UNIVERSITÀ DEGLI STUDI DI ROMA TRE
Dipartimento di Informatica e Automazione
Via della Vasca Navale, 79 – 00146 Roma, Italy

Mediterranean Fiber Cable Cut (January-February 2008) Analysis of Network Dynamics

A. ANTONY[1], L. CITTADINI[2], D. KARRENBERG[1], R. KISTELEKI[1] T. REFICE[2], T. VEST[1], R. WILHELM[1]

RT-DIA-124-2008 **March 2008**

(1) Science Group,
RIPE NCC,
Amsterdam, The Netherlands.
{antony,dfk,robert,tvest,wilhelm}@ripe.net
(2) Dipartimento di Informatica e Automazione,
Università di Roma Tre,
Rome, Italy.
{ratm,refice}@dia.uniroma3.it

ABSTRACT

On the morning of 30 January 2008, two submarine cables in the Mediterranean Sea were damaged near Alexandria, Egypt. The media reported significant disruptions of Internet and phone traffic in the Middle East and South Asia. About two days later, a third cable was cut, this time in the Persian Gulf, 56 kilometers off the coast of Dubai. In the days that followed, more news on other cable outages came in.

We looked at the impact these events had on Internet connectivity by analyzing the data collected by the *Routing Information Service* (*RIS*) [7] of *RIPE NCC* [5] and using publicly available tools developed by the *Compunet Research Group* of *Rome Tre University* [2] and by the RIS.

This document has also been published as a RIPE NCC's document at [6].

1 Background

The history of submarine telecommunications cables goes back to 1850, when the first international telegraph link between England and France was established. Eight years later, the first trans-Atlantic telegraph cable linked Europe to North America. In the 20th century, telephony became the driving force for submarine cable deployments. TAT-1, the first trans-Atlantic telephone cable, was installed in 1956. It had the capacity to transmit 36 analog phone channels simultaneously. These days fibre-optic submarine cables carry the bulk of the trans-oceanic voice and data traffic. Maximum capacity is now in the order of 1 Tb/s, equivalent to 15 million old analog phone channels. Compared to satellites, submarine cables offer higher capacity and, because of the shorter distance, feature much better latencies. Cable systems are also more cost effective on major routes. In the past two decades, many of these cables have been deployed, primarily triggered by the explosive growth of Internet traffic.

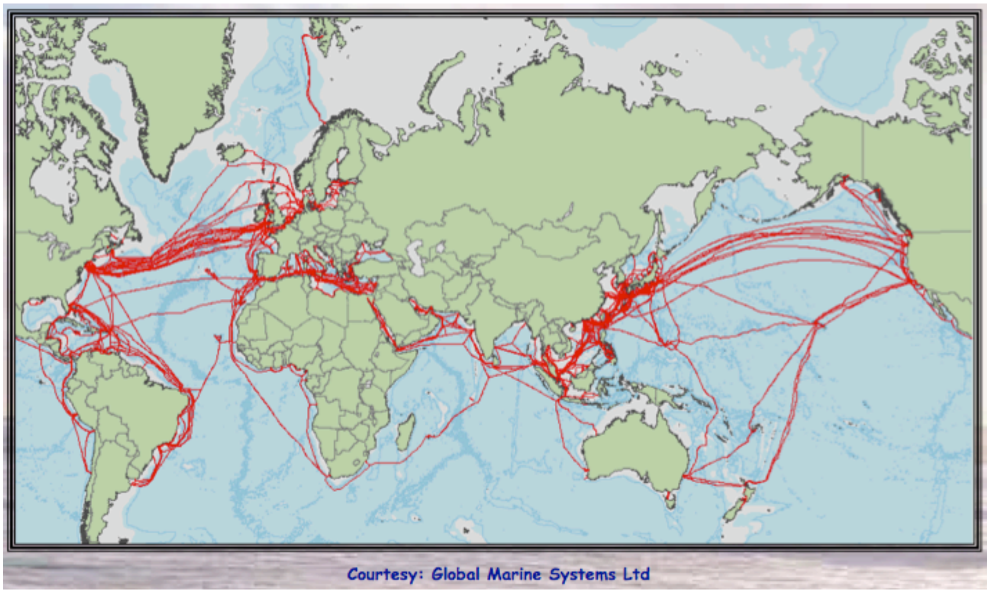


Figure 1: World map of submarine cable systems.

The world map of cable routes (Figure 1) shows that Europe, North America and East Asia are well connected. Numerous cables connect continents and countries. However, Africa, the Middle East and South Asia have far fewer cable systems. Looking at the available bandwidth or capacity in these cables, the differences become even more apparent. Faults in cables connecting these regions therefore have a higher impact than comparable faults in trans-Atlantic cables.

Although they rarely make news headlines, cable faults are not uncommon. Global Marine Systems, a company active in submarine cable installations and repairs, reported more than 50 failures in the Atlantic alone in 2007. A study [4] on behalf of the Submarine Cable Improvement Group shows 75% of all faults are caused by external aggression (physical damage). Of these, three out of four were attributable to human activities such as fishing, anchors and dredging. Natural hazards such submarine earthquakes, density currents and extreme weather were responsible for the remainder.

To limit the impact of such faults, cable systems often have build-in redundancy. Ring structures, for example, cross the ocean twice, each cable segment taking a geographically different route. When one segment breaks, signals can still reach a destination over the other segment(s). Repairs can then take place without much media attention.

1.1 Location of the Mediterranean Cables

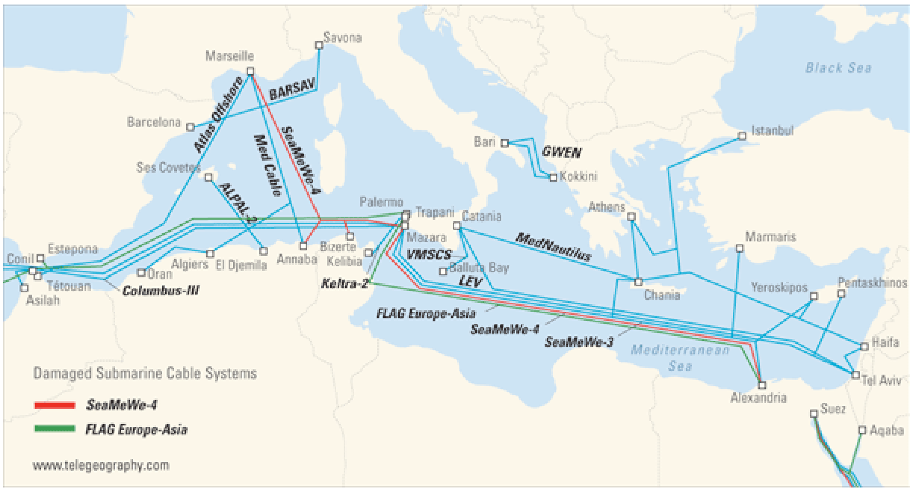


Figure 2: Mediterranean cables.

The International Cable Protection Committee provides lists of deployed cable systems on their website [3]. Several smaller cables connect most regions bordering the Mediterranean. Spain, Morocco, France, Algeria, Italy, Tunisia, Greece, Libya, Cyprus, and Israel have multiple connections. However, Figure 2 shows that only three cables connect Europe to Egypt, the Middle East and Asia: Flag Europe Asia, SEA-ME-WE4 and its predecessor SEA-ME-WE3. When the first two failed on 30 January 2008, the low capacity SEA-ME-WE3 was left as the only cable system providing a direct route from Europe to the region. All other options for rerouting traffic involved much longer cable routes or satellite systems.

1.2 Effects of a Cable Cut

When a communications cable system used for IP connectivity fails, two things can happen:

- Networks become unreachable, meaning they disappear from the Internet; or
- Traffic is rerouted.

Every individual IP link set up over the failed cable will be subject to one of these two options. Both options, however, can refer to a number of specific scenarios:

- Networks may become unreachable, because no action is taken, meaning packets fall in a “black hole”
- Networks may become unreachable, because the (only) upstream provider withdraws route announcements.
- Traffic may be rerouted on the IP level, either by manual reconfiguration or by routing protocols like BGP reacting to a loss of IP connectivity to previously preferred routers.
- Traffic may be rerouted on the data link layer. The (virtual) circuits on which an IP link has been set up are changed to follow a different physical path.

In the analysis below, we see evidence that all of the above occurred after the outages on the two Mediterranean cables. The outages on the FALCON cable are not visible in our data. If the FALCON cable failures had significant effects on Internet connectivity, these were obscured in our data by the network outages, the network congestion and the rerouting activities triggered by the problems in the Mediterranean.

2 Event Locations

Figure 3 shows where the events were located.

Figure 3: Google map of the cable cuts.

3 Event Timeline

The following events are known/confirmed:

Wednesday, 23 January 2008 (exact time unknown) FALCON cable, segment 7b damaged (Persian Gulf) Note: This is one week prior to the Mediterranean outages.

Wednesday, 30 January 2008, 04:30 (UTC) SEA-ME-WE-4 cable, segment 4/Alexandria-Marseilles, 25 kilometers from Alexandria, Egypt.

Wednesday, 30 January 2008, 08:00 (UTC) FLAG Europe-Asia cable (FEA), segment D (EG-IT) cut approximately 8.3 kilometers from Alexandria, Egypt.

Friday, 1 February 2008, 05:59 (UTC) FALCON cable, segments 2 and 7a (AE-OM) cut approximately 56 kilometers from Dubai, UAE.

Friday, 1 February 2008 (exact time unknown) Unidentified cable, between Halul (QA) and Das (UAE).

Friday, 8 February 2008 (exact time unknown) SEA-ME-WE-4 repair completed.

Saturday, 9 February 2008, 18:00 (UTC) FEA segment D repair completed.

Sunday, 10 February 2008, 10:00 (UTC) FALCON cable repair completed.

Thursday, 14 February 2008 Doha-Halul part of the unidentified QA-UAE cable "to be operational soon".

Note: Date and time for FEA segment D and FALCON segments 2 and 7a cable outages were reported by FLAG Telecom on their website. Other dates are from news reports. The SEA-ME-WE4 operators have not published exact times of cable failure and repairs. The timestamps for SEA-ME-WE4 failure are from clear observations in measurement data (both BGP monitoring and active measurements).

4 Dataset

Our analysis focused on publicly available data from the following RIPE NCC [5] services:

Routing Information Service (RIS)) [7] The RIS collects Border Gateway Protocol (BGP) routing information messages from 600 peers with 16 collection boxes called "Remote Route Collectors" ("RRCs") in near real time. Results are stored in a database for further processing by tools such as BGPlay [1], a visualisation tool. Three times a day, the route collectors take snapshots of their respective Routing Information Bases (often referred to as "RIB dumps"). Figure 4 shows the location of the RIS Remote Route Collectors.

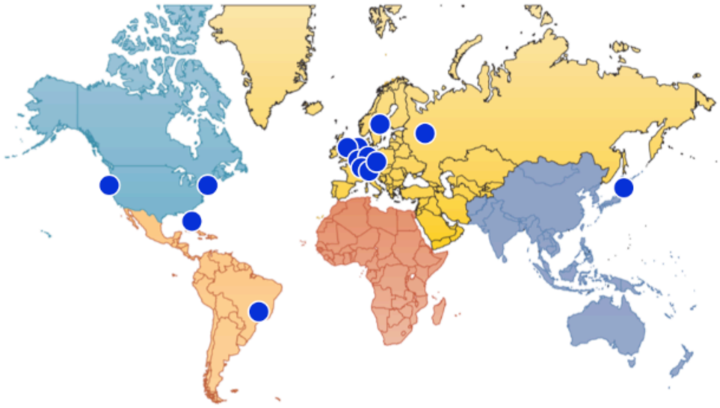


Figure 4: Location of RIS Remote Route Collectors.

Test Traffic Measurements (TTM) TTM measures key parameters of connectivity between a site and other points on the Internet. Traceroute vectors, one-way delay and packet-loss are measured using dedicated measurement devices called "test-boxes". Configured in an almost full mesh (that is, fully inter-connected), the TTM test-boxes continuously monitor end-to-end connectivity between the hosting sites. Figure 5 shows the location of the TTM nodes.

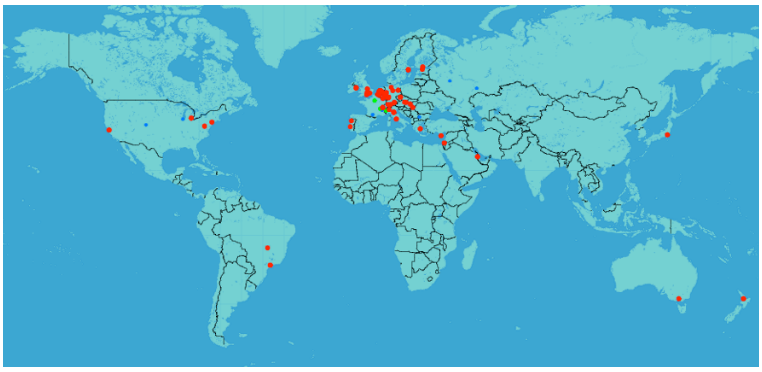


Figure 5: Location of TTM nodes.

DNS Monitoring Service (DNSMON) DNSMON provides a comprehensive, up-to-date overview of the service quality of root name servers, as well as participating country codes and generic Top Level Domains (TLDs). Using the infrastructure of TTM probes, DNSMON measures query response times to 196 name servers. Therefore DNSMON provides another perspective on possible effects the cable outages may have had on the Internet.

5 Analysis

In this section, we discuss the results of each of the services described above. However, we want to stress that the bigger picture only emerges when we combine the knowledge gained in the separate fields. Where BGP looks at routing information, active measurements by TTM and DNSMON provide a glimpse on how networks performed during the outage. Results from one service thus help us understand results from the others.

5.1 BGP Overall

The impact of the outages on BGP largely depended on how the cables were used in peering sessions.

- When peering between a Middle East-based provider and their transit in the West was established over the cables, or when the peering depended on receiving reachability information over the cable through some interior routing protocol, the sessions were dropped immediately after the failure. If the provider had no other way to route its IP traffic, the networks disappeared from the global routing tables at 04:30 or 08:00 (UTC) respectively.
- In cases where peering happens in the West, or where Middle Eastern networks are announced statically in the West, the sessions would not be expected to go down immediately after a cable failure. Instead, visibility in BGP would depend on manual network administrator action. This explains why some networks disappeared from the global routing tables several hours after the cable failures.

Where backup paths did exist, BGP explored and used them. We observed significant changes in the peering usage between larger transit providers. However, due to limited bandwidth and increased demand on the backup paths, sessions weren't always stable. In those cases BGP had difficulties converging on alternate topologies.

5.1.1 Prefix Counts

The number of prefixes seen in RIS is an indication of global network reachability. If a prefix cannot be seen and is not covered by an aggregate, it is likely the network in question has become unreachable. Therefore, total prefix count is a good starting point for BGP analysis. Using the data from the eight hourly RIB dumps, we graphed the total number of prefixes seen by RIS over time (Figure 6). The result does not show any clear drop or increase in the amount of prefixes near the known failure and recovery events. On a global scale, the cable outages only affected a small percentage of all networks.

However, when we assign country codes to each prefix, based on the information from RIR delegation statistics [8], we see various countries with significant reductions in the number of announced prefixes. Figure 7 compares, for each country, the number of prefixes seen in RIS over time with the number of prefixes visible at 00:00 (UTC), 30 January 2008. This gives an indication of the impact of the cable failures on network reachability in BGP. Egypt, Sudan and Kuwait were amongst the hardest hit, with drops of up to 40% in prefix visibility.

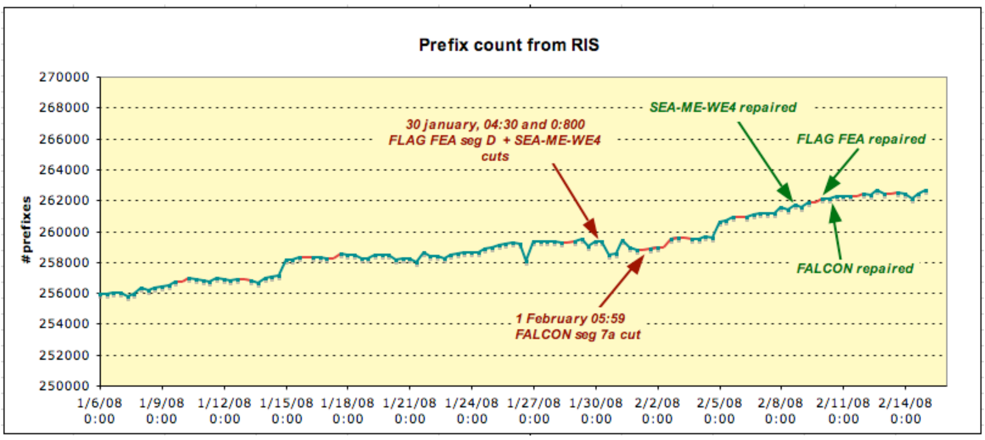


Figure 6: Total prefix count vs. time.

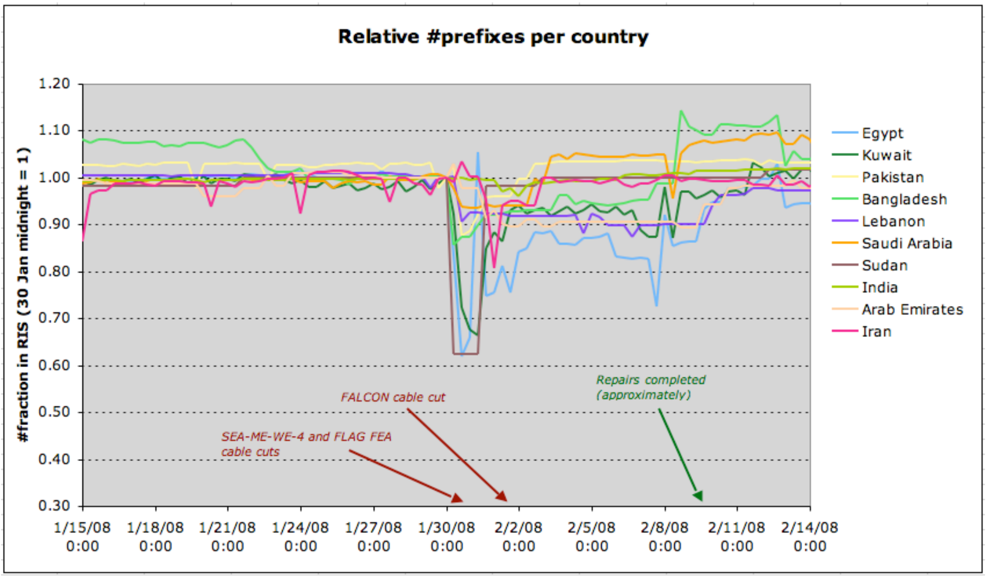


Figure 7: Relative prefix count for most affected countries.

5.1.2 Analysis of AS Path Changes

Next we looked at changes in Autonomous System (AS) paths. In relation to the cable cuts, the two main reasons for change are:

- Networks disappearing from the routing tables (this reduces the total number of AS paths)
- Networks rerouted on IP level (when the preferred transit is unavailable, BGP will try backup paths; these are likely to be longer)

Again starting from the RIB dumps, we compared consecutive dumps from the largest RIS route collector, RRC03, and determined the changes in distinct AS paths. Because AS path changes capture more of the event, we thought that graphing global level changes would clearly and unambiguously show when the cables snapped. That hypothesis did not hold. Although the cable cuts triggered a flux of changes, it is not the only such signal in a one month time period. Other events, either in BGP or related to RIS collector peers, triggered comparable levels of change.

As with the prefix counts, the correlation with the cable outages only becomes clear when we restrict the comparison to those AS paths where one or more of the constituent AS Numbers is registered to the region by an RIR.

Figure 8 shows the total number of distinct AS paths and the relative amount of change in distinct AS paths associated with the region. The changes seen align quite well with the event timeline.

Figure 9 again shows the total number of distinct AS paths associated with the region, this time augmented with the average length of those AS paths. As expected, the average AS path length increases around the time period of the cuts.

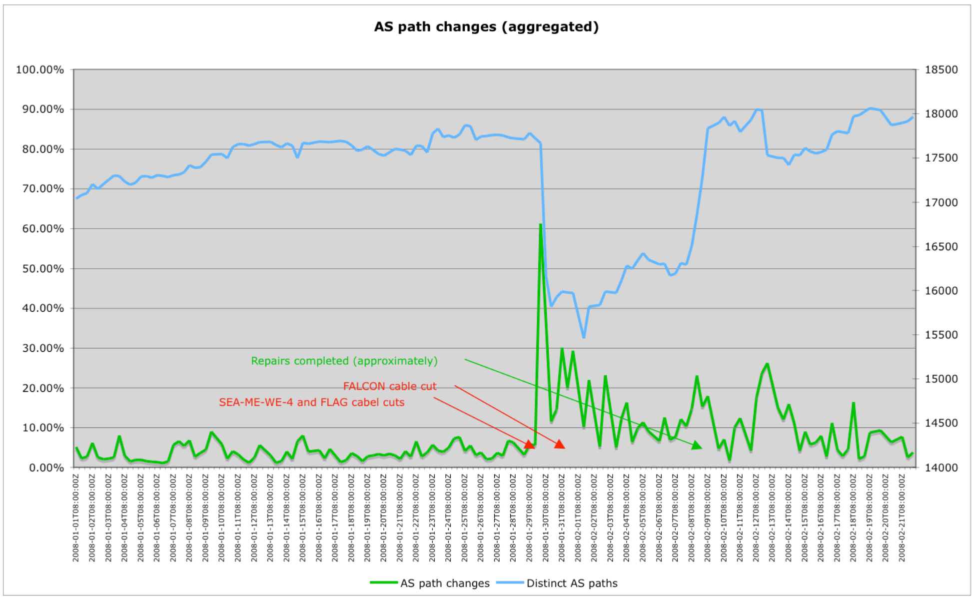


Figure 8: AS path changes (Aggregated).

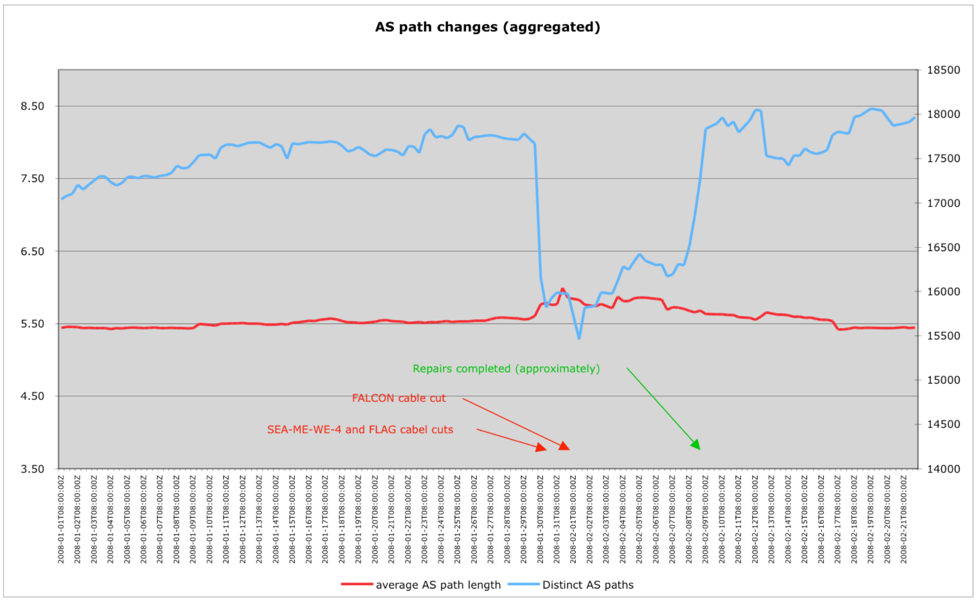


Figure 9: AS path changes (Aggregated).

5.1.3 Affected BGP Peerings

During the fibre cuts, we observed significant changes in BGP peerings usage. Some peerings went down and were not used again before the repair of the fibre(s). Others were selected as backup paths, which meant that their usage exploded. The metric we used to approximate the usage of a BGP peering between two Autonomous Systems is the size of the sets of all the prefixes routed through the peering, as seen by all the available collectors at once.

From our data sources, we extracted the BGP peerings which, from 16:00 (UTC), 29 January to 08:00 (UTC), 3 February, matched the following criteria:

- Carried, at some time, a significant number of prefixes (that is, more than 50 distinct prefixes) as observed from all the available collector peers
- Routed no traffic (that is, zero prefixes world-wide) for more than 12 hours

590 distinct BGP peerings involving 309 distinct Autonomous Systems matched the criteria described above. The ratio between the number of peerings and the number of Autonomous Systems suggests a high correlation of events.

As expected, AS15412 (FLAG Telecom) was one of the most involved Autonomous Systems, being present in more than 40 peerings. Other highly affected Autonomous Systems included AS4788 (TMNET, Malaysia), AS7575 (AARNET, Australia) and AS7473 (Singapore Telecommunications Limited), which were involved in more than 70 peerings overall.

The most affected AS was AS8966 (Emirates Telecommunications Corporation), which accounted for more than 100 peerings. From communications on the LINX-ops mailing list, we learned that AS8966 stopped advertising prefixes at the London Internet Exchange because of lack of capacity on back haul links. This meant that all peers at LINX explored alternative paths towards AS8966.

Note that the selection criteria for the peerings will give us the failing links, links which went down on purpose and links which were used for backup connectivity.

Backup links As shown in Figures 10 and 11, some previously unused peerings suddenly began to route a large number of prefixes, probably due to the unavailability of other (preferred) routes.

Large sets of prefixes were rerouted and subsequently withdrawn within a few hours, thus BGP convergence likely slowed down.

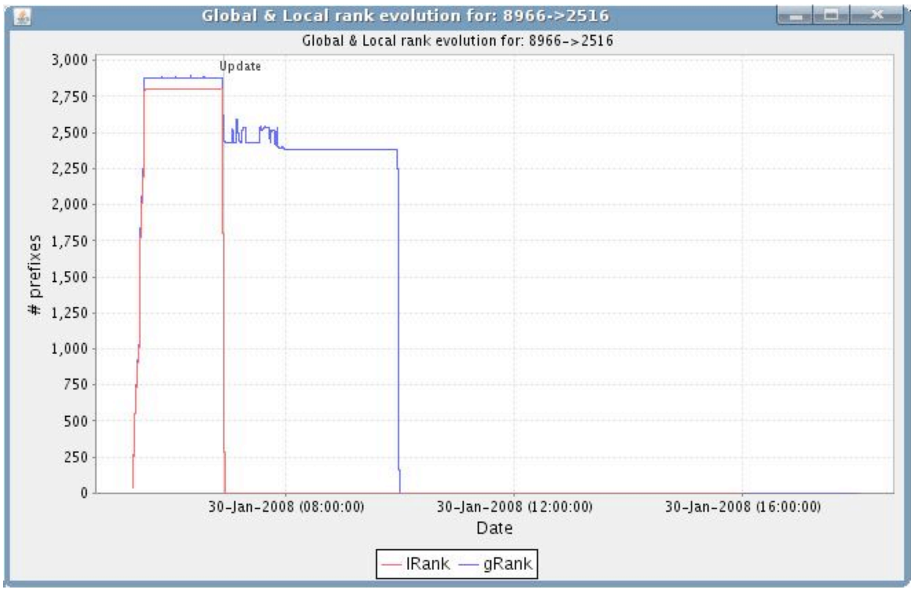


Figure 10: More than 25 hundreds prefixes were suddenly routed through the peering AS8966-AS2516 (Emirates Telecommunications Corporation - KDDI). This continued for a couple of hours.

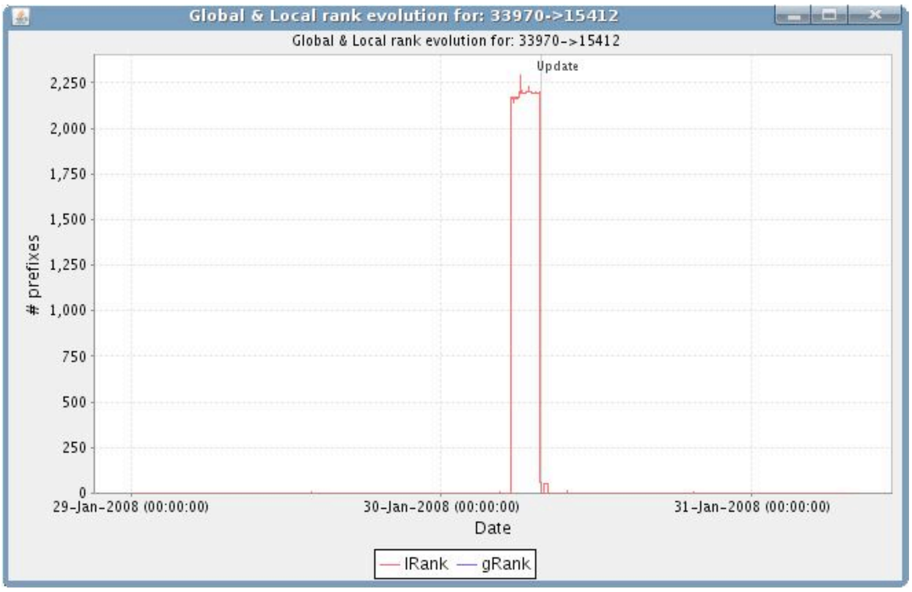


Figure 11: AS33970 (OpenHosting) temporarily routed a large set of prefixes through FLAG Telecom (AS15412). It took more than two hours to return to the previous usage level.

Failing links As expected, some of the peerings experienced a drop in the number of prefixes routed right after the faults.

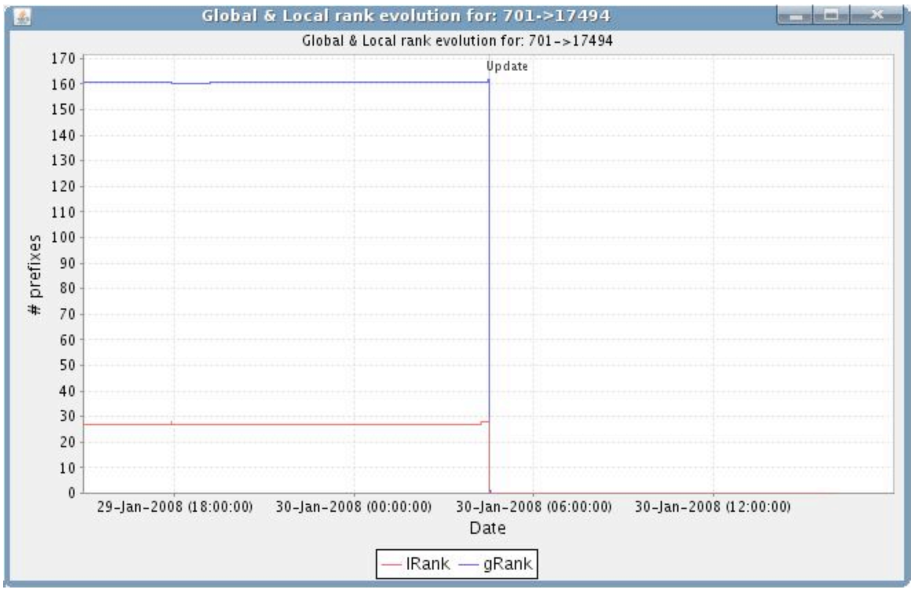


Figure 12: The peering between AS701 (UUNet) and AS17494 (Bangladesh Telegraph and Telephone Board) dropped by over 100 prefixes.

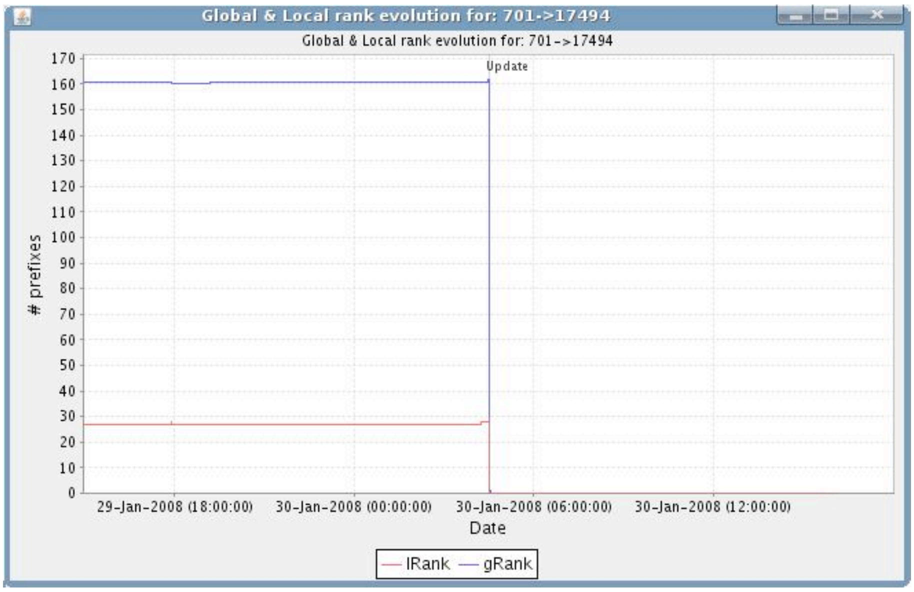


Figure 13: The AS701-AS17494 peering again, observed from a different collector peer. The local rank differs because the two peers see different numbers of prefixes with AS701 in their AS path. Obviously, the global rank evolution is the same as in Figure 12 (but note the slightly different time range on the x-axis).

5.1.4 Analysis of BGP dynamics (Case studies)

To better understand the routing dynamics caused by the fibre cuts, we analysed some specific cases in detail, as samples of different patterns in routing changes.

The fibre cuts meant that some prefixes were unreachable for a significant period of time. Section 7.1 shows that the link between AS20484 (Yalla Online, Egypt) and AS8452 (TEDATA, Egypt) experienced a major network event; the prefixes usually routed through this link either were unreachable or changed their routes for most of the period.

On the other hand, some prefixes did not change their routes. Section 7.2 demonstrates that the network hosting TTM box 138 (in Bahrain) was rerouted at sub-IP level. Thus its BGP routing didn't change, while the traffic experienced some major delays and packet loss; evidently, the backup bandwidth is much less than the original path over SEA-ME-WE4.

Other prefixes changed their routes for a significant period of time. Section 7.3 illustrates how AS17641 (Infotech, Bangladesh) changed one of its upstream providers.

Finally, we observed prefixes which experienced long periods of BGP instability. Section 7.4 shows how OmanTel triggered over 10,000 BGP update messages in RIS in a 90 hour (3.5 day) time period. Routes were flapping constantly for several peers, making it questionable whether BGP path exploration always converged.

Further details on each of these case studies can be found in the appendix.

5.2 Active Measurements

5.2.1 Test Traffic

Looking at statistics and plots published on the Test Traffic Measurement (TTM) website, we learned that from the approximately 75 active probes, only one had serious trouble as a result of the cable outages: node TT138, installed with 2connect in Bahrain, became unreachable from all other nodes at 04:30 (UTC) on 30 January.

After two and a half days, basic Internet connectivity was restored. However, the latencies and packet loss, especially for traffic going to Bahrain, were much higher. With peaks in one-way delay of 1.2 full seconds on 6 and 7 February, the network's performance would have been rather poor for end users. Finally, in the evening of 8 February, when repairs to the SEA-ME-WE4 cable had been completed, latencies returned to normal conditions.

Figures 14 and 15 show packet delay (black dots) and number of hops in traceroute (red lines and dots) to and from the RIPE NCC over time. Similar patterns were seen in plots to and from other TTM probes.

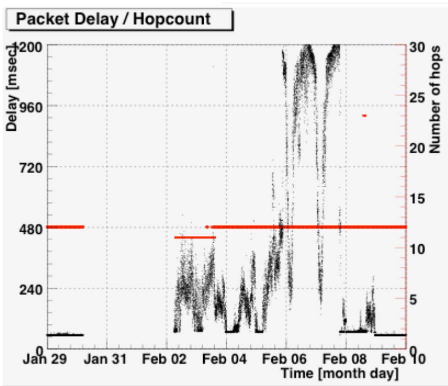


Figure 14: Delay from tt01 to tt138.

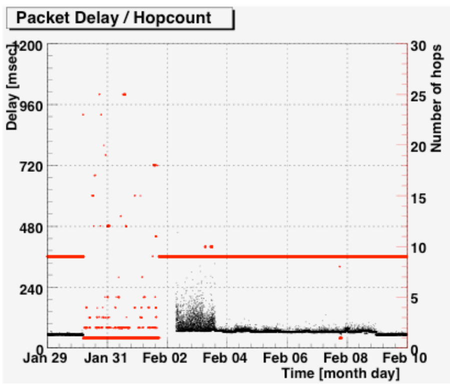


Figure 15: Delay from tt138 to tt01.

As illustrated in Section 7.2, the BGP routing information was relatively stable during this entire period. Even after the site became unreachable, BGP continued propagating routing information for 2Connect's prefix. We see this reflected in the tracreoutes conducted by TTM: right after the cable cut, traceroutes from TT01 did start tracing the route to TT138, but replies stopped coming in after the eighth hop, in AS6453.

Traceroute results from Amsterdam to Bahrain on January 30th:

hop	IP	origin AS
1	193.0.0.238	3333
2	195.69.144.110	1200/31283/30132/12989
3	4.68.120.15	3356
4	4.68.110.226	3356
5	80.231.80.5	6453
6	80.231.80.30	6453
7	195.219.195.6	6453
8	195.219.189.106	6453
9	80.88.244.33	35313
10	no response	
11	80.88.240.121	35313
12	80.88.240.14	35313

Table 1: Before the cable cut.

hop	IP	origin AS
1	193.0.0.238	3333
2	195.69.144.110	1200/31283/30132/12989
3	4.68.120.15	3356
4	4.68.110.226	3356
5	80.231.80.5	6453
6	80.231.80.30	6453
7	195.219.195.6	6453
8	195.219.189.106	6453
9	no response	
10	no response	
11	no response	
12	no response	
...		
30	no response	

Table 2: After the cable cut.

Before the cut, traceroute entered the destination AS at hop 9. The return-trip times (RTT) returned in manual tracreoutes to TT138 suggest this hop is located in Western Europe. Interestingly, a traceroute to the hop 9 address, 80.88.244.33, returns with last hop IP 195.219.189.62. Because this address is from a network registered in RIPE Database as LONDON-TGB/Teleglobe's backbone, we conclude this network node is a Teleglobe router in London configured with IP addresses from Bahrain-based 2Connect. Teleglobe are also likely to be the ones injecting the 80.88.240.0/20 prefix into BGP on behalf of AS35313. This is done irrespective of the state of the link(s) to Bahrain, presumably using statically configured routes.

When basic end-to-end connectivity had been restored on 2 February, traceroutes to Bahrain showed a slightly different path in the final three hops. However, traceroutes from Bahrain don't show any change. We conclude that there was a temporary recovery from the SEA-ME-WE4 cut, created by Teleglobe setting up a different data link for IP traffic to Bahrain, and possibly augmented by a minor internal IP level change in Bahrain.

5.2.2 DNSMON

Because the DNSMON service uses the TTM infrastructure, problems with the Bahrain node are also showed in Figure 16.

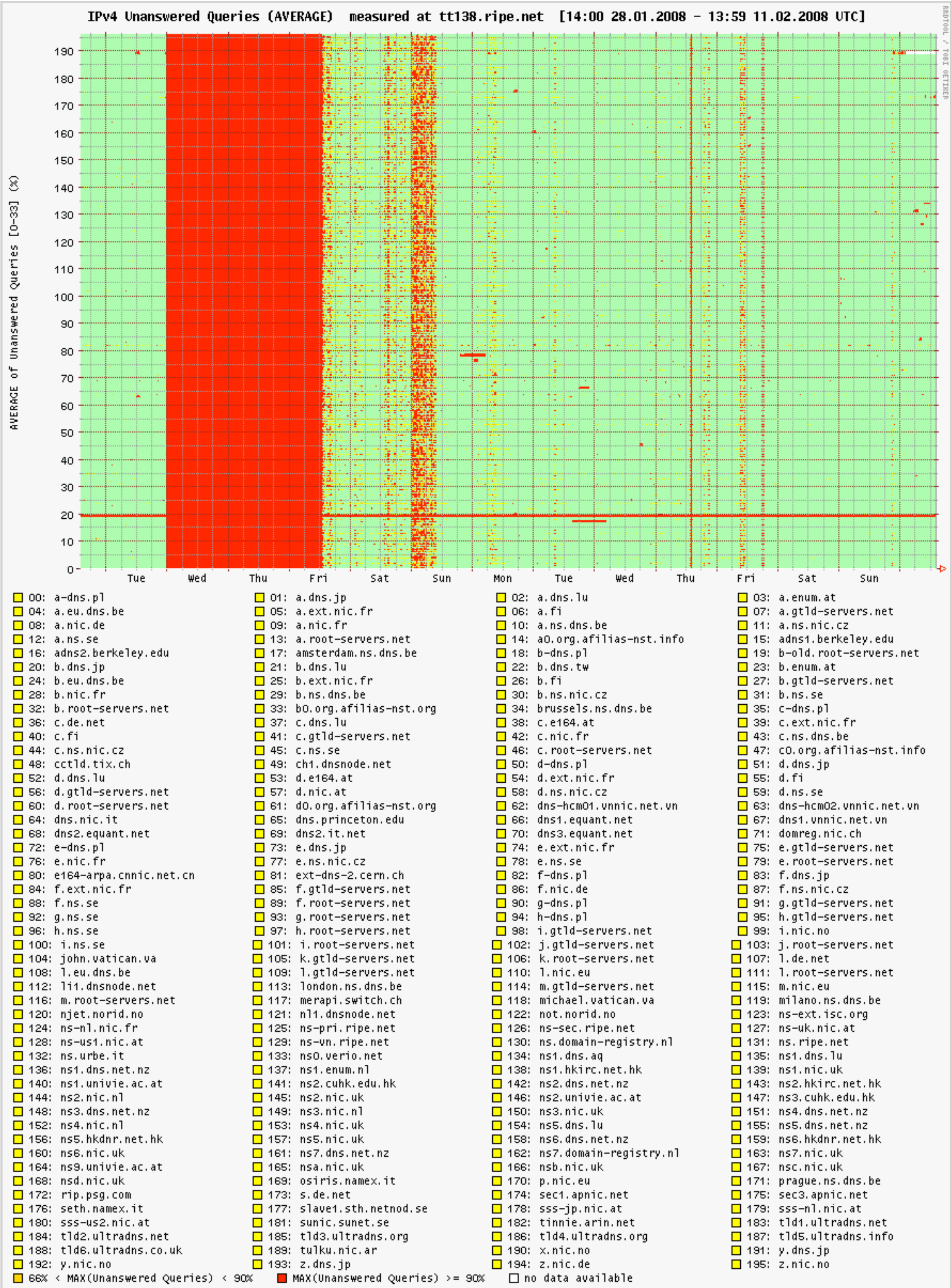


Figure 16: DNSMON graph for TT138. Each horizontal line corresponds to one monitored name server. Red dots mark unanswered queries.

The plot shows that none of the 196 monitored name servers answered queries from TT138 in the period for which TTM already reported problems. This is further evidence that, despite the routing information carried in BGP, the ISP was actually cut off from most of the Internet for two and a half days.

At first glance the list of monitored name servers and TLDs in the graph suggests DNSMON had no measurement targets in the Middle East or South Asia. However, this is not completely true. Some of the name servers are implemented with anycast, with a number of servers connected at different locations worldwide, all using the same global IP address. Which of these is actually used by a client depends on BGP policies, which affect how the name server's network is announced and received at various places.

The K-root name server, operated by RIPE NCC, is one example of an anycast server. Three instances are deployed in the region affected by the cable cuts: two of these, with Qtel in Doha (Qatar) and with Emix in Abu Dhabi (United Arab Emirates), are intended as local nodes; routing announcements for these K-root sites should not be propagated by the local peers to the global Internet. The third one in Delhi, India, is a global node; this means the peers can pass the announcements to their upstreams, thereby making the node an option to choose from for anyone who receives the route to Delhi.

In normal reporting DNSMON does not show which anycast instance was used by the probes. However, the raw data also includes queries into the identity of the server. Using this raw data, we looked for correlations between k.root-servers.org instance changes on each DNSMON probe and the known cable outage events. Of the 75 probes, only four show strongly correlated and unexpected correlations: on 30 January at 05:25 UTC, less than 1 hour after the SEA-ME-WE4 cut, the test-boxes hosted by AMS-IX all switched to using the local node hosted by EMIX in Dubai, UAE. The situation lasted for more than seven days, until 20:00 (UTC), 5 February. These observations provide positive confirmation that EMIX was reachable from Amsterdam during the entire cable outage period. However, the highly increased response times for the DNS queries do indicate congestion on the backup links. Also, the unfortunate leaking of the EMIX local node route announcements caused deteriorated service for those who received and preferred that route over a global node's announcement.

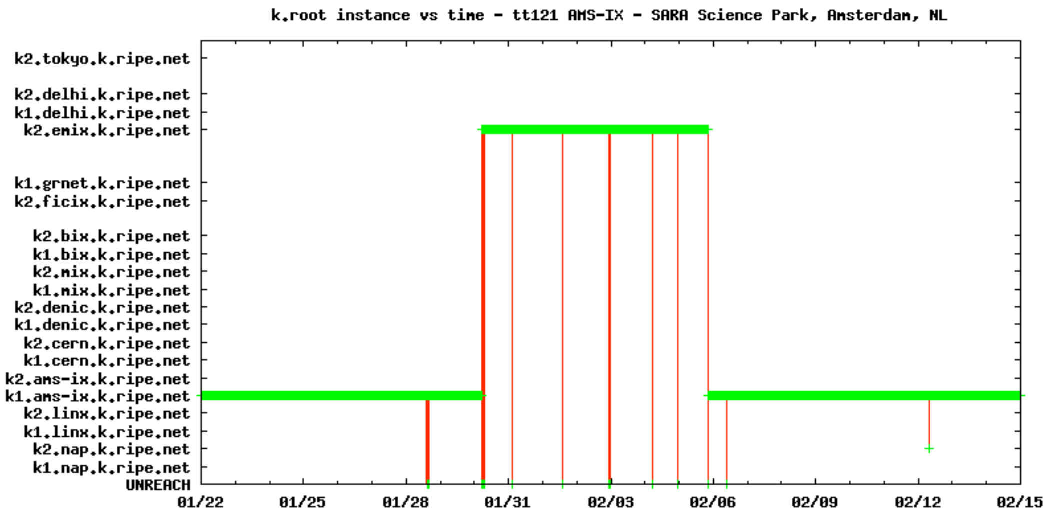


Figure 17: Results from DNSMON probe tt121: k.root instance used in a two-week time interval.

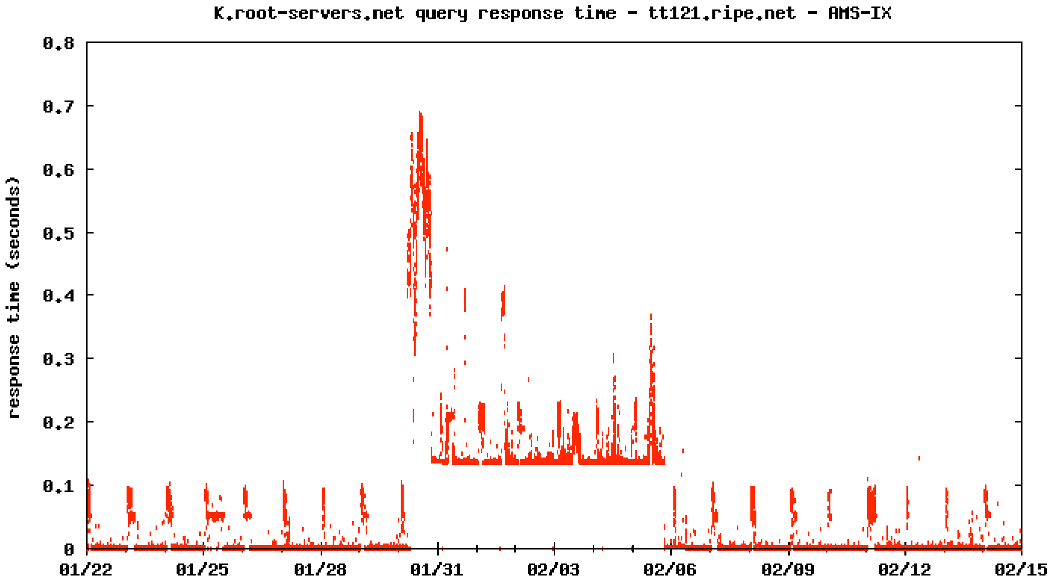


Figure 18: Results from DNSMON probe tt121: k.root query response time for TT121 in the same period.

6 Conclusions

Combining data from different measurement/monitoring systems, our analysis provides insight into how the cable outages affected Internet connectivity:

- Immediately following each cable cut, networks became unreachable, either because routes were withdrawn in BGP or because back haul links went down.
- Sites that had arranged for multiple transit providers observed massive rerouting in BGP, such as moving to satellite providers. Other sites were rerouted on the sub-IP level, moving to circuits set up over other, lower bandwidth or longer distance cable systems. Both types of back-ups experienced increased latencies and congestion, significantly impacting end users and likely causing instability in BGP.

The Mediterranean cable crisis demonstrates the importance of adequately dimensioned redundant connectivity, ideally following different geographical paths.

7 Appendix: Selected BGP case studies

7.1 Case Study 1 - Unreachable Prefixes From BGP Point of View (Egyptian Prefix)

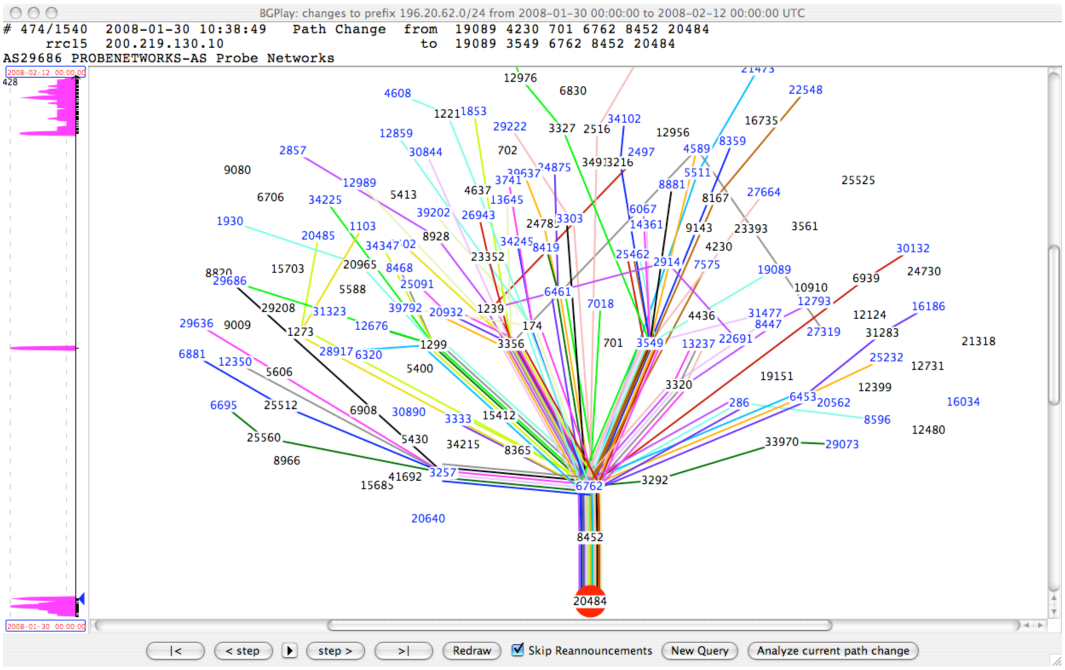
7.1.1 Introduction

Due to the fibre cuts, some prefixes were unreachable for a significant time period. We analysed the effects of the FEA fibre cut on the prefixes originated by AS20484 (Yalla Online, Egypt) (Yalla Online, Egypt).

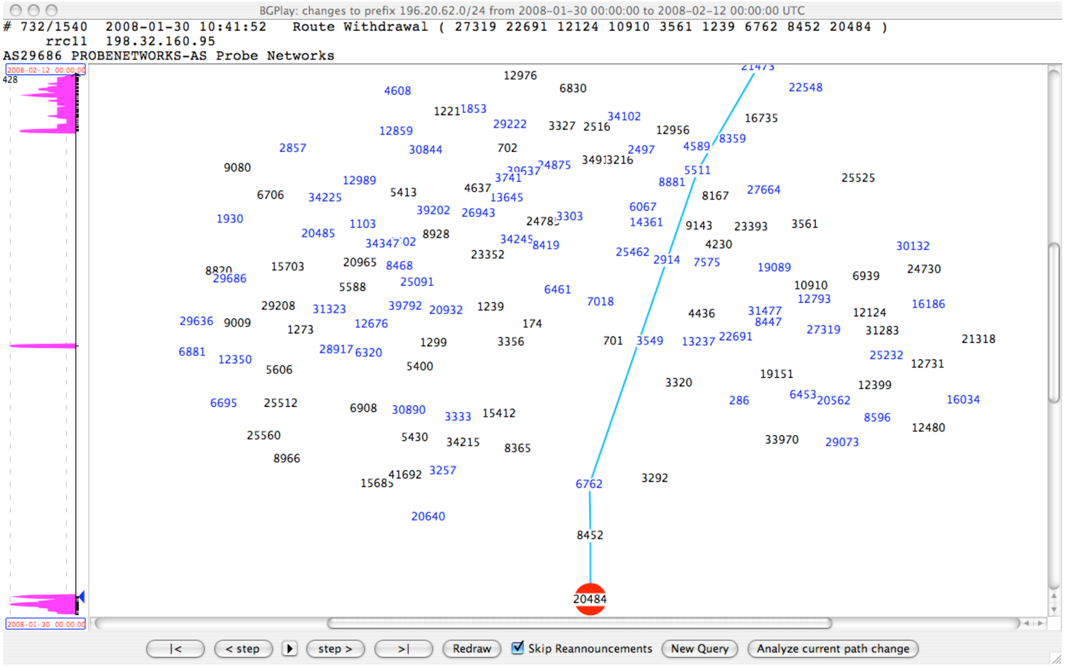
7.1.2 Routing States of a Prefix Originated by AS20484 - BGPlay screenshots

We looked at the routing dynamics of the prefix 196.20.62.0/24 (originated by AS20484), using BGPlay. The following figures show some of the key routing changes the prefix underwent. Overall, the prefix was completely disconnected from the Internet for about 11 days.

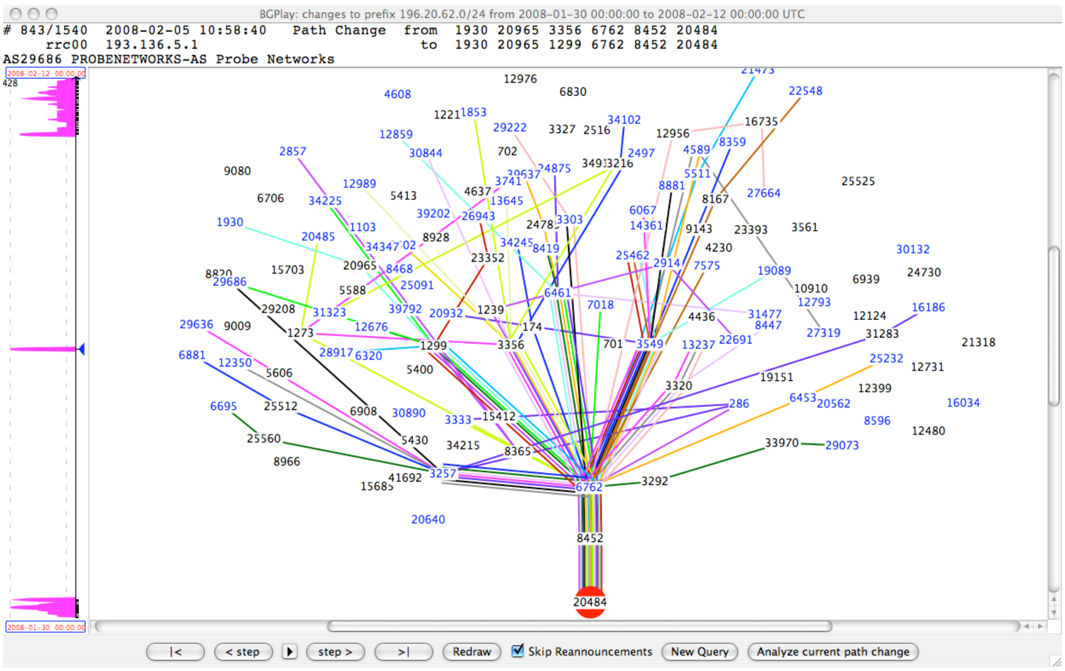
10:38 (UTC), 30 January 2008 Some hours after the fibre cut, the prefix is still reachable by most of the RIS peers.



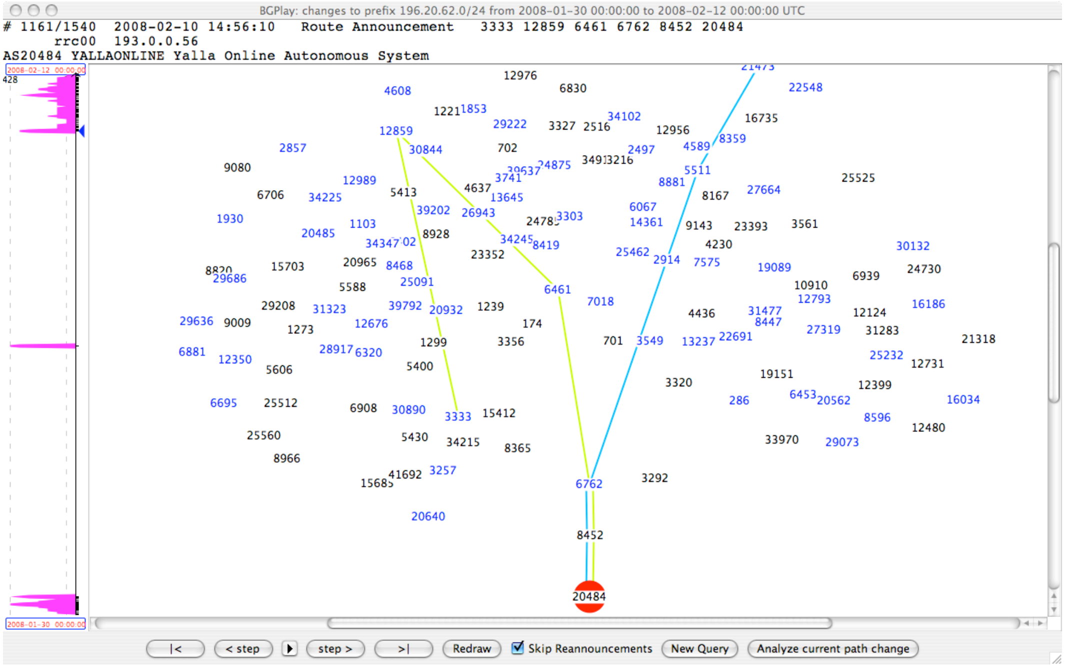
10:41 (UTC), 30 January 2008 In a few minutes the prefix has lost all its paths towards RIS peers. The path still shown connecting AS20848 and AS21475 (ISP Global Ukraine LAN Lviv, Ukraine) is the last to go. It was withdrawn some seconds after the time of this snapshot.



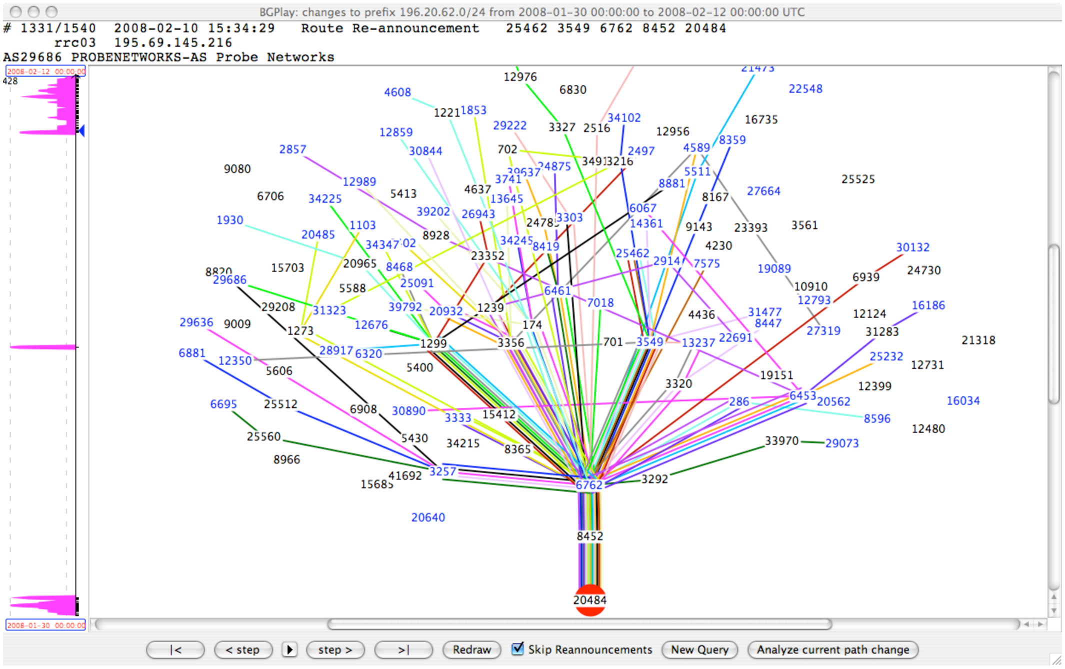
10:58 (UTC), 5 February 2008 For a very short time period, the prefix temporarily regained its connectivity to the Internet.



14:56 (UTC), 10 February 2008 First signs of recovery, route announcements start to arrive at the RIS peers.



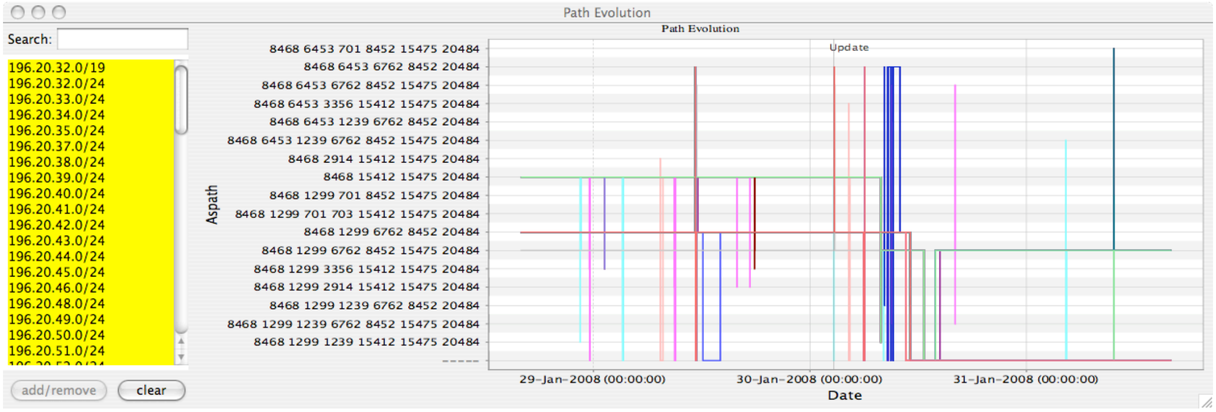
15:34 (UTC), 10 February 2008 All the RIS peers are observing the prefix again.



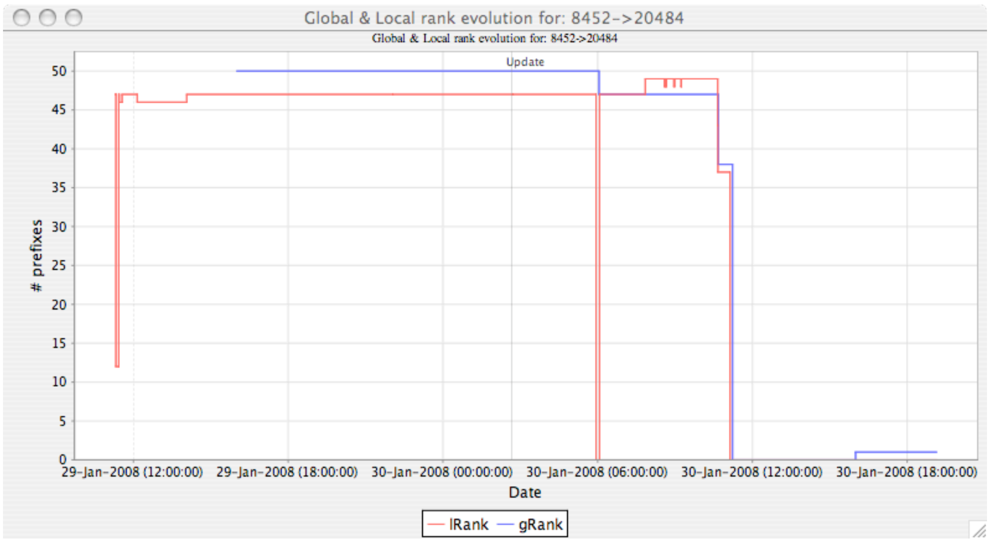
7.1.3 Cross Prefix Analysis - BGPath screenshots

We also looked at the routing dynamics of all the prefixes originated by AS20484, and other prefixes' usage of the AS20484-AS8452 link, using the BGPath tool.

Path Evolution of All the Prefixes Originated by AS20484 All the prefixes originated by AS20484 and using AS8452 (TEDATA) as upstream provider either changed their paths or completely lost their connectivity to the Internet for the whole period.



Evolution of the Number of Prefixes on Specific AS-links The AS-link 20484-8452 usually carries about 50 prefixes. The figure below shows that all these prefixes stopped using the link during the morning of 30 January 2008. Because the timing of that event does not align with the known start times of the cable outages, the link may have been shutdown by human operator intervention.



7.1.4 Conclusions

This analysis shows that from 30 January 2008 to 10 February 2008, the AS-link 20484-8452 experienced some major event. All the prefixes usually passing through this link (originated by AS20484 and other ASes) either were unreachable or changed their routes for most of the period.

7.2 Case Study 2 - BGP Still Carries Routes While Traffic is Black Holed (Bahrain)

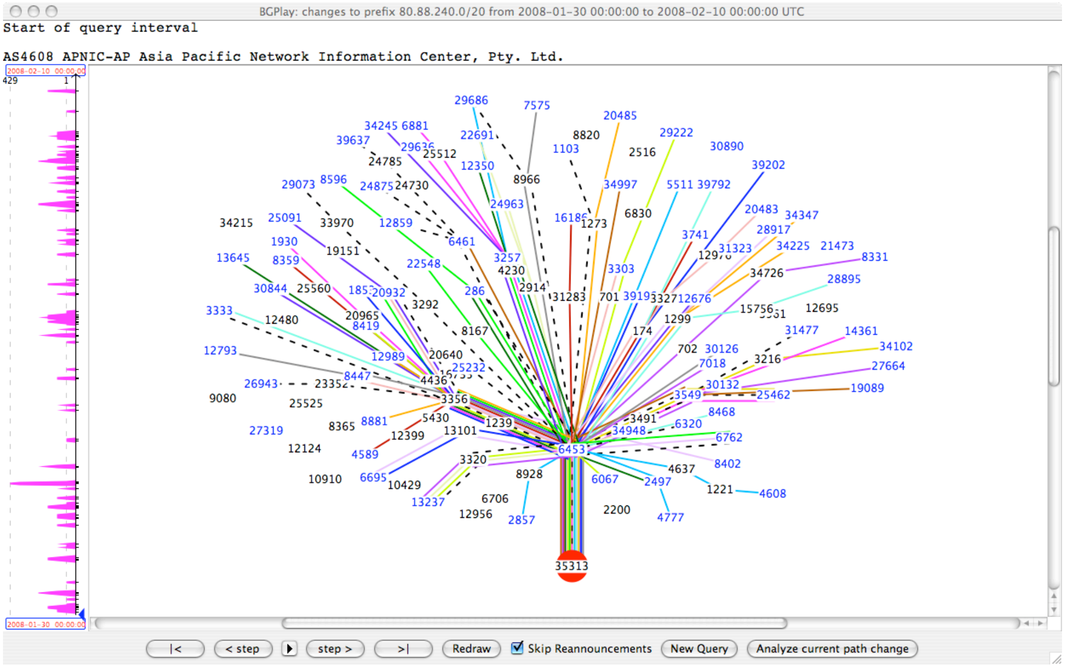
7.2.1 Introduction

As shown in Section 5.2.1, the TTM box in Manama, Bahrain was not reachable by any other TTM box between 30 January and 2 February because of the outage on the SEA-ME-WE-4 cable. If we look into RIS routing tables there is no indication of an outage, instead route announcements were rather stable for the prefixes originated by AS35313 (2Connect), which hosts the test-box.

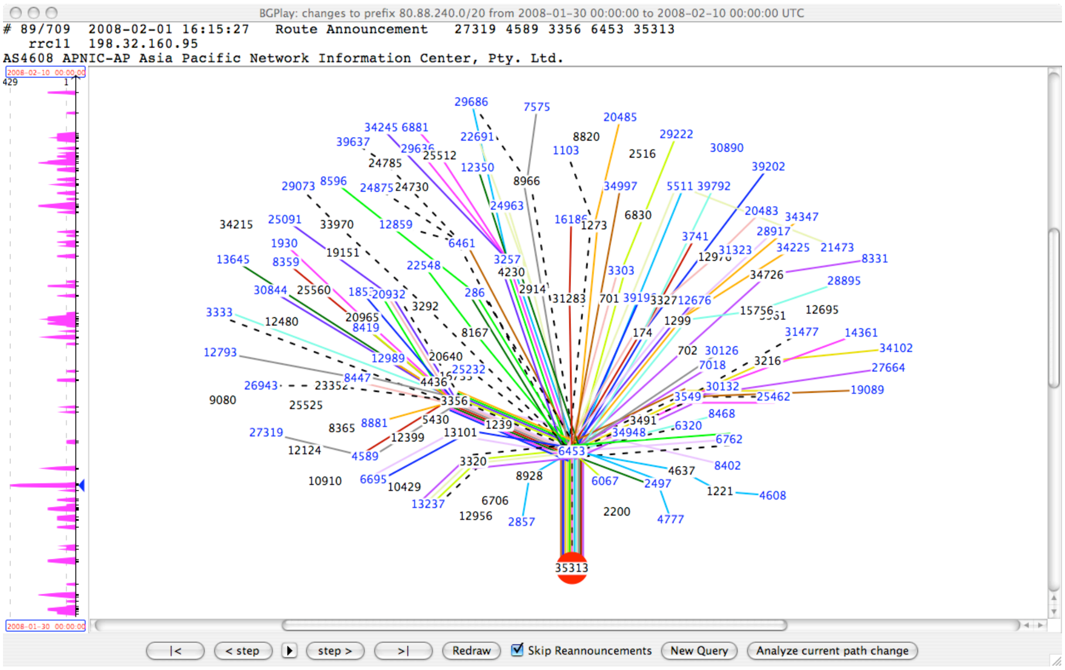
7.2.2 Routing States of a Prefix Originated by AS35313 - BGPlay Screenshots

AS35313 announces two prefixes (80.88.240.0/20, 80.88.244.0/22), and both of them underwent almost the same routing changes during the fibre outage time period. Therefore, we only show some key routing changes of one prefix (80.88.240.0/20) with the following BGPlay screenshots.

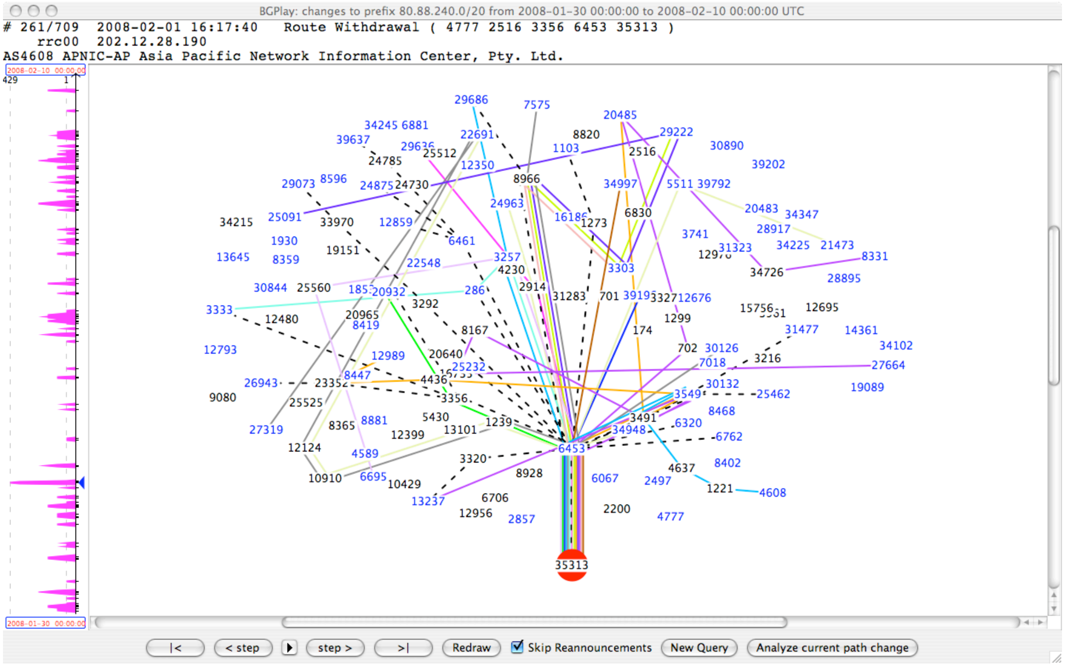
00:00 (UTC), 30 January 2008 Before all the fibre outages.



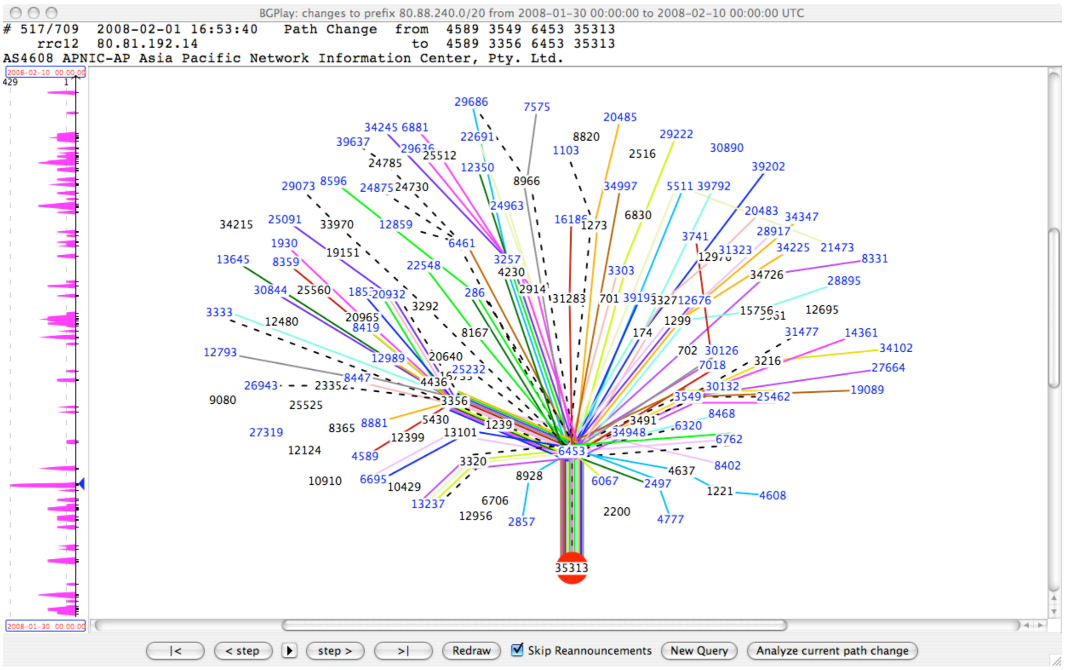
16:15 (UTC), 1 February 2008 More than a day after the last the fibre outage, the paths are almost the same. AS35313 starts losing some paths.



16:17 (UTC), 1 February 2008 In two minutes, AS35313 reaches the most disconnected status during the fibre cut period.

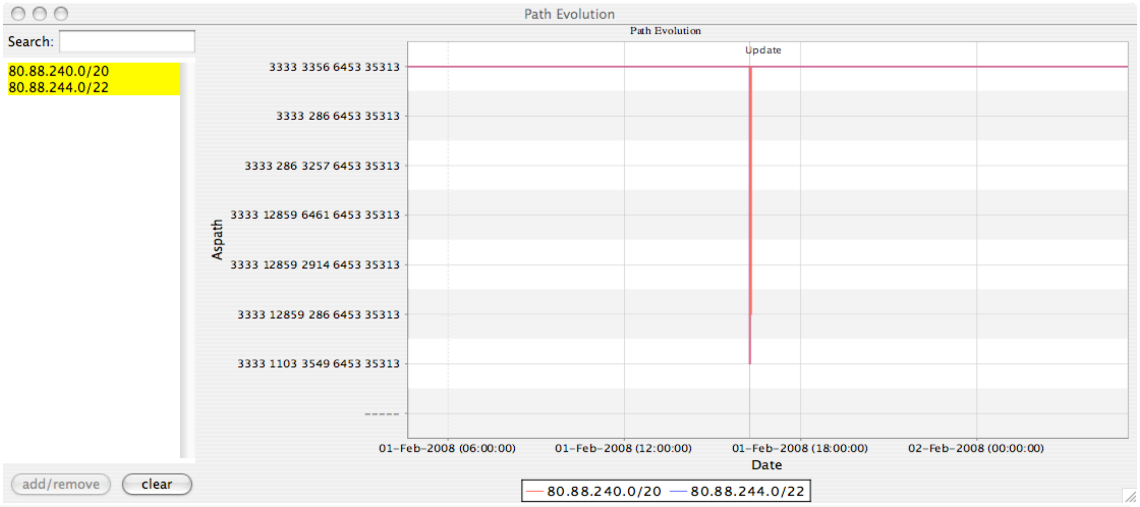


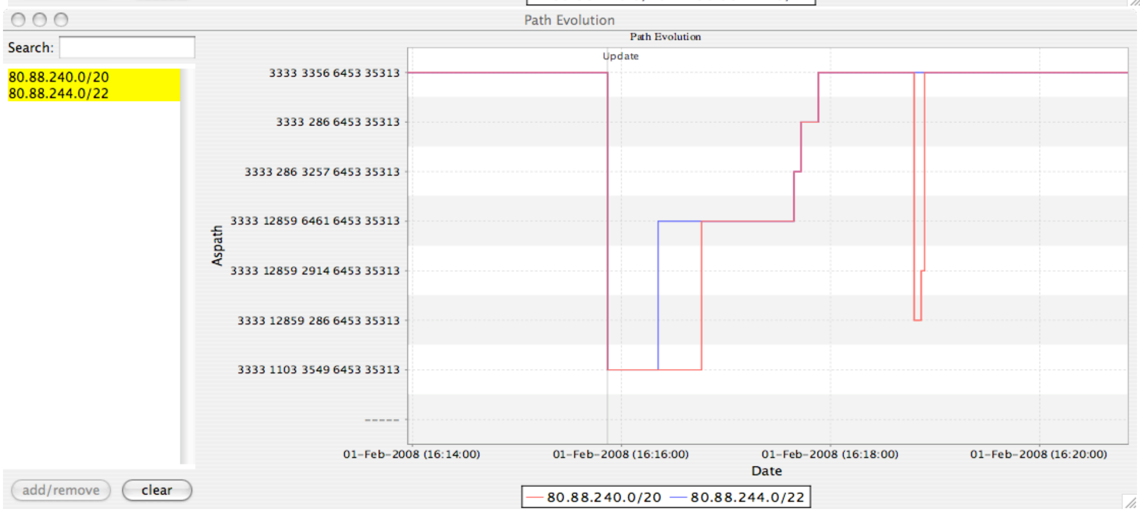
16:53 (UTC), 1 February 2008 About 30 minutes later, practically all of the paths have been recovered. Most likely this short-lived event was caused by operator intervention (a manual router reset or similar event). The results from TTM traceroutes indicate that end-to-end connectivity was restored another 35 minutes later, around 17:29 (UTC).



7.2.3 Path Evolution of All the Prefixes Originated by AS35313 - BGPath Screenshots

From the point of view of peers at RRC00 (RIPE NCC) and RRC03 (AMS-IX), AS35313's prefixes were reachable at all times during the cable outage period. The routes were also stable almost all time, except for about five minutes when both prefixes experienced some path changes. The following BGPath screenshots illustrate this.





7.2.4 Conclusions

The TTM data show that test-box 138 in Bahrain had no connectivity to any other TTM test-box for 2.5 days, starting 04:30 (UTC), 30 January. However during this period, hardly any changes were seen in BGP for the prefixes originated by AS35313 (the site which hosts the box). This shows that the presence of a route in BGP is no guarantee of a working Internet connection.

As explained in Section 5.2.1, we suspect the prefix was (statically) originated by a router located in London, hosted or owned by Teleglobe. Thus the failure of the submarine cable did not trigger a withdrawal of the prefix from BGP routing tables. Later, the Teleglobe-Bahrain traffic usually carried by the SEA-ME-WE4 cable was rerouted via a different submarine cable. Because this had no effect on how the prefix was announced in BGP, the RIS collectors see no changes in ASpath.

7.3 Case Study 3 - BGP Rerouting of Prefixes

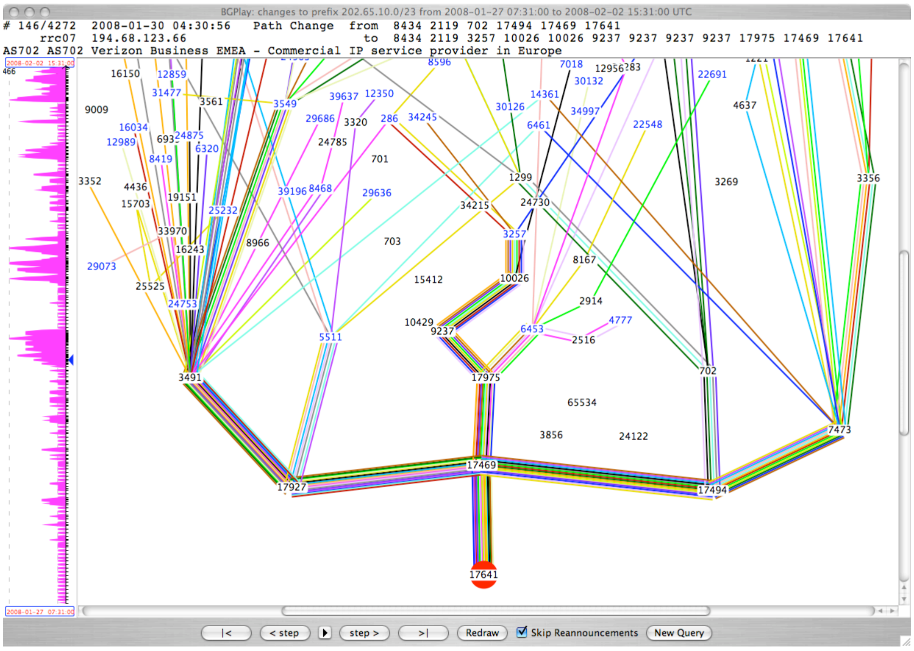
7.3.1 Introduction

Due to the cable outages, some prefixes changed their routes for a significant time period. We analysed the effects of the SEA-ME-WE4 outage on the prefixes originated by AS17641 (Infotech, Bangladesh).

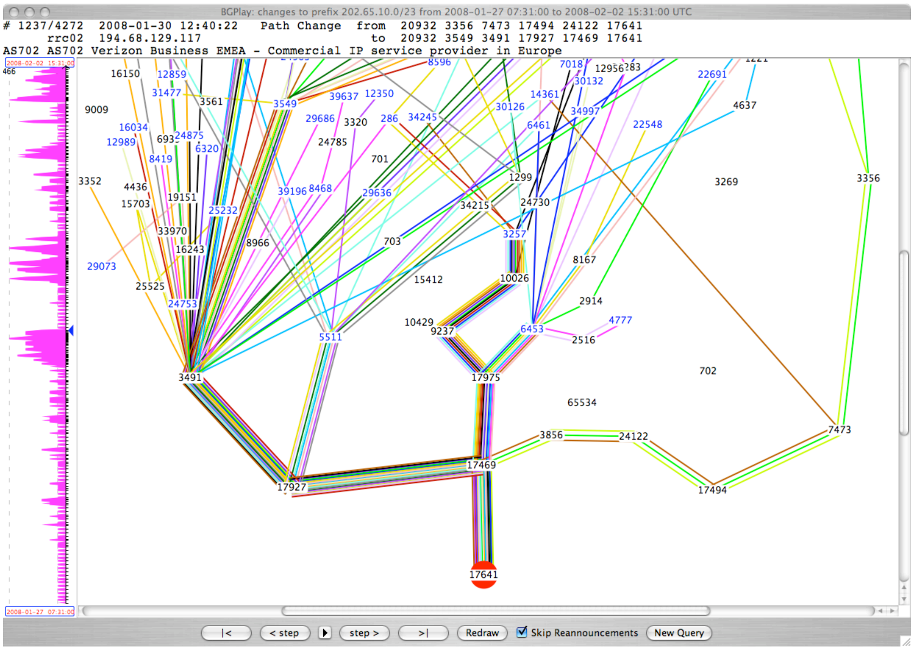
7.3.2 Routing States of a Prefix Originated by AS17641 - BGPlay Screenshots

We looked at the routing dynamics of the prefix 202.65.10.0/23 (originated by AS17641), using BGPlay. The following figures show some of the key routing changes the prefix underwent.

4:30 (UTC), 30 January 2008 Until this moment, not much has happened in terms of BGP messages for this prefix. The purple histogram at the left of the graph shows we are at the start of a period with high activity, triggered by one of the cable faults.



12:40 (UTC), 30 January 2008 Five hours after the event, BGP temporarily stabilises. Note that AS702 (UUNET Europe) is no longer seen in any route to AS17641. Also, only a few routes remain using the links through AS17494 (Bangladesh Telegraph and Telephone Board). All others switched to using AS17927 (WEBSATMEDIA PTE LTD, Satellite Over IP, Singapore) and AS17975 (APT Telecom Services Ltd., Hong Kong) for transit.

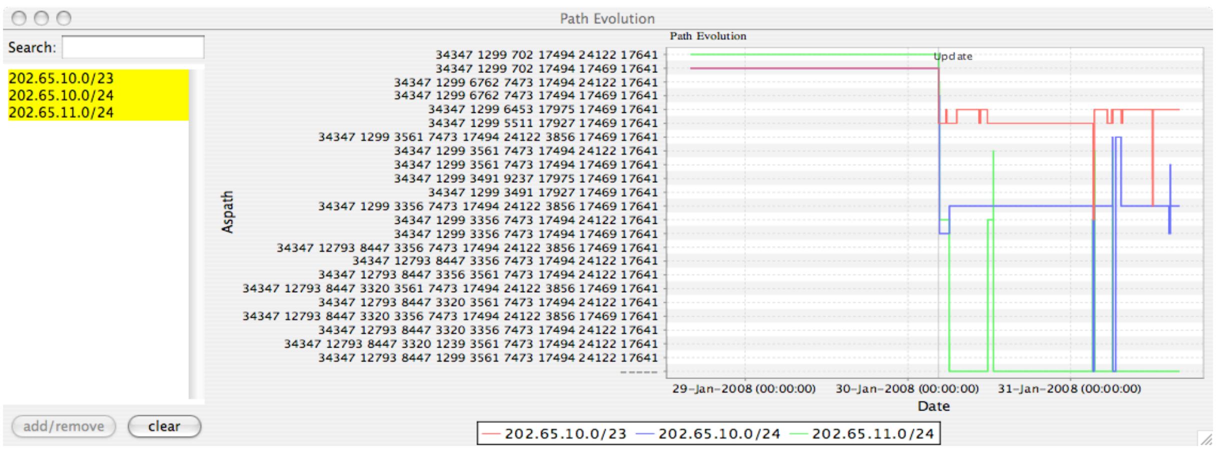


7.3.3 Cross Prefix Analysis - BGPath Screenshots

We also used the BGPath tool to look at the routing dynamics of all prefixes originated by AS17641 as well as the dynamics of some inter-AS links seen in the AS paths ending at AS17641.

Path Evolution of All the Prefixes Originated by AS17641 AS17641 announces three prefixes.

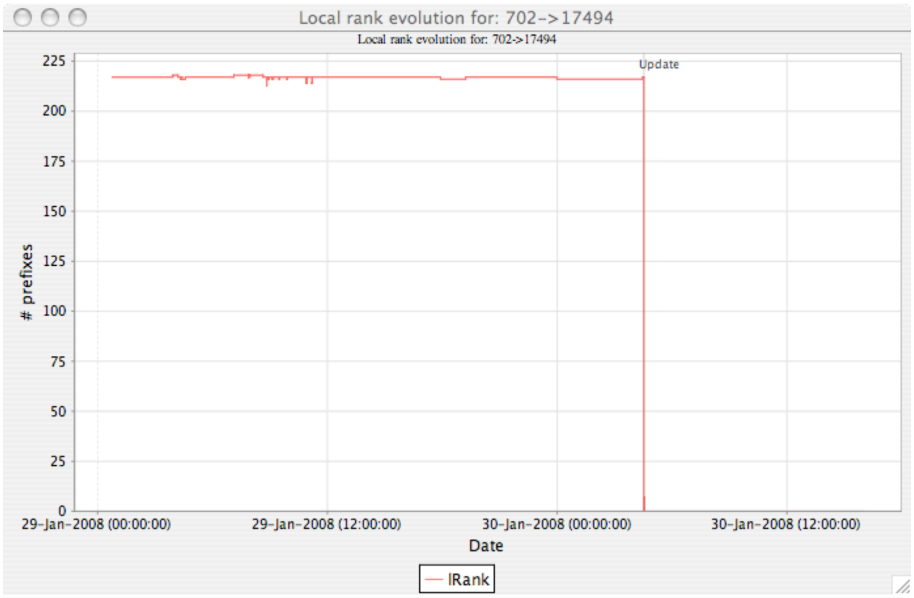
The prefixes (202.65.10.0/23 and 202.65.10.0/24), whose paths usually pass through AS17494, have been rerouted through AS17927 and AS17975. The prefix 202.65.11.0/24, which passes through AS24122 (BDCOM Online Limited, Bangladesh), completely lost its connectivity.

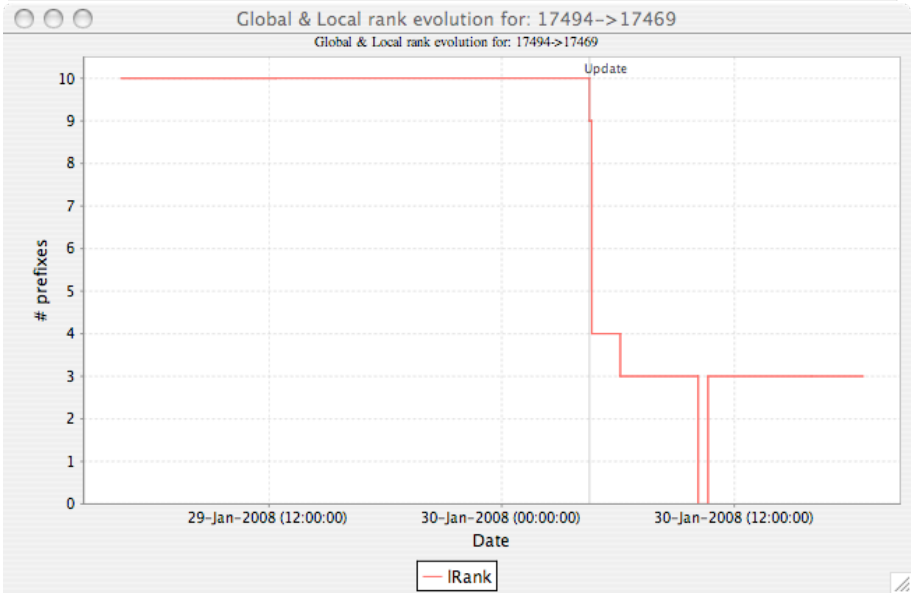


Evolution of the Number of Prefixes on Specific AS Links We analysed how the number of prefixes through AS links "related" to AS17641 (that is, adjacent to AS17641's upstream providers and announcing AS 17641's prefixes) changed.

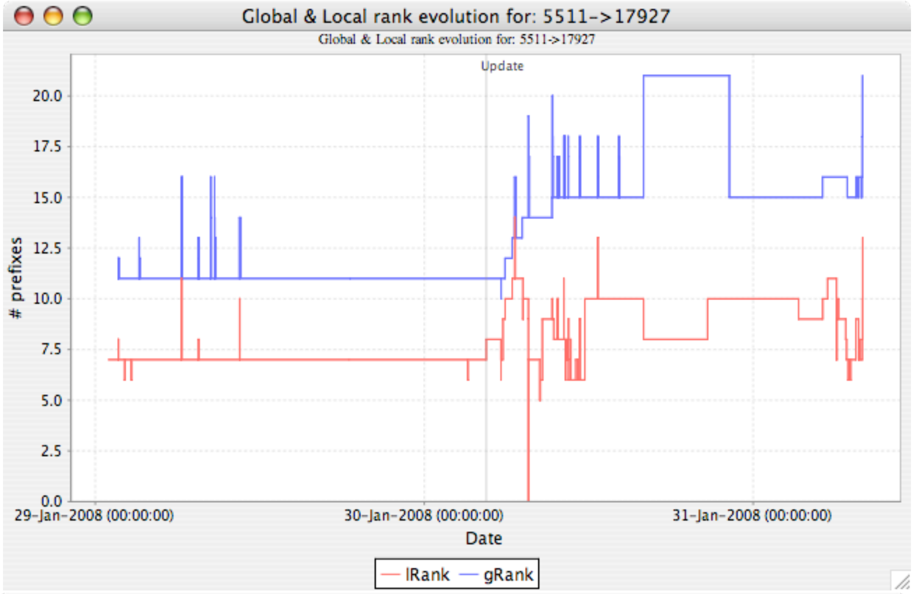
30 January 2008 Fault

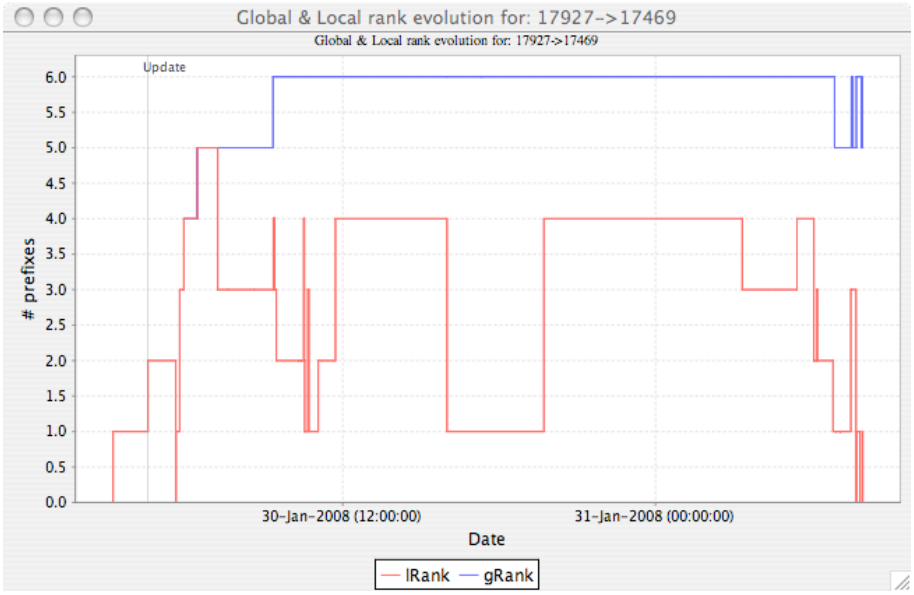
The BGPlay plots showed how AS17641 stopped using paths through 702-17494-17469 after the cable cut. At the same time, many other prefixes stopped being routed over these paths too. The following figures show how the number of prefixes announced by AS17641 and others through the AS links 702-17494 and 17494-17469 suddenly dropped. In particular, the AS link 702-17494, usually used by more than 200 prefixes, is not used by any of them for a significant time period. This is evidence of a major network event, something that affects the connectivity between AS702 and AS17494.



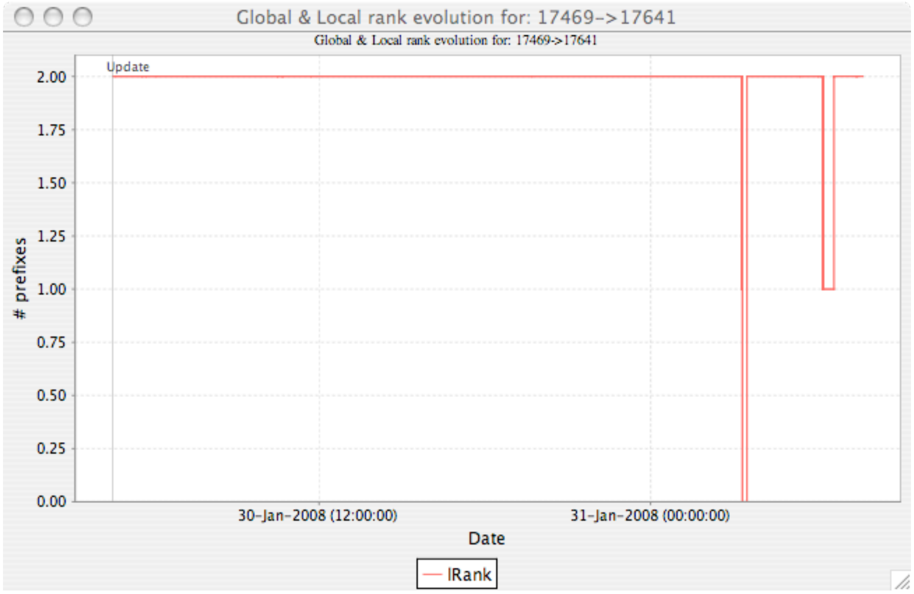


The BGPlay plots also showed that, as a consequence, AS17641's prefixes moved from 702-17494-17469 to 5511-17927-17469. Other prefixes also changed paths, preferring these two Autonomous Systems after the fault. As illustrated in the figures below, the number of prefixes announced by AS17641 and others through the AS links 5511-17927 and 17927-17469 increased.



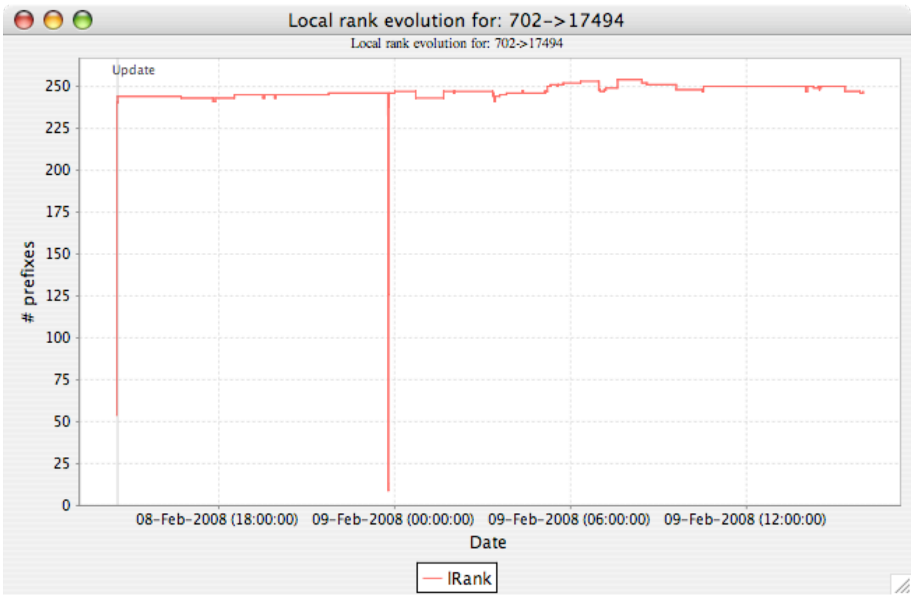


Because almost all the prefixes affected by the event changed their paths without losing their connectivity, almost no change was observed on the AS link 17469-17641.



8 February 2008 Recovery

When the SEA-ME-WE4 cable was repaired and the original routing state recovered, we observed opposite changes in the number of prefixes on the AS links. As shown in the figure below, the link 702-17494 sees a sudden increase to it's old level of about 250 routed prefixes in the afternoon of 8 February 2008.



7.3.4 Conclusions

This analysis shows that from 30 January 2008 to 8 February 2008, the AS-link 702-17494 experienced some major event, and all the prefixes (whether originated by AS17494 or not) usually passing through this link were rerouted through different links. The timing of this event, 04:30 (UTC), aligns well with the recorded outages of the TTM box in Bahrain and other events in BGP. Because the FEA cable went down at 08:00, we conclude the AS link 702-17494 is set up over the SEA-ME-WE4 cable.

7.4 Case Study 4 - OmanTel: Explosion in AS Path Count, Hours of BGP Churn

The AS path changes graph for Oman caught our attention as being quite different from the general pattern: instead of going down, because of loss of connectivity, the number of distinct AS paths, as observed from RRC03 peers in the eight hourly RIB dumps, went up. The average AS path length also increased for the duration of the cable outages.

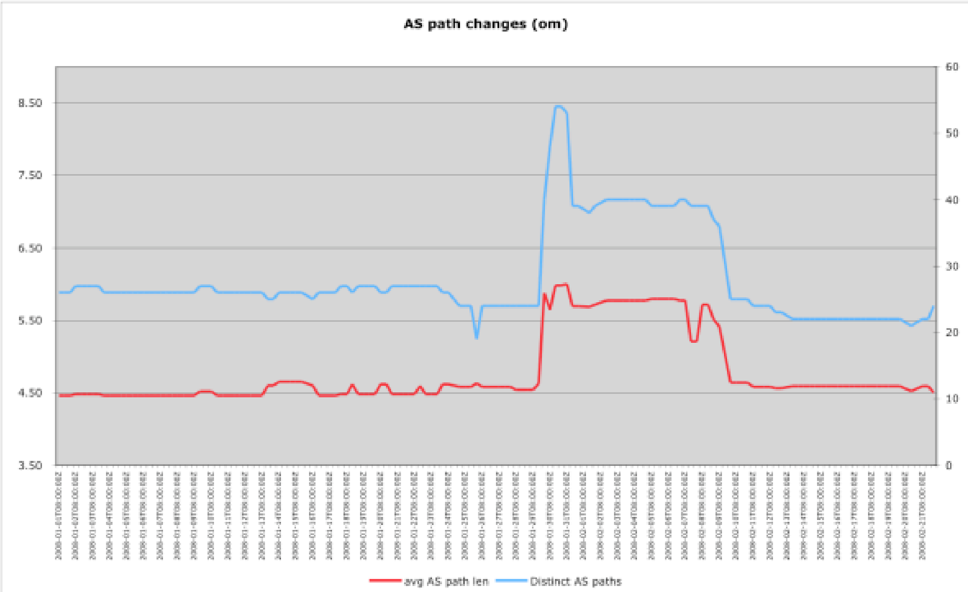


Figure 19: Number of AS Paths And Average Path Length For Oman.

The RIR stats files show only one AS assigned directly to a provider in Oman: AS28885 - OmanTel NAP. In January 2008, RIS saw 26 prefixes originated by this AS. Looking at the raw data, we noticed these prefixes are usually announced to the RIS peers in batches. One BGP update message carries the bulk of OmanTel prefixes, one or two other updates carry the rest. From RIS data alone is not possible to deduce with 100% certainty the reasons behind the observed behaviour; however, we imagine specific routing policies related to the networks served by the prefixes could play a role.

In relation to the cable outages we see the following:

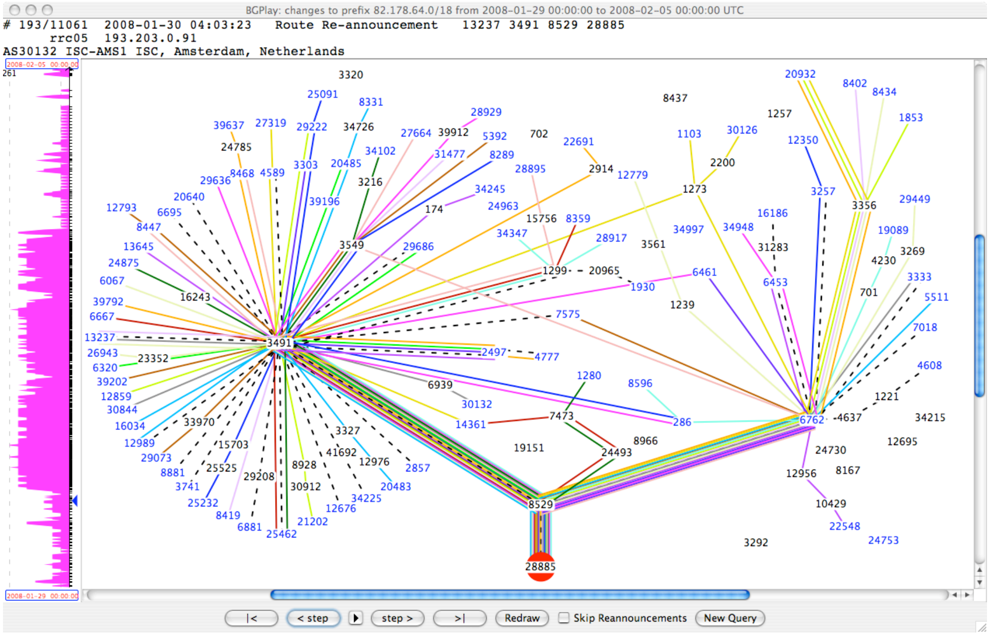
Before the cable fault Each collector peer received the same AS path in all update messages for AS28885. This indicates the same routing policy for all prefixes.

During the cable outage The origin AS apparently used a different policy for different sets of prefixes. So the collector peers receive different AS paths on the update messages for the full set of Oman prefixes.

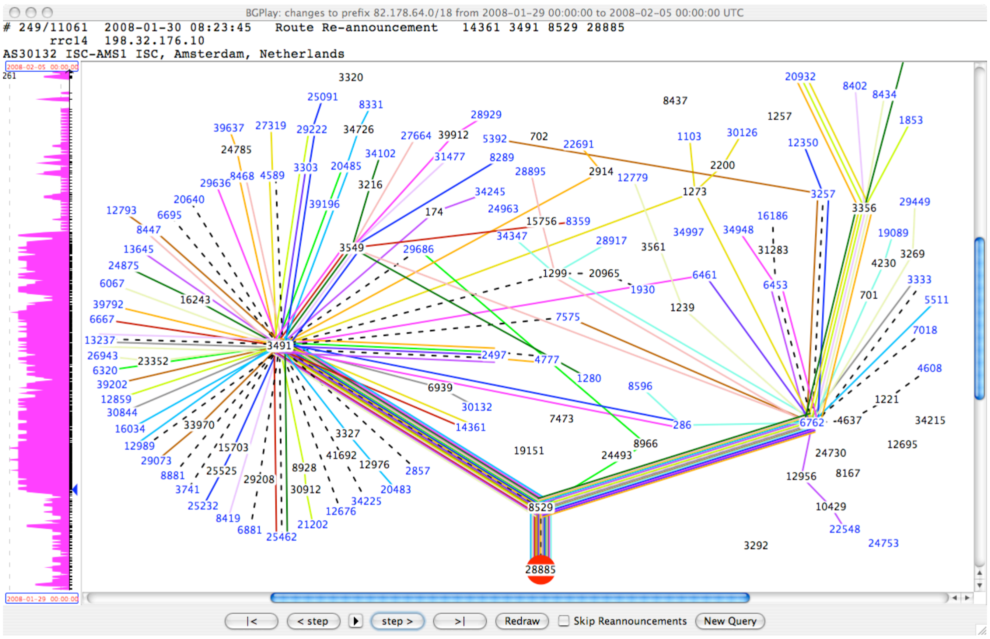
7.4.1 Routing States of a Prefix Originated by AS28885 - BGPlay Screenshots

We looked at the routing dynamics of the prefix 82.178.64.0/18 (originated by AS28885), using BGPlay.

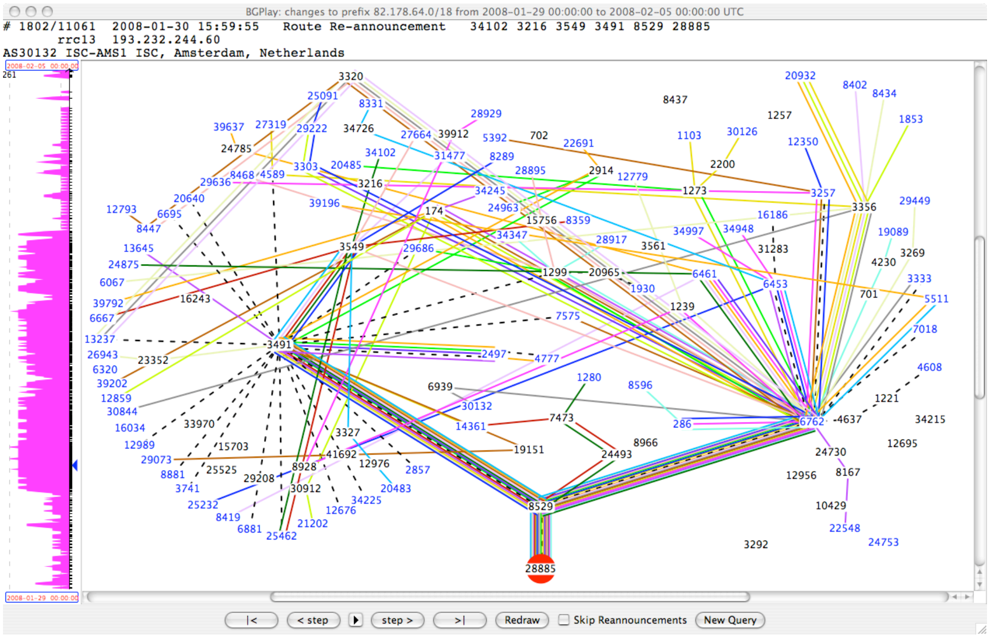
04:03 (UTC), 30 January 2008 Before the cable outages; primary transits for Oman are AS3491 (PCCW) and AS6762 (Telecom Italia Sparkle). Note how the purple histogram on the left indicates a period of prolonged, continuous BGP updates is about to begin. Over 10,000 messages were recorded in 90 hours, which means that on average the collective of all RIS peers saw an announcement or withdrawal every 30 seconds.



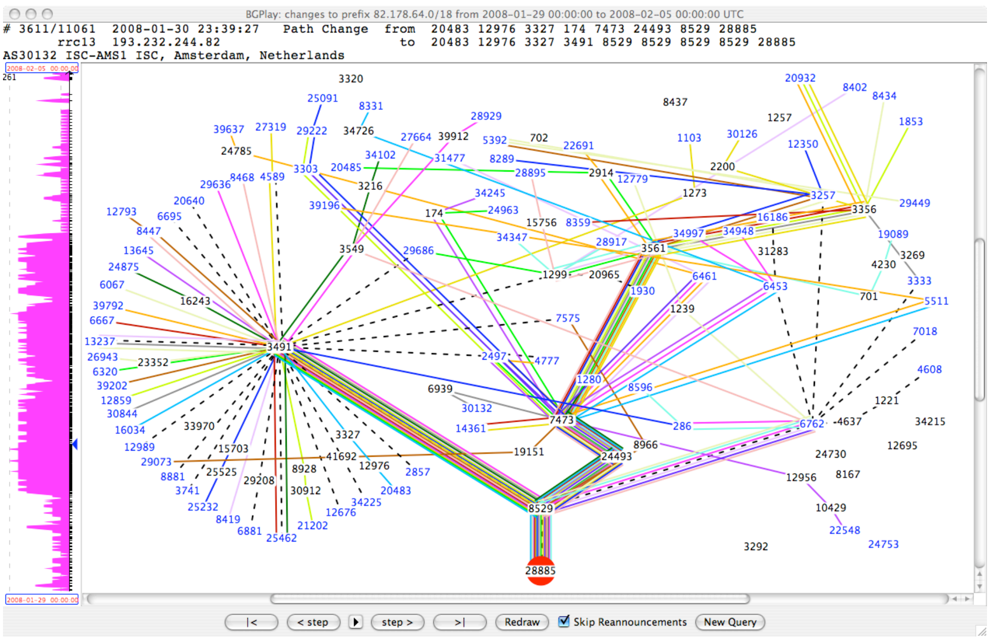
08:23 (UTC), 30 January 2008 Shortly after the second cable went down; first signs of rerouting are visible.



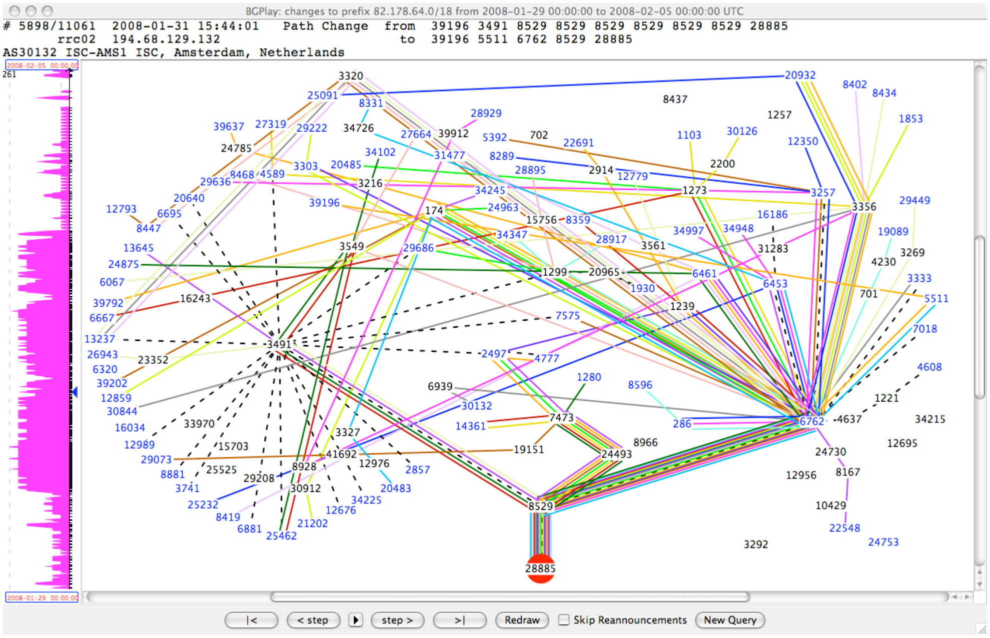
15:59 (UTC), 30 January 2008 Immediately before the second RIB dump of the day. Many peers have switched from using AS3491 to AS6762 as transit to Oman. From the graph we can see most of these peers need more AS hops to reach AS28885, thus the average AS path length increases.



23:39 (UTC), 30 January 2008 Before the last RIB dump of the day. AS24493 (STIXLITE Transit Service Provider Singapore) has taken over the transit for most peers who first used AS6762 (Telecom Italia Sparkle).



15:44 (UTC), 31 January 2008 Yet another routing state. AS6762 is used by more peers than ever, AS24493 (Singapore) is still strong and AS3491 (PCCW) is the least preferred transit provider.



7.4.2 Conclusions

The case of OmanTel shows how a combination of (likely) routing policy and an explosion in BGP activity increase the routing topology entropy for AS28885. The number of observed distinct AS paths for the prefixes announced from Oman doubled and the average AS path length increased by 20%. Because there was a constant high rate of changes, we ca not be sure if BGP ever converged in that period, or if the routes which were seen could really be used.

References

[1] RIS BGPlay. `http://www.ris.ripe.net/bgplay/`.

[2] ROMA TRE Compunet Research Group. `http://www.dia.uniroma3.it/~compunet/`.

[3] Mediterranean, Red Sea and Black Sea Region. `http://iscpc.org/cabledb/Mediterranean_and_Red_Sea_Cable_db.htm`.

[4] Submarine Cable Improvement Group: Worldwide Trends in Submarine Cable System faults. `http://www.scig.net/Section11a.pdf`.

[5] RIPE NCC. `http://www.ripe.net`.

[6] Mediterranean Fibre Cable Cut - a RIPE NCC Analysis. `http://www.ripe.net/projects/reports/2008cable-cut/index.html`.

[7] Routing Information Service. `http://www.ripe.net/ris/`.

[8] RIR delegation statistics. `ftp://ftp.ripe.net/pub/stats/ripencc/`.