



Como otimizar o balanceamento de serviços L2VPN em redes metro utilizando o FAT PW

20 de março de 2020

ÍNDICE DA APRESENTAÇÃO

- Sobre a PS Network Experts
- Quem sou eu
- Evolução das redes metro
- O problema
- ECMP: escolha do caminho
- Solução para otimizar o balanceamento
- Demonstração



SOBRE A PS



A PS Network Experts é uma prestadora de serviços profissionais que atua nas áreas de:



REDES DE DADOS E VOZ



PROVEDORES DE SERVIÇO DE INTERNET



DATA CENTER



SEGURANÇA



COLABORAÇÃO



Diferencial: desenvolvemos, implantamos e suportamos projetos **em ambiente multivendor, sem qualquer restrição a fabricantes**

- Trabalha há 7 anos na PS Network Experts
- Há mais de 10 anos desenvolve e implanta projetos de backbone IP/MPLS e redes metro em grandes operadoras, operadoras regionais e provedores
- Detentor de certificações Cisco CCIE



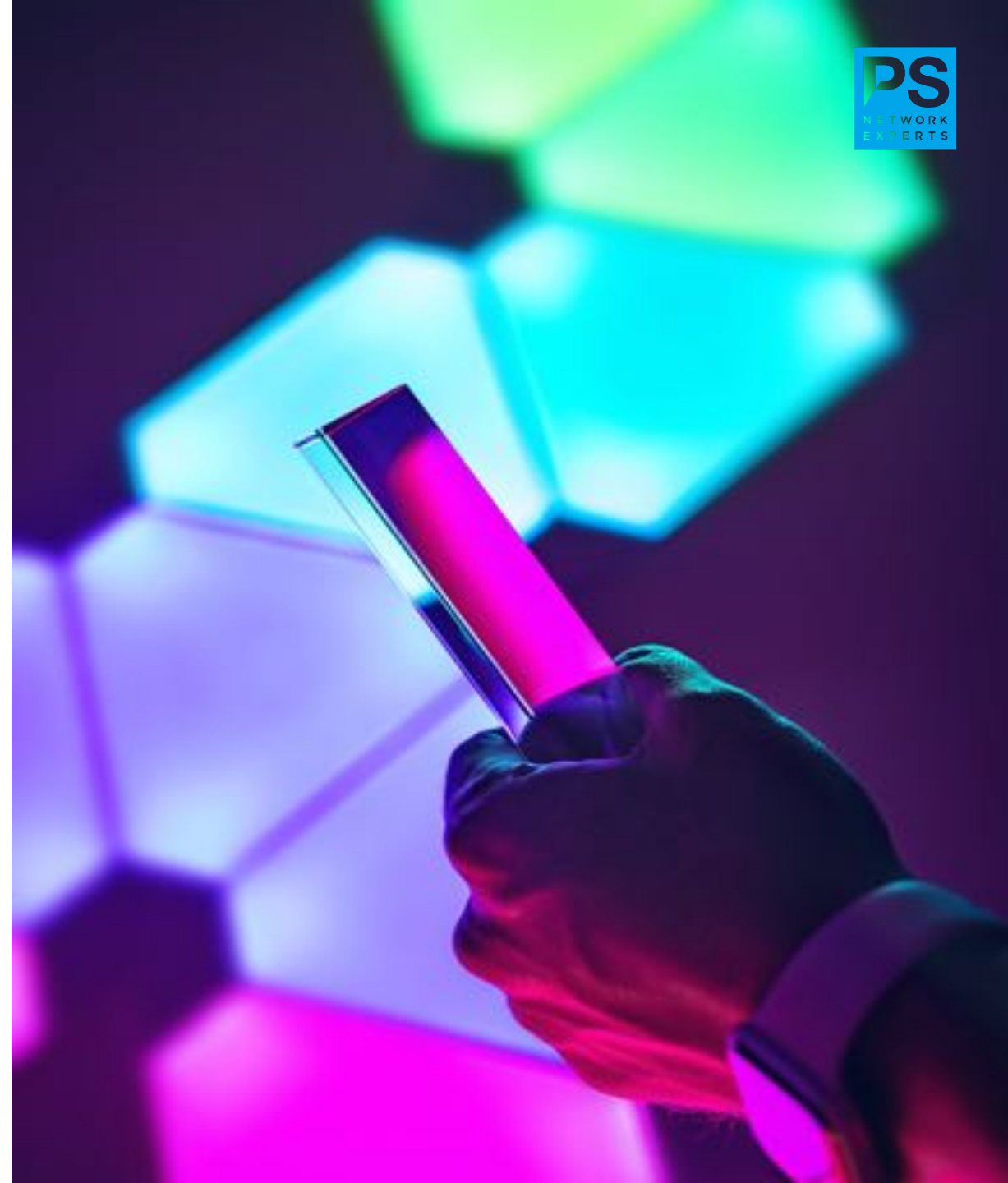
Evolução das redes metro

- Migração do rádio para fibra obrigou a implementação de redes metro
- A redução do custo de aquisição de equipamentos permitiu a implementação de MPLS nas redes metro
- A concentração de serviços em roteadores de borda e em BRAS foi simplificada pela utilização de VPWS ou VPLS
- Devido ao alcance de interfaces de 10Gbps o custo de iluminar fibras com switches mostrou-se muito eficiente

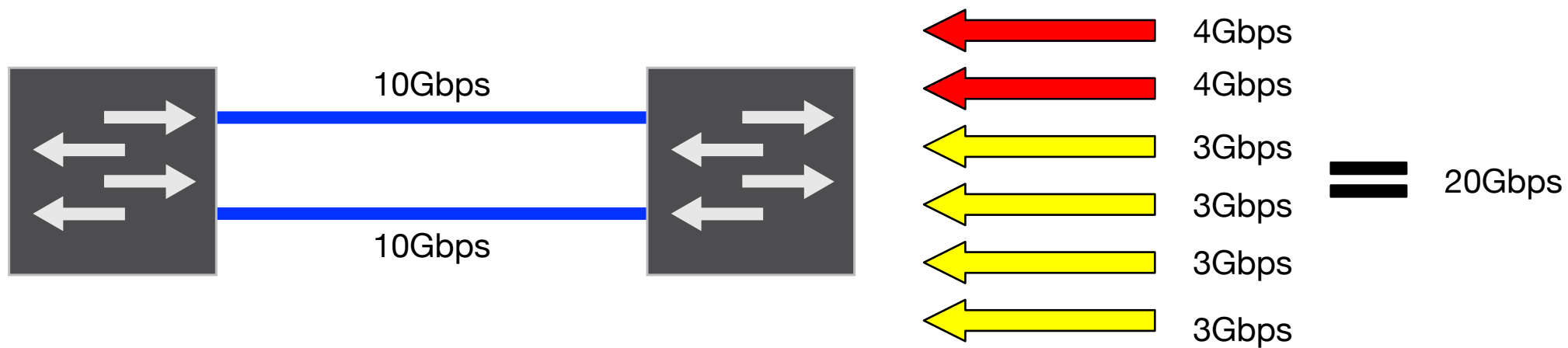


O problema

- Crescimento exponencial da demanda por mais banda
- PW individuais consumindo cada vez mais banda
- Congestionamento de links, mesmo existindo caminhos paralelos (ineficiência do balanceamento de tráfego)



Cenário do problema



O caminho, quando há caminhos paralelos (ECMP ou Equal Cost Multi-Path), é determinado pelo cálculo do hash

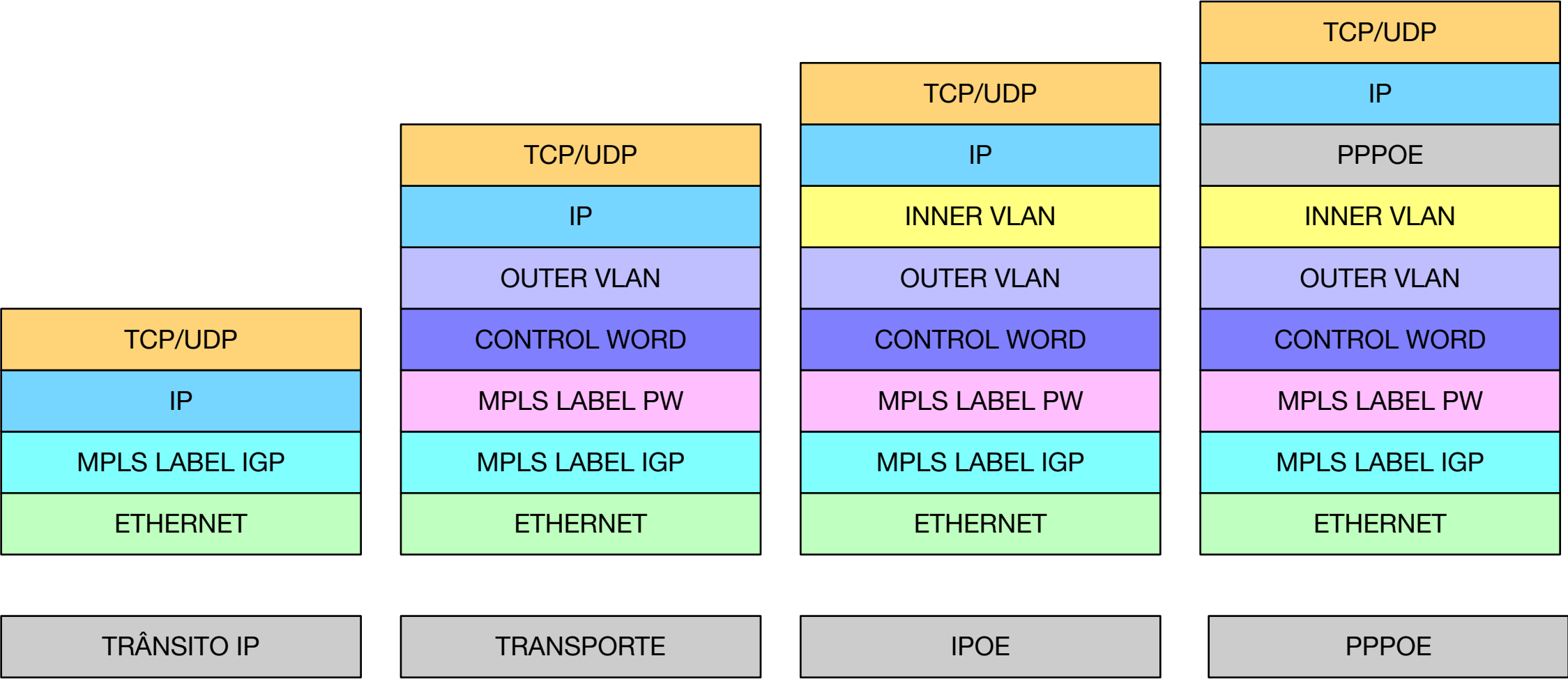
- O hash é computado sobre campos dos pacotes
- Os campos escolhidos dependem do tipo de tráfego sendo transportado
- O tráfego IP oferece muita variação (fluxos), principalmente porque as redes de provedores e operadoras possuem centenas de milhares de fluxos
- Balanceamento baseado em IP (layer 3), normalmente, permite balanceamento otimizado

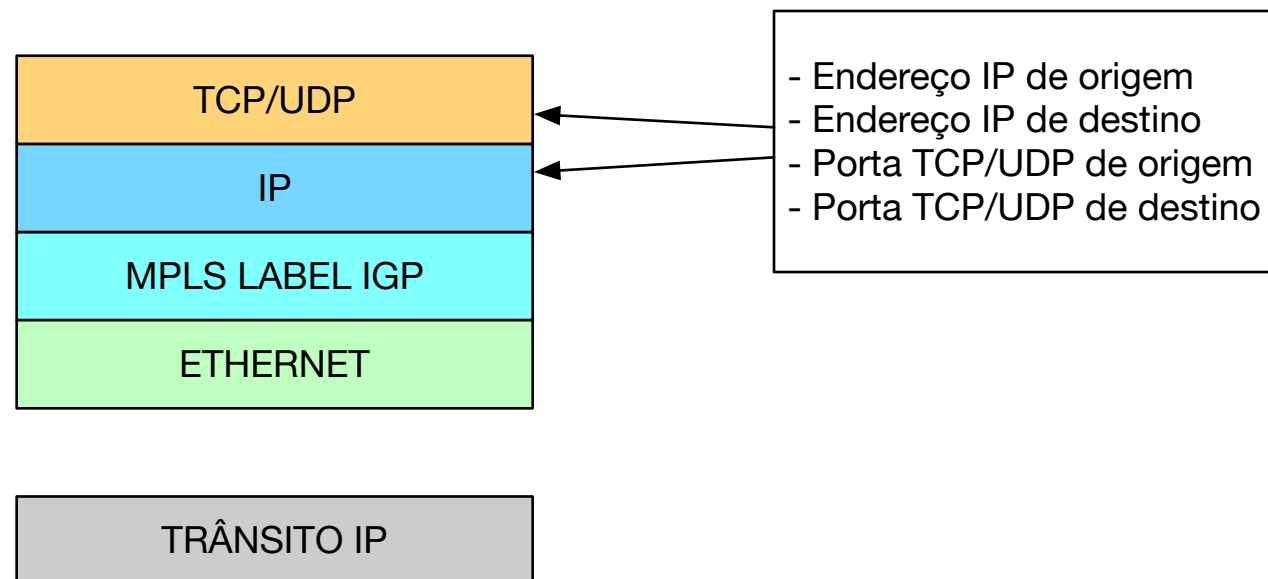
Escolha dos campos para computação do hash*

Tipo do tráfego	Campos utilizados para computação do hash
IPv4	Endereço de origem, endereço de destino, porta de origem e porta de destino
IPv6	Endereço de origem, endereço de destino, porta de origem e porta de destino
IP sobre MPLS com menos de 4 labels	Endereço de origem, endereço de destino, porta de origem e porta de destino
IP sobre MPLS com 4 ou mais labels	Último label da pilha (inner most label)
Pacote não IP	Último label da pilha (inner most label)

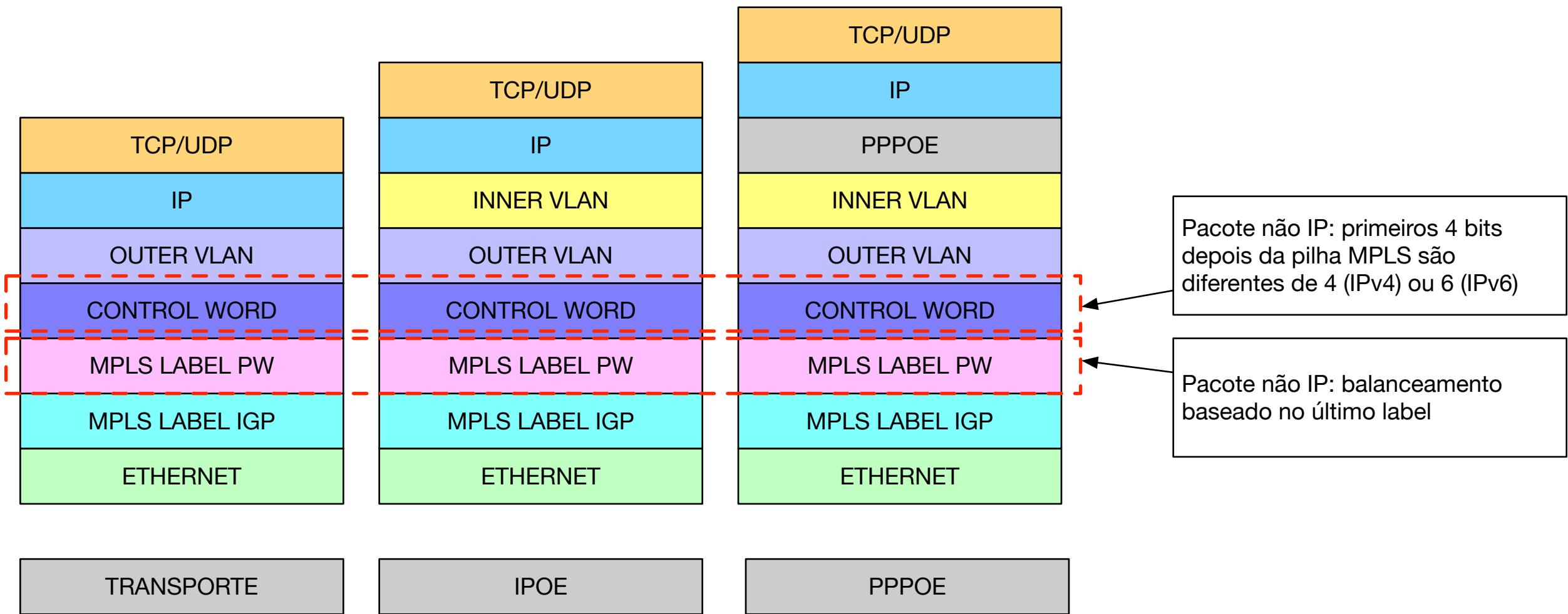
*<https://community.cisco.com/t5/service-providers-documents/asr9000-xr-load-balancing-architecture-and-characteristics/ta-p/3124809#field>

Exemplos de pilhas de protocolos





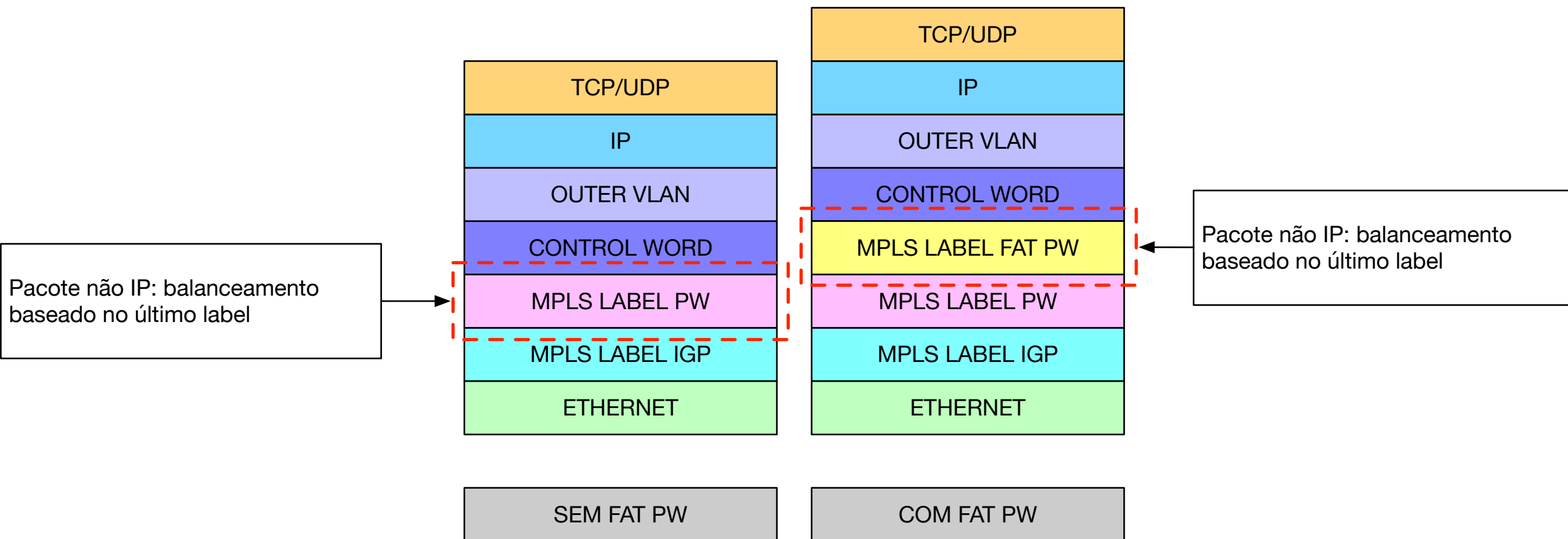
O que são pacotes não IP



RFC 6391 FAT PW (Flow-Aware Traffic of Pseudowires)

- Dependente do hardware nos equipamentos de borda
- Transparente para nós intermediários (roteadores P e outros PE)
- Poucos comandos para ativar
- Adiciona um label MPLS extra, cujo valor é baseado nos campos do pacote original (cada fabricante possui seu algoritmo)
- Equipamento que termina o túnel (PW) descarta o label extra

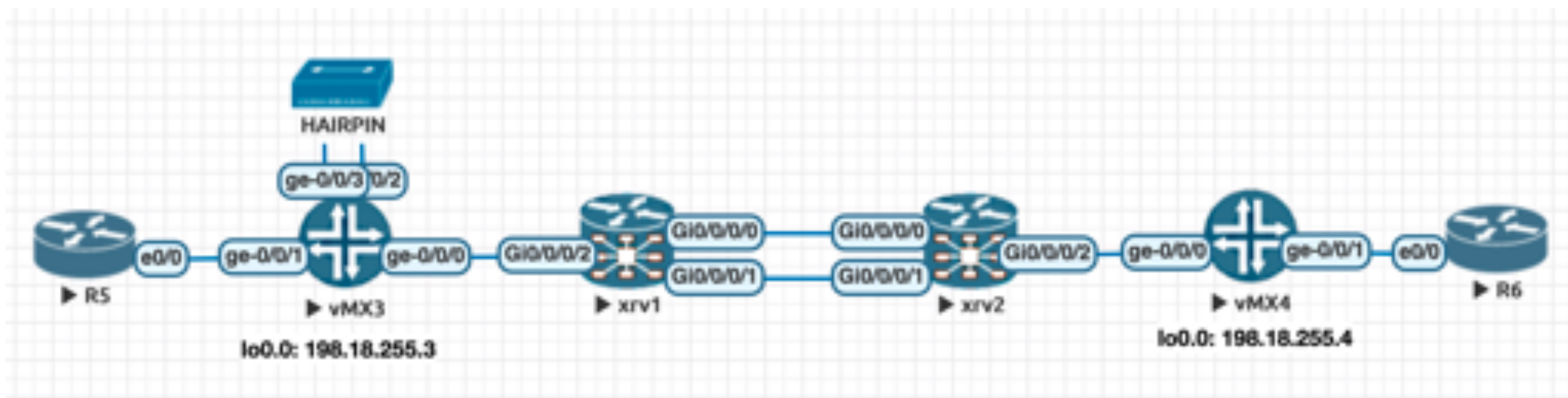
Solução para otimizar o balanceamento (pacotes com e sem o FAT PW)





Demonstração

- EVE-NG
- Juniper vMX (roteador PE)
- Cisco XRv (roteador P)
- Cisco IOL (gerador de tráfego)
- Wireshark (captura de pacotes)




```
root@vMX3> show l2circuit connections
```

```
Layer-2 Circuit Connections:
```

```
Neighbor: 198.18.255.4
```

Interface	Type	St	Time last up	# Up trans
ge-0/0/3.0(vc 500)	rmt	Up	Mar 9 19:44:42 2020	1

```
Remote PE: 198.18.255.4, Negotiated control-word: Yes (Null)
```

```
Incoming label: 299824, Outgoing label: 299824
```

```
Negotiated PW status TLV: No
```

```
Local interface: ge-0/0/3.0, Status: Up, Encapsulation: ETHERNET
```

```
Flow Label Transmit: No, Flow Label Receive: No
```

```
RP/0/0/CPU0:xrv1#show mpls forwarding labels 16002
```

```
Mon Mar 9 20:18:26.046 UTC
```

Local Label	Outgoing Label	Prefix or ID	Outgoing Interface	Next Hop	Bytes Switched
16002	16002	198.18.255.4/32	Gi0/0/0/0	198.18.0.1	36226
	16002	198.18.255.4/32	Gi0/0/0/1	198.18.0.3	21534372

```
RP/0/0/CPU0:xrv1#show mpls forwarding exact-route label 16002 bottom-label 299824
```

```
Mon Mar 9 20:15:45.237 UTC
```

Local Label	Outgoing Label	Prefix or ID	Outgoing Interface	Next Hop	Bytes Switched
16002	16002	198.18.255.4/32	Gi0/0/0/1	198.18.0.3	N/A

```
Version: 10, Priority: 1
```

```
Via: Gi0/0/0/1, Next Hop: 198.18.0.3
```

```
MAC/Encaps: 14/18, MTU: 1586
```

```
Label Stack (Top -> Bottom): { 16002 }
```

Wireshark interface showing a packet capture filter: `ip.src == 10.5.5.0/24`

No.	Time	Source	Destination	Protocol	Length	Info
7491	684.003292	10.5.5.114	172.31.6.114	UDP	1076	50114 → 50114 Len=1000
7492	684.103080	10.5.5.115	172.31.6.115	UDP	1076	50115 → 50115 Len=1000
7493	684.201535	10.5.5.116	172.31.6.116	UDP	1076	50116 → 50116 Len=1000
7494	684.302688	10.5.5.117	172.31.6.117	UDP	1076	50117 → 50117 Len=1000
7495	684.401830	10.5.5.118	172.31.6.118	UDP	1076	50118 → 50118 Len=1000
7496	684.503175	10.5.5.119	172.31.6.119	UDP	1076	50119 → 50119 Len=1000
7498	684.603600	10.5.5.120	172.31.6.120	UDP	1076	50120 → 50120 Len=1000
7499	684.702755	10.5.5.121	172.31.6.121	UDP	1076	50121 → 50121 Len=1000
7500	684.802224	10.5.5.122	172.31.6.122	UDP	1076	50122 → 50122 Len=1000
7501	684.900506	10.5.5.123	172.31.6.123	UDP	1076	50123 → 50123 Len=1000
7502	685.000747	10.5.5.124	172.31.6.124	UDP	1076	50124 → 50124 Len=1000

Frame 1: 1076 bytes on wire (8608 bits), 1076 bytes captured (8608 bits) on interface 0

- Ethernet II, Src: 50:00:00:01:00:02 (50:00:00:01:00:02), Dst: 50:00:00:02:00:02 (50:00:00:02:00:02)
- MultiProtocol Label Switching Header, Label: 16002, Exp: 0, S: 0, TTL: 254
- MultiProtocol Label Switching Header, Label: 299824, Exp: 0, S: 1, TTL: 255
- PW Ethernet Control Word
- Ethernet II, Src: aa:bb:cc:00:50:00 (aa:bb:cc:00:50:00), Dst: aa:bb:cc:00:60:00 (aa:bb:cc:00:60:00)
- 802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 500
- 802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 1005
- Internet Protocol Version 4, Src: 10.5.5.174, Dst: 172.31.6.174
- User Datagram Protocol, Src Port: 50174, Dst Port: 50174
- Data (1000 bytes)

Packet bytes: 50 00 00 02 00 02 50 00 00 01 00 02 88 47 03 e8 P...P...G..

Wireshark_vun0_1_2-20200309-165303_20200309165308_ZVioYypcapng

Packets: 7502 - Displayed: 6851 (91.3%)

Profile: Default


```
root@vMX3# show | compare
[edit protocols l2circuit neighbor 198.18.255.4 interface ge-0/0/3.0]
+      flow-label-transmit;
+      flow-label-receive;
```

```
root@vMX4# show | compare
[edit protocols l2circuit neighbor 198.18.255.3 interface ge-0/0/1.0]
+      flow-label-transmit;
+      flow-label-receive;
```

Wireshark capture window showing a packet list and packet details for a flow.

Filter: `ip.src == 10.5.5.0/24`

No.	Time	Source	Destination	Protocol	Length	Info
364...	3586.102344	10.5.5.135	172.31.6.135	UDP	1080	50135 → 50135 Len=1080
364...	3586.201556	10.5.5.136	172.31.6.136	UDP	1080	50136 → 50136 Len=1080
364...	3586.302253	10.5.5.137	172.31.6.137	UDP	1080	50137 → 50137 Len=1080
364...	3586.401971	10.5.5.138	172.31.6.138	UDP	1080	50138 → 50138 Len=1080
364...	3586.600829	10.5.5.140	172.31.6.140	UDP	1080	50140 → 50140 Len=1080
364...	3586.902129	10.5.5.143	172.31.6.143	UDP	1080	50143 → 50143 Len=1080
364...	3587.100970	10.5.5.145	172.31.6.145	UDP	1080	50145 → 50145 Len=1080
364...	3587.301765	10.5.5.147	172.31.6.147	UDP	1080	50147 → 50147 Len=1080
364...	3587.400315	10.5.5.148	172.31.6.148	UDP	1080	50148 → 50148 Len=1080
364...	3587.500686	10.5.5.149	172.31.6.149	UDP	1080	50149 → 50149 Len=1080
364...	3587.701893	10.5.5.151	172.31.6.151	UDP	1080	50151 → 50151 Len=1080

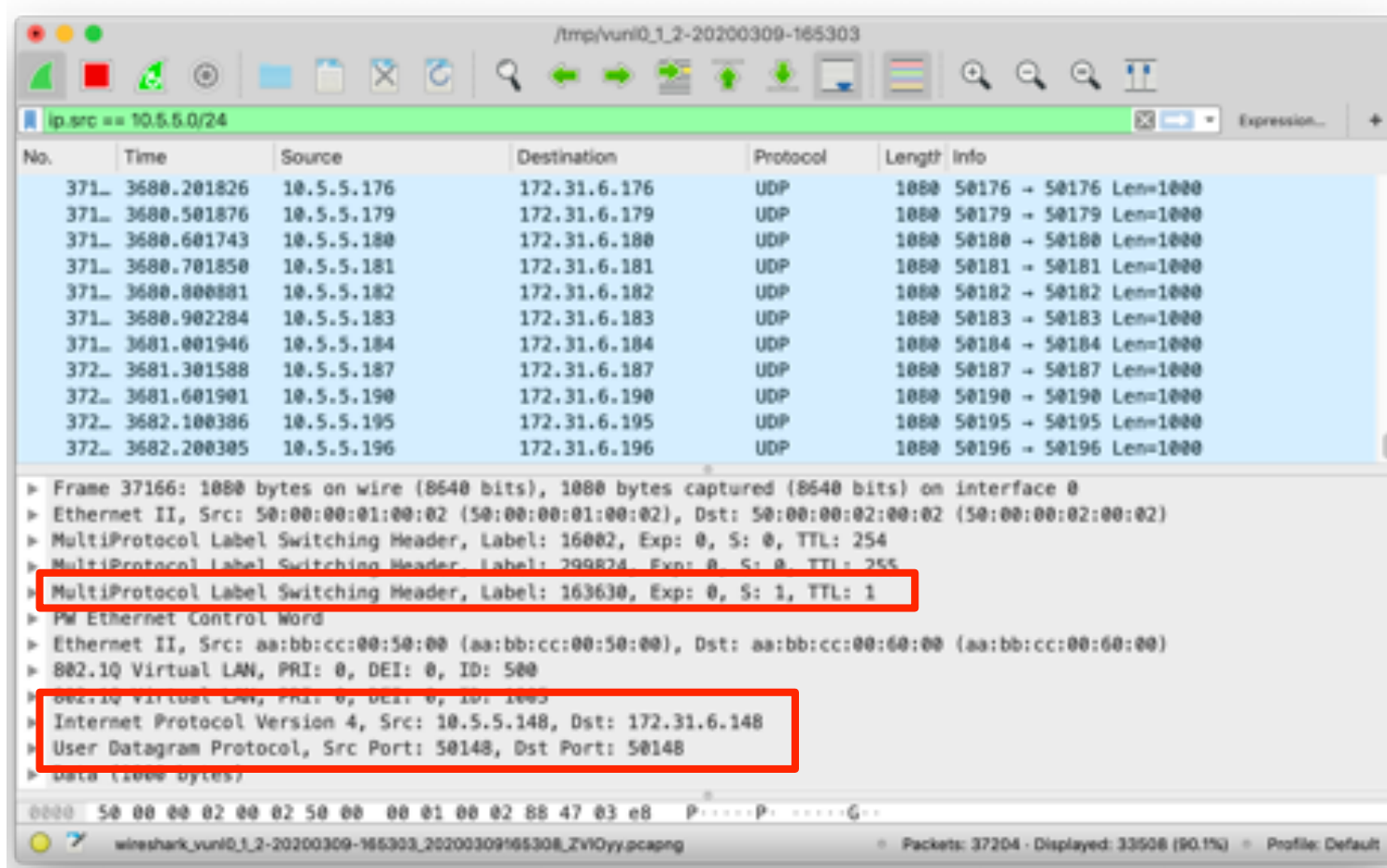
Packet 36319 details:

- Frame 36319: 1080 bytes on wire (8640 bits), 1080 bytes captured (8640 bits) on interface 0
- Ethernet II, Src: 50:00:00:01:00:02 (50:00:00:01:00:02), Dst: 50:00:00:02:00:02 (50:00:00:02:00:02)
- MultiProtocol Label Switching Header, Label: 16002, Exp: 0, S: 0, TTL: 254
- MultiProtocol Label Switching Header, Label: 299824, Exp: 0, S: 0, TTL: 255
- MultiProtocol Label Switching Header, Label: 879444, Exp: 0, S: 1, TTL: 1
- PW Ethernet Control word
- Ethernet II, Src: aa:bb:cc:00:50:00 (aa:bb:cc:00:50:00), Dst: aa:bb:cc:00:60:00 (aa:bb:cc:00:60:00)
- 802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 500
- 802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 1005
- Internet Protocol Version 4, Src: 10.5.5.108, Dst: 172.31.6.108
- User Datagram Protocol, Src Port: 50108, Dst Port: 50108
- Data (1000 bytes)

Packet bytes: 0000 50 00 00 02 00 02 50 00 00 01 00 02 88 47 03 e8 P:.....P:.....G..

Wireshark: `wireshark_vun0_1_2-20200309-165303_20200309165308_ZvOyy.pcapng`

Packets: 36499 - Displayed: 32912 (90.2%) - Profile: Default



Filter: ip.src == 10.5.5.0/24

No.	Time	Source	Destination	Protocol	Length	Info
371..	3680.201826	10.5.5.176	172.31.6.176	UDP	1080	50176 → 50176 Len=1080
371..	3680.501876	10.5.5.179	172.31.6.179	UDP	1080	50179 → 50179 Len=1080
371..	3680.601743	10.5.5.180	172.31.6.180	UDP	1080	50180 → 50180 Len=1080
371..	3680.701850	10.5.5.181	172.31.6.181	UDP	1080	50181 → 50181 Len=1080
371..	3680.800881	10.5.5.182	172.31.6.182	UDP	1080	50182 → 50182 Len=1080
371..	3680.902284	10.5.5.183	172.31.6.183	UDP	1080	50183 → 50183 Len=1080
371..	3681.001946	10.5.5.184	172.31.6.184	UDP	1080	50184 → 50184 Len=1080
372..	3681.301588	10.5.5.187	172.31.6.187	UDP	1080	50187 → 50187 Len=1080
372..	3681.601901	10.5.5.190	172.31.6.190	UDP	1080	50190 → 50190 Len=1080
372..	3682.100386	10.5.5.195	172.31.6.195	UDP	1080	50195 → 50195 Len=1080
372..	3682.200305	10.5.5.196	172.31.6.196	UDP	1080	50196 → 50196 Len=1080

Packet 37166 details:

- Frame 37166: 1080 bytes on wire (8640 bits), 1080 bytes captured (8640 bits) on interface 0
- Ethernet II, Src: 50:00:00:01:00:02 (50:00:00:01:00:02), Dst: 50:00:00:02:00:02 (50:00:00:02:00:02)
- MultiProtocol Label Switching Header, Label: 16082, Exp: 0, S: 0, TTL: 254
- MultiProtocol Label Switching Header, Label: 290824, Exp: 0, S: 0, TTL: 255
- MultiProtocol Label Switching Header, Label: 163630, Exp: 0, S: 1, TTL: 1
- PW Ethernet Control Word
- Ethernet II, Src: aa:bb:cc:00:50:00 (aa:bb:cc:00:50:00), Dst: aa:bb:cc:00:60:00 (aa:bb:cc:00:60:00)
- 802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 500
- 802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 1000
- Internet Protocol Version 4, Src: 10.5.5.148, Dst: 172.31.6.148
- User Datagram Protocol, Src Port: 50148, Dst Port: 50148
- Data (1000 bytes)

Packet bytes: 0000 50 00 00 02 00 02 50 00 00 01 00 02 88 47 03 e8 P.....P.....G..

Wireshark_vun0_1_2-20200309-165303_20200309165308_ZvOyy.pcapng

Packets: 37204 · Displayed: 33508 (90.1%) · Profile: Default

Dúvidas?

Apresentação e os scripts podem ser encontrados em:
<https://github.com/avargasn/fatpw>

<https://go.oncehub.com/FaleComExpert>

Schedule now



Online appointment scheduling

Pick a date and time that works for you -
it's fast and easy.

go.oncehub.com



Obrigado

Alberto Noronha

anoronha@psrv.com.br
+55 11 97061-6523