enisa
European Network and Information Security Agency

Good Practices in

Resilient Internet Interconnection

[June 2012]

About ENISA

The European Network and Information Security Agency (ENISA) is a centre of network and information security expertise for the EU, its member states, the private sector and Europe's citizens. ENISA works with these groups to develop advice and recommendations on good practice in information security. It assists EU member states in implementing relevant EU legislation and works to improve the resilience of Europe's critical information infrastructure and networks. ENISA seeks to enhance existing expertise in EU member states by supporting the development of cross-border communities committed to improving network and information security throughout the EU. More information about ENISA and its work can be found at http://www.enisa.europa.eu.

Authors

The report was co-authored by:

Christian DOERR, **Delft University of Technology – Network Architectures and Services**

Razvan GAVRILA, **ENISA, CIIP and Resilience Unit**

Fernando KUIPERS, **Delft University of Technology – Network Architectures and Services**

Panagiotis TRIMINTZIOS, **ENISA, CIIP and Resilience Unit**

Acknowledgements

Over the course of this study, we have talked extensively with technical and managerial staff at Internet service providers, network operators, Internet exchange points and research labs. Some of our sources requested that we do not explicitly list their contributions. We therefore collectively thank all our contributors, workshop participants, and participating experts for their valuable input.

Contact Information

E-mail: resilience@enisa.europa.eu

Executive Summary

The interconnection system is the most fundamental building block of the Internet, as it enables independent operators to link up and extend their networks, reaching universal global coverage. Important enough to actually coin the term "Internet", it is the prerequisite for the availability of the global Internet, as well as other services such as utilities, transport systems, or financial services that now depend on it.

The 'Internet interconnection ecosystem' holds together all the Autonomous Systems that make up the Internet. The ecosystem is complex and has many interdependent layers. This system of connections between networks occupies a space between and beyond those networks and its operation is governed by their collective self-interest, staffed with technicians who can leap into action when trouble occurs[1].

Given the importance of the Internet and consequently the interconnection system to our society, it is critical to protect it and increase its resilience against disruptions. This report investigates the development and current state of the interconnection system, the economic incentives that will engage market players, modify or terminate individual interconnection links, as well as on-going trends that can be expected to evolve and shape the ecosystem as a whole.

Based on an analysis of past incidents and an investigation of the current state of the art, this report further outlines good practices to create a more resilient Internet interconnection system, and which provide concrete steps that can be taken by Internet Service Providers (ISPs) and network operators to strengthen their infrastructures against accidental failures and malicious attacks.

The good practices that have been identified are:

- Good Practice 1: Deploy link protection schemes
- Good Practice 2: Investigate common dependencies and shared risk groups
- Good Practice 3: Overprovision network elements by a factor of 2
- Good Practice 4: Introduce independent availability regions capable of securing network operation
- Good Practice 5: Instantiate path restoration techniques
- Good Practice 6: On backbones with traffic engineering, use an alternative path method
- Good Practice 7: Multi-home networks with dependency-less connections
- Good Practice 8: Utilize Border Gateway Protocol (BGP) attributes to select entry/exit points
- Good Practice 9: Create resiliency at the interconnection border

[1] *ENISA's Report on the Resilience of the Internet Interconnection Ecosystem*

https://www.enisa.europa.eu/activities/Resilience-and-CIIP/networks-and-services-resilience/inter-x/interx/report

- Good Practice 10: Maintain internal records about the availability and reliability of remote routes and use this information to select between inter-domain routes
- Good Practice 11: Develop business continuity management aligned with the organization's expectations and resources
- Good Practice 12: Prime a traffic management scheme for mixed service infrastructures to prioritize critical traffic during crises situations
- Good Practice 13: Utilize tags, padding and prefix splitting to control outgoing prefix announcements, and BGP filters to sanitize announcements
- Good Practice 14: Register and maintain up-to-date information about an Autonomous System (AS) prefixes within regional Internet registry (RIR)
- Good Practice 15: Foster strong relations with compatible providers to give and receive assistance during crises

In addition to the identified good practices, the following recommendations are made:

- Recommendation 1: Avoid policy making against temporary developments and market culture
- Recommendation 2: Establishing a platform and community to disseminate important information
- Recommendation 3: Stimulate the usage of Route Information Registries
- Recommendation 4: Foster change by creating market pull
- Recommendation 5: Develop an infrastructure to bring together and safely share information on networks, infrastructure and network-related incidents
- Recommendation 6: Define acceptable practices for traffic prioritization
- Recommendation 7: Establish a traffic prioritization standard across multiple operators or administrative zones
- Recommendation 8: Foster and protect engineering community efforts to deal with multi-operator outages
- Recommendation 9: Implementation of inter- and intra-domain good practices
- Recommendation 10: Develop techniques to accurately measure the structure of the Internet
- Recommendation 11: Investigate the structural properties of the Internet in a changing provider ecosystem

Contents

List of Tables and Figures

1 Introduction

Regardless of size and market share, no network – whether it is a local or regional network provider, a national incumbent or a multi-national backbone network – is large enough to provide its customers service to all remote destinations they expect to be able to reach. To allow customers to gain access to all parts outside their own network and in turn be reachable from remote end-points, individual telecom and network operators need to establish connections to those remote networks outside of their current range. By linking their respective networks and thereby gaining mutual access to each other's network infrastructure, such interconnections provide operators with two benefits:

1. They are able to extend their overall reach towards universal, global coverage
2. They can strengthen and diversify their connectivity towards important destinations so that working links remain even after one or more infrastructure failures.

Interconnection is thus one of the most fundamental building blocks of the Internet, a principle so important to actually coin the term Internet – the interconnection of networks – itself. Internet interconnection is the foundation for the availability of Internet service, and with the transition of other products and services on this Internet Protocol (IP) based infrastructure also for the availability of many other services, such as telephony, TV, financial transactions, or supply chain management.

Given the general importance of the Internet and Internet-enabled services to our society, protecting Internet interconnections as the enabling mechanism of this critical infrastructure is key and it is imperative to remove any obstacles that may hamper the resilience of the interconnection system in addition to providing further incentives to strengthen it. This study provides an overview of past incidents that impaired the Internet's interconnection fabric, and discusses good practices to limit or avoid the impact of future crises events.

1.1 Policy Context

This report and its presented findings complement the initiatives on the implementation of Article 13a of the EU Telecom Package[2], by identifying a number of steps to protect the integrity of the Internet as a public communication service.

The study is part of ENISA's overall activity portfolio within the area of Critical Information Infrastructure Protection (CIIP) and Resilience, continuing the previous efforts on interconnection resilience such as the "Resilience of the Interconnection Ecosystem"[3] and "Secure routing: state-of-

[2] *Directive 2002/21/EC of the European Parliament and of the Council, of 7 March 2002, on a common regulatory framework for electronic communications networks and services (Framework Directive), as amended by Directive 2009/140/EC and Regulation 544/2009.*

[3] *http://www.enisa.europa.eu/activities/Resilience-and-CIIP/networks-and-services-resilience/inter-x/interx*

the-art deployment and impact on network resilience"[4] studies and aligning with stocktaking and exercising initiatives such as the European Public-Private Partnership for Resilience (EP3R) and the multi-national and multi-stakeholder cyber exercises.

1.2 Targeted Audience

The work presented in this report is mainly directed at two groups: Internet Service Providers (ISPs) seeking to improve the resiliency of their infrastructure and interconnections and National Regulatory Authorities (NRAs) defining the operating context for the Internet ecosystem. The study and its findings have also wider applicability and are of interest to for example consumer and industry groups, academic and industrial research labs as well as standardization bodies. Some of the recommendations for improving the resiliency of the interconnection system are also directed at these stakeholders.

1.3 Scope: What is a Resilient Internet interconnection System?

Before beginning the discussion on the interconnection system and its resilience, it is first necessary to clarify the scope and goal. Throughout this report, we aim to investigate threats that impact and mechanisms to increase the resilience of general Internet services, i.e., *providing and maintaining the availability of any application running on the Internet's best-effort infrastructure during crises*. IP-based networks serve a variety of different services and customer segments, ranging from specialized business-oriented IP solutions to general Internet connectivity, that are designed with different requirements and therefore react differently to crises and challenge situations.

The admittedly large, ambitious focus on general best effort IP connectivity, instead of solely on the physical Internet interconnections infrastructure, however gives consideration to the large weight that general Internet service has obtained in today's society, visible in the wide-spread usage of applications ranging from email, chat, Voice over IP, to specific web services across all social and demographic classes. With the adoption of these services into everyday life however also comes along the common expectation that these services are also available during crises, similarly as traditional communication means have been in the past.

1.4 Objectives of the Study

The aim of this study is to survey the current state of the art and to identify good practices for resilient Internet interconnections, thereby considering the following five subjects:

- Market and economical forces that drive the interconnection of Autonomous Systems
- Technical good practices that (would) make the interconnection system more resilient

[4] *http://www.enisa.europa.eu/act/res/technologies/tech/routing/state-of-the-art-deployment-and-impact-on-network-resilience/at_download/fullReport*

- The existence and terms of inter-operator mutual aid assistance agreements
- Policy and regulatory issues that regulate interconnection
- Aspects of traffic prioritization

These issues play a role to a varying degree in the connectivity and resiliency of the Internet and on how traffic within the Internet is treated (for instance in the presence of failures).

1.5 Methodology & Approach

The study comprises results obtained from a variety of sources: interviews with practitioners and domain experts active in the study's four subject areas, a desktop research and literature search, and a formative workshop with working groups comprised of community representatives.

To obtain a comprehensive overview of the techniques and practices currently used within the Internet ecosystem, an expert group was formed and interviewed on methods, techniques and tools used to create resilient interconnections, the causes and impacts of past incidents and on-going and future trends. To this forum, representatives from Tier1, Tier2 and Tier3[5] providers were invited, ranging from regional providers, national research and education network operators (NRENs), national incumbent operators, to multi-national networks. Internationally recognized researchers and independent specialists further complemented the expertise of the expert group.

Finally, a validation workshop presented the preliminary findings of the study and four working group sessions contributed further feedback on current usage of good practices identified within this study throughout the ISP industry, discussing on-going issues and open challenges for resilient Internet interconnections.

1.6 Organization of this Document

In the following chapters this document provides a broad overview of the interconnection ecosystem and investigates the economic incentives for operators to form and maintain it (Section 2). Starting from a review of past failures of interconnections resulting in Internet outages (Section 3), current technical and operational good practices will be reviewed to protect the network operators and interconnection system from future challenge situations (Section 4). Section 5 summarizes the identified good practices and presents a list of recommendations to stakeholders on how to further develop resiliency within the ecosystem.

[5] *See ENISA's Full Report on the Resilience of the Internet Interconnection Ecosystem,* section: **3.5.2**, The 'Traditional' Tier Structure, pp. 82-85, https://www.enisa.europa.eu/activities/Resilience-and-CIIP/networks-and-services-resilience/inter-x/interx/report

2 Economic Analysis of the Interconnection System

All interconnections, among the some 40,500 Autonomous Systems (AS, a network under single administrative control) in the Internet (as of 2012[6]), can be classified in two basic categories: transit and peering, which means that network operators are either buying the capability to forward traffic to other destinations in the Internet from a provider against monetary payment (and thereby entering a vendor-customer "transit" relationship), or are entering a bilateral "peering" relationship with another network operator to directly exchange traffic between the two parties. Figure 1, which will be developed further throughout this report, visualizes these two types of interconnections.

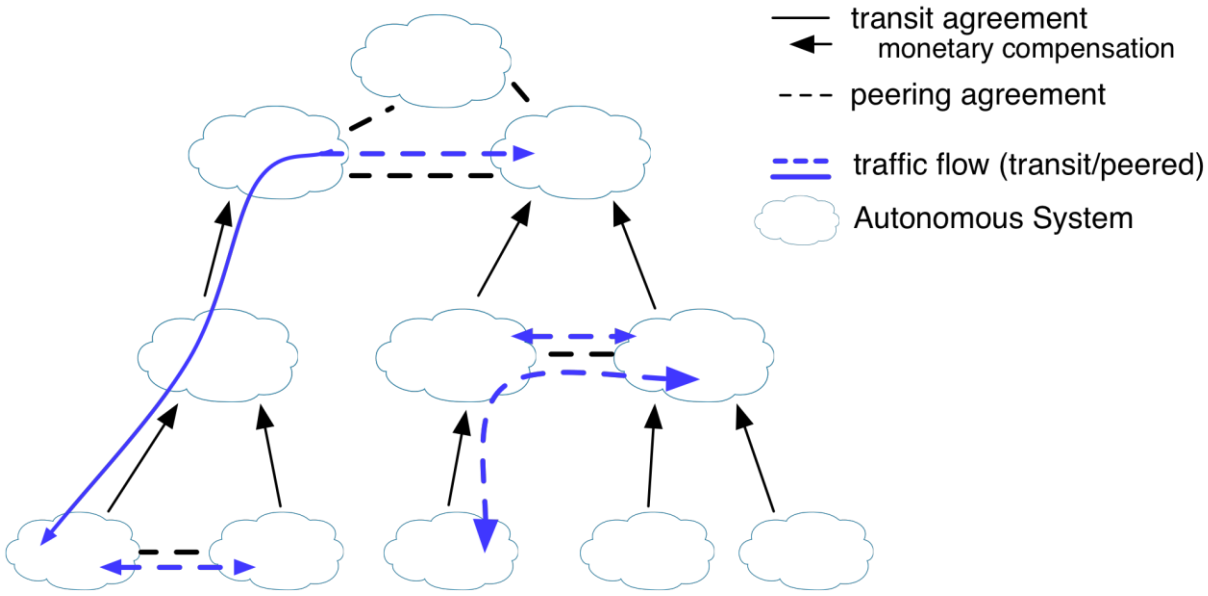


Figure 1: Network operators seek to enhance the overall areas they can provide service to by creating interconnections with other networks, usually referred to as autonomous systems (AS). These relationships are either monetary (transit) or reciprocal (peering).

2.1 Historical Development of Interconnection

The construction of the current Internet into hierarchical transit agreements and horizontal peering agreements can be traced back to the organization and strategic decisions made for the early predecessors of today's Internet, the ARPANET (1969) as well as the CSNET (1981) and the later introduced NSFNET (1986)[7]. Originally conceptualized as a means to connect users to the few distributed computing facilities in the U.S., access to the ARPANET was restricted by the Advanced Research Project Agency (ARPA) to the U.S. Department of Defense and its contractors only. Excluded

[6] *T. Bates, P. Smith, and G. Huston, CIDR Report, http://www.cidr-report.org/, retrieved April 2nd, 2012.*

[7] *J. Ryan, A History of the Internet and the Digital Future, Reaktion Books, 2010.*

from using ARPANET, similar networks were subsequently developed by other government agencies and organizations that recognized the value and benefit provided by a networked environment. Of most significance were the two networks funded by the U.S. National Science Foundation aimed at connecting the scientific community: CSNET, linking computer science departments, which broadened six years later as NSFNET towards academic institutions in general.

Based on the previous limited access experiences with ARPANET, which created pressure by dividing academia and industry into "haves" and "have-nots" (depending on whether they were a Department of Defence contractor or not), the newly funded CSNET was designed with universal interoperability at its core: it had to be able to interface and provide the means to exchange data with the other major networks at the time, which remained autonomous units but could directly interact using a common protocol[8]. This was the starting point of peering between autonomous systems and the principles of these early interconnection agreements live on until today.

With the rapid growth of the subscriber base and the geographical coverage area of the initial individual sub-networks, the expansion of CSNET towards a general audience under the umbrella of NSFNET also brought in architectural changes. Instead of continuing to build a network by linking individual sites as equals, NSFNET introduced hierarchies into network design: institutions connected by this new infrastructure would be part of smaller local and regional networks, which were linked at network access points (NAP) to a long-distance backbone providing access to other parts of the country. This hierarchical organization into tiers, with Tier1 being highest up the hierarchy, is a core feature of today's interconnection landscape.

While NSFNET was introduced to provide equal access to those excluded before, it did also introduce some access restrictions on its own. One of those restrictions was that commercial, academic and research organizations could access and be part of the network at the local and regional deployments, but only research-oriented and education-related traffic was permitted to use the backbone connecting the individual regionally distributed parts. Thus, industrial actors could get first-hand experience of the potential utility such an inter-networked system could provide, but to use the full potential the commercial sector had to build up alternative long-distance backhaul networks between the already established network access points. It was NSF's intended goal to stimulate through such restrictions in acceptable use policies the development of an environment of competing commercial long-haul networks[9], many of which have evolved into today's Tier1 networks.

---

[8] *P. J. Denning, A. Hearn, and C. W. Kern, "History and overview of CSNET," in* ACM SIGCOMM*, 1983.*

[9] *B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolff, "A brief history of the internet," tech. rep., Internet Society, 2011.*

2.2 Economic Factors for Interconnection

While in the past interconnections were heavily driven by policy decisions, today, the decision how and where to interconnect is predominantly based on economic considerations. In principle, interconnection to other networks has become a necessity to be pursued by any operator, ranging from the small local network to a global backbone operator, since (a) no network has a large enough subscriber market share (anymore) to remain a closed system and (b) customers are expecting complete reachability to any service and part of the Internet. Full access and interoperability to all Internet destinations is a prerequisite in the market, and over the last 10 years, all proprietary networks and applications (such as for example CompuServe, MSN Network or AOL) have, regardless of their size or market share, either vanished or adopted inter-operable systems, applications and protocols.

How a network operator provides universal reachability is essentially determined by the provider's relative position, as expressed by its network size, geographical coverage, and type and magnitude of traffic. The simplest and easiest solution to obtain full connectivity is through transit agreements (see solid arrows of contracts and traffic flows in Figure 1), in which a network obtains the capability to receive and forward from another operator against monetary payment, who in turn has obtained the capability to reach all other destinations either through paid transit or peering agreements. The settlement amount should include the costs incurred by the upstream network to operate and maintain its infrastructure as well as ideally include some margin to realize a profit. However, ever since the widespread adoption of the Internet, prices have been in free fall as shown in Figure 2. This deterioration can be attributed to two main factors:

1. There exists a large supply of international transit - both in terms of competing providers as well as overall capacity - a remnant of the "dot com boom" in the late 1990s/early 2000s.
2. Internet access has turned into a commodity market (which does not provide room for any proprietary offer). In such commodity markets – especially in an oversupplied market, pricing typically becomes the major differentiator between otherwise indistinguishable offers, increasingly driving down the price level and revenues.

Over the past years, the several actors in the oligopoly transit market have suffered significant losses and a consolidation of actors towards fewer, but larger competitors has begun[10,11].

[10] http://dealbook.nytimes.com/2011/04/11/level-3-to-buy-global-crossing-in-3-billion-deal/

[11] http://dealbook.nytimes.com/2011/04/27/centurylink-to-buy-savvis-for-2-5-billion/

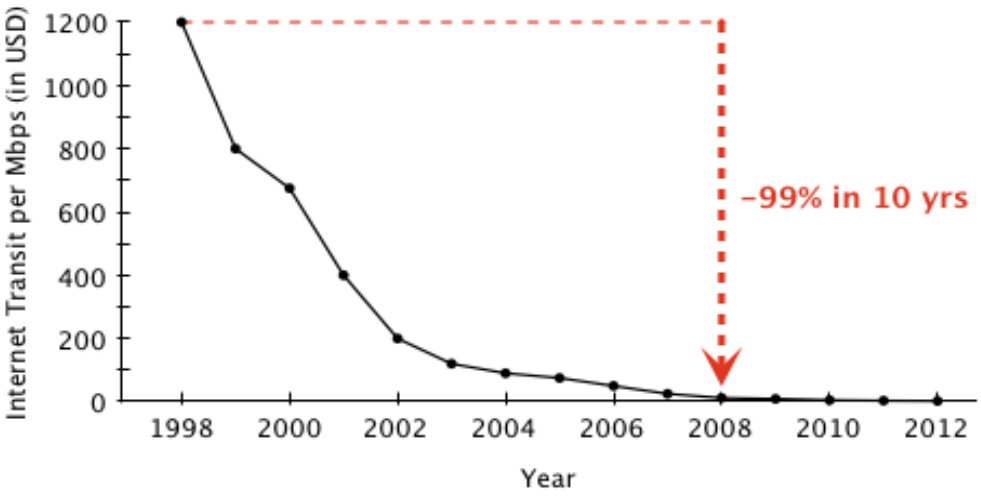


Figure 2: The prices[12] for whole-sale transit have decreased by on average 30% per year, leading to a price cut of over 99% within a 10-year time frame.

When being linked through a transit interconnection, the involved operator pair establishes a client-vendor relationship: the vendor agrees to forward all traffic of the client via its other available interconnections, announces the customer's prefixes as well as receives and delivers incoming data to the client against transit fees. There exists a variety of transit pricing methods and contract terms contain a sizable amount of flexibility[13]. Typically, transit fees are billed by the 95% percentile of the forwarded traffic[14], meaning that from the volume of regular traffic measurements – frequently assessed in 5-minute intervals – the highest 5% of the samples are discarded. This eliminates to some extent the impact of Denial of Service (DoS) attacks and unusual traffic behaviour from a customer's typical usage. For a calendar month about 36 hours of extraordinary high traffic can be tolerated without any impact on the 95% percentile. For example, suppose that in the schematic topology of Figure 3, autonomous systems (AS) 1 and 3, 2 and 3, as well as 3 and 8 have established transit agreements (as indicated by the solid arrows). As a consequence AS1 can now reach Tier1 networks such as AS8 and AS9 via its upstream provider AS3.

[12] *data source:* www.drpeering.net

[13] *E. G. Manning and G. C. Shoja, "A Framework for Quality of Service Control Through Pricing Mechanisms," Network Operations and Management Symposium (NOMS), 2006.*

[14] *X. Dimitropoulos, P. Hurley, A. Kind, and M. P. Stoecklin, "On the 95-percentile billing method," in* Proceedings of the 10th International Conference on Passive and Active Network Measurement*, 2009.*

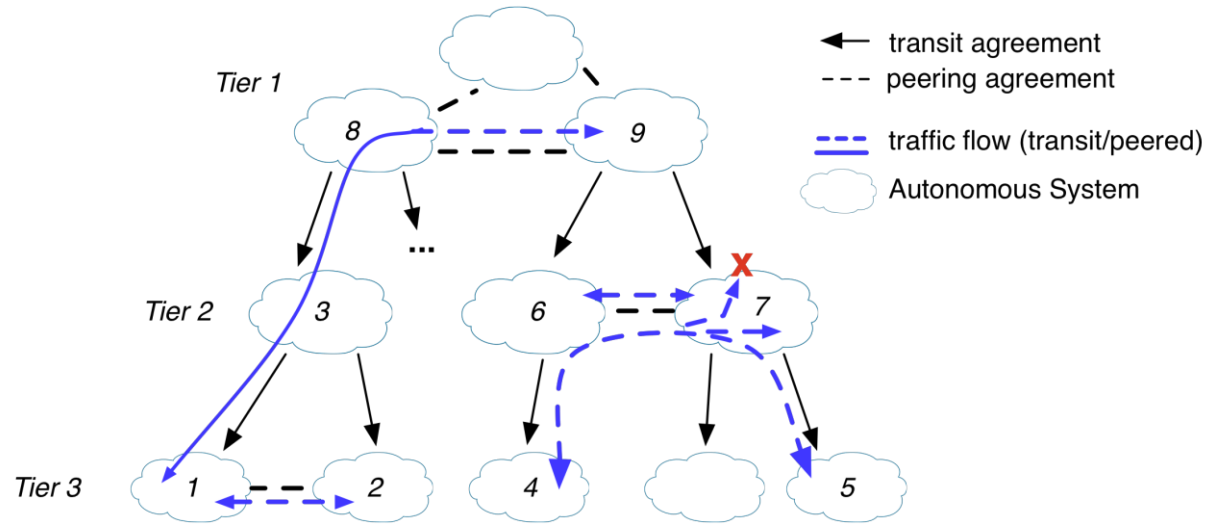


Figure 3: Schematic structure of the interconnection types between autonomous systems and resulting traffic flows.

Instead of obtaining connectivity only through a paid transit agreement, network operators may also be in a position to secure some level of reachability through peering agreements, driven by the economic incentive to reduce their costs. In the example of Figure 3, all traffic of AS1 not terminating in its own network is being routed via its paid transit connection to AS3. Suppose that AS1 and AS2 are two networks that happen to exchange a significant amount of traffic between them. In a transit-only situation, this traffic will be paid for twice: AS1 will pay AS3 for its transit, as does AS2. If for example the two networks are reasonably geographically close and/or the amount of traffic exchanged between them is large enough, it may make economic sense for these two networks to enter a peering agreement between them, typically at several points-of-presence (POPs). By establishing a direct connection, both networks are now able to deliver any traffic terminating in the respective other network. Instead of both paying AS3 for traffic exchanged, the saved expenditures can be applied to the dues of the direct peering connection. Clearly, establishing a direct peering is only economically feasible if the costs of transit are larger than the overhead of a direct peering. When traffic is exchanged through peering, this new interconnection only becomes applicable for the traffic sourced or destined for autonomous systems involved and their customers, such as AS1 and AS2, or between AS6 (and its customer AS4) and AS7 (and its customers such as AS5) in Figure 3. Peered traffic is never forwarded along paths terminating neither in the AS itself nor at its customers (such as AS6 → AS7 → AS9 in the figure), otherwise the peering network would pay transit for traffic it or its customers are neither sourcing nor the destination.

2.3 Peering Practices

It is commonly argued that peering typically requires the participating networks to have a similar size

or generate comparable amounts of traffic[15]. Closer inspection however reveals that factors such as comparable size and traffic appear to be only circumstantial and seem to be repeatedly ignored by market actors. Examples include the case where a Tier1 provider enters a settlement-free arrangement with a local or regional network operator (such as MCI/UUNET[16]) or when networks with largely different volumes of incoming and outgoing traffic peer (as observed in peering agreements between content providers or content delivery networks and mostly consumer based, so-called eyeball, networks). These apparent contradictions can be easily explained if we assume a pure cost-driven perspective to peering: two networks will peer if doing so will provide a more cost efficient alternative to transit. The decision is further simplified if the efficiency gains are symmetric, for example if the cost from incoming traffic is roughly the same as the one incurred by outgoing traffic. In case of moderate asymmetric gains, settlement-free peering agreements, in which traffic is exchanged without any direct compensation and the cost of the connection is shared between both partners, are frequently superseded by paid peering arrangements, where the costs of the connection are mainly covered by the lead beneficiary and/or mismatches in traffic volumes are compensated through an additional cash payment. Peering arrangements may even be established in client-provider transit situations, if the revenue lost from potential or previous transit traffic is outweighed by shed costs and the change from transit to peer does not have any net difference on the previous transit provider's operating costs, a zero-sum settlement. This particular form is however rare and often banned by company policies as coined in the expression: "Once a customer, never a peer"[17].

The same economic consideration can be seen as the driving force for peering interconnections ranging from small local networks to global networks: Milgrom et al. (2000)[18] for example demonstrate how this cost-driven view will naturally lead to the emergence of peering between backbones. Thus, it is more likely that peering networks have a similar subscriber base, customer group and similar regional extent, and thereby implicitly feature similar traffic patterns and cost structures - leading to a mutually efficient peering opportunity -, than that network operators are making a wilful strategic decision to peer only with those of equal size. Peering therefore is in effect a pure micro-economical optimization of an operator's interconnection portfolio.

2.3.1 Organizational Aspects of Peering

As discussed above, there are two major strategies how providers pursue interconnection, either

---

[15] *L. Chapin and C. Owens, "Interconnection and peering among internet service providers," tech. rep., Interisle Consulting Group, LLC, 2005.*

[16] *A. R. Metzger Jr. and S. D. Blumenfeld, "Supplemental internet submission, cc docket 99-333," tech. rep., Federal Communication Commissions, 2000.*

[17] *J. Hess, "Why don't ISPs peer with everyone?," NANOG Mailing List, June 2011.*

[18] *P. Milgrom, B. M. Mitchell, and P. Srinagesh, "The Internet Upheaval – Raising Questions, Seeking Answers in Communications Policy," Chapter in Competitive Effects of Internet Peering Policies, MIT Press, 2000.*

through transit or by peering agreements. When networks decide to peer, additional aspects come into play, ranging from how they will peer, where the interconnection is being made and who will pay for it. Consequently, in the following, we will analyse these aspects separately in order to understand interconnection incentives and circumstances that drive the selection process in choosing one peer or transit provider over the other. Figure 4 visualizes these three components and their typical choices.

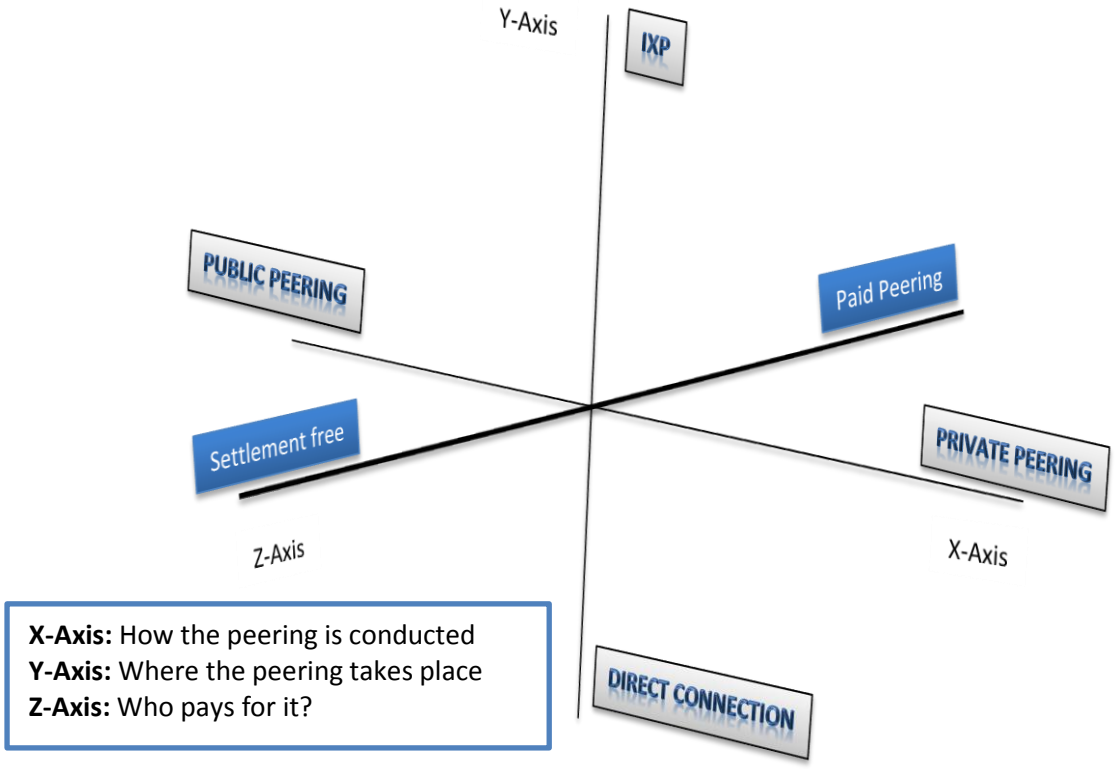


Figure 4: Classification of peering agreements by the "how" and "where" of peering, as well as "who will pay for it?"

How is peering conducted: with one network or many?

The first aspect that characterizes a peering arrangement is how a network will peer with other operators, especially if it privately interconnects with only one network or a selected few on an individual basis or whether the network publicly seeks as many connections as possible. The former case is referred to as "private peering", the latter one as "public peering".

Whether a network enters public or private peering agreements predominantly depends on the size of the operator and the parties it connects to. Larger networks, such as Tier1 and Tier2 operators,

conduct the majority of their interconnection through private peerings. Public peerings are typically sought after by smaller ISPs and usually established at Internet Exchange Points (IXP).

There presently exists only a limited amount of publicly available information about the process that network operators use to evaluate interconnection options. It has been found that this is due to two main causes:

1. Some networks wish to maintain some room to be able to enter peering agreements on an individual basis.
2. The existence of on-going contracts and especially their terms and conditions are frequently kept under non-disclosure agreements (NDA) to minimize leakage to other paid peering or transit customers who might then use this knowledge to renegotiate for alternative terms.

In contrast, Tier1 network operators are frequently publishing a minimum baseline of requirements they have before they consider peering requests. This can be seen as a first filter to reduce the number of inquiries from networks such an operator is not interested in peering with, as it will not provide significant benefit and could take potential business away. Table A.1 (Appendix) lists a summary of the peering requirements of 10 major Tier1 network operators, in terms of the requirements on the partner's network itself, the mechanics of the possible peering, as well as business and operational requirements. Note that 7 out of 10 list very concrete specifications. For example, all network operators with a public peering policy demand a network operation centre (NOC) as a central point of contact and coordination from a partner, which is staffed around-the-clock. Most networks also place some minimum demand on the network's bandwidth, the majority expecting at least a 10 Gbps backbone. All providers require establishing interconnections in several locations, and have a minimum traffic volume threshold that must be exceeded and state a maximum inbound/outbound traffic ratio to become eligible. To prevent abuse, all networks regulate the route advertising and forwarding policies of a peering, such as hot-potato routing or identical route advertisements at different interconnection points. Two of the networks explicitly exclude current customers from entering a peering relationship.

As the terms of private peering options and even mere existence of private peerings between networks is not public information (compared to for example to databases of public peerings at IXPs[19]) and Tier1 and Tier2 providers predominantly make use of this interconnection option, the actual state of the Internet interconnection ecosystem is likely to vary drastically from the maps routinely generated from the publicly available data sources. Analyses of the Internet graph and the derived conclusions should therefore be evaluated carefully.

[19] *For example www.peeringdb.com*

Where peering takes place: interconnection at IXPs vs. direct interconnects

The second component that determines a peering agreement is where the interconnection will be established, either at the facilities of an IXP or via an optical fibre interconnect connecting the networks of the two operators directly.

For private peering, interconnections at IXPs are preferred when the amount of traffic exchanged via each individual peering relationship is relatively small and the bandwidth provided by an optical link between the two peers (commonly 10 Gbps) would not be sufficiently utilized to justify the recurring expenses of the connection. Instead of establishing dedicated fibre links with all its peers, a network could instead opt to join one or more IXPs, using the IXP's facilities to reach several of its private peers simultaneously, thereby consolidating several lower-volume traffic streams to achieve a higher utilization and larger return-on-investment of its connection to the IXP. Even though the interconnection takes place at an IXP (which most commonly facilitates public interconnections), the peerings in such arrangements are still private (they will not be visible to other networks at the exchange point). In these circumstances, the IXP simply acts as a facilitator.

As networks and the volume exchanged via peering agreements further grow, peering via direct optical links becomes the default choice, as at a certain traffic volume a leased line is more cost effective than the costs associated with an IXP presence. For a simplified, illustrative example, consider a situation where a leased 10 Gbps cross-European fibre (either to an IXP or between two networks) costs 3000 Euros/month, while the fees associated with an IXP-connection at a 10 Gbps port are 1500 Euros/month. If two operators are engaging in a private peering relationship via the IXP's facilities, a share of each their connections to the IXP will be spent on their private peering for which both need to pay. In such comparative cost system, moving a private peering onto a dedicated link makes sense when the combined relative share of the IXP costs outweighs the dues of a direct connection (and the freed IXP link bandwidth can be filled otherwise). In this situation, an exchange of more than 3.3 Gbps via a single peering relationship will make it worthwhile to pursue a direct link (33% of 10 Gbps * (3000 Euro/mo+1500 Euro/mo)*2 = 33%*9000 Euro/mo > 3000 Euros/mo). In practice, larger networks with a significant traffic volume therefore conduct their private peerings nearly exclusively via direct interconnects and only rely on IXPs for private peerings in remote markets where the exchanged volume is comparatively low or no other feasible or economical options are available. Smaller networks on the other hand conduct a larger share of private peering relationship at IXPs, as the ability to consolidate several agreements into one physical link makes this option economically viable.

Public peering on the other hand can by definition only take place at IXPs, where the presence of multiple providers allows the simultaneous exchange between several networks at once. Comparatively low fibre costs now also make it feasible for network operators to connect at Internet exchange points distant from the operator's core deployment. With several dozen IXPs offering their services within Europe alone, network providers have to develop a strategy how to commit their resources and where to connect. Besides access and operational issues, such as (i) whether there exists sufficient capacity to connect to and build into an IXP, (ii) the exchange is in a financially stable

situation and has enough staff to accommodate requests, the decision to participate at an IXP is mainly driven by cost-value propositions and to a lesser extent by peering policy considerations.

In its 2003 annual report[20], the Amsterdam Internet Exchange (AMS-IX) presented a model by Bill Norton to assess the value provided by an IXP. According to this assessment, the

$$\text{value of an IXP} = \frac{\text{avg. traffic maximum} \times \text{avg. cost of transit}}{\text{avg. number of members}},$$

which the AMS-IX delivers to its members was derived for the year 2003 as 16 Gbps · 75 Euro/Mbps / 160 = 7680 Euro/month. Following this model, new members should continue to subscribe until the added value is outweighed by the subscription costs of currently 1500 Euros per 10 Gbps network port per month[21], when several competing options are available it is in the best interest of ISPs to pursue that offer generating the highest additional value from joining.

This additional value can be obtained by comparing the financial aspects of an IXP participation with the operational implications. Consider the situation that a network provider is already present at a number of IXPs, and is considering to build into an additional IXP which according to the value analysis of the AMS-IX pricing model would realize cost savings. If the new addition however contains the same set of networks to which the operator already peers at other IXPs (or the disjunctive networks are not significant sources or sinks of traffic), there is no positive benefit, except for possibly shorter routings and added resilience, which might not be enough to justify the associated expenses and overhead. Thus, the incentive to join an IXP is driven by a combined cost-value analysis, commonly approached as how many unique routes with their corresponding traffic are being added.

Who pays for peering? Settlement-free peering vs. paid peering

Finally, the third aspect that describes peering arrangements is who will bear the cost of the agreement when two networks are interconnecting. When networks peer publically or do so privately at an exchange point, the incurred costs are implicitly shared by the networks as each is paying for their own dues. If networks peer privately via a direct interconnect, the associated costs become a factor in the negotiations.

In the past, peering arrangements were commonly conducted settlement-free, meaning that both networks equally share the costs for establishing and maintaining the link between them. Derivations from this common practice have been observed when one of the partners either perceived the arrangement to not benefit the networks equally well (for example if the incoming/outgoing ratio is skewed), or when the economic incentive to pursue peering instead of transit was borderline. In these

[20] *"Annual Report 2003", technical rep., Amsterdam Internet Exchange, 2003.*

[21] *http://www.ams-ix.net/pricing/*

situations, networks may either deviate from a 50-50 split of the interconnection costs towards a more one-sided distribution, or establish a paid peering relationship where one network pays the other some fee (not necessarily limited to the actual costs of the interconnection) to be able to peer with the other network directly. Motivations to do so include for example the ability of the paying network to gain better (in most circumstances lower latency) access to the other operator's subscribers.

2.4 Challenges and Future Trends

From a resiliency and connectivity perspective, the interconnection system would converge to a comparatively dense structure, if it would not be for the costs of maintaining individual interconnections. To reduce and consolidate peering expenditures, connections among many – today predominantly smaller networks – have in the past been mostly established at IXPs. IXPs provide the facilities for world-wide, continental, regional and local networks to connect, and the externalities associated with the placement of such infrastructure and presence of major networks has led to a significant concentration of peering activities at these points. Internet exchanges are typically operated by a flat-fee model based on the number of rack units of networking equipment placed by a particular operator at the facilities. In Europe however, IXPs are frequently run as a non-profit cooperative, where the costs of operating the exchange points are shared and divided up among the participating network providers. In the past, it has been reported that cooperative IXPs were endangered or fell victim to the tragedy of the commons[22]: it is in the best interest of everyone to draw as much traffic and benefit out of the common infrastructure, while contributing as little as possible for the maintenance and upgrade to cope with the increasing exploitation.

The commoditization of Internet access, diminishing Internet access costs paid by customers and increase in traffic volume has led and can be expected to continue to lead at least in the near future to (i) an increased concentration of ISPs at exchange points and (ii) an increasing market domination of the current market leaders, an aspect further discussed in Section 2.5. In pursue of cost reductions, ISPs at Tiers 2 and 3 will attempt to peer directly with as many other networks as possible.

From an economical viewpoint, future interconnection strategies will depend upon the interplay between three major cost factors in interconnection: the triangle between the development of transit prices, the price trend in leased lines, and the cost of IXP membership and usage. The current interconnection landscape we have today has been shaped by relatively high costs of transit compared to leased fibre and IXP connections: as transit was expensive, ISPs peered whenever possible to shed costs leading to a very concentrated market.

Decreasing transit and fiber prices weaken the position of IXPs

Within the last decade however, transit prices have dramatically declined (see Figure 2), while leased

[22] *G. Huston, "Interconnection, peering and settlements,"* The Internet Protocol Journal*, vol. 2, no. 1, 1999.*

fibre has experienced only a moderate decline. These two developments now shift the economical incentives of interconnection. At low transit prices compared to relatively high IXP costs, cost reductions via public peering become less tangible or infeasible. Suppose a network operator is present at an IXP with a 10 Gigabit port, and the monthly cost of this engagement (fees, rental, energy, equipment) is in the order of 7,000 – 8,000 Euros. Just to recover the direct costs, in 2005 the provider had to offload 130 Mbps through publicly peering networks at the IXP, which otherwise needed to be routed via transit providers for an approximate 2.50 Euros per Mbps (see Figure 2). With the price decline, in 2011 more than 3000 Mbps would have been needed to be routed via its peerings to recover the monthly IXP costs. Interpolating from the current devaluation, the public peering will become more costly than the alternative transit price for this example within the next 3 years, i.e., more than the full switching capacity of the IXP port would be continuously needed to still realize a cost reduction through public peering.

It should be noted at this point that there exist some significant differences between the IXP landscape and operating principles in the United States and Europe. Historically, Internet exchange points in Europe have evolved from community efforts, and are typically run as a non-profit association or by academic and government organizations. These setups are geared towards the provisioning of universal IX service instead of revenue maximization, thus participation is relatively inexpensive as usually only the shared operating costs are distributed among the involved operators. As IXPs are run, supported and controlled by the community, cooperation among IXPs is wide-spread and the installations are carrier and co-location neutral, i.e., ISPs can choose among several competitors all servicing the same IXP. In the United States on the other hand, IXPs have commonly been deployed by private players and are therefore maintained for-profit, where the entire installation and location is owned by the IXP operator. Since all control is in the hands of the particular IXP maintainer, there is no direct influence by participating networks (who are customers instead of members), little incentive to cooperate between different IXPs and no competition at an IXP for example for co-location space. In effect, the prices are not cost-driven but market-driven, and will be as high as the particular local market will bear. This difference also leads to different peering considerations and dynamics in the US and EU in general and at US vs. EU IXPs in particular. The extent to which this development will manifest itself is further influenced by the rate of future technological development, such as whether the value-proposition of the IXPs can be maintained simply by providing cheaper and faster exchange through technical advancements.

If such technical developments cannot provide sufficient support, we might experience a shift in Internet interconnection in the long run from public peerings at exchange points either back to transit or towards private peering agreements, where two operators are establishing a direct (leased) fibre link between their networks, as it can already be seen among larger network providers. This alternative is now also increasingly in reach for smaller networks due to the relatively low costs of

leased fibre: for an average cost[23] for typical intra-European OC3 lines of 1.90 Euros per kilometre, maintaining direct private peering connections to several other networks on the same continent, as an alternative to peering with them at a major European IXP, is economically viable. Clearly, such interconnections would only be sought after with the most important ISPs (in terms of off-loadable traffic), and it is reasonable to conclude that the interconnection network would become thinner regionally.

Spread of private peering

Private peering initiatives could emerge at a global scale. The average retail price of a transatlantic 10 Gbps wavelength declined to a monthly leasing cost of 10,935 Euros, or 1.92 Euro/km. ISPs with a large enough subscriber base and traffic volume could reach out and establish more international presence by peering in major foreign markets. The viability of this step depends on the future development of prices for transit and fibre leases, which will depend (among other factors) on the further consolidation of the Tier1 market. If operating own long distance links becomes economically lucrative for Tier2 (and 3) operators, this development will lead Internet network operators on a similar path as telephony network operators, where local providers extended into long-distance markets, thereby undermining the established transit fee model of the international backbone operators. A similar step in the ISP domain would further put pressure on the existing transit market and likely lead to an unsustainable state for international transit. The general existence and possible outcome of such instabilities repeatedly became visible over the last decade in a chain of peering disputes, commonly leading to a temporary or permanent de-peering, i.e., a cut of the direct connection between two networks and a halt on any traffic forwarded between the two parties and all their respective customers via this link. Although a number of providers have encountered de-peerings, the long-distance Tier2 operator Cogent Communications has generated wide-spread attention and publicity over these type of business disputes, notably for its conflicts with AOL (2003), Level 3 (2005), France Telecom (2006), TeliaSonera (2008) and Sprint (2008), which due to the size of the involved parties caused slowdowns and service disruptions for a significant customer base and in practice cut off some parts of the Internet. While the affected parties have put forth a variety of arguments to explain or blame these disruptions, such as most recently network neutrality, these conflicts and subsequent de-peerings can typically be reduced to the same economic principles that led to the emergence of peering connections in the first place: the mutually beneficial cost savings of a direct connection. In all of the above cases of recent peering conflicts, the network operator initiating the de-peering has named lack or diminishing benefit of an existing connection as a cause for its termination, usually on the basis that the provider has to handle much more revenue-less incoming traffic than it is able to offload to its partner at no cost. This is particularly true in cases of peering arrangements with consumer-dominated (eyeball) networks, which by the nature of their customers and typical consumption patterns have an unbalanced traffic ratio and could be seen as more unstable. Kuri and Smith[24] however also bring up the alternative, indirect economic explanation specifically for the case

[23] http://www.telegeography.com/research-services/wholesale-bandwidth-pricing-database-service/index.html

[24] *J. Kuri and R. W. Smith, "France telecom severs all network links to competitor Cogent," Heise Online, 2005.*

of Cogent Communications, which has acquired a controversial reputation among operators due to its very low bandwidth pricing, that competing networks might strategically use de-peering to prevent the further growth of such competitors and the continuing erosion of traffic margins.

Emergence of paid peering to counter price erosion

Finally, it may be expected that paid private peering for eyeball networks[25] (i.e., networks with a significant residential subscription base) will emerge and become broadly acceptable, not primarily due to an urge to save costs (as argued above), but due to market pressure by content providers. With the widespread popularity of content delivery networks (CDN), the major Tier1 networks have seen a steep decline in long-distance traffic over the past years: while there exist literally hundreds of CDNs today, the market leading CDN[26] in 2012[27] is carrying nearly 20% of all Internet volume, traffic and revenue that is now bypassing the global Tier1 providers. The increased competition has however already started to erode prices in the CDN market[28] and it can be foreseen that CDN will also enter the stage of commoditization soon. While Internet transit currently only competes on price, Content Delivery Networks – with their focus on video and real-time applications – still have the opportunity to differentiate on the Quality of Service (QoS). Providing the best QoS – and thereby being able to still charge a premium – will however require to be as close as possible to the major consumer groups. It is currently common place that CDNs pay for peering opportunities due to the unbalanced traffic ratio. Direct and prioritized access to the eyeball networks can provide a strategic advantage, which these networks may attempt to convert into cash beyond the current practice of paid peering by offering high priority or better access to the end customers, possibly even through exclusivity agreements. The market will force content networks to explore these options if offered, in the competition for last mile latency and Quality of Service.

Given their relatively recent advent, the role of CDNs and their impact on the ecosystem are not fully clear yet. Still, their rapid spread and increasing market power has certainly already triggered several changes in the ISP landscape. Interestingly, CDNs violate the traditional economics of Internet best practices that have evolved among operators, such as hot potato routing in which a network tries to hand traffic over to the next-hop network as soon as possible to facilitate better routing and decrease its own transportation cost. CDNs on the other hand reach in their quest for better last mile access deep into operator's networks, holding on to data traffic as long as possible, frequently even deploying their own infrastructure at the core of operator's systems. In this respect, CDNs can be considered to increase the resilience of Internet services – at least those served by a particular CDN to a given operator -, as content delivery networks in practice duplicate and place resources close to the end

---

[25] *See ENISA's Full Report on the Resilience of the Internet Interconnection, section* **3.5.1** *'Eyeball' and Content, p. 82, https://www.enisa.europa.eu/activities/Resilience-and-CIIP/networks-and-services-resilience/inter-x/interx/report*

[26] *A.-M. K. Pathan and R. Buyya, "A taxonomy and survey of content delivery networks," tech. rep., University of Melbourne, 2007.*

[27] *Visualizing Akamai, http://www.akamai.com/html/technology/dataviz3.html, retrieved 2nd of April 2012.*

[28] *V. Bonneau, "Evolution of the CDN market," presented at* CDN World Summit*, 2010.*

users. Being close to the network's end customers is of vital importance to a CDN and the customer it serves – an access controlled by the network operator. CDNs however simultaneously also possess significant power, as they control a sizeable share of Internet content and without a good connection to the major CDNs the experience of end customers will suffer, potentially letting ISPs lose customers subscribing to a better connected network instead. Who in the relationship has control over the longer lever – CDNs or ISPs – is not clear yet.

The established practices and developments in interconnection described above can be expected to exhibit mixed influences on the resiliency of the interconnection ecosystem, depending on the relative position of a network operator within this landscape.

Due to the very specific requirements of large network peerings, such as for example extensive geographic coverage and interconnection at several spatially separate points, the resiliency - especially of larger providers - towards equipment failures is of a lesser concern. The current practices of over-provisioning intra-operator links by at least a factor of 2, make wide-scale outages resulting from equipment malfunction in Tier1 or Tier2 networks unlikely. Instead, a higher risk and larger impact on the overall ecosystem resilience could result from operational and technical considerations and design choices, specifically:

- *Peering disputes* leading to sudden connectivity changes impacting a large user base. Depending on the location and positioning of the affected providers, de-peerings could temporarily partition the Internet and not be remediated well by typical good practices such as multi-homing.
- *Thinning out of backup options*. Increasing price erosion can further drive providers to cut cost through excessive peering. A pronounced focus on peering however could move some operators to thin out their portfolio of then unclaimed transit bandwidth, which would – in case of operational or technical failure – provide a backup option.

2.5 Concentration and diversification of the IXP landscape

The current Internet Exchange Point (IXP) market within the EU is dominated by three major players, the "Deutscher Commercial Internet Exchange" (DE-CIX) in Frankfurt, Germany; the "London Internet Exchange" (LINX) in London, United Kingdom; and the "Amsterdam Internet Exchange" (AMS-IX) in Amsterdam, the Netherlands. These three IXPs handle the bulk of the European public peerings and carry on average nearly 8 times more traffic volume than the next largest European competitor[29] [30] [31] [32].

[29] *De-Cix – Traffic Statistics, http://www-old.de-cix.net/content/network/Traffic-Statistics.html, retrieved 2nd April 2012.*

[30] *Ams-IX – Traffic Statistics. https://www.ams-ix.net/statistics/, retrieved 2nd April 2012.*

[31] *Linx – Traffic Statistics. https://www.linx.net/pubtools/trafficstats.html, retrieved 2nd April 2012.*

[32] *Aggregate Traffic Graphs for NetNod, http://www.netnod.se/ix-stats/sums/, retrieved 2nd April 2012.*

This significant concentration of IXPs within Europe is criticized by some as undesirable, as a large fraction of the European Internet economy – particularly smaller Internet service providers – rely on comparatively few players for their general operation and financial well-being. This current dominance of a few exchange points is the result of the historical development of the Internet economy in Europe: it is where historically fibre connections came in and in the beginning stages of the Internet major networks took up residence, that over the evolution of the Internet ecosystem also exchange points sprung up around these network concentration points. Other European exchange points, such as France-IX (Paris, France), MIX (Milan, Italy), NIX (Oslo, Norway), ESPANIX (Madrid, Spain) which were founded in the same time frame, did not have these strategic advantages and today play only a minor role in the IXP landscape, carrying a fraction of the European public peering traffic.

Increasing future concentration

This current concentration of exchange point traffic is unlikely to change in the near future, and can even be expected to further intensify. The principal driving force for the initial development and current popularity of IXPs is the ability of networks to offload some of their traffic via peering relations and thereby avoid the transit dues.

For this value proposition to manifest and offset the costs associated with an IXP presence (such as membership fees, equipment costs, connection to the IXP etc.), a sufficient number of networks, with which a particular newcomer is currently exchanging traffic, already need to be present at such a facility. Less populated exchange points provide less opportunity to do so, and may not generate enough value to justify the costs. Thus, networks tend to connect to the largest market players offering the highest value proposition, resulting in an increasing concentration. Figure 5 schematically depicts the value an Internet exchange point can provide as a function of the participating networks and traffic volume, visualizing the “startup hump” and non-linear value proposition hindering the establishment and development of new and smaller IXPs.

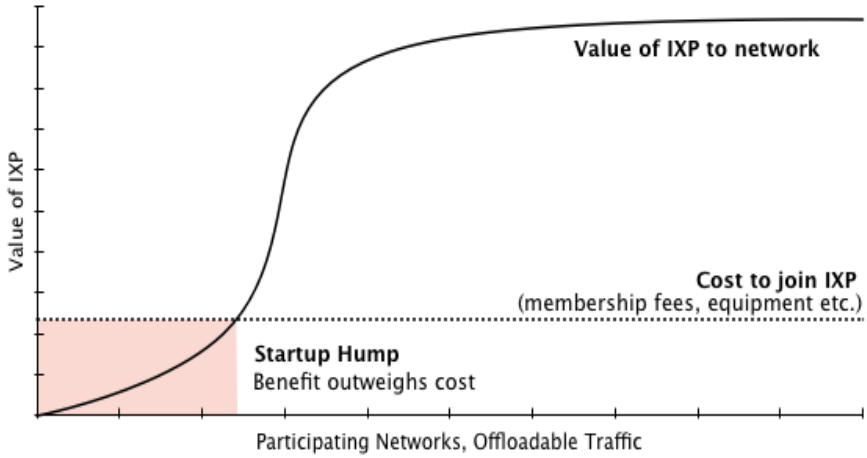


Figure 5: Value of an Internet Exchange Point (IXP) as a function of its participating networks[33].

While in the early stages of the Internet, a relatively high cost of fibre leases supported the use case of regional exchange points, the large low-cost availability of optical connectivity is further accelerating this concentration: Instead of establishing or investing into an exchange point to develop it into a large enough presence, it is much easier even for smaller ISPs to directly connect to LINX, DE-CIX or AMS-IX. With the availability of peering options and lower-cost transit prices in nearly all cases offsetting the costs of long-distance fibre connections and ISPs, thus avoiding the administrative and financial burden and risk in the development of new IXPs, the current ecosystem and recent trends are unlikely to change organically. Providers will choose the path of least resistance and the financially most rewarding option.

Publically funded diversification?

A potential response to countervail the current concentration of public peering could be to stimulate the development of new IXPs, for example in southern Europe, by financing the initial infrastructure deployment and early maintenance costs through public funds. While this approach would seem a logical option, all experts participating in this study were highly sceptical about the lasting impact of such an initiative.

Besides the "start-up hump", i.e., the economical unattractiveness of newly established or exchange points below critical mass that could be overcome by public funding, the success of an IXP also largely depends on the acceptance within the local community and motivations of the players within the local or regional target market. Many of today's poorly utilized IXPs have failed to foster this community aspect, assuming that the recipe to success is mostly technology-driven ("build it and they will come").

[33] *Modelled after DrPeering, Internet Service Providers and Peering, http://drpeering.net/white-papers/Internet-Service-Providers-And-Peering.html*

Whether new IXPs can in principle emerge in regional environments (and reach a sustainable scale) is instead determined to a significant extent by the motivation of local and regional players to peer directly, in other words whether there is a strong justification for eyeball networks and content providers to peer directly in the region rather than outside at an established IXP. Unless large quality improvements in the delivery of content to end users (typically latency) from such a local peering could be realized and the consumer base served by the eyeball networks is substantial enough for content providers to be attracted and establish a local presence within this regional market, such critical mass may likely not be achieved.

Too much diversification weakens the landscape

Besides the expected low likelihood of success, such initiatives could also potentially weaken the existing European ISP ecosystem as it increases costs and introduces additional operating complexity. In a survey of peering coordinators conducted by one expert in our study, all polled networks indicated a strong preference to limit the number of exchange points within a given region and thereby benefit from a high network concentration.

Nearly 50% of all network operators considered a situation with two IXPs per region (e.g. US east, US west, Japan, Central Europe) as desirable, ideally each housed in different data centres, supplied by different utilities and operating on different vendor hardware, in order to achieve resiliency. According to these peering coordinators, the presence of more than two IXPs per region would however only unnecessarily fracture the community and require the installation of resources at several locations yielding increased costs and management overhead without providing a noteworthy benefit. The other half of the polled networks recommended even the presence of one IXP per region only, as such setup would lead to a universal accessibility of all a region's ISPs at a single facility and thereby maximizing the benefit and return on investment. For these networks, resiliency is of no concern as it is otherwise realized within their system by multi-homing or interconnection across multiple regions.

3 Past Interconnection Incidents and Vulnerabilities

By analysing past incidents that disrupted the interconnection system in general or impaired the interconnection between network operators in particular, a number of good practices (presented in Sections 4 and 5) to avoid similar incidents in the future might be identified. Non-technical problems, like peering conflicts, are not discussed in this section. Our goal is to determine the root causes of example incidents of the most important vulnerabilities. Unfortunately, it has been difficult to reveal any reliable descriptions of concrete actions taken by operators during crises, since there is little benefit and significant risk for operators in exposing these techniques in a public forum. We therefore recommend (Section 5, Recommendation 4) to develop an infrastructure to bring together and safely share information on networks, infrastructure and network-related incidents. The few crises responses that have been published suggest an ad-hoc approach. A forum that is popular among network operators is NANOG[34], the North American Network Operators' Group. Via NANOG people from the network research community, the network operator community and the network vendor community come together to identify and solve the problems that arise in operating and growing the Internet.

In the following we present incidents and vulnerabilities related to:

- Border Gateway Protocol (BGP)
- Human error Problems and Maintenance
- Interconnection Equipment
- Physical Network Disruptions
- Malicious Intent

3.1 BGP Vulnerabilities

Given that the Border Gateway Protocol (BGP) is *the* mechanism to manage and maintain Internet interconnections between operators (no other protocols are used), the technical security of the interconnection ecosystem is largely driven by the security of BGP.

Given its key importance to the Internet, BGP is surprisingly susceptible to malfunctions, and despite recent technical advances and significant attention devoted to its vulnerabilities, BGP is still considered by most providers and experts “the Achilles’ heel of the Internet”. Indeed, most of the recent interconnection-related incidents and Internet outages have revolved around BGP failures and vulnerabilities, and as BGP disorders tend to spread fast within this protocol monoculture, many incidents can have regional, national or even international impact. An excellent survey of BGP security issues and proposed solutions has been provided by Butler et al.[35]

---

[34] *http://www.nanog.org/*

[35] *K. Butler, T. R. Farley, P. McDaniel, and J. Rexford, “A Survey of BGP Security Issues and Solutions,” Proceedings of the IEEE, vol. 98, no. 1, January 2010.*

BGP's role in networks and for the interconnection system in general is to collect and distribute information on how particular ranges of IP addresses and their associated autonomous systems might be reached. When two operators interconnect, they set up a session between two designated "BGP speakers" located within their networks. The speakers will continuously maintain a "BGP session" through which all inter-domain routing information will be exchanged: routes can be announced, i.e., distributed when they are available, or withdrawn when the route to the destination becomes unavailable. The information collected by the BGP speaker throughout its various sessions with other autonomous systems is then used internally to facilitate the routing of packets towards other destinations in the Internet. If BGP is disrupted, an ISP essentially loses its capability to communicate with anyone outside the operator's own network and the parts of the Internet affected by this problem will disintegrate into a collection of unconnected islands.

Figure 6 depicts a high-level overview of these BGP communication processes between autonomous systems. If AS3 is the owner and origin for the block of IP addresses in the range 14.2.1.0 – 16.5.7.255, it communicates this information via its outgoing BGP sessions. Until otherwise specified, for example by internal rules or inter-operator agreements, each connected network, in this case *AS2*, will pick up this announcement and further distribute it via its own links, prepended by its own AS number *AS2*. In consequence, the original route continues to spread across all connected networks, with a growing path vector of AS numbers indicating the route to get to the origin.

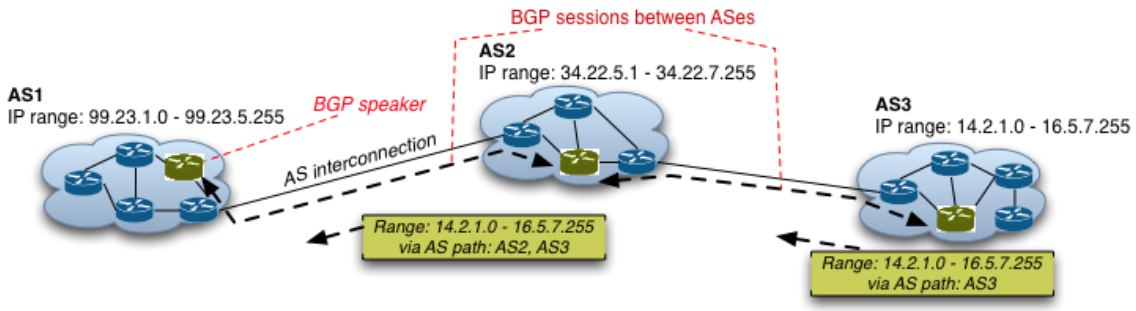


Figure 6: Announcements and distribution of inter-domain routes.

Route Forging

As long as there are no means within BGP to validate the correctness of an announcement (beyond syntax), a fundamental security problem arises as in principle any BGP router can announce an arbitrary route within the Internet.

With BGP making a routing decision primarily based on the granularity of the advertised IP range and the length of the announced path, misconfigured BGP speakers have the power to divert the entire traffic for particular IP ranges within certain regions away from the actual receiver, thereby creating a so-called "black hole". This "prefix hijacking", i.e., the announcement of prefixes actually not owned by

the misbehaving AS, could also be deliberately taken. In the past, several prefix hijacking incidents occurred, with regional (YouTube-Pakistan incident[36]) to global (AS7007[37]) impact. More recently, on February 23, 2012, one of Australia's major Internet providers faced a route leak[38]: *A routing leak can happen when a small ISP X buys transit from ISP A and also from ISP B. ISP X receives a full BGP routing table from A and, because of incorrect filtering, relays these messages to ISP B. As a result ISP B now learns all Internet routes via ISP X to ISP B and ISP X (the customer) now becomes an upstream provider for ISP B.*

Currently, many network operators therefore engage in a practice called "route filtering" (Sections 4 and 5, Good Practice 14) at their interconnection points, where they only accept BGP announcements if the particular IP ranges match the allocations and permissions in a separate validation database. While this mechanism requires keeping filters up to date, which is not trivial, it does partially protect peers against prefix hijacking. BGP routers can filter announcements they believe to have incorrect information. There are several rules or good practices to configure these filters, such as filtering announcements containing exceptionally long paths, but care has to be taken not to filter too aggressively (which could cause a black hole). A good practice is to apply filters at both sides of a customer-provider boundary. However, informing and updating a provider filter can take between 24 hours to a week. If something needs to be announced quickly, this presently still constitutes a problem. In the recent Australia route leak incident a filter was set to drop the BGP session when more than a certain amount of prefixes were advertised, which caused the interconnection to be dropped.

Origin check, not path validation

In the recent past, several ideas have been proposed to overcome BGP security problems. Recently, most efforts were conducted within IETF's working group on Secure Inter-Domain Routing[39] (SIDR) and its solutions are gradually being adopted. The recent RFC 6480[40] describes a Resource Public Key Infrastructure (RPKI) to list, digitally sign, and verify prefixes. With an RPKI it becomes possible to track whose prefixes are owned by whom. When a remote autonomous system then receives an incoming route offer, it can verify based on this digital signature whether the origin AS is entitled to distribute such route. This allows networks to detect whether an AS attempts to hijack another network's address blocks, or a currently unused block of IP addresses.

This mechanism is however only partially effective in addressing this problem. While an RPKI-based signature mechanism could reject the accidental leakage of foreign prefixes (which most previous

---

[36] *RIPE NCC, "YouTube Hijacking: A RIPE NCC RIS case study", March 17, 2008, http://www.ripe.net/news/study-youtube-hijacking.html*

[37] *V.J. Bono, "7007 Explanation and Apology", April 26, 1997, http://www.merit.edu/mail.archives/nanog/1997-04/msg00444.html.*

[38] *A. Toonk, "How the Internet in Australia went down under", February 27, 2012, http://bgpmon.net/blog/?p=554.*

[39] *IETF Secure Inter-Domain Routing Charter, http://datatracker.ietf.org/wg/sidr/.*

[40] *M. Lepinski and S. Kent, "An Infrastructure to Support Secure Internet Routing", IETC RFC 6480, February 2012, http://datatracker.ietf.org/doc/rfc6480/.*

hijacking attempts were), this origin-centric method is unable to help in detecting and avoiding any intentional attacks on inter-domain routing. Consider the situation in Figure 7, where the route announcement of *AS3* in Figure 6 has finally reached *AS0*, offering a path via *AS1* and *AS2* to the destination entitled to announce this range of prefixes.

As an RPKI only secures the origin of an announcement, a maliciously acting *AS5* could forward a copy of *AS3*'s signed announcement, even though it does not have a connection to AS3. To *AS0* this path seems much more preferable as it is shorter, and – as *AS3* can be verified as the rightful owner of the IP range – *AS0* will forward traffic for *AS3* directly to *AS5*. *AS5* could destroy the incoming information or copy before forwarding it to *AS3*. This threat remains possible when one only validates the source of a route offering, not the actual validity of the path. Hence, BGPSEC[41] is being developed to work with an RPKI to protect the integrity of the path and ensure that the announcement came the way it claims (by way of AS-PATH attribute).

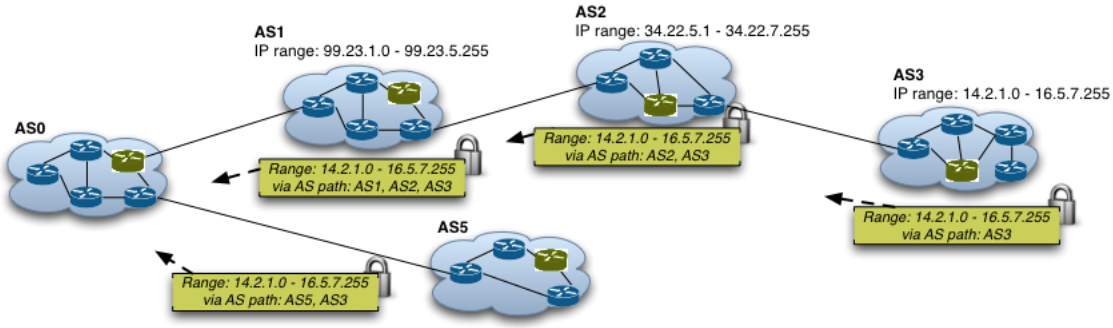


Figure 7: A signed route announcement would not provide resilience against a maliciously acting AS spoofing a path.

Dependence on trust anchors

A core ingredient to any public key infrastructure system is the existence of trust anchors that can vouch for and confirm the validity of a particular identity. The trust anchors are implicitly trusted by all actors within the system and thus allow a remote verification of any signature used within the system: as shown in Figure 8, the identity of *AS3* is verified by *trust anchor B*, the validity of its route announcements can be directly confirmed by both *AS1* and *AS2*, as the keys used in signing the routes can be checked against the issuing anchors.

Consulted experts envisaged that the RPKI deployment will begin with 5 Routing Information Registries (RIRs) as trust anchors. Each trust anchor will equip network operators with the necessary rights to issue and sign routes, and enable all others to verify these announcements. As universally trusted entities, these trust anchors gain definitive power. In 2011, the certificate authority Diginotar, a trust anchor for the PKI infrastructure responsible in issuing SSL-certificates, was breached. Given its ultimate power, the attacker was able to create any forged identities (in this case impersonating

[41] *M. Lepinski (ed.) "BGPSEC Protocol Specification", IETF Draft, March 12, 2012, http://datatracker.ietf.org/doc/draft-ietf-sidr-bgpsec-protocol/?include_text=1.*

various Google and Microsoft services[42]), which given Diginotar's approval stamp were immediately trusted by any major web browser. For the case of BGPSEC, an attacker could similarly use a compromised trust anchor to create a key able to sign any arbitrary prefix, thus black-holing traffic despite the existence of a secured BGP system.

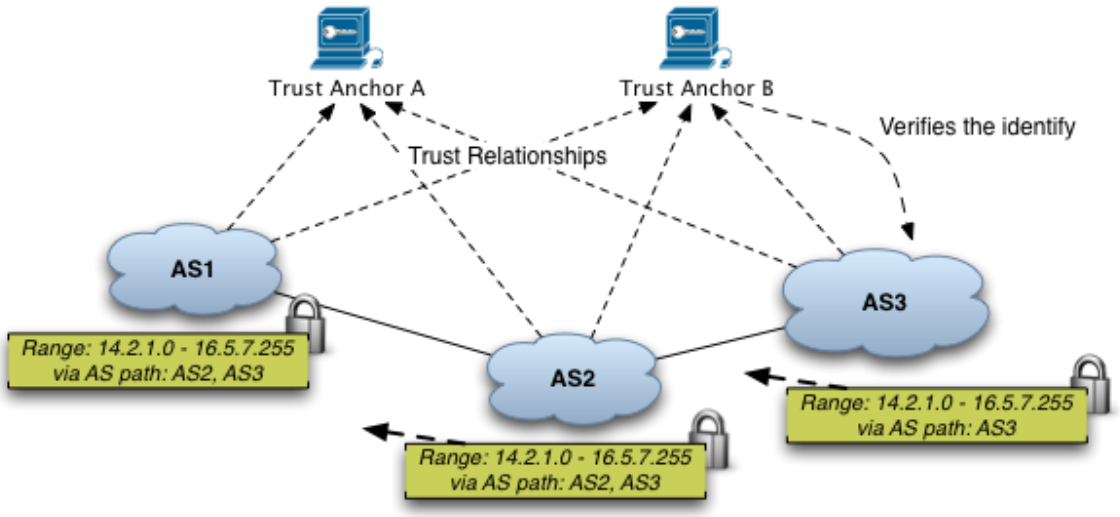


Figure 8: Within a PKI system, the trust anchors gain definitive power.

Given that the trust anchors are physical organizations each located within a particular jurisdiction, a second issue could arise. Suppose a trust anchor located in the Netherlands has issued a certificate to a provider in some other country, enabling this ISP to announce its IP range and become reachable from the rest of the Internet. As the Dutch trust anchor is subject to local laws and regulation, it might be forced by a local legislative body to revoke the previously issued certificate, for example if the activities of the foreign ISPs are illegal in the Netherlands, even though this entity has no jurisdiction over the foreign ISP. The revocation would render all the ISP's routes invalid, and in an RPKI-secured inter-domain routing universe would effectively disconnect the ISP from the Internet.

Similar conflicts have for example occurred in the case of domain names[43], where a US court issued a subpoena against a US-based infrastructure provider to seize the domain name of a Canadian company.

3.2 Human error and Maintenance

Many incidents have been caused by unintentional human errors. When humans are in the loop, such problems can and will occur. Fully automating network configuration and operation, if at all possible,

[42] *G. Markham, DigiNotar Compromise, http://blog.gerv.net/2011/09/diginotar-compromise/*

[43] *http://blog2.easydns.org/2012/02/29/verisign-seizes-com-domain-registered-via-foreign-registrar-on-behalf-of-us-authorities/*

could remedy human error problems, but in general is not supported by network engineers who feel that human intervention remains crucial. Proper education, multiple pairs of eyes, test configurations, etc. are ways of minimizing Human errors.

Maintenance of network equipment may temporarily decrease the network's resiliency and when insufficient backup measures can be taken should be communicated to customers potentially at risk.

3.3 Interconnection Equipment Incidents and Vulnerabilities

Clearly, physical interconnection equipment, like routers and switches, are of direct and crucial importance to the interconnection system. We identify three types of vulnerabilities related to network equipment:

- *Legacy Equipment*: Routing policies influence how BGP route attributes (e.g., local preference, AS path length, origin type, and multi-exit discriminator) are set. However, these settings may be misused. For instance, an AS may artificially inflate the length of the AS path to make the route look less attractive to other ASes. This happened on February 16, 2009, when a small Czech provider SuproNet (AS47868) announced its one prefix to its backup provider with an extremely long AS path[44]. It had prepended its own AS number 251 times. While this should not have been a problem, some older BGP routers could not handle the long path, which led to a wave of updates (many times higher than normal), causing instability.
- *Hardware failure*: Although network equipment is designed to operate continuously-on for long periods of time, failures may occur, either due to ageing or for any unforeseen event.
- *Software bugs*: On November 7, 2011, a bug in a (BGP-related) router firmware update caused a network failure at backbone (Tier1) provider Level 3 Communications, resulting in service disruptions for Time Warner Cable, Research In Motion's BlackBerry services and various ISPs, among others[45]. The affected routers experienced core dumps and reloads. A potential related concern is the dominance of few network equipment vendors, where a bug might affect a large portion of the equipment portfolio of a dominant vendor.

3.4 Physical Network Disruptions

Physical network disruptions may have many causes, ranging from fires, to earthquakes, to floods, etc. The precise cause is not of relevance to this report, but the network failures, such as disconnectivity, as a result of the events are. In the following we list three examples of typical physical network

---

[44] *E. Zmijewski, "Reckless Driving on the Internet", February 16, 2009, http://www.renesys.com/blog/2009/02/the-flap-heard-around-the-worl.shtml.*

[45] *F.Y. Rashid, "Bug in Juniper Router Firmware Update Causes Massive Internet Outage", November 7, 2011, http://mobile.eweek.com/c/a/IT-Infrastructure/Bug-in-Juniper-Router-Firmware-Update-Causes-Massive-Internet-Outage-709180/*

disruptions:

- On March 31, 2010, after a flooding at a British Telecom exchange[46], tens of thousands of customers in parts of North and West London were at risk of losing broadband and/or telephony service. Customers needing to make calls to emergency services and who had problems using their phones were advised to do so by using their mobile phone.
- On November 10, 2009, there was a transmission problem[47] with power from the Itaipu dam, one of the world's largest hydroelectric plants. In total 18 Brazilian states were affected and 40% of Brazil's total energy was cut. The involvement of hackers was suspected when the Brazilian National Electricity System Operator (ONS) confirmed that – two days after the massive outage – someone had managed to gain unauthorized access to the corporate IT network. Upon detecting and correcting the security hole, the ONS insisted that the network used for the real-time operation of the national grid be completely isolated from the Internet. This example also illustrates that the resiliency of the Internet interconnection ecosystem relies in part on the resiliency of the energy infrastructure.
- On January 30, 2008, several communications cables (Sea-Me-We 4, Sea-Me-We 3, and FLAG FEA) were (partly) cut, affecting Internet traffic in the Middle East and Indian subcontinent[48]. At least 2840 prefixes suffered outages during this interval, with Egypt taking most of the impact: over 1400 of Egypt's globally routed prefixes suffered outages, roughly 80% of their prefix count.

While one typically protects against physical network disruptions through redundancy (see Section 4), there might be a tipping point when adding more redundancy will actually weaken the system. For example with diversification through Ethernet switches, after a certain number of switches the overall throughput from an extra unit actually decreases for the entire system.

3.5 Malicious Intent

Terrorist attacks, either directly aimed at a network infrastructure or indirectly affecting it, may take out a significant portion of a regional communications infrastructure. Although knowing the details of a physical network is important for risk analysis (see Section 5, Recommendation 4), that information should not be disclosed to the public.

Besides attacks on the physical infrastructure, BGP itself might also be targeted, for instance via Denial-of-Service (DoS) attacks[49], because routes learned from a peering session are withdrawn as soon as the associated TCP connection is lost. Moreover, if a connection is being reset multiple times, the Route Flap Dampening (RFD) protection mechanism of BGP will start suppressing routes. There are many examples of DoS attacks, but mostly the attack is meant to take down some server or disrupt

[46] *http://www.datacenterknowledge.com/archives/2010/03/31/flooded-exchange-disrupts-bt-service/*

[47] *http://www.guardian.co.uk/world/2009/nov/11/brazil-power-cut-rio-madonna*

[48] *http://en.wikipedia.org/wiki/2008_submarine_cable_disruption*

[49] *Y. Zhang, Z.M. Mao, and J. Wang, "Low-Rate TCP-Targeted DoS Attack Disrupts Internet Routing", In Proc. 14th Annual Network & Distributed System Security Symposium, 2007.*

traffic. It is not clear if DoS has ever been used to cut specific BGP connections. Another type of attack is where the data plane is used to attack the control plane[50].

A recent threat[51] of the hackers group anonymous to protest against the Stop Online Piracy Act[52] (SOPA) was to bring down the DNS root servers (in charge of translating IP numbers to human-readable names and vice versa) on March 31, 2012. Although strictly speaking, this would not affect the Internet interconnection, it might have seriously affected the web browsing experience of Internet users. Lately, given the importance of the Internet and its applications to society, much attention is given to cyber security to deal with hacking or DoS attacks.

[50] *M. Schuchard, E.Y. Vasserman, A. Mohaisen, D.F. Kune, N. Hopper, and Y. Kim, "Losing Control of the Internet: Using the Data Plane to Attack the Control Plane", in Proc. of the 17th ACM conference on Computer and Communications Security, 2010.*

[51] *http://pastebin.com/XZ3EGsbc*

[52] *L.S. Smith, "Stop Online Privacy Act", H. R. 3261, October 26, 2011, http://hdl.loc.gov/loc.uscongress/legislation.112hr3261.*

4 Good Practices

This section summarizes and provides an overview of currently used good practices to secure Internet interconnections and provide a resilient Internet ecosystem. The good practices identified within this study are targeted at different components of the interconnection system and come into play at various points in the design and life cycle of operator networks.

A possible classification to describe these good practices is shown in Figure 9, which categorizes 10 different classes of good measures discussed within this study into:

- **Network design practices**, employed during the planning and network growth, redesign and development phases,
- **Technology-related protection techniques**, securing specific protocols and technologies of deployed networks against failures and increase their resiliency against disruptions,
- **Operational principles**, defining procedures and actions to react appropriately during crises situations.

Each section briefly highlights good practices utilized within the industry within their specific application context; Section 5.1 summarizes the identified practices in a descriptive overview.

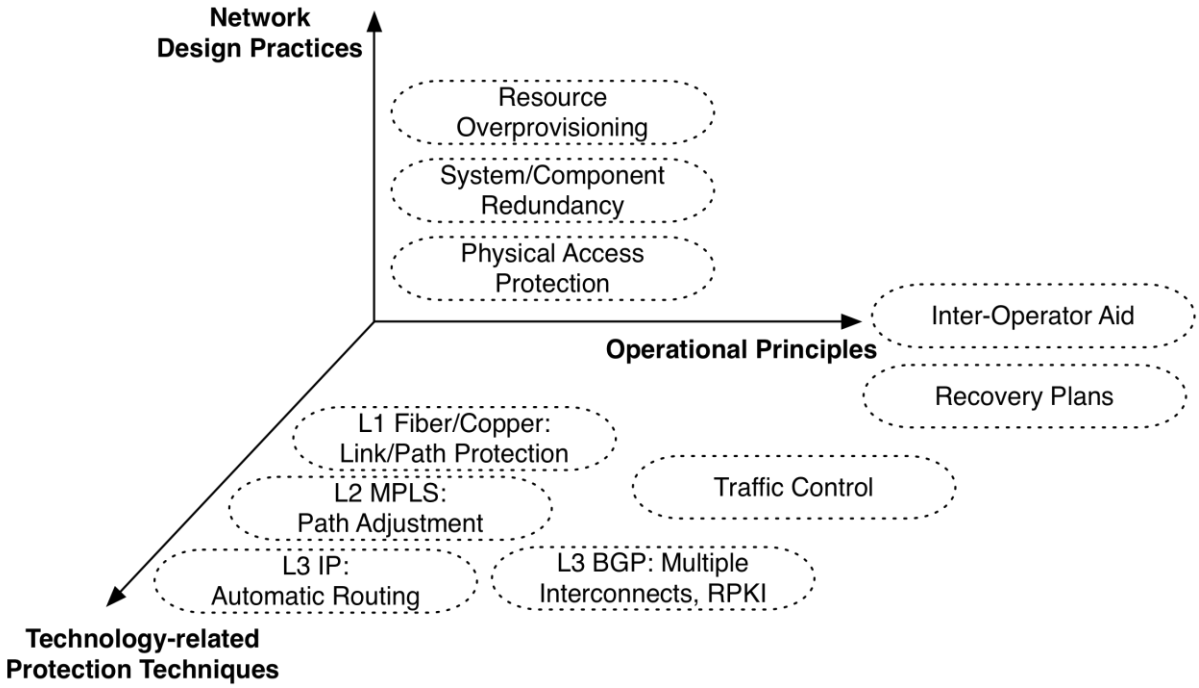


Figure 9: Classification of good practices by their application area and deployment stage.

4.1 Network Design Practices

4.1.1 Physical Asset Protection

When investigating physical protection practices for Internet interconnections, we need to consider the two main components networks need in order to link up, namely a physical connection between the two operators' backbones as well as facilities in which the connection terminates. While for private interconnections these two components can be clearly separated, they are typically fused in case of public peering as the Internet exchange point provides both the location as well as the switching fabric used for the cross connect.

Link Redundancy

In principle, link redundancy (protection) requires setting up several connections between the two end points, each providing sufficient capacity such that the backup path could completely carry the entire traffic load after a failed primary connection. Link protection can be achieved using a variety of techniques; Section 4.2 provides an overview of available methods.

A network is often represented as a graph G(N,L), with N nodes (which for instance represent routers) and L links (which for instance represent optical fibre lines). A graph is said to be connected if there exists a path between each pair of nodes in the graph, else the graph is called disconnected. The notion of connectivity may be further specified as k-connectivity, where at least k disjoint paths exist between each pair of nodes. Depending on whether these paths are node or link disjoint, we may discriminate between node and link connectivity. The link connectivity λ(G) of a graph G is the smallest number of links whose removal disconnects G. Correspondingly, the node connectivity κ(G) of a graph is the smallest number of nodes whose removal disconnects G. We have that

$$\kappa(G) \leq \lambda(G) \leq \delta(G),$$

where δ(G) is the minimum degree (i.e., the number of neighbours) in the network.

For instance, a common design practice for resilient access networks on the other hand is the deployment of ring structures, which offer protection against one link failure, i.e. they are 2-connected. Some operators were found to deploy two parallel running rings, with unidirectional flows, to increase the connectivity.

Connectivity properties may be less obvious when applied to multi-layered networks, like IP-over-WDM. In multi-layer networks, a single failure in a lower layer may induce multiple failures at higher layers. For instance, often several fibre links may share a same duct, while this sharing is not seen at a higher layer. A damaging of the duct may therefore affect multiple links, and we say those links are part of the same Shared Risk Link Group (SRLG).

When network connectivity is considered over time, network availability is used to denote the

percentage of time that the network is connected[53].

Network operators typically plan for at least 2N redundancy, meaning that for every connection there is at least one equal backup available. Some networks increase the protection requirement to 3N for the most important parts of their backbone, for example for links between major metropolitan areas.

Besides the deployment of redundant links, which every network operator participating in the study considered good practice, the desired level of protection from redundancy can only be fully achieved when all connections are truly independent from each other and do not share any common dependency. Such dependencies could arise from a variety of causes, such as primary and backup connections powered by the same electricity feed, both connections (partially) sharing the same duct or routing/amplifier equipment or chassis. In such cases, redundancy could provide insufficient protection as failures may affect primary and backup resources simultaneously.

While every consulted network operator did use redundancy to protect its network, the usage and extent of dependency analyses varied drastically between networks. Some smaller networks did not or only partially perform such risk analysis, because they either achieved sufficient redundancy through other means or the required data was not available. According to the experts, performing a thorough analysis of risks is often challenging as suppliers either do not have or do not want to disclose detailed information about the location and dependencies of their assets due to security reasons. Some network operators have recently begun to require such information from their vendors during procurement and some experts believed that regulation to compile and disseminate such information could provide an added protection.

Larger networks on the other hand can perform a deeper analysis and evaluation of their link protection as they frequently have built or own the physical infrastructure themselves.

The following good practices have been identified in this section:

Good Practice 1: Deploy link protection schemes

Good Practice 2: Investigate common dependencies and shared risk link groups

End Point Protection

The end points of interconnections between individual networks take place in data centres, which themselves encompass a wide variety of practices to increase resiliency. The level of redundancy and

[53] *W. Zou, M. Janic, R. Kooij, and F. Kuipers, "On the availability of networks," in Proc. of BroadBand Europe, 2007.*

protection against typical failures[54] is described by tiers, with specific guidelines what practices must be implemented for a data centre to meet these levels and be certifiable as such. There currently exist three commonly used tier standards, the "Telecommunications Infrastructure Standards for Data Centres" (ANSI/TIA-942)[55], the "Tier Classifications Infrastructure" (UpTime Institute)[56] and the "Criticality Levels" by the Syska Hennessy Group[57]. These standards list specific requirements for the design, construction and operation of facilities; a brief listing of their coverage for the example of the "Tier Classifications Infrastructure" can be found in Table 1.

Table 1: Feature comparison of the UpTime Institute data centre tier standards

	Tier 1 Basic	Tier 2 Redundant Components	Tier 3 Concurrently Maintainable	Tier 4 Fault Tolerant
Power and Cooling Paths	1 active	1 active	1 active, 1 passive	2 active
Redundant Components	N	N+1	N+1	N after any failure
Concurrently Maintainable	No	No	Yes	Yes
Fault Tolerance	No	No	No	Yes
Annual Downtime	28.8 hours	22 hours	1.6 hours	0.4 hours
Availability	99.67%	99.75%	99.98%	99.995%

The experts consulted in the study reported that their operations followed these established or comparable guidelines that may go beyond general data centre standards. AMS-IX for example has extended these available standards and further refined them into a list of 141 minimum baseline requirements for the data centres providing service to the Amsterdam Internet Exchange. Data centres meeting minimum availability levels (99.995%, MTTF = 5 years, accomplished only through a Tier 4 data centre), have implemented technical design specifications, operational requirements and possess business continuity plans can become certified and together with AMS-IX currently 11 other data centres are part of the foundation that facilitates interconnections at Europe's second largest public peering point.

[54] *B. J. Elliot, "A data centre Standard for the Amsterdam Internet Exchange," Fifth Issue, 19/10/2009.*

[55] *Telecommunications Industry Association, "Telecommunications Infrastructure Standard for Data Centers," April 2005.*

[56] *Uptime Institute, "Data Center Site Infrastructure Tier Standard: Topology," 2009.*

[57] *http://www.syska.com/*

4.1.2 System Redundancy and Resource Over-provisioning

In addition to the protection of individual sites and links through backup systems, it is also standard practice to introduce additional resiliency at a more structural, network-design level through system and component redundancy and resource over-provisioning. These practices go beyond the local protection of specific assets (such as particular links or nodes) and are predominantly concerned with enabling the operator to maintain an acceptable level of network operation during crises with a potentially wide-spread impact.

Similarly to other good practices (see Sections 4.1.1 and 4.2.3), the available mechanisms to the network-level redundancy are built upon the main principle of resource duplication, i.e., deploying several independently operating units with sufficient capacity to individually handle the overall typical peak demand. This practice of resource over-provisioning throughout the entire system thus directly protects the network against any unspecific outages as well as unforeseen high demand during crises. A common minimum baseline among network operators for over-provisioning is a factor of 2, meaning that every resource should only be utilized up to a threshold of 50% of its overall available capacity. In case normal utilization crosses this threshold, the affected components are upgraded, at the low end their capacity is doubled, at the high end increased by an order of magnitude to provide an additional safety margin for future growth. Over-provisioning is furthermore intensified to especially critical network parts (as seen either from a revenue, service level, or public utility perspective) to a tripling of resources.

Problems related to equipment (routers) are usually dealt with on two levels. The first action is to quickly restore connectivity, for instance by using alternative paths. The second step is to notify the hardware vendor who likely will disseminate the information widely and possibly (at a later time) present a patch to its customers at risk. Clearly, there is also a responsibility for the network equipment vendors in doing what is possible to avoid equipment errors, for instance by:

- Following sound and secure software engineering and development processes
- Adopting automated patch management
- Deploying well-designed security baselines
- For large-scale border routers, following development standards similar to “mission critical” systems (e.g., software for aircraft)

Here we briefly summarize the recommendations by Kitamura et al.[58] for lessons learnt from the Internet-related aftermath of the Asia Pacific network failures caused by the Taiwan earthquake on December 2006:

[58] *Y. Kitamure, Y. Lee, R. Sakiyama, and K. Okamura, “Experience with Restoration of Asia Pacific Network Failures from Taiwan Earthquake,” IEICE Trans. Commun., vol. E90-B, no. 11, November 2007.*

- First, surviving BGP routers automatically detoured traffic along redundant (mostly narrow-bandwidth) routes. This again argues for a resilient interconnection ecosystem in which the network is well connected with sufficient backup capacity.
- Because the alternative routes chosen by BGP were of low quality, the operators manually reconfigured the routing policies to obtain better quality routes. The authors therefore argued for a QoS-aware BGP.

Kitamura et al. also state that having a good fault detection mechanism that can quickly process all information (and potentially also simulate new configurations) would have facilitated the work greatly. Since communication between NOCs was hampered, the authors also suggested using a decoupled and reliable emergency communications system.

Contrary to the large Internet failures after the Taiwan earthquake, the Japanese earthquake in March 2011, while having devastating consequences, had surprisingly limited impacts on the structure and routing dynamics of the regional Internet. According to James Cowie of Renesys[59]:

> *Of roughly 6,000 Japanese network prefixes in the global routing table, only about 100 were temporarily withdrawn from service — and that number has actually decreased in the hours since the event. Other carriers around the region have reported congestion and drops in traffic due to follow-on effects of the quake, but most websites are up and operational, and the Internet is available to support critical communications.*
>
> *…*
>
> *The engineers who built Japan's Internet created a dense web of domestic and international connectivity that is among the richest and most diverse on earth, as befits a critical gateway for global connectivity in and out of East Asia. At this point, it looks like their work may have allowed the Internet to do what it does best: route around catastrophic damage and keep the packets flowing, despite terrible chaos and uncertainty.*

Redundancy in the Backbone

At the network level, this over-provisioning is also further extended to duplicate resources between different geographical areas and create independent availability regions. In case of national or multi-national network operators for example, enough resources are placed within each key deployment area (such as metropolitan area or customer concentrations) that the network demand could still be accommodated after a complete or partial loss of the network infrastructure within one or two geographic regions.

As a further resiliency mechanism within the core of the network, these key sites are heavily interconnected, typically within a full mesh configuration to mitigate the risk from individual link

[59] *http://www.renesys.com/blog/2011/03/japan-quake.shtml*

failures. Good practices for the placement and planning of links (see Section 4.1.1) and high availability routing solutions (Section 4.2.3) are applied here as well. Note that a full mesh configuration is usually only applied to the backbone and key sites of the network, as a high link density negatively affects the routing convergence times. While adding more links improves the resilience of the network during crises, it is traded-off against efficiency of the dynamic network adaptation processes (for example taking place after the onset of a crises event), which is desired to be as short as possible so that network recovery does not interfere with currently on-going services and thus remains invisible to the end user.

Redundancy in the Access Networks

For this reason, the redundancy in the access parts of a network, which can have extensive geographic coverage and subscriber numbers, remains limited. The most commonly used design strategy for access networks among the interviewed network operators were duplicate ring topologies, each connecting the individual customer locations to one or more backbone nodes. Each using independent hardware and implemented with different directions of traffic travel, a 2N redundancy can also be achieved.

In principle, the level of resiliency provided by these techniques can be arbitrarily scaled up if necessary, but the super-linear near exponential cost increase of additionally introduced availability (e.g. going from “5-9s”=99.999% to “6-9s”=99.9999%) limits the economic attractiveness of more intense protection schemes. All providers investigated in the study therefore design their networks to meet a minimum baseline requirement, and add additional resiliency based on specific customer requirements.

Good Practice 3: Overprovision network elements by at least a factor of 2

Good Practice 4: Introduce independent availability regions capable of securing network operation

4.2 Technology-related Protection Mechanisms

This section lists an overview of existing protection mechanisms proposed in the research literature and discusses to what extent they have been employed by industry. As discussed in the previous sections, network nodes and/or links may go down due to malicious attacks, unintentional fibre cuts, or planned maintenance. Resilient, Fault-tolerant, Survivable, Reliable, Robust, and Dependable, are different terms that have been used by the networking community to capture the ability of a communications system to maintain operation when confronted with such network failures.

Unfortunately, the terminology has overlapping meanings or contains ambiguities, as elaborated on by Al-Kuwaiti et al.[60] We use the term survivable routing to refer to routing mechanisms that, when a component fails and the network is well-connected, may "survive" by finding alternative paths that circumvent the failed component. The path that carries traffic during normal operations is referred to as the primary path, whereas the path that is used to reroute traffic when the primary path fails is called the backup path. Our goal is not to provide an extensive survey on survivability (excellent books and surveys already exist[61]), but rather to capture some of the notions and terminology related to survivability.

4.2.1 Survivable routing techniques

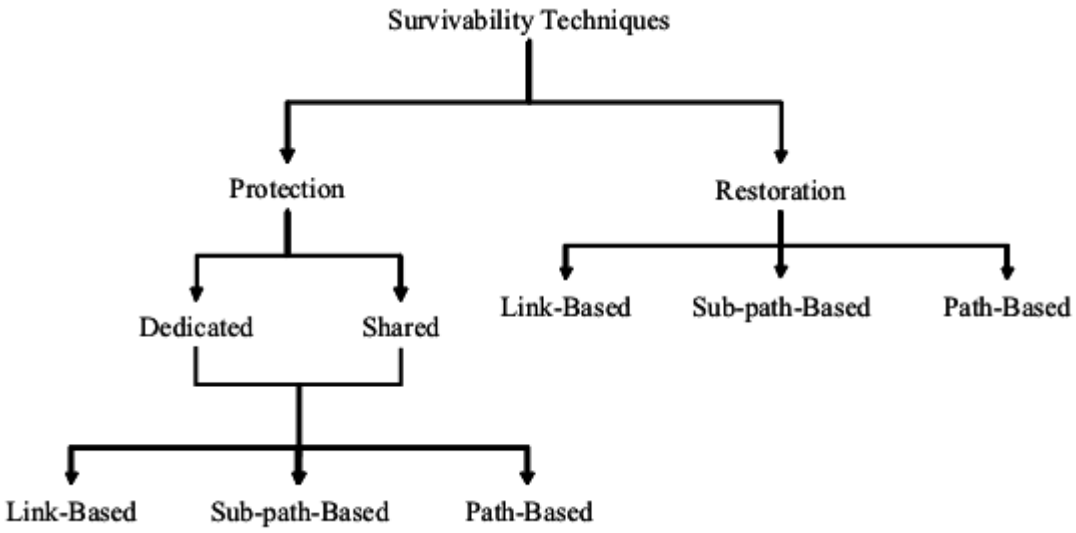


Figure 10: Survivability techniques.

We focus on survivability techniques for short-term failures. Long-term network dynamics, such as nodes/links joining or leaving, are often dealt with through the routing protocols that try to maintain routing tables with the "best" paths toward destinations. For short-term failures that are likely to get fixed within reasonable time, depending on whether backup paths are computed before or after a failure of the primary path, survivability techniques can be broadly classified into restoration or protection techniques (see Figure 10)[62]:

---

[60] *M. Al-Kuwaiti, N. Kyriakopoulos, and S. Hussein, "A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability,"* IEEE Communications Surveys & Tutorials*, vol. 11, no. 2, 2009.*

[61] *R. Bhandari, Survivable Networks: Algorithms for Diverse Routing, Kluwer Academic, 1999. M. Sivakumar, R. Shenai, and K. Sivalingam, Emerging Optical Network Technologies, Chapter in A survey of survivability techniques for optical WDM networks, Springer, 2005.*

[62] *B. Mukherjee,* Optical WDM Networks, *Optical Networks Series, Springer, 2006.*

- *Protection scheme*: Protection is a proactive scheme, where backup paths are precomputed and reserved in advance, and traffic is rerouted along the backup path upon the failure of the primary path.

- *Restoration scheme*: Restoration is a reactive mechanism that handles a failure after it occurs. Thus, the backup path is not known a priori. Instead, a backup path is computed only after the failure in the primary path is sensed.

In general, protection has a shorter recovery time since the backup path is precomputed, but it is less efficient and less flexible. Restoration, on the other hand, provides increased flexibility and efficient resource utilization, but it may take a longer time towards recovery and there is no guarantee that a backup path will be found.

Depending on how rerouting is done after a failure in the primary path, there are three categories of survivability techniques: path-based, link-based, and sub-path based.

- *Path-based protection/restoration*: In path-based protection, a link- or node-disjoint backup path is precomputed and takes over when the primary path fails. A popular algorithm for computing two disjoint paths, for which the sum of both path lengths is minimum, is the Suurballe-Tarjan algorithm[63]. In path-based restoration, a new path is computed between the source and destination nodes of the failed path. If the algorithm fails to find a backup path, the request is blocked.

- *Link-based protection/restoration*: In link-based protection, each link is pre-assigned a local route that is used when it fails, and in link-based restoration, the objective is to compute a detour between the two ends of the failed link for all paths that are using that link. Since link-based protection/restoration requires signalling only between the two ends of the failed link, it has a smaller recovery time than path protection/restoration, which requires end-to-end signalling between the source and destination nodes. However, in link-based protection/restoration, the backup paths may be circuitous, and the backup path is forced to use the same wavelength in wavelength-selective WDM networks since the rest of the primary path is retained[64].

- *Sub-path-based protection/restoration:* The sub-path based scheme is a compromise between path-based and link-based schemes. Thus, in sub-path-based protection, backup routes are

[63] *J. Suurballe and R. Tarjan, "A quick method for finding shortest pairs of disjoint paths,"* Networks*, vol. 14, no. 325–336, 1984.*

[64] *O. Gerstel and R. Ramaswami, "Optical layer survivability-an implementation perspective,"* IEEE Journal on Selected Areas in Communications*, vol. 18, 2000.*

precomputed for segments of the primary path. In sub-path-based restoration, a detour of the segment containing the failed link is computed following a failure. Note that a necessary condition for the existence of a disjoint primary-backup pair is that the network should be 2-connected. If this is not the case, one could opt for finding maximally disjoint paths[65], which can also be considered a sub-path-based protection scheme.

Depending on whether sharing of resources is allowed among backup paths, protection schemes can be of two types: dedicated and shared.

- *Dedicated protection:* In this scheme, links are not shared among backup paths, and are exclusively reserved for a given path request.

- *Shared protection:* In this scheme, backup paths may share (resources on some) links as long as their primary paths do not share links. The shared scheme provides better resource utilization, however it is more complicated and requires more information, such as the shareability of each link. Another notion of sharing is obtained when the capacity of the backup paths is used to transport low-priority traffic during normal network operation. In case of network failure, the low-priority traffic is pre-empted, where needed, by the high-priority traffic that ran on the failed primary path. A backup network refers to a part of the entire network that has been reserved in the network configuration phase solely for routing traffic in case of failures. This backup network should provide protection to any possible traffic pattern and as such is also a shared protection scheme. In case of a link failure, the failed link is bypassed via the backup network.

Ramamurthy et al.[66] have shown that in terms of required capacity, path protection significantly outperforms link protection, and shared protection performs significantly better than dedicated protection. However, path protection is more susceptible to multiple link failures than link protection, and so is shared protection compared to dedicated protection. The choice of survivability techniques is therefore an engineering decision and may depend on the following factors:

- *Type of network*: The underlying network may affect the choice for a survivability mechanism. The most common topologies considered in the literature are ring and mesh topologies. Rings are typical choices for metropolitan area networks, while mesh topologies are commonly used in wide area networks. Survivability schemes in ring topologies have been widely studied due to the fact that they are relatively better understood and some of the schemes, such as embedded protection cycles (p-cycles), can be extended to mesh topologies[67].

---

[65] *R. Banner and A. Orda, "The power of tuning: a novel approach for the efficient design of survivable networks,"* IEEE/ACM Transactions on Networking*, vol. 15, no. 4, 2007.*

[66] *S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM mesh networks,"* Journal of Lightwave Technology, *vol. 21, no. 4, 2003.*

[67] *O. Gerstel and R. Ramaswami, "Fault tolerant multiwavelength optical rings with limited wavelength conversion,"* IEEE Journal on Selected Areas in Communications*, vol. 16, no. 7, 1998.*

- *Centralized vs. distributed*: Survivability techniques can be centralized or distributed. Centralized survivability techniques generally require detailed information on all existing paths as well as all node/link capacities, which may not be scalable. In addition, the centralized approach may lead to a single point of failure. On the other hand, distributed survivability techniques may not be able to take advantage of shareability information, thereby leading to inefficient use of resources.

- *Intra-domain vs. Inter-domain*: As far as survivability is concerned, most work done in the literature focuses on intra-domain networks. Unlike intra-domain networks, in inter-domain networks, there is limited information on the amount and type of traffic exchanged between domains due to scalability as well as privacy requirements. Interconnection and BGP adaptability are most prominent to reach resiliency on the inter-domain level.

- *Different Objectives*: Survivability techniques can be required to satisfy certain objectives, such as:
  - The recovery time (or restoration time), which is defined as the downtime that the connection experiences due to a failure, is important since it determines the amount of data and revenue losses.
  - Capacity utilization, which is defined as the measure of additional backup resources that have been reserved by the specific survivability scheme.
  - Blocking ratio, which can be defined as the ratio of the total number of failed connections over the total number of requests.
  - Restoration guarantee or Quality of Protection (QoP), which is the extent to which a protection/restoration mechanism can restore a failed connection, where dedicated protection mechanisms provide 100% restoration guarantee.

Good Practice 5: Instantiate path restoration techniques

4.2.2 Protection of Intra-Domain Routing

A resilient – in terms of connectivity – network only truly provides resiliency if the network processes running on top of it can actually quickly find alternative paths in case of failures. Most network operators do not use such protection schemes, but rather rely on the longer-term reconfiguration capabilities of routing (intra- or inter-domain).

On occasion, multiprotocol label switching (MPLS) is used to establish specific paths to route or detour traffic. In conjunction with MPLS an MPLS fast reroute mechanism can be used that, as the name suggests, provides the ability to switch over in sub-second time from a failed primary path to an

alternate (backup) path. The fast reroute mechanism is specified in RFC 4090[68], May 2005, and has already been implemented by several vendors. RFC 4090 defines RSVP-TE extensions to establish backup label-switched path (LSP) tunnels for local repair of LSP tunnels. The backup path could either be configured to protect against a link failure or to protect against a node failure. Since the backup paths are pre-computed, no time is lost in computing backup paths or performing signalling in the event of a failure. The fast reroute mechanism as described in RFC 4090 assumes that MPLS (primary and backup) paths are computed and explicitly routed by the network operator. These paths are still often allocated by hand and the application of these methods to LSPs that dynamically change their routes, such as LSPs used in unicast IGP routing or via the various survivability algorithms, is not common practice.

An analogous technique to MPLS fast reroute is IP fast reroute as described in RFC 5714[69], January 2010. The concept of quickly and locally rerouting via backup paths is the same, but the mechanisms employed for the backup routes in pure IP networks are different. RFC 5714 lists three basic categories of backup paths:

1. *Equal cost multi-paths (ECMP). Where such paths exist, and one or more of the alternate paths do not traverse the failure, they may trivially be used as repair paths.*
2. *Loop-free alternate paths. Such a path exists when a direct neighbour of the router adjacent to the failure has a path to the destination that can be guaranteed not to traverse the failure.*
3. *Multi-hop repair paths. When there is no feasible loop-free alternate path it may still be possible to locate a router, which is more than one hop away from the router adjacent to the failure, from which traffic will be forwarded to the destination without traversing the failure.*

Contrary to MPLS fast reroute, IP fast reroute is only starting to be used.

Good Practice 6: On backbones with traffic engineering, use an alternative path method

4.2.3 Protection of Inter-Domain Routing

Key to inter-domain routing resiliency is the establishment of redundancy at multiple physical end points and if possible also across multiple levels. The most fundamental inter-domain protection concept shared by all networks investigated in the study was the establishment of multi-homing, i.e., the presence of at least two distinct uplink connections towards non-local destinations.

How concretely multi-homing is achieved depends on the size and position of a particular network within the interconnection system. Smaller operators such as Tier3 networks that need to purchase

---

[68] *P. Pam, G. Swallow, and A. Atlas (eds.), "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", IETF RFC 4090, May 2005, http://tools.ietf.org/html/rfc4090.*

[69] *M. Shand and S. Bryant, "IP Fast Reroute Framework", IETF RFC 5714, January 2010, http://tools.ietf.org/html/rfc5714.*

the majority of their reachable prefixes from transit providers typically establish multiple transit agreements with different service providers, each transit connection provisioned to be able to carry the full bandwidth requirements of the entire network to provide full backup option in case of failures. To realize the maximum possible resiliency from such setup, the critical dependencies of the upstream providers should ideally be investigated (such as where the transit providers' fibres run along, from which grid their equipment is powered, or where they interconnect), obtaining a comprehensive view of this is however frequently difficult. Larger operators delivering significant portions or even the majority of their traffic via peering agreements, either augment the resiliency of their interconnection portfolio via additional transit agreements to be used as a backup in times of crises or overutilization, or establish their desired level of inter-domain routing resiliency through interconnecting with their peers at several geographically separated points. As discussed before, such multiple-point/multiple-region interconnection practices are frequently an elementary requirement for larger networks to consider a peering agreement in the first place.

In case a network operator has established several interconnection points with another ISP, the BGP protocol provides additional means to manage and thereby strengthen the interconnection. By tuning the individual BGP configuration at each location and influencing through which points traffic should enter or exit the autonomous system, such as the BGP multi-exit discriminators, local preferences or path attributes, providers can obtain a fine level of control on the traffic flows between networks, privileging or relieving particular hardware over other.

Besides providing backups to the actual interconnecting links, it is also common practice to establish resiliency from the inside of an autonomous system at the interconnection border. Here, resource multiplication is also being used, placing several border routers each with enough capacity to support the entire outgoing traffic volume within a load-balanced or hot-standby high availability (HA) - setup.

Finally, providers have the opportunity to actively manage and influence the inter-domain routes they and their customers will utilize based on past and current resiliency levels. Based on active measurements, networks can build up an assessment and track record of "how good" connections via a particular autonomous system are. As each network operator typically has a variety of possible paths to reach each destination, those paths providing only a "poor" quality and lacking the desired levels of reliability and availability can be actively avoided by use of AS path prepending (see Section 3).

The survey by Butler et al.[70] lists several steps that need to be taken to secure BGP:

- Prefix hijacking asks for being able to validate that the AS announcing an address range is indeed the rightful AS, i.e. origin authentication.

[70] *K. Butler, T. R. Farley, P. McDaniel, and J. Rexford, "A Survey of BGP Security Issues and Solutions," Proceedings of the IEEE, vol. 98, no. 1, January 2010.*

- Protecting the connection between two BGP-speaking routers relies on both protecting the underlying TCP connection and the BGP session itself. Cryptographic techniques could be used and several other mechanisms have been proposed like MD5[71] and Generalized TTL Security Mechanism[72] (GTSM), which are fairly easy to implement and therefore often in place.
- Adopting trustworthy routing registries might provide a global view of correct routing information, although setting up such registries is challenging. For example, any provider can register prefixes within these databases, thereby potentially making errors (mall-intended or not), which boils down to a trust problem. The recent RFC 6480[73] describes a Resource Public Key Infrastructure (RPKI) to list, digitally sign, and verify prefixes.
- Monitoring and looking for abnormal behaviour is of vital importance. One such monitoring tool is BGPmon[74]. On the corresponding website operators can define their ASs and prefixes and which policies are being used. The system monitors and notifies if the route announcements change. Over one thousand active users are utilizing the service, ranging from Internet providers, to corporations like banks, large ISPs, and research networks.

Good Practice 7: Multi-home networks with dependency-free connections

Good Practice 8: Utilize BGP attributes to select entry/exit points

Good Practice 9: Create resiliency at the interconnection border

Good Practice 10: Maintain internal records about the availability and reliability of remote routes and use this information to select between inter-domain routes

4.3 Operational Practices

This part discusses the issue of crises management, and particularly investigates how network operators, providers and IXPs deal with challenged network conditions. This operational aspect is viewed from three main perspectives:

- Business Continuity Management: what procedures and structures are to be followed during crises situations,
- Traffic Control: how is network operation altered during crises to cope with shifted requirements and demands,

---

[71] *R. Rivest, "The MD5 Message-Digest Algorithm," April 1992, http://www.ietf.org/rfc/rfc1321.txt.*

[72] *V. Gill, J. Heasley, D. Meyer, P. Savola (ed.), and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)," October 2007, http://tools.ietf.org/html/rfc5082.*

[73] *M. Lepinski and S. Kent, "An Infrastructure to Support Secure Internet Routing," February 2012, http://tools.ietf.org/rfc/rfc6480.txt.*

[74] *http://bgpmon.net/*

- Inter-operator mutual aid agreements: how do networks cooperate to overcome crisis situations.

4.3.1 Business Continuity Management

An operational method towards the mitigation of crises situations and the maintenance of normal operation are business continuity plans. The intensity and detail at which business continuity plans were developed among the networks investigated in this study varied significantly.

Especially smaller service providers were found to avoid intensive pre-planning on potential internal and external operational threats, and rather respond in an ad-hoc manner to a given crisis situation. The type of organizational crises response was found to deliver good results by the interviewed experts within smaller ISPs, as it allowed the network engineers much flexibility in the way they approach and solved challenge situations. Given the presence of small teams where all engineers had good command over all technologies used within the organization, solution approaches could be holistically implemented without the evocation of predefined processes resulting in a fast turnaround time.

Larger service providers on the other hand relied on more formalized processes within their business continuity management, although even here a relatively wide spectrum of options was encountered. On the less formalized end of the spectrum, network operators possessed a list of general scenarios that would threat the normal operation of the organization. Noting that most crises situations were fairly unique and required original problem solutions, these plans served mainly as a basic scaffolding to guide during crises, while still providing large amounts of flexibility to the responsible network engineers to make the final call. On the more formalized end, network operators had developed an entire internal infrastructure for developing and evaluating organizational risks in frequent periodic intervals, and specifying appropriate remediation strategies.

Besides the general solution strategy, also the extent and magnitude of risks considered within the business continuity plans differed considerably among operators, ranging from a clear focus on geographically confined but potentially comparatively frequent operational risks to large-scale disaster scenarios. Again, generally smaller ISPs restricted the level of detail of their crisis planning, considering predominantly only situations where some limited parts of their overall serviced area would be impacted by a problem, for example through a local or regional power outage. It must however be clearly emphasized that this focus was not the result of negligence or lack of planning, but rather a conscious business decision taken based on such provider's service offerings, customer base and customer expectations.

The most elaborate risk assessments were encountered among incumbent operators, which given their specific market position and their implicit role in providing and maintaining a public utility had to

take special precautions. The threats considered within their infrastructure planning and operation covered a broad range of aspects, from local, regional to even national catastrophic disasters disabling infrastructure and service dependencies to assessments of political and economical stability in remote countries where the organization either had assets, partnerships or business interests.

Good Practice 11: Develop business continuity management aligned with the organization's expectations and resources

4.4 Traffic Control and Network Neutrality

Another operational practice commonly used is traffic control. Network traffic control as a generic concept comprises a number of different techniques such as traffic management, traffic shaping, or traffic prioritization each targeted towards a specific issue in telecommunication networks: whereas traffic management techniques are typically targeted to direct and steer the traffic flows in a network, traffic shaping aims at restricting the bandwidth available to network flows. Traffic prioritization techniques on the other hand distinguish between the types of network flows, and assign resources according to a predefined importance scheme. As each method differs in their application area and their actual usage in the field, they will be discussed separately.

4.4.1 Traffic management

Traffic management is the collection of all tools and methods to actively influence and change the paths data flows would take in a network compared to the natural, unmodified routing decisions taken by the network's routing algorithms. Typical scenarios for traffic management techniques are for example explicit routing by avoiding congested areas or simplifying routing between larger metro areas, running networks at higher utilization level (run it "hotter"), resource/cost optimization and the ability to enable Quality of Service (QoS)[75].

Intra-domain

The usage of traffic management techniques within operators' networks was mixed; if a network applied them, they predominantly ran their network on MPLS. Some of the interviewed network operators indicated that the solution obtained via regular IP shortest path routing was "good enough", and the benefits that a traffic management technique such as MPLS would provide did not offset the operational costs of adding and maintaining the extra complexity.

The other group of sampled networks did use MPLS to steer flows in the intra-domain. Although not enough network operators were interviewed to prove this at a statistically significant level, usage of MPLS seemed more accepted in larger and topologically more complex networks. How MPLS routing

[75] *T. Telkamp, "Deploying MPLS & DiffServ,"* Telecommunications Development Asia - Pacific*, 2001.*

policies are applied inside an operator network cannot be distilled into universally applicable, simple patterns. MPLS paths are defined based on the specific network characteristics and requirements of an operator. It is not uncommon to find thousands, if not tens of thousands of custom paths within a network. Beside general manually specified patterns, software tools are in place to plan and assess the utility of particular traffic management options.

This development however is a very recent trend, as sentiments on the low perceived utility of traffic engineering were initially common also among most network operators. In 2010, Cariden Technologies Inc., a company developing IP traffic management solutions, conducted a survey[76] among 49 network providers worldwide. While the study may be over-representing the adopters of traffic engineering (as Cariden's customers by definition have active deployments), the results indicated that traffic prioritization has gained traction: both RSVP-TE and Label Distribution Protocol (LDP) approaches were being used by more than half of the surveyed networks (61% and 57% respectively); providers offering Internet access and business services together were more likely to deploy traffic engineering tools than those targeting the commodity Internet access segment. The results also revealed the very long adoption cycle in the industry. Of those using RSVP for example, only 20% adopted the technology prior to 2003. 50% of deployments were rolled out between 2003 and 2008, and 30% of providers followed after 2009. Thus, in combination with the general usage statistics of RSVP, only about 12% of the operators adopted it in the first 3 years after its introduction[77]. As of 2012, Cariden reports that 8 out of 11 international Tier 1 providers use their products to conduct MPLS-based traffic engineering[78].

Some networks have also been found to apply DiffServ, typically when offering a wider product spectrum with value-added-services such as IPTV or VoIP in addition to vanilla Internet access products. In these cases, DiffServ is being used to distinguish between an operator's different service offerings and for example prioritize any data belonging to real-time applications over the remaining best-effort Internet traffic to provide an acceptable QoS to end-users and assure the continued availability of phone and emergency communication services even during crises and heavy resource overutilization.

Good Practice 12: Prime a traffic management scheme for mixed service infrastructures to prioritize critical traffic during crises situations

Inter-domain

At this point, traffic management techniques such as MPLS are limited to the intra-domain, due to

---

[76] *A. Maghbouleh, "Ten+ years of MPLS: A retrospective deployment survey," tech. rep., Cariden Technologies, 2010.*

[77] *D. Awduche, L. Berger, D. Gan, V. Srinivasan, and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP tunnels," tech. rep., Internet Engineering Task Force (RFC3209), 2001.*

[78] *About Cariden Technologies, http://www.cariden.com/about/customers/, retrieved 2nd April 2012*

their provider-specific definition of labels and lacking means to plan and assign flows across multiple autonomous systems. Towards a solution, a draft architectural specification and requirement proposal was introduced as RFC 4105[79] into the community, but this RFC has to this point not spawned significant activity.

Thus, in the inter-domain area, there exists to this date no solution for fine-grained traffic engineering as provided by MPLS in the intra-domain. The inter-domain BGP routing protocol provides a limited tool set for managing the routes inter-domain traffic takes, and the interviewed operators confirmed that they are using one or more of these available options in practice. Figure 11 displays the general operation principle of BGP path padding, prefix splitting and tagging.

BGP path padding is commonly used to artificially increase the length of routes, as BGP prefers to route traffic via the shorter path. In the example of Figure 11(a) AS1 thus inserts its own AS number into its prefix advertisement to AS2, making the route less attractive to remote ASes which can see and choose between several alternatives, such as AS6. AS6, receiving two routes via AS2 and AS3/AS4 will choose the latter but shorter path. Network operators reported to employ this technique when balancing traffic between multiple transit providers.

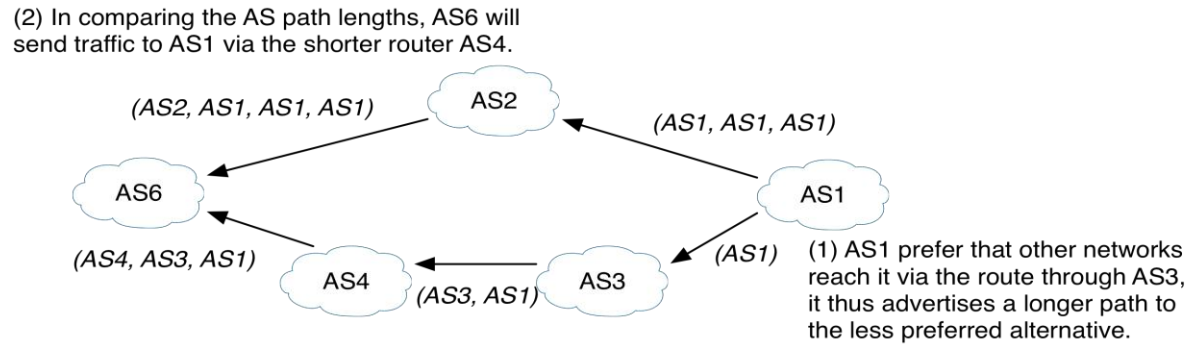
The same technique could in principle be used to further distinguish geographically and steer traffic at a finer level. A network could announce differently padded routes at the individual interconnection points between itself and another network, thereby forcing the peer's routers to bear most of the cost of the traffic flows. This practice is however frowned upon in the industry and all peering policies investigated over the course of this study explicitly require homogeneous announcements at all interconnection points between two peers.

A further method to direct traffic flows in the inter-domain is by use of prefix splitting, as BGP also prefers more specific routes over general routes. This practice is for example used by network operators to establish resiliency through multi-homing, but at the same time offload as much traffic as possible through public peering connection at IXPs. As shown in Figure 11(b), network AS1 could announce all its routes via an upstream transit provider as well as an Internet exchange point. Splitting the prefix into smaller blocks at the IXP will however let those networks reachable via the IXP use the more specific route instead. Traffic will only flow via the transit route if the public peering session at the IXP breaks. Splitting prefixes into a few smaller blocks is used as a common practice in BGP traffic management. Making heavy use of this technique by fracturing a block into a large number of separate announcements is however not viewed benevolently in the community as it leads to an unnecessary increase in the size of the inter-domain routing tables.

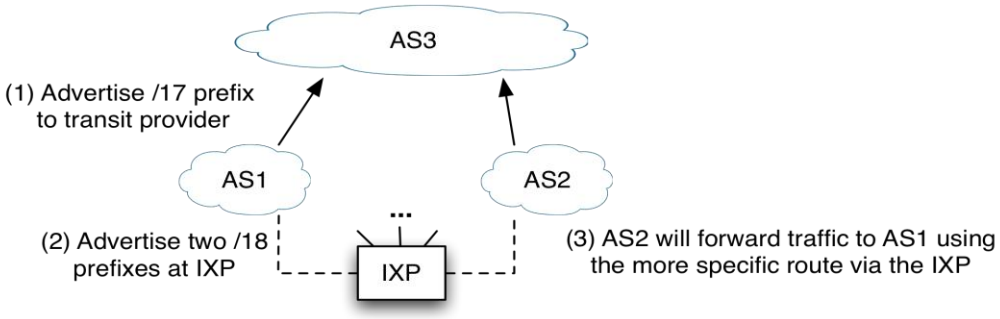
---

[79] *J.-L. L. Roux, J.-P. Vasseur, and J. Boyle, "Requirements for inter-area MPLS traffic engineering (rfc4105)," tech. rep., Internet Engineering Task Force, 2005.*

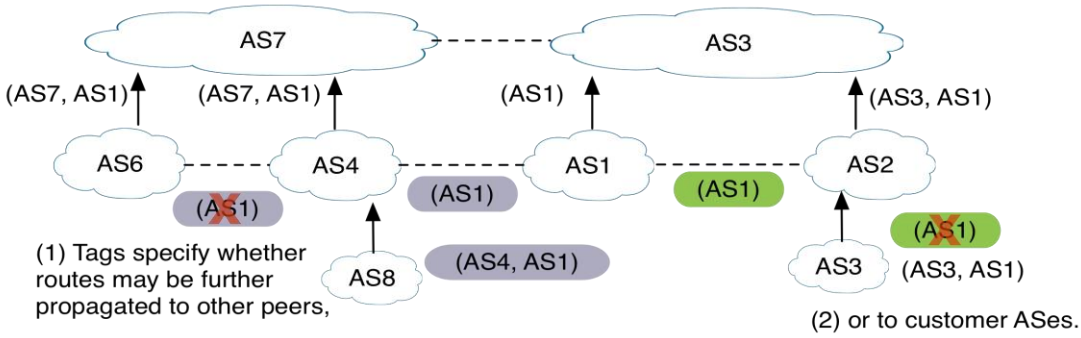
Finally, BGP offers prefix announcements to be tagged, specifying how and where a route might be further distributed. These tags can for example determine that an announcement may only be used directly by a peering AS but not its customers, an AS and its customers, used locally and not be distributed via other peering links or be freely distributed. Figure 11(c) visualizes an example. The tags are used on a per-interconnection basis (BGP session) between two individual providers and have an operator-specific semantics which is typically publicly communicated on a network's website.



(a) **AS path padding.** BGP routing will provide a shorter path, however paths can be made artificially made longer by inserting one's own AS number into the advertised path.



(b) **Prefix splitting.** BGP will prefer a more specific route towards a destination. By splitting up a provider's network into smaller chunks, these alternative routes will be preferred by other ASes.



(c) **Route tags.** Prefix announcements can be tagged, these tags determine how a peer should distribute a route.

Figure 11: Inter-domain traffic management options using BGP.

Good Practice 13: Utilize tags, padding and prefix splitting to control outgoing prefix announcements, and BGP filters to sanitize announcements

Good Practice 14: Register and maintain up-to-date information about AS prefixes within Routing Information Registries

4.4.2 Traffic prioritization

While traffic management techniques are used to alter the path data transmissions will take through the network, traffic shaping and traffic prioritization actively influence the actual transmission itself: traffic shaping limits the amount of traffic or bandwidth a customer can consume and is typically enforced at the edges of a network, traffic prioritization distinguishes between classes of traffic based on preconfigured importance values and allocates resources accordingly.

IP backbone

All experts from network operators participating in the study stated that no traffic prioritization schemes were currently in place within the IP backbone and that at this stage there is no concrete action plan to deploy such techniques. Select participants noted that traffic shaping techniques may be occasionally used to enforce traffic limits with other networks.

Recently, a discussion in the industry on traffic prioritization was initiated by the 2006 Taiwan cable cuts (see Section 4.1.2), where the total bandwidth in and out of the area was greatly reduced and demand heavily outweighed supply. As the non-prioritizing IP networks treated all incoming traffic equally, the vast share of video and streaming traffic on today's Internet also consumed the bulk of the still available resources within the impaired networks. In consequence, the relatively low bandwidth banking and payment services did not receive the necessary allocations and a breakdown of these services was observed[80]. Traffic prioritization techniques could address this problem during future crises in that such data flows could be allocated before servicing the remaining general class, best-effort IP traffic. All participating experts were also very conscious about the possible implications and negative effects the introduction of traffic differentiation and prioritization might have on the ecosystem itself.

The reports by experts on the usage of traffic shaping and traffic prioritization matched the outcomes of previous measurement studies[81] on the usage of such practices among Internet service providers. A number of studies conducted on the m-lab platform have concluded that traffic shaping techniques are applied within a number of U.S. ISP networks. There is no mention of any usage of traffic shaping techniques in Europe, although an upcoming unpublished study by the Body of European Regulators for Electronic Communication reports usage of such practices in Europe[82].

---

[80] *Sandvine Inc., Global Internet Phenomena Report: Fall 2011, Sep. 2011.*

[81] *Partha Kanuparthy and Constantine Dovrolis, "DiffProbe: Detecting ISP Service Discrimination," in Proc. of the IEEE Conference on Computer Communications (INFOCOM) 2010.*

[82] *Body of European Regulators for Electronic Communication, BEREC preliminary findings on traffic management practices in Europe show that blocking of VoIP and P2P traffic is common, other practices vary widely, Press Release, March 6, 2012.*

Mobile data networks

In contrast to fixed IP networks, traffic shaping and prioritization is frequently practiced within mobile data networks. Here, the increased usage of data-intensive smart phone applications combined with a limited supply of bandwidth by base stations, create occasions for network operators to limit and throttle access and traffic within these networks to maintain their proper functioning. These traffic shaping techniques are conducted both fully automatic by the network as well as in severe interventions manually. In the UK for example, mobile users are being assigned to 10 user classes and, once network load surpasses a critical loading threshold, service to users within a particular class is temporarily blocked. In certain crises situations, for example during the 2005 London terrorist attacks, manual restrictions for in- and outgoing flows were enabled to prevent a breakdown of the network from increased volume while still providing emergency services.

4.4.3 Mutual aid agreements

Despite the protection of networks via technical and operational good practices as discussed above, there may be extraordinary circumstances and adverse operating conditions when the means available within a single network operator's reach fail to meet the service demands put on the network. To recover from such events, affected networks may temporarily receive technical assistance from unaffected neighbouring operators.

All network operators participating in the study, ranging from small local, regional to international backbone operators, mentioned inter-operator assistance as an established good practice to overcome crises situations and confirmed that they had either given or received such disaster relief themselves in the past, or that such procedure was a feasible practice they could draw upon in case of a future hypothetical crisis.

In case of such inter-operator aid, one option for the assisting operator is to temporarily take over part of the network traffic that cannot be carried anymore by the impaired network, and deliver it via its own peering or transit connections. For this method to be effective, networks should and are interconnecting at a number of geographically dispersed interconnection points as well as have sufficient bandwidth available to temporarily transport the extra load. Both multiple-point peering and over-provisioning of network resources (typically by a factor of 2) are considered standard practices within the industry, however *mutual* aid can only be feasible between networks of similar size and traffic volumes. This was also reflected in expert interviews - network operators only engage in such practices or consider parties as possible candidates for mutual aid when the other autonomous system is comparable in footprint and traffic volume and thus assistance could be safely provided without impairing one's own operation.

While in principle available as a means to respond to crises situations, inter-operator mutual aid was

viewed as an extraordinary measure by network engineers, a resource that may and should only be tapped as a last resort to ensure connectivity and proper functioning of the network. It was considered prime duty to plan for and introduce means to respond to all foreseeable events within one's own network and realm, and fall back on outside help when a crisis exceeded planned situations and disaster recovery preparations.

Regardless of their potential importance to crisis response, the study did not reveal any formalized mutual aid agreements in the field. This could be attributed to two main factors:

1. The practice of providing relief to distressed networks is primarily conducted between network engineers and not an initiative at higher managerial levels. These network engineers are tied together by a strong honour code and professional ethics with the primary goal of doing everything to "keep the Internet running", even when temporary help created monetary or organizational expenditures and inconveniences at one's own organization. Network engineers reported that formal contracts were not needed to enable such practices: The community is small enough for the respective network engineers to be directly acquainted, and these personal contacts are the driving factor for inter-operator crisis response. Thus, being present and well connected within the community of network engineers can be seen as the underlying foundation for this good practice.

2. Formalizing inter-operator aid agreements will create obligations that might result in legal liabilities or penalties if they cannot be granted. This is further aggravated by that fact that the protective and monetary value of such aid can neither for the giving nor the receiving party be easily quantified, and such agreements may therefore not be supported by upper management or an operator's controlling unit. In addition, as mutual aid by definition and design will only appear in crisis situations where the magnitude of the crisis was larger than planned for and established good practices of resiliency have already failed, the system and resources of the operator who would in principle come for aid might as well be constrained. Whether or not aid could actually be provided as contractually defined will remain uncertain. Thus, instead of specifying terms and conditions as done in service level -, transit - or peering agreements, inter-operator aid agreements are viewed as unspecified "gentleman's agreements". If drawn upon, it will be quickly and un-bureaucratically provided, and if necessary compensation for incurred expenditures will be negotiated afterwards.

Good Practice 15: Foster strong relations with compatible providers to give and receive assistance during crises

5 Recommendations and Good Practices

This section lists good practices and recommendations to create and enhance resilient Internet interconnections. While the former is aimed at concrete technical and organizational action items implementable within, the latter describes medium and long term efforts that may be undertaken by policy makers, standardization bodies, industry and academia to support and strengthen the interconnection ecosystem.

5.1 Good Practices

Good Practice 1: Deploy link protection schemes
The connections between two network points should be protected through the introduction of a backup link. Several strategies such as link, sub-path and path protection are available to tune the level of resiliency while adjusting for the costs of this extra resource.

Good Practice 2: Investigate common dependencies and shared risk groups
Protection of the deployment of backup components can only be fully achieved if the primary asset and its backup do not share any common dependency and could potentially fail simultaneously in a challenge situation. An investigation of the dependencies and the shared risk groups between the network components can to the most part eliminate these issues. Obtaining a detailed enough view might however be difficult in cases of leased infrastructure.

Good Practice 3: Overprovision network elements by a factor of 2
Loading network elements only up to a maximum utilization of 50% provides additional protection against the unforeseen failure of a (potentially otherwise unprotected) network component. Higher levels of resource overprovisioning should be used in especially vulnerable or important parts of a network.

Good Practice 4: Introduce independent availability regions capable of securing network operation
The main footprint areas of the network (as measured in customer base, traffic volume, revenue etc.) should be designed in such a way that they can operate independently from each other so that a failure of a key site only affects the traffic going in and out of this particular zone without impairing any other flows within the network. To achieve this, the resources at the key sites should be dimensioned in such a way to be able to withstand the failure of at least one main area. Critical network equipment is best deployed in a high-availability setup, such as a load balance or hot standby configuration.

Good Practice 5: Instantiate path restoration techniques
For situations where proactive protection schemes have failed, path restoration techniques can provide a dynamic response to return to unperturbed operating conditions. Given the lower response

time (and potentially unsuccessful result) of path restoration techniques, such mechanisms are used in conjunction with other resiliency techniques.

Good Practice 6: On backbones with traffic engineering, use an alternative path method
Networks augmenting vanilla IP-routing tables with label-switched path approaches should protect these specifically defined paths via alternative path methods, capable to react within the timescale appropriate for the label-routed services.

Good Practice 7: Multi-home networks with dependency-less connections
Autonomous systems should secure the reachability of their prefixes by investing in multi-homing, ideally using different providers not sharing any common dependency such as power grids, fibres, exchange point equipment etc. Interconnection to other networks should be conducted at as many geographically dispersed points as possible and economically feasible.

Good Practice 8: Utilize BGP attributes to select entry/exit points
The load of network elements and the distribution of incoming and outgoing traffic can be dynamically managed through the manipulation of BGP attributes at the various interconnection points between two autonomous systems.

Good Practice 9: Create resiliency at the interconnection border
Infrastructure required to maintain inter-domain connectivity should be redundantly available and placed in a high-availability configuration.

Good Practice 10: Maintain internal records about the availability and reliability of remote routes and use this information to select between inter-domain routes

While many prefixes may be reachable via several possible paths, not all paths offer the same level of stability and resiliency. By monitoring the availability and reliability of individual routes and influencing the route selection within BGP through path padding, routing can be conducted over the most robust routes available.

Good Practice 11: Develop business continuity management aligned with the organization's expectations and resources
A business continuity management plan exploring the potential threats to the normal operation of the network can help identify unmitigated vulnerabilities. The extensiveness of such plan should match the organization's target market and product description, customers' requirements, expectation and available personnel and financial resources, and be regularly exercised.

Good Practice 12: Prime a traffic management scheme for mixed service infrastructures to prioritize critical traffic during crises situations
In case of networks servicing a variety of different applications, a traffic management scheme can help

to reserve the minimum required share of resources to maintain critical services during crises and adverse operating conditions, such as emergency or telephone services.

Good Practice 13: Utilize tags, padding and prefix splitting to control outgoing prefix announcements, and BGP filters to sanitize announcements

Route tags, AS path padding and prefix splitting provide opportunities to manage the routes other autonomous systems will take to deliver incoming traffic, and provide valuable tools for load balancing and establishing backup connections only to be used during crises. Both incoming and outgoing prefix announcements should always be sanitized and checked against routing information databases to prevent accidental prefix hijacking and prefix leakages.

Good Practice 14: Register and maintain up-to-date information about an AS prefixes within RIR

A successful avoidance of future prefix hijacking incidents – which were the predominant disruption in the interconnection system in the recent years – depends both on the wide-spread usage of BGP prefix filtering and the availability of high quality up-to-date information about the prefixes actually in use by an AS. Currently, such information is not comprehensively accessible.

Good Practice 15: Foster strong relations with compatible providers to give and receive assistance during crises

Inter-operator aid during crises is typically organized spontaneously at the level of network engineering staff, among the members of a small community of personal acquaintances. Establishing and maintaining strong relations and good visibility within this community is key to give and receive assistance among providers, and should be supported both by the ISP's management and the organizational culture.

5.2 Recommendations

a) Policy/Public Sector

Recommendation 1: Avoid policy making against temporary developments and market culture

Successfully changing the Internet ecosystem via regulation is a challenging task for two reasons: First, the process of creating new regulation is comparatively slow to the speed at which trends are evolving in the Internet. Within just a few years, the industry has seen the rise, feared domination and subsequent fall of particular individual actors, strategies and entire business models - ranging from single players to market segments such as Tier1 providers and in the recent past CDNs. Considering the cycle time of these trends in the past, regulation addressing them would often not be able to keep up and stimulate change. Second, regulators should align their efforts with the private sector, as in the past policy decisions working against the established practices of the market were eventually circumvented by the market actors. The recommendations put forth in the following therefore built on

these past lessons learned and instead suggest long term strategies to increase interconnection resilience, by building an environment and conditions to trigger change together with the market.

Recommendation 2: Establishing a platform and community to disseminate important information

There exists a wide spread in the adoption of current good practices, techniques and tools within the ISP and IXP community. Some providers are heavily investing into these while others concentrate only on a few items. Some experts in our study attributed this heterogeneity to the different availability of personnel, financial resources and exchange opportunities across the service providers: while large incumbent operators have dedicated staff for standardization activities, budget for R&D, attend operator meetings and are in close contact with other large-scale networks. Such resources and in consequence access to important information distributed via these channels is often not in reach for smaller networks. The European Commission could help by establishing a platform and foster a strong ISP/IXP community aimed at the timely dissemination and evangelization of important information, such as current good practices, training or vulnerability/exploit information.

Recommendation 3: Stimulate the usage of Route Information Registries

Several important proposals to secure the interconnection system can currently not be implemented due to the lack of information, such as a comprehensive record of prefix assignment to autonomous systems which would enable a number of protection mechanisms. The participation in such documentation efforts could be stimulated by the public sector, for example by defining a road map for the European-wide adoption of these techniques, encouraging European ISPs to engage in a code of conduct to populate and use such tools, or even by requiring the implementation of such practices by all ISPs with a European presence.

Recommendation 4: Foster change by creating market pull

Instead of intervening by regulation, the public sector could bring change to the ecosystem by leveraging its buying power and creating market pull. By for example requiring that all suppliers to the government need to have implemented a certain minimum baseline of good practices (for example use and participate in RIRs – see recommendation 2), ISPs would be enticed to step up their efforts to meet this new demand. In circumstances where market demand is not enough to stimulate change, the public sector could additionally fund initial investments in specific infrastructure if needed.

Recommendation 5: Develop an infrastructure to bring together and safely share information on networks, infrastructure and network-related incidents

Currently much information valuable to peers or to the ISP community in general is not directly shared between operators in order not to expose existing vulnerabilities or the concrete protection steps taken in securing their assets. Furthermore, the implementation of some good practices is hindered by the non-availability of infrastructure-related information. The interviewed networking experts did not

express any concerns about introducing a sharing scheme, as long as confidentiality (and anonymity in dissemination) can be ensured. The public sector could develop a platform that would enable experts to safely share information, practices and incident-related information and remedies.

Recommendation 6: Define acceptable practices for traffic prioritization

Traffic prioritization offers benefit in maintaining critical services during crises, but once established may also find applicability in other application contexts. As traffic prioritization automatically implies a discrimination of particular content, a debate needs to take place to define acceptable practices under which circumstances and by what means such differentiation and discrimination is acceptable.

b) Standardization

Recommendation 7: Establish a traffic prioritization standard across multiple operators or administrative zones

Traffic engineering techniques are being employed by operators to manage and mitigate issues intra-domain. If similar techniques are to be used towards a more resilient interconnection system, mechanisms supporting cross-operator mitigation actions will need to be developed.

c) Industry

Recommendation 8: Foster and protect engineering community efforts to deal with multi-operator outages

Current inter-operator remediation mechanisms rely to a large extent on informal personal relationships between network engineers of individual operators. This has proven to be effective in crisis management, and has been voiced by engineers, designers and managers to be a preferred mode of operation as incident remediation typically requires a significant amount of improvisation and creativity. Industry should protect and support these structures.

Recommendation 9: Implementation of inter- and intra-domain good practices

Quick wins in the resiliency of the interconnection system could be achieved by universally adopting a small number of good practices, such as the by default filtering of routes or multi-homing. For example, most prefix hijacking incidents in recent years where due to accidental mis-configurations, which could have been avoided by an industry-wide adoption of route checks. European networks could assume a leadership role to drive the worldwide adoption of this and other protection schemes by setting a mid-term roadmap by introducing requirements on their peering and transit partners and thereby creating incentives for other networks to follow suit.

d) Research

Recommendation 10: Develop techniques to accurately measure the structure of the Internet

Current methodologies used to map the interconnection system rely on BGP monitors of which too few are deployed to derive a broad view of the Internet. Additionally, the viewpoint of these monitors is heavily skewed by their placement in access networks, missing private peering connections by Tier1 and Tier2 networks. As most peering relations between Tier1 and Tier2 networks are however private, the actual structure of this system remains largely unknown and topology-driven research initiatives might arrive at incorrect conclusions. Improved techniques are necessary to be able to map out the networks of providers and their interconnections to gain a deeper understanding of the Internet structure.

Recommendation 11: Investigate the structural properties of the Internet in a changing provider ecosystem

What can be expected for the future importance of Tier1 operators in providing global connectivity? To what extent will the resiliency of the Internet in terms of multi-path and spare capacity be affected by changing peering habits? More effort is needed in a technological-economical investigation of the Internet ecosystem to improve the assessment of ongoing and future trends on the development of the provider landscape.

A Tier1 - Peering Requirements

Table A.1: Peering requirements for 10 Tier1 network operators

Network	Public Policy	Network Requirements		Peering Requirements			Operational Requirements	
		Backbone Speed	Min. Connections	Min. Volume	Traffic Ratio	Routing Req.	Excl. Cust.	24-7 NOC
AT&T	yes	10 Gbps	3 points	avg. 7 Gbps	< 2:1	yes	yes	yes
AOL	yes	> 10 Gbps	4 points with >1 Gbps	>970 Mbps	< 2:1	yes		yes
Verizon	yes	>0.6-10 Gbps	>50% netw. overlap, connect at 8 pts in US	>0.03-1.5 Gbps	< 1.8:1	yes		yes
CenturyLink	yes	>40 Gbps	6 points with >1 Gbps	>10 Gbps	< 1.5:1	yes		yes
nLayer	yes	> 10 Gbps				yes		yes
VSNL	no		yes			yes		yes
NTT	no		yes			yes		
Level 3	yes	> 10 Gbps	5 US regions		yes			yes
Deutsche Telekom	no							
AboveNet	yes	> 5 Gbps	2 locations, > 1 Gbps	> 50 Mbps	< 2:1	yes	yes	yes

P.O. Box 1309, 71001 Heraklion, Greece
www.enisa.europa.eu