

# Vérification de performances dans les tunnels multidomaines

Stage recherche 2017 - master 2

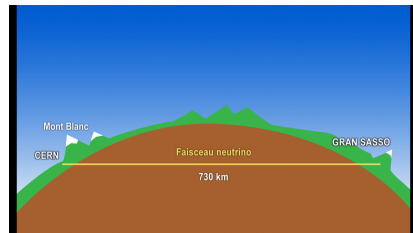
Thibaut Ehlinger

Mardi 29 août 2017



# Introduction

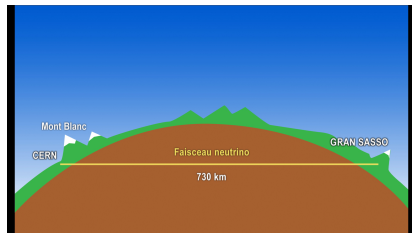
- Expérience OPERA, 23 septembre 2011 :  
Neutrino **7,4 km/s** plus rapides que la vitesse de la lumière



1. <https://www.ipnl.in2p3.fr/spip.php?article1464>

# Introduction

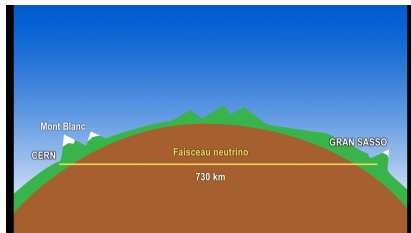
- Expérience OPERA, 23 septembre 2011 :  
Neutrino **7,4 km/s** plus rapides que la vitesse de la lumière
- *"due au branchement défectueux d'un câble de synchronisation optique des horloges de précision"*<sup>1</sup>



1. <https://www.ipnl.in2p3.fr/spip.php?article1464>

# Introduction

- Expérience OPERA, 23 septembre 2011 :  
Neutrino **7,4 km/s** plus rapides que la vitesse de la lumière
- *"due au branchement défectueux d'un câble de synchronisation optique des horloges de précision"*<sup>1</sup>
- Juin 2017 :  
Des paquets ont remonté le temps



1. <https://www.ipnl.in2p3.fr/spip.php?article1464>

# Un projet européen

- **Projet Horizon 2020**  
80 milliards d'euros alloués à la recherche et l'innovation
- **Projet GN4-2**  
destiné à maintenir le réseau de GÉANT et ses NREN à la pointe de l'innovation  
Plusieurs thématiques :
  - Network Activities (NA)
  - Service Activities (SA)
  - Joint Research Activities (JRA)
    - Network Services Development (JRA2)
- **Équipe JRA2T4**  
équipe du thème recherche de GN4-2
- **Objectif :**  
déployer un dispositif de vérification de performances de tunnels multidomains de GÉANT



FIGURE – Horizon 2020

## Sujet

# Supervision



# La supervision : un domaine à enjeux

- Problème : Internet très distribué
  - Comment mesurer les performances d'un réseau ?
- Exemple : les SLA *Worldcom* - 2002<sup>2</sup> :
  - 99.8% de disponibilité
  - Délai moyen de bout en bout

---

2. Jim MARTIN et Arne NILSSON. On service level agreements for ip networks. In **INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, tome 2**, pages 855–863. IEEE, 2002

# La supervision : un domaine à enjeux

- Problème : Internet très distribué
  - Comment mesurer les performances d'un réseau ?
- Exemple : les SLA *Worldcom* - 2002<sup>2</sup> :
  - 99.8% de disponibilité
  - Délai moyen de bout en bout
- Protège peu les utilisateurs :
  - Disponibilité  $\neq$  fluidité
  - Délai moyen englobe périodes actives **et** inactives
  - Délai moyen peut diminuer malgré une augmentation du taux d'utilisation moyen des liens.

---

2. Jim MARTIN et Arne NILSSON. On service level agreements for ip networks. In **INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, tome 2**, pages 855–863. IEEE, 2002



# La supervision : un domaine à enjeux

- Problème : Internet très distribué
  - Comment mesurer les performances d'un réseau ?
- Exemple : les SLA *Worldcom* - 2002<sup>2</sup> :
  - 99.8% de disponibilité
  - Délai moyen de bout en bout
- Protège peu les utilisateurs :
  - Disponibilité  $\neq$  fluidité
  - Délai moyen englobe périodes actives **et** inactives
  - Délai moyen peut diminuer malgré une augmentation du taux d'utilisation moyen des liens.
- SLA *NTT Communications* - 2017 :
  - délai **maximum** par continent
  - taux de pertes **maximum** 0.1%
  - Gigue moyenne 250  $\mu$ s et moyenne des 0.1% < 10ms

---

2. Jim MARTIN et Arne NILSSON. On service level agreements for ip networks. In **INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, tome 2**, pages 855–863. IEEE, 2002

# Analyse de performances

Indicateur de performances	Métrique	Outil(s) de mesure	Objects analysés
Délai	+	(tokyo-)ping, owamp	chemins ECMP aller/retour chemins dirigés (synchro ?) vs. liens/segments dirigés
Gigue Déséquencement	+	QoSmet	variations délais (n,n+1) (dé)composition bout en bout approximative / dégradé si pas de numérotation interne
Taux de pertes/succès	*	Idem délai/gigue	Même calcul que Délai et gigue
Capacités	min	pathload, iperf, nuttcp, bing	brut/résiduel vs. utile/applicatif tcp vs udp saut par saut aller/retour ?
Topologie	Composition	(paris-) traceroute	Changement de chemins

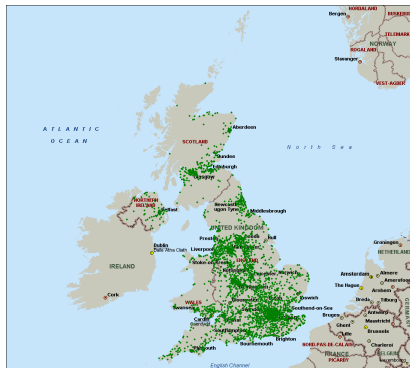
# Outils et plateformes

## Plateformes de sondage :

- (En général) Sondage **actif** :  
BP, gigue, pertes, routes, délai
- SamKnows (2009)  
5000+ boîtiers chez des utilisateurs
- BISmark  
Sondage **passif**
- RIPE Atlas (1997)  
9000+, dans les réseaux de cœur

## Plateformes de supervision :

- PerfSONAR (2004) :
  - **supervision**
  - Objectif : surveillance Worldwide LHC Computing Grid, historiquement supervision GÉANT, aujourd'hui<sup>3</sup>
  - Surveillance passive **SNMP**

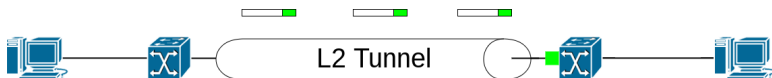


3. 75 pétaoctets entre 2012 et 2015

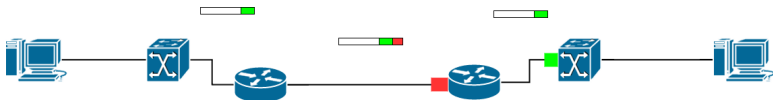
## Sujet

# Supervision des tunnels

# Circuits



# VPWS : *virtual private wire service*



# Tunnels et VPN

## Objectifs des tunnels :

- Étanchéité ⇒ Sécurité ⇒ Simplicité d'administration
- Qualité de service
- Performances

## Caractéristiques des tunnels :

- De niveau 2 ou 3
- Point à point/multipoints
- Multidomaines

NOM	DATE	Type	Protocoles	Objectif
VLAN	1998	Sous-réseau	VLAN - 802.1Q	Découpage LAN
VPN BGP MPLS	1999	Overlay IP	BGP, MPLS, LDP	Réseau privé IP distribué
MPLS TE LSP	2001	Tunnel IP	MPLS, RSVP TE	Qualité de service
VPWS	2005	Tunnel ethernet	VPWS	Tunnel couche 2
Q-in-Q	2005	Overlay VLAN	802. 1ad	Transport VLAN
VPLS	2007	Overlay ethernet	VPLS	Overlay niveau 2
Mac-inMac	2008	Overlay VLAN	802.1ah	Transport VLAN ciblé

## Sujet

# Supervision des tunnels multidomaines



# MDVPN : *carrier of carriers*

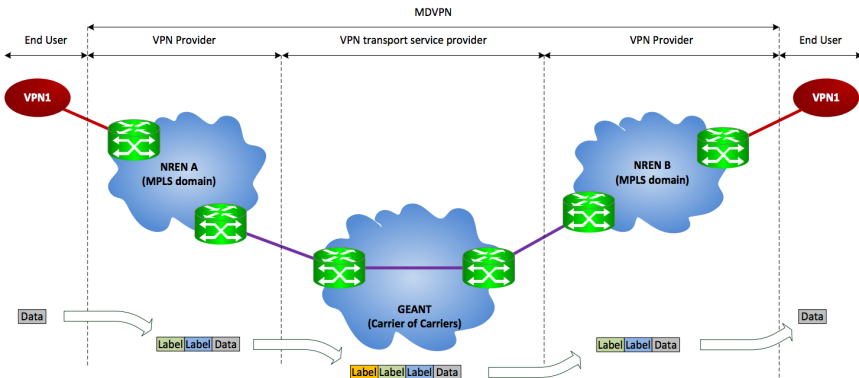
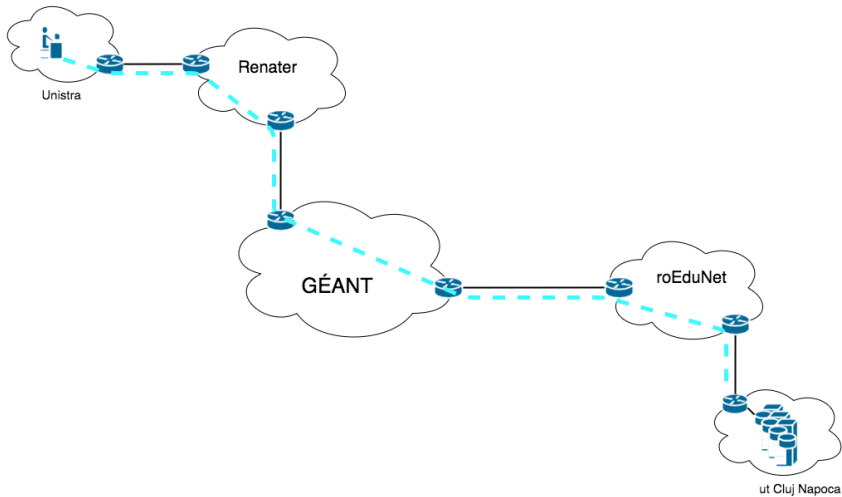
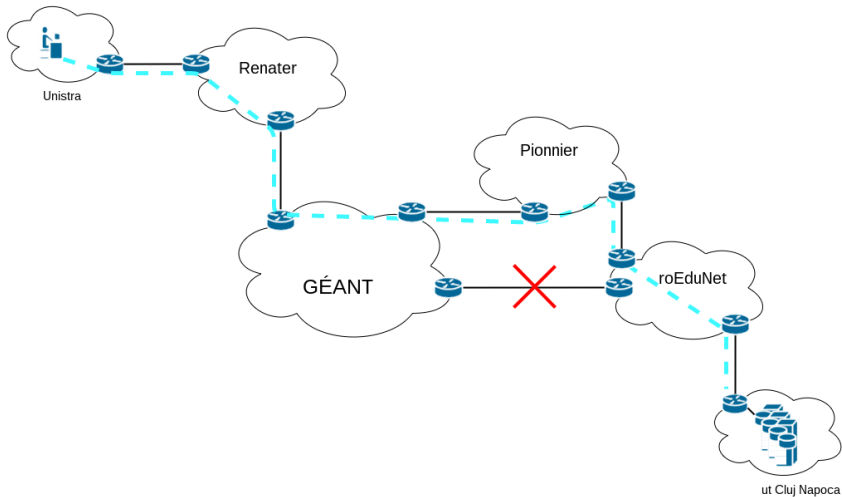


FIGURE – GÉANT MDVPN<sup>4</sup>

4. Xavier JEANNIN et T SZEWCZYK. Gn3plus sa3t3-multi domain vpn-technical architecture. In **2nd TERENA Network Architects Workshop (Prague)–14th Nov, 2013**





# Objectifs : PathCap

## Pour l'utilisateur :

tableau de bord de l'état du tunnel  
( *red/amber/green* )

## Pour les administrateurs des domaines intermédiaires :

- Statistiques de chaque tunnel en quasi-temps réel sur leur domaine

## Originalités :

- Capture très haut débit
- Localisation de défaillances



# Objectifs : PathCap

## Pour l'utilisateur :

tableau de bord de l'état du tunnel  
( *red/amber/green* )

## Pour les administrateurs des domaines intermédiaires :

- Statistiques de chaque tunnel en quasi-temps réel sur leur domaine

## Originalités :

- Capture très haut débit
- Localisation de défaillances

**Solution :** Mesures passives à partir de multiples points de capture



# Solution : *PathCap*

## Développement incrémental des fonctionnalités

Trois modes :

- 1 Mode 3 : prélèvement du trafic réel
  - Comportement réel des tunnels, historique QoS, localisation de défaillance
  - Confidentialité, performances, persistance des *tags*

# Solution : *PathCap*

## Développement incrémental des fonctionnalités

Trois modes :

- 1 Mode 3 : prélèvement du trafic réel
  - Comportement réel des tunnels, historique QoS, localisation de défaillance
  - Confidentialité, performances, persistance des *tags*
- 2 Mode 2 : injection et prélèvement de paquets forgés (plan B)
  - Localisation de défaillance
  - QoS pas complètement fiable

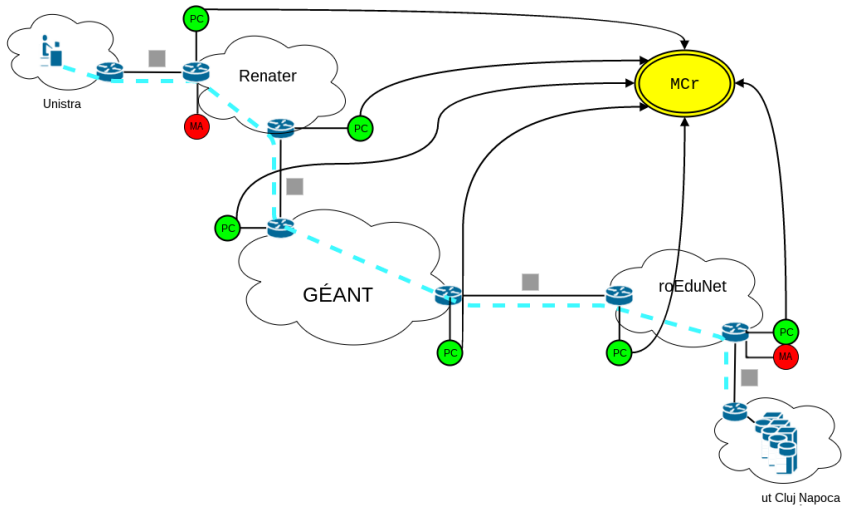
# Solution : *PathCap*

## Développement incrémental des fonctionnalités

Trois modes :

- 1 Mode 3 : prélèvement du trafic réel
  - Comportement réel des tunnels, historique QoS, localisation de défaillance
  - Confidentialité, performances, persistance des *tags*
- 2 Mode 2 : injection et prélèvement de paquets forgés (plan B)
  - Localisation de défaillance
  - QoS pas complètement fiable
- 3 Mode 1 : sondage actif de bout en bout (plan C)
  - Simple à mettre en place
  - Pas de localisation de défaillance





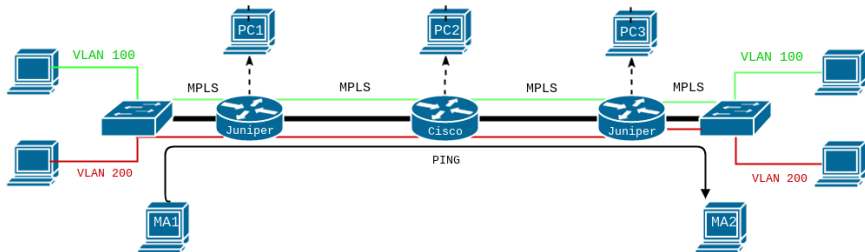
# Preuve de concept : banc d'essai

## GÉANT Testbed Service (GTS)

- service offert par GÉANT
- 9 "PoDs" en Europe (réseau+serveurs virtuels)

### Notre banc d'essai :

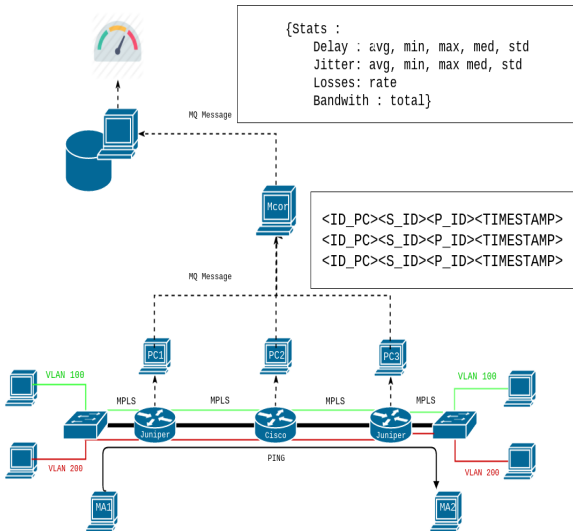
- 1 Point de distribution
- 2 tunnels
- 3 routeurs
- **Hétérogène**



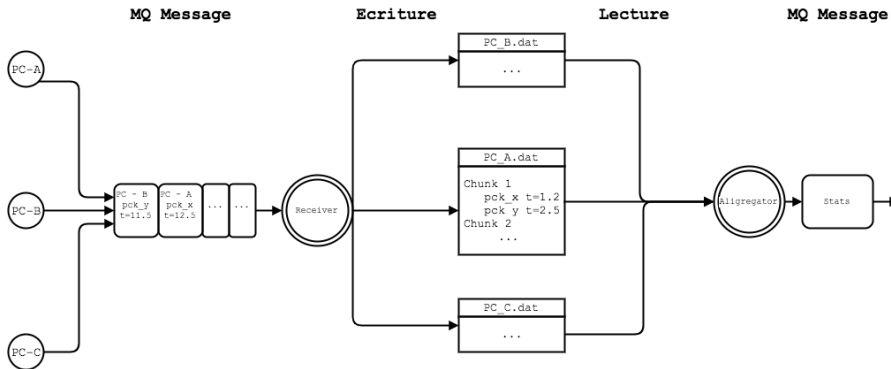
# Modes 2/3

## Mode 2 (David Schmitz - LMU, Munich)

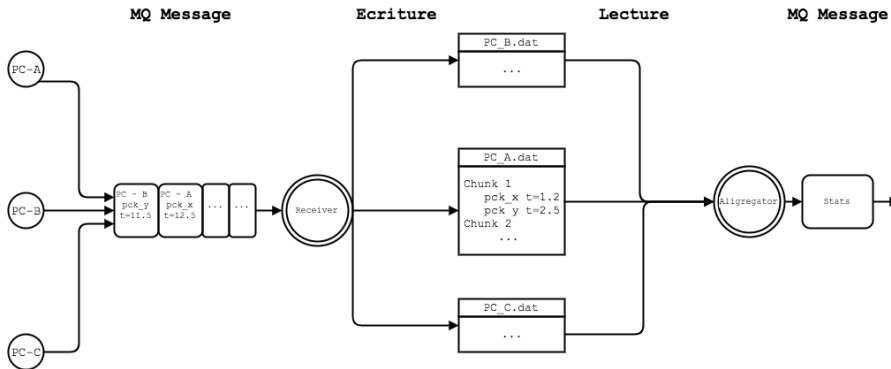
- Trafic OWAMPD interdépôts moyenne exponentielle
- *Magic cookie* capturé par MA3,4,5
- Traité par **aligregator**



# Aligregator (Aligregator : Listening aGREGATOR)

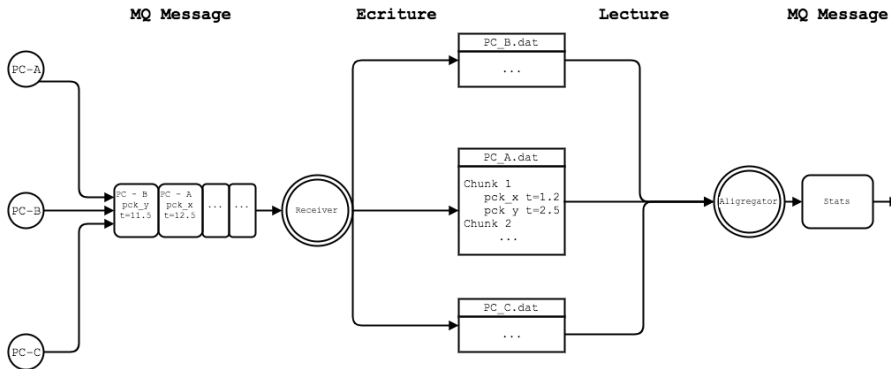


# Aligregator (Aligregator : Listening aGREGATOR)



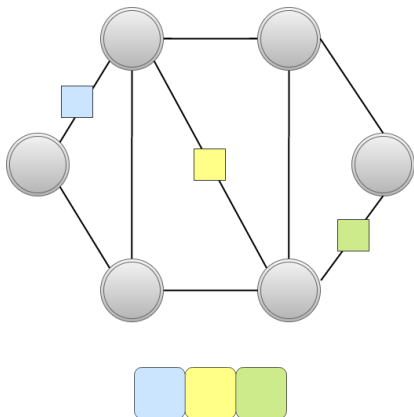
## ■ Délais négatifs

# Aligregator (Aligregator : Listening aGREGATOR)

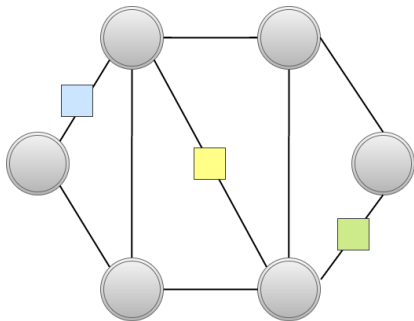


- Délais négatifs
- **Hypothèse 1** : synchronisation des horloges

# Reconstitution topologique

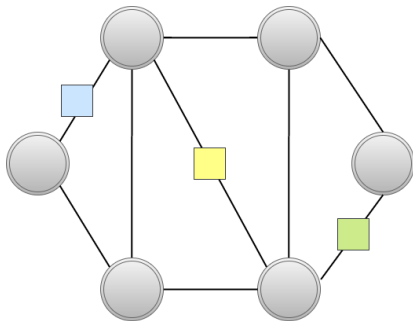


# Reconstitution topologique



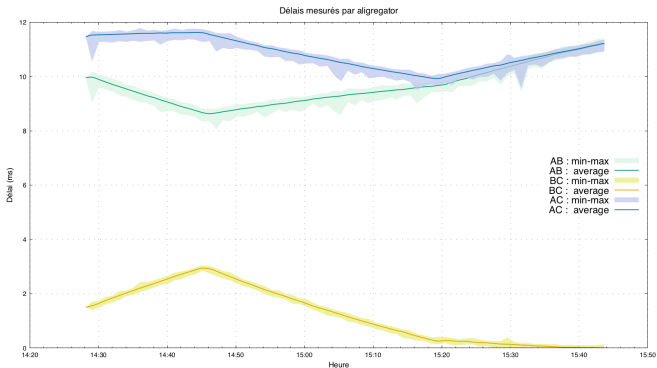


# Reconstitution topologique



- **Hypothèse 2** : un point de capture en terminaison du tunnel

# PoC : résultats préliminaires et perspectives



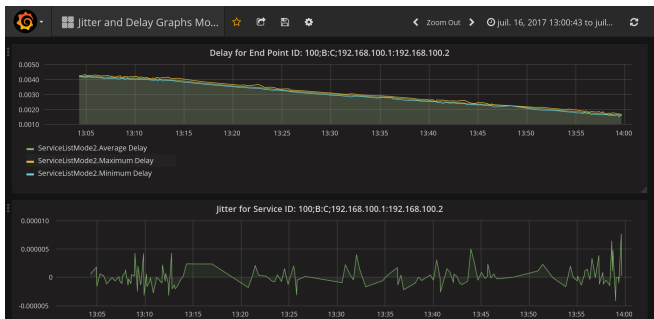
# Conclusion

## Aligregator

- Fonctionnel en continu
- Calcul complet de statistiques sur délai/gigue/pertes/débit
- Taille des fenêtres paramétrable
- Reconstitution topologique sur des tunnels statiques

## Perspectives

- Gestion dynamique de :
  - L'ajout/retrait de PC
  - L'ajout/retrait d'identifiants à superviser
- Trafic réel (mode 3)
- Essai sur tunnels dynamiques



# Capture traffic 100Gb/s : volume et débit

## Réduction des données :

- Hash de chaque paquets
- Compression par flux

## Performances de capture :

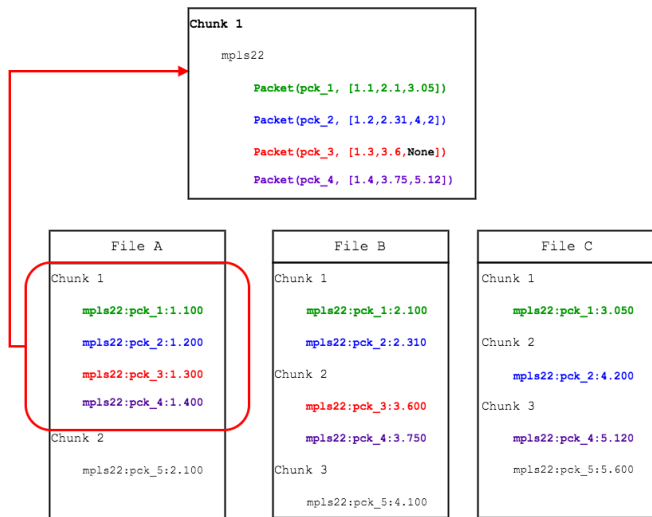
- Utilisation de *network taps*
- Programmation bas niveau

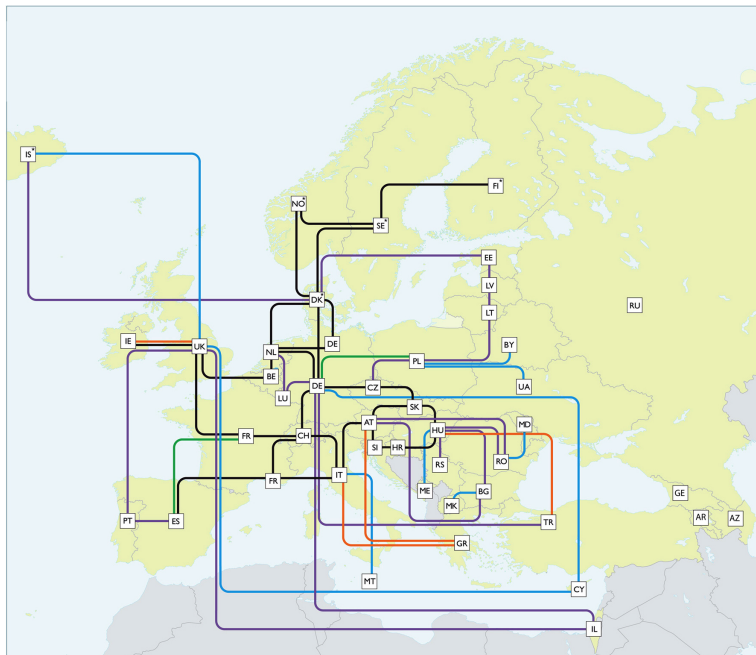
Problématiques de confidentialité

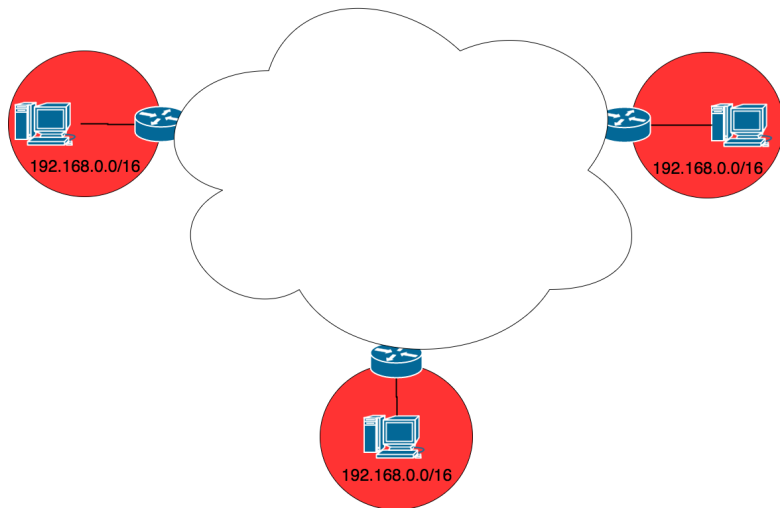


FIGURE – Network tap

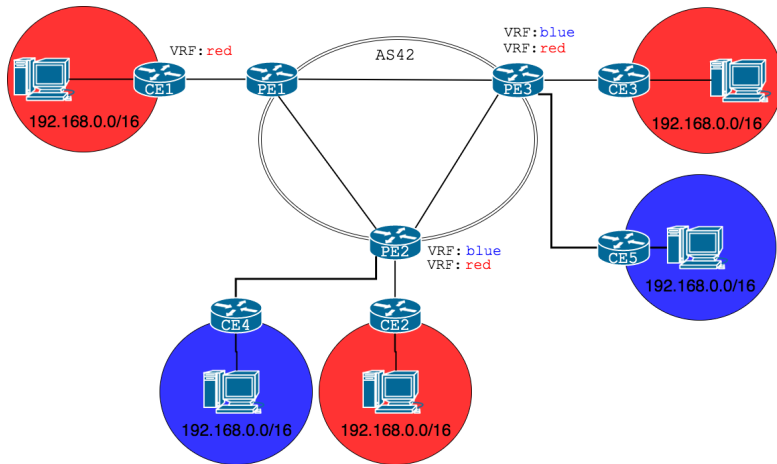
# Reconstitution topologique





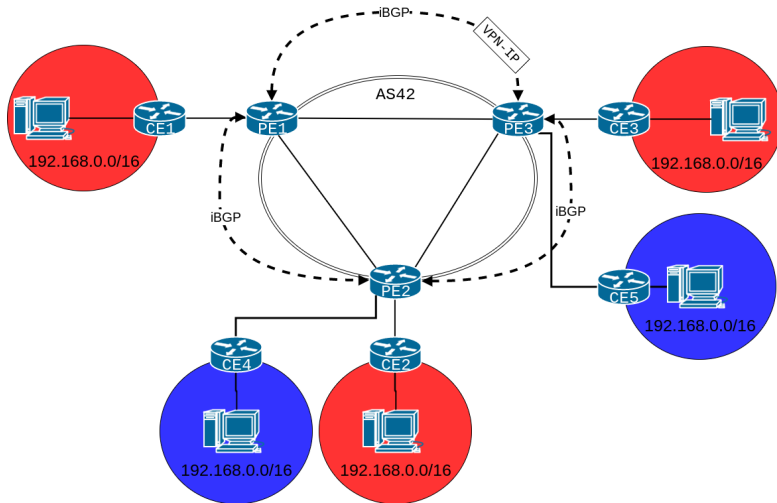


# VPN BGP MPLS : *virtual routing and forwarding (VRF)*

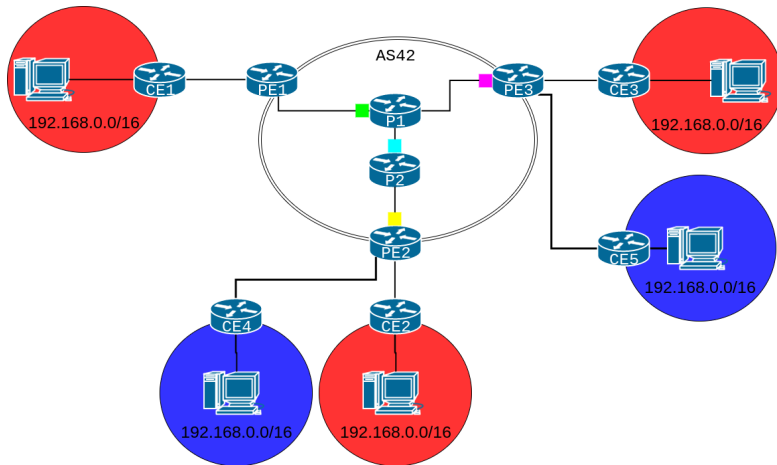




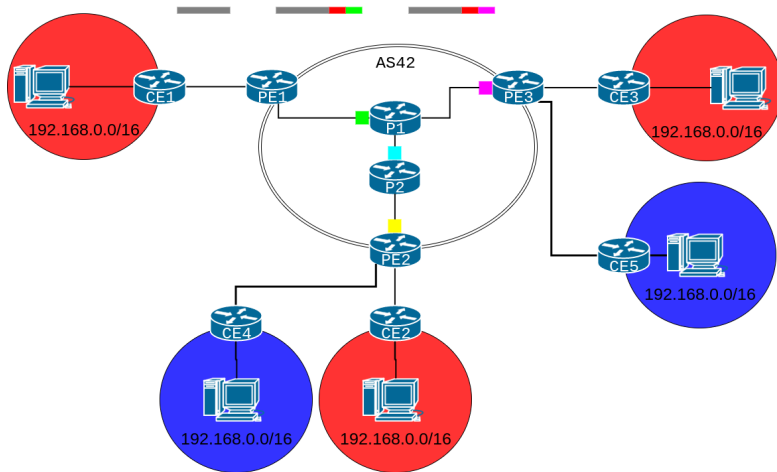
# VPN BGP MPLS : échange de routes



# Routage de cœur : en MPLS



# Plan de données



## RECHERCHE

**Département  
Informatique recherche**  
*Dir. adj : Thomas NOEL*

**Équipe 1 - IGG**  
*Resp : D. BECHMANN*  
*Adj. : J-M DISCHLER*

**Équipe 2 Réseaux**  
*Resp : T. NOËL*  
*Adj : S. CATELOIN*

**Équipe 3 - ICPS**  
*Resp : P. CLAUSS*  
*Adj : J. GUSTEDT*

**Équipe 4 -SDC**  
*Resp : N. LACHICHE*  
*Adj : C. WEMMERT*

**Équipe 5 - CSTB**  
*Co-resp : O. POCH*  
*Co-resp : P. COLLET*

**Équipe 6 - MIV**  
*Resp : F. HEITZ*  
*Adj : C. RONSE*

**Département Imagerie,  
Robotique,  
Téledétection et Santé**  
*Dir. adj: Fabrice HEITZ*

**Équipe 6 - MIV**  
*Resp : F. HEITZ*  
*Adj : C. RONSE*

**Équipe 7 - AVR**  
*Resp : P. RENAUD*  
*Adj : J. GANGLOFF*

**Équipe 8 - TRIO**  
*Resp : F. NERRY*  
*Adj : P. GRUSSENMEYER*

**Équipe 9 - IMIS**  
*Resp : J-P ARMSPACH*

**Département  
Électronique du Solide,  
Systèmes et Photonique**  
*Dir. adj: Daniel MATHIOT*

**Équipe 10 - MaCEPV**  
*Resp : T. HEISER*  
*Adj : F. Le NORMAND*

**Équipe 11 - SMH**  
*Resp : C. LALLEMENT*  
*Adj : W. UHRING*

**Équipe 12 - IPP**  
*Resp : P. MONTGOMERY*  
*Adj : S. LECLERC*

**Département  
Mécanique**  
*Dir. adj: Robert MOSÉ*

**Équipe 13 - MécaFlu**  
*Co-resp : J. DUSEK*  
*Co-resp : I. CHARPENTIER*

**Équipe 14 - MMB**  
*Co-resp : R. WILLINGER*  
*Co-resp : N. BAHLOULI*

**Équipe 15 - GCE**  
*Resp : C. FOND*  
*Adj : F. FEUGEAS*

**Équipe 16 - CSIP**  
*Resp : D. CAVALLUCCI*  
*Adj : V. GOEPP*