# Structures et Dynamique des Réseaux

## Robustesse et métrologie dynamique

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

prenom.nom@lip6.fr

# Plan

# Plan

# Outline

# Contexte

Phénomènes de propagation :

- Internet
- Réseaux sociaux (maladies, rumeurs, ...)
- Échanges d'e-mails (virus informatiques)
- ...

Deux types d'atteintes

- Pannes
- Attaques

But : Comprendre ces phénomènes pour pouvoir :

- Prédire
- Construire des stratégies d'attaque efficaces
- (Protéger)

# Évaluation des dégâts

Être capable de déterminer à quel point le réseau est endommagé :

## Plusieurs critères possibles

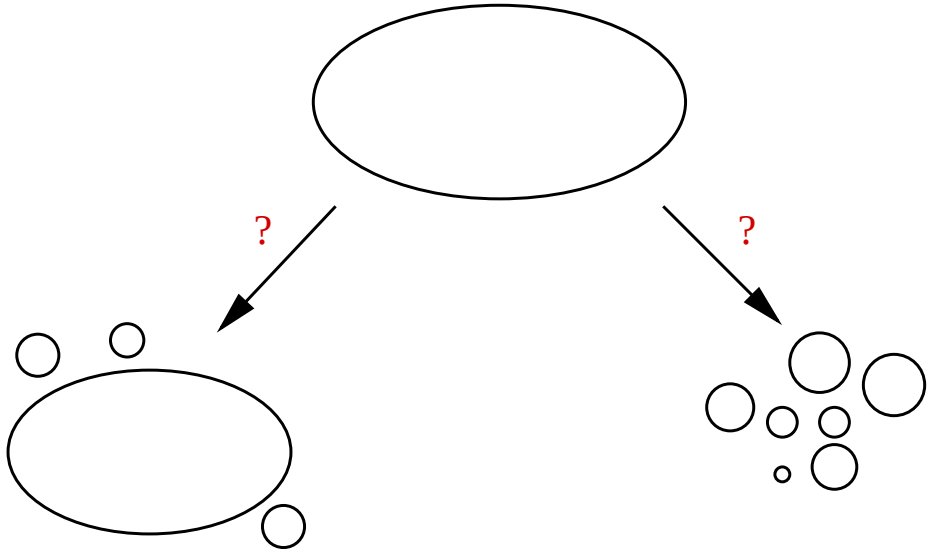
- Critères fondés sur la distance
- Tailles des composantes connexes

# Évaluation des dégâts

Être capable de déterminer à quel point le réseau est endommagé :

## Plusieurs critères possibles

- Critères fondés sur la distance
- Tailles des composantes connexes



# Modélisation

- Pannes
- Attaques
- Réseaux

# Modélisation

- Pannes
  - Suppressions aléatoires de nœuds ou de liens
- Attaques
- Réseaux

# Modélisation

- Pannes
- Attaques
  - Suppression de nœuds ou liens choisis
- Réseaux

# Modélisation

- Pannes
- Attaques
- Réseaux :

## Étude de l'influence des degrés

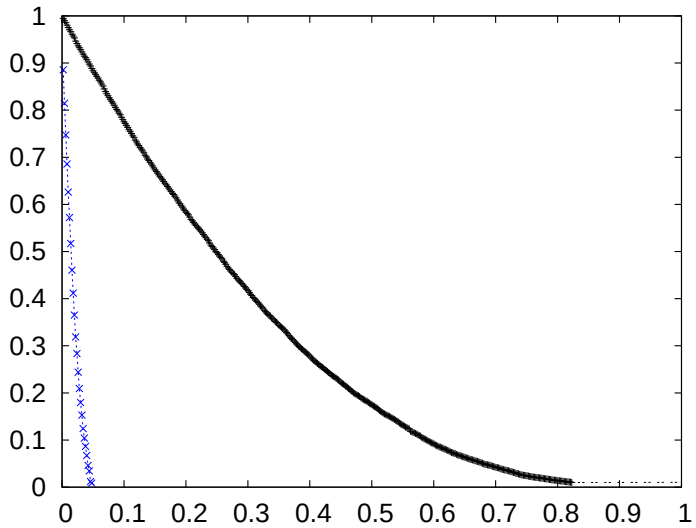
- Graphes de terrain
- Graphes aléatoires (Erdös et Rényi)
  Tous les nœuds sont équivalents
- Graphes sans échelle ($p_k \sim k^{-\alpha}$)
  Présence de nœuds de fort degré

# Outline

# Simulations

[Albert, Jeong et Barabási, 2000]

— Pannes de nœuds
— Attaques (nœuds de plus fort degré)

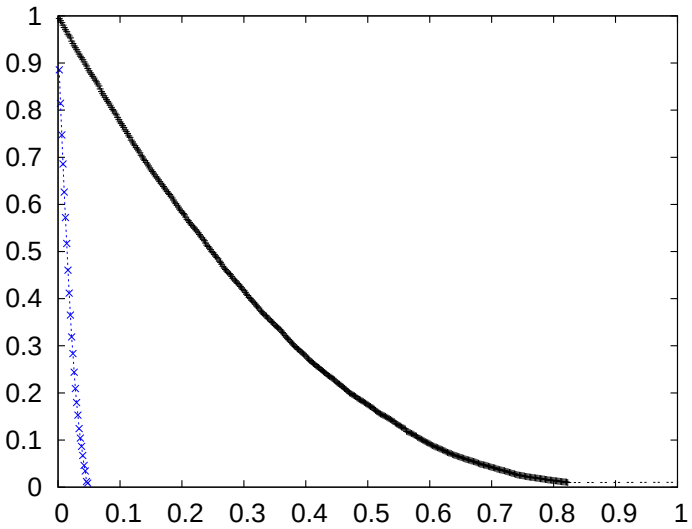



Mesure de l'internet

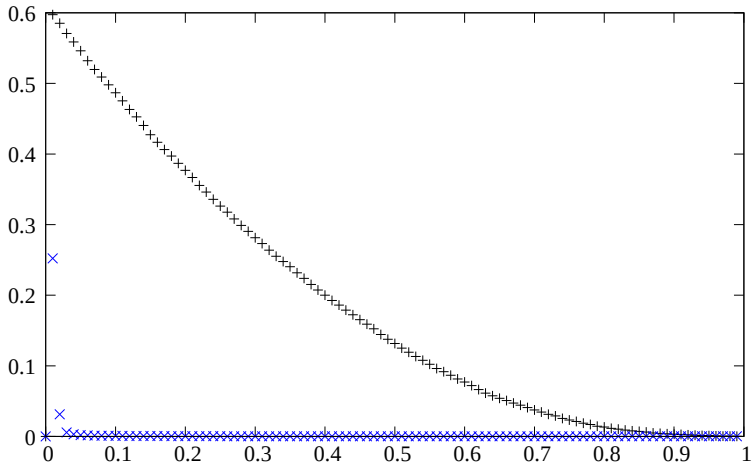
# Simulations

[Albert, Jeong et Barabási, 2000]

— Pannes de nœuds
— Attaques (nœuds de plus fort degré)




Mesure de l'internet

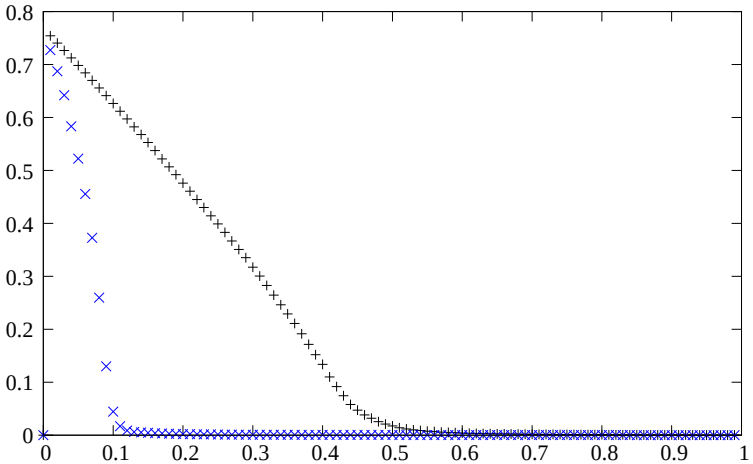



Graphe sans échelle

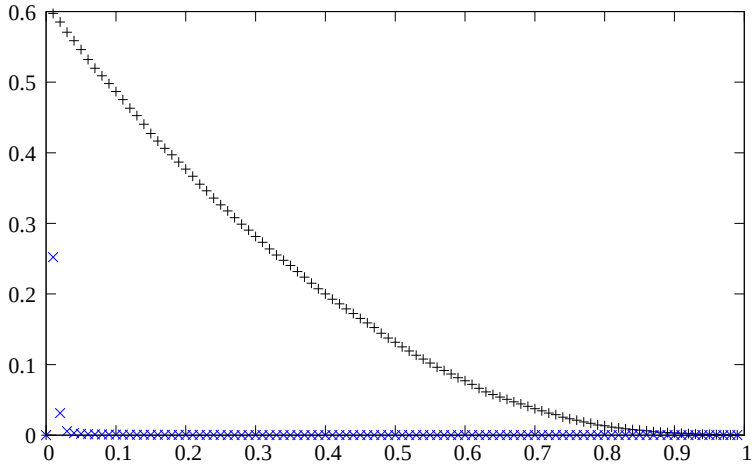
# Simulations

[Albert, Jeong et Barabási, 2000]

— Pannes de nœuds
— Attaques (nœuds de plus fort degré)




Graphe aléatoire




Graphe sans échelle

# Observations

## Graphes réels :

- résistants aux pannes

  supprimer tous les nœuds pour déconnecter

- sensibles aux attaques

## Graphes aléatoires :

- plus sensibles aux pannes

  supprimer une fraction des nœuds pour déconnecter

- moins sensibles aux attaques

# Observations

Différences : distribution des degrés

## Graphes aléatoires

Degrés homogènes, pannes $\sim$ attaques

## Graphes sans échelle

Différences dans les degrés
$\rightarrow$ différences entre pannes et attaques

$\rightarrow$ modélisation par des graphes avec degrés fixés :

- plus de souplesse qu'utiliser les graphes réels
- permet d' expliquer les phénomènes

comparaison avec graphes aléatoires

# Pannes, résultats théoriques

[Cohen *et al*, 2000]

Graphe aléatoire avec distribution distribution des degrés $p_k$.
Suppression aléatoire d'une fraction $p$ des nœuds :

$$p_k(p) = \sum_{k_0=k}^{\infty} p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k}$$

# Pannes, résultats théoriques

[Cohen *et al*, 2000]

Graphe aléatoire avec distribution distribution des degrés $p_k$.
Suppression aléatoire d'une fraction $p$ des nœuds :

$$p_k(p) = \sum_{k_0=k}^{\infty} p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k}$$

## Théorème [Molloy, Reed, 1995]

Composante géante ssi :

$$\langle k(p)^2 \rangle - 2\langle k(p) \rangle = \sum_{k=0}^{\infty} k(k-2) p_k(p) \geq 0$$

# Pannes, résultats théoriques

[Cohen *et al*, 2000]

En remplaçant $p_k(p)$ dans le théorème :

$$p_c = 1 - \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

# Applications (1/1)

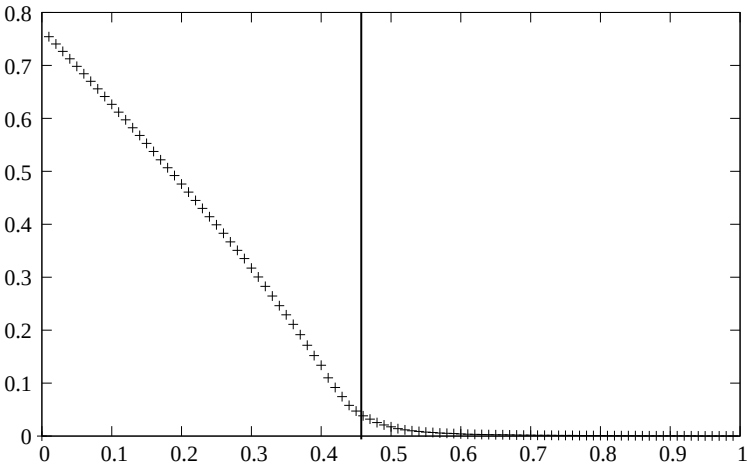
$$p_c = 1 - \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

Graphes $\longrightarrow \infty$ :

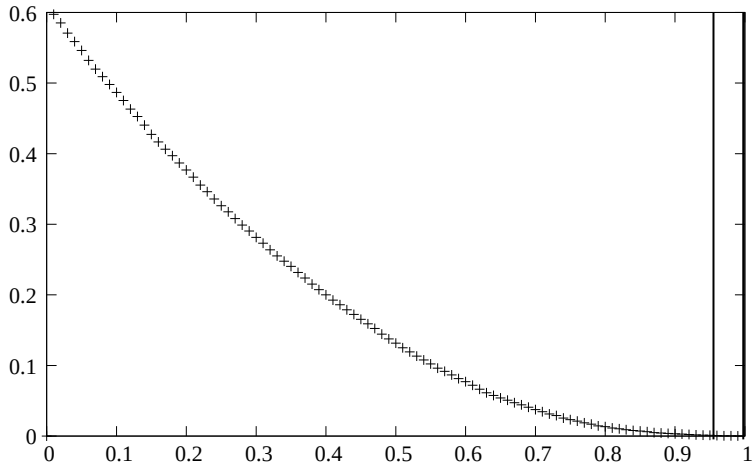
- aléatoires :
  - $\langle k \rangle = z$
  - $\langle k^2 \rangle = z^2 + z$
- sans échelle $(2 \leq \alpha \leq 3)$ :
  - $\langle k \rangle = \zeta(\alpha - 1) < \infty$
  - $\langle k^2 \rangle = \zeta(\alpha - 2) = \infty$

# Applications (1/1)

$$p_c = 1 - \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$




ER




SF

# Applications (2/2)

On aimerait comparer les seuils théoriques aux seuils observés en pratique

Problème : pas de définition du seuil en pratique
Notion qui a du sens seulement pour les graphes dont la taille tend vers l'infini

# Pannes de liens

Suppression aléatoire de liens.

# Pannes de liens

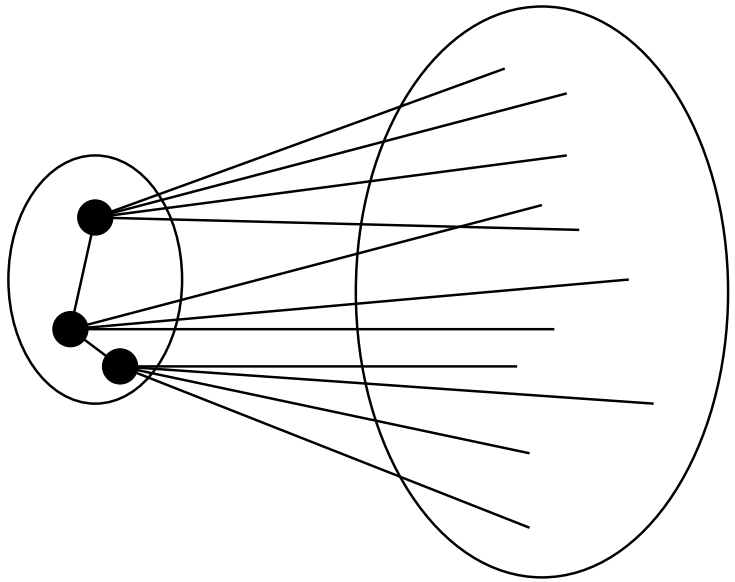
Suppression aléatoire de liens.
Même effet sur la distribution des degrés que les suppressions de nœuds.
$\longrightarrow$ mêmes seuils théoriques que pour les pannes de nœuds.

# Attaques - 1

Suppression d'une fraction $p$ des nœuds de plus fort degré.

Deux effets :

- Borne sur le degré maximum $K(p)$
- Modification de la distribution des degrés



# Attaques - 1

Suppression d'une fraction $p$ des nœuds de plus fort degré.

Deux effets :

- Borne sur le degré maximum $K(p)$

$$p = \sum_{k=K(p)+1}^{\infty} p_k$$

# Attaques - 1

Suppression d'une fraction $p$ des nœuds de plus fort degré.

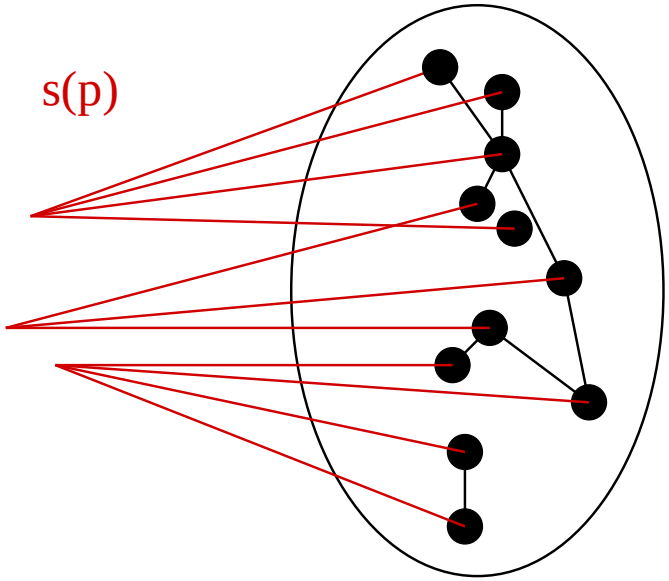
Deux effets :

- Modification de la distribution des degrés

Fraction des demi-liens appartenant aux nœuds supprimés :

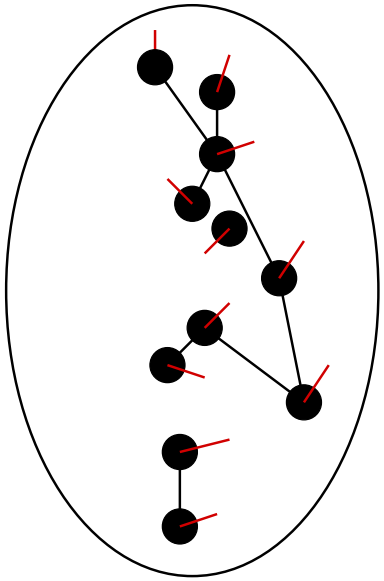
$$s(p) = \frac{\sum_{k=K(p)+1}^{\infty} kp_k}{\langle k \rangle}$$

# Attaques - 2



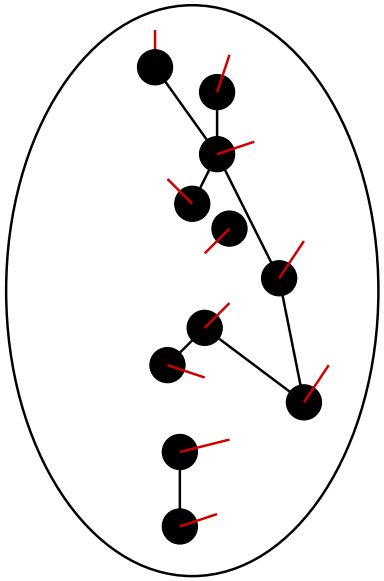
# Attaques - 2

s(p)



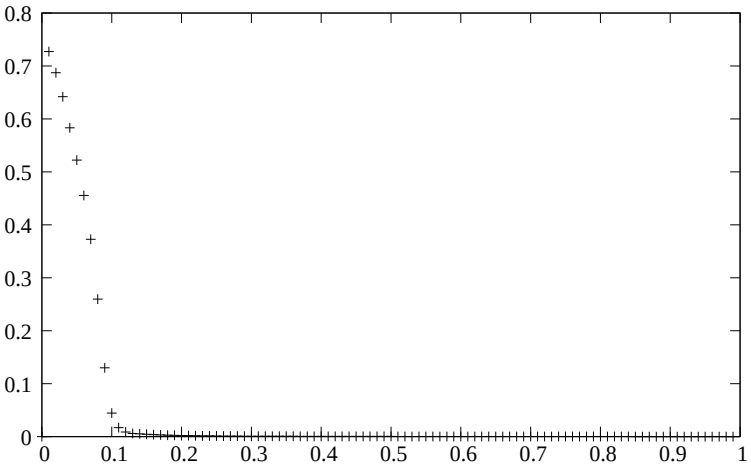
# Attaques - 2

s(p)

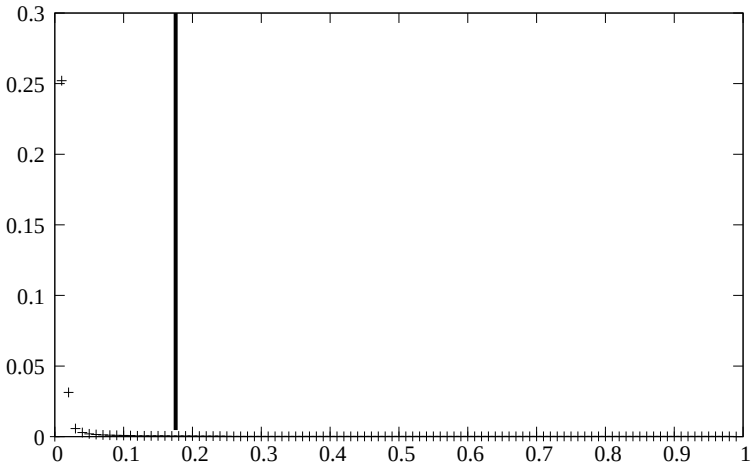


Équivalent à des pannes de liens
$\rightarrow$ possible de calculer le seuil

# Attaques - 3

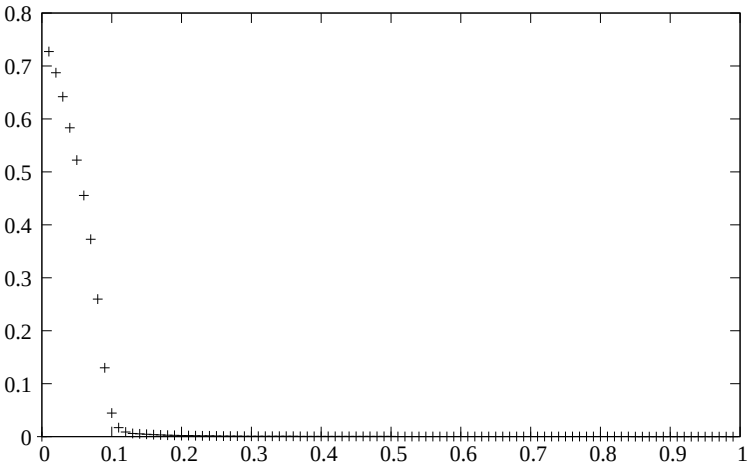



ER

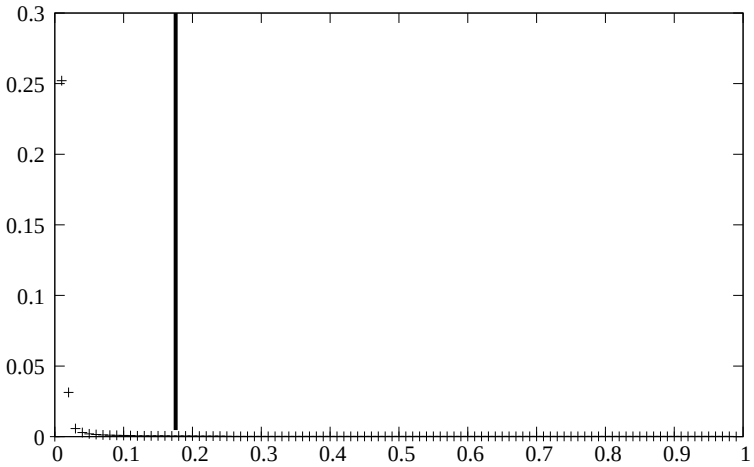



SF

# Attaques - 3




ER

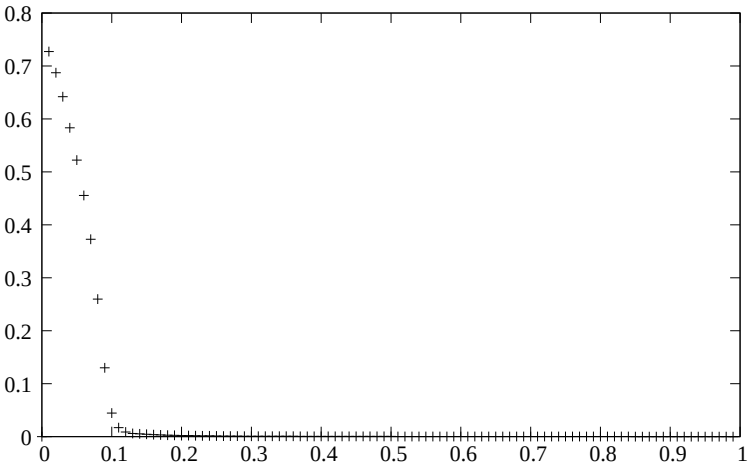



SF

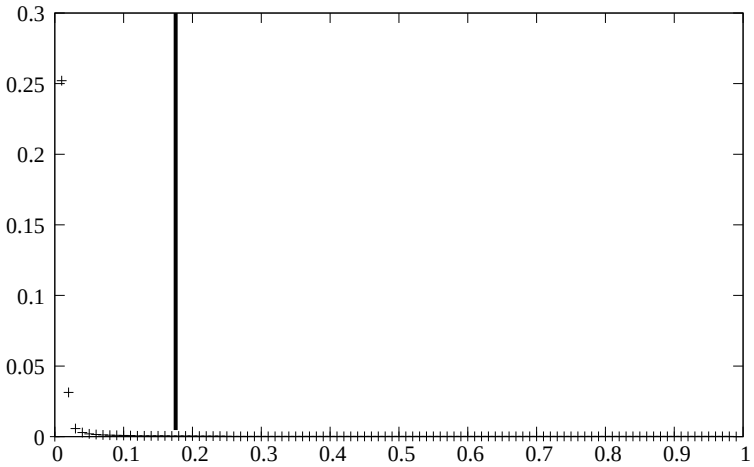
**Graphes sans échelle :**

- Possible de donner des formules pour le seuil
- Seuils théoriques faibles

# Attaques - 3




ER




SF

## Graphes aléatoires

- Même idée théorique
- Problème pour les calculs

# Outline

# Attaques vues du point de vue des liens

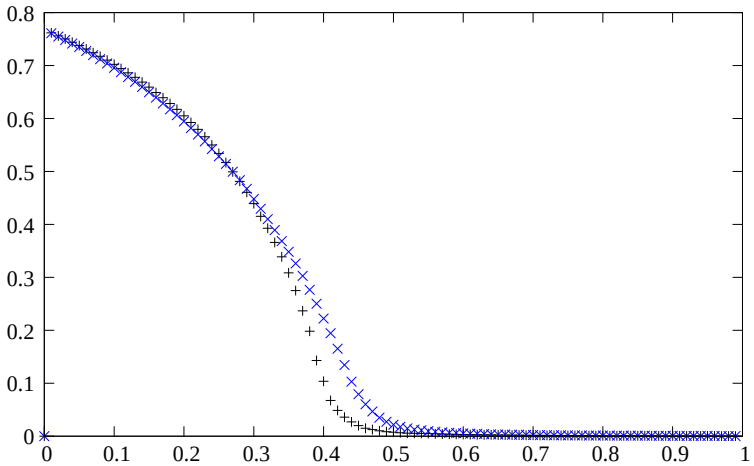
Attaques sur les nœuds $\longrightarrow$ stratégie d'attaque sur les liens

Grands nombre de liens supprimés :
Explication de l'efficacité des attaques ?

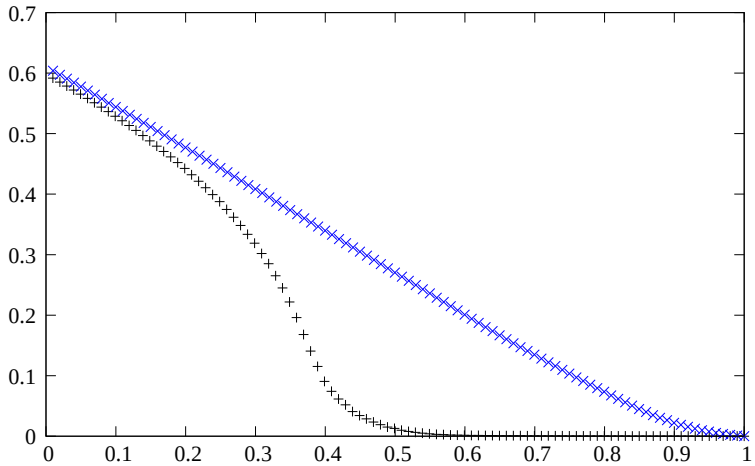
# Attaques vues du point de vue des liens

Attaques sur les nœuds $\longrightarrow$ stratégie d'attaque sur les liens

- Attaques vues du point de vue des liens
- Pannes de liens




ER




SF

# Conclusion

Efficacité des attaques :
pas uniquement dûe au grand nombre de liens supprimés

# Outline

# Nouvelles stratégies d'attaque

[Molloy et Reed, 1995]

Composante géante ssi :

$$\langle k^2 \rangle - 2\langle k \rangle \geq 0$$

$$\sum_{k=0}^{\infty} k(k-2)p_k \geq 0$$

# Nouvelles stratégies d'attaque

[Molloy et Reed, 1995]

Composante géante ssi :

$$\langle k^2 \rangle - 2\langle k \rangle \geq 0$$

$$\sum_{k=0}^{\infty} k(k-2)p_k \geq 0$$

$$p_1 \leq \sum_{k=3}^{\infty} k(k-2)p_k$$

# Nouvelles stratégies d'attaque

[Molloy et Reed, 1995]

Composante géante ssi :

$$\langle k^2 \rangle - 2\langle k \rangle \geq 0$$

$$\sum_{k=0}^{\infty} k(k-2)p_k \geq 0$$

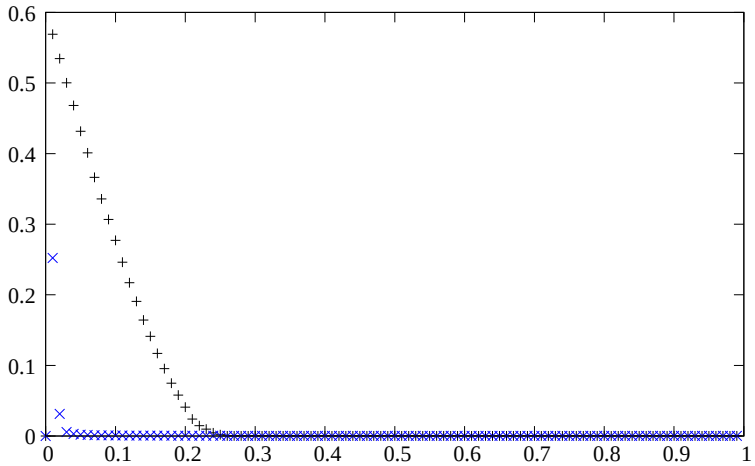
$$p_1 \leq \sum_{k=3}^{\infty} k(k-2)p_k$$

$\longrightarrow$ nouvelles stratégies d'attaque

# Nœuds

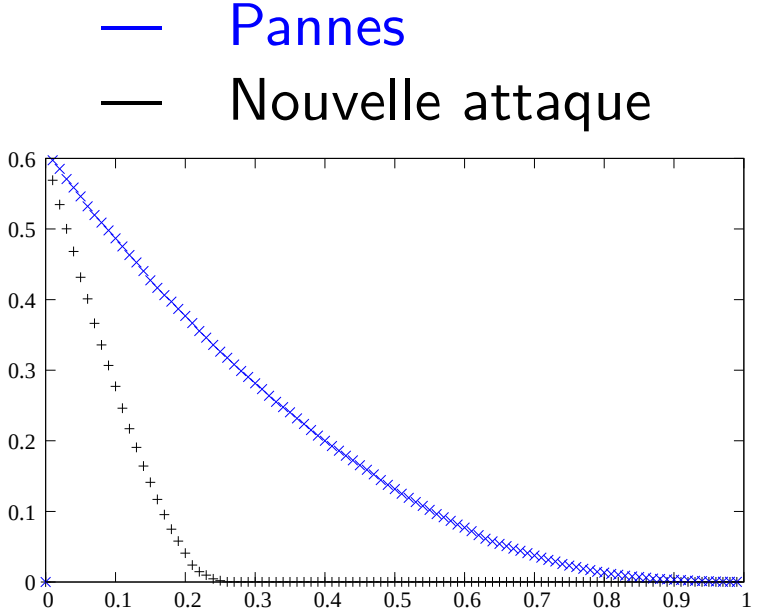
Stratégie : Suppression aléatoire de nœuds de degré $\geq 2$

— Attaques classiques
— Nouvelle attaque



# Nœuds

Stratégie : Suppression aléatoire de nœuds de degré $\geq 2$



# Seuil

Borne supérieure pour le seuil :

$$1 - p_1$$

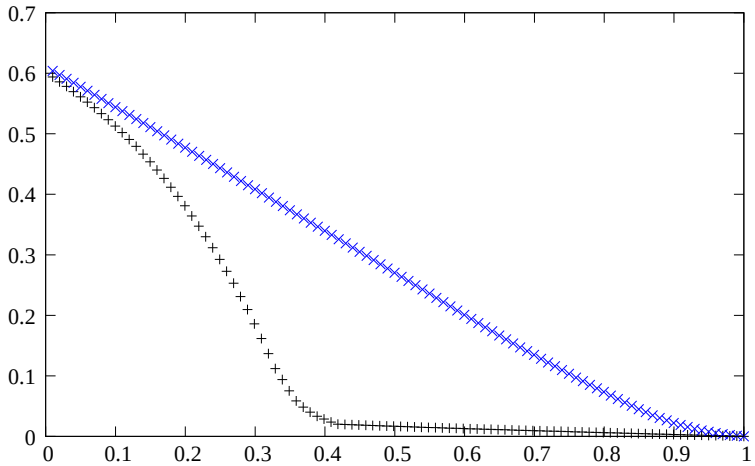
# Liens

Stratégie :

Suppression aléatoire de liens entre des sommets de degré $\geq 2$.

— Pannes de liens

— Nouvelle attaque



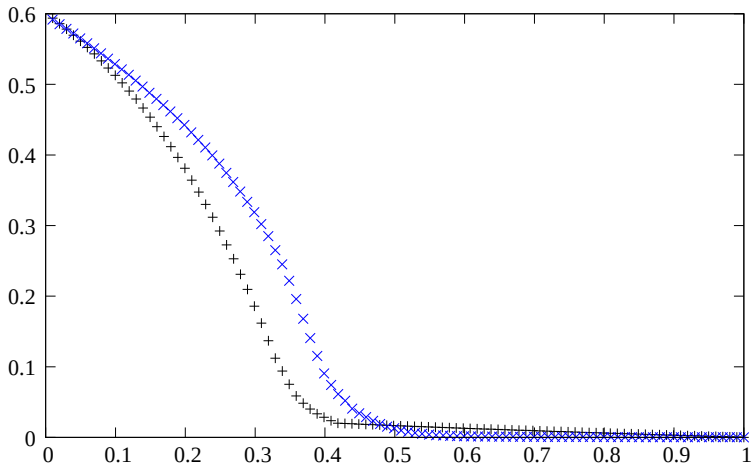
# Liens

Stratégie :

Suppression aléatoire de liens entre des sommets de degré $\geq 2$.

— Attaques classiques

— Nouvelle attaque



# Seuil

Borne supérieure pour le seuil :
supprimer tous les liens entre deux nœuds de degré $\geq 2$.

Nombre de liens incidents à au moins un sommet de degré 1 :

$$N_1 - N_{(1,1)}$$

# Conclusion

## Pannes et attaques classiques

- Attaques plus efficaces que les pannes
- Grande différence entre graphes aléatoires et sans-échelle

# Conclusion

## Pannes et attaques classiques

- Attaques plus efficaces que les pannes
- Grande différence entre graphes aléatoires et sans-échelle
- Mais attention au cas fini

# Conclusion

## Pannes et attaques classiques

- Attaques plus efficaces que les pannes
- Grande différence entre graphes aléatoires et sans-échelle
- Mais attention au cas fini

## Efficacité des attaques liée à :

- Pas de suppression de nœuds de degré 1
- Pas à la quantité de nœuds supprimée
- Pas de suppression de liens adjacents à des nœuds de degré 1.

# Conclusion

## Pannes et attaques classiques

- Attaques plus efficaces que les pannes
- Grande différence entre graphes aléatoires et sans-échelle
- Mais attention au cas fini

## Efficacité des attaques liée à :

- Pas de suppression de nœuds de degré 1
- Pas à la quantité de nœuds supprimée
- Pas de suppression de liens adjacents à des nœuds de degré 1.

## Impact d'autres propriétés ?

- Clustering
- Corrélations entre les degrés
- ...

# Plan

# Principaux travaux de référence

Fabien Viger, Brice Augustin, Xavier Cuvellier, Clémence Magnien, Matthieu Latapy, Timur Friedman et Renata Teixeira

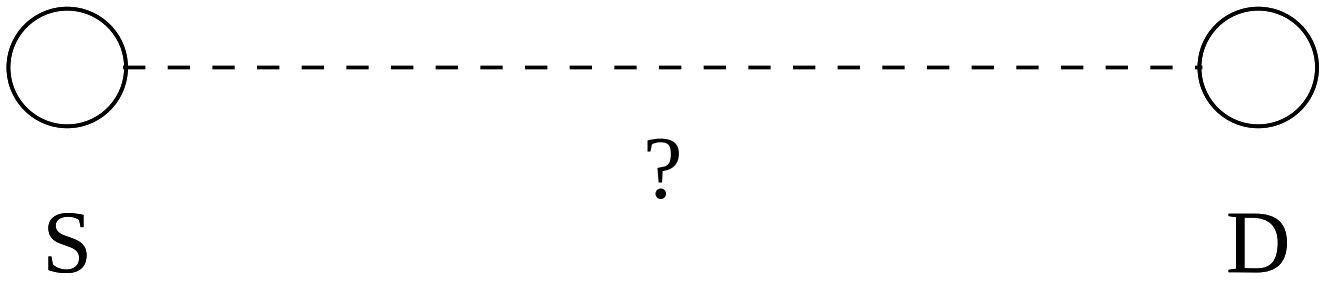
Detection, understanding, and prevention of traceroute measurement artifacts
Computer Networks 52-5, 2008.

# Topologie de l'Internet au niveau IP

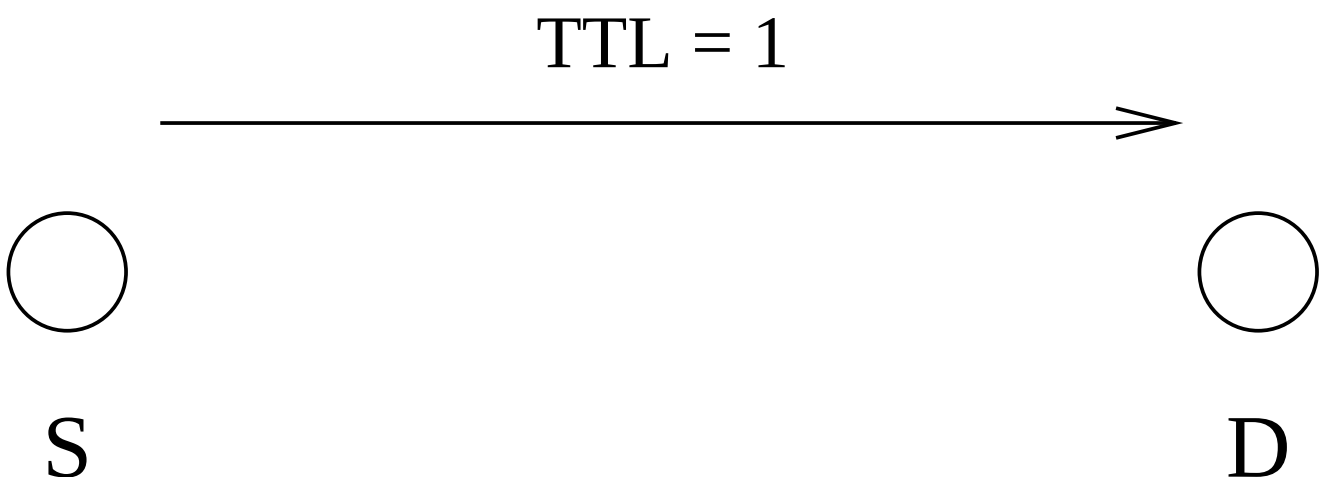
- Adresses IP
- Sauts au niveau IP

# Outline

# Mesure avec `traceroute`: rappels

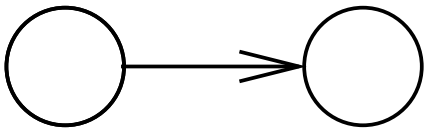


# Mesure avec `traceroute`: rappels

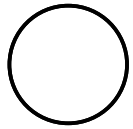


# Mesure avec `traceroute`: rappels

# Mesure avec `traceroute`: rappels

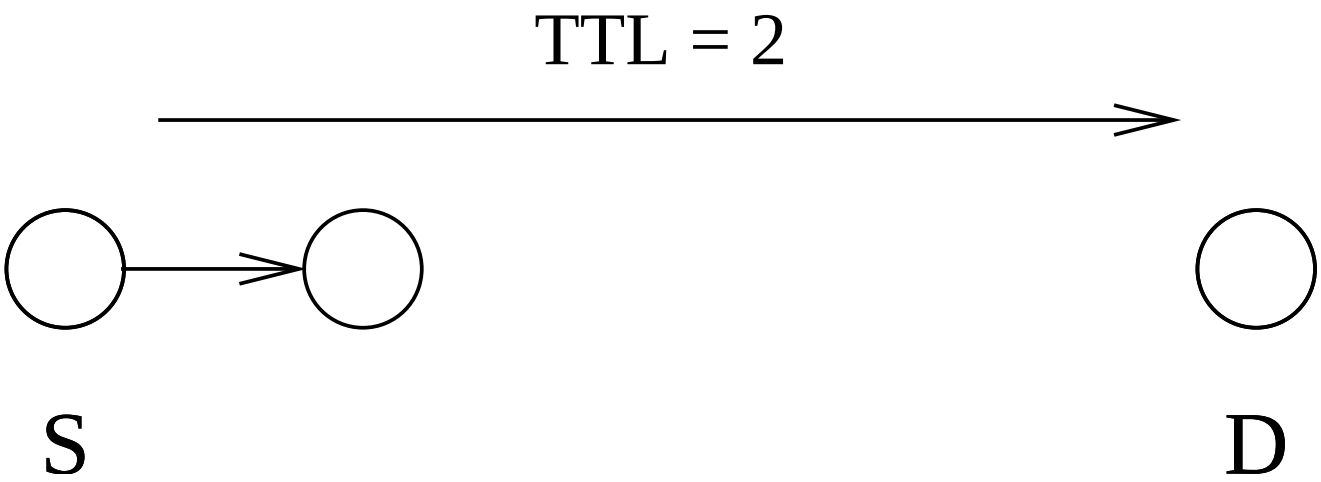


S

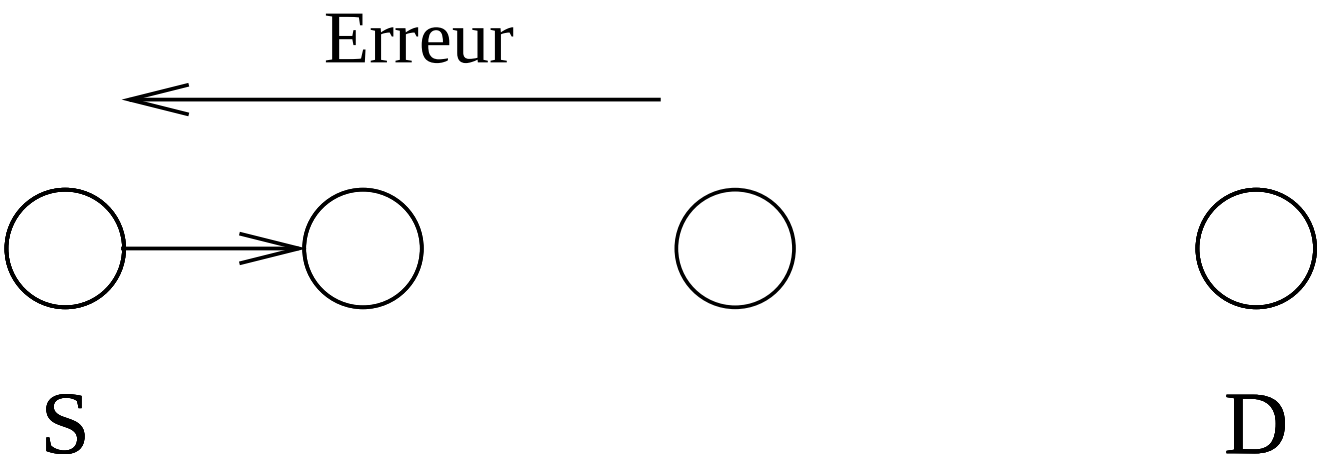


D

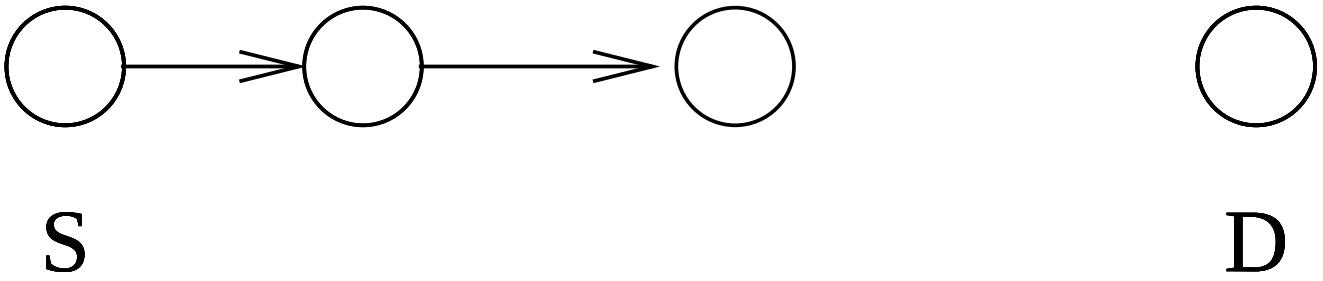
# Mesure avec `traceroute`: rappels



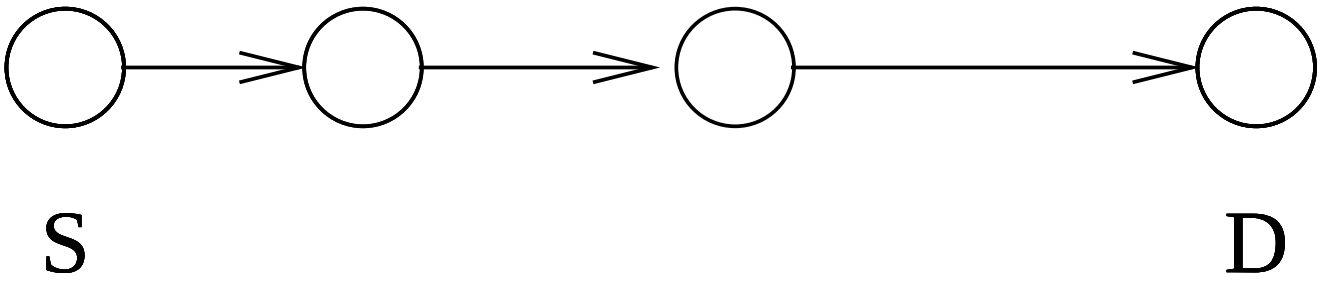
# Mesure avec `traceroute`: rappels



# Mesure avec `traceroute`: rappels



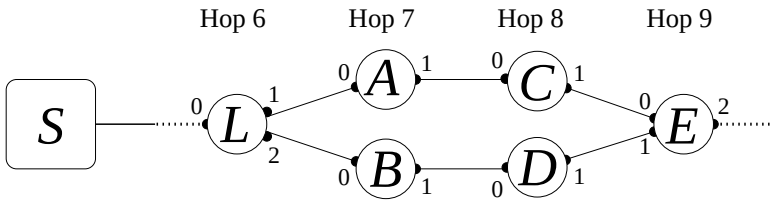
# Mesure avec `traceroute`: rappels



# Outline

# Load-balancing

Load-balancing (équilibrage de charge)



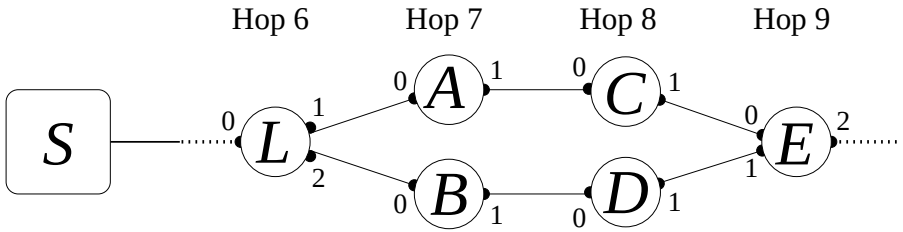

$L$ envoie les paquets pour $E$ tantôt vers $A$, tantôt vers $B$.

# Problème causé par le load-balancing

Traceroute $\longrightarrow$ Information manquante ou fausse

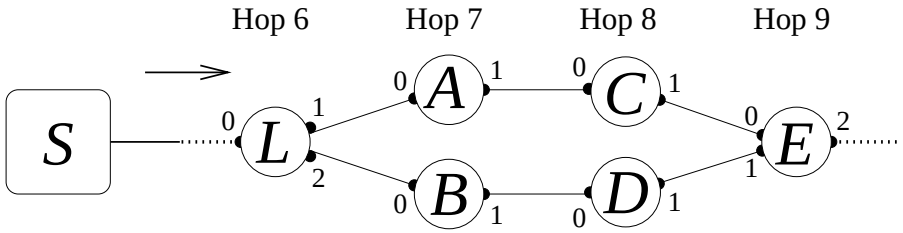
# Problème causé par le load-balancing

Traceroute $\longrightarrow$ Information manquante ou fausse



# Problème causé par le load-balancing

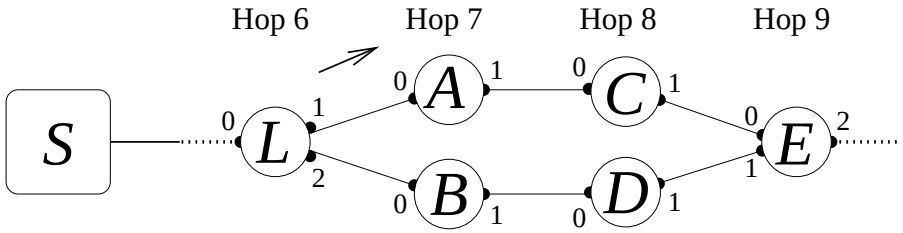
Traceroute $\longrightarrow$ Information manquante ou fausse

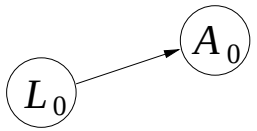



$L_0$

# Problème causé par le load-balancing

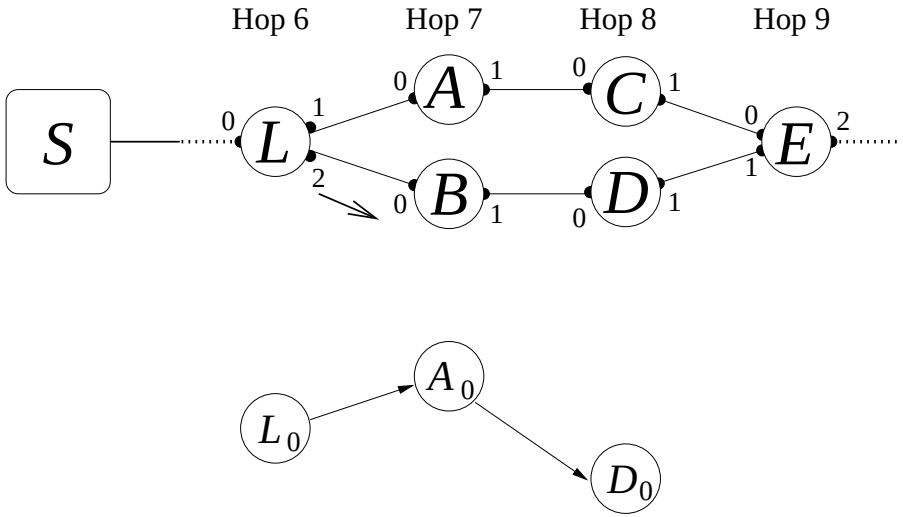
Traceroute $\longrightarrow$ Information manquante ou fausse



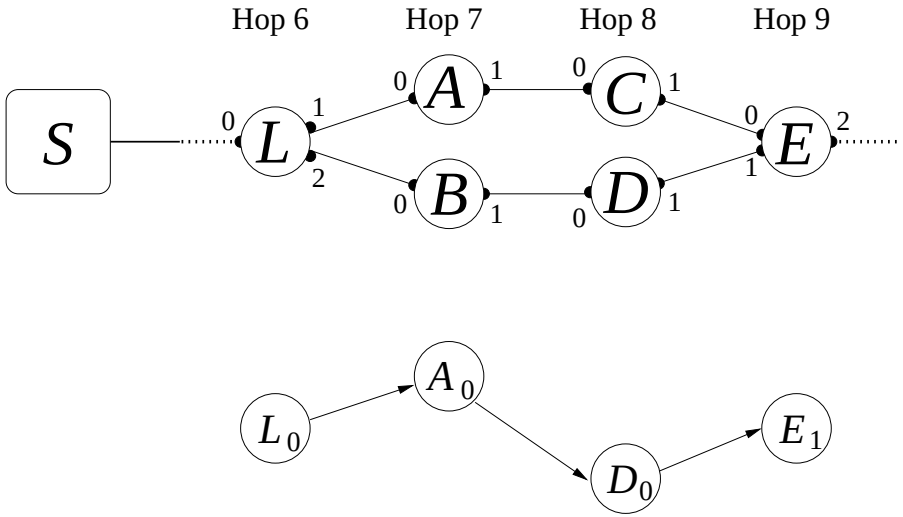
# Problème causé par le load-balancing

Traceroute $\longrightarrow$ Information manquante ou fausse



# Problème causé par le load-balancing

Traceroute $\longrightarrow$ Information manquante ou fausse





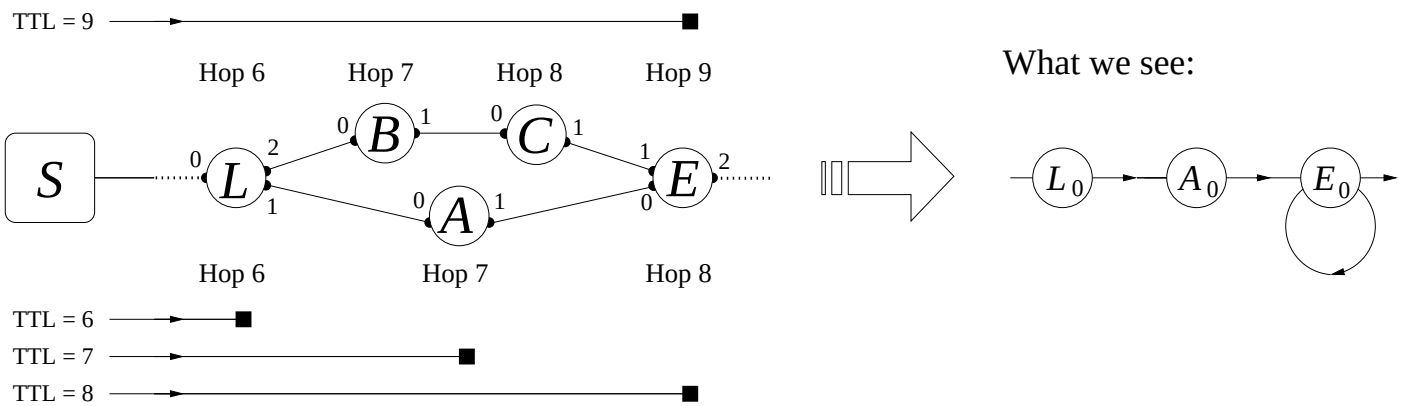


Faux liens

Nœuds non vus

# Problèmes – 2

## Boucles:



# Problèmes – 3

Cycles:

Différence de longeur entre routes $= k$
$\rightarrow$ cycle de longueur $k$

# Load-balancing

Plusieurs types de load-balancing :

- par paquet
  → aléatoire, *round-robin*
- par flot
  → paquets d'une même application sur le même chemin
- par destination
  → paquets pour une même destination sur le même chemin

# traceroute et load-balancing

- par paquet $\rightarrow$ aucun contrôle
- par flot
- par destination

# traceroute et load-balancing

- par paquet → aucun contrôle
- par flot
- par destination

Flot = IPs + ports source & destination, protocole

traceroute varie systématiquement le port destination
→ les paquets suivent systématiquement des chemins différents

# traceroute et load-balancing

- par paquet → aucun contrôle
- par flot
- par destination

Pas de biais dans les mesures avec traceroute

# Paris-traceroute

Variante de traceroute :
les ports source/dest sont gardés constants pendant toute la mesure
(utilisation d'autres champs de l'en-tête pour identifier les paquets)

→ pas de problème avec le load-balancing par flot

# Paris-traceroute

Variante de traceroute :
les ports source/dest sont gardés constants pendant toute la mesure
(utilisation d'autres champs de l'en-tête pour identifier les paquets)

→ pas de problème avec le load-balancing par flot

# Question

Quel est l'impact en pratique du load-balancing ?

Comparer des mesures avec traceroute et paris-traceroute

# Mesures

- 5000 destinations
- une passe :
  - un traceroute vers chaque destination
  - un paris-traceroute vers chaque destination
- 1 465 passes (74 jours)

# Boucles

avec traceroute : $\sim 7\%$ des IP ont une boucle

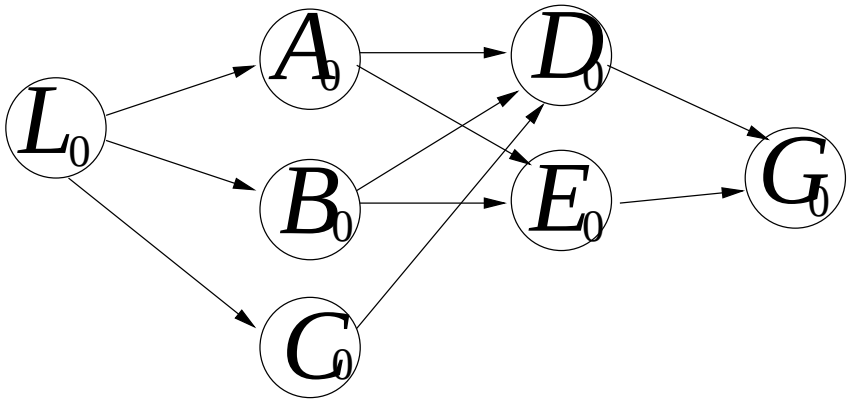
$\rightarrow$ 90% des boucles disparaissent avec paris-traceroute

# Cycles

Cycles : rares avec traceroute (fréquence 0.6%)
$\rightarrow$ 80% des cycles disparaissent avec paris-traceroute

load-balancing ou événements rares ?

# Diamants




$(L_0, D_0) \rightarrow$ taille 3
$(L_0, E_0), (A_0, G_0), (B_0, G_0) \rightarrow$ taille 2

Le load-balancing crée des diamants

# Diamants

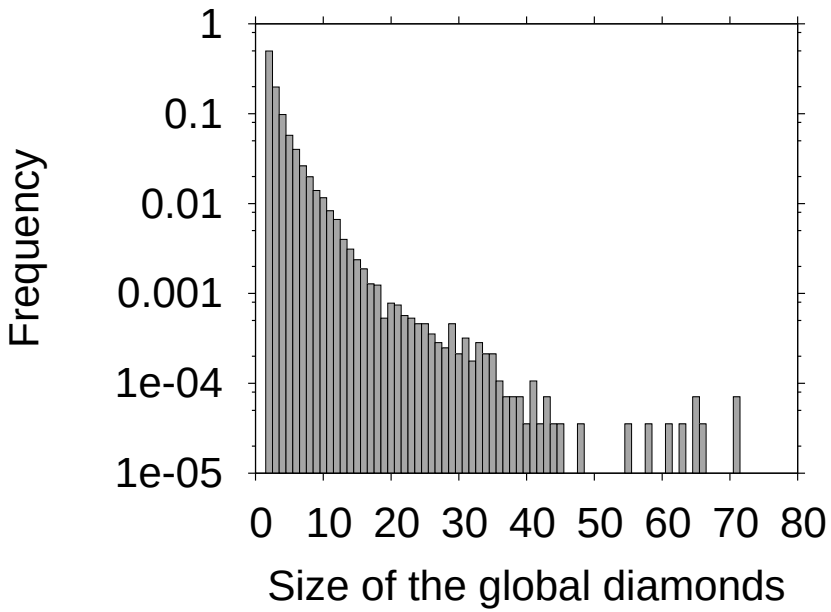
traceroute :

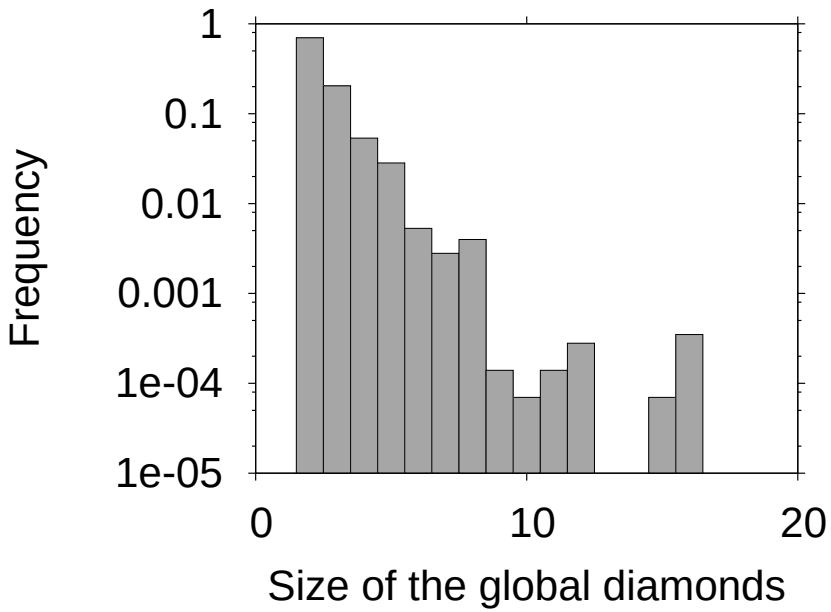
- 28 231 diamants
- diamants vers 90% des destinations
- 21.3 diamant / destination en moyenne

paris-traceroute :

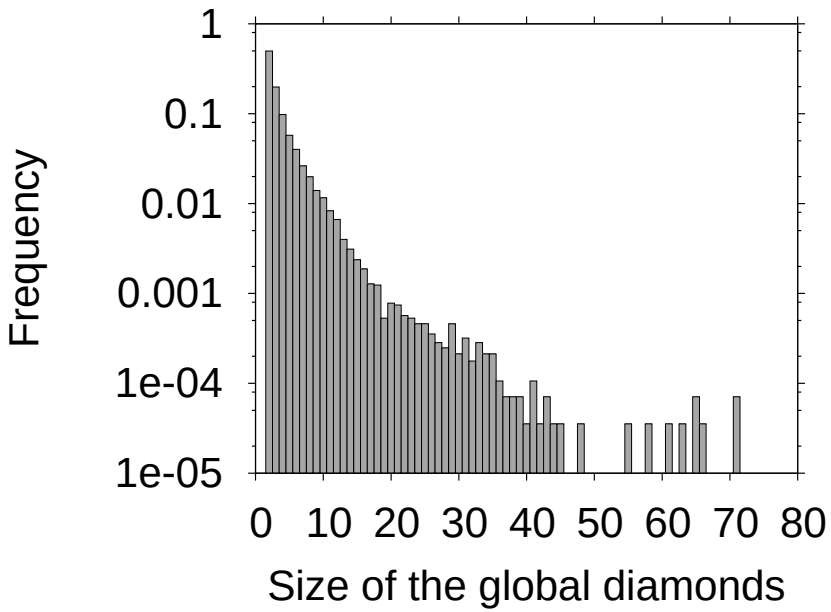
- 52% des diamants disparaissent
- diamants vers 89% des des destinations
- 9.7 diamant / destination en moyenne

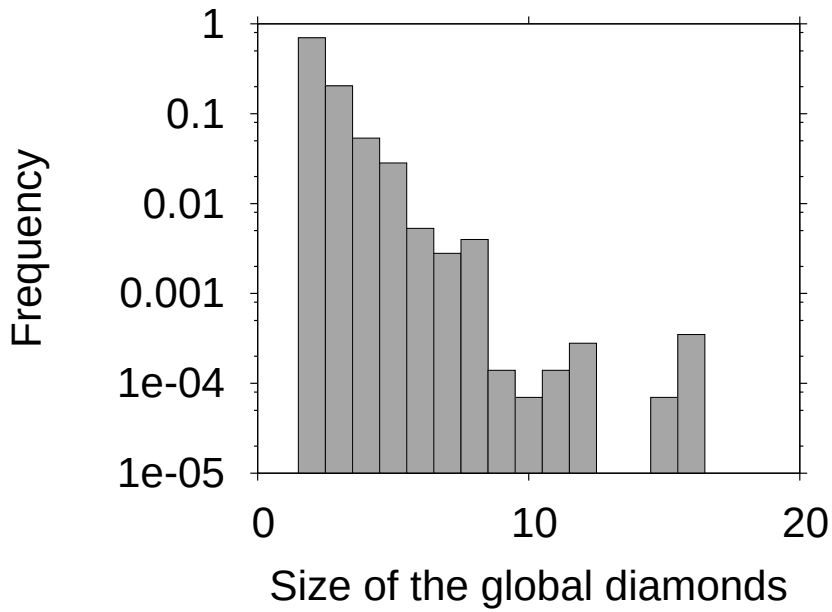
# Diamants


Frequency
1
0.1
0.01
0.001
1e-04
1e-05
0 10 20 30 40 50 60 70 80
Size of the global diamonds


Frequency
1
0.1
0.01
0.001
1e-04
1e-05
0 10 20
Size of the global diamonds

# Diamants




- Moins de diamants avec paris-traceroute
- Diamants moins gros

le load-balancing par flot crée de faux liens

# Diamants – questions

Diamants restants : load-balancing par paquet ou vrais liens ?

Certains diamants sont vus avec paris-traceroute mais pas traceroute

événements ?

# Plan

# Approches

Plusieurs approches pour détecter les événements

## Par signature

Si on sait à quoi ressemble un événement (signature)
$\rightarrow$ détection de la signature dans les données

## Top-down

Pas d'*a priori* sur la dynamique
$\longrightarrow$ on cherche à détecter les événements différents de la dynamique normale

# Méthodologie

[Detecting Events in the Dynamics of Ego-centered Measurements of the Internet Topology

Hamzaoui, Latapy, Magnien, 2010]

**Une statistique, trois cas possibles :**

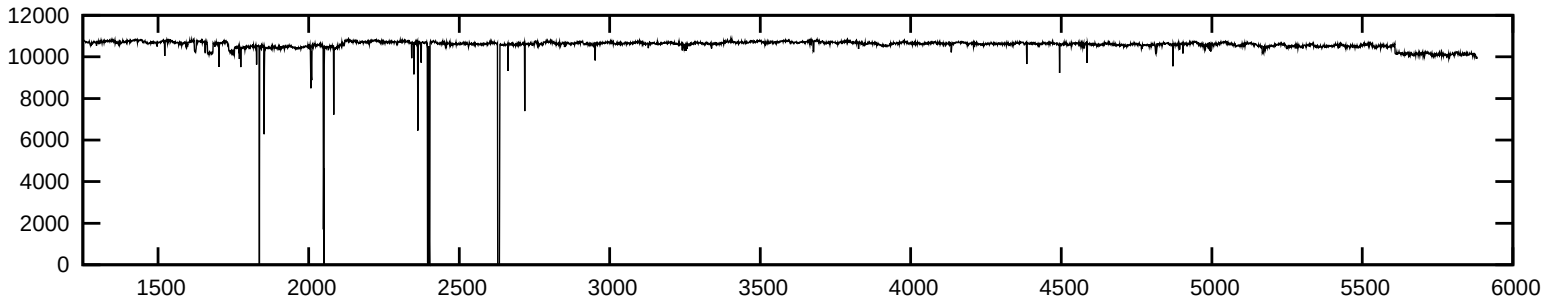
- homogène
- hétérogène
- homogène avec outliers

# Données

## Données étudiées

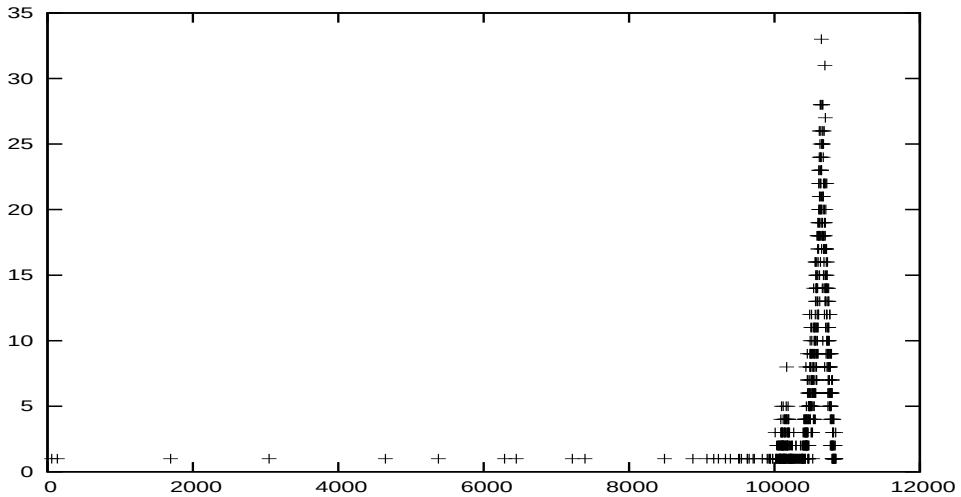
Mesure radar

# Nombre de nœuds par passe



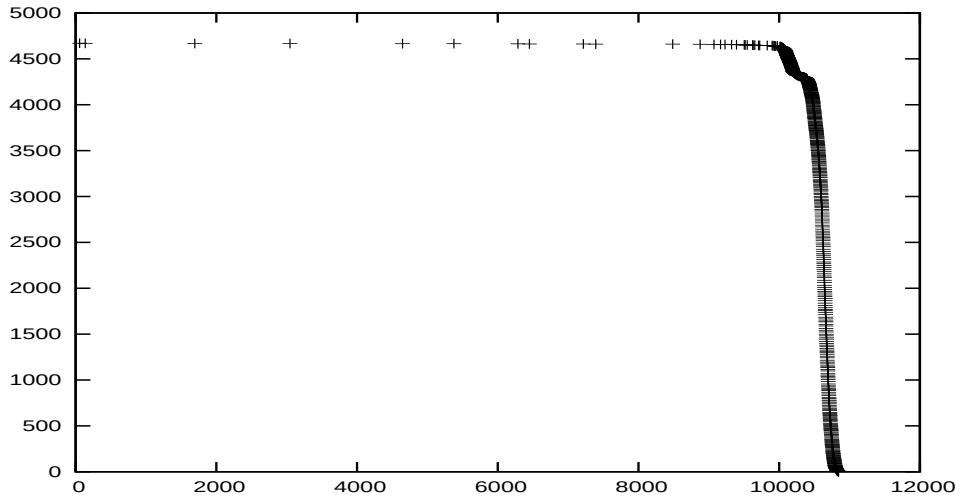

Statistique au fil du temps

# Nombre de nœuds par passe



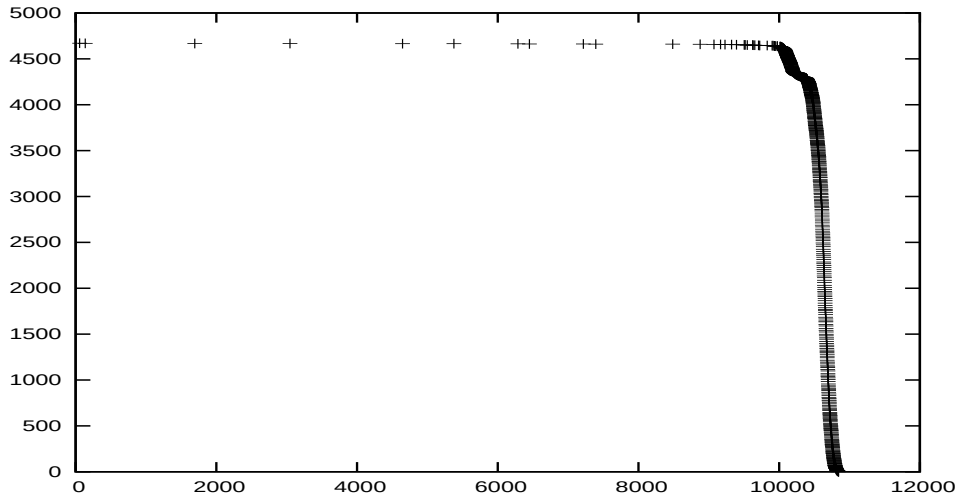

Distribution

# Nombre de nœuds par passe




Distribution cumulative inverse

# Nombre de nœuds par passe

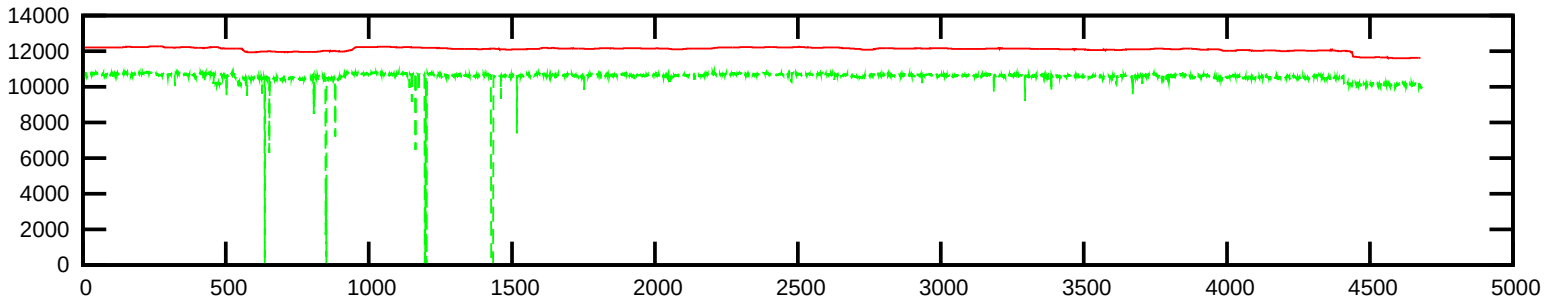



Distribution cumulative inverse

## Conclusion

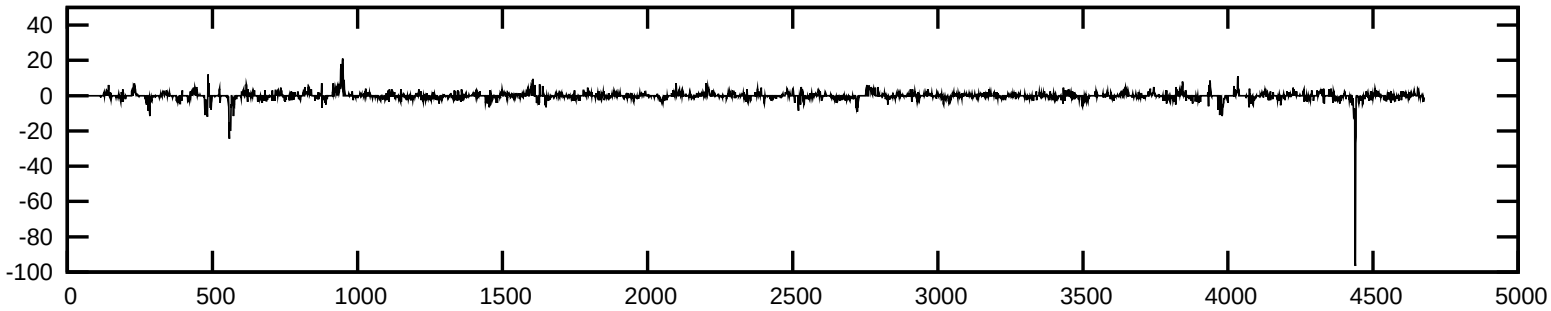
- Pics vers le bas significatifs
- Événements non intéressants (pertes de connexion locales)

# Changements de régime



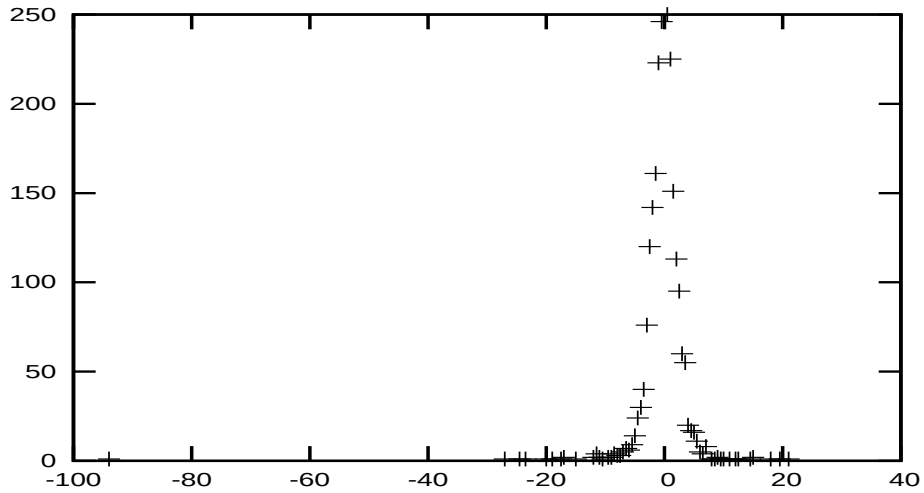

- $N_i$ : nombre de nœuds par passe
- $M_i$ : médiane $N_i \rightarrow N_{i+100}$ (courbe décalée)

# Changements de régime



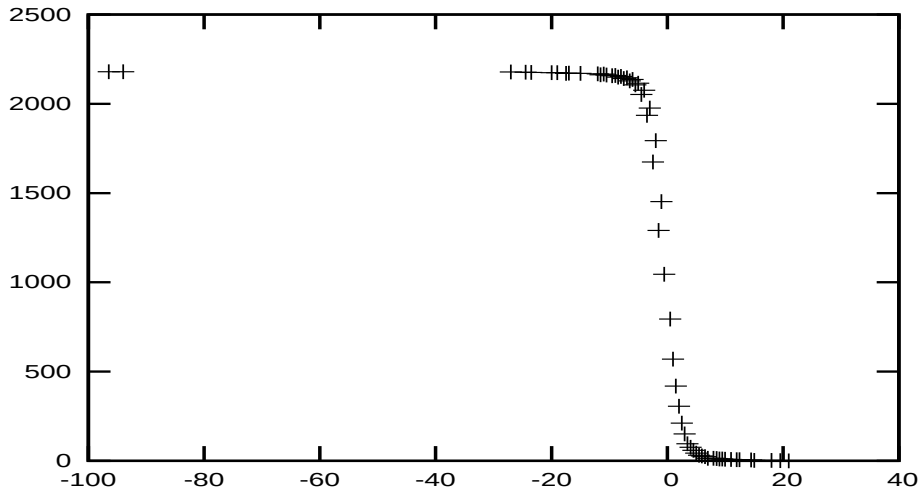

$M_i - M_{i-1}$

# Changements de régime



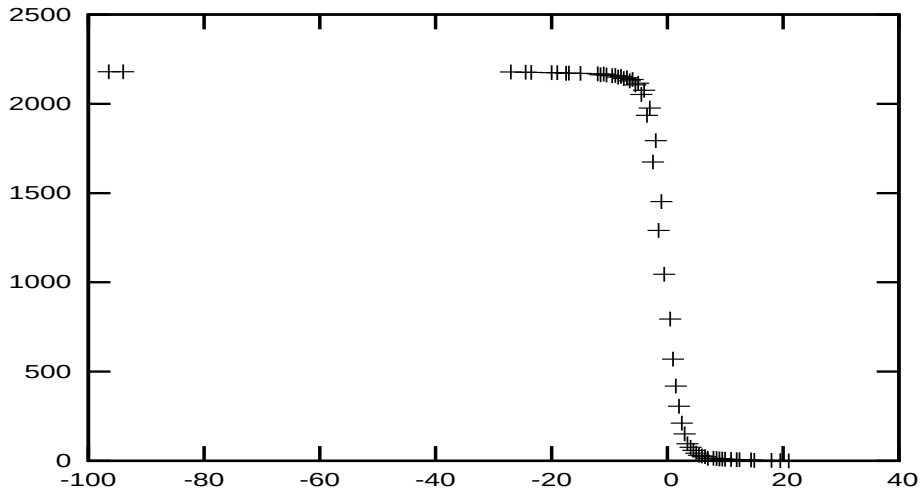

Distribution des $M_i - M_{i-1}$

# Changements de régime




Distribution cumulative inverse des $M_i - M_{i-1}$

# Changements de régime



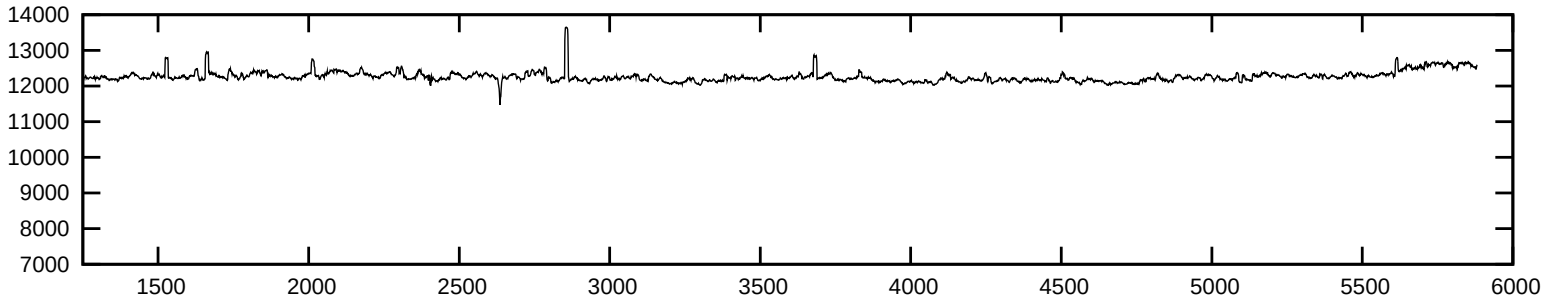

Distribution cumulative inverse des $M_i - M_{i-1}$

## Conclusion

- Présence d'outliers
- Changements de régime détectés rigoureusement

# Nombre de nœuds dans plusieurs passes consécutives

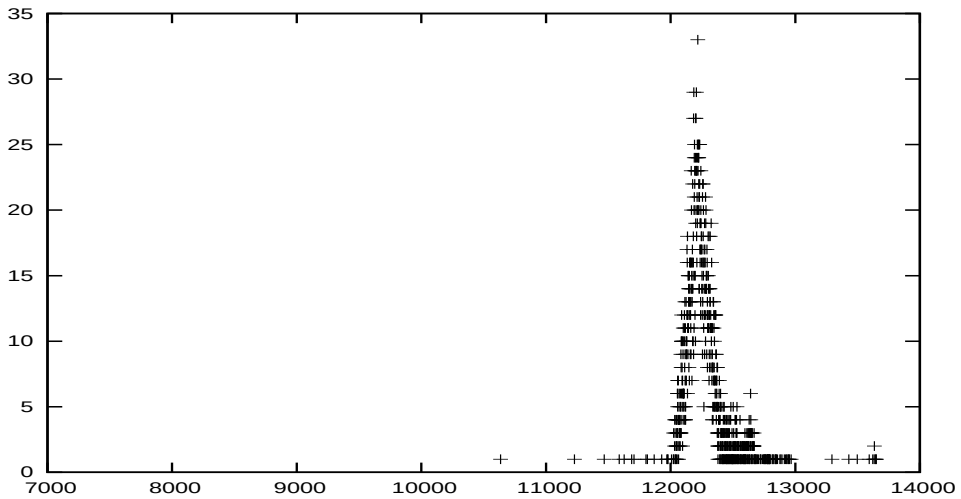
Statistique : nombre d'IP distinctes dans l'union de 5 passes




Statistique au fil du temps

# Nombre de nœuds dans plusieurs passes consécutives

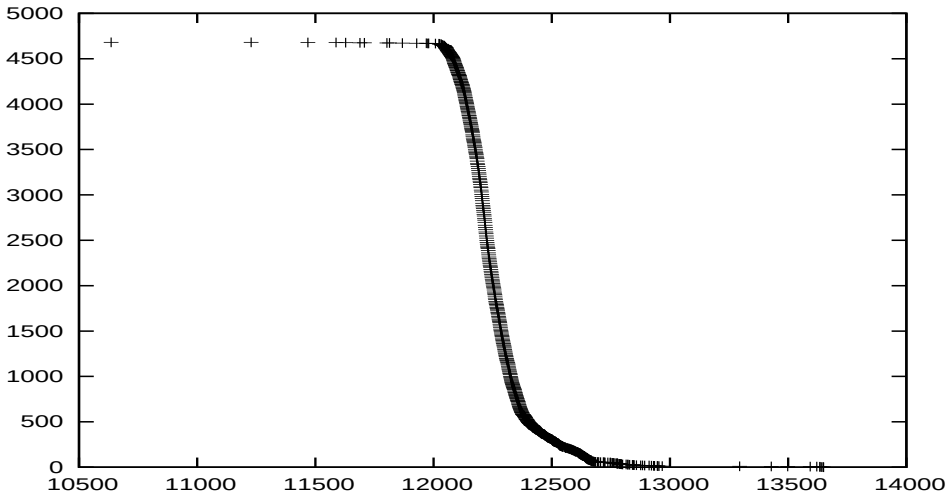
Statistique : nombre d'IP distinctes dans l'union de 5 passes




Distribution

# Nombre de nœuds dans plusieurs passes consécutives

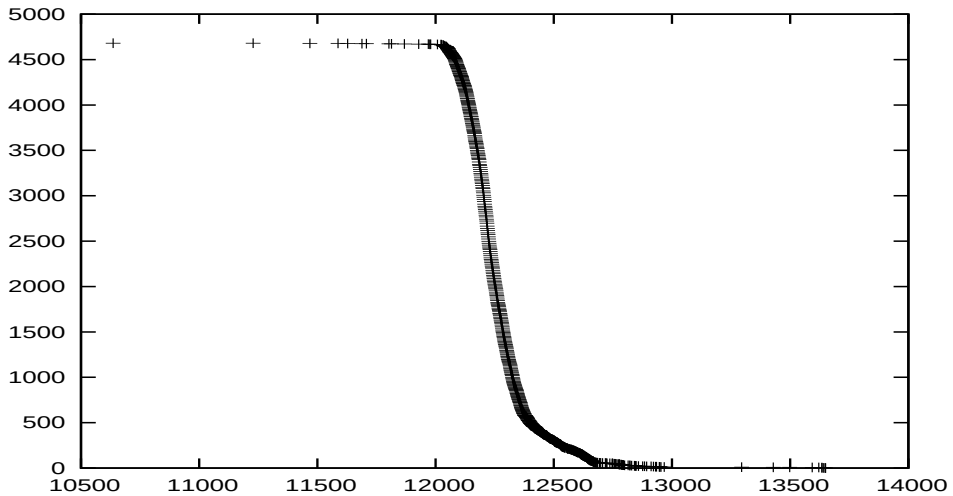
Statistique : nombre d'IP distinctes dans l'union de 5 passes




Distribution cumulative inverse

# Nombre de nœuds dans plusieurs passes consécutives

Statistique : nombre d'IP distinctes dans l'union de 5 passes

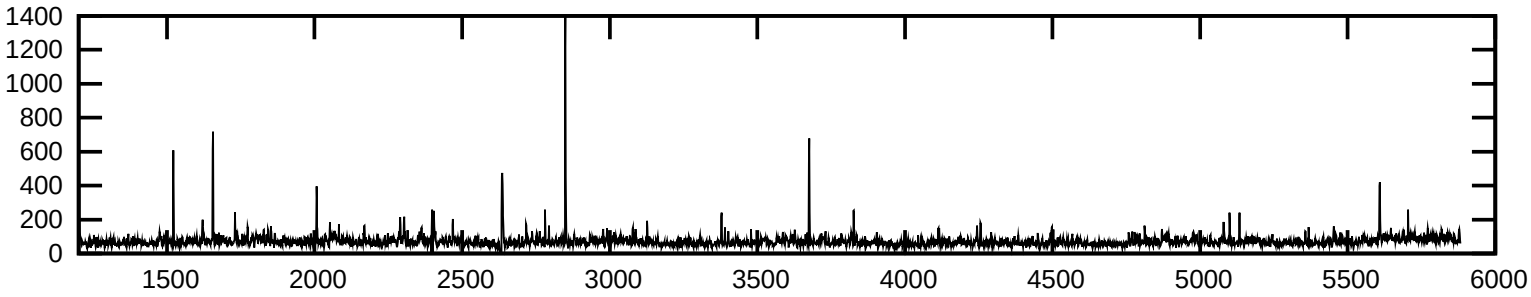



## Conclusion

- Pics significatifs
- Pics vers le haut $\rightarrow$ changements
- Pics vers le bas significatifs ? (perte de connexion locale longue ?)

# Nombre de *nouveaux* nœuds

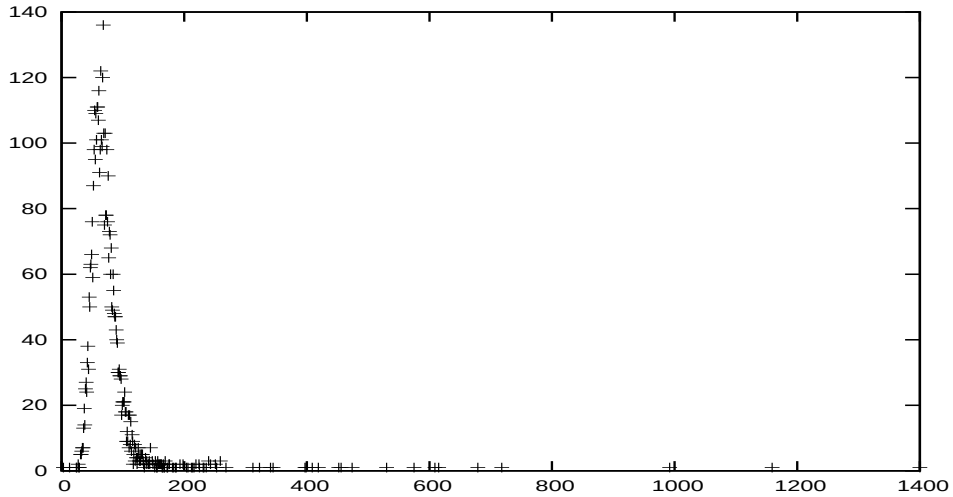
Statistique : nombre d'IP vues dans l'union de 2 passes non vues dans les 10 passes d'avant




Statistique au fil du temps

# Nombre de *nouveaux* nœuds

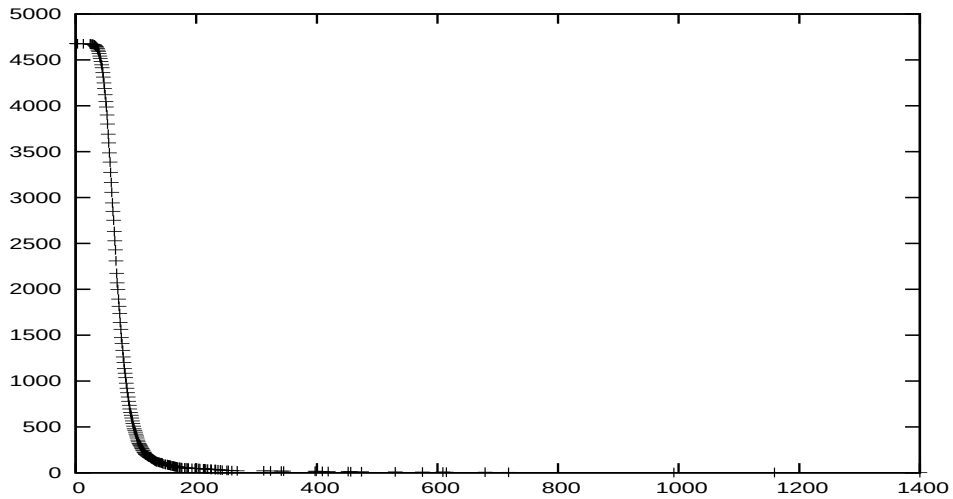
Statistique : nombre d'IP vues dans l'union de 2 passes non vues dans les 10 passes d'avant




Distribution

# Nombre de *nouveaux* nœuds

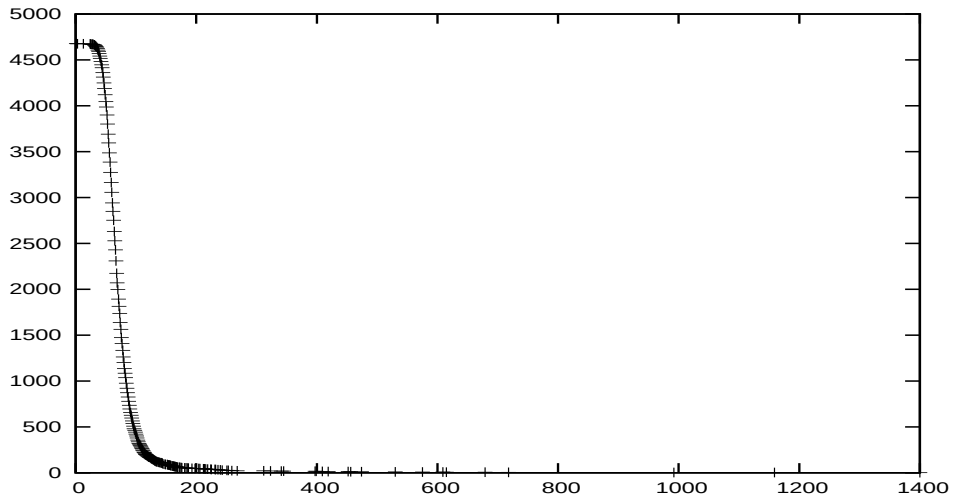
Statistique : nombre d'IP vues dans l'union de 2 passes non vues dans les 10 passes d'avant




Distribution cumulative inverse

# Nombre de *nouveaux* nœuds

Statistique : nombre d'IP vues dans l'union de 2 passes non vues dans les 10 passes d'avant



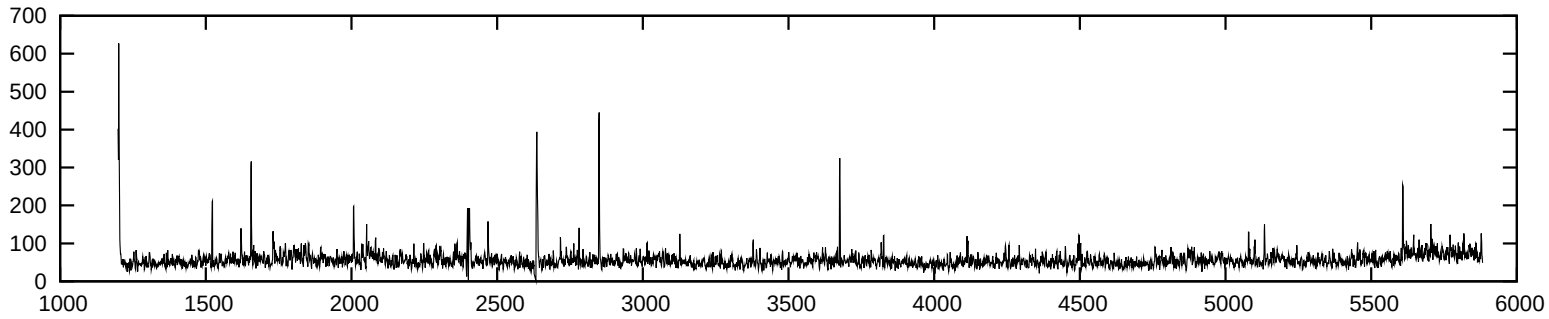

Distribution cumulative inverse

## Conclusion

- Pics significatifs
- Pics vers le haut → changements

# Nombre de composantes connexes de *nouveaux* nœuds

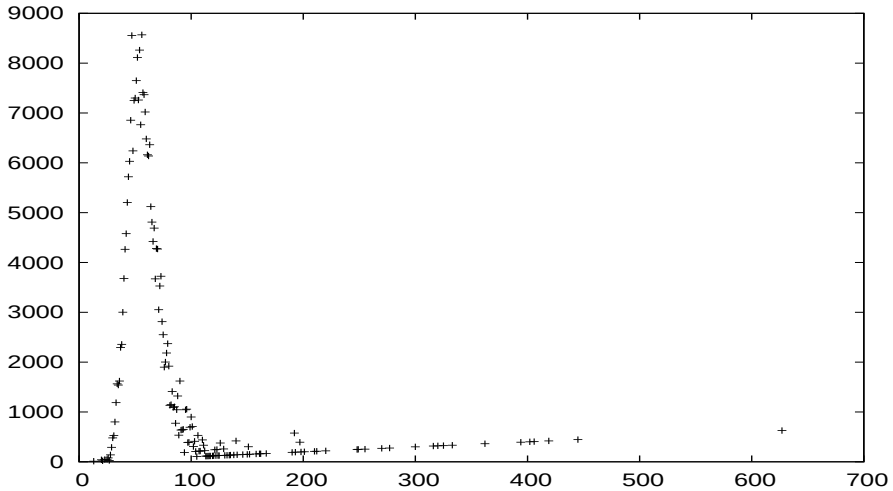
Statistique : nombre composantes connexes formées de nouveaux nœuds




Statistique au fil du temps

# Nombre de composantes connexes de *nouveaux* nœuds

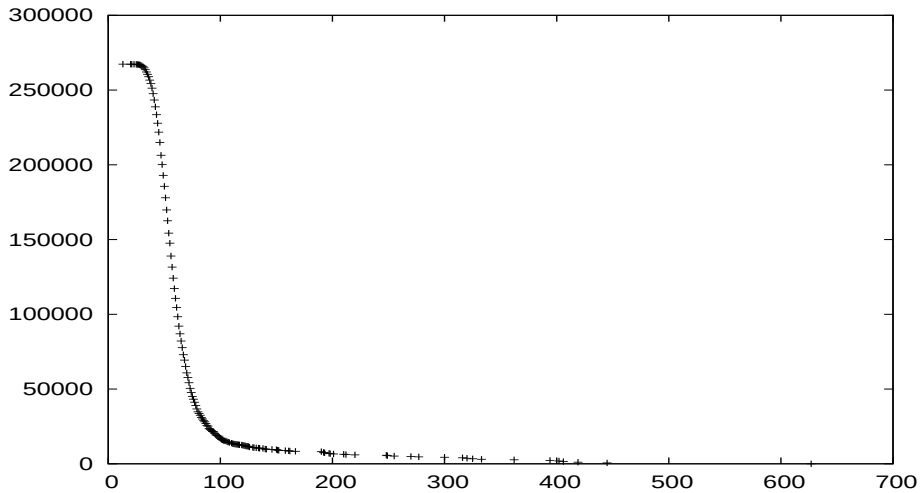
Statistique : nombre composantes connexes formées de nouveaux nœuds




Distribution

# Nombre de composantes connexes de *nouveaux* nœuds

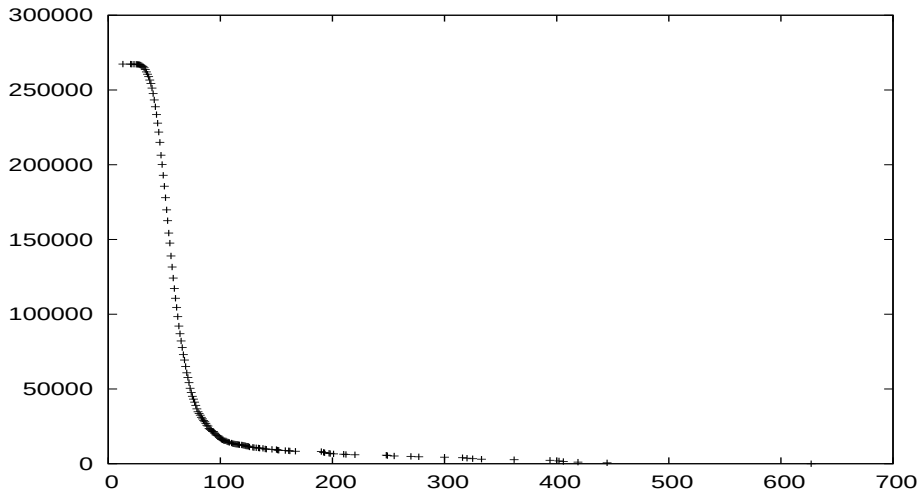
Statistique : nombre composantes connexes formées de nouveaux nœuds




Distribution cumulative inverse

# Nombre de composantes connexes de *nouveaux* nœuds

Statistique : nombre composantes connexes formées de nouveaux nœuds

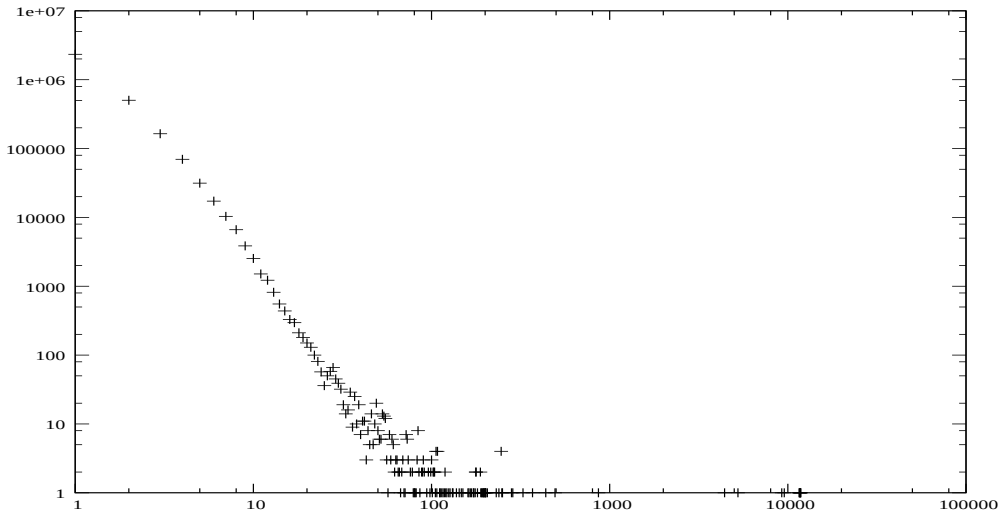



## Conclusion

- Pics significatifs
- Pics vers le haut $\rightarrow$ changements
- Mêmes événements que le nombre de nouveaux nœuds

# Taille des composantes connexes de *nouveaux* nœuds

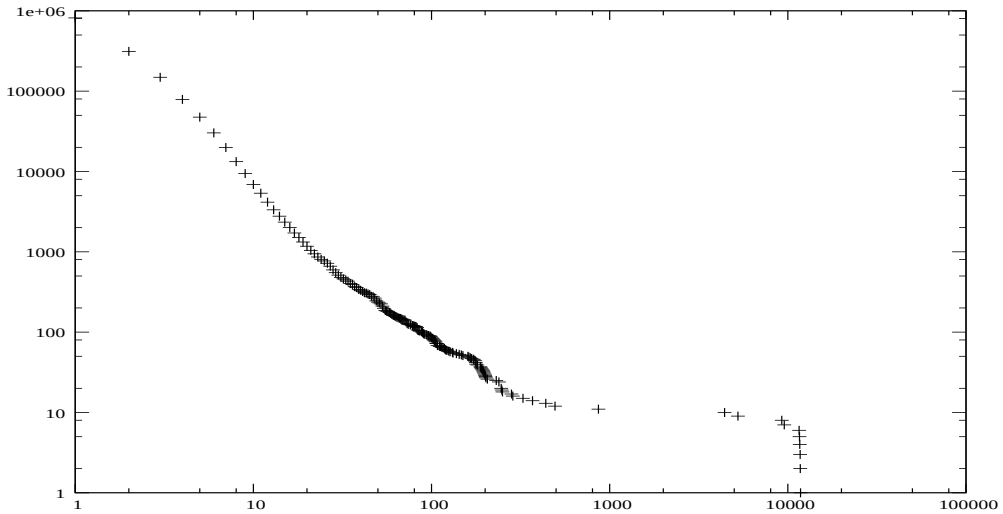
Statistique : nombre composantes connexes formées de nouveaux nœuds




Distribution

# Taille des composantes connexes de *nouveaux* nœuds

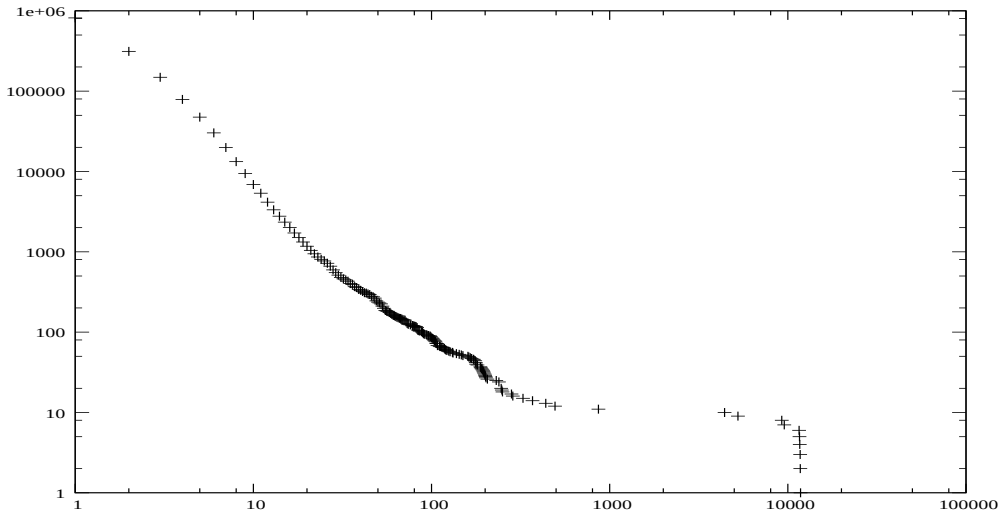
Statistique : nombre composantes connexes formées de nouveaux nœuds




Distribution cumulative inverse

# Taille des composantes connexes de *nouveaux* nœuds

Statistique : nombre composantes connexes formées de nouveaux nœuds




Distribution cumulative inverse

## Conclusion

- Distribution hétérogène
- Pas de notion de normal vs anormal

# Interprétation des événements

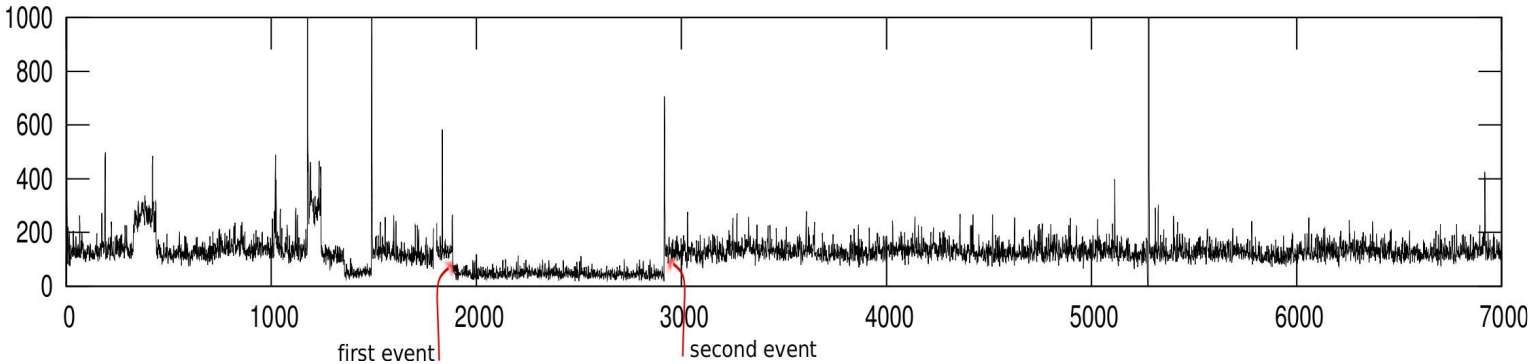
Événement statistiquement significatif détecté
→ Interprétation ?
But : compréhension, pertinence de l'événement

## Deux approches

- Corrélations avec des événements connus
- Dessin

# Tickets d'incidents Abilene

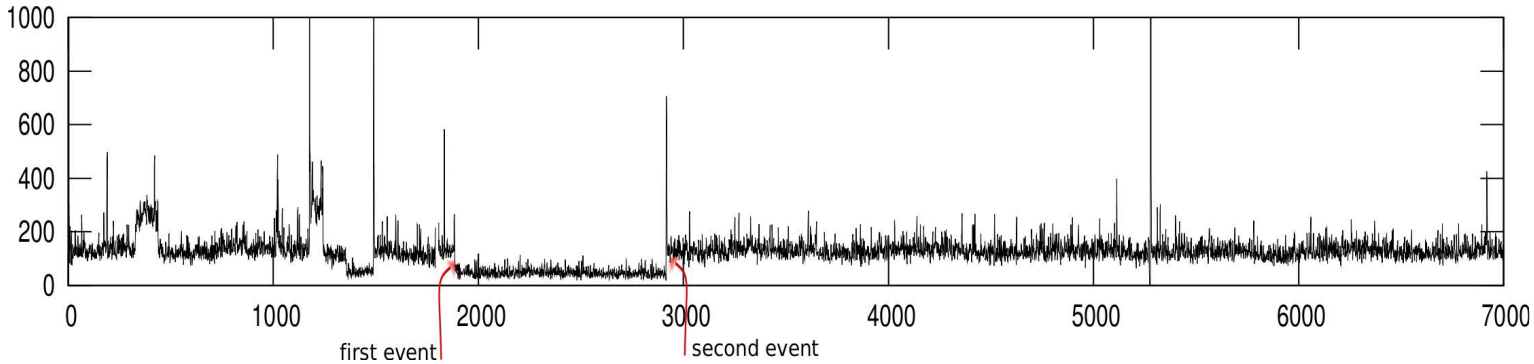
Dans certains cas, correspondance entre un incident détecté et un ticket d'incident




```
AFFECTED:       Peer SINET (CHIC)
STATUS:         Unavailable
START TIME:     Thursday, May 17, 2007, 11:47 AM (1147) UTC
END TIME:       Pending
DESCRIPTION:    Peer SINET's connection the Internet2 IP
                Community is unavailable.  SINET Engineers
                have been contacted, however, no cause of
                outage has been provided yet. SINET is multi-homed.
TICKET NO.:     10201:45
TIMESTAMP:      07-05-18 00:40:43 UTC
```

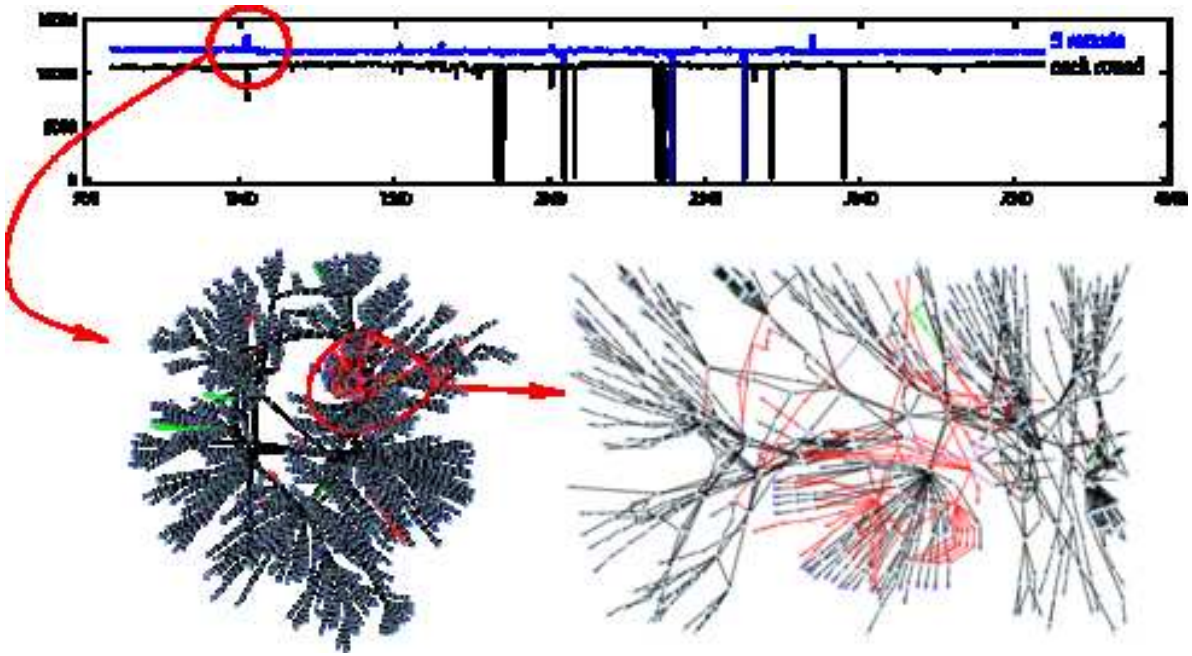
# Tickets d'incidents Abilene

Dans certains cas, correspondance entre un incident détecté et un ticket d'incident




```
STATUS:          Available
START TIME:      Thursday, May 17, 2007, 11:47 AM (1147) UTC
END TIME:        Friday, May 18, 2007, 3:51 AM (0351) UTC
DESCRIPTION:     Peer SINET was unavailable to the the Internet2 IP
                 Network Community.   SINET Engineers reported the
                 reason for outage was due to a fiber cut in New York.
                 SINET is multi-homed.
TICKET NO.:      10201:45
TIMESTAMP:       07-05-18 07:39:16 UTC
```

# Dessin



# Pour aller plus loin

- Caractérisation automatique du type de distribution

  fit

- Détection automatique et rigoureuse des outliers

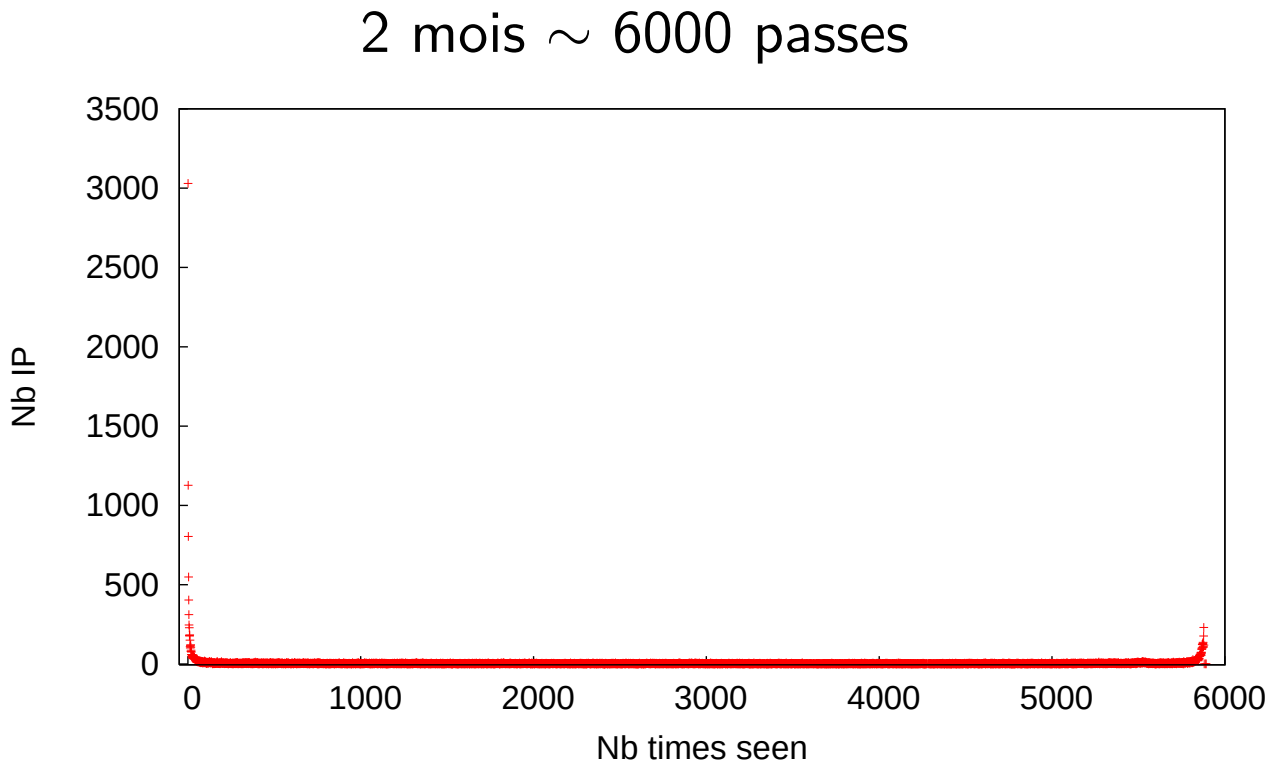
# Plan

# Présence des IP

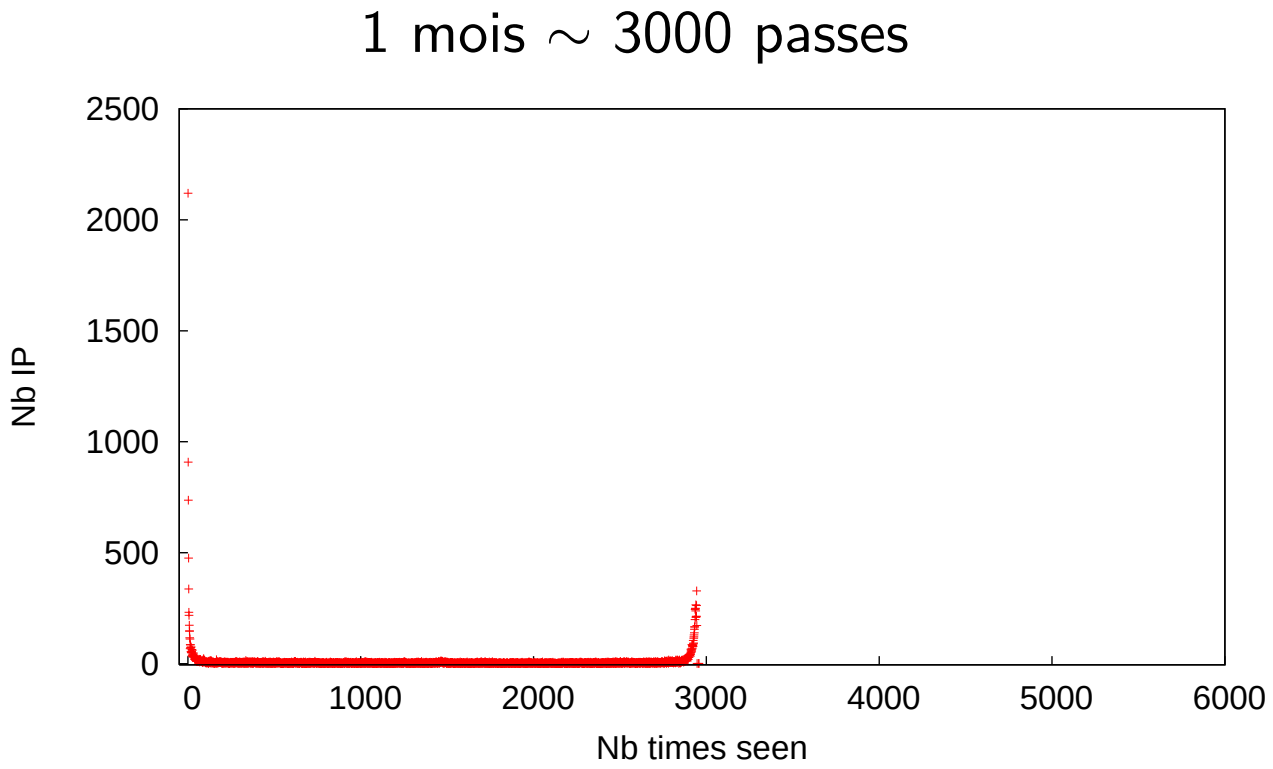
Distribution du nombre de présences

Toutes les IP vues pendant la mesure ($\sim 29\,000$)
Pour chaque IP : nombre de passes où on l'a vue
$\longrightarrow$ Distribution

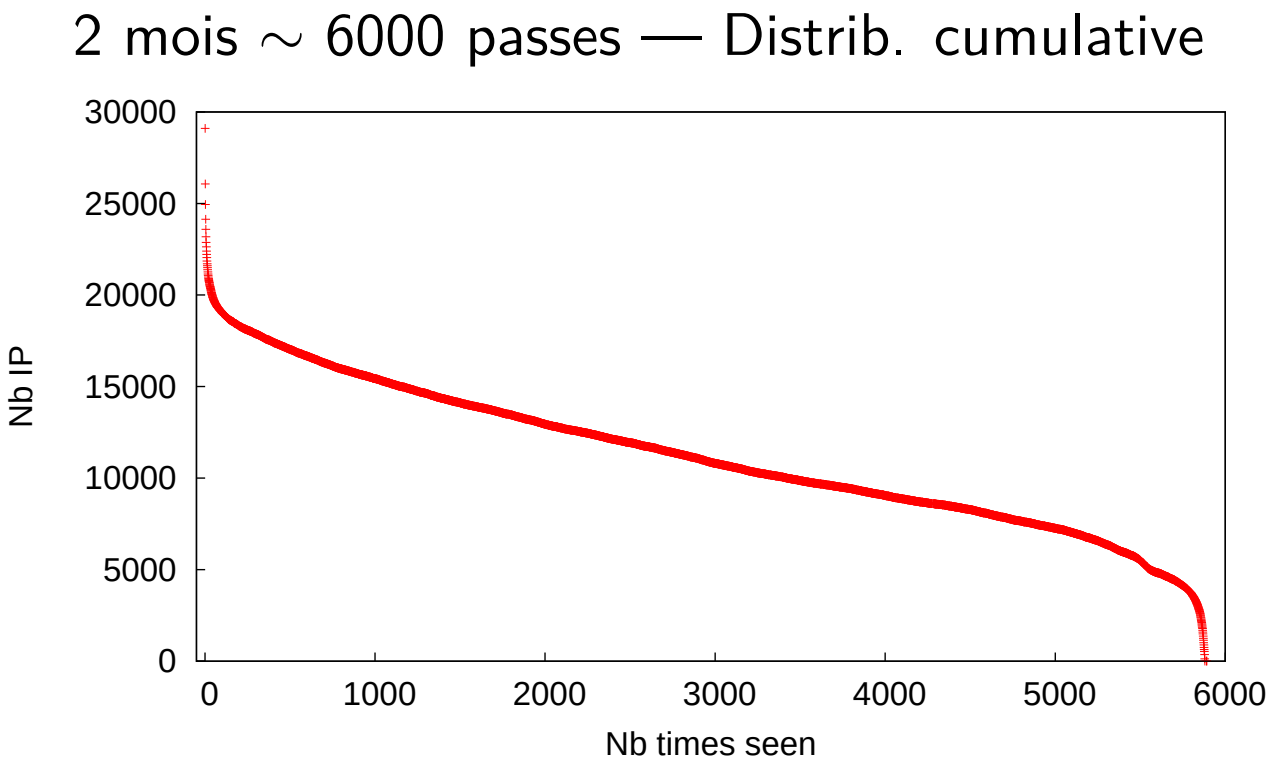
# Présence des IP



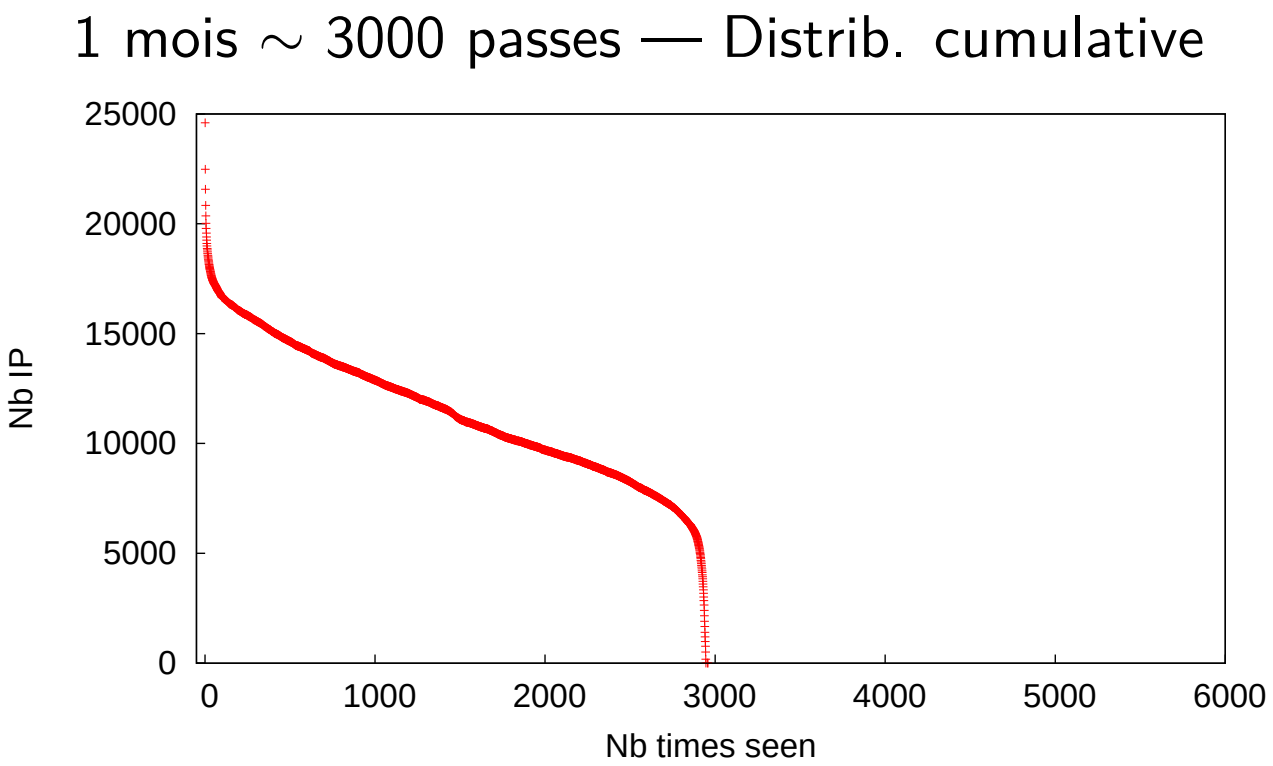
# Présence des IP



# Présence des IP



# Présence des IP



# Observations dépendantes de la durée de la mesure ?

Mesure plus longue :

- Moins d'IP stables
- Plus d'IP éphémères

Conclusion ?

## Méthodologie

- Augmenter la taille de la fenêtre de mesure
- Évolution de la distribution

# Distributions normalisées

## Problème :

- Valeur max en x : nombre de passes
- Valeur max en y : nombre total d'IP vues en $N$ passes

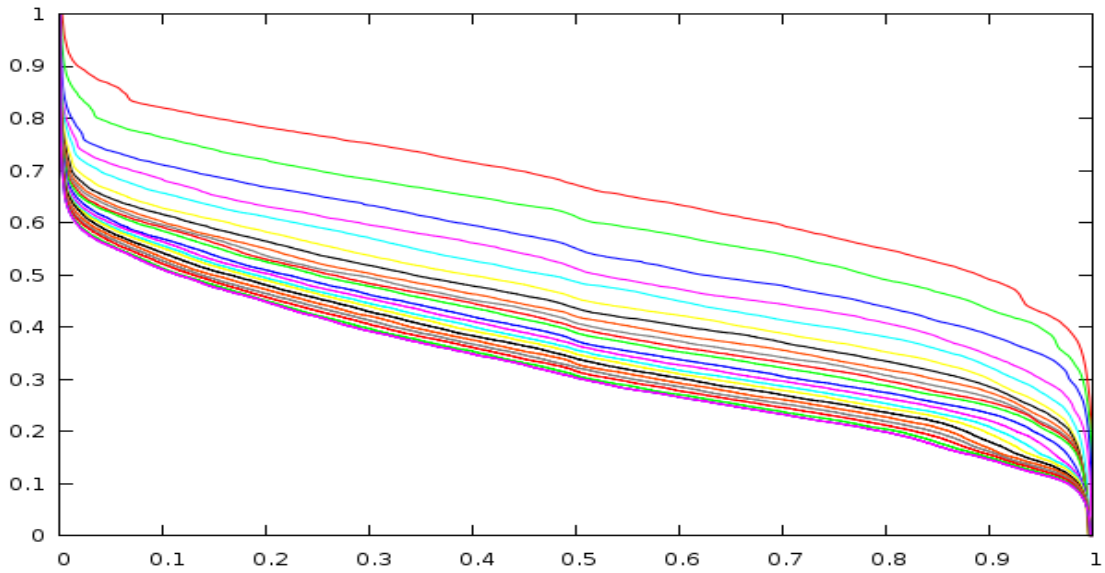
Les courbes ne sont pas directement comparables

## Normalisation

- $y$ : fraction des IP observées
- $x$ : fraction des passes

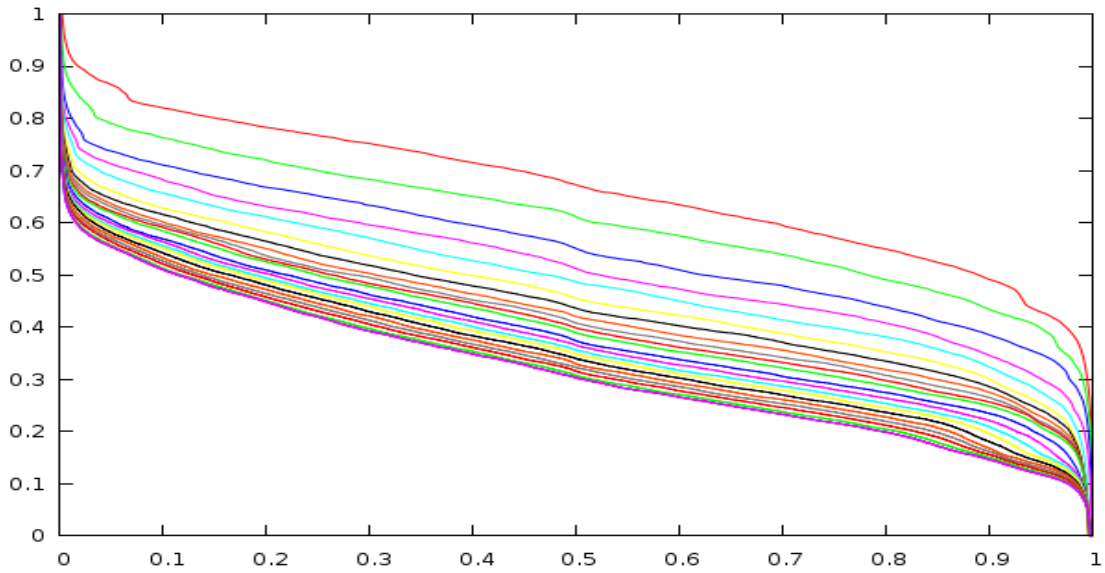
# Résultats

10 000 passes $\rightarrow \sim$ 100 jours



# Résultats

10 000 passes $\rightarrow \sim$ 100 jours



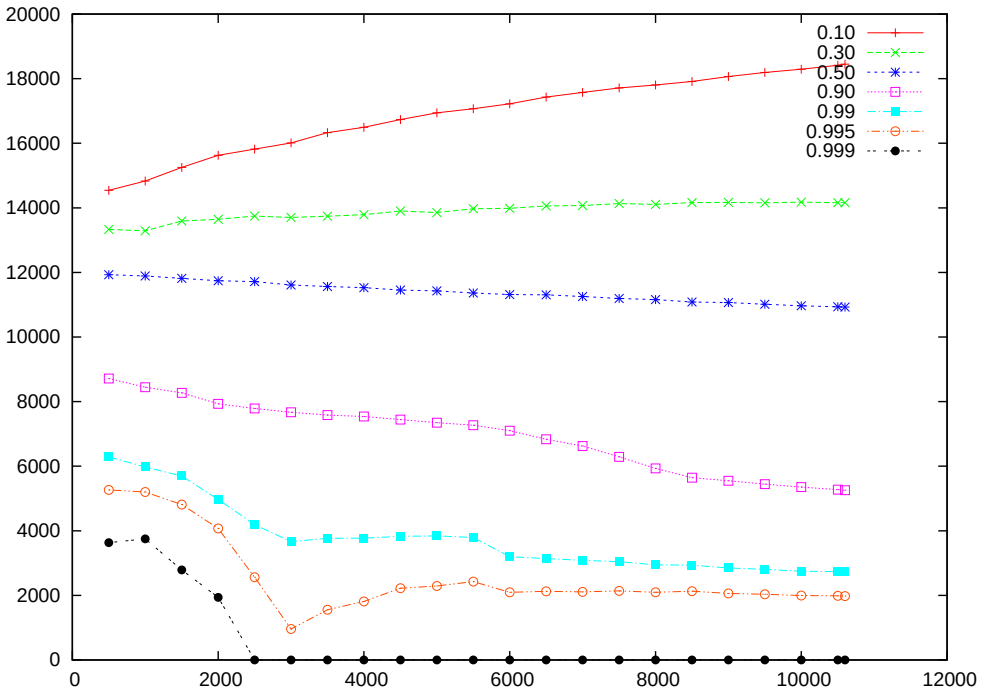

Mesure plus longue :

- Moins d'IP stables
- Plus d'IP éphémères

Pas de convergence

# Étudier la convergence

Nombre d'IP vues au moins $x$% des passes,
en fonction de la durée de mesure

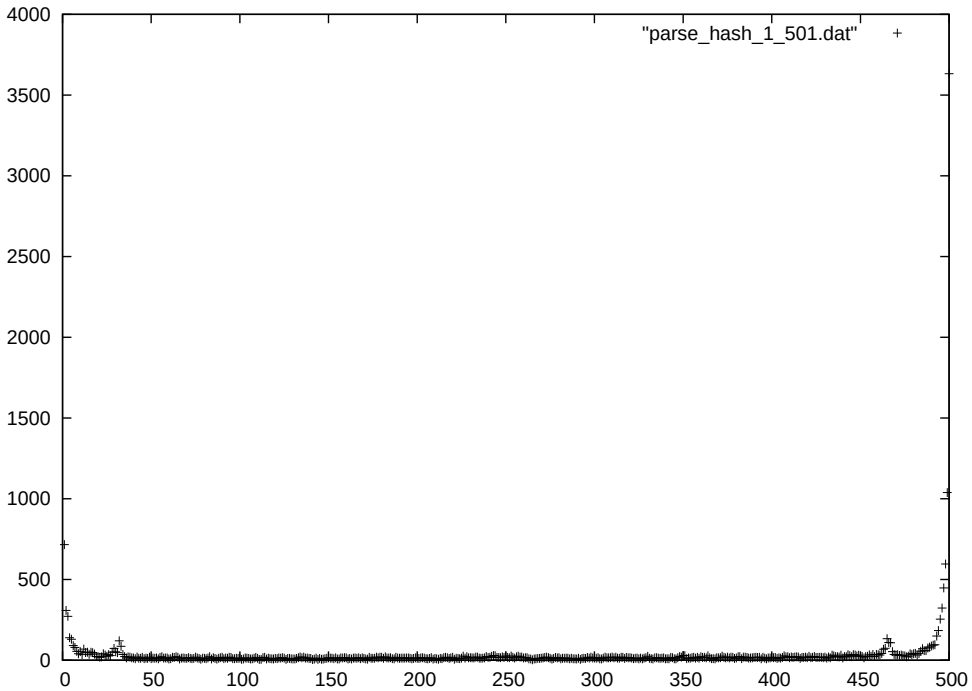



$\sim$ 2 000 IP présentes à 99.5% des passes

$\rightarrow$ noyau stable ?

# Lien avec la détection d'événements

Pour certains moniteurs / certaines périodes




500 passes au total

Beaucoup d'IP vues 25 ou 475 passes

événement de durée 25 passes

vidéo events.avi

# Prochains cours

Prochaine séance : 20 janvier (cours + TP)

Dernière séance : 27 janvier 8h30-12h30 – 14-15/503

Conflits EdT ?