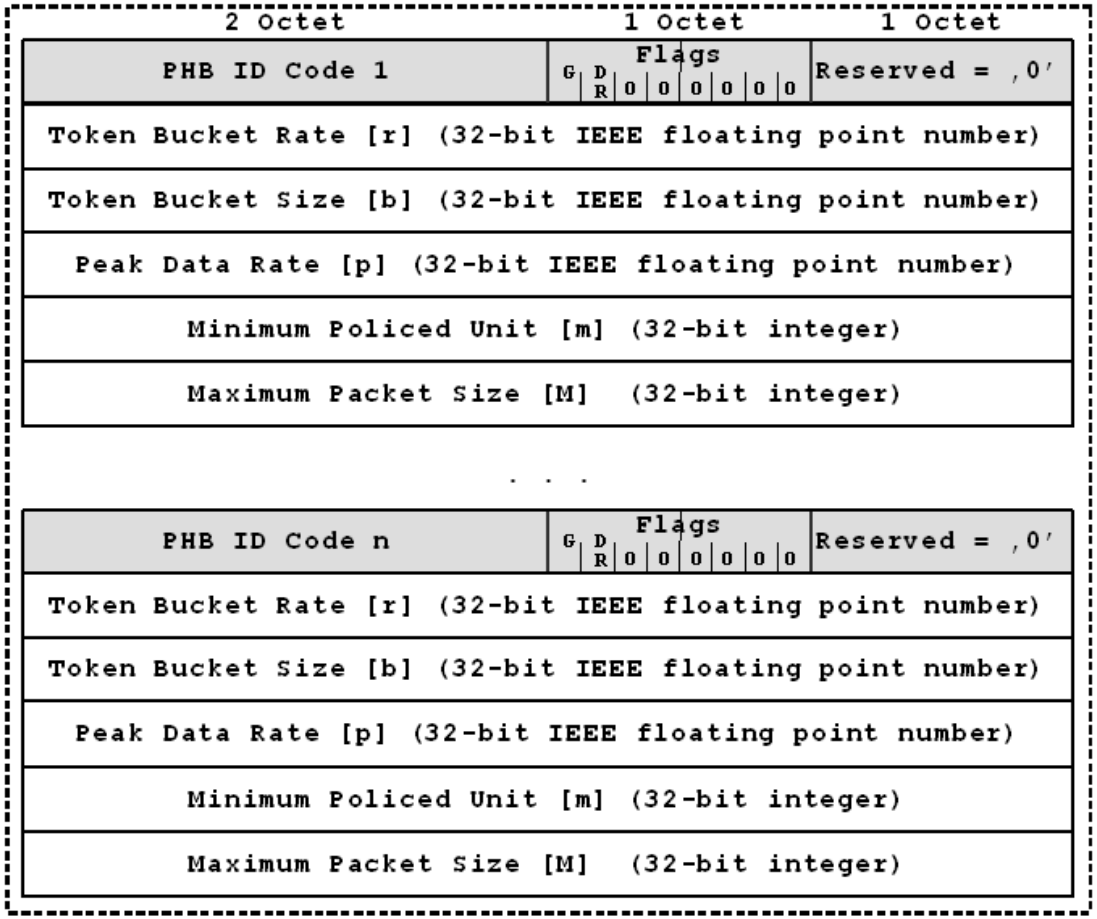
Definition of the CoS Parameter Attribute

Ext. Community Attribute http://tools.ietf.org/html/draft-knoll-idr-cos-interconnect-00

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

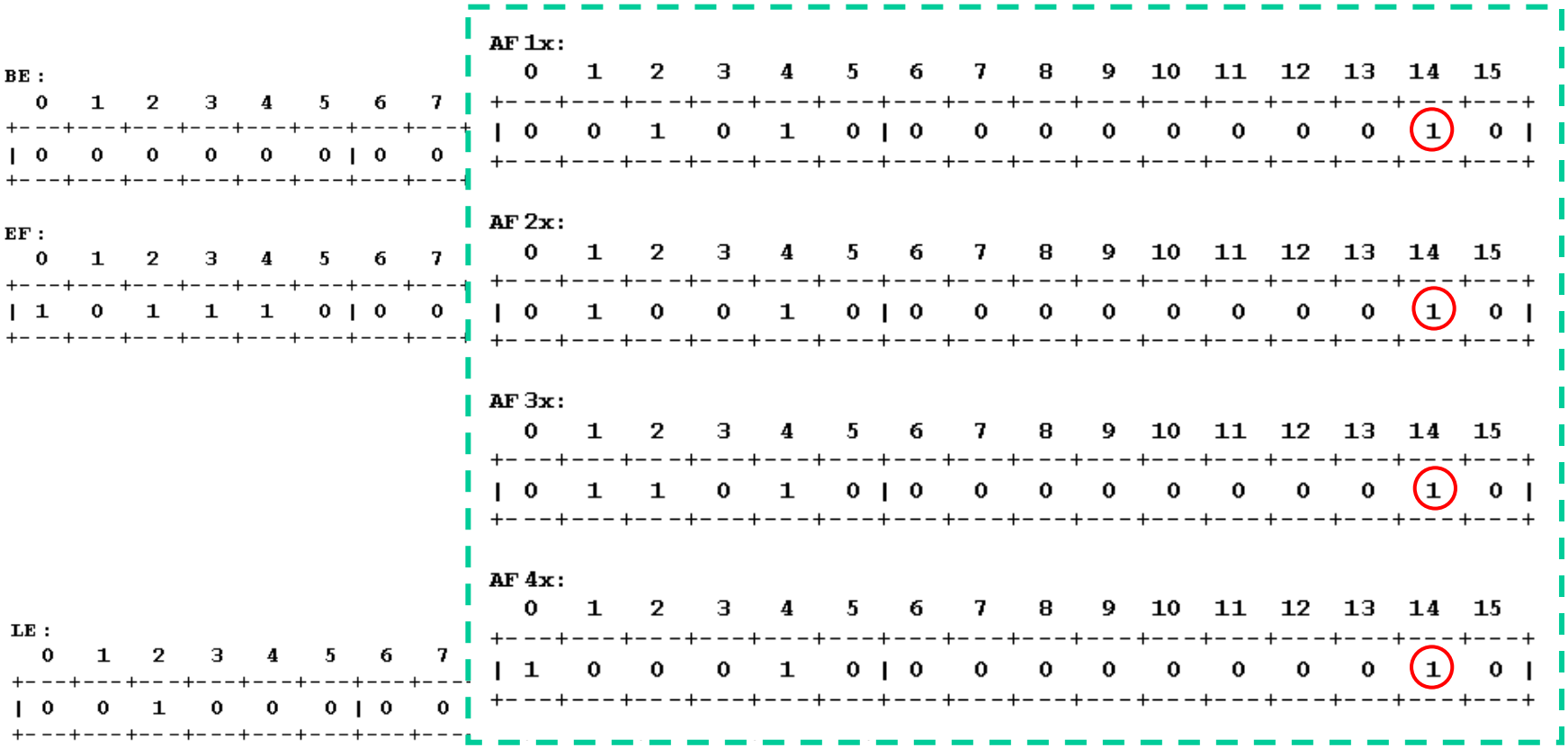
G flag … globally or NLRI local

DR flag… drop / remarking



Definition of the CoS Parameter Attribute

Ext. Community Attribute http://tools.ietf.org/html/draft-knoll-idr-cos-interconnect-00



Definition of the CoS Cap./Parameter Attribute

Optional non-transitive Attribute

- Smooth integration and locally used at AS peering points
- Ignoring of those attributes does not introduce instability or other harm

Binary signalling of the supported classes

- Simple processing of the few bit flags
- PHB groups should relate to the QoS markings, if both approaches are combined

Token Bucket parameters + consequences

- Clear and understandable mechanism to prevent class overload and perform predictable punishment on excess traffic.

Global/local association

- Simplify the signalling of generally applied classes and measurement parameters

Remarks

General comments

- Distinction between direct peering and transit peering (avoid remarking)
  → favour tunnelled transport for transit traffic
- Define a general Technology Type enumeration for cross-protocol (service) consistent numbering -> started
- L1 priority -> encompass QoS path/media selection for seamless interworking with optical and radio networks
- Usage of the Class Set information and the processing count analysis for the best path selection process -> ongoing debate about process changes and multi-path inter-AS peerings
- Usage of the Class Set information and the processing count analysis for PCE calculations
- High need for a consistent Class of Service concept

Remarks

Class Set Definition

1. Best solution: fixed standard including metering, enforcement and allocation
   e.g. using ITU parameters [Y.1541]
2. Free choice + class signalling + recommendations
   e.g. using “Configuration Guidelines for DiffServ Service Classes”, [RFC4594]
3. Free choice without signalling (confidential status) → not wanted
4. No Class Set support → not wanted

Class Mapping / Encoding Mapping

1. Best solution: fixed cross-layer standard including encoding and mapping
2. Free choice + class encoding signalling -> DSCP as anchor point
   (eases tunnelling and provides “inferred” QoS treatment
3. Free choice without signalling (confidential status) → not wanted
4. No cross-layer Class Set support → not wanted

Remarks

http://www.iana.org/assignments/bgp-extended-communities

Conflicting number assignment until October 16th

Border Gateway Protocol (BGP) Data Collection Standard Communities

Registry Name: Two-octet AS Specific Extended Community
Reference: [RFC4360]

Range	Registration Procedures
0x0000-0x00ff Transitive communities	First Come First Served
0x4000-0x40ff Non-transitive communities	First Come First Served

Registry:

Type Value	Name	Reference
0x0002	two-octet AS specific Route Target	[RFC4360]
0x0003	two-octet AS specific Route Origin	[RFC4360]
0x0005	OSPF Domain Identifier	[RFC4577]
0x0008	BGP Data Collection	[RFC4384]
0x0009	Source AS	[draft-ietf-l3vpn-2547bis-mcast-bgp]

0x00	QoS Marking Attribute	[Knoll]	2008-06-10

Likely to become
- **0x04** for **QoS Marking Attribute**
- **0x44** for **CoS Capability Attribute**

Reason:
- 0x00 … 0x03 in use as high bytes
- 0x04 & 0x44 sort of correspond

Registry Name: BGP Extended Communities Type - regular, non-transitive
Reference: [RFC4360]

Range	Registration Procedures
0xd0-0xff	Standards Action/Early IANA Allocation
0x40-0x7f	First Come First Served

Registry:

Type Value	Name	Reference	Registration Date
0x40	CoS Capability Attribute	[Knoll]	2008-07-09

Remarks

Quagga Extended Community implementation

- Linux routing suite provides Extended Community Attribute support.
- Currently, no support for regular type ext. communities
- Currently, AS4 encoding high byte value not officially assigned as yet

```
#define ECOMMUNITY_ENCODE_AS4          0x02
```

- Modified Quagga source will be available at bgp-qos.org

Summary

- The proposed approach enables a general QoS based forwarding which allows for **informed marking and possibly routing** decisions. It is **optimized for ease of deployment** and adopted to the current inter-AS forwarding model.
- The concept aims for a **consistent and widely adopted CoS approximation**, which encompasses cross-layer and cross-domain traffic class handling from L1 to at least L3 as generally offered CoS treatment.
- The concept **incorporates a confidentiality option** that allows operators the distinction between an secluded internal and the advertised external CoS support.
- **More sophisticated QoS concepts** are **not prohibited** and will always exist, which results in future “better quality islands/path”.

Summary cont.

Vision:

2 ... 4 class traffic separation Internet globally available & open to everyone

Interest in this concept can be expressed as follows

- Direct email contact
- RIPE routing wg email lists
- IETF IDR email list

Thank you –

Questions ?