Violation of interdomain routing assumptions

Riad Mazloum
UPMC Sorbonne Universités

Marc-Olivier Buob
UPMC Sorbonne Universités

Jordan Augé
UPMC Sorbonne Universités

Bruno Baynat
UPMC Sorbonne Universités

Timur Friedman
UPMC Sorbonne Universités

Dario Rossi
TELECOM ParisTech *

ABSTRACT

We challenge a set of common assumptions that are frequently used to model interdomain routing in the Internet. We draw assumptions from the scientific literature and confront them with routing decisions that are actually taken by ASes, as seen in BGP feeds. We show that the assumptions are too simple to model real-world Internet routing policies. We also show that ASes frequently route in ways that are inconsistent with simple economic models of AS relationships. Our results should introduce a note of caution into future work that makes these assumptions and should prompt attempts to find more accurate models.

1. INTRODUCTION

Figure 1 shows a few of the Internet's autonomous systems (ASes), and the connections between them, gleaned from publicly-available BGP feeds, which are reports from routers that run the BGP inter-domain routing protocol. This figure illustrates a case of what is known as *multi-exit routing*. AS6762, which is Telecom Italia's Sparkle, the world's 9th most important AS as reported by CAIDA's AS Rank service [1], has two routes to reach the address prefix 103.11.245.0/24, which is advertised by AS5845. (Each of the Internet's more than 40,000 ASes is identified by a unique AS Number (ASN).) The first route goes via AS10026, Pacnet Global Ltd., while the second goes via AS1299, TeliaNet Global Network. CAIDA's AS Relationships database [2] tells us that Pacnet is a client of Sparkle, wheras TeliaNet is Sparkle's peer. The common assumption made in the research community in such a case would be that Sparkle will route all of its traffic for this prefix through Pacnet, since Pacnet pays Sparkle for traffic that it receives, whereas it receives no revenue from TeliaNet. Indeed, the BGP feeds tell us that Sparkle only advertises the first route, via its client Pacnet, to its neighbor AS262589, INTERNEXA Brasil Operadora de Telecommunicaes S.A., which makes sense. However, the BGP feeds also tell us that Sparkle advertises *only the second route, via its peer, TeliaNet*, to another neighbor, AS26615, Tim Cellular S.A.. This is strange.

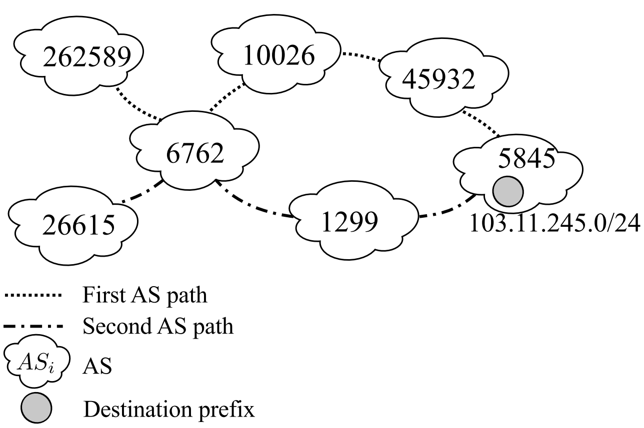

Figure 1: Multi-exit routing example

What is wrong? Could it be an error in CAIDA's AS Relationships database? In the case of multi-exit routing, we would expect both neighbors of an AS to have the same relationship to that AS. Both would be peers, for instance, or both would be clients, or both providers (the three main possibilities, along with sibling relationships, which do not alter this reasoning). Imagine they are both peers. Then some other assumptions that are common in the research community should apply: that Sparkle will only advertise the route via TeliaNet because it has the shorter AS Path (sequence of AS hops to the destination prefix), unless it has set a parameter called `local_pref` to override this preference, in which case it will only advertise the route with the longer AS Path, via Pacnet. Here, it is advertising both routes, which is, again, strange.

The only conclusion that we can draw from this example is that none of the commonly-held assumptions hold. When we try to apply them, they conflict with the reality of this AS topology, as seen through the BGP feeds. This paper looks at 4 million routes that we collected from BGP feeds, and in particular at 170 thousand instances of multi-exit routing, or *multi-exits*, that

*Joint collaboration through the LINCS lab (www.lincs.fr).

those feeds reveal. In 30% of the cases, the assumption about routing preferentially to customers over peers and to peers over providers was not coherent with the relationships, as described by CAIDA. In fully 55% of the cases, the path length assumption did not hold.

This paper proceeds in Sec. 2 by providing some BGP background for readers who are not familiar with the details. In this context, it formalizes four commonly-held assumptions, and cites examples in the literature where they are made. (The assumptions described above are composites of these four assumptions.) Sec. 3 describes our methodology for confronting the assumptions with the data. Results appear in Sec. 4. The paper wraps up with related work (Sec. 5) and a conclusion pointing to future work (Sec. 6).

Our contributions are to formalize commonly-held assumptions about interdomain routing and AS relationships and quantify the extent to which none of them are adequate to model the data provided in publicly-available BGP feeds. Our results should introduce a note of caution into future work that makes these assumptions and should prompt attempts to find more accurate models.

2. BGP AND OUR SET OF ASSUMPTIONS

2.1 BGP background

BGP is the interdomain routing protocol that allows an AS to learn how to route to destinations in other ASes. A BGP route describes the *AS Path*, or sequence of ASes, to be traversed on the way to a *prefix*, which is a set of continugous IP addresses. The *BGP next-hop* is the egress point to use at the IP level in order to follow the route. Routes are exchanged between routers in the same AS using *iBGP*, and between routers in different ASes via *eBGP*.

In the general case, a BGP router learns several routes toward a given destination. These routes are said to be *concurrent*. The router elects one route among them (the *best route*) by following the *BGP decision process*. This consists of a sequence of *steps*, shown in Table 1, that successively reduce the set of concurrent routes. At each step, routes dominated by at least one concurrent route are discarded. When, after one of these steps, there remains just one element in the set, then this element is the best route.

When a BGP router receives a route, the router is free to accept it or to modify it. And when the decision process elects the best route, the router is also free to select to which of its neighboring routers it will forward it. One modifiable parameter that directly affects the choice of best routes is the *local preference* value.

If a router receives two routes ρ and ρ' toward the same destination with a higher value of `local_pref` assigned to ρ, then ρ is preferred to ρ'. Since an AS typically wishes to control its own routing policy, its routers will, prior to starting the decision process, set the `local_pref` values of all routes received via eBGP sessions in accordance with that policy.

1. Highest local preference
2. Shortest AS path length
3. Lowest origin type
4. Lowest multi exit discriminator (MED)
5. eBGP over iBGP
6. Lowest IGP cost
7. Tie break rules

Table 1: The BGP decision process.

Gao [9] introduces a model based on economic relationships between ASes, identifying three common types of relationship:

- *Customer-provider (c2p)*: a customer pays a provider for transit service to the rest of the Internet for its traffic and its customers' traffic.
- *Peering (peer)*: a pair of ASes transit traffic between each other and their respective customers, free of charge.
- *Mutual-transit (s2s)*: a pair of ASes (called *siblings* in this case) transit traffic for each other and for their respective clients to every destination in the Internet, free of charge.

Most subsequent work is based on this classification.

2.2 Set of assumptions

This section describes four assumptions about interdomain routing that are frequently found in the literature.

(A1) iBGP valid.
The assumption is that any BGP route has the potential to be propagated within an AS to all of that AS. In other words, route propagation is *only* governed by routing decisions taken by the different routers in the AS and there are no parts of the AS to which a route cannot be forwarded.

This assumption is usually made by default since an AS should guarantee this property in order to assure that all of its routers are selecting the best routes [3, 12, 13, 17]. Nevertheless, there are some examples of violations that can appear [8].

(A2) The routing policy of an AS is only implemented in eBGP sessions.
This means that routers in iBGP sessions do not modify the attributes of the routes, all modifications being made by routers having interdomain sessions. We focus in this assumption on the `local_pref` value, since this value is the first to be checked by the decision process.

If a router modifies this value for some or all of the routes in an iBGP session, this may affect the choices of all routers in the AS to which this route is eventually forwarded.

This assumption is made to simplify the model of route propagation in an AS and has been made in previous research [3, 12, 13]. However, a BGP router may in reality modify the `local_pref` value of some of its routes in an iBGP session in order to change the preference of routes. Neudorfer et al. [14] are, for instance, aware of such cases.

(A3) An AS always prefers to send traffic toward a client over a peer over a provider.
This assumption is based on the presumed economic benefits. Sending traffic through a client means that the client will pay for it, while sending though a provider means paying the provider. An AS will implement this hierarchy through its local-preference settings. For example, AS_1 has a client AS_2 and a provider AS'_2. It assigns a value of `local_pref` to routes that are learned from AS_2 that is higher than that assigned to routes learned from AS'_2, in this way always preferring routes learned from AS_2.

This assumption does not always hold, as shown, for example, by Mühlbauer et al. [13].

Note that this assumption might seem at first quite similar the well-known *valley free* assumption[1] [9], since both of them are based on economic benefits. The difference is that valley free is only concerned with the propagation of routes, i.e., which routes are allowed and which are filtered, while the assumption (A3) is concerned with the preference among the received candidate routes, whether they are valley free or not.

(A4) There is only one type of relationship established between two ASes.
In most of the literature, each AS interconnection is modeled as a single business relationship [4–6, 9, 12, 13, 16, 18, 19]. This assumption is convenient for most of the previous work since the main source of data used is AS paths carried on BGP routes. These paths provide only AS level information. We do not know which egress points are traversed, for example.

This assumption is not respected in some cases. We can find for example an AS_1 that is a customer of an AS_2 for the set of prefixes $\mathcal{P}_1$ and at the same time is a peer for the set $\mathcal{P}_2$. We can also find ASes that interconnect in two places, where AS_1 is a customer of AS_2 in Europe, for instance, while they peer in the United States.

Multiple articles have shown that considering one relation per AS pair is too simple, and that there are other models that could better represent the relationships [10, 14].

[1] The valley-free property expresses the idea that no AS will provide connectivity between two other ASes without benefits. The validity of this assumption has been studied earlier [11].

3. METHODOLOGY

It would be possible to challenge, and possibly invalidate, the individual assumptions described in the previous section if we had detailed knowledge of the routing decisions made by BGP routers, which is not available to us. However, the publicly-available BGP feeds do allow us to challenge combinations of assumptions.

Our methodology is novel in the way that we use multi-exits, the general case of which is illustrated in Fig. 2. We infer multi-exits from the data and then extrapolate from them to discover the routing decisions that must have been taken within ASes.

Sec. 3.1 precisely states the implications of multi-exit routing observations for each assumption. Sec. 3.2 presents a multi-exit detection methodology. Finally, Sec. 3.3 identifies criteria for detecting assumption violations.

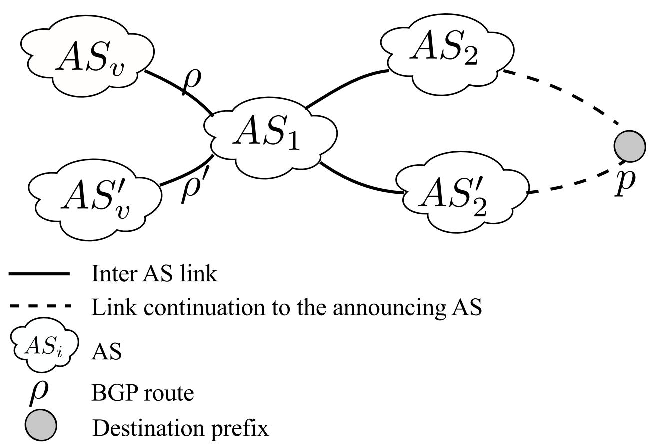

Figure 2: AS_1 is doing multi-exit routing through AS_2 and AS'_2 by announcing ρ and ρ'

3.1 Deductions

Multi-exits and AS path length (A1, A2)

Going back to the routing decisions taken by BGP routers in an AS under (A1) and (A2), let us consider two BGP routers R and R' belonging to the AS AS_1 that announce two different routes ρ and ρ' to AS_v and AS'_v toward the same destination prefix p, as illustrated in Fig. 2. Since there is no one dominant route in the AS, both routes must have the same `local_pref` values and the same AS path length. This clearly holds as, otherwise, the route having the highest `local_pref` or the shortest AS path would have been chosen by all of the routers in AS_1.

It follows that if (A1) and (A2) jointly hold, if an autonomous system AS_1 is observed to do multi-exit routing through multiple different next-hop ASes such as AS_2 and AS'_2, then the routes ρ and ρ' related to these ASes *must* have the same AS path length.

Multi-exits and AS relationships (A1, A2, A3, A4)

In addition to (A1) and (A2) we now bring in (A3) and (A4). According to (A3), an AS assigns higher values of `local_pref` to its customers than those of its peers, which are in turn higher than those of its providers. At the same time, according to (A4) there is only one relationship between two ASes, which means that there is one value of `local_pref` per AS and that this value further corresponds to the type of the relationship.

In other words, if two routes ρ and ρ' toward the same destination have the same `local_pref` value in an autonomous system AS_1 then they have been received from two ASes AS_2 and AS'_2 that have the same type of relation with AS_1.

It follows that, if an AS AS_1 is observed to do multi-exit routing through two different ASes AS_2 and AS'_2 then AS_2 and AS'_2 *must* have the same type of relation with AS_1 (e.g., they are both clients of AS_1).

3.2 Multi-exit observation

A BGP snapshot at a given instant t is simply the set of all of the BGP routes being used by the vantage points at this instant.

When we consider a snapshot, we need first to prune AS paths in order to remove AS number repetitions that are due to AS path prepending[2]. Actually, we can see the AS path as a sequence of AS numbers (AS_1, ..., AS_i, ..., AS_k). Now, for each AS AS_i of the AS path and for each destination prefix p related to this path, we extract the next-hop AS AS_{i+1} used by AS_i to reach the destination p. In this way we build the set of *BGP triplets*, $\mathcal{T}_{BGP} = \{(AS_i, AS_{i+1}, p)\}$.

Looking at these triplets, a multi-exit is detected whenever we observe two (or more) triplets of the form (AS_i, AS_{i+1}, p) and (AS_i, AS'_{i+1}, p). Intuitively, these triplets correspond to the fact that AS_i routes its traffic to p through both AS_{i+1} and AS'_{i+1}.

3.3 Validation

Multi-exits and AS path length (A1, A2)

We can easily identify cases when (A1) and/or (A2) are violated by checking the AS path length of routes identified in a multi-exit. If an AS announces two routes ρ and ρ' with an AS path length of ρ different from the one of ρ', we deduce that assumptions (A1) and/or (A2) are violated.

Further notice that, in case path lengths are equal, we cannot conclude that (A1) and (A2) are not violated – rather, no definitive conclusions can be made in this case.

Multi-exits and AS relationships (A1, A2, A3, A4)

Once we have extracted the multi-exits from the BGP triplets, as described in Sec. 3.2, suppose that we can label the relationships between the ASes as *c2p* or *peer*. (Without affecting our conclusions, we set aside the special case of *s2s*.) Cases in which the ASes of a multi-exit exhibit different AS relationships imply violation of at least one of the assumptions (A1, A2, A3, A4).

[2]An AS number is repeated multiple times to make some AS paths longer. This technique is usually employed to announce backup routes.

N° routes	3,948,447
N° stable routes	3,493,673
N° prefixes	459,532
N° vantage points	35 routers in 32 ASes
N° triplets	13,852,998
N° unique triplets	8,257,351

Table 2: Snapshot statistics

4. RESULTS

4.1 Data sources

Our study is based on two types of data, BGP and AS relationships.

We parsed BGP updates from BGPmon, which gathers data provided by RouteViews[3] and peers to some other BGP routers [20]. We stored these updates in their sequential time order, allowing us to extract snapshots for any instant we need.

We ran our analysis on snapshots taken in August 2012, then January and March 2013. Results presented here are based on a snapshot taken the 24$^{\text{th}}$ of March 2013 at 10:00:00 GMT. The other snapshots presented similar results. To avoid measurement anomalies, such as considering routes seen through one vantage point that may not be yet observed in others, if any, we consider only stable routes, i.e., routes that have not changed during the past 24 hours. Table 2 lists some statistics concerning the snapshot.

The choice of the set of AS relationships is harder since this data is not available publicly. There are multiple projects that offer inferred relationships, among which we chose CAIDA's relationship dataset since it is the only dataset we know to be public and verified against a ground truth [2].

4.2 Multi-exit quantification

[3]`http://www.routeviews.org/`

Before using multi-exits to examine assumption violations, we have first to check if multi-exits are indeed frequently observed, and then to check in what type of ASes they occur, because is it useless to challenge the assumptions based on infrequent observations. At the same time, it is also important to know in which types of ASes such behaviors are observed.

We start by extracting the BGP triplets as described in Sec. 3.2 and then select those that form multi-exits. Here we find that there are 6,776 ASes in the Internet that have at least one next-hop AS, and 22% of these are observed to perform multi-exit routing.

This result confirms our intuition that multi-exits are indeed frequent. Then, checking the classification of these ASes [7] we find that 38 ASes fall into Tier-1, and 556 ASes into Tier-2. This means that lessons that we learn from multi-exits will tend to apply to important parts of the internet.

4.3 Violations of AS path length assumptions

We find that 55% of the multi-exit routes involve unequal AS Path lengths.

We return to the example provided in the Introduction (Fig. 1). AS6762 (Sparkle) announces two routes to reach the prefix `103.11.245.0/24`. Fig. 1 shows that this AS announced two different routes, with two different path lengths, to ASes AS262589 and AS26615.

As a result, among the 1,486 ASes that are observed to do multi-exit routing, 489 are observed to violate (A1) and/or (A2). This could be caused, for instance, by an AS's iBGP router tweaking the `local_pref` value.

4.4 Violations of AS relationship assumptions

Regardless of the results concerning the unequal AS path length, multi-exits and AS relationships can still be crossed since we expect that even if an AS tweaks the `local_pref` value, this tweaking should reflect economic relationships. To illustrate this, let us consider an AS that has two clients c_1 and c_2 accessible through two different egress points e_1 and e_2. This AS can tweak the `local_pref` to force routers near e_1 to choose the route through c_1 and the routers near e_2 to choose e_2, respecting (A3) and (A4) while violating (A2).

We consider three datasets in turn. The first one consists of all of the triplets that form multi-exits. The second has only triplets forming multi-exits with equal AS path lengths. The third, only triplets forming multi-exits with unequal AS path lengths.

In the first dataset, fully 29% of the multi-exits are to ASes with differing relationships, which should not be surprising knowing that there are violations of (A1) and/or (A2). In the second dataset, only 10% of the multi-exits differ. This result is more coherent with the assumptions, but still high. The result also contrasts with Mühlbauer et al.'s statement [13] that AS relationships do not reveal information about local preference. Here, once paths with unequal AS path lengths have been removed, local preference seems to reveal AS relationships in 90% of the cases. Finally, 44% of the multi-exits differ in the third dataset although the AS path lengths are not equal. This implies that an AS sometimes may violate assumptions (A1) and (A2) respecting (A3) and (A4).

To summarize, we cannot limit the reasons for mismatching to just one assumption, since multiple assumptions can be violated for the same reasons. Our added value is that we show here that this set of assumptions held together are violated in 30% of the cases.

4.5 Discussion

This section discusses some possible reasons for assumptions to be violated.

Misconfigurations

Misconfigurations can appear on multiple levels. For instance, setting wrong values of `local_pref` would favor a provider over a client. As Giotsas et al. show [11], 50% of the times when the valley free assumption is violated, it is not a misconfiguration. This means that we should look for other reasons that can lead to inconsistent observations. Misconfigurations violate all of our assumptions.

Traffic engineering

In some cases, an AS might prefer sending its traffic through a peer over a client or through a provider over a peer, for all or some of the prefixes. This may happen when the client has insufficient bandwidth or when a router is geographically situated in an area that is far from the rest of the AS. Regarding geographically large ASes, the network administrator may prefer to send traffic through a closer neighboring peer AS (although it may seem more expensive) in order to avoid sending all of the traffic across long paths within the AS (which are indeed expensive) to reach the client AS. Like misconfigurations, traffic engineering can violate the assumptions.

CAIDA AS relationship dataset

A drawback of the CAIDA dataset is that it provides only one relation type per AS pair. However, this cannot explain all mismatch cases. This is why we manually verified some of the cases of mismatching presented in Sec. 4.4 by checking the routing policies available on IRR [4]. For these cases, IRR was consistent with CAIDA. This indicates that contradictions we found in AS relationships were because of assumption violations and not errors in the AS relationships inferred by

[4] IRR is a declarative routing policies database. `http://www.irr.net/`

CAIDA.

IXPs

In some cases, the AS number of an IXP (internet exchange point) is added to the AS path, which may introduce strange behavior. Let's take for example an IXP AS_{IXP} that adds its ASN, and suppose that this IXP provides connectivity between two AS pairs, (AS_1, AS_2) and (AS'_1, AS'_2). AS paths will contain patterns such $AS_1 - AS_{\text{IXP}} - AS_2$ and $AS'_1 - AS_{\text{IXP}} - AS'_2$. As a result, we may observe a false multi-exit at AS_{IXP} towards AS_2 and AS'_2. There are more scenarios, but we do not have space to describe them here.

5. RELATED WORK

Violation of interdomain routing assumptions has been a subject of discussion during the past few years. Looking for a model to represent routing policies, Mühlbauer et al. [13] show that business relationships do not reveal information about the preferences. Giotsas et al. [10] show that AS relationships differ depending upon IP version (IPv4 vs. IPv6) in 13% of cases. Roughan et al. [15] summarize lessons about modeling ASes based on an extensive study of common assumptions. They observe, notably, that modeling an AS interconnection by a single connection is insufficient.

6. CONCLUSION AND FUTURE WORK

This paper formalized a set of assumptions about interdomain routing in the Internet that are commonly used in the literature. It employed a data-based methodology to challenge them, showing that these assumptions, when combined, are frequently violated. As a result, future work should be more cautious about using these assumptions. The research community would also benefit from more accurate models.

This work opens the door to deeper analysis of our dataset, taking into account the presence of IXPs in the topology, and also making use of ways to identify ASes by their geolocalization and their ranking. We have found that instances of multi-exit routing can be revealing, and we expect that studying their dynamics will tell us more about how ASes perform traffic engineering. We also believe that much more can be revealed by crossing the BGP data with IP level measurements.

7. ACKNOWLEDGEMENTS

We thank Martin Levy of Hurricane Electric for taking the time to impress upon us the weaknesses in assumptions (A2) and (A3).

8. REFERENCES

[1] The CAIDA AS Ranking service, http://as-rank.caida.org/.

[2] The CAIDA AS Relationships dataset, http://www.caida.org/data/active/as-relationships/.

[3] Buob, M.-O., Meulle, M., and Uhlig, S. Checking for optimal egress points in iBGP routing. In *Proc. DRCN* (2007).

[4] Di Battista, G., Patrignani, M., and Pizzonia, M. Computing the types of the relationships between autonomous systems. In *Proc. Infocom* (2003).

[5] Dimitropoulos, X., Krioukov, D., Fomenkov, M., Huffaker, B., Hyun, Y., claffy, k., and Riley, G. AS relationships: inference and validation. *SIGCOMM CCR. 37*, 1 (Jan. 2007), 29–40.

[6] Dimitropoulos, X., Krioukov, D., Huffaker, B., claffy, k., and Riley, G. Inferring AS relationships: Dead end or lively beginning? vol. 3503 of *LNCS*. Springer Berlin Heidelberg, 2005, pp. 113–125.

[7] Dimitropoulos, X. A., Krioukov, D. V., Riley, G. F., and Claffy, K. C. Revealing the autonomous system taxonomy: The machine learning approach. *CoRR abs/cs/0604015* (2006).

[8] Feamster, N., and Balakrishnan, H. Detecting BGP configuration faults with static analysis. In *Proc. NSDI* (2005).

[9] Gao, L. On inferring autonomous system relationships in the internet. *IEEE/ACM Trans. Netw. 9*, 6 (Dec. 2001), 733–745.

[10] Giotsas, V., and Zhou, S. Detecting and assessing the hybrid IPv4/IPv6 AS relationships. *SIGCOMM CCR. 41*, 4 (Aug. 2011), 424–425.

[11] Giotsas, V., and Zhou, S. Valley-free violation in internet routing — analysis based on BGP Community data. In *Proc. ICC* (2012).

[12] Mühlbauer, W., Feldmann, A., Maennel, O., Roughan, M., and Uhlig, S. Building an AS-topology model that captures route diversity. *SIGCOMM CCR. 36*, 4 (Aug. 2006), 195–206.

[13] Mühlbauer, W., Uhlig, S., Fu, B., Meulle, M., and Maennel, O. In search for an appropriate granularity to model routing policies. *SIGCOMM CCR. 37*, 4 (Aug. 2007), 145–156.

[14] Neudorfer, L., Shavitt, Y., and Zilberman, N. Improving AS relationship inference using PoPs. In *Proc. TMA* (2013).

[15] Roughan, M., Willinger, W., Maennel, O., Perouli, D., and Bush, R. 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *IEEE JSAC 29*, 9 (2011), 1810–1821.

[16] Subramanian, L., Agarwal, S., Rexford, J., and Katz, R. Characterizing the internet hierarchy from multiple vantage points. In *Proc. Infocom* (2002).

[17] Teixeira, R., Shaikh, A., Griffin, T., and Voelker, G. M. Network sensitivity to hot-potato disruptions. In *Proc. SIGCOMM* (2004).

[18] Weinsberg, U., Shavitt, Y., and Shir, E. Near-deterministic inference of AS relationships. In *Proc. Infocom Workshops* (2009).

[19] Xia, J., and Gao, L. On the evaluation of AS relationship inferences. In *Proc. Globecom* (2004).

[20] Yan, H., Oliveira, R., Burnett, K., Matthews, D., Zhang, L., and Massey, D. BGPmon: A real-time, scalable, extensible monitoring system. In *Proc. CATCH* (2009).