A Carrier-Grade L2BSA Gateway with Netgraph

Lutz Donnerhacke
IKS Service GmbH
Personal background

• Grown up with a C64 in Eastern Germany
• Short excurse to MSDOS and Win3.11
• Switched to Linux in 1992
• Individual Network e.V. in Germany
  • grassroot Internet for ordinary people
• Build regional ISP in 1996
  • infrastructure provider of two regional Broadband-ISPs in Thuringia
• Switch to FreeBSD in 2013
  • Replace commercial LNS by netgraph/MPD
Broadband in Germany

• 1983 – CDU (German conservatives) choose copper over fiber
  • Probably due to corruption via the minister Schwarz-Schilling

• 1995 – CDU privatization created Deutsche Telekom
  • Got the whole cable network as gift, hence a (copper) monopoly

• 1998 – Unbundling of last mile to competitors

• 2009 – Vectoring for increased bandwith over copper (DSL)
  • All lines need to be handled by a single DLSAM
  • Unbundling not longer possible
  • To break the local monopoly, regulated bitstream access was introduced

• We are still a copper DSL based country!
  • Fiber is rolled out 2022 but expected to be available not before 2024
FTTH/B European Ranking as for September 2020

Including:
- Countries with more than 1% FTTH/B household penetration
- Economies with at least 125,000 Households

Source: IDATE DigiWorld for FTTH Council Europe - May 2021
Layer2 Bitstream Access (regulated)

• Solution for the layer1 monopoly created by vectoring
  • DSL port in sync: Hand over all layer2 traffic to a different ISP
  • Requires a transport layer which maps 1:1 to the port
  • No opportunity for smart networks (like Multicast distribution)
  • Problem: How to handle bandwidth discrepancy?

• Technical approach (A10NSP)
  • Add a VLAN tag to all traffic from the port (IEEE 802.1ad)
    Can only handle up to ~4000 customers per aggregation line
  • Add current line speed in DHCP option or PPP LCP option
    Shape downstream traffic to signaled speed (or pay fines)
Commercial aspects

• Numbers
  • 1 Mio households in Thuringia => ~260 interconnection points
  • About 50 interconnections needed for DTAG operated areas
  • Most rural regions are operated by us (not DTAG)

• You need a line to each area, where your customer is
  • pay for a 1G/10G WAN + interconnection fee
  • get a couple of customers switching to you

• We need a very cheap termination device
  • Commercially available: Large PPPoE appliances
  • We like to have DHCP and IPv6 in multiple VLANs (i.e. separate VoIP)
What happens on wire?

<table>
<thead>
<tr>
<th>Customer</th>
<th>A10NSP Interconnection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payload (IP)</td>
<td>Payload (IP)</td>
</tr>
<tr>
<td>C-VLAN Tag</td>
<td>C-VLAN Tag</td>
</tr>
<tr>
<td>Source-MAC</td>
<td>S-VLAN Tag</td>
</tr>
<tr>
<td>Destination-MAC</td>
<td>Source-MAC</td>
</tr>
<tr>
<td></td>
<td>Destination-MAC</td>
</tr>
</tbody>
</table>
What really happens on wire

<table>
<thead>
<tr>
<th>Customer</th>
<th>A10NSP Interconnection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payload (IP)</td>
<td>Payload (IP)</td>
</tr>
<tr>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>01:02:03:04:05:06</td>
<td>27</td>
</tr>
<tr>
<td>01:02:03:07:08:09</td>
<td>01:02:03:04:05:06</td>
</tr>
<tr>
<td></td>
<td>01:02:03:07:08:09</td>
</tr>
</tbody>
</table>
What really happens on wire?

<table>
<thead>
<tr>
<th>Customer</th>
<th>A10NSP Interconnection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payload (IP)</td>
<td>Payload (IP)</td>
</tr>
<tr>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>01:02:03:04:05:06</td>
<td>01:02:03:04:05:06</td>
</tr>
<tr>
<td>01:02:03:07:08:09</td>
<td>01:02:03:07:08:09</td>
</tr>
<tr>
<td></td>
<td>35</td>
</tr>
</tbody>
</table>

The table shows the payload details for both customer and A10NSP interconnection.
What really happens on wire!

- On every DSL resync
  - A new, dynamic S-VLAN is used
  - Traffic send to the old S-VLAN is lost immediately
- PPP will drop the line and resync
- DHCP does not even notice
## Simplify by adding complexity

### Table: QoS per interface and service

<table>
<thead>
<tr>
<th>Customer</th>
<th>A10NSP</th>
<th>Interface</th>
<th>Rotate</th>
<th>Service</th>
<th>Interface</th>
<th>Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payload</td>
<td>Payload</td>
<td>Payload</td>
<td>Payload</td>
<td>Payload</td>
<td>Payload</td>
<td>Payload</td>
</tr>
<tr>
<td>C-Vlan</td>
<td>C-Vlan</td>
<td>C-Vlan</td>
<td>S-Vlan</td>
<td>S-Vlan</td>
<td>S-Vlan</td>
<td>Src-MAC</td>
</tr>
<tr>
<td>Src-MAC</td>
<td>S-Vlan</td>
<td>S-Vlan</td>
<td>A-Vlan</td>
<td>A-Vlan</td>
<td>Src-MAC</td>
<td>Dst-MAC</td>
</tr>
<tr>
<td>Dst-MAC</td>
<td>Src-MAC</td>
<td>A-Vlan</td>
<td>C-Vlan</td>
<td>Src-MAC</td>
<td>Dst-MAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dst-MAC</td>
<td>Src-MAC</td>
<td>Src-MAC</td>
<td>Dst-MAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dst-MAC</td>
<td>Dst-MAC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **QoS per interface**
- **QoS per service**
- **Customer traffic handled per service**

*Static*
Todo list for netgraph

- **Blue/Red:** Existing nodes
- **Green:** Insufficient node
- **Orange:** Missing node
- **Yellow/Rose:** Userspace scripting
Necessary modifications to FreeBSD

• Mandatory
  • D21846: Add 802.1Q ethertypes used in the wild
  • D24179: ixl: Permit 802.1ad frames to pass though the chip
  • D22076: netgraph/ng_vlan_rotate: IEEE 802.1ad VLAN manipulation netgraph node type

• Bridging
  • D21803: netgraph/ng_bridge: Replace NG_BRIDGE_MAX_LINKS with unlimited links
  • D23840: netgraph/ng_base: Allow larger BINARY2ASCII conversions
  • D23963: netgraph/ng_bridge: Introduce "uplink" ports without MAC learning
  • D28123: netgraph/ng_bridge: become multithreaded

• QoS
  • D22140: netgraph/ng_tag: Variable length data can not be set for all length
  • D22110: netgraph/ng_car: Add color marking code

• Debugging
  • D21968: netgraph/ng_source: Allow ng_source to inject into any netgraph network
  • D21965: usr.sbin/ngctl: Generate more compact GraphWiz output
Real world problems

• Same MAC in different VLANs
  • Tagged and untagged from the same customer device (CPE)
  • Overlapping MACs from different CPEs due to MAC increase per VLAN

• Different CPE configuration
  • Same service may come in different C-Vlans (or untagged)
  • Need a decision mechanism to choose the right one

• Erroneous transmission of data rates
  • No DHCP exchange due to signaled 0 bps (enforced!)

• Broken interface cards (ixl) do not transmit 802.1ad
  • Customer untagged is an IEEE violation, but works
Real world services
Real world VoIP QoS
Real world interface
### Real world bridge table

<table>
<thead>
<tr>
<th>MAC</th>
<th>Link</th>
<th>Line</th>
<th>S-Vlan</th>
<th>C-Vlan</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:96:d7:46:fa:ec</td>
<td>hsi:link61</td>
<td>ixl0</td>
<td>2</td>
<td>140</td>
</tr>
<tr>
<td>08:96:d7:46:fa:ed</td>
<td>voip:link4249</td>
<td>ixl0</td>
<td>2</td>
<td>142</td>
</tr>
<tr>
<td>2c:91:ab:bb:4b:76</td>
<td>hsi:link7554</td>
<td>ixl10</td>
<td>10</td>
<td>140</td>
</tr>
<tr>
<td>2c:91:ab:bb:4b:77</td>
<td>voip:link3536</td>
<td>ixl10</td>
<td>10</td>
<td>142</td>
</tr>
<tr>
<td>3c:a6:2f:fe:ac:03</td>
<td>voip:link2299</td>
<td>ixl10</td>
<td>18</td>
<td>142</td>
</tr>
<tr>
<td>3c:a6:2f:fe:ac:02</td>
<td>hsi:link3507</td>
<td>ixl10</td>
<td>18</td>
<td>untagged</td>
</tr>
<tr>
<td>2c:3a:fd:dd:39:2a</td>
<td>hsi:link3588</td>
<td>ixl1</td>
<td>49</td>
<td>140</td>
</tr>
<tr>
<td>2c:3a:fd:dd:39:2b</td>
<td>voip:link5014</td>
<td>ixl1</td>
<td>49</td>
<td>142</td>
</tr>
<tr>
<td>00:01:21:01:44:2d</td>
<td>hsi:link9805</td>
<td>ixl8</td>
<td>5</td>
<td>180</td>
</tr>
<tr>
<td>00:01:21:01:44:2d</td>
<td>voip:link4426</td>
<td>ixl8</td>
<td>5</td>
<td>182</td>
</tr>
</tbody>
</table>
Questions?

https://lutz.donnerhacke.de/Blog
https://www.thyotec.de/
https://www.netkom.de/Privatkunden/Internet