Methods of interconnecting MPLS Networks

NANOG31, May 2005
San Francisco

Cable & Wireless
Internet Engineering
Udo Steinegger

What this talk is about

- **General**
  - This presentation covers technologies on how to possibly interconnect MPLS networks of different carriers, that support RFC2547bis VPN‘s
  - Little view in the future
    - what is going on at IETF regarding MPLS VPN‘s
    - considerations
- **Report from the real life implementation that C&W have done**
  - The interconnection method that C&W have chosen and why
  - The issues that C&W have found during the implementation
  - How the issues have been addressed.

Agenda

- **Technologies to interconnect MPLS networks that support VPN‘s according to RFC2547bis**
- **A real life report**
  - The method C&W have chosen and why.
  - Details - bits & pieces

Agenda (cont‘d)

- **Stuff C&W considers to support soon**
  - Carrier supporting Carrier (CSC)
  - OAM support
  - PWE3
- **Future Stuff**
  - Current actions from the IETF working groups
- **Questions**

Multi-AS Backbone Interconnections

VRF-to-VRF connections at the AS Border Routers

- **Using this architecture, a PE router in one AS attaches directly to a PE router in another AS.**
- **The two PE routers will be attached by multiple sub-interfaces (at least one for each of the VPNs spanning both AS's).**
- **Each PE router treats the other PE router as if it were a CE router and attaches a VRF to the sub-interface.**
- **Any iBGP-learned prefixes associated with that VRF are subsequently advertised in eBGP as an unlabelled prefix to the other PE.**
  - No requirement for MPLS support between the PE routers
- **Relies on OSI layer 2 for VPN separation (Frame Relay, ATM, 802.1q VLAN).**

Multi-AS Backbones

Multi-hop eBGP redistribution of labelled VPNv4 routes between AS's, with eBGP redistribution of labelled IPv4 routes from AS to neighbouring AS.

- **Using this architecture, VPNv4 prefixes are not advertised by ASBRs.**
- **An ASBR must maintain labelled IPv4 /32 routes to the PE routers within its AS. It uses IPv4 eBGP to distribute these routes to other AS's.**
  - Results in the creation of a label switched path from the ingress PE router to the egress PE router.

Multi-AS Backbones

Multi-hop eBGP redistribution of labelled VPNv4 routes between AS's, with eBGP redistribution of labelled IPv4 routes from AS to neighbouring AS. (cont'd)

- **PE routers in different AS's can subsequently establish multi-hop eBGP sessions to each other and exchange labelled VPNv4 prefixes over those connections.**
  - Potential to use multi-hop eBGP sessions between Route-Reflectors in each AS to avoid meshing issues.

Multi-AS Backbones

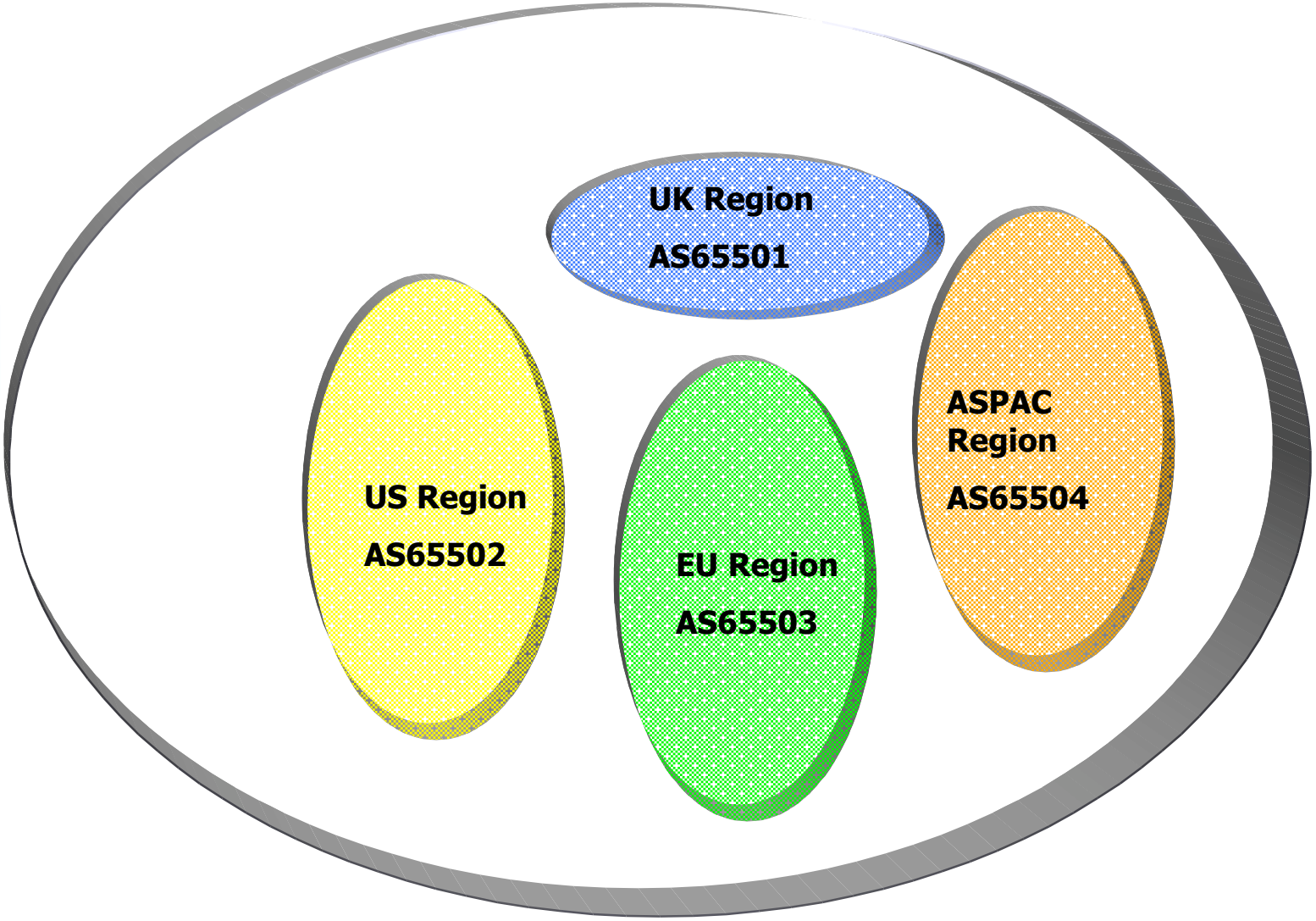
eBGP redistribution of labelled VPNv4 prefixes from AS to neighbouring AS

- **Using this architecture, the ASBR learns VPNv4 prefixes within its AS through iBGP.**
- **The ASBR then uses eBGP to redistribute those labelled VPNv4 prefixes to an ASBR in another AS, which in turn distributes them to the PE routers in that AS using iBGP.**
- **Label-switched interface exists between ASBRs.**
  - ASBR should never accept a labelled packet from an eBGP peer unless it actually distributed the top label to that peer.

What C&W has chosen for their MPLS network - a real life report

- C&W has built and operates a seperate global MPLS network for IP VPNs.
- This network is not part of C&W‘s public IP network, though there are links in between these.
- eBGP redistribution of labelled VPNv4 prefixes from AS to neighbouring AS

Confederated AS4445



Details - bits and pieces

On the following pages there's a discussion in detail about things that we had to think of when planning and testing the interconnection methods.

The most important things we have found are things:

- **Route distinguisher values**
- **Route target values**
- **Route target filtering**
- **QoS Continuity**
- **Resilience**
- **Security**

Details - Bits & Pieces

Route distinguisher numbering

- **When prepended to an IPv4 prefix, it is the 8-byte Route-Distinguisher that makes a unique VPNv4 prefix.**
  - Providing the Route-Distinguishers themselves are unique
- **Cable & Wireless Route-Distinguishers take the form**
  - Global Administrator sub-field (2 octets)
    - Autonomous System Number (ASN) assigned by IANA. The C&W IP-VPN has the registered ASN 4445
  - Local Administrator sub-field (4 octets)
    - The organisation identified by the Global Administrator sub-field can encode any information in this field. C&W assigns unique decimal integers to each VRF within the IP-VPN.
- **When interconnecting with other Service Providers MPLS-VPN networks, we should attempt to ensure that a single Route-Distinguisher is used across both Autonomous Systems.**

Details - Bits & Pieces

Route distinguisher numbering

- **To understand rationale for this, it is necessary to understand th BGP decision process when PE routers receive a VPNv4 prefix**
  - Take all routes with the same Route-Target as any of the configur "import" statements within the VRF.
  - Consider all routes that have the same Route-Distinguisher as the one assigned to the VRF being processed.
  - Create new BGP paths with a Route-Distinguisher that is equal to the Route-Distinguisher configured for the VRF that is being processed.
  - All routes are now comparable, and at the point the conventional BGP path selection algorithm can be executed.

- **Point 3 is critical……create new BGP paths……..**
- **BGP prefixes consume valuable (finite) PE memory. With differe Route-Distinguishers**
  - 1 BGP prefix consumes 2 BGP prefixes worth of memory
  - 1000 BGP prefixes consumes 2000 BGP prefixes worth of memo
  - etc…….

Bits & Pieces

Route distinguisher numbering

- **Route-Distinguishers need to be agreed and reconciled between Service Providers.**
- **A simple proposal is as follows.**
  - If the customer is C&W customer whose reach is being extended through another SP's network, then a C&W Route-Distinguisher should be used.
  - If the customer is another SP's, and that customers reach is being extended through C&W's IP-VPN, then that SP's Route-Distinguisher should be used.
- **Unfortunately this does create some issues with provisioning systems that appear to "hard-code" Route-Distinguishers.**
  - Manually provision if necessary to avoid prefix duplication and needless PE memory consumption.

Bits & Pieces

Route Target Numbering

- **The BGP Extended Community attribute "Route-Target" is used to determine whether a prefix is accepted by other PE routers.**
- **Cable & Wireless Route-Targets take the form**
  - Global Administrator sub-field (2 octets)
    - Autonomous System Number (ASN) assigned by IANA. The C&W IP-VPN has the registered ASN 4445
  - Local Administrator sub-field (4 octets)
    - The organisation identified by the Global Administrator sub-field can encode any information in this field. C&W assigns unique decimal integers to each VRF within the IP-VPN.
- **it is possible to "re-write" Route-Target values.**
  - Why not only use our own Route-Target values within the C&W IP-VPN using this "Route-Target re-write" feature at the ASBR ?

Bits & Pieces

Route Target Numbering

- Potentially, we could…..
- So, we receive a VPNv4 prefix with a neighbouring SP's Route-Distinguisher and Route-Target….
  - Re-write the Route-Target(s) to a C&W Route-Target that our PE routers will import.
  - Impossible to re-write Route-Distinguisher
- C&W PE routers import VPNv4 prefix, but because SP's Route-Distinguisher is different from C&W Route-Distinguisher configured on PE routers, VPNv4 prefix must be duplicated.
  - Memory consumption issue.

- **Route-Target values need to be agreed and reconciled.**
- **Same proposal as for Route-Distinguisher values**
  - If the customer is C&W customer then a C&W Route-Target should be used.
  - If the customer is another SP's, then that SP's Route-Distinguisher should be used.

Bits & Pieces

Route Target Filtering

- **When a PE router receives a VPNv4 prefix with a Route-Target for which it has no configured "import" statements, it will silently discard the update.**
  - Known as Automatic Route Filtering (ARF).
- **When a neighbouring SP has a customer with "B" end requirements in C&W's IP-VPN, C&W will configure a VRF on the ASBR with the relative (agreed) Route-Target and Route-Distinguisher values.**
  - No requirement for Route-Target "export" statements on ASBR as both VPNv4 iBGP updates (within own AS) and VPNv4 eBGP updates (from neighbouring ASBR) already have BGP Extended Community attribute "Route-Target" attached.
  - Requirement for "import" statement only
  - All other VPNv4 prefixes discarded using ARF.
- **Some providers may (and do) use other Route-Target filtering mechanisms - beware !**

Bits & Pieces

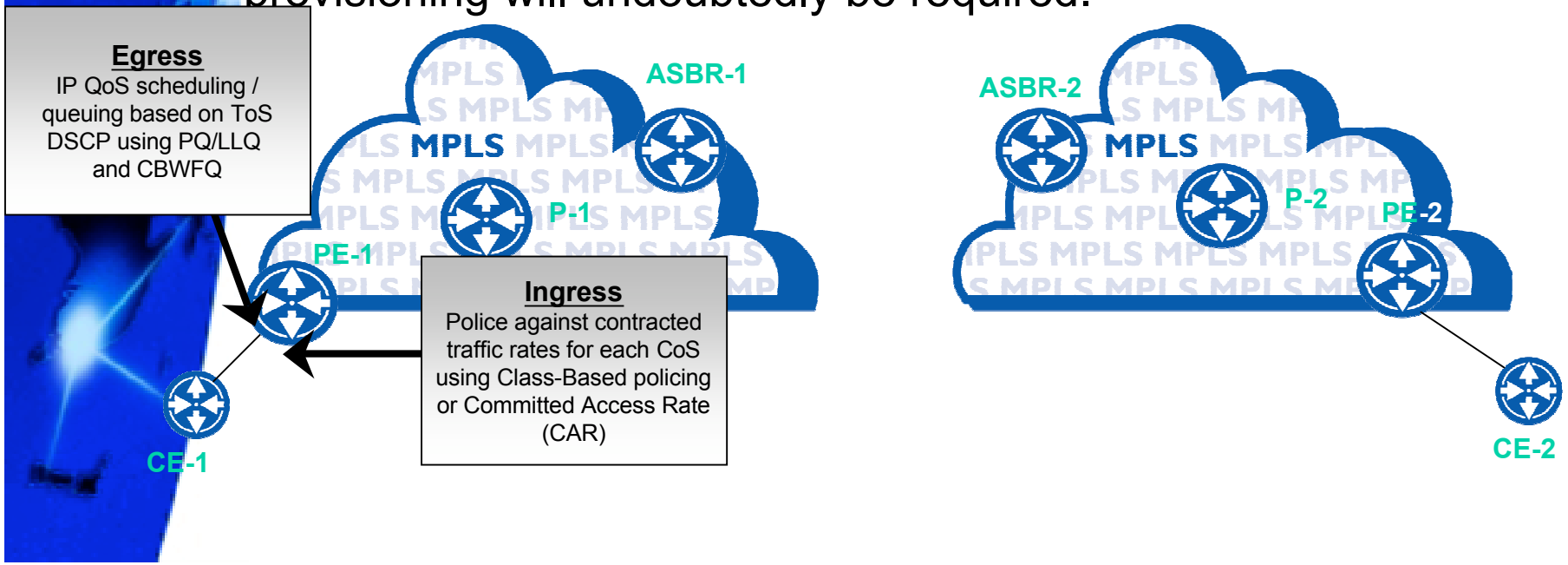
QoS Continuity

- **Each Service Provider operating an MPLS-VPN will offer a differing number of CoS, and will use different IP Precedence/DiffServ Codepoints to represent these CoS.**
  - But all will use common queuing/scheduling tools available in IOS to implement their QoS model (PQ, CBWFQ etc).
- **As far as C&W or other SP's MPLS-VPN networks are concerned, packet is exposed at IP layer twice only**
  - PE ingress
  - PE egress
- **Everything between these two points is label-switched, hence no manipulation of layer 3 fields (such as ToS/DSCP bits) is possible.**
  - No QoS mediation/reconciliation available at PE-ASBR

Bits & Pieces

QoS Continuity

- Implication of this is that C&W PE router must
  - Support other SP's IP QoS model on PE router egress port
  - Police against contracted traffic levels for each CoS on PE router ingress port
    - "B" end is effectively "unmanaged with QoS".
- Neither of the above should represent a problem, although manual provisioning will undoubtedly be required.



Bits & Pieces

QoS Continuity

Each Service Provider operating an MPLS-VPN will offer a differing number of CoS, and will use different IP Precedence/DiffServ Codepoints to represent these CoS.

- It follows therefore that traffic from a neighbouring PE-ASBR will have EXP bits set which are derived from the IP Precedence or DiffServ Codepoints in use in that Autonomous System

For example, assume SP#1 implements three CoS; Gold, Silver, and Bronze using IP precedence 5, 4, and 0 respectively.

- By default, PE-ASBR will not modify EXP bits, therefore traffic will be passed across C&W core with EXP bits 5, 4, and 0 for each CoS

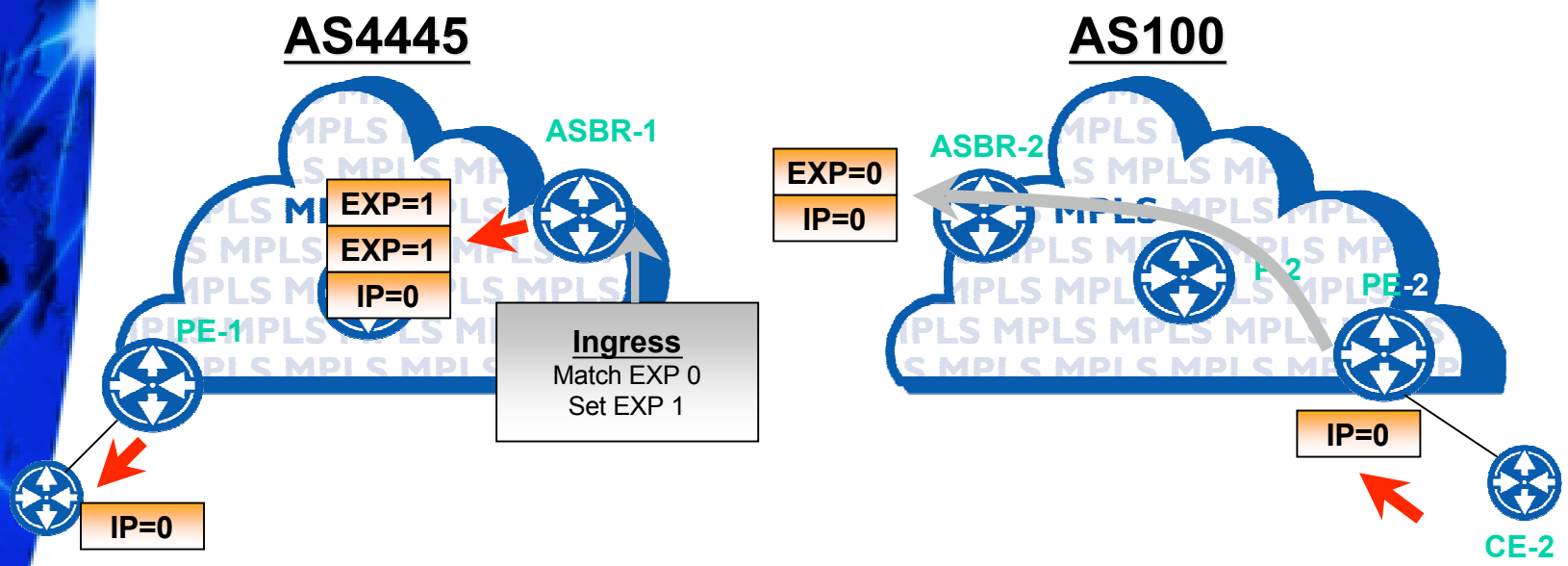
Currently, C&W do not implement queuing and/or MPLS traffic engineering in the core using EXP bits, however, it is likely that this will happen in the future

- Likely to be based around C&W classes of service using EXP 5, 3, and 1 respectively for Premium, Enhanced, and Standard

Bits & Pieces

QoS Continuity

- **Therefore, we need to ensure that SP#1's Gold/Silver/Bronze is mapped with C&W Premium/Enhanced/Standard to avoid SP#1s traffic receiving "unfair" treatment in C&W core in the presence of EXP-based queuing.**
- **EXP bit manipulation required at PE-ASBR**
  - Match EXP *x*--->Set EXP *y*



Bits & Pieces

QoS continuity

- **Manipulation of MPLS labels can only be done on ingress interfaces (not possible on egress interfaces)**
  - Responsibility therefore lies with SP to map EXP settings from neighbouring SP to those in use in own AS
  - Below example maps
    - EXP 5→EXP5 in Gold class (not really required but shown for clarity)
    - EXP 4→EXP3 in Silver class
    - EXP everything else→EXP 1 in Bronze class.

```
class-map match-all Gold
 match mpls experimental topmost 5
class-map match-all Silver
 match mpls experimental topmost 4
class-map match-all Bronze
 match mpls experimental topmost 0  1  2  3  6  7
!
policy-map EXP
 class Gold
  set mpls experimental topmost 5
 class Silver
  set mpls experimental topmost 3
 class Bronze
  set mpls experimental topmost 1
!
interface Serial2/0
 description E3 link to neighbouring PE-ASBR
 ip address 172.25.0.5 255.255.255.252
service-policy input EXP
```

“inbound” QoS policy

Bits & Pieces

Resilience

- **When interconnecting with other SP‘s networks multiple links can/might be considered and leaves some things to think of:**
  - Suboptimal
  - asymetric routing
  - Etc.

Bits & Pieces

Resilience

- **Most sub-optimal and asymmetric routing can be avoided by manipulating BGP attributes at the PE-ASBR (for example, BGP MED or Local-Pref).**
  - In order to “set” an attribute, we need to “match” against something in the BGP update first.
    - VPNv4 prefix. Matching against VPNv4 prefix is a) not possible in IOS today, and b) not scalable/manageable anyway.
    - Route-Target. Route-Targets are generally “VPN-wide”. Matching against BGP Extended Community Route-Target may resolve sub-optimal routing issues for Site#1-->Site#2, but break optimal routing Site#1-->Site#3
    - BGP Standard Community. Possible. CE routers could set BGP Standard Community value which PE-ASBRs could “match” against and manipulate BGP attributes. This is possible, however, don’t assume that all other Service Providers will pass BGP Standard Communities as only BGP Extended Communities are required in an MPLS-VPN.
- **Careful consideration needs to be given to where the ASBRs are located.**
  - If they are geographically close, then most of these issues can be avoided.

Bits & Pieces

Security

- **Security of the C&W IP-VPN is a major concern.**
  - If we connect our private network to another SP, it could be argued that it's no longer private !
- **From a security perspective, there are two major concerns**
  - The point-to-point link that connects the PE-ASBRs.
    - Requirement to protect from undesirable IP traffic sourced from within other SP's network.
    - Point-to-point link is label-switched. Therefore, deny all IP except BGP and ICMP.
  - Managed CE routers
    - If C&W managed CE router is in other SP's network, requirement is to protect CE from attack through serial interface.
      - Use Access-Lists in the same way as we do today on PE ingress
    - If other SP managed CE router is connected to C&W's network, requirement is to protect PE from attack through serial interface.
      - Use conventional Access-Lists on PE-->CE interface like we do today.

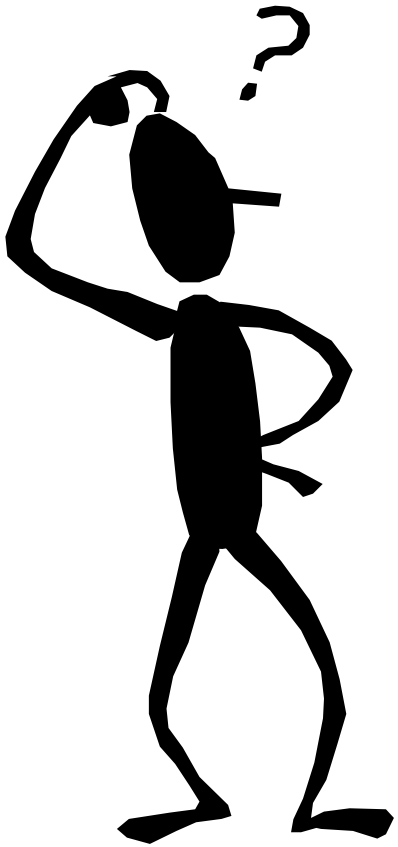
Stuff that C&W will support soon

- Carrier supporting Carrier (CSC)
- OAM support
- Draft Martini L2 VPNs

Future stuff

- **What is currently going on at IETF**
  - The most hot thing (as I see it) is in the working group PWE3
    - „Pseudo wire edge to edge emulation“
    - VCCV

Any Questions ???



Either write a mail to udo@cw.net

Or

Bring beer and talk to me :-)