

Dynamique des vues ego-centrées de la topologie de l'internet : analyse et modélisation

Clémence Magnien
Amélie Medem, Fabien Tarissan, Sergey Kirgizov

Équipe *Complex Networks*
LIP6 - CNRS - Université Pierre et Marie Curie

8 mars 2012

Outline

- 1 Définition et mesure
- 2 Analyse
- 3 Modélisation

Topologie de l'Internet au niveau IP

- Routeurs : adresses IP
- Liens au niveau IP

→ Internet vu comme un **graphe**

Pourquoi ?

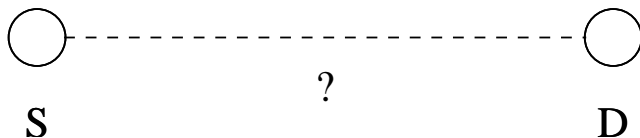
Étudier la structure

- Identifier des faiblesses
- Faire des simulations
- Preuves, approches formelles

Mesure de la topologie

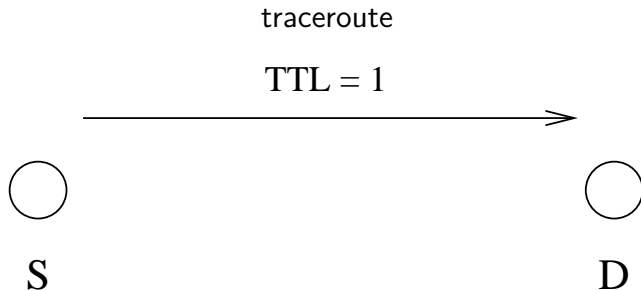
Pas de carte officielle disponible → Besoin de **mesurer**

traceroute



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**

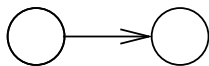
traceroute



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**

traceroute



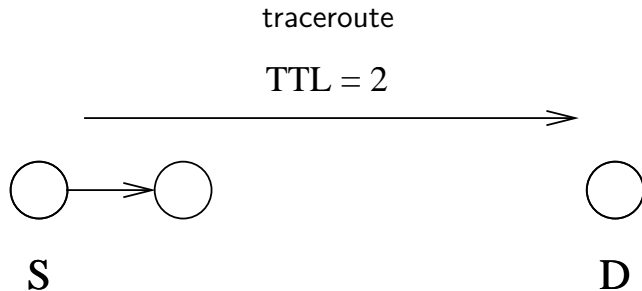
S



D

Mesure de la topologie

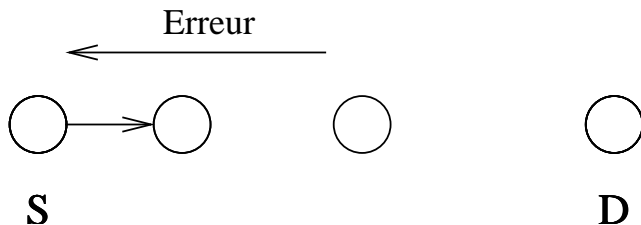
Pas de carte officielle disponible → Besoin de **mesurer**



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**

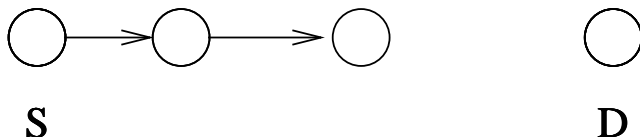
traceroute



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**

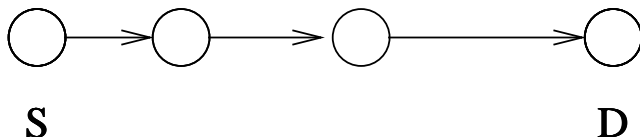
traceroute



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**

traceroute



Mesure de la topologie

Pas de carte officielle disponible → Besoin de **mesurer**

traceroute

Plus d'informations

Multiplier

- Les destinations (\sim millions)
- Les sources (\sim quelques dizaines)

Coûteux !

Biais dans les propriétés observées

[Infocom 02, Infocom 03, STOC 05, ...]

- Mesure \longrightarrow échantillon \neq graphe réel
- Propriétés échantillon \neq Propriétés réelles

\longrightarrow Travaux théoriques et empiriques pour :

- Évaluer le biais
- Évaluer de façon sûre certaines propriétés

Pas de méthode parfaite

Dynamique globale de la topologie : difficultés

Coût

traceroute

- depuis plusieurs sources
- vers un grand nombre de destinations

→ Répétition, **fréquence faible**

Biais

Représentativité des données (partielles) obtenues ?

→ Représentativité de la dynamique ?

Un radar pour l'internet

Approche orthogonale : vue ego-centrée

Routes entre :

- un moniteur
- plusieurs destinations fixes

Avantages :

- objet bien défini
- taille réduite → mesure rapide

→ fréquence élevée (radar)

Un radar pour l'internet

Approche orthogonale : vue ego-centrée

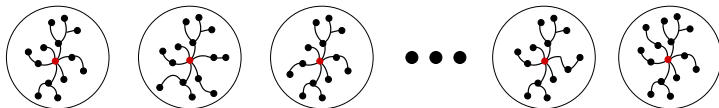
Routes entre :

- un moniteur
- plusieurs destinations fixes

Avantages :

- objet bien défini
- taille réduite → mesure rapide

→ fréquence élevée (radar)



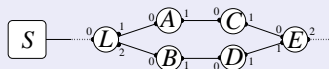
Outline

- 1 Définition et mesure
- 2 Analyse
- 3 Modélisation

Intuitions sur la dynamique

Dynamiques attendues

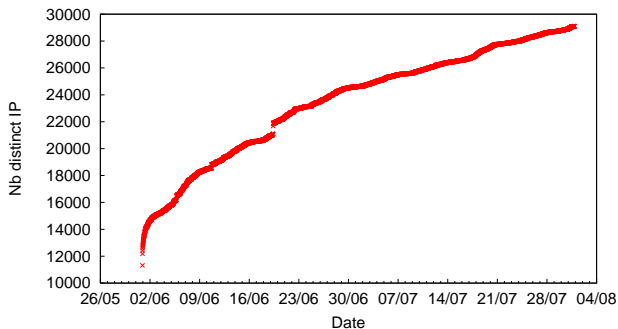
- *load-balancing*
- évolution de la topologie
- événements



Vitesse de la dynamique

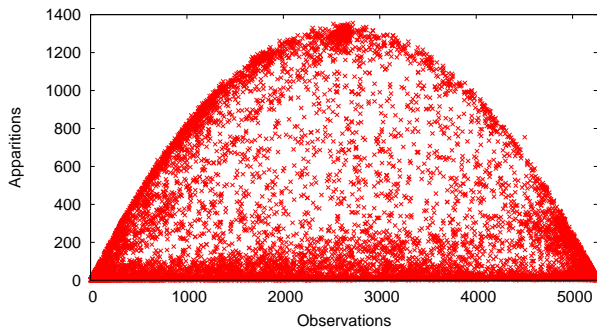
[Magnien et. al, 2009]

Nombre d'adresses IP distinctes vues depuis le début de la mesure



**découverte en permanence de nouvelles adresses IP
à une vitesse élevée**

Observations vs apparitions

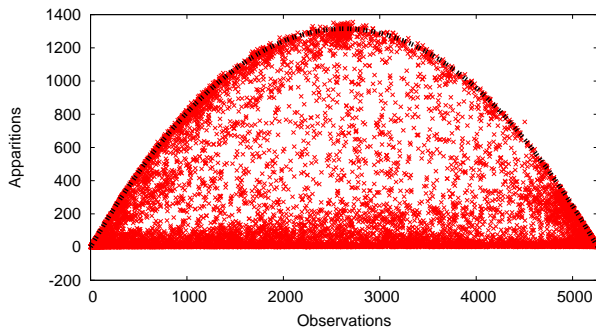


1 2 3 4 5 6 7 8

● ● ● ● ●

5 observations, 2 apparitions

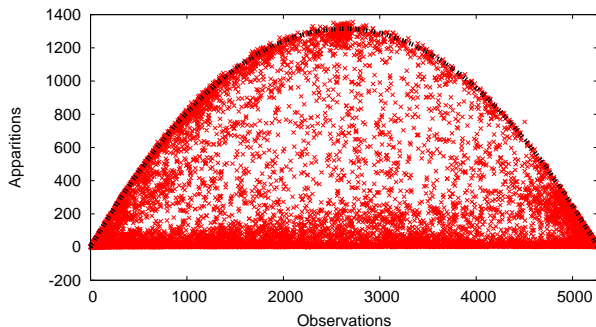
Observations vs apparitions



Parabole : *load-balancing*

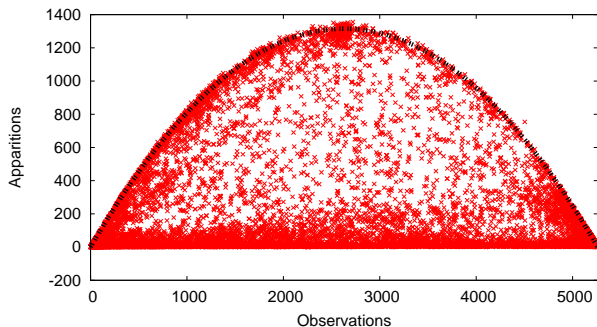
$$p = 1/2 \rightarrow x = n/2, y = 1/2 * 1/2 * n = n/4$$

Observations vs apparitions



$$p \text{ quelconque : } x = np, y = p(1 - p)n$$
$$\longrightarrow y = n * x/n * ((n - x)/n)$$

Observations vs apparitions

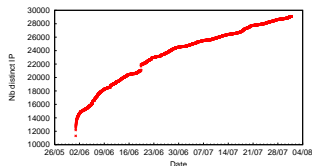


Deux classes différentes

parabole : *load-balancing*

proche de l'axe des x : *addresses stables*

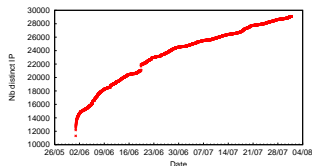
Impact des différents facteurs



Évolution des adresses causée par

- le *load-balancing*
 - plus de passes → plus d'IP
- des changements de routage
 - plus de temps → plus d'IP

Impact des différents facteurs



Évolution des adresses causée par

- le *load-balancing*
 - plus de passes → plus d'IP
- des changements de routage
 - plus de temps → plus d'IP

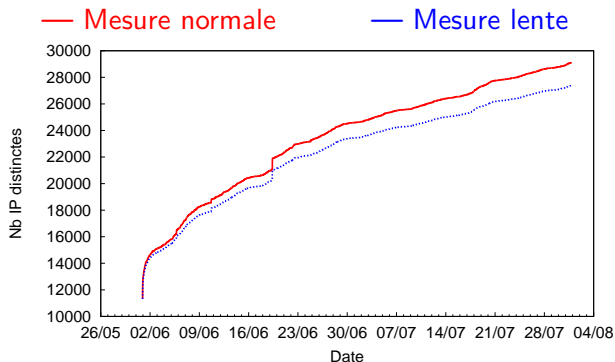
Importance du **temps** et du **nombre de passes** : fréquence

Mesure normale et mesure lente

Deux mesures

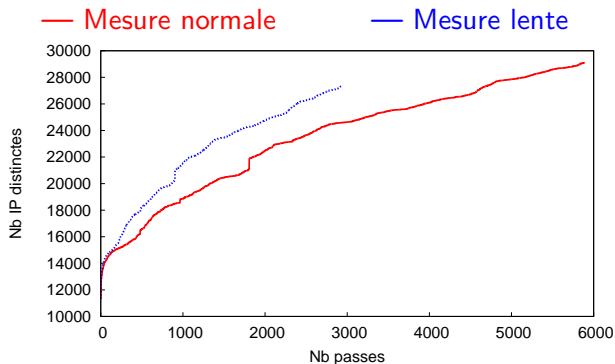
- fréquences différentes
 - même moniteur/destinations
 - même moment
-
- Mesure radar normale
 - → une passe sur deux : mesure lente

Mesure normale et mesure lente



Temps égal, plus de passes : plus d'IP \rightarrow *load-balancing*

Mesure normale et mesure lente



Nombre de passes égal, plus de temps : plus d'IP
 → changements de routage

Outline

- 1 Définition et mesure
- 2 Analyse
- 3 Modélisation**

Modélisation – routage et load balancing

Route : plus court chemin

Parcours en largeur aléatoire

Tant que file non vide :

$v = \text{défiler}(F)$

Pour chaque voisin u de v :

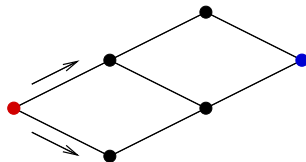
Si u non marqué, alors

$\text{pere}[u] \leftarrow v$

marquer u

enfiler u

ordre aléatoire



Modélisation – dynamique de la topologie

Topologie = graphe aléatoire (Erdős-Rényi)

Paramètre : nombre d'arêtes à changer k

À chaque étape : k *swap*

- supprimer une arête
- ajouter une arête aléatoire

Modélisation – dynamique de la topologie

Topologie = graphe aléatoire (Erdős-Rényi)

Paramètre : nombre d'arêtes à changer k

À chaque étape : k *swap*

- supprimer une arête
- ajouter une arête aléatoire

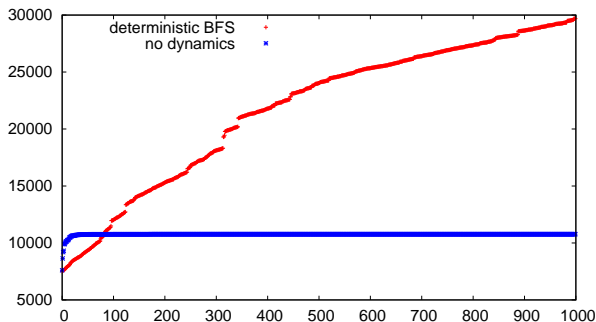
Au total

- Générer un graphe aléatoire
- Choisir une source et d destinations au hasard
- À chaque étape :
 - parcours en largeur aléatoire
 - k *swaps*

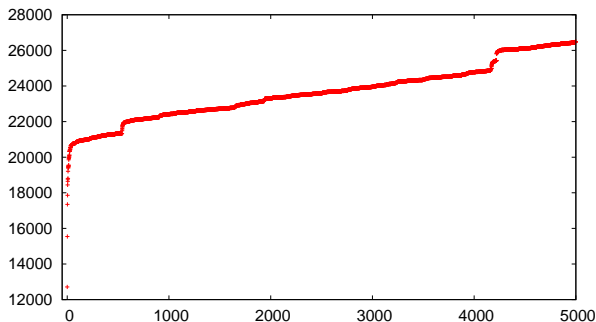
Principe – Illustration

— Parcours en largeur déterministe

— Pas de dynamique

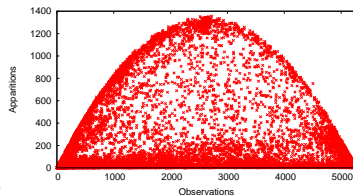
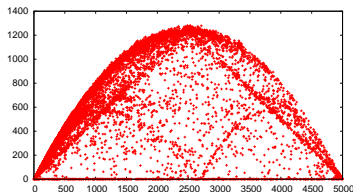


Premières observations



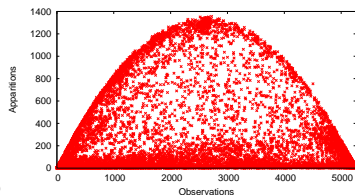
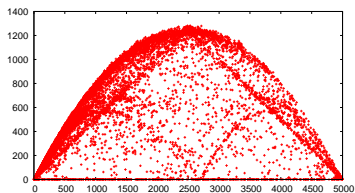
- $n = 500,000$
- $m = 1,000,000$
- 3000 destinations
- 5 *swaps* / passe

Premières observations



- $n = 500,000$
- $m = 1,000,000$
- 3000 destinations
- 5 *swaps* / passe

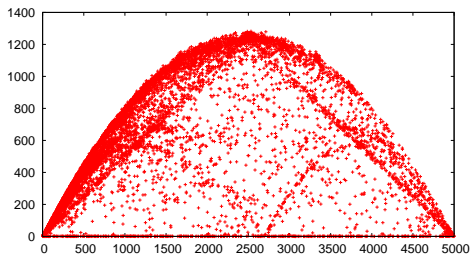
Premières observations



Comparaison

- parabole : identique
- concentration de points près de l'axe des x : **sur** l'axe des x

Autres caractéristiques



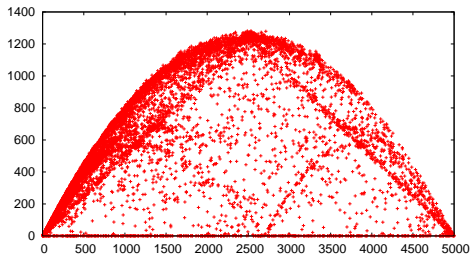
Triangle

Sommet présent pendant k passes avec $p = 1/2$:

- $x = k/2$
- $y = k * 1/2 * (1 - 1/2) = k/4 = x/2$

Présent mais moins visible dans les données réelles

Autres caractéristiques



Petites paraboles

Changement majeur : unique chemin pendant $\sim n/2$ passes puis *load-balancing* :

- parabole sur $[0, n/2]$
- parabole sur $[n - n/2, n]$

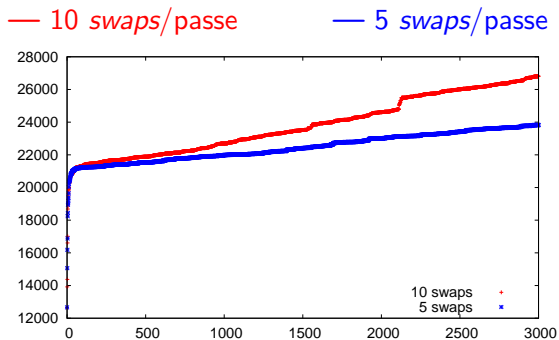
Présentes dans certains cas dans les données réelles

Résumé

Capture des comportements observés

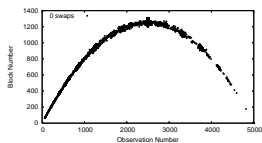
- Nombre de sommets découverts
 - Croissance initiale rapide
 - Longue croissance soutenue
- Observations vs apparitions

Influence des paramètres – nombre de *swaps*

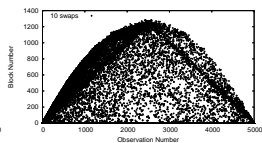


Plus de *swaps* → pente plus élevée

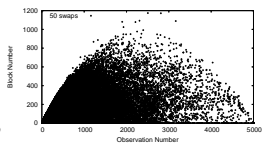
Observations vs. apparitions



$s = 0$

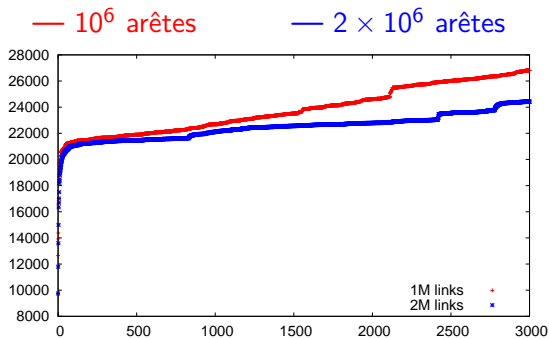


$s = 10$



$s = 50$

Influence des paramètres – densité

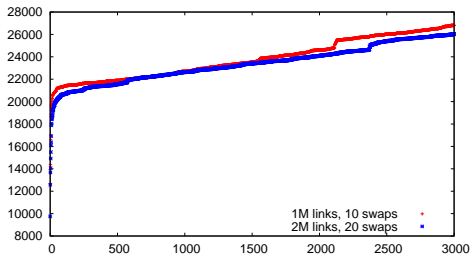


Plus d'arêtes → pente moins élevée

Invariants

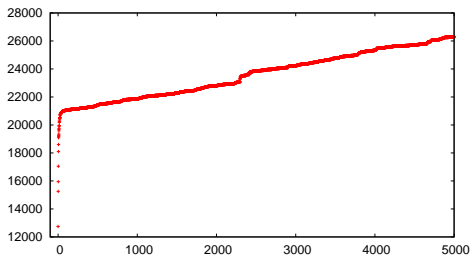
— 10^6 arêtes, 10 *swaps*

— 2×10^6 arêtes, 20 *swaps*



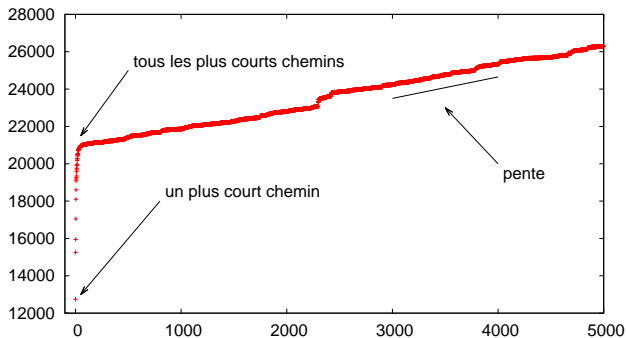
Compromis entre nombre d'arêtes et nombre de *swaps*

Besoin de quantifier l'influence des paramètres



Besoin de **moyenner**

Prédictions théoriques ?



Conclusion

Conclusion

- Analyse de la dynamique des vues égo-centrées de l'internet
- Isolé les facteurs de la dynamique
- Modèle simple reproduisant les comportements observés
→ **caractère explicatif**

Suites

- Quantifier l'influence des paramètres \rightarrow moyenne
- Études formelles
- Retour sur le cas réel

- Reproduire d'autres propriétés
- Influence de la distribution des degrés