

Génération de Topologies Réalistes pour la Simulation du Routage Interdomaine

J.M. Fourneau et H. Yahiaoui

Laboratoire PRiSM, UVSQ, 45, av. des États-Unis, 78035, Versailles - FRANCE

La simulation du routage interdomaine nécessite l'utilisation de topologies se rapprochant le plus possible de la taille et de la conformation du réseau réel. L'utilisation de topologies à granularité plus fine que le graphe de domaines de routage, permet d'observer d'avantage d'instabilités. Nous avons élaboré un algorithme de génération de topologies de sessions BGP pour la simulation réaliste de BGP et l'avons implémenté dans un outil général de manipulation de topologies interdomaine. Cet algorithme infère les topologies de sessions BGP en utilisant des topologies, édifiées à partir de données brutes de l'Internet, et d'informations de connectivité des domaines de routages, au niveau routeur.

Mots Clé: Routage Interdomaine, Simulation de BGP, Topologie Internet

1 Introduction

Durant plus de dix ans, BGP (Border Gateway Protocol) a été le protocole *officiel* de routage interdomaine. Cependant, des études ont montré le caractère instable du routage par BGP. Pour tenter de comprendre et de corriger ces instabilités, plusieurs simulateurs ont déjà été conçus [5][6][7]. Nous avons imaginé un simulateur à large échelle du protocole, qui a pour ambition de reproduire les instabilités de façon réaliste. Une composante importante d'un tel simulateur est la topologie utilisée. En effet, reproduire les instabilités observées sur Internet exige l'utilisation de topologies les plus réalistes possible. La majorité des simulateurs existant fonctionnent avec la représentation classique de la topologie interdomaine : un graphe d'ASs (Autonomous Systems) interconnectés par des liens BGP. Chaque AS est abstrait et simplifié pour constituer l'entité communicante de base. Des études récentes [4] ont montré qu'une partie importante de l'instabilité constatée de BGP était dûe, en partie, à la complexité interne des ASs et à la multiplicité des liens existant entre certains couples d'ASs. Il apparait, donc, que la simulation réaliste de BGP à grande échelle nécessite des topologies concernant cette complexité.

C'est pour répondre à ce besoin qu'ont été imaginées les topologies de sessions BGP. Basés sur une représentation plus fine du réseau interdomaine, ces topologies permettront de générer les comportements d'instabilité dûs à certaines politiques de routage alliées à la complexité interne des ASs. Cependant, la pauvreté des informations à propos de la structure interne des ASs, qui transpire à l'extérieur ne permet pas de construire une topologie fiable de l'Internet. Il a, donc, fallu mettre au point une méthode de génération, permettant, à partir de la topologie des ASs et d'informations complémentaires, de créer des topologies de sessions BGP qui, si elle ne correspondent pas exactement à la réalité, possèdent au moins un caractère de réalisme suffisant pour produire les instabilités de BGP qui nous intéressent.

Le présent article se poursuit en décrivant dans la section 2 les topologies d'AS, les moyens de les obtenir, ainsi que leurs limitations pour la simulation à large échelle. Puis, la section 3 introduit la notion de topologie de sessions BGP. Cette partie expose un algorithme de génération de topologies de sessions BGP à partir d'information de connectivité aux niveau AS et de connaissances complémentaires. Enfin, la section 4 expose l'implémentation de cet algorithme et son paramétrage.

2 Topologies de Systèmes Autonomes

BGP permet de hiérarchiser le routage sur Internet en considérant la connectivité d'entités complexes : les systèmes autonomes. Chaque système autonome est un domaine de routage indépendant, administré indépendamment des autres AS, en utilisant des protocoles et des métriques de routage, uniquement définis à l'intérieur du domaine. Il apparaît alors, que du point de vue de BGP, l'Internet en entiers se comporte comme un graphe, dont les nœuds sont les ASs et dont les arcs sont schématisés par les sessions BGP établies entre ASs. C'est, justement, cette vision de l'Internet qui est majoritairement utilisé pour la simulation du protocole interdomaine. En effet, l'adoption des topologies de simulation basées sur le graphe de connectivité des ASs de l'Internet permet de répondre aux impératifs d'abstraction exigés dans la simulation. De plus, l'extraction d'informations de topologies à partir de données disponibles sur Internet permet de produire assez facilement des topologies d'AS. Alors qu'il est pratiquement impossible d'obtenir une vision plus détaillée de l'intérieur des AS (une vision au niveau *speaker* BGP). L'obtention de telles topologies utilise, dans la majorité des simulateurs ou des projets de cartographie de l'Internet, la même méthode. Celle-ci se base sur la présence, sur Internet, de certains speakers spéciaux. Ces speakers, souvent dévolues à l'observation du routage interdomaine, procèdent régulièrement à une sauvegarde de leurs tables de routage. Ces *logs* peuvent, alors être utilisés pour extrapoler une topologie de l'Internet actuel. Par exemple, d'un fragment de *log* comme celui-ci (extrait des logs BGP du projet RouteViews [2]) :

```

Network Next Hop      Metric LocPrf Weight Path
*> 4.0.0.0 134.24.127.3          0 1740 1 i
*          194.68.130.254 0 5459 5413 1 i

```

On pourra déduire la présence d'au moins un lien BGP (une sessions BGP établie entre un speaker de chaque AS) entre chacun des couples (5459, 5413), (5413, 1), . . . L'analyse et la synthèse de telles informations, récupérées à partir de différents point de l'Internet à différents instants, produira une carte presque complète des liens entre ASs.

Si ce type de topologie a été largement exploité, il nous est apparu comme incomplet pour le type de simulations que l'on veut réaliser. En effet, l'étude du volume et des types de messages BGP échangés sur Internet a montré que certaines instabilités du routage interdomaine étaient dues à des causes internes aux ASs [4]. Par exemple, certaines politiques de routage, dites de *mapping MED-IGP*, attachent aux annonces de routes des attributs MED (Multi_Exit_Descriptor) dont les valeurs correspondent aux métriques IGP interne aux ASs. Ces politiques peuvent, donc, produire des instabilités sur le routage interdomaine à partir de variations internes aux ASs. En simulant des topologies aux ASs simplifiés, il est impossible de reproduire de tels comportements. Ce qui diminue le réalisme des instabilités que l'on voudrait obtenir. De plus, les topologies d'AS simulées possèdent une limitation due à leur structure interne : un AS de ces topologies ne peut être relié que par un seul lien vers un autre AS. Cette disposition ne reflète pas du tout la réalité où les ASs, surtout de grandes tailles, nécessitent souvent l'établissement de plusieurs sessions BGP entre eux pour ne pas provoquer de surcharge de trafic sur certains liens. En conséquence de quoi, il peut arriver que différents speakers d'un même AS choisissent des routes différentes pour les mêmes destinations. Pour produire des topologies d'ASs à partir de tels réseaux, la solution classique consiste à séparer les ASs arborant de tels comportements en **ASs logiques**. Cette méthode résout une partie uniquement des problèmes de simulation liés à la complexité intra-AS. Les difficultés évoquées plus haut n'y trouve aucune solution.

Pour pallier à ces manques, nous avons décidé d'utiliser des **Topologie de Sessions BGP**. Dans ces topologies, le nœud élémentaire n'est plus l'AS, mais le **speaker** ; un AS pouvant être composé de plusieurs speakers interconnectés. Bien que moins abstraites que les topologies d'ASs, donc plus difficiles à simuler pour de grandes tailles de topologies, ces topologies nous permettront de reproduire au plus près la hiérarchie de connectivité du routage interdomaine. Ainsi, la présence de sessions multiples entre ASs devient évidente à produire. Et l'impact des différents timers et délais est mis en valeur par la présence de plusieurs speakers, communiquant à des rythmes différents

dans un même AS. Il devient aussi possible d'observer les conséquences de certaines politiques sur le routage interdomaine, ainsi, que les répercussions de la transmission d'instabilités **internes aux ASs** vers l'extérieur de ces ASs.

3 Topologies de Sessions BGP inter-Speakers

Les topologies de simulation que nous avons proposés sont, donc, des **topologies de sessions BGP**. Moins abstraites que les topologies d'ASs, ces topologies prennent en compte la complexité des phénomènes au sein des ASs et permettent de prendre en compte les cas de multiplicité des connexions entre ASs, qui est une configuration plus que fréquent sur l'Internet d'aujourd'hui.

Les nœuds de ces topologies sont les speakers BGP des ASs. L'adoption d'une granularité plus fine que l'AS a pour but d'observer des phénomènes d'instabilité dont la source est, par exemple, l'intérieur des AS, mais dont les effets se voient sur la structure global du routage interdomaine. Les arcs reliant les nœuds sont les sessions BGP établies entre les speakers des différents ASs. Etant donné la multiplicité des speakers appartenant à un AS, il devient possible d'établir plusieurs sessions entre chaque paire d'AS. Il suffira pour cela de sélectionner des paires de speakers différents, tel que chaque speaker appartient à l'un des AS, et d'établir une sessions EBGP entre eux. Les liaisons entre les speakers d'un même AS deviennent indispensables. En effet, l'unicité des speakers au sein des ASs des topologies d'AS permettait d'ignorer la transmission des routes interdomaines entre speakers d'un même AS. En adoptant un topologie plus complexe, nous devons permettre aux speakers d'un même AS de se communiquer leurs routes respectives, afin de conserver la cohérence du routage. La connectivité que nous avons adopté pour cela est de deux type : les ASs relativemnt "petits" (petit nombre de speakers) disposeront d'un graphe en clique de sessions IBGP (BGP adapté pour la transmission intra-AS). Les ASs plus "grands" seront élaborés en disposants des réflecteurs de routes pour diminuer le nombre de sessions IBGP attaché à chaque speaker de l'AS.

Les topologies de Sessions BGP répondent aux exigences de la simulation à grande échelle du routage interdomaine. Cependant, ce type de topologies ne peut être directement produit comme le sont les topologies d'AS. En effet, la structure interne des ASs (nombre de speakers, leur disposition et les sessions qu'ils établissent) est rarement connue de l'extérieur de l'AS. Il est, donc, indispensable de disposer d'une méthode d'**inférence** de topologies de sessions BGP à partir des topologies d'ASs. L'algorithme d'inférence que nous proposons utilise une topologie d'ASs **pondérée** pour générer la topologie de sessions BGP. La pondération est une opération qui consiste à assigner des poids aux liens entre ASs à partir d'informations de niveau routeur. La méthode la plus simple consiste à pondérer le lien par le nombre de lien routeur-routeur reliant chaque paire d'ASs. Puis, l'algorithme opère par :

1. La réduction des poids des liens en les divisant par un facteur commun : Pour tenter d'estimer le nombre de sessions BGP existantes entre chaque paire d'AS, chaque poids de le topologie est divisé par un facteur *factor* :

$$factor = \frac{\sum_i PoidsLien_i}{NbrLiens \div (1 - PourcentageLiensManquants)}$$

où , *NbrLiens* est le nombre de liens dans le graph des ASs et *PourcentageLiensManquant* représente une estimation de la fraction de liens BGP qui n'apparaissent pas dans le graphe des ASs.

2. La création, dans chaque AS, d'un nombre de speakers égale au poids maximum des liens cet AS. À chaque lien de l'AS est associé un compteur initialisé avec le poids du lien.
3. La sélection dans l'AS d'un speaker n'ayant aucune session. Pour chaque compteur de lien non nul, on associe au speaker une session avant de décrementer ce compteur. Chaque session sera aussi associée à l'un des speakers de l'AS voisin ne possédant pas de sessions avec un speakers de l'AS courant. Cette opération est répétée tant que subsiste des compteurs non nuls.
4. Lorsque toutes les sessions sont construites, créer les sessions IBGP au sein des ASs.

4 Implémentation

L'algorithme d'inférence de topologies de sessions BGP décrit ci-dessus a été implémenté dans un outil de manipulation de topologies, associé au projet de développement du simulateur des instabilités de BGP. L'algorithme exige en entrée une topologie d'AS (figure 1-a), des informations de niveau routeur (nombre de liens routeur-routeur existant entre chaque couple d'ASs ayant établie au moins une sessions BGP entre eux) et une valeur numérique représentant la proportion de liens supplémentaires désirée. La sortie de cet algorithme est une topologie de Sessions BGP disposant d'un nombre de sessions correspondant à la proportion indiquée (figure 1-b pour une proportion=30%, figure 1-c pour une proportion=50%).

Malheureusement, nous ne disposons pas encore d'information fiable de niveau routeur pour la totalité d'Internet, ni de méthode pouvant facilement fournir ces détails (l'exploration par *traceroute* étant trop contraignante pour le réseau global). L'algorithme est, donc pour l'instant, approvisionné en information de niveau routeur grâce à un générateur aléatoire. Ce générateur attribue des poids aux arêtes de la topologie d'AS avant qu'elle ne soit introduite dans le générateur.

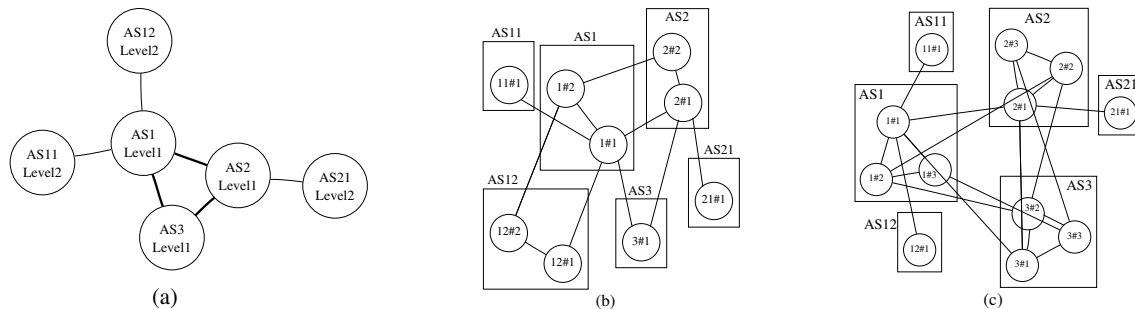


Fig. 1: Génération de topologies de Sessions BGP

Les topologies ainsi produites, peuvent différer des topologies réelles de sessions BGP (les topologies réelles complètes sont difficilement obtenus par cartographie). Cependant, elles possèdent le degré de réalisme nécessaire à la simulation des instabilités de BGP à large échelle.

Références

- [1] Y. Rekhter et T. Li : *A Border Gateway Protocol (BGP-4)*. RFC 1771, Mars 1995.
- [2] D. Meyer : *Route Views Project*. <http://antc.uoregon.edu/route-views/> University of Oregon.
- [3] W. Li : *Inter-Domain Routing : Problems and Solutions*. <http://cite-seer.ist.psu.edu/li03interdomain.html>, 2003.
- [4] C. Labovitz, R. Malan et F. Jahanian : *Origins of Internet Routing Instability*. Proceedings of IEEE INFOCOMM '99 Conference, March 1999.
- [5] B. Quoitin : *C-BGP User's Guide*. CSE Departement, UCL, Belgique, Avril 2004.
- [6] Hao et Koppol : *An Internet Scale Simulation Setup for BGP*. ACM SIGCOMM, Volume 33, Numéro 3, Juillet 2003.
- [7] B.J. Premore : *An Analysis of Convergence Properties of BGP using Discrete Events Simulation*. Dartmouth College, Hznover, New Hamshire, Mai 2003.
- [8] S. Agarwal : *Characterizing the Internet Hierarchy from Multiple Vantage Points*. <http://www.cs.berkeley.edu/sagarwal/research/BGP-hierarchy/>.
- [9] B. Cheswick et H. Burch : *The Internet Mapping Project*. <http://research.lumeta.com/ches/map/>.
- [10] H. Chang and S. Jamin and W. Willinger : *Inferring AS-level Internet topology from router-level path traces*. Proceedings of SPIE, Vol 4526, Juillet 2001.