Internet Routing Stability

Jussi Vähäpassi
ICL Data Oy
P.O.BOX 458, 00101 Helsinki, Finland
jussi.vahapassi@icl.fi

Abstract

While the Internet is already, and in the future even more so, a serious business tool, pressure to increase its availability is rising. The new technologies such as VoIP, streaming video and e-commerce require QoS features from the Internet. But first the stability of the "Net" must be observed, and increased if possible. This paper introduces traditional routing protocols for use within service provider networks, and focuses on inter-domain routing protocol stability (BGP-4). Sources of excessive routing traffic, and performance of inter-domain routing protocol convergence are studied, and practical measurements that have been made recently are presented.

1 Introduction

We are almost daily dealing with Internet routing stability although we may not realize it. The availability on Internet routes is not even close to the availability of PSTN and transmission networks. Delayed web site responses, unreachable sites, "Error 404" and such are familiar to all net users. The web has not earned the nickname "World Wide Wait" for nothing. While these anomalies are sometimes caused by server overload or browser unresponsiveness, a major factor that is not easy to centrally manage is instability of Internet routing.

What makes Internet routing particularly difficult is that the Internet is a "network of networks", consisting of many parties who run their networks independently. What is worse, the benefits of these parties are sometimes conflicting. Solving routing anomalies and optimizing routes to the best of all is next to impossible. It is not uncommon to find service providers accusing each other of connectivity failures, while the customer itself has no clue where the problem lies. Despite the many conflicts of interests, the common goal of the service providers is to have connectivity to as many places as possible. The routing protocol that has been adopted in the Internet to accomplish this between ISPs is Border Gateway Protocol (BGP). While it hides the internals of the ISP network from other ISPs, it also enables the creation and enforcement of various policies. As the Internet consists of many interconnected networks, the routing convergence in case of failure in some part of the network is of greatest importance. Also the scalability to accommodate the tens of thousands of routes in the Internet default-route free core makes great demands to the routing protocol. BGP, and specifically its current version BGP-4, is the answer to these demands.

In smaller networks, particularly inside ISP and enterprise networks, a faster protocol is required. The requirement for topology hiding is no longer an issue inside a single service provider, but rapid convergence is most important. OSPF, IS-IS, EIGRP and RIP have all proven useful in these limited environments due to rapid convergence and ease of administration. They certainly have their differences, benefits and disadvantages, so there is still demand for all of these. While all of these are under constant development, a "boost" has been injected to the development by the emergence of multicast, demands for Quality of Service and IPv6.

In the following a short introduction to the most important protocols is presented. Routing protocols are divided into two classes: IGP protocols for intra-AS applicability, and EGP protocols for inter-AS routing. Next, types of instabilities are presented. While in practical engineering it is commonly understood that instability is equivalent to oscillation, here instead anything that causes route changes qualifies as instability. After that some sources that cause instability or excessive routing update traffic are examined. Finally, methods to reduce the instabilities are surveyed.

2 IGP protocols

Interior Gateway Protocols (IGPs), or interior protocols, carry network reachability information within a single Autonomous System (AS) or even within a smaller entity. The protocol operates as a distributed process, in which all routers participate in network topology discovery and find a route to all destinations. The routing information is then injected into each router's forwarding table. Examples of IGPs are RIP, IS-IS, IGRP, EIGRP and OSPF.

The most important requirement for IGPs is rapid convergence in case of network reachability information change. Convergence is the process of agreement, by all routers, on optimal routes. The criteria for optimal routes through the network to destination may vary. Simpler

protocols, such as RIP, use hop count as the only criteria; i.e. the path with least routers is the best path. A more advanced protocol, such as OSPF, works with configurable link metric that can be adjusted to represent the favorability of the link. The demand for QoS requires that several link metrics are transported in updates and used as criteria to select the route that satisfies the required QoS.

RIP, IGRP and EIGRP are examples of *distant vector protocols* that broadcast their complete routing table periodically, regardless of whether the routing table has changed. When the network is stable, distance vector protocols behave well but waste bandwidth because of the periodic sending of routing table updates, even when no change has occurred. When a failure occurs in the network, distance vector protocols do not add excessive load to the network, but they take a long time to reconverge to an alternate path or to flush a bad path from the network.

Link-state routing protocols, such as OSPF and IS-IS were designed to address the limitations of distance vector routing protocols (slow convergence and unnecessary bandwidth usage). Link-state protocols are more complex than distance vector protocols, and running them adds to the router's overhead. The additional overhead (in the form of memory utilization and bandwidth consumption when link-state protocols first start up) constrains the number of neighbors that a router can support and the number of neighbors that can be in an area. When the network is stable, link-state protocols minimize bandwidth usage by sending updates only when a change occurs. A hello mechanism ascertains reachability of neighbors. When a failure occurs in the network, link-state protocols flood link-state advertisements (LSAs) throughout an area. LSAs cause every router within the failed area to recalculate routes. The fact that LSAs need to be flooded throughout the area in failure mode and the fact that all routers recalculate routing tables constrain the number of neighbors that can be in an area ([1]).

2.1 RIP

RIP (Routing Information Protocol) is one of the simplest and most widespread routing protocols. It is defined in RFC2453 ([2]) and is an Internet Standard STD0056. Although some refer it to as "a broken protocol", it has proven useful in small networks or customer routers where the CPU power and amount of memory is low and the number and size of subnets is small. The primary reason for the ubiquitous availability of RIP is its deployment routing program "routed" in Berkeley distribution of UNIX. RIP is under extensive development in IETF in area of IPv6 and multicast.

RIP is a distance vector algorithm and it uses Bellman-Ford algorithm, and converges less rapidly than link state protocols. The protocol is well known and easy to configure.

RIP classifies routers as active and passive (silent). Active routers advertise their routes (reachability information) to others; passive routers listen and update their routes based on advertisements but do not advertise. Typically, routers run RIP in active mode, while hosts use passive mode.

RIP prevents routing loops from continuing indefinitely by implementing a limit on the number of hops allowed in a path from the source to a destination. The maximum number of hops in a path is 15. To adjust for rapid network-topology changes, RIP specifies a number of stability features that are common to many routing protocols. RIP implements the split-horizon and poison-reverse mechanisms to speed up the convergence in case two routers are engaged in a mutual loop. Simple "split horizon" scheme prevents the advertising of routes back to the interface from which they were learned. "Split horizon with poisoned reverse" includes such routes in updates, but sets their metrics to infinity (=16). These mechanisms improve the convergence, but "poison reverse" increases the amount of routing update messages ([2]). All router implementations include "split horizon", but "poison reverse" is not mandatory.

Split horizon can not stop a loop where three or more routers are involved. This loop will only be resolved when the metric reaches infinity. "Triggered updates" try to speed up convergence in this case. A router that sees the route metric changed will send an update immediately. While this will speed up convergence significantly in most cases, it still leaves a change for "count to infinity". This happens if some router, which has not gotten the triggered update yet, sends a scheduled update containing its outdated route. This could reestablish an orphaned remnant of the faulty route. If triggered updates happen quickly enough, this is very unlikely. However, counting to infinity is still possible ([2]).

The most notable advantages of RIP version 2 over version 1 include:

- Next hop: the next hop need not be the router sending the update.
- Network mask (VLSM): RIP v2 adds the ability to explicitly specify the network mask for each network in a packet, instead of having the same mask in a given network
- Simple authentication: packets may contain an authentication string that can be used to verify the validity of the supplied routing data. The security level is low, however.

Some of the limitations of RIP include:

- The largest network diameter supported is 15 hops
- The protocol depends upon "counting to infinity" to resolve certain unusual situations
- The metrics cannot be dynamic, such as delay or available bandwidth, but instead the metrics are fixed

2.2 OSPF

OSPF is an Interior Gateway Protocol (IGP). It is developed by IETF's OSPF working group as RFC 2328 ([3]) and STD0054, and thus it can be used to exchange routing information between routers from different vendors. As an IGP, OSPF distributes routing information between routers belonging to a single AS. The OSPF protocol is based on shortest-path-first, or link-state, technology.

OSPF provides very fast convergence. Given N link state packets the number of calculations required is proportional to

$$N*\log(N).$$

OSPF uses the Dijkstra algorithm ([4]) to find shortest path. OSPF allows hierarchical domain structure, in which all subdomains connect to backbone area. OSPF implements its own specific reliable transport protocol and flooding mechanism over IP for routing update propagation. Naturally, OSPF is capable of classless routing; it supports variable length subnet masks (VLSM) and route aggregation.

In a multi-access network, such as Ethernet, that has at least two attached routers OSPF routers elect a designated router and a backup designated router. The designated router floods a link-state advertisement for the multi-access network. The designated router concept reduces the number of adjacencies required on a multi-access network so that not every router has to make adjacency to all other routers, but instead they make an adjacency to a virtual router.

All OSPF protocol exchanges are authenticated. Authentication guarantees that routing information is imported only from trusted routers to protect the Internet and its users. A variety of authentication schemes can be used.

For a complete reference on OSPF, refer to [5].

3 EGP protocols

3.1 EGP

While EGP is a name for a family of protocols, it is also a name of a protocol itself. The Exterior Gateway Protocol (EGP) is an exterior routing protocol used for exchanging routing information with gateways in other autonomous systems. Unlike interior protocols, EGP propagates only reachability indications, not true metrics. EGP updates contain metrics, called distances, which range from 0 to 255.

Before EGP sends routing information to a remote router, it must establish an adjacency with that router. This is accomplished by an exchange of Hello and I Heard You (I-H-U) messages with that router. Computers communicating via EGP are called EGP neighbors, and the exchange of Hello and I-H-U messages is referred to as acquiring a neighbor.

EGP is now considered historic and is replaced by BGP.

3.2 BGP

The original BGP is specified in RFC1105 ([6]). BGP-2 is described in RFC 1163 ([7]), and BGP-3 in RFC 1267 ([8]). BGP-4 is defined in RFC 1771 ([9]), and it is the routing protocol used currently in the highest level hierarchy of the Internet.

BGP is related to EGP but operates with more capability, greater flexibility, and requires less bandwidth.
For BGP an important concept is Autonomous System (AS) that is a collection of networks that share a common routing policy. BGP is used as a routing protocol specifically between ASs. BGP-4 supports a crucial feature for today's Internet: classless aggregation of addresses. BGP Versions 2 and 3 are quite similar in capability and function, but they will only propagate classed network routes, and the AS path is a simple array of AS numbers. A mixture of BGP versions in not allowed within a single autonomous system BGP conversations [RFC1772].

The protocol in itself is quite simple. It relies on TCP for reliable transport. This eliminates the need to implement explicit update fragmentation, retransmission, acknowledgement, and sequencing. BGP uses AS_PATH attribute to prevent loops. AS_PATH is an ordered sequence of AS values to reach the originating AS. The art of BGP is in determining the routing policy and applying announcements and filtering accordingly.

BGP supports two basic types of sessions between neighbors, internal (sometimes referred to as iBGP) and external (eBGP). The configuration where BGP speaker communicates with BGP speakers of different ASs is termed eBGP. If the local AS has multiple exit points in separate routers, the routers need to communicate with each other internally using iBGP. This is required since if an IGP were used to communicate between border routers by redistributing routes to IGP then BGP attributes would be lost. iBGP and eBGP are the same protocol; just different rules.

The BGP process selects a single autonomous system path to use and to pass along to other BGP-speaking routers. The algorithm for path selection is as follows ([10]):

- Do not consider iBGP path if not synchronized
- Do not consider path if no route to next hop
- Highest weight (local to router)
- Highest local preference (global within AS)
- Shortest AS path
- Lowest origin code IGP < EGP < incomplete
- Lowest MED
- Prefer eBGP path over iBGP path
- Path with shortest next-hop metric wins
- Lowest router-id

Generally, iBGP speakers do not readvertise the iBGP learned route to other iBGP speakers because the routers do not pass routes learned from internal neighbors on to other internal neighbors, thus preventing a routing information loop. Thus BGP requires that full mesh connectivity be established between iBGP neighbors. This is not scalable to large networks and may lead to instability. To overcome this two solutions have been introduced: *route reflectors* and *confederations*.

With route reflectors, all iBGP speakers need not be fully meshed because there is a method to pass learned routes to neighbors. In this model, an internal BGP peer is configured to be a route reflector responsible for passing iBGP learned routes to a set of iBGP neighbors (clients). Since complicated reflector systems with backup reflectors are allowed there is an additional mechanism to prevent loops. Non-transitive IDs are added to updates so that loops can be detected.

Another way to reduce the iBGP mesh is to divide an autonomous system into multiple autonomous systems and group them into a single confederation. To the outside world, the confederation looks like a single autonomous system. Each autonomous system is fully meshed and has a few connections to other autonomous systems in the same confederation. Even though the peers in different autonomous systems have eBGP sessions, they exchange routing information as if they were iBGP peers. Specifically, the next-hop, MED and local preference information is preserved. This retains a single IGP for all of the autonomous systems ([11]).

4 Types of instability

Paxson ([12]) has made practical measurements on end-to-end routing behavior in the Internet. He describes the following routing pathologies that have occurred:

- routing loops
- erroneous routing
- rapidly changing routing ("fluttering")
- infrastructure failures
- excessive hops
- temporary outages

Routing algorithms are designed to avoid loops. Some algorithms are more effective in this than others, but the goal is common to all. Loops tend to form when connectivity changes take place. Routing protocols try to minimize the lifetime of loops since they represent total connectivity failures. Even short lived loops cause problems since they cause outbound traffic to return back to the router, thus stressing its routing and forwarding capacity. Paxson ([12]) notes that BGP loop suppression virtually eliminates inter-AS looping, but stable inter-AS routing does not guarantee stable end-to-end routing because AS's are entities capable of significant internal instabilities.

Erroneous routing is not very common in the Internet. Paxson observed one instance, where packets between Connecticut and London traveled via Israel. Whether this is accidental, we never know, but there is a lesson: you can never be sure about the security on the Internet!

Fluttering is the term used to describe rapidly changing routing. This may be due to load splitting, which may also have harmful effect if the round trip delay is very different in each path, so TCP performance is stressed. Also the path suffers from asymmetry, and network monitoring tools such as "traceroute" become confused, which makes troubleshooting difficult.

Infrastructure failures contain failures in underlying PSTN or other transmission facility. Failure may also be in router hardware or local connectivity.

The hop count in e.g. RIP is limited to 15, and traceroute defaults to 30. Paxson estimated that the mean path length is about 16 hops, and the diameter has grown beyond 30 hops. IP datagrams contain Time to Live (TTL) field that prevents datagrams from looping forever. The TTL field should be large enough to cover largest Internet diameter. It must be noted that hop count does not always reflect the geographical distance.

Temporary outages are more short-lived than infrastructure failures. Congestion may lead to an outage of a few seconds where packets are dropped.

5 Sources of instability

[13] shows that most of the delay in Internet path restoration stems from BGP processing delay. The adoption of path vector (BGP) is widely and incorrectly believed to provide BGP with significantly improved convergence properties over traditional distant vector protocols, such as RIP ([2]). In [13] Labovitz et al. study the convergence of BGP, and propose enhancements to

router vendor software that would enable more rapid convergence.

Labovitz et al. have studied the BGP behavior in Internet environment. They noticed that the number of pathological (redundant) BGP updates was dramatically higher than expected. They concluded that this was caused by software bugs and specific implementation decisions. Once vendors had corrected their software and most ISPs have updated their software, the level of BGP update instability was reduced by orders of magnitude ([14]).

In 1998 Labovitz et al. analyzed the sources of Internet outages within a large ISP's network in USA. Results are presented in [15], and recorded outages are presented in Table 1.

Table 1: Outages in observed network

Outage Category	Number of outages	%
Maintenance	272	16,2
Power Outage	273	16,0
Fiber Cut/Circuit/Carrier Problem	261	15,3
Unreachable	215	12,6
Hardware problem	154	9,0
Interface down	105	6,2
Routing Problems	104	6,1
Miscellaneous	86	5,9
Unknown/Undetermined/No problem	32	5,6
Congestion/Sluggish	65	4,6
Malicious Attack	26	1,5
Software problem	23	1,3

Most of the problems are not directly related to routing instability. Maintenance (planned and unplanned), power outages and circuit failures form the largest portion of the trouble tickets assigned.

In the following certain categories that can be seen as causing routing instability are studied in more detail.

5.1 Software bugs

Until 1998 the Internet showed high volume of BGP updates. Labovitz et al. studied this phenomenon in [14] and noticed that a vast majority of BGP traffic was pathological. This traffic consisted of redundant withdrawals of routes that were already withdrawn earlier. This was caused by a router feature that they called *stateless BGP withdrawal*. The processing burden for withdrawals is low so this may not be a significant problem. A certain router vendor had made a time-space tradeoff implementation decision: the router would not maintain a state regarding information advertised to router's BGP peers. Upon receiving any topology change the router would send withdrawals to all neighbors regardless of whether it has previously send the peer an announcement for the route. By summer 1998 all major service providers had deployed the fixed software and the number of BGP traffic decreased dramatically.

Another major component of BGP traffic found in [14] was duplicate route announcement in which a route is withdrawn and immediately replaced with a duplicate of the original route. This is redundant behavior. Labovitz et al. found two sources for this. The first was the inability of the software to determine whether a change in non-transitive attributes actually change the announcement to external peer. A non-transitive attribute is an attribute that only affect intra-AS routing behavior, such as LOCAL_PREF, and this attribute must not be included in updates that are sent to BGP speakers located in a neighboring autonomous systems ([9]). In one implementation a border router did not detect that a filtered route – absent the changed attribute – is a duplicate of the previously advertised route. The second source was the combination of minimum advertisement timer and stateless BGP implementation. There may be many route updates during the timer interval, and if the last update is the same as the one that was sent to peers just before the timer was initiated, an update should not be sent. But this implementation did not notice the similarity but instead it noticed that there had been updates so the last one should be sent.

Despite the AS_PATH attribute of BGP that is used to prevent loops, BGP does not converge very quickly. Experiment has shown that withdrawing a route in one part of the Internet causes several announcements of stale information in other parts of the network ([16]). It will take some time before the routing stabilizes, and this may trigger route damping, which would prevent the route becoming available again.

5.2 Router overload

The processing power of routers is crucial for the performance of the Internet. If processing performance for routing updates is insufficient, the router may fail under heavy routing instability. Also when links are congested the KeepAlive messages do not get through, and connectivity to neighbors is lost. This can initiate "route storm". When a router is unreachable under heavy load it is no longer reachable by its neighbors, who search for alternative routes by the failed router. When the router comes available again it sends large state dump to its neighbors. This dump may be too much for the neighbors to handle and they may fail. The initiated storm may affect even larger portions of the Internet ([15]).

New router generation is less prone to such failures, since they implement features that give higher precedence to BGP updates and KeepAlive messages, so these messages always get through. Also the packet forwarding is done fully in customized ASICs, so this leaves more power for the processor to handle the routing updates.

5.3 Route flapping

Route flapping can be defined as rapid fluctuation of route withdrawal followed by route reannouncement ([19]). This causes stress to routers, which have to recalculate routes every time a route availability information changes. This calculation consumes CPU and memory resources, and bandwidth since routing updates are passed to reflect changed conditions. Furthermore, it leads to loss of connectivity and packet loss.

Flapping may be caused by many reasons, e.g. configuration errors, software malfunctioning (bugs) and bad network connections that produce high error rate. In the worst case this may lead to successive collapsing of the routers' software in portions of the Internet. The most flapping routes are generally originated from non-CIDR systems, and their prefix is /22 or longer. Short prefix routes do not tend to flap.

5.4 IGP injection to EGP

Human error is always a considerable source of instability. Configuration errors in router configuration may cause large portions of the network to receive faulty information. This is partly helped by routing policies, and filtering received route announcements. Directly injecting IGP information is discouraged since errors in intra-AS network will show as routing updates outside the AS. The common practice is to aggregate intra-AS-routing information into larger prefix at the domain boundary. This leads into a smaller set of prefixes to be announced and thus to a greater possibility of achieving stability. The largest number of instabilities occur for /24 routes, which also constitutes the largest portion of prefixes.

An example in [14] is the IBGP mapped MED (MULTI_EXIT_DISC [9]). The value of this attribute may be used by a BGP speaker's decision process to discriminate among multiple exit points to a neighboring autonomous system. If the value for MED is determined from the intra-AS IGP metrics, the internal routing topology or policy may become visible to external peers and changes in internal network affect changes in intra-AS routing.

6 Dealing with Instabilities

6.1 CIDR

In the early 1990's it became clear that the class-based addressing hierarchy is not scalable for the Internet. Even the largest enterprises could not utilize the class A addresses efficiently, and class A was running out. Class B was in use and was running out, too. Class C was too small for many enterprises and advertising several class C's separately was going to be too much for backbone routers.

CIDR (Classless Inter Domain Routing) dampens route flaps by absorbing the flap within a larger aggregate, and is one of the driving mechanisms for improving routing in the Internet. It is a mechanism to remove the division of address space to classes A, B and C (and D, E), and permit aggregation at any bit boundary. The subnetting is also possible on any bit boundary, and it is called variable length subnet masks (VLSM). The number of prefixes in Internet routers' routing tables has decreased since the advent of CIDR. CIDR definitely has its advantages but due to changes in Internet topology, it was not able to noticeably affect the size of the global routing table. This is due to growing amount of private peering that tend to make effective aggregation difficult. This is nowadays called the "myth of CIDR". CIDR is specified in [17] and [18].

The interior (intra-domain) routing protocols that support CIDR are OSPF, RIP II, Integrated IS-IS, and E-IGRP. The exterior (inter-domain) routing protocol that supports CIDR is BGP-4. Protocols like RIP, BGP-3, EGP, and IGRP do not support CIDR.

6.2 BGP route flap damping

Route flap damping is a method of ignoring routing updates for routes that have been unstable. Damping only occurs on individual eBGP routes before routes are injected into IGP or redistributed via iBGP ([14]). Cisco's routers implement a penalty system in which a predefined amount of penalty is inserted to route entry at each flap. When the amount of penalty exceeds a predefined threshold, the route is suppressed and no more updates are accepted or sent for the particular route. The penalty is exponentially decreased and as another threshold is crossed the route is reconsidered. Route flap damping is specified in RFC 2493 ([20]).

Damping is not a perfect solution. While it does reduce the number of announcements in unstable situations, it does not improve the quality of updates. Damping may break connectivity since new and correct updates are not allowed anymore. Damping consumes router memory. Missing routes tend to cause calls to service providers' help desks. And the configuration complexity is added.

Despite these, damping has had a significant effect on decreasing the level of Internet routing instability.

Damping is applied as a consequence of certain BGP parameters. These generally include only route announcements and withdrawals. Some implementations also use AS_PATH and MED attributes. They do not provide dampening of IGP or iBGP routing information.

7 BGP Convergence

In [13] Labovitz et al. developed the lower and upper bounds for BGP convergence.

The adoption of path vector in BGP provides solution to count-to-infinity problem present on distance vector protocols. In [13] it is shown that in complete graph the path vector algorithm has an upper bound that is limited by the number of possible paths that traverses all neighbors though the network. In the worst case the algorithm may, with correct ordering of messages, explore all possible paths of all possible lengths. As there are k! permutations of paths of length k, it is shown that upper bound on convergence is

$$O(e*(n-1)!).$$

The best case convergence can be achieved in

$$O(n)$$

steps.

BGP, as a practical path vector implementation, does not exhibit in practice convergence problems of this scale. A parameter called *MinRouteAdvertisementInterval* determines the minimum amount of time that must elapse between advertisement of routes to a particular destination from a single BGP speaker. This parameter is defined as 30 s (jittered). The timer introduces a dampening effect as follows: a router will select a new route after the failure of the old one, and after sending an update it will not send further updates until the timer has expired ("penalty box"). It will continue receiving updates from neighbors and processing them even while in the "penalty box".

MinRouteAdvertisementInterval enhances the convergence significantly by introducing synchronization to the system. The number of messages and states is reduced significantly because a node must process all announcements from neighbors before sending a new announcement. The number of steps to achieve convergence is

$$O(n).$$

While the timer gives great relief to BGP convergence, it also artificially creates multiple 30 s rounds. The authors of [13] suggest a modification to BGP implementations so that loop detection is performed both on the sender and receiver side. RFC1771 does not define where the loop detection should take place, but generally it is not implemented on both sender and receiver sides. If the detection would be implemented on both sides, the convergence would in best case happen in one round. Labovitz claims that they have had success in convincing router vendors that this feature should be implemented.

The following tables present simulation results with no timer applied (Table 2), *MinRouteAdvertisementInterval* applied (Table 3), and a loop detection for both sender and receiver side applied (Table 4). It should be emphasized that Table 3 describes the behavior of current routers in the Internet.

Table 2: Simulation results [13]: No timer

Nodes	Time	States	Messages
4	N/A	12	41
5	N/A	60	306
6	N/A	320	2571
7	N/A	1955	23823

Table 3: Simulation results [13]: Timer applied

Nodes	Time	States	Messages
4	30	11	26
5	60	26	54
6	90	50	92
7	120	85	140

Table 4: Simulation results [13]: modified loop detection

Nodes	Time	States	Messages
4	30	11	26
5	30	23	54
6	30	39	92
7	30	59	140

8 Conclusion

BGP convergence has not been experimentally studied widely. In this text a few of the rare experimental studies have been examined. It has been observed that BGP implementations suffer from engineering tradeoffs. Protocol efficiency has sometimes been traded with CPU cycles and implementation easiness. This led to pathological BGP traffic. BGP convergence lower and upper bounds have been found. The limit of minimum update interval as specified in RFC1771 has been shown to be of utmost importance for the convergence of BGP, since otherwise the upper bound on convergence is extremely high. Types and sources of instability have been analyzed, and ways of dealing with them. Since the size of Internet is growing and the hierarchy is getting

flat, these kinds of studies are of more and more importance. Especially practical measurements that may last for months provide valuable insight into the behavior of the Internet.

References

[1] Internetwork Design Guide: Designing Large-Scale IP Internetworks, Cisco Systems, 2000, www.cisco.com.

[2] G. Malkin: RIP Version 2, RFC 2453, STD0056, November 1998.

[3] J. Moy: OSPF Version 2, RFC 2328 and STD0054, April 1998.

[4] R. Bhandari: Survivable Networks: Algorithms for Diverse Routing, Kluwer Academic Publishers 1999, ISBN 0-7923-8381-8.

[5] J. Moy: OSPF: Anatomy of an Internet Routing Protocol, Addison-Wesley 1998.

[6] K. Lougheed, Y. Rekhter: Border Gateway Protocol (BGP), RFC1105, June 1989.

[7] K. Lougheed, Y. Rekhter: Border Gateway Protocol (BGP), RFC 1163, June 1990.

[8] K. Lougheed, Y. Rekhter: Border Gateway Protocol 3 (BGP-3), RFC 1267, October 1991.

[9] Y. Rekhter, T. Li: A Border Gateway Protocol 4 (BGP-4), RFC 1771, March 1995.

[10] A. Freedman: Scaling the Network, Presentation in NANOG15, Denver, USA, January 1999 http://www.nanog.org/mtg-9901/ppt/bgp102/

[11] Cisco IOS Software Configuration, Cisco Systems, 2000, www.cisco.com.

[12] V. Paxson: End-to-End Routing Behaviour in the Internet, IEEE/ACM Transactions on Networking, Vol.5, No.5, pp. 601-615, October 1997, www.aciri.org/vern/papers.html.

[13] C. Labovitz, A. Ahuja, A. Bose, F. Jahanian: An Experimental Study of Internet Routing Convergence, Microsoft Research Technical Report MSR-TR-2000-08, February 2000.

[14] C. Labovitz, G. Malan, F. Jahanian: Origins of Internet Routing Instability, University of Michigan Technical Report CSE-TR-368-98, July 1998, www.eecs.umich.edu/home/techreports/cse98.html.

[15] C. Labovitz, A. Ahuja, F. Jahanian, Experimental Study of Internet Stability and Wide-Area Backbone Failures, University of Michigan Technical Report CSE-TR-382-98, November 1998.

[16] C. Labovitz, A. Ahuja, F. Jahanian: Analysis and Experimental Measurements of Internet BGP Convergence Latencies, NANOG Meeting October 1999, www.nanog.org/mtg-9910/converge.html

[17] Y. Rekhter, T. Li: An Architecture for IP Address Allocation with CIDR, RFC 1518, September 1993.

[18] V. Fuller, T. Li, J. Yu, K. Varadhan: Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy, RFC 1519, September 1993.

[19] G. Huston: ISP Survival Guide, John Wiley & Sons, 1999, ISBN 0-471-31499-4.

[20] C. Villamizar, R. Chandra, R. Govindan: BGP Route Flap Damping, RFC 2439, November 1998.