BGP-Level Topology of the Internet: Inference, Connectivity and Resiliency

Rémi Varloot

École Normale Supérieure de Paris

Supervisors

Mathieu Feuillet and Guillaume Valadon
ANSSI

Agence nationale de la sécurité des systèmes d'information

Observatoire de la résilience de l'internet français

Résumé

Ce rapport s'intéresse à l'interconnectivité entre les divers réseaux de l'Internet français. Les réseaux désignent ici les Systèmes Autonomes (AS), tels qu'ils sont identifiés pas le protocole BGP.

Dans un premier temps, il s'agit de définir un modèle théorique pour représenter les relations entre ces différents AS. Ce modèle doit tenir compte de la nature des accords commerciaux entre les organisations derrières chaque réseau, et de leurs impacts sur le routage BGP et la connectivité vis-à-vis de l'Internet.

On implémente alors un algorithme permettant de reconstituer un carte de l'Internet selon ce modèle. Celui-ci utilise des données de routage publiquement disponibles afin d'inférer la nature des accords entre les différents organismes. Le résultat se présente sous la forme d'un graphe, dans lequel sont présents l'ensemble des AS visibles, ainsi que la nature des connections entre ceux-ci.

Dans un second temps, afin d'étudier l'Internet français, l'on isole tout d'abord la partie du réseau jouant un rôle clef dans l'interconnexion entre les AS français. On y identifie des AS critiques, dont la disparition isolerait des AS français du reste d'Internet, ainsi que les AS français courant un tel risque de déconnexion. On y montre notamment que deux fournisseurs suffisent à assurer un accès permanent à Internet, même en présence d'une faille dans le réseau.

Pour finir, l'on propose plusieurs améliorations possibles du modèle, tout en détaillant les difficultés que cela engendrerait. On illustre aussi comment la carte de routage BGP pourrait permettre, dans le domaine de la supervision des réseaux, d'identifier des usurpations d'AS, jusqu'à maintenant difficile à détecter.

Abstract

In this paper, we study the resiliency of the French Internet from the point of view of network interconnectivity. We define a model for representing the BGP-level topology of the Internet which takes into account the business relationships between ASes, and implement an algorithm to construct such a map from publicly available routing information.

The portion of the Internet responsible for the connectivity of French ASes is then established, and the risk of disconnection is assessed. To this end, we identify a set of critical ASes whose suppression would disconnect other ASes from the Internet.

Finally, we give some insight as to how the model could be expanded and the difficulties this would introduce. We also give an example of how a BGP-level map of the Internet can be used to actively monitor the network.

1 Introduction

In the 21st, the Internet has become the most import means of exchanging data. One of its decisive features over the last decades has been its ability to scale as the number of users has increased. This was made possible by the distributed nature of the Internet: many organizations have worked on its deployment at the same time, each one extending its own network, all of them interconnecting to form the Internet as we know it.

There are two key aspects to the Internet that are often differentiated: data transfer and routing. Data transfer covers the issue of forwarding traffic between end-users. It includes considerations such as transfer rates, delays, and congestion. Routing, on the other hand, deals with informing routers of where packets should be sent next in order to reach a given machine. For machines communicating over the Internet, the most important of these is IP routing, through which the network learns how to reach individual IP addresses. These come in two categories: intra-network routing protocols, such as OSPF, and inter-network routing protocols, such as BGP.

The French Internet Resilience Observatory [2] was created in 2011 to study the state of the Internet today, focusing on what can be considered as the French portion of the Internet: those networks considered to be operated in France or by French organizations. One of its key missions is to assess the resiliency of the French Internet, that is to say its aptitude to overcome and recover from deficiencies in the overall network, retaining a functional state at all times. Multiple factors are taken into account, such as the deployment of protocols like IPv6 and DNSSEC, the declaration of `route` objects and the usage of RPKI. This has led to the creation of a number of resiliency indicators, which the Observatory improves and refines over time.

This work concentrates on one of these indicators: interconnectivity. We consider the Internet at a network level, and aim to evaluate the level of interconnection between the different networks that make up the French Internet, all of which are operated by independent organizations. Connection focuses here on routing policies — and more precisely on the existence of known paths between French networks — rather than, for instance, on the available bandwidth. This indicator is important for understanding the risk of some networks becoming isolated from rest of the Internet due to another network failing.

The first part of this work, detailed in section 2, consists in a proper understanding of the mechanisms behind inter-network IP routing, and on the underlying protocol: the Border Gateway Protocol (BGP). In section 3, we look at how to construct a map of the BGP-level topology of the Internet that is compliant with the financial agreements between organizations. This is largely inspired by the works of the Cooperative Association for Internet Data Analysis (CAIDA) [9]. We then explain in section 4 how this map can be used to analyse the resiliency of the French Internet, and suggest some other possible applications.

2 A Brief Overview of BGP

Since the Internet is composed of thousands of interconnected networks operated by independent organizations, IP routing has to be achieved in a decentralized manner. More specifically, whereas each organization engineers its interior IP routing as it sees fit, inter-network routing needs to rely on a common distributed protocol. Though multiple routing protocols have existed, such as the Exterior Gateway Protocol [10], introduced in 1982, the only one to be widely used today is BGP [12].

2.1 Autonomous Systems and Route Propagation

BGP allows interconnected networks, called *Autonomous Systems* (AS) to exchange knowledge of paths to various IP addresses, be they their own or those to which they know a path. Each AS is given a unique *AS number*, which we sometimes identify with the AS. IP addresses are grouped in IP subnetworks called *prefixes*. A prefix is a set of IP addresses beginning with the same bits. For

example, the 1.1.0.0/23 prefix contains the IPv4 addresses whose 23 first bits are equal to those of 1.1.0.0, i.e. the 1.1.0.0–1.1.1.255 range.

BGP is a point-to-point protocol: BGP sessions are established between two connected BGP routers, who can then exchange routing information. We differentiate two cases: internal BGP, or iBGP, where the two routers belong to the same AS, and external BGP, or eBGP, where they belong to different ASes. Unless stated otherwise, the use of BGP links refers here to eBGP sessions.

A BGP session between two routers allows them to exchange *routes*. A route is a set of *path attributes* [12] used to describe the path to a given prefix. The most common path attributes include the following:

AS_PATH An ordered list of AS numbers corresponding to the ASes crossed to reach the prefix. The last AS corresponds to the AS from which the prefix originates.

NEXT_HOP This IP address corresponds to the next destination for traffic headed for the destination prefix. It is often, but not always, the IP address of the last router from which the route is learned.

LOCAL_PREF This value is used in iBGP to define preferences between multiple routes. This is used in network engineering, for example to prioritize cheaper routes.

COMMUNITIES This optional attribute, defined in [1], contains a list of values, whose meanings are defined freely by network administrators. These are generally used to filter routes or set the local preferences, and can include information such as where the route was learned, from who, etc.

BGP routers can propagate the routes they learn by announcing them to other neighbors. For eBGP sessions, the AS number associated to the router is prepended at least once to the AS path. An example of route propagation is given in figure 1. This process eventually leads to every BGP router learning the path to every announced prefix.

An exception to BGP prepending appears at certain Internet eXchange Points, or IXPs. These are locations meant to facilitate the creation of peering links between ASes. Most of these use special BGP routers that serve as eBGP route reflectors: they establish BGP sessions with the peering ASes, receive routes, and redistribute these routes to all the ASes, without prepending their own AS number. This instantly creates a full mesh of peering ASes.

When a router learns two routes for a given prefix, as in the case of AS444 in figure 1, it selects the path it will use for routing packets, and only propagates the corresponding route to its neighbors; this avoids a prohibitive increase in routing information as more ASes are added to the Internet. The choice of the route is based on multiple factors, considered in a specific order. To begin, the local preferences are compared, and only routes with the highest value are retained. This is followed by comparing the AS path lengths: the route with the shortest AS path is preferred. As a consequence, if an AS has a preference between two links — for example if one of them is cheaper or has better bandwidth —, a common practice is to prepend the AS number multiple times on the AS path announced along the least favorable link, so as to reduce the chances of that route being selected. For example, AS111 appends its AS number twice when it announces its prefix to AS333, resulting in AS444 selecting the other route when it learns the two routes. Other criteria exist for routes with the same local preferences and AS path lengths, as explained in [12].

In order to avoid routing loops, two constraints must be respected: BGP routers must reject all paths in which their AS number already appears, and BGP routers establishing iBGP sessions within a same AS must always form a full mesh. The first condition ensures there are no loops over multiple ASes, and the second one that all the routers of a given AS learn all the routes without there being a risk of a loop appearing within the AS. Note that the full mesh constraint can be avoided by using iBGP route reflectors; a central router over which a virtual full mesh is constructed.

The design of BGP is such that the protocol scales very well. Since the creation of BGP-4 in 1994 [11], when the Internet was composed of less than a thousand ASes, only a few modifications

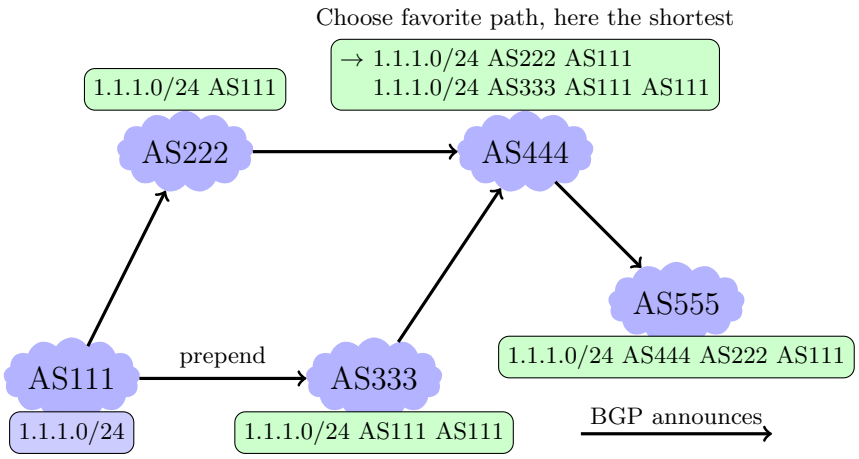


Figure 1: Path propagation with BGP

were required in order to allow its use to spread to the current 46000 ASes [3]. One of the most notable ones was the passage from 2 to 4-octet AS numbers in 2007 [14]. On the downside, BGP was not conceived with security in mind, and there are namely no means of guaranteeing that a route learned from a neighboring AS has not been tampered with, or that a prefix announced by an AS is legitimate. Prefix and AS spoofing have become a common observation [8], and their detection is important in network monitoring (such as that done by the Observatory [2]).

2.2 Routing Policies

One of the most important points regarding BGP is that ASes have no obligations to transmit known routes: though the protocol enables them to send routing information to a neighbor, this is not compulsory. In particular, ASes are free to exchange only some of the routes they know, as illustrated in figure 2.

It is common practice, for example, for an AS to only inform its neighbors of its own prefixes. In this manner, the AS is sure never to receive traffic to transfer to other ASes. This is the case of most smaller ASes, such as AS333 in figure 2. That AS's neighbors are generally either Internet Service Providers (ISPs) or transit providers, i.e. companies whose trade is the interconnection of ASes. These companies generally announce all of their known routes, thus effectively offering connectivity to the rest of the Internet, in exchange for payment. In this situation, the first AS is called a *customer*, and the second a *provider*. When a customer has customers of its own, as is the case for AS789, we call *customer cone* the set of ASes composed of itself, its customers, their customers, etc. These ASes generally announce routes to all the prefixes of their customer cone to their providers.

Another common situation is *peering* agreements: two ASes who exchange much traffic agree to share routes to their respective customer cones so as not to exchange their transit through a common provider. Originally, neither of the two peers would pay the other, as traffic was balanced. This is no longer true, however: peering agreements between ISPs and content providers generally yield uneven traffic, and financial compensations can be negotiated. For example, in figure 2, AS123 and AS456 have a peering agreement, and can exchange traffic without passing through AS1000.

We give some details on the three routes announced by AS456 to AS222 in figure 2:

1.1.1.0/24 AS456 has chosen the peering link over the route through its provider, as it is probably cheaper. As a consequence, AS222 only learns that route.

3.3.3.0/24 AS456 directly serves as a transit provider between AS222 and AS333, and is likely to be payed by both ASes.

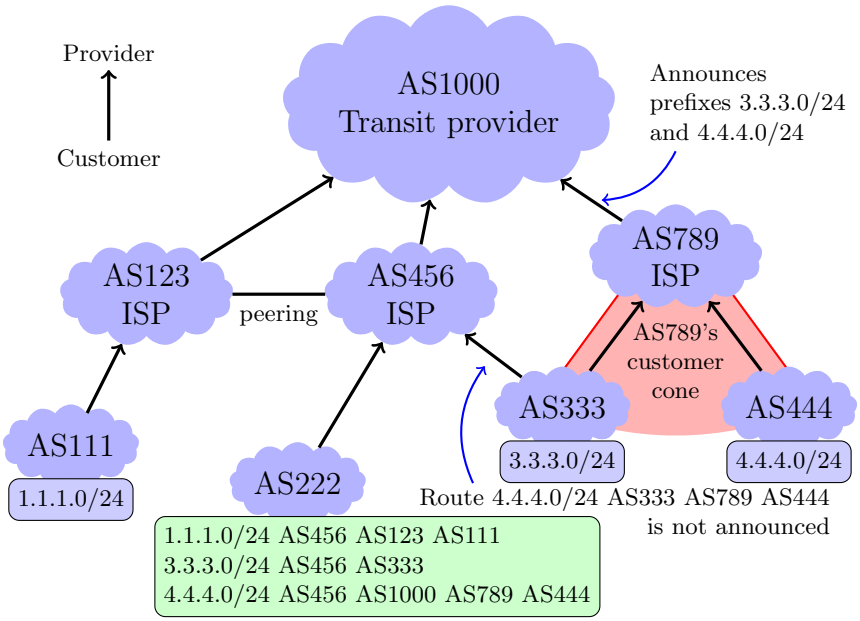


Figure 2: Routing policies and their consequences on route propagation

4.4.4.0/24 AS333 does not inform AS456 of the route to AS444. AS456 knows only the route through the transit provider, and relays it on to AS222.

As a general rule, what routes are to be exchanged is decided as part of the contract between two ASes when an interconnection is created. There is no mechanism in BGP to enforce this, however, and routing policies have no technical existence in BGP exchanges.

The aim of this work is first to construct a map of the Internet similar to that of figure 2, i.e. with explicit routing policies, and then to see what can be learned from it. We will hereon refer to this map as the AS or BGP-level network topology of the Internet.

3 A Tour of Scientific Literature

Since Lixin Gao's cornerstone paper on valley-free paths in [7] in 2001, much work has gone into finding efficient inference algorithms for the AS-level topology of the Internet. We give here a brief description of the most common model and methods, as well as a detailed description of an algorithm devised by the Cooperative Association for Internet Data Analysis (CAIDA) [9] and used in our work.

3.1 Academic Representation of Commercial Agreements

Due to the apparently unconstrained nature of BGP exchanges, it is difficult to define an adequate model for an AS-level network topology. As we have seen in section 2.2, the simple, naive model in which we suppose ASes exchange routes for all known prefixes is not satisfying, as it does not reflect the impact of routing policies. The network topology would only consist in an unoriented graph of interconnections, which is not enough to determine whether two given ASes can communicate or not.

Consider the example in figure 2. If the link between AS1000 and AS789 goes down, then AS444 is no longer reachable from AS222, despite the existence of a physical path between the two.

A more reasonable approach is to categorize the different types of agreements according to the different policies described in section 2.2: customer-to-provider and peering relationships.

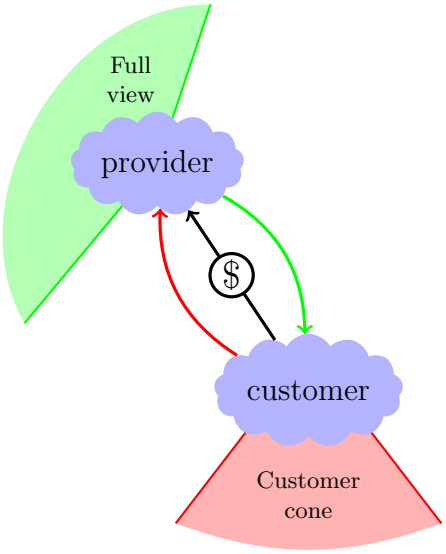


Figure 3: Customer-to-provider relationship

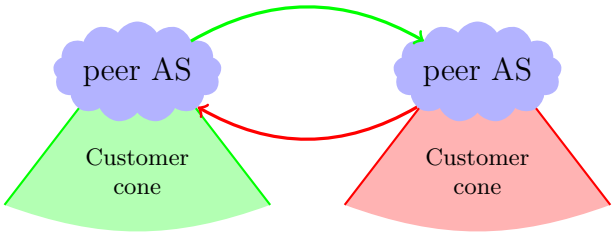


Figure 4: Peering relationship

It is generally accepted throughout literature [7, 9] that these suffice for a vast majority of the agreements found on the Internet.

The main interpretation of this model comes from two observations. The first concerns traffic forwarding, i.e. an AS allowing two neighbors to communicate by passing through its network. We consider that this is only done in exchange for payment. The second observation is that, reciprocally, if an AS pays for a connection, it expects to have access to every route known to the connected AS. As such, when two ASes interconnect, if the first one receives payment from the second, it announces routes to all its known prefixes (called a full view) over the link, otherwise only routes to its own prefixes and those of its customer cone.

This then gives the two types of relationships: if one AS pays the other, we have a customer-to-provider relationship (figure 3), otherwise we have a peering relationship (figure 4). Sometimes, we speak of a provider-to-customer relationship rather than a customer-to-provider relationship, either because it is viewed from the provider's perspective, or because it is in an AS path in which the provider is placed before the customer.

Notice that an AS can allow traffic to transit between a customer and a provider. In this case, it receives payment from its customer and in turn pays its provider.

Other models of policies have been introduced in the literature, but are generally more debated or complex. The most common of these is sibling relationships [7], in which two ASes effectively accept to provide transit for each other's traffic. In terms of routing, they exchange their full views. An example of where such an agreement can be found is the case where two companies merge, but two separate ASes are maintained. This type of relationship is often left aside, however, because such cases are both rare and very difficult to detect [5].

Another policy that does not fit in this model is prefix-specific peering or customer-to-provider agreements, in which only a subset of the available prefixes are announced.

It must also be noted that all of these models suppose that ASes can be considered as atomic structures. This is not the case, however, as almost all ASes have multiple BGP routers. Two ASes can, for example, be connected at multiple datacenters. In some cases, the policies can even differ depending on the location, such as a customer and provider peering at an IXP, i.e. a peering only facility. For such cases, the model is insufficient, as it supposes that the routing policy between two ASes is unique.

Regardless, unlike with the simpler model, the resulting network topology provides enough information to determine whether two given ASes can communicate or not, as illustrated in figure 5.

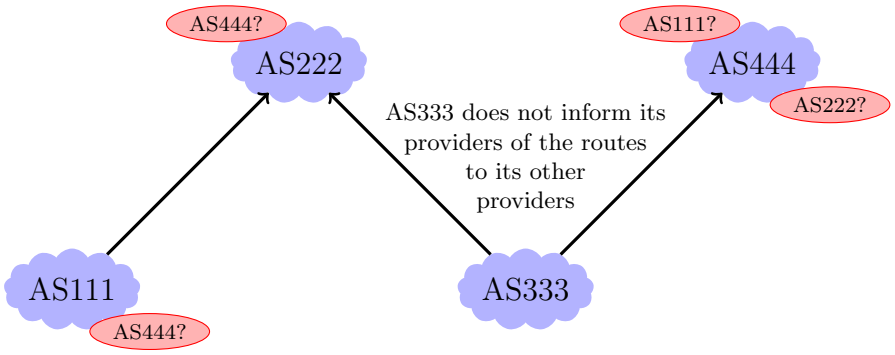


Figure 5: Impacts of commercial agreements on connectivity

3.2 Inference Algorithms

Many algorithms have been suggested for the discovery of the BGP-level topology of the Internet. These can be categorized in multiple ways. The first major criterion is the type of data used: measurements, or data provided by the AS themselves?

As far as measurements are concerned, a natural classification exists between active measures and passive measures.

Active measures Designates any type of data collection that requires actually impacting the network, generally by generating traffic. The most common data source in this case is probing, namely with tools such as `traceroute`.

Passive measures Measurements that do not affect the dynamics of the network. The most common example is the use of BGP collectors: these routers are connected in datacenters, where they generate no traffic but receive BGP announcements that are then made publicly available.

Other sources of data exist as well, such as the information given by AS maintainers. These include the `aut-num` declarations in the `whois` databases for which import and export attributes have been specified.

The context of this work requiring we not use active measures, and the reliability of declarative information being difficult to assess, this work is based solely on passive measurement. More precisely, we consider only algorithms whose input can be derived from datasets gathered from BGP collectors.

In practice, the only path attribute of use in routing information is the AS path. Other information is either of little use (next hop), unavailable (local preference), or difficult to handle because its meaning changes from one AS to the next (communities, though some studies have been done [6]).

We now give the main result used throughout literature when attempting to map the BGP-level topology of the Internet:

Definition (Valley-free)**.** *A path between two ASes in routing graph is called* valley-free *if and only if it passes through any number of customer-to-provider links, followed by at most one peering link, and then any number of provider-to-customer links (cf. figure 6).*

The following properties hold:

Property (Valley-free [7])**.** *Two ASes can communicate over a given path, i.e. the corresponding routing information has been propagated both ways, if and only if the path is valley-free.*

Property (Transit)**.** *Every AS in the middle (neither first nor last) of an AS path is the provider of either the preceding AS or the following one, possibly both.*

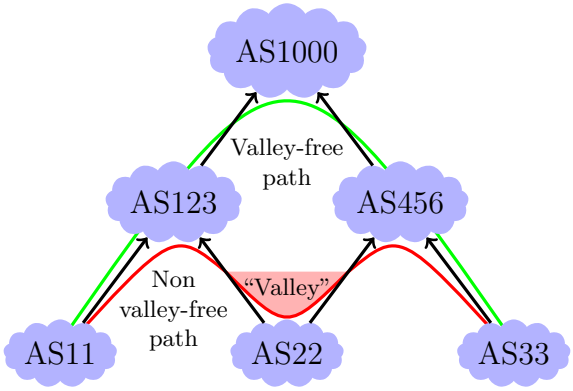
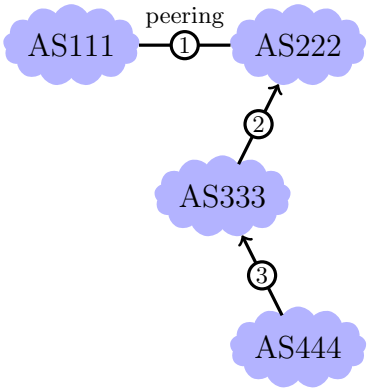


Figure 6: The valley-free property



AS Path: AS444 AS333 AS222 AS111

① Suppose AS111 and AS222 have a peering relationship.

② The transit property applied to AS222 requires that it be a provider of AS333.

③ Applying the transit property again to AS333 gives that it must be a provider of AS444.

Figure 7: Example of top down inference

In theory, every AS path observed at collectors should be valley-free. In practice, this is not always the case [9], either because some relationships do not fall into our model, or — and this is not uncommon — because of routing errors or path spoofing [8]. These paths are called *poisoned*, and the difficulty for inference algorithms is to find the correct types of relationships between ASes without making errors due to poisoned paths.

Two types of passive measurement inference methods exist: *type of relationship optimization problems*, or *ToR problems*, and *top-down inference algorithms*.

ToR problems were first introduced in [13]. They consist in finding the set of routing policies which minimizes the number of non-valley-free paths. The problem is NP-hard, but multiple approximation algorithms exist [4].

Top down inference relies strongly on the transit property, and more specifically on the fact that, if three ASes a, b and c appear consecutively in a valley-free path, and a is either a provider or peer of b, then c is necessarily a customer of b. This implies that, for a given valley-free AS path, if one interconnection is known to be a provider-to-customer link, then so are all the ones after it (see figure 7). The main caveat of this technique is the possibility of erroneous inferences due to poisoned paths, which can easily be propagated. It also requires knowledge of at least one link's type of relationship in order to initialize the inference.

In the following section, we present the top-down inference algorithm conceived by CAIDA and used in our work.

3.3 The CAIDA Algorithm

The use of the algorithm described by CAIDA in [9] was motivated by the work they have put into validating their algorithm: not only does this confirm it yields some of the best known results, but it also allows us to quantify the reliability of the output.

The algorithm itself is a top-down inference algorithm, requiring no prior knowledge, that attempts to limit the impact of poisoned paths through multiple heuristics. The key feature of their algorithm is the notion of *transit degree*. A simplified version of the algorithm is given in algorithm 1. For full details, see the paper by CAIDA [9].

Algorithm 1 Summary of the algorithm

1. Compute transit degrees.
2. Find the top-most ASes, known as clique ASes.
3. Apply a top-down inference with constraints on the transit degree.
4. Use partial viewpoints to infer more customer-to-provider agreements.
5. Identify valid relationships opposed to the transit degree constraint.
6. Find customers for ASes with no provider.
7. Infer stub-clique relationships as customer-to-provider.
8. Resolve ties for consecutive uninferred links.
9. Designate all remaining links as peering relationships.

Compute transit degrees We say an AS forwards traffic for one of its neighbors if it appears between that AS and another on an AS path. The transit degree of that AS is the number of neighbors for whom the AS forwards traffic. Intuitively, ASes with a larger transit degree are higher up in the AS "hierarchy".

We compute the transit degree of each AS by going through all AS paths and counting the number of different ASes for which they forward traffic.

Clique inference Since top-down inference requires some knowledge of the uppermost providers, we begin by finding these. In the Internet, these ASes correspond to Tier 1 networks: a limited number of large, provider-less ASes that all peer together. In this paper, we refer to them as clique ASes.

In order to find these clique ASes, we begin by taking the 10 ASes with highest transit degree, and compute the largest clique amongst them. We then take each AS in order of decreasing transit degrees, and see if it shares a link with every clique AS. If it does, we add it to the clique, otherwise, we ignore it and continue.

This gives us a set of big ASes which are guaranteed to all be interconnected. Supposing these correspond to Tier 1 ASes, we infer that they all have peering agreements and no providers.

Top-down This is the main step of the algorithm. It consists in applying the generic top-down inference method, but with the following restriction: a customer-to-provider link can only be inferred if the provider has a higher transit degree than the customer.

This heuristic serves as a mean of limiting poisoned path propagation, and is based on the intuition that providers generally have a bigger network then their customers. This is called the *degree constraint*.

For the rest of the algorithm, whenever a new relationship is inferred, we automatically test to see if more top-down inference is possible using this new information. The degree constraint is maintained in these circumstances.

Validation by CAIDA shows that, at the end of this stage, approximately 90 % of the customer-to-provider links have been correctly inferred.

Partial viewpoints Some ASes connected to collectors only transmit routes to prefixes from their customer cone. Others have a default route to their providers. A common feature in these two cases is that the number of origin ASes in the announced paths is small. Whereas it is not possible to further distinguish between the two, and therefore the first link can be of any nature, the second one is necessarily a provider-to-customer relationship.

Bottom-up inference Sometimes customers have a higher transit degree than their provider. In order to correctly infer these relationships, we now look at the links whose relationship could not previously be inferred due to the degree constraint.

If ever a path finishes with a link that could have been inferred as provider-to-customer, but did not satisfy the degree constraint, we now infer it as such. Requiring that the AS be at the end of the path limits the risk of errors, as we assume that faulty ASes do not announce their own prefixes.

ASes with no providers At this point, non-clique ASes with no providers have not had any of their links inferred. They appear on no top-down path, otherwise they would have been inferred as another AS's customer. Since top-down inference requires an AS to have at least a provider or peer, it cannot have been applied to provider-less ASes.

In order to overcome this, we infer what we believe is the most likely peering relationship for that AS: a peering link with its neighbour of highest transit degree for whom it forwards traffic.

Stub-clique links We consider all uninferred links between clique ASes and ASes of transit degree zero, i.e. ASes that only appear at the end of paths. Since the former are very large ASes, and the later very small, these links are inferred as corresponding to provider-to-customer relationships.

This case arises when all routes leading to the stub AS pass through only one clique AS, its neighbour. The link preceding the newly inferred one is a customer-to-provider link, and is therefore insufficient to infer the provider-to-customer relationship.

Resolve ties The final step, right after this one, consists in inferring all uninferred links as being peering relationships. Suppose we have three consecutive ASes a, b and c on an AS path p, such that neither of the relationships between a and b or between b and c have been inferred. With the next step, b will not respect the transit property, and p will not be valley-free.

This step aims at resolving some of these ties so as to limit the number of errors when inferring peering relationships. We consider the cases where a is always seen before b on AS paths, and c always after b. In this configuration, c is most likely seen through a provider, and is therefore inferred as being a customer of b.

For other configurations, we do not have enough information to break the tie, so nothing is done.

Infer peering links All remaining links are inferred as peering links. This step has more than a 98 % success rate.

We have implemented our own version of the CAIDA algorithm, in order to have a better understanding of it, but also so as to be able to adapt it to our needs. The algorithm takes as input a list of AS paths, derived from the information retrieved from collectors, and outputs a completely oriented graph of the Internet, along with the list of clique ASes. It is also possible

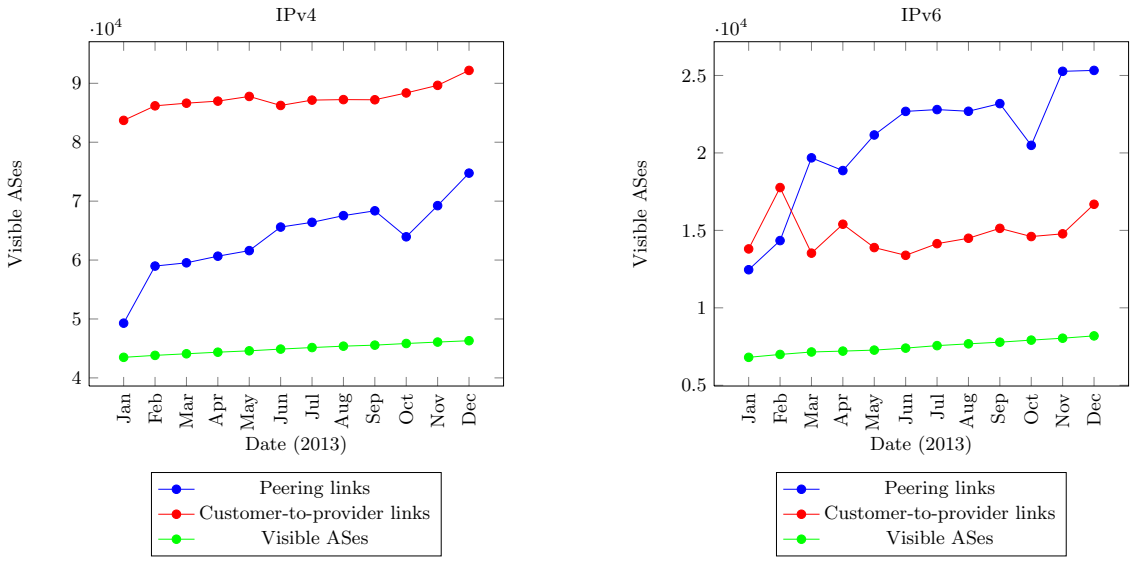


Figure 8: Inference results in 2013

to specify known relationships, fix the clique, or give a list of IXPs that should be removed from paths. On this last point, a list of 93 European IXPs were extracted from the Euro-IX website[1]. Though this list is incomplete, our main priority concerns the French Internet, for which we feel this data is sufficient.

We used this implementation to create BGP-level maps of the Internet throughout the year 2013, using data imported from all available collectors (RIS[2] and RouteViews[3]). For each month, we download all the available information over the five first days, combine them, and extract all AS paths for IPv4 and IPv6 routing. Some paths are immediately identified as abnormal [9], and rejected. The rest form the input to our algorithm.

In this way, we obtain BGP-level topologies of the Internet throughout the year 2013, for both IPv4 and IPv6. We give the number of visible ASes and links in figure 8. It should be noted that, though the clique inference yields satisfactory results for IPv4, the results obtained for IPv6 are often off, possibly because IPv6 is not yet sufficiently deployed. Furthermore, the inferred clique evolves over the year. Although it more or less corresponds to the usual list of Tier 1 ASes, there are exceptions, and this could serve as an alternative definition for what a Tier 1 AS is.

4 A Look at the Internet

We now explain in what ways the maps obtained were useful for analysing the Internet resiliency in France, and give some insight as to how they could also be used for active monitoring. Resiliency is here defined as the Internet's ability to remain functional in presence of faulty behavior.

4.1 Connectivity

To begin, we introduce our main indicator for resilience: *connectivity*. We say an AS is connected to the Internet if there exists a path of customer-to-provider links between that AS and a clique AS. These clique ASes are considered to represent Tier 1 ASes, and being able to reach these gives access to every other connected AS. If no such path exists, we say the AS is disconnected from the Internet.

[1] https://www.euro-ix.net/
[2] http://www.ripe.net/data-tools/stats/ris/
[3] http://archive.routeviews.org/

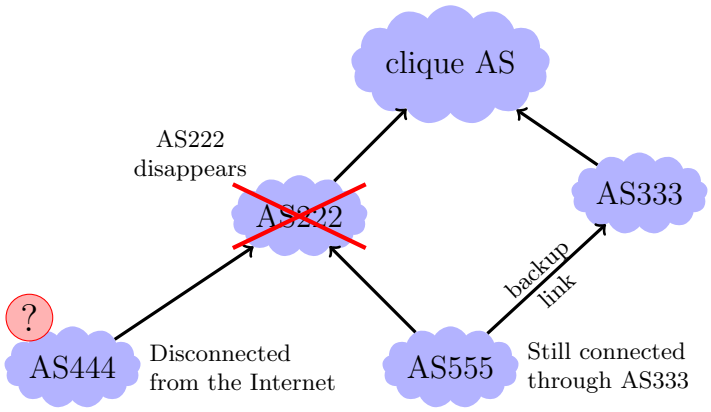


Figure 9: Impact on connectivity of an ASes disappearance

We consider the impact of an AS's "disappearance" from the Internet — such as might result from a configuration error — on the connectivity of other ASes. More specifically, an AS whose disappearance disconnects at least one other AS from the Internet is called a *critical AS*. In a similar fashion, any AS that can be disconnected from the Internet in this way is said to be *at risk*.

In figure 9, the disappearance of AS222 disconnects AS444 from the Internet. AS222 is therefore a critical AS, and AS444 is at risk.

We compute the set of ASes at risk and critical ASes by checking, for every AS c, that every AS in the customer cone of c connected to the Internet is not disconnected from the Internet when c is removed. If at least one such AS is disconnected, we flag c as being critical, and every disconnected AS as being at risk.

One of the main difficulties when looking at connectivity is backup links, i.e. secondary customer-to-provider links used when the primary ones fail, such as the one between AS555 and AS333. When not used, the routes passing through these links are not propagated over the Internet, and are therefore rarely visible from the collectors. As such, most backup links do not appear on the inferred network topology. An AS can therefore appear to be at risk when it is not, simply because we cannot see its backup link. Nonetheless, the ASes computed as being at risk form a worst case scenario: ASes that actually are at risk are always flagged as such.

4.2 The Case of France

This work concentrates on the resiliency of the French Internet, as defined by means of a list of French ASes provided by the Observatory [2]. It is therefore important to define a set of relevant ASes, i.e. a set of ASes which play an active role in the connectivity of French ASes. For example, an AS with no French AS in its customer cone has no impact on resiliency, regardless of whether it is a critical AS or not. Reciprocally, any AS with a French customer is most likely to play a part in connecting that customer to the Internet. This set of relevant ASes will be called the *convex envelope* of the French Internet.

Let N be the set of ASes present in the Internet, C be the set of clique ASes and S be a subset of N (such as the set of French ASes). We begin by giving three possible definitions of the convex envelope of S:

Definition (1). *An AS $a \in N$ is in the convex envelope $E_1(S)$ of S if there exists ASes $s \in S$ and $c \in C$, and a path p of customer-to-provider links from s to c, such that $a \in p$.*

Definition (2). *An AS $a \in N$ is in the convex envelope $E_2(S)$ of S if there exists ASes $s \in S$ and $t \in S$, and a valley-free path p from s to t, such that $a \in p$.*

Definition (3). *For every AS $a \in N$, call $P(a)$ the* provider cone *of a, i.e. the set of ASes that*

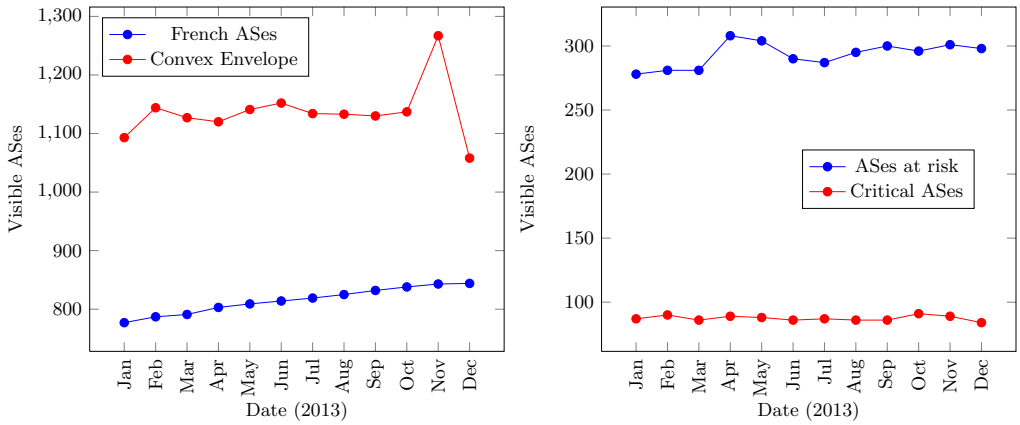


Figure 10: Evolution of the French Internet in 2013

have a in their customer cone. The convex envelope $E_3(S)$ of S is defined as

$$\bigcup_{s \in S} P(s).$$

Definitions 1 and 2 both have simple interpretations: for definition 1, the ASes are those responsible for connectivity to the Internet, and for definition 2, they are those responsible for reachability between members of the AS set. Definition 3 is somewhat more theoretical, but has the advantage of being easily computable. From a theoretical perspective, these definitions differ greatly. In practice, they are very similar. We explain why this is, and justify the use of the third definition to define the convex envelope of the French Internet.

To begin, notice that, if an AS a is in E_2, i.e. on a path between two ASes from S, then at least one of these two ASes is in its customer cone. This implies that a is in the provider cone of at least one AS in S. As a direct consequence, we have that $E_2 \subseteq E_3$.

We compute the number of French ASes in the customer cones of the clique ASes, and observe the following: each clique AS has at least one French AS in its customer cone, but never all French ASes. As a result, they all appear on a path between two French ASes: one from their customer cone, and one not from their customer cone. Every AS in E_1 is therefore also on a path between two ASes in S, i.e. $E_1 \subseteq E_2$.

We now look at the set of ASes in E_3 that are not in E_1. Notice that, if no such AS exists, then $E_3 \subseteq E_1$ and all definitions yield identical sets. This is not the case. However, all these ASes are naturally disconnected from the Internet. Indeed, suppose this is not the case, and that one such AS a is connected to the Internet. Its provider cone then contains at least one clique AS c, and a is in the provider cone of an AS $s \in S$. As a consequence, a is on a path of customer-to-provider links from s to c, i.e. $a \in E_1$. This contradiction concludes the proof.

Examples of such ASes include IXPs who add their AS number on AS paths or regional research. Whether to include these ASes or not in the convex envelope is debatable. Since these are never critical ASes — they grant no connectivity to begin with —, they do not play an important role in our study of the French Internet. We decide to include them in order to use the third, more practical definition.

The evolution of the French Internet in 2013 is given in figure 10. As we can see, as the number of French ASes increases, so does the number of ASes at risk. The number of critical ASes, however, does not appear to increase.

One of the most important results concerning the French Internet is that the only ASes at risk are those with a unique provider. That these ASes should be at risk is intuitive, since the disappearance of their provider would immediately disconnect them. What is less intuitive, yet positive, is that any AS with two or more providers is not at risk. In practice, this implies we never

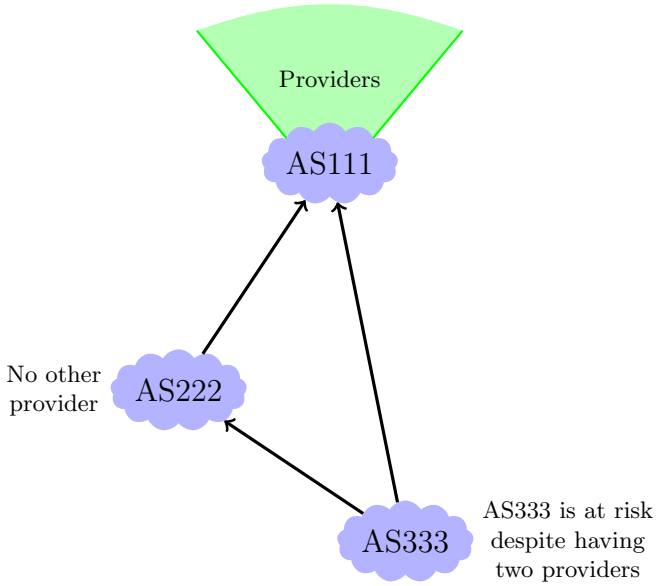


Figure 11: AS at risk despite multiple providers

have the case described in figure 11 in which AS333 has multiple providers, but is at risk because it becomes disconnected from the Internet if AS111 disappears.

This raises the question of the reliability of the information provided in figure 10, however. Some of the ASes considered as being at risk may have "invisible" backup links, as explained in section 4.1. The number of ASes actually at risk may therefore be quite less. This value can nonetheless serve as an upper bound: unseen links can only improve connectivity.

4.3 Perspectives

The study of the French Internet was our priority. Of course, our model is incomplete, and there is still much that can be done to improve it. We give here some possible adjustments we have briefly begun exploring. Furthermore, we believe there are other applications to BGP-level representations of the Internet and give an example at the end of this section.

As mentioned in section 3, one of the most common additions to this model found throughout literature is siblings [7]. The problem with sibling relationships is that they are hard to detect. Considering the CAIDA algorithm, the reason is apparent: if we attempt to identify them before the initial top down inference, we have very little information to work with, yet once the top down inference is over, it is most likely the link will have been inferred as customer-to-provider in one direction. In the same manner, the AS with lowest transit degree provides traffic for the other one, which is contrary to the degree constraint.

A common approach to this problem in literature is the use of the overall interpretation for sibling links: two ASes belonging to a same organization [5]. This technique has its caveats, however, as we have observed cases in which only one of the ASes provides transit for the other; this is then simply a customer-to-provider relationship in regards to our model, despite the absence of payment.

We give another approach, in which we use a technique similar to that of CAIDA for inferring customer-to-provider links that do not satisfy the degree constraint: we ask that both ASes announce paths *originating* from the others AS. In other words, for two ASes s and t, if we find two paths that end respectively with the links st and ts, and such that we can guarantee they were learned respectively from providers of s and t, then we infer s and t as being siblings. This technique currently has a low success rate, but has so far enabled us to find some examples of

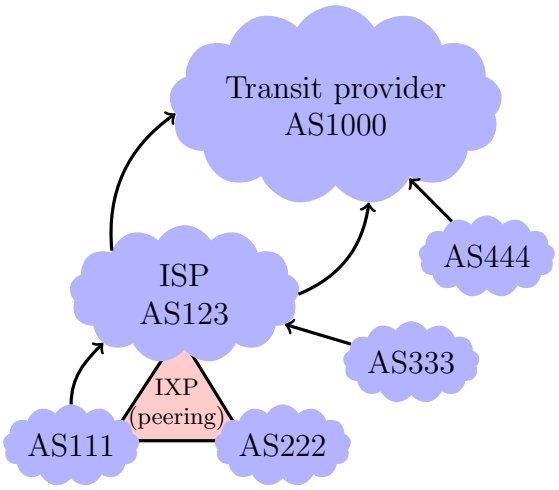


Figure 12: Points of presence and IXPs

sibling links between ASes from two distinct organizations.

Another aspect of BGP mapping of the Internet that we believe can be improved is AS fragmentation. So far, ASes have been considered as having a unique network. This is almost never the case. As mentioned in section 3, some ASes establish peering sessions at different routers, and the type of relationship can even vary between these links. Figure 12 illustrates what a router-level BGP map would look like (iBGP sessions are not represented, but always form a full mesh). Notice, for example, how AS111 would likely use its peering link with AS123 at the IXP to send packets to AS333, but must use the customer-to-provider link when contacting AS444.

Being able to represent ASes as multiple routers would greatly improve our understanding of connectivity risk. So far, we have considered the case of an entire AS disappearing. This can occur with configuration errors, where the BGP announcements from the deficient AS are faulty and therefore ignored. Other types of failures, however are not considered: a single link being disconnected due to a router failing, for example, or a power outage in a facility containing routers from numerous ASes.

In the first case, the number of ASes at risk would be greatly reduced. Consider figure 12. AS123 has a unique provider, but two links to that provider. It therefore cannot be disconnected from the Internet by a single router malfunction. This allows us to spot new behaviours as well: if the customer-to-provider link between AS111 and AS123 is removed, AS111 can no longer reach AS444 or AS123, but still has access to AS333 through that same AS.

In the second case, an AS can loose access to multiple providers, provided all the connections are in a same physical location. Rather than requiring that ASes have two different providers in order for them not to be at risk, the constraint would now be that they that two provider links — possibly to the same provider — in different facilities. In figure 12, a power outage at the IXP would result in both the disconnection of AS222 and the necessity for AS111 to use the customer-to-provider link to contact AS333. This requires not only fragmenting ASes into BGP routers, but also being able to group these by physical locations, called Points of Presence (PoP).

We began to try a simple heuristic in order to regroup ASes by PoPs. The intuition is that two ASes with a common provider are approximately of the same size, and therefore likely to peer. This requires them to be present in a same facility, however, so turning the argument around yields that, if two ASes have a common provider, they are more likely *not* to have any common facilities. We then attempted to use available clustering algorithms on graphs. The results were relatively inconclusive, though most French ASes were inferred as having common PoPs.

We finish by presenting a possible application for a BGP-level map of the Internet, that could not be tested at this time. We consider the problem of AS hijacking: a BGP routers establish a BGP session with a neighbor, but passes itself off as another AS, using the hijacked AS's AS

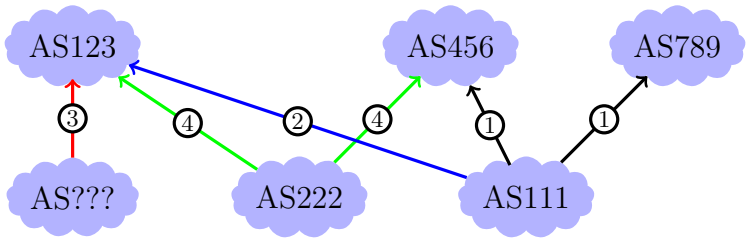


Figure 13: Detection of AS spoofing

number. We wish to detect possible AS hijacks. Consider the following, illustrated in figure 13:

Suppose AS111 has only two providers, AS456 and AS789 ①. Surely these connections will either be at the same or facility, or in two close ones. Suppose now that a new AS path is announced in which AS111 appears to have a new provider, AS123 ②. It is possible that a new transit agreement has been signed. If the connection takes place on another continent, however, this is most unlikely, and more closely resembles an AS hijack in which a distant AS has tried to pass off as AS111 ③. Of course, the lack of geographic data in our inferred topology means such an analysis is not immediately possible. Nonetheless, a simple heuristic could be used instead. For example, ASes with a strong geographic proximity are bound to have connections at common facilities. Inversely, two ASes separated by an ocean should not have any common provider, save for huge transit providers. Therefore, if we can find an AS222 that is a customer of AS123 and of any of AS111's providers ④, then the new route seems to be feasible. On the opposite, if no such AS exists, then this new route is likely to be an AS hijack.

This last consideration has not been tested on AS hijacking, but geographic heuristic has been used to find some ASes that were incorrectly considered as French by the observatory, giving us hope that this method should perform well.

5 Conclusion

In this paper, we give a rapid overview of BGP, and of the impact that routing policies between independent networks have on reachability. We then explain and implement the CAIDA algorithm in order to obtain a BGP-level map of the Internet. The result is used to improve our understanding of the interconnectivity of the French Internet, and thus assess its resilience.

Through the use of a new indicator, connectivity, we are able to examine the impact a deficient network will have on the reachability of the Internet by French ASes. We have shown that all ASes that take the precaution of having at least two providers, possibly a primary provider and a backup link, are not prone to being disconnected from the Internet by the disappearance of a unique faulty AS.

We also give a list of possible improvements and uses for this model. We believe a more detailed map would greatly improve our understanding of the Internet, and give us more insight on the consequences of other network failures, such as facility power outages. We also feel it can be used to actively monitor the network for AS spoofing.

References

[1] R. Chandra, P. Traina, and T. Li. BGP Communities Attribute. IETF RFC 1997, August 1996.

[2] François Contat, Mathieu Feuillet, Pierre Lorinquer, Guillaume Valadon, Stéphane Bortzmeyer, Samia M'timet, Mohsen Souissi, and Xavier Beaudouin. Résilience de l'Internet français, June 2013.

[3] A. Dhamdhere and C. Dovrolis. Twelve Years in the Evolution of the Internet Ecosystem. *IEEE/ACM Transactions on Networking*, 19(5):1420–1433, Sep 2011.

[4] Giuseppe Di Battista, Thomas Erlebach, Alexander Hall, Maurizio Patrignani, Maurizio Pizzonia, and Thomas Schank. Computing the Types of the Relationships Between Autonomous Systems. *IEEE/ACM Trans. Netw.*, 15(2):267–280, April 2007.

[5] Xenofontas Dimitropoulos, Dmitri Krioukov, Marina Fomenkov, Bradley Huffaker, Young Hyun, kc claffy, and Geogre Riley. AS Relationships: Inference and Validation. *ACM SIGCOMM Computer Communication Review (CCR)*, 37(1):29–40, Jan 2007.

[6] Benoit Donnet and Olivier Bonaventure. On BGP Communities. *SIGCOMM Comput. Commun. Rev.*, 38(2):55–59, April 2008.

[7] Lixin Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Trans. Netw.*, 9(6):733–745, 2001.

[8] Lixin Gao, Jian Qiu, Supranamaya Ranjan, and Antonio Nucci. Detecting Bogus BGP Route Information: Going Beyond Prefix Hijacking. *SecureComm 2007*, pages 381–390, 2007.

[9] Matthew Luckie, Bradley Huffaker, kc claffy, Amogh Dhamdhere, and Vasileios Giotsas. AS Relationships, Customer Cones, and Validation. In *Internet Measurement Conference (IMC)*, October 2013.

[10] D.L. Mills. A Border Gateway Protocol 4 (BGP-4). IETF RFC 904, April 1984.

[11] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4). IETF RFC 1654, July 1994.

[12] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). IETF RFC 4271, January 2006.

[13] Lakshminarayanan Subramanian, Sharad Agarwal, Jennifer Rexford, and Randy H. Katz. Characterizing the Internet Hierarchy from Multiple Vantage Points. Technical report, Berkeley, CA, USA, 2001.

[14] Q. Vohra and E. Chen. BGP Support for Four-octet AS Number Space. IETF RFC 4893, May 2007.