Scalable Route Selection for IPv6 Multihomed Sites

Cédric de Launois *, Steve Uhlig **, and Olivier Bonaventure

Department of Computing Science and Engineering
Université catholique de Louvain (UCL), Belgium
{*delaunois*, *suh*, *bonaventure*}@info.ucl.ac.be

Abstract. We propose the use of an improved Internet coordinate system to allow multihomed IPv6 sites to select the source and destination address pair that provides the lowest delay path. We describe the new coordinate system and evaluate its use on the RIPE test box data set.

1 Introduction

More and more ISPs and corporate networks are multihomed for reliability and performance reasons. Nowadays, at least 60% of stub domains are multihomed. It is expected that many sites will also require to be multihomed in IPv6. In order to preserve the size of the BGP routing tables in the Internet, every IPv6 multihoming solution is required to allow route aggregation at the level of their providers. Most IPv6 multihoming mechanisms proposed at the IETF rely on the use of several IPv6 provider-aggregatable prefixes per site, see [1] and the references therein. It has been shown that the use of multiple provider-aggregatable prefixes increases the number of paths available between multihomed sites [2], such that the number of paths available between two stub ASes is often higher than or equal to 4. The end-to-end delay is a major component of the performance of a path as it reflects its length and bandwidth. Also, the TCP throughput increases when the delay decreases. Thus, choosing low delay paths is important for most applications, not only for interactive real-time applications.

We propose to use a network coordinate system in order to select, for each source and destination pair, the IPv6 addresses to be used that will lead to a low delay path, without actively probing them. Our goal is to avoid all paths with really bad delays, and to use the lowest delay path as much as possible.

2 The use of Network Coordinate Systems to Identify paths with Lower Delays

Synthetic coordinate systems have been originally developed to allow an Internet host to predict the round-trip delays to other hosts, without having to contact

* Supported by a grant from FRIA (Fonds pour la Formation à la recherche dans l'Industrie et dans l'Agriculture, Belgium). This work is also partially supported by the Walloon Government within the WIST TOTEM project.
** Supported by the FNRS (Fonds National de la Recherche Scientifique, Belgium).

them first. Each host computes its synthetic coordinates such that the distance between the coordinates of two hosts predicts the RTT between them. For example, if two hosts have coordinates x and y respectively, the distance $\|x - y\|$ is a good predictor of the RTT between them. IPv6 hosts currently arbitrarily choose between multiple global-scope IPv6 addresses. We propose that hosts in IPv6 multihomed sites base their source and destination IPv6 address selection on the delays predicted by synthetic coordinates. The selection of those addresses is made once at the start of each flow (e.g. TCP connection) between two hosts.

A host might compute and publish its own coordinates in the Domain Name System. However, in enterprise or campus networks, hosts whithin a single IPv6 prefix will typically end up with approximately the same coordinates. In such a case, the coordinates can be associated to the prefix itself, as a good estimate of the coordinates of any host within this prefix. When this is not true, the network can be divided in blocks, each with its own coordinates, shared by all the hosts whithin the block. We propose that the DNS server of a site computes the coordinates for the few prefixes assigned to the site, and publishes them in the DNS. The coordinates can be advertised using a new DNS resource record. This resource record can possibly be associated directly with the domain name of a host, so that the coordinates and the domain name of a host can be provided together in a single DNS response message, when the DNS name is resolved.

The best path can thus be predicted using a single DNS request, instead of performing multiple series of delay measurements, each using several probes. Since the coordinates should not change frequently, they can be cached in the DNS resolvers, further reducing the cost associated to the prediction of the best path. Another option is that the DNS server makes the choice in behalf of the host by removing bad addresses from the DNS response message. In this case, no modification is required for the hosts.

In this paper, the coordinates are computed by the Vivaldi algorithm, because it is simple and fully decentralized [3]. A detailed description of the solution and of the evaluation is available [4].

2.1 Computing Stable Synthetic Coordinates

Unfortunately, the Vivaldi algorithm does not produce stable coordinates for the nodes when latencies do not satisfy the triangle inequality [4]. This happens quite often in the Internet, for example due to policy-based routing with BGP. Such behaviour is unacceptable in our case since it would require to constantly update the DNS with new coordinates. We propose two modifications to the original Vivaldi algorithm. The first is to improve the local error estimate, so that each node computes a better estimation of the accuracy of its coordinates. The second is to introduce a *loss* factor to prevent the system from oscillating indefinitely. The new algorithm, called SVivaldi, is presented in Alg. 1 and detailed in [4]. It can be shown that the new local error estimator allows to find better solutions, and that the use of a *loss* factor allows to produce stable coordinates, see [4].

Alg. 1. The SVivaldi algorithm

1: SVivaldi(rtt_{ij}, x_j, e_j)
2: $w = e_i/(e_i + e_j)$
3: $Neigh_i = Neigh_i \cup \{j\}$
4: $Rtt_i = Rtt_i \cup \{rtt_{ij}\}$
5: $e_i = \frac{\sum_{j \in Neigh_i} |\|x_i - x_j\| - rtt_{ij}|}{\#Neigh_i} \times c_e + e_i \times (1 - c_e)$
6: $loss = c_l + (1 - c_l) * loss$
7: $\delta = c_c \times w \times (1 - loss)$
8: $x_i = x_i + \delta \times (rtt - \|x_i - x_j\|) \times \frac{(x_i - x_j)}{\|x_i - x_j\|}$

3 Evaluation of the Quality of the Route Selection

We evaluate here the quality of the routes that are selected using the coordinates computed by SVivaldi. We use delay measurements provided by the RIPE NCC Test Traffic Measurements Service in order to compute the coordinates of 58 nodes. IPv6 multihoming with multiple prefixes is not currently deployed in the Internet. In order to simulate IPv6 multihoming, we follow a similar methodology to the one used in [5]. A virtual multihomed site is emulated by using collectively a few RIPE nodes in the same metropolitan area. This method models IPv6 multihoming where the provider-dependent prefixes advertised by the virtual site are aggregated by its providers. 13 multihomed sites are emulated using this method, a number similar to the study of Akella et al. on multihoming [5]. 10 sites are dual-homed, 1 is 3-homed, 1 is 4-homed and a last one has 8 providers.

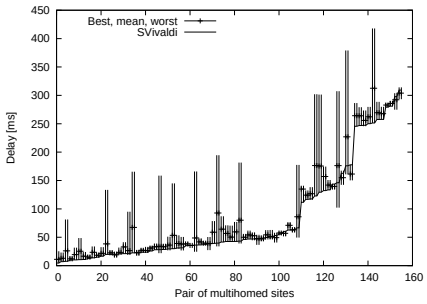

Fig. 1. The delay of the path chosen by SVivaldi for each pair of multihomed sites, in the RIPE data set.

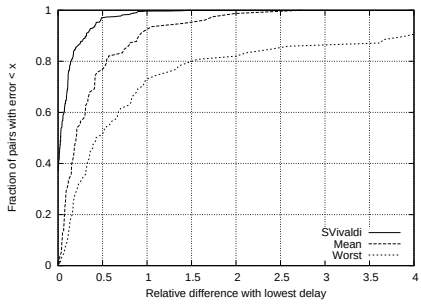

Fig. 2. Differences between the delay of the best path and the delay of the path selected using SVivaldi coordinates.

Fig. 1 shows the delay of the path chosen by SVivaldi for each pair of multihomed sites, sorted by increasing delay. The bars indicate the delay of the best, mean and worst path. We can see that the delay of the worst paths can sometimes be several times larger than than the delay of the best path. In this data set, SVivaldi never selects those really bad paths. For the large majority of multihomed pairs, SVivaldi even manages to select almost the best path. When SVivaldi does not select the best paths, the difference between the delay of the path selected and the best delay is not that large. Fig. 2 shows the relative difference between the path with the lowest delay and the path selected by different

path selection algorithms. It shows $f(x)$, the fraction of pairs of multihomed sites where a relative difference lower than x is observed. We see that SVivaldi finds the absolute best path in about 40% of the time, and selects a path with a delay at most 20% worse than the best delay for more than 85% of the pairs of multihomed sites. Note that in IDMaps [6], a path selection is considered correct if the delay of the selected path is within a factor of 2 times the delay of the best path. Following this criteria, SVivaldi practically never selects a wrong path. Fig. 2 confirms that SVivaldi succesfully manages to avoid all really bad paths, i.e. paths where the delay is more than twice the best delay. According to Fig. 2, the worst delay is more than twice the best delay for about 25% of IPv6 multihomed sites pairs, so these bad paths are not unusual.

4 Conclusion

With IPv6, the use of multiple prefixes increases the number of paths available to a multihomed site. Selecting the path with the lowest delay is important for many applications. A first contribution is to propose the use of a network coordinate system as an efficient and scalable way to help IPv6 hosts select the best source and destination IPv6 prefixes. We have shown that the use of synthetic coordinates is a scalable way to select good paths and to avoid all really bad paths. Our experiments with the RIPE data set have shown that we can select paths with a delay at most 20% worse than the lowest delay for more than 85% of the pairs of multihomed sites. A second contribution is SVivaldi, an improved version of the Vivaldi algorithm for computing synthetic coordinates. We have introduced two modifications in order to stabilize and produce more accurate coordinates.

Acknowledgments We thank the RIPE NCC for providing the Test Traffic Measurements Service.

References

1. Huston, G.: Architectural approaches to multi-homing for IPv6. Internet Draft, IETF (2004) <draft-ietf-multi6-architecture-02.txt>, work in progress.
2. de Launois, C., Quoitin, B., Bonaventure, O.: Leveraging Network Performances with IPv6 Multihoming and Multiple Provider-Dependent Aggregatable Prefixes. In: QoS-IP 2005, LNCS 2856, pp.118-179, Catania, Italy (2005)
3. Dabek, F., Kaashoek, F., Morris, R.: Vivaldi: A decentralized network coordinate system. In: Proceedings of ACM SIGCOMM'04, Portland, Oregon, USA (2004)
4. de Launois, C., Uhlig, S., Bonaventure, O.: Scalable route selection for IPv6 multihomed sites. `http://www.info.ucl.ac.be/people/delaunoi/networking05/` (2005)
5. Akella, A., et al.: A comparison of overlay routing and multihoming route control. In: Proceedings ACM SIGCOMM'04. (2004)
6. Francis, P., Jamin, S., Jin, C., Jin, Y., Raz, D., Shavitt, Y., Zhang, L.: IDMaps: A global internet host distance estimation service. In: Proceedings of IEEE/ACM Transactions on Networking. (2001)