Scalable Inter-domain Routing Architecture

Frédéric Olive
2002-12-15

Abstract

The current BGP routing architecture has several flaws that threaten the overall Internet stability. Some of them have been successfully addressed in the late 90's, but instabilities still remain and even tend to increase. This suggests that the BGP architecture may not be able to cope with the inexorable growth of the Internet. Other alternative architectures appear to be more scalable and reliable in very large growing networks.

Introduction

As the Internet grows in size and in diversity of service requirements, providing a reliable inter-domain routing architecture becomes an increasingly crucial issue. So far, the current architecture relying mostly on BGP 4 manages to keep up with the pace of the Internet expansion, and the increasing constraints on reliability and service quality. However, predictions on the Internet growth, which, at a minimum, should double in size each year, in conjunction with traffic studies on BGP instabilities, suggest that some major changes will have to be implemented to avoid the Internet to collapse the future.

The BGP Routing Table

Almost all of Internet related metrics, among which the number of domain names registered each year, the number of connected computers, the amount of traffic exchanged and so on, show the same exponential growth pattern. Moreover, it seems that this expansion would be sustained by the introduction of new mobile Internet devices and services, such as the 3rd generation mobile telephones. This trend puts high pressure on the routing architecture, and is reflected on the size of BGP routing table at some point in the exterior routing domain of the Internet. Since new networks connected to the Internet announce their prefix into this table, its study, like in [3], gives an idea of what is happening in the routing architecture.

The graph on Fig. 1 shows the exponential growth of the BGP table. Some improvement in the routing system managed to slow down the expansion. The most relevant were CIDR and route aggregation, which allowed providers to considerably reduce the number of routes announced and stored. Hierarchical routing is indeed a really effective technique to manage a scalable routing architecture. However, the size of the BGP routing table resumed its exponential growth in 1998, doubling each year. The main reason for that is multi-homing. Since the cost of routers and communication has considerably decreased during the 90's, providers can
now afford a richer connectivity mesh to improve the resiliency of their services. This common behaviour lead to a lot more path between ASs, which results in more entries in the routing table.

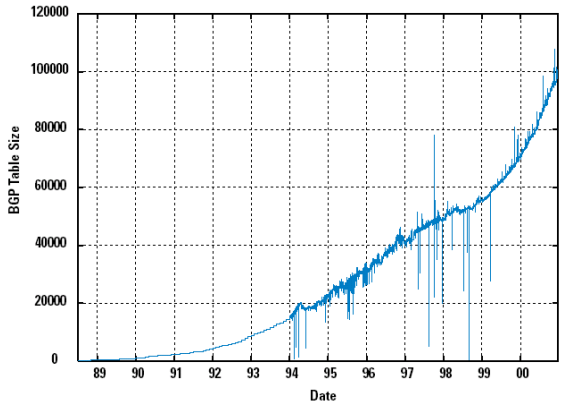

Fig. 1 : BGP Table growth (1988-2000), [3]

Area and Landmark Architecture

In order to cope with the unbearable growth of the BGP table, some alternative architectures have been proposed. The area and the landmark architectures, as described in [5], could be a elegant solution to make the BGP architecture scalable.

The strength of these architectures is their highly hierarchical routing scheme, based on several levels of area or Landmark routers. These routers are responsible for knowing routes within a certain portion of the Internet. Each node inside these regions has routes to them, which in turn have routes to other routers in exterior regions. The routing scheme is based on the source sending packets to its top area or top Landmark router, which knows how to forward packets in different parts of the Internet.

Introducing independent structures that are able to handle routes within the scope their interior network is quite similar to the concept of ASs and aggregation in the current architecture. However, the area and the Landmark architecture are not bound to one level of abstraction. Therefore, they are more suitable for very large networks. The size of the routing table of top-level routers would be equivalent to the size of the current BGP table, but the table maintained in lower level routers will considerably be reduced, since they must only keep routes to their higher level routers. The penalty paid for the savings in network resources is non-optimal paths.

BGP Instabilities

Even if routing protocols and algorithms have been widely studied, their actual deployment on a large scale reveals some unexpected pathological behaviours. Traffic studies on BGP updates in one of the largest public NAP ([1]) pinpoint some faulty software implementations responsible for most of the pathological excess in update traffic. Thanks to this study, most vendors corrected their router software, and by 1998 most of the BGP pathological flaws have been addressed, reducing BGP traffic as shown on Fig 2. However, routing instabilities still remain, without any provider or router vendor being at fault. There is still a important part of persistent route oscillations, policy and topology changes, and insecure routing configurations (such as the MED-IGP mapping policy), that generate legitimate but still excessive update traffic and can create temporary loss of connectivity.

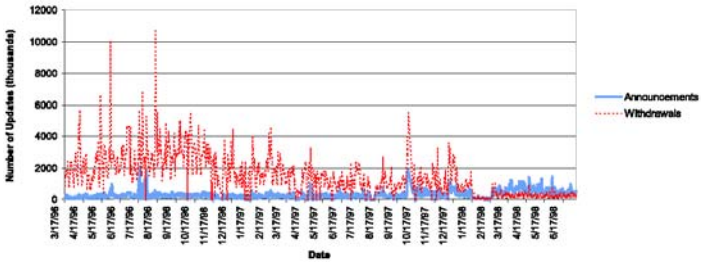

Fig. 2 : BGP update traffic at Mae-East, [1]. The BGP traffic significantly decreased after software update deployment in 1998.

On a user point of view, [4] concludes that end-to-end connectivity is usually guaranteed : the likelihood of path encountering some major routing pathological failure is around 1.5%. However, this figure has more doubled between 1994 and 1995, suggesting that the current architecture will be soon overwhelmed by the growth of the Internet.

The reasons for these increasing routing instabilities are not well clarified yet. Some trends in routing configuration appear to threaten the global routing stability. For example, coupling inter and intra-domain configurations is likely to create Internet-wide instabilities. The lack of hierarchy and abstraction in the architecture emphases the risk of failure, since configurations tend to be globally visible.

An other default of BGP revealed in [6] is that routes tend to have stable paths, despite the large number of BGP update messages exchanged about them. Moreover, a majority of updates concerns prefixes that do not receive much traffic. This enforces the overall idea that BGP is not scalable to very large network.

The Nimrod architecture

The Nimrod routing architecture is currently being developed to address the BGP architecture flaws. It is also designed to provide additional services, especially service-specific routing, i.e. based on service requirements, which is absent from BGP.

A full description of the Nimrod architecture is given in [2]. The essential differences between Nimrod and BGP are that Nimrod is highly hierarchical, with several level of abstraction, '*nodes*'; it unifies inter and intra-domain routing; and provides mechanisms to limit the amount of routing information that needs to be maintained throughout the network. It actually allows router to fetch information on demand, based upon service requirements of the traffic.

These few new features allow the Nimrod architecture to address most of the pathological issues of the BGP architecture revealed by the traffic studies. The nodes hierarchy, similar to an area hierarchy, limits the routing table size, and therefore make the system scalable to very large networks. Unifying inter and intra.-domain routing, thanks to several level of abstraction, avoids local failures to propagate to the global routing context. Finally, idea of routers acquiring on demand the routing information when generating routes efficiently reduce the need of repetitive routing update traffic.

Security issues

Security is today one of the prime concerns in network engineering. However, routing systems have not been immune against denial of service attacks. In this sense, the Nimrod architecture as described in [2] does not implement any security mechanisms to cope with this new threat.

Securing the Nimrod architecture, [7], is a crucial stake, since it is expected to be the next generation of routing architecture. Providing a safe and reliable infrastructure, in association with the current effective end-to-end security mechanisms, will make the overall Internet a secure and resilient system.

Conclusion

The BGP architecture now shows some sign of weakness as the Internet grows. It does not appear scalable to the size of the Internet. Some unexpected pathological flaws still overload the network with excessive update traffic, and end-to-end stability degrades. Alternative architectures are under development in order to replace it. Among them, the Nimrod architecture seems to be the most adapted to the next generation of Internet services requirements.

References

[1] Labovitz, Malan, Jahanian, "Origins of Internet Routing Instability", September 1997.

[2] Castineyra, Chiappa, Steenstrup, "The Nimrod Routing Architecture", RFC 1992, August 1996.

[3] Huston, "Analysing the Internet BGP Routing Table", February 2001.

[4] Paxson, "End-to-End Routing Behavior in the Internet", August 1996.

[5] Tsuchiya, "The Landmark Hierarchy", 1988.

[6] Rexford, Wang, Xiao, Zhang, "BGP Routing stability of popular destinations", AT&T, 2002.

[7] Sirois, Kent, "Securing the Nimrod Routing Architecture", IEEE 1997.