View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Lancaster E-Prints

It Bends but Would it Break? Topological Analysis of BGP Infrastructures in Europe

Sylvain Frey, Yehia Elkhatib, Awais Rashid, Karolina Follis, John Vidler, Nicholas Race, Christopher Edwards
Security Lancaster Research Centre, Lancaster University, United Kingdom
{*s.frey, y.elkhatib, a.rashid, k.follis, j.vidler, n.race, c.edwards*}*@lancaster.ac.uk*

***Abstract*—The Internet is often thought to be a model of resilience, due to a decentralised, organically-grown architecture. This paper puts this perception into perspective through the results of a security analysis of the Border Gateway Protocol (BGP) routing infrastructure. BGP is a fundamental Internet protocol and its intrinsic fragilities have been highlighted extensively in the literature. A seldom studied aspect is how robust the BGP infrastructure actually is as a result of nearly three decades of perpetual growth. Although global black-outs seem unlikely, local security events raise growing concerns on the robustness of the backbone. In order to better protect this critical infrastructure, it is crucial to understand its topology in the context of the weaknesses of BGP and to identify possible security scenarios. Firstly, we establish a comprehensive threat model that classifies main attack vectors, including but non limited to BGP vulnerabilities. We then construct maps of the European BGP backbone based on publicly available routing data. We analyse the topology of the backbone and establish several disruption scenarios that highlight the possible consequences of different types of attacks, for different attack capabilities. We also discuss existing mitigation and recovery strategies, and we propose improvements to enhance the robustness and resilience of the backbone. To our knowledge, this study is the first to combine a comprehensive threat analysis of BGP infrastructures with advanced network topology considerations. We find that the BGP infrastructure is at higher risk than already understood, due to topologies that remain vulnerable to certain targeted attacks as a result of organic deployment over the years. Significant parts of the system are still uncharted territory, which warrants further investigation in this direction.**

1. Introduction

The Internet has become a critical infrastructure. It is now essential to fuel the world's economy, to the functioning of developed societies, not to mention its pervasive importance in the daily and social life of billions of people. Disrupting it would have devastating consequences, similar in importance to disruptions to the power grid or to transport networks. So far, however, Internet disruptions have been local and temporary, the resilience of the network of networks is often cited as a remarkable emergent phenomenon. This study approaches this idea from multiple perspectives and shows that more investigation and security improvements are necessary for the Internet to be considered truly resilient against targeted attacks.

The Border Gateway Protocol (BGP) is the de facto standard for routing between large IP networks, called Autonomous Systems (AS). ASs advertise their existence and the range of addresses they own to the rest of the Internet. BGP ensures all other ASs know about the various subnets and how to reach them. Without BGP, each sub-network would be isolated and unreachable.

Since its introduction in the 1980's, BGP has grown to become an important part of running the Internet's core. The infrastructure, both logical – i.e. AS routing tables – and physical – i.e. BGP routers and the traffic highways connecting them – is thus quite a critical one. However, like many other key internet technologies, BGP as a protocol was not designed with security requirements. There is plenty of evidence to show that it is intrinsically fragile (e.g. [1]–[3]) and significant disruptions to parts of the backbone raise the question of its robustness as a whole. The BGP infrastructure has grown over the years into what is now a particularly complex system. Such an organic growth results notably in a scale-free topology that is known to be particularly resilient against random failures but remains vulnerable to targeted attacks [4].

This paper aims at shedding some light onto the security of BGP infrastructures in Europe. Our methodology is reproducible for other portions of the network, and the findings of the paper are of high importance to network operators and regulators both within and outside European countries. Understanding the Internet as a critical infrastructure from a security perspective is essential. Such a relatively young system that has grown mostly unsupervised is a tempting attack target. Technical operators are often limited to the narrow scope of their own sub-systems and require a better knowledge of the overall picture.

To our knowledge, this study is the first to combine a comprehensive threat analysis of BGP infrastructures with advanced network topology considerations. On the one hand, existing work in the domain has extensively investigated the logical topology of the BGP infrastructure. Important properties of the network (e.g. its hierarchical, scale-free nature) have been identified by several contributions (cf. [5], [6]). However, such analysis offers a network perspective and does not take security aspects into consideration. On the other hand, works proposing security-driven analysis of BGP topologies often focus on a particular type of attack, such as BGP hijacks (e.g. [7], [8]), focusing solely

on the vulnerabilities posed by the protocol specification and not how such vulnerabilities interplay with real world deployments outside of high profile incidents. This study considers the whole spectrum of possible threats to put topology analysis into perspective. In this paper:

- We establish a comprehensive threat model (section 3) encompassing known BGP weaknesses but also physical vulnerabilities and non BGP-specific attack vectors. We also discuss existing improvements that can address these vulnerabilities.
- We investigate the topology of backbone infrastructures (section 4) in the light of this threat model and we highlight possible disruption scenarios (section 5) that leverage topological weaknesses via multiple attack vectors, for variable attack means.
- We propose strategies for improving the topology (section 6) of the backbone and enhance its robustness and resilience against such targeted attacks.

2. Background

We begin by giving an overview of the BGP protocol and presenting backbone infrastructures.

2.1. BGP Route Announcement

Individual computers, both servers and clients, are connected to the wider Internet using a web of various networking devices, including switches and routers, that ensure delivery of traffic to its intended destination. When a network of these devices is managed by one organisation, such as an Internet Service Provider (ISP), they are commonly referred to as an *Autonomous System* (AS). Routers within a single AS would be under a unified administrative control, following the same routing policy and using the same routing protocol(s).

Different protocols are in use for the intercommunication of routers either within an AS or in between different ASs. BGP is the main protocol used for routing inter-AS traffic [9]. BGP routers – or gateways – communicate with each other over TCP and advertise accessible addresses in their respective ASs and available routes towards other ASs.

Accessible addresses are announced via **IP prefixes** that combine an IP address (for instance 1.0.0.0 for IPv4) and a prefix length measuring a number of bits in this address. For instance, 1.0.0.0/8 identifies all IP addresses of which the first 8 bits match the first 8 bits of 1.0.0.0, i.e. $\{1.0.0.0, 1.0.0.1, 1.0.0.2, \ldots, 1.255.255.255\}$. The longer the prefix, the smaller the corresponding set of IP, the more specific the prefix is. An AS can therefore delegate one of its subprefixes to another AS via commercial agreements. For instance, in Fig. 1, AS 2 is the owner of prefix 2.0.0.0/8 and delegates the 2.3.0.0/16 prefix to AS 3.

BGP announcements combine a prefix with a **path** towards the destination AS. Gateways advertise a direct path towards their own AS, and they can also learn from other gateway announcements and advertise indirect routes

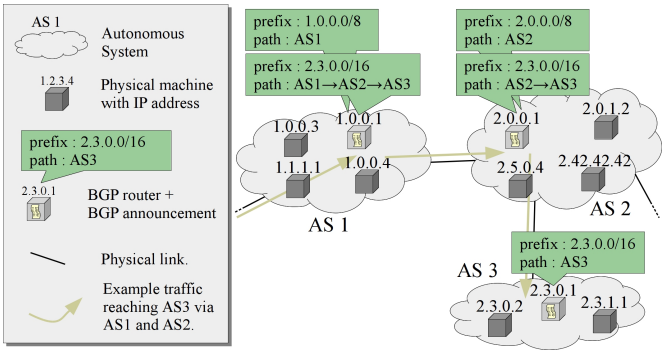

Figure 1: Example BGP routing: traffic reaches AS3 via a route advertised by AS3, AS2 and AS1 successively.

towards remote prefixes. For instance, in Fig. 1, AS3 announces the 2.3.0.0/16 prefix as a direct route toward itself. Its provider AS2 announces the same prefix with the path AS2→AS3. AS2's peer AS1 then announces the prefix via its own network with a AS1→AS2→AS3 path, and so on.

Announcements are thus propagated from neighbour to neighbour, according to network topologies and commercial agreements that can be divided roughly into two categories:

- *Peering* agreements are symmetrical and result in two ASs advertising each other's routes and sharing each other's traffic.
- *Customer/provider* agreements introduce an asymmetrical relation where one AS provides access to another – the latter *buying transit* from the former.

2.2. BGP Path Selection

BGP routes are being constantly advertised, updated or withdrawn. Each BGP router monitors these announcements and selects the best routes to use amongst all possible alternatives, according to a number of criteria. The path selection algorithm first ensures that a number of basic assumptions hold, such as checking that the next hop in a path has a valid route. Then the algorithm attaches weights to the paths in order to rank them, preferring locally originated and shortest paths, and longer (i.e. more specific) prefixes. Rules could be applied to the said path selection process in order to influence it in some way in order to balance cost, performance, reliability and other factors. This is known as a *routing policy*, which is not part of the protocol but a deployment configuration choice.

The routing policy allows each AS operator to set their own criteria for weighing paths. A common objective within routing policies, for instance, is to reduce costs through favouring peering routes (usually free of charge) against costly transit routes. Latency and congestion are other possible criteria that network engineers can take into account when weighing paths. Gateways then select the best paths accordingly, and install such paths in their forwarding tables which are subsequently used to route traffic.

At this stage, it is worth noting that there is no explicit path negotiation in BGP: communication between BGP routers is limited to path and prefix announcements only, and each gateway administrator is responsible for establishing their own routing policy and BGP announcements. In particular, there is no authentication procedure and no path validation mechanism that would allow a router to verify a path beyond its direct connections with other gateways. Certain heuristics commonly used by path selection algorithms – such as preference towards short paths and specific prefixes – can easily be abused, as shown in the next sections.

2.3. Logical Topology of the Backbone

The logical topology of BGP infrastructures has been extensively investigated by network analysts. Publications in the domain propose taxonomies that differ in their denominations, but globally it is agreed that BGP infrastructures follow a structure in three layers [7], [8], [10]:

- *Tier-1* (also: *transit*, *core*) ASs that can access the entire address space without buying transit from any other AS. The list of tier-1 networks has been relatively stable over the last decade, major networks entering the club or leaving it depending on commercial agreements (cf. Table 1).
- *Tier-2* ASs buy transit from at least another (tier-1 or tier-2) AS and peer with or provide transit to other ASs. The largest Tier-2 ASs are comparable in size and in importance to tier-1 networks, at least in the regions where they operate.
- *Tier-3* (also: *leaf*, *edge* or *marginal*) ASs connect to the Internet by peering with or buying transit from higher-tier ASs exclusively.

Name	Country	AS number
AT&T	US	7018
CenturyLink	US	209, 3561
Cogent	US	174
Deutsche Telekom	Germany	3320
Level 3	US	3356, 3549
NTT	Japan	2914
Sprint	US	1239
Tata	India	6453
Telecom Italia Sparkle	Italy	6762
TeliaSonera	Sweden - Finland	1299
Tinet/GTT	US - Italy	3257
Verizon	US	701, 702, 703

TABLE 1: Tier-1 Autonomous Systems as of 2015.

It is generally recognised that the graph of BGP nodes follows a scale-free distribution [5], [6] typical of organically-grown networks, where newcomers connect preferably to existing high-degree nodes. This means in particular that high-degree tier-1 ASs concentrate a majority of the edges of the graph.

2.4. Physical Topology of the Backbone

ASs connect physically with each other at *colocation points* where their routers can directly exchange traffic via dedicated high-performance infrastructures. Historically, international carriers provide transit at private *Points of Presence* (PoP) whereas public *Internet eXchange Points* (IXP) are dedicated to peering. In practice the distinction is not always as clear, notably in Europe where IXPs are now becoming mainstream transit providers [11], [12].

Colocation points are physical hubs linked by traffic highways, and notably submarine cables (Fig. 2). These large-scale, international infrastructures are owned by tier-1 and large tier-2 ASs, either as a monopoly or as a consortium. These networks are exploited either via AS-level transit agreements or via direct leasing. Submarine cables constitute major choke points of the physical infrastructure, for instance, in Europe:

- A limited number of cables (a dozen) join the two sides of the Atlantic ocean: despite the complexity of routing tables, all traffic between Europe and America eventually goes through one of these cables.
- All cables tend to converge to the same colocation points: London for instance is visited by a majority of transatlantic and UK-Europe cables (Fig. 3).

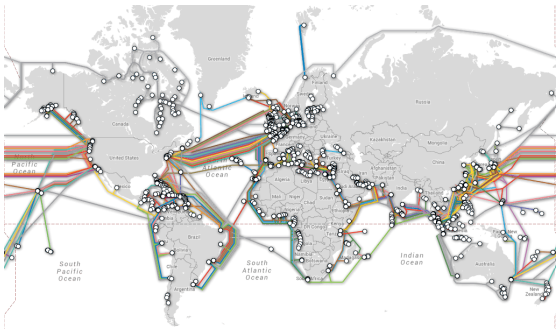

Figure 2: Submarine cables, current and planned [13].

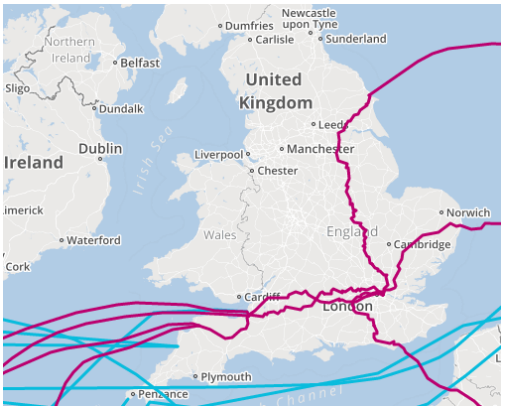

Figure 3: Major carrier landlines converging to the same colocation points [14].

3. BGP Threat Model

This section provides a short review of different BGP threats described in the literature, not limited to the specific shortcomings of BGP itself. Firstly we classify possible attackers in terms of their nature and corresponding offensive capabilities. Then we identify four categories of threats:

- Hijacks, exploiting the lack of authentication and path validation in BGP to manipulate routing tables.
- Gateway attacks exploiting either the BGP protocol itself or generic router vulnerabilities.
- Denial of Service (DoS) attacks.
- Physical attacks to colocation points, landlines and undersea cables.

We describe each threat, in particular in terms of their accessibility for different types of attackers, we comment on countermeasures, and we explore recovery methods. This work is based on a systematic literature review (SLR) that surveyed a total of 4,248 papers from six major online digital libraries[1], 579 of which were manually screened culminating in a final set of 66 related publications.

3.1. Attacker Taxonomy

The nature of attackers and the means at their disposal are an important factor to consider in a threat model. Almost all security incidents in the history of the Internet backbone have been attributed to non-malicious mistakes. However, the current security scene is composed of a variety of potential attackers with various offensive capabilities:

- Isolated individuals, with very limited capabilities and opportunistic behaviour (in general).
- Small groups of diverse motivations, with limited capabilities and variable degrees of expertise.
- Full-fledged organisations (criminal and non-criminal), with substantial capabilities.
- Nation-states, some with unprecedented offensive capabilities and supporting infrastructures.

The potential of some attacks presented in this section is correlated with the means of the attacker. For instance, DoS attacks grow in power according to the size of the attacker's infrastructure. There is little foresight of what could happen in case large organisations or states started to use DoS attacks openly as an act of war. The same goes for BGP hijacks: if a significant number of ASs started to maliciously and repeatedly try to disrupt each other, the behaviour of the backbone would certainly become unpredictable. In both cases, engaging unprecedented means into an attack could turn the tables in the global Internet landscape.

For some kinds of attacks however, large-scale effects could be at hand for very limited attackers. Generic router security flaws and "ping of death" attacks do allow a single individual to take down routers individually. Due to the concentration of high numbers of core routers, landlines and undersea cables at a handful of key locations, local physical attacks could also prove very efficient at taking down large parts of the BGP infrastructure. Such attacks do require expertise and specific personnel, yet they can be carried out by limited groups without the infrastructure and financial capabilities of nation-states or large international organisations.

1. ACM Digital Library, IEEE Xplore, ScienceDirect, Springer Link, Web of Science, and Wiley Online Library.

3.2. BGP Hijacks

Description. A prefix hijack is where an AS advertises false prefixes, claiming to have routes to address spaces that do not belong to it [3], [15]–[32]. When such BGP advertisements are propagated, potentially large amounts of traffic are redirected to the advertising network rather than to the intended destination. For instance, defensive "blackholing" hijacks allow administrators to drop traffic and isolate a network under attack [33]. A prefix hijack is indistinguishable from legitimate traffic engineering operations where network administrators rely on advertising additional prefixes to optimise their routing tables. Hence, the stability of the whole system relies on practice-based heuristics and mutual benevolence between ASs.

Prefix hijacks can either be intentional (and possibly malicious) or unintentional (i.e. misconfigurations, cf. [15], [34], [35]). In either case, the result of a prefix hijack is having rogue entries in the routing tables which may in turn cause disruptions or denial of service. Furthermore, the attacker (if any) could exploit the redirected traffic to carry out other attacks such as spamming [27], phishing [28], or Man In The Middle (MITM) operations [25], [26], [36]. In many cases, the recipient of the misdirected traffic collapses under the unexpected load.

In practice, almost all prefix hijacks are attributed to either misconfiguration (i.e. human error) or software malfunction [15], [19]. This includes commonly cited incidents such as those relating to AS7007 [22], AS3561 [23], as well as the global YouTube outage in 2008 [32]. Very few deliberate attacks have ever been documented until very recently (cf. [36]–[39]) which in itself is remarkable.

By nature, hijacks are sophisticated attacks, accessible only to organisations able to either build and maintain an AS of significant size or to take control of such an infrastructure via other means.

Protection. The BGP protocol is fundamentally insecure since prefix origins are not authenticated: there is no way to handle or even detect prefix hijacks with standard BGP alone. This lack of authentication has been identified in the literature [40] and several extensions of the protocol propose to address this issue. The main contenders here include BGPSEC [41], S-BGP [42], HI [43], soBGP [44], psBGP [45], IRV [46], OA [47], SPV [48], and many more. Aside from the effects on router performance [49] and partial security improvements [50], several works have studied the efficacy of partial and mixed deployments of the proposed

secure BGP variants [1], [18], [51]–[53] and new attacks defeating secure extensions have been proposed [54].

A recent study [21] has identified that the key to improving BGP security is to ensure routing policies that complement authentication protocols. In other words, a whole BGP deployment could be remarkably undermined if the involved ASs do not prioritise security when dealing with insecure BGP advertisements, which is not at all uncommon according to a recent survey [55]. Whitelist-based prefix filtering and path validation are the two main complements to authentication identified so far [1].

Recovery. Hijack recovery consists in restoring legitimate routing tables by withdrawing bogus routes and / or overriding it by preferable routes (i.e. routes with longer prefixes and shorter paths). The hijacked AS may advertise new specific routes and try to override the faulty conflicting announcements. Most major prefix hijacking events are handled manually in a few hours of time, which is the typical delay for corrected routing announcements to diffuse over the entire network.

Since there is no explicit recovery procedure in BGP, mitigation of past incidents has relied on collaboration between the involved parties: the rogue routes were withdrawn as soon as possible, and all the ASs involved took measures to rectify their systems. However, there is no occurrence of an openly conflictual situation where a hijacker would continuously hijack a prefix and refuse to collaborate.

Malicious ASs that repeatedly attempt to hijack other ASs can be manually identified and blacklisted. However, large-scale malicious scenarios where large ASs would be in open confrontation with each other and keep on advertising conflicting routes are yet to be investigated. In particular, the lack of a global authority able to resolve such conflicts and ensure overall coherence could prevent the entire network from stabilising as long as the conflict continues.

3.3. Gateway Elimination Attacks

Description. BGP routers can be attacked via any of their operational platforms, which includes common operating system and firmware vulnerabilities. Listing such attacks is out of the scope of this study. BGP routers are owned and managed by different companies following different administration policies. However, they often rely on the same physical hardware and software basis of their systems. As a consequence, a security breach in one of these components may affect a significant number of systems and make them fail in the same way due to a single type of attack – so-called *common mode failures*. Dependencies to common assets are difficult to identify and trace back, as such implementation details are often not published for confidentiality reasons. Sometimes even administrators are not fully aware of the hardware and software basis their systems run on. Yet, the transversal diffusion of these assets can compromise large numbers of systems, the 512k event [56], the "Packet of Death" [57] or the Shellshock vulnerability [58] being relevant examples. Whether or not a 0-day could effectively threaten the BGP infrastructure is yet to be demonstrated, but for lack of evidence this should be considered as a potential vulnerability.

Sophisticated BGP-specific attacks have been discussed in the routing community for years, but never practically proven to be possible until recently [24], [59]–[63]. The attacker uses a combination of specific BGP messages to overload the victim gateway router. For instance, the attacker would use a large number of deaggregated IP prefixes in order to create a significant volume of routing updates which costs the victim in CPU time (similar tactic to a DoS attack, cf. section 3.4). The attacker could also cause the victim's neighbours to stop responding to the victim, causing the victim's network buffers to overflow. These two factors are enough for some routers to stop functioning until restarted. These techniques are not necessarily new for attacking neighbouring BGP peers (cf. [64], [65]). However, when coupled with a simple targeting mechanism, they could be used to single out a victim that could be anywhere in the Internet backbone [62], [66].

BGP-specific gateway eliminations are sophisticated attacks that require a backing BGP infrastructure, thus they are reserved to the high end of the attacker spectrum. On the other hand, some generic router vulnerabilities can be exploited by individuals, which significantly increases the risk associated with this kind of threat.

Protection. Up-to-date patching of vulnerabilities is the main protection against generic attack vectors – again, detailing these goes beyond the scope of this study. BGP-specific gateway elimination attacks rely on the lack of path validation and origin authentication. As with hijacks, RPKI-based authentication, whitelist-based prefix filtering and path validation can help thwart malicious message propagation. However, these mechanisms have a cost in terms of network and CPU consumption that can be leveraged by attackers and result in other resource depletion techniques. In addition, recent work has discussed how "best practices" could in fact be self-defeating [62].

Recovery. Assuming the victim router does not suspect an attack and accepts all the attacker's updates, it quickly runs out of memory and needs to be restarted. Upon restart, the router would need further time in order to build its forwarding tables. Failure to detect the attack would make the router vulnerable to be eliminated again soon after restarting. Thus, detection is important for both efficient protection and rapid recovery. The literature does not detail detection mechanisms but lessons could be learned from SYN flood countermeasures [67].

3.4. Denial of Service Attacks

Description. A Denial of Service (DoS) is a disruption attack that consists in flooding a network or a particular machine with malicious traffic [2], [68]–[76]. This type of attack is well understood and documented in the literature, they are discussed here for the sake of completeness. A

DoS intends to disrupt and possibly blackout a system by two different means:

- Overwhelming the target network with an unmanageable amount of traffic.
- Saturating the processing capabilities of the target machines via excessive amounts of data.

DoS could happen unintentionally and without malicious intent. This is when underprepared websites become extremely popular (e.g. due to breaking news or via social media) and fail to cater to unexpected numbers of simultaneous service users [73].

Saturating the network or the processing capabilities of an online service is usually no easy feat (although not impossible, cf. [77]), in particular when the target benefits from a solid, large-scale backing infrastructure. A single machine is unlikely to have the capabilities to perform such a task, and common defence systems can easily identify a single DoS source and block its incoming attacks. Therefore, DoS attackers often rely on multiple, distributed attack sources, often in the form of large clusters of infected machines – so-called *botnets*. When thousands of machines controlled by an attacker contribute to a *Distributed DoS* (DDoS), the aggregation of their individual power allows to reach the large debits required for the attack, and the source distribution prevents the victim from identifying and defending efficiently from a single hostile source.

Botnets require significant amounts of resources to be built and maintained. An alternative for achieving a DDoS attack consists in diverting existing systems to perform the attack. In this variant, called reflection and amplification attack, the attacker queries multiple different providers while redirecting the responses to its target. Any publicly accessible service with weak authentication and providing large responses to small queries can be exploited for this purpose. The UDP protocol in particular lacks source validation, which makes UDP-based protocols such as DNS, NTP, SNMPv2, NetBIOS, etc. to be particularly exposed to this kind of exploitation [78]. For instance, a DNS server could provide a 3KB response to a 64 byte query, in effect generating 50 times the traffic it receives from the attacker towards an unsuspecting victim. The amplification mechanism makes DoS attacks accessible even to attackers with limited offensive capabilities.

Finally, the BGP-specific attack presented in section 3.3 exploits specificities of the BGP protocol to achieve the same effect.

Depending on the effectiveness of the attack method and on the victim's defences, a (D)DoS attack may:

- Have no visible effect on the victim from the customer's perspective, while generating extra maintenance and protection costs.
- Partially degrade services provided by the victim.
- Blackout the victim, making it unavailable from the network as long as the attack continues.
- Severely damage the victim even after the attack has finished, imposing costly repairs and maintenance operations.

Protection. DoS protection usually relies on detecting and filtering malicious sources. However, this is a difficult task since malicious traffic can be indistinguishable from legitimate traffic. Using reflection also helps prevent efficient filtering while hiding the true origin of the attack, as traffic is generated by multiple, apparently legitimate sources such as public services. A DoS attack based on redirected traffic via a BGP hijack would also naturally appear to come out of many legitimate sources.

Recovery. Efficient DDoS recovery requires extreme scalability, with no bottleneck nor single point of failure, allowing the network to cope with the attack traffic [72] by absorbing and dispatching traffic surges. Such scalability capabilities are so far reserved to the largest infrastructures and require heavy preparation on both the nodes and the topology of the infrastructure, discussed further in section 6.

3.5. Physical Attacks

BGP infrastructures are susceptible to a range of physical attacks. In addition to direct destructive attempts to BGP routers, attackers can target the power infrastructure, cooling systems, or other parts of the network fabric (such as links and switches) that connect a router to the rest of the infrastructure. Although an exhaustive taxonomy of such attacks is beyond the scope of this study, we discuss here the interplay between physical attacks and the logical topology of the backbone.

Mapping the logical topology of the backbone to its physical layout is difficult, as there is no straightforward way to associate a given route to a particular line or colocation point. Transatlantic cables for instance are usually shared by several carriers, which means that a single cable cut would cause trouble to several ASs simultaneously. In particular, peering or buying transit from several providers does not guarantee physical redundancy, as these networks could favour the same cable to route their traffic. Since tier-1 networks peer with each other, a single cable cut would not severely disrupt the backbone, as automatic re-routing would shift the traffic onto other transit providers using different cables. However, the case of several simultaneous cable cuts is unprecedented in Europe. In other regions where a single undersea cable carries a majority of the traffic, its disruption usually results in massive losses of connectivity [79].

Major cables converge to colocation points that concentrate large numbers of key physical assets. In Europe in particular, public (IXP) and private (PoP) colocation tends to share common premises [11] which increases even more the centrality of such physical hubs. As a result, complex routing tables such as the ones shown in section 5 map down to a handful of physical locations and lines between them, while it remains difficult to determine exactly the physical trajectory of traffic and vice-versa. Routine failures of individual equipments has made physical redundancy a strong requirement for such locations. However, it is unclear how significant physical attacks carried on one or several key colocation points would impact the entirety of the

backbone, as such attacks would simultaneously affect a large number of routes at the logical layer.

4. Mapping the BGP Infrastructure in Europe

This section presents the data sources used for this study and discusses their relevance and limitations.

4.1. Data Sources

RIPE (Réseaux IP Européens) is the regional Internet registry for Europe, the Middle East and Central Asia. RIPE maintains the Routing Information Service (RIS) which provides global routing updates (in 32-bit format) advertised in a number of major IXPs in European cities such as Amsterdam, Frankfurt and London since 2001 [80]. The following sections present an analysis of the data from 5 different Remote Route Collectors (RRC), labelled henceforth as rrc$x$, where x varies from "a" to "e". The exact identity of these vantage points will not be disclosed.

4.2. Data Interpretation & Relevance

It is important to note that the AS graphs presented in this paper – based on our preliminary investigations [81] – are a collection of the routes seen from a particular (European) vantage point in the network (example shown in Fig. 4). The data presented here should *not* be seen as a complete view of the network; it is nonetheless a representative section of significant portions of several national backbones in Europe and of their international connections – for instance, to tier-1 ASs from the US. The data varies across the different RRCs due to the specific topology – logical and physical – of the corresponding IXPs; nonetheless, the similarity of the analysis' results (section 6) on these different sources supports the relevance of the patterns it exhibits.

The coverage that our vantage points provide is difficult to ascertain. Complete routing tables from a single RRC – e.g. the one shown in Fig. 4, AS count on Table 2 – contain nearly the entire Internet (more than 50,000 ASs as of July 2015 [82], [83], cf. Fig. 5). More precisely, such graphs contain all existing ASs and routes to reach them from a particular vantage point, but not necessarily all possible routes between any two ASs. Therefore these maps provide a local perspective on the global AS routing table, notably in terms of access to remote transit. We decided to simplify the dataset and limit our analysis to the first hop of the AS routes recorded in the RIPE dataset, since any second hop in such a route would either:

- Be a link between two arbitrarily remote ASs, away from the local IXP where the route was advertised, and therefore not relevant to a local perspective.
- Be a local hop between two ASs in the same IXP and therefore a duplicate of the first hop of another route announced at the same IXP.

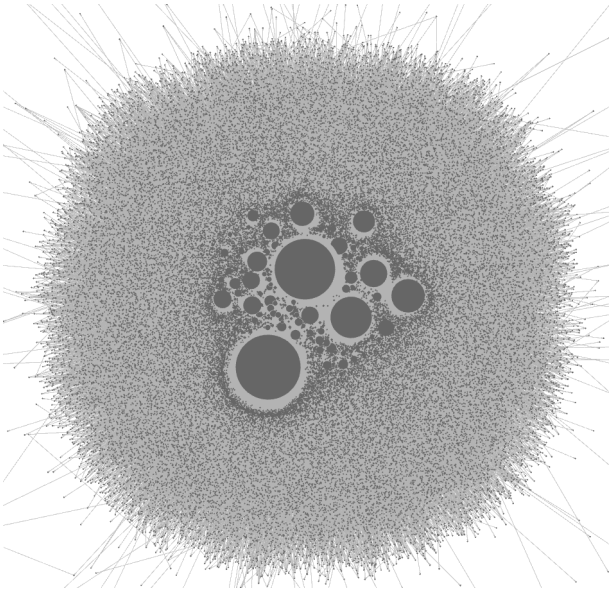

Figure 4: Example AS routing table for one RRC [80]: vertices are ASs, edges are routes, vertices' size is proportional to their degree, i.e. the number of routes going through them.

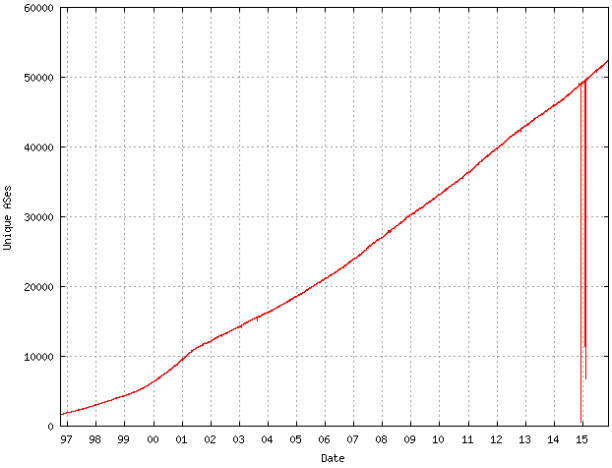

Figure 5: AS count for the last 18 years [82].

Table 2 shows the difference between complete routing tables and simplified routing tables for the 5 RRCs. Table 4 shows global statistics for the simplified dataset, i.e. for ASs one hop away from a given vantage point. Such ASs could still be anywhere in the world, as international carriers could virtually peer with any part of the Internet. Special cases such as VPNs excepted, ASs appearing on the routing tables are therefore "one physical network" away from the vantage point – be it a local ISP or a global carrier.

We checked the consistency of the routing tables in two ways. Firstly, we computed the intersection between the five routing tables (both complete and simplified): Table 2 shows that they are indeed significantly overlapping. Secondly, we took the following approach:

RRC	# ASs	# 1-hop ASs
rrca	51313	36635
rrcb	51360	32928
rrcc	51231	25111
rrcd	51408	38809
rrce	51458	31569
rrca ∩ rrcb ∩ rrcc ∩ rrcd ∩ rrce	51113	22838

TABLE 2: Cross-validation of datasets.

ASN	Name	Tier-1	Average rank	Average degree	Caida Rank [84]
3356	Level3	*	1.8	3968	1
174	Cogent	*	2.2	3781	2
6939	HE		2.8	3309	8
6461	Abovenet		7.8	1261	16
3257	Tinet/GTT	*	9.6	1102	5
9002	RETN		11.8	1353	19
1299	Telia	*	12.8	939	3
2914	NTT	*	13.8	1060	4
7018	AT&T	*	18.8	1074	17
4323	TW telecom		19.2	1149	21
8220	COLT		20.2	603	55
3216	Vimpelcom		20.6	606	24
209	Qwest		20.8	791	18
13030	Init7		21.0	717	45
12389	RosTelecom		21.0	573	22
20485	TransTelecom		21.8	749	20
3549	Level3	*	25.6	587	9
701	Verizon	*	25.6	571	13
2828	XO		25.6	552	10
43531	IX Reach		26.0	570	209

TABLE 3: Highest average-rank ASs (ranked by degree in each RRC).

- We computed the degree of each AS in the 5 RRCs.
- We ranked ASs by degree in the 5 RRCs.
- We compared the highest-rank ASs with several indicators, such as Caida's AS-rank (measuring the number of customers worldwide) and the list of tier-1 ASs (Table 1).

Table 3 shows an average result of this ranking on the five RRCs. All ASs represented here are either tier-1 or top tier-2 ASs, as denoted by their high rank on the CAIDA ranking system (based on the number of ASs in the customer cone). This demonstrates that the routing tables are consistent with other established sources of data. Also, this strongly indicates that ASs with a very high rank on average on the five sources are likely to be major interconnectivity providers between the European IXPs where the routing tables were recorded.

4.3. Data Limitations

The data is not free of biases however. Some notable networks are missing in Table 3: one would expect major European networks such as Deutsche Telekom (AS3320), Telecom Italia Sparkle (AS6762), British Telecom (AS5400 and AS2856) or France Telecom (AS5511) to be well represented in the routing tables. Also, when considering RRCs individually, some of the highest-degree ASs are peripheral regional networks (not shown here). This can be explained by several reasons:

- RIPE data is gathered on the basis of voluntary contributions that tend to exaggerate the importance of the contributors in the routing tables. This explains notably local biases towards some particular local networks.
- The routing tables are recorded in IXPs and do not capture private peering and customer agreements at private colocation points, that some transit providers favour over public IXPs.

This means that the view provided by the RRCs captures only the tip of the iceberg and misses important parts of the infrastructure: a significant consideration to keep in mind when interpreting the following disruption scenarios. However partial, the data is representative enough: the number of ASs captured in the routing table is significant (Table 4), the degree distributions are similar (Fig. 6) and the topology analysis presented in the following sections yield similar results for all five sources with no particular outlier. The results we present are therefore relevant and insightful despite the limitations of the dataset.

To achieve a comprehensive analysis of the entire European infrastructure, considerable cooperation between service providers would be required to allow a multi-vantage-point view of the network. Since BGP messages are exchanged via point-to-point TCP connections, they are not easy to monitor from a unique remote listener. Ideally, by positioning several BGP-message listening servers at key locations in the infrastructure (determining the exact locations of which would also require coordination with ISPs), it would be possible to collect enough information to build a near-complete map of Europe's Internet. Significant efforts are carried by RIPE in this direction [85].

5. Topology Analysis & Attack Scenarios

In this section, we study the topology of the backbone in the light of known attack vectors and we investigate disruption scenarios. Section 3 established that ASs could be taken down – individually or collectively – via a number of means:

- Physical attacks on colocation points.
- Disruptions of major landlines and undersea cables.
- Traffic disruptions (hijacks, DDoS).
- Other attacks, such as common mode failures.

In this section, we first look at the high end of the threat spectrum: we study the global topology of the backbone and we investigate global blackout scenarios necessitating heavy attack capabilities. Then we consider the low end of the spectrum: individual AS disruptions – achievable with more limited offensive resources – and their potential effect on the rest of the backbone in terms of traffic re-routing. Finally we discuss the domain in between, in particular regarding the risk of cascading failures.

RRC	# ASs	# links	average degree (± 0.1)	links (in proportion) held by the 40 highest-degree ASs
rrca	36634	69261	1.9	40519 (59%)
rrcb	32927	60069	1.8	36503 (61%)
rrcc	25110	42759	1.7	27541 (64%)
rrcd	38808	81062	2.1	44665 (55%)
rrce	31568	61139	1.9	39222 (64%)

TABLE 4: Global statistics for the 5 routing tables.

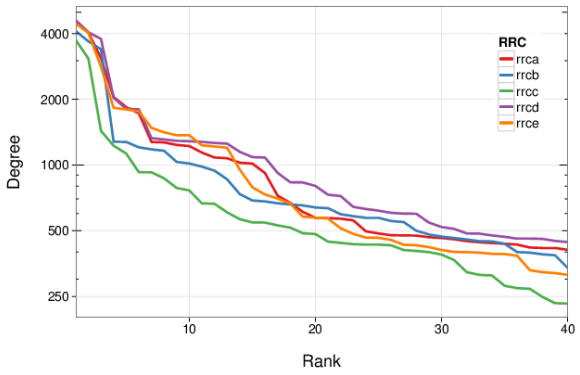

Figure 6: Head of the decreasing degree distribution.

Network topology analysis for the sake of security seems to be limited to the study of the possible impact of BGP hijacks, e.g. [7], [8], [86]. These works discuss how to exploit the three-layer logical topology of the infrastructure to maximise the impact of prefix hijacks. To our knowledge, our study is the first to extend this type of approach to other types of threats – such as DoS and physical attacks – and to combine a study of the logical and physical topologies of the infrastructure.

5.1. Large-scale Disruptions

Large-scale blackouts can be achieved by taking down all transit providers via simultaneous attacks. Fig. 6 shows the head of the degree distribution for all five RRCs – i.e. ASs with the highest number of links. As expected in scale-free graphs, high-degree vertices concentrate a majority of the graphs's edges. As shown in Tables 4 and 6, for each RRC, the 40 highest-degree ASs (i.e. the top 0.1%) hold more than half of the links in the graph, while the degree of the 40th AS is less than 10% of the degree of the first AS. The average degree in the routing tables is around 2, which is due to a large majority of tier-3 ASs being connected to only a couple of access providers.

Such a scenario would require significant offensive capabilities and planning, which makes it unlikely outside of global crisis situations. The exact consequences of large-scale failures are hard to determine, as it is not clear how the rest of the Internet interfaces with the particular subsets of the BGP infrastructure investigated here. There is little experience in the domain, the best guess one can make then is that national failures would certainly have knock-on effects on other national backbones.

AS rank (by degree)	orphan nodes	re-routable links
1	1188 (1.7%)	4588 (6.6%)
2	838 (1.2%)	4045 (5.8%)
3	410 (0.6%)	3083 (4.5%)
4	946 (1.4%)	2049 (3.0%)
5	718 (1.0%)	1840 (2.7%)
6	150 (0.2%)	1726 (2.5%)
7	105 (0.2%)	1275 (1.8%)
8	454 (0.7%)	1269 (1.8%)
9	257 (0.4%)	1236 (1.8%)
10	121 (0.2%)	1221 (1.8%)

TABLE 5: Effect of taking down individual transit providers in rrca. Percentages measure the proportion with respect to the total number of routes in the routing table.

Recovery from global failures is a completely uncharted territory with no precedents. The lack of explicit coordination procedures between ASs could be problematic, since uncontrolled traffic could disrupt the restoration of connectivity, especially in case all transit providers have been taken down: connectivity would have to be rebuilt progressively to prevent the first restored transit providers from crumbling under excessive load. Again, simulation is the only way to envision such scenarios, as a large-scale failure situation is unprecedented.

5.2. Local Disruptions

Taking all core ASs down simultaneously may not be necessary to significantly disrupt the BGP infrastructure. Given that high-degree transit providers concentrate a large portion of the traffic, shutting down a single one of them would re-route a traffic surge through other transit providers. Leaf ASs that depended on the dead AS only would be orphaned and lose connectivity with the rest of the infrastructure. In this section, we investigate the effects of such a targeted attack and we estimate the corresponding traffic surge (or *load repercussion*) on the rest of the backbone.

Taking down a large (possibly tier-1) AS still requires significant attack means, as these networks span over several continents. Section 3 has shown a number of means that can be exploited to achieve such a result:

- Logical attacks such as hijacks can disrupt access to and from an entire AS.
- Targeted attacks (DoS, packet of deaths) can take down routers at particular choke points.
- Physical attacks on choke points such as undersea cables and IXPs can literally sever connections large sections of the backbone rely on.
- Combinations of such attacks can also be considered, for instance: altering routing tables to concentrate traffic towards a particular physical link or place that is then physically attacked.

Such attacks carried with unprecedented means can affect targets comparable in size to international carriers – although in practice, several networks or entire geographical regions would be affected. The rest of this section considers

taking down individual ASs since it is simpler to consider from a simulation point of view while still relevant in terms of the order of magnitude of the disruption.

Without precise traffic information, it is difficult to evaluate accurately the effect of an AS being taken down. As a first estimate, Table 5 shows, for each transit provider, the number of orphan nodes that would end up disconnected, the number of lost links that could be re-routed by the rest of the infrastructure, and corresponding proportions these numbers represent compared to the total number of routes in the network. This proportion gives a first estimate of the amount of traffic that would be re-routed by the rest of the infrastructure. We complement this global consideration with local traffic estimates based on the logical topology around the victim AS.

Load repercussion model: The direct knock-on effect of taking an AS down can be estimated more precisely by measuring how big is the traffic surge that would have to be re-routed through other transit providers. In the absence of precise traffic data, we propose the following simple model, illustrated in Fig. 7. Each edge in the AS graph – i.e. each direct route between two ASs – is supposed to carry the equivalent of one unit of traffic. Then, the *load* of each node – i.e. the amount of traffic the AS handles – is equal to the sum of all traffic on its edges, in other words its degree.

This estimate does not account for the direction of the traffic (coming from or going towards neighbours), nor for the capability of ASs to handle traffic. Supposing that all routes carry the same amount of traffic is an oversimplification that cannot bring meaningful results on a small scale: this traffic model is only useful when considering orders of magnitude on a large number of routes, where averages start being statistically significant.

The approximation helps estimating the traffic volumes to be re-routed during disruptive attacks: supposing that one core AS has been taken down, its neighbours (peers and customers) will shift their traffic on alternative paths and create an overload that can be estimated in terms of number of routes. Let deg_i denote the degree of a node i and v a victim AS being taken down, then each AS n in the neighbourhood N_v of v will generate an overload $\Delta_{n\to c}$:

$$\Delta_{n\to c} = \frac{1}{deg_n - 1}$$

on each of its own neighbours $c \in N_n$ (i.e. one unit of traffic split equally among its own $deg_n - 1$ remaining neighbours). Then, any AS c in the neighbourhood of the neighbours of victim v (and in particular, when c is another core AS) would receive a total overload of:

$$\Delta_c = \sum_{n \in N_v \cap N_c} \Delta_{n\to c}$$

Then this overload can be compared to the usual traffic handled by c via the ratio:

$$\Delta\%_c = \frac{\Delta_c}{deg_c}$$

Let us illustrate the model in Fig. 7. If node B is taken down, nodes AB1 and AB2 each re-route one link onto node A ($\Delta_{AB1\to A} = \Delta_{AB2\to A} = 1$), node BC1 re-routes one unit of traffic onto C($\Delta_{BC1\to C} = 1$), while node ABC1 re-routes one link onto A and C ($\Delta_{ABC1\to A} = \Delta_{ABC1\to C} = 1/2$). This estimate preserves the total amount of traffic in the graph before and after disruption (minus orphan nodes the disruption disconnects completely, such as nodes A1 to A3 if node A were taken down). In total, A receives an extra traffic of $\Delta_A = 1 + 1 + 1/2 = 2.5$, which is approximately one third of its usual load ($\Delta\%_A = \frac{2.5}{8} = 0.3125$). However simple at such a small scale, the model becomes relevant when applied to the thousands of routes handled by core transit providers.

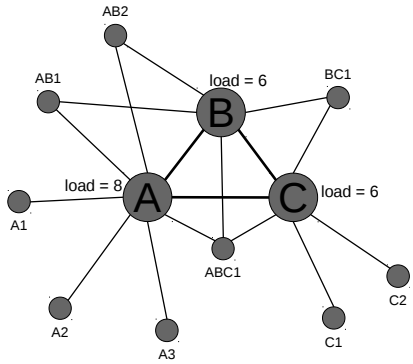

Figure 7: Load model illustration: one edge = one unit of traffic.

Model discussion: Counting the number of affected ASs and paths is the usual model for BGP disruptions in the literature [7], [8], [25], [27], [29], [30], [32]. This method is sometimes refined by counting affected prefixes [27], [29], [30] which allows to measure their length [30] or count affected IPs [29]. The refinements allow to discriminate ASs by their size, however this type of analysis has a cost: reconstructing the prefix graph is non-trivial. Also, there is no evidence that such refinements yield significantly different results. Determining whether or not it is significant is beyond the scope of this work.

AS-level analysis can be coupled with traceroute data and latency measurements between individual ASs [37]. Such models are based on an important monitoring infrastructure that is not available to the public. Similar public traceroute data (e.g. [87]) is fragmented and does not map to areas covered by the AS routes studied in the paper.

The way traffic would be redistributed cannot be predicted without access to detailed routing policies which determine which alternative paths are used when a route fails. These policies are implemented internally in BGP routers, as they depend on commercial agreements and internal technical decisions. In practice, the average degree in the graph is 2: this means that most customer ASs buy transit from 1 to 3 providers: one being the main link whereas the others are used as backups. In case the main provider fails, traffic is redirected on the backups, depending on an unknown order of preference.

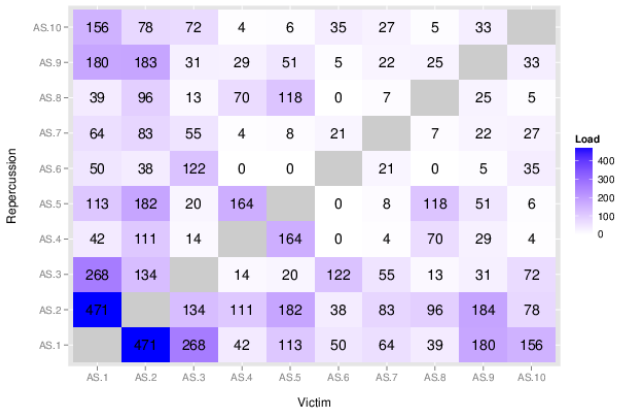

Figure 8: Load repercussion when taking down individual core ASs, in terms of absolute traffic.

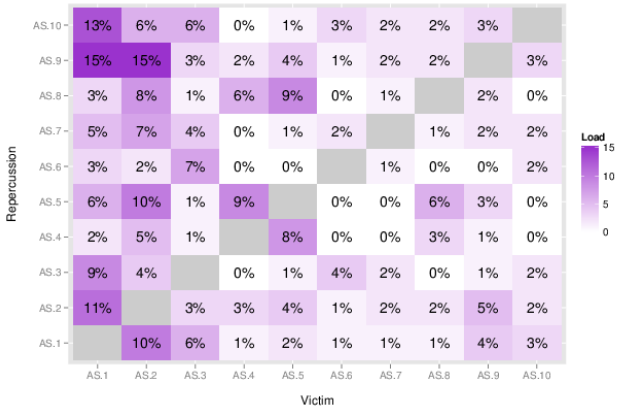

Figure 9: Load repercussion when taking down individual core ASs, in proportion to each AS's normal traffic.

In the absence of routing policies, the model does not discriminate between main provider and backups and assumes an equal distribution of traffic between remaining routes, as a statistical average would do. This approach does yield consistent results: an alternative simulation redirecting all traffic to one single remaining link (chosen at random) produced exactly the same results (less than .1% difference).

Simulation results: Figures 8 and 9 show the application of this model to the core of rrca, according to its topology as of July 2015, in terms of absolute traffic and relative to usual traffic respectively. The core we considered is constituted of the 10 highest-degree ASs in the routing table (cf. Fig. 6; the simulation could be extended to a larger core, for instance the first 20 or 40 ASs: the results would be the same for the top 10 anyway). The death of each transit provider is considered, and for each of them, the estimated knock-on effect Δ on other core ASs is computed, in terms of absolute traffic and ratio to the initial traffic handled by the AS. For instance, in case AS 1 is taken down, AS 2 would undergo a load surge of 471 units of traffic (equivalent to 471 routes), which represents 11% of its normal traffic. Most load repercussions in the table represent less than 5% of the normal traffic, which is likely not to constitute an issue. Repercussions seem to be proportional to the degree of the victim AS: the highest load surges would be created by the death of the two highest-degree ASs and would represent up to 15% of the usual traffic of other core ASs.

The same model was applied to the four other RRCs and produced similar results in terms of order of magnitude. Table 6 shows a summary of these simulations, in terms of average and maximal load surge following the death of a transit provider, both in absolute traffic and relative to the normal traffic. These results are similar across the five vantage points.

RRC	average load repercussion	maximum load repercussion
rrca	72 (3.3%)	471 (15%)
rrcb	62 (3.1%)	422 (14%)
rrcc	52 (3.9%)	250 (20%)
rrcd	76 (3.2%)	445 (14%)
rrce	73 (3.4%)	537 (13%)

TABLE 6: Summary of load repercussions following the death of a single transit provider, in terms of absolute traffic and proportionally to normal traffic (percentage).

5.3. Escalation Scenarios

In the absence of precise quantitative information, the main qualitative conclusion one can draw is that disrupting a single transit provider is likely to cause moderate traffic variations on other ASs. The order of magnitude provided by our estimate seems to rule out the possibility that the disruption of a single transit provider could trigger a cascading failure. However, the continuum between (unlikely) high-cost attacks taking down all core ASs at once and lower-cost attacks taking down a single one of them is a grey area.

Simultaneous disruption of several transit providers, via physical attacks on landlines and undersea cables, could still generate significant traffic variations that would go beyond the accumulated effect of individual disruptions. Modelling these variations with more precision is key to being able to anticipate possible cascading failures in a backbone. In the meantime, it should be considered possible that a handful of targeted attacks could generate a traffic re-routing surge that proves too much of a load for the rest of the network to handle. The resulting scenario would be a cascading failure where a successful attack on key networks happens to take down the whole infrastructure. This could in turn escalate into a global failure.

6. Topology Improvements

In this section we discuss how studying and improving the topology of the backbone can help mitigate attacks.

Decentralisation is key to the robustness and resilience of both physical and logical topologies. We propose two indicators to measure decentralisation:

- A local indicator: the number of links between non-core ASs and transit providers. The more an AS has

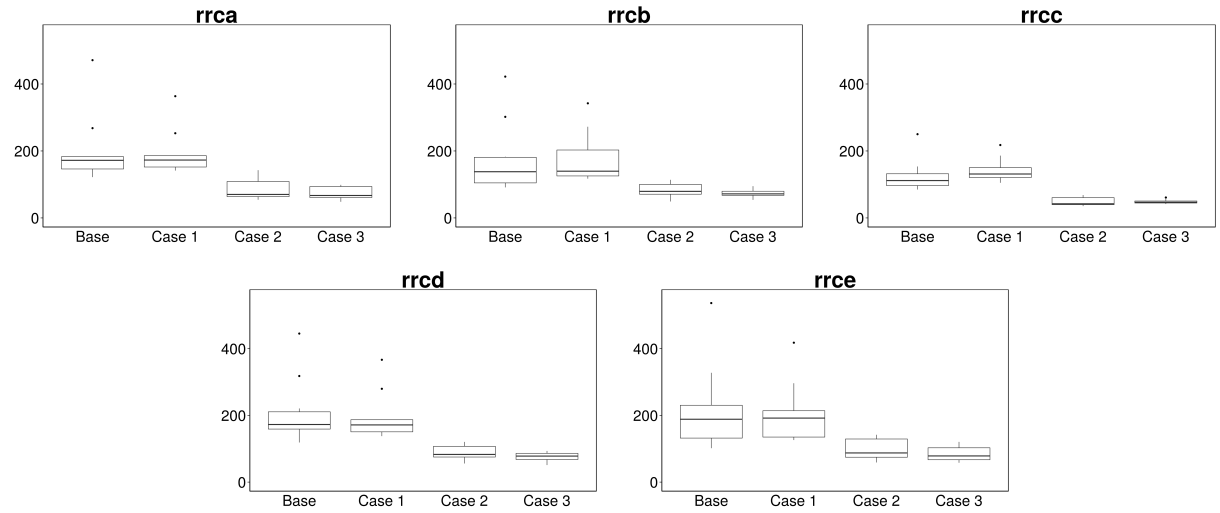

Figure 10: Effect of increasing core-connectivity (case 1), increasing core size (case 2) and both improvements combined (case 3) on the distribution of the 10 highest load surges.

direct routes to transit providers, the more backups are available in case one of them is disrupted[2].

- A global indicator: the number of core ASs in a backbone. The larger this number is, the smaller load repercussions will be in case one of them is taken down.

These indicators provide heuristics for increasing the backbone's robustness and resilience. These concepts are illustrated in Fig. 11: starting from a low-connectivity, small-core network (bottom left), one can increase the number of transit providers (top left), the number of links towards transit providers (bottom right) or both (top left).

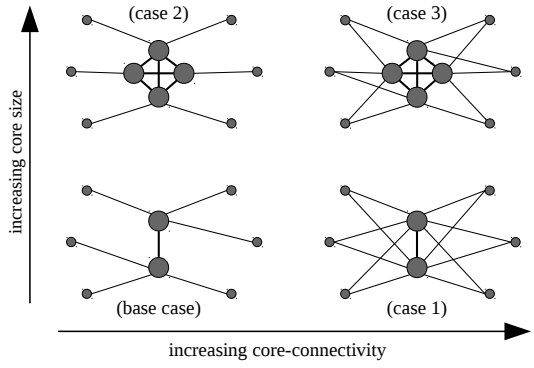

Figure 11: Illustration of improvement to backbone topology according to two orthogonal metrics.

2. Here, one must recall that the average degree in all routing tables is around 2 (cf. Table 4) which means that there is significant room for improvements in this department.

	case 1	case 2	case 3
reduction of average repercussion	-0%	-57%	-62%
reduction of maximum repercussion	-19%	-72%	-78%

TABLE 7: Summary of improvement experiments, averaged over the five RRCs.

6.1. Experiment 1: Increasing Core-connectivity

We have simulated the effect of increasing the number of links between core ASs and the rest of the backbone, such as illustrated by going from left to right topologies in Fig. 11. In this experiment, each non-core AS builds an additional link towards a randomly chosen core AS it is not already connected to (in case the AS is already connected to all identified core ASs, then no link is created).

Due to the increase of core-connectivity, a victim's traffic is spread on more neighbours-of-neighbours – since each of the victim's neighbours has an additional route to consider as a backup – and individual load surges on these are less important. Globally this transformation increases massively the degree of core ASs while shifting the degree distribution one unit higher for all other ASs; in particular, it removes all potential "orphans" from the network, as all nodes are connected to at least 2 transit providers.

Fig. 10 (case 1) shows the results of applying this procedure to the last known topology of the five RRCs. The maximum load repercussion is reduced in all cases, while the midspread remains stable. The minimum repercussion increases slightly in all cases, as a consequence of the increased degree of all transit providers. Table 7 (case 1) shows a summary averaging the improvements over the five RRCs: average load repercussions remain unchanged while maximum repercussions are decreased by 19%.

6.2. Experiment 2: Increasing Core Size

We have simulated the effect of increasing the number of core ASs in the backbone. For each core AS, a new twin AS is created and connected to all existing core ASs, and all the original AS's non-core neighbours are re-connected to the newly-created twin with a probability $1/2$.

Increasing core size literally divides each core AS into two parts: the death of a (half-)core is therefore half as important in terms of affected traffic as before the improvement. Also, the total number of links in the backbone is unchanged (except for a few links between the newly created AS and other core ASs), and the total number of core ASs is doubled, such as illustrated by going from bottom to top topologies in Fig. 11.

Fig. 10 and 7 (case 2) shows the results of applying this procedure: on average, average and maximum load repercussions are reduced by 57% and 72% respectively.

6.3. Experiment 3: Increasing Both Core Size and Core-connectivity

In the last experiment, the two procedures of the two previous experiments were applied sequentially: first the number of core ASs was increased, then the connectivity of all ASs to this extended core was increased by one link.

These two mechanisms are orthogonal: the first one reduces the initial load surge, the second one spreads it on more core ASes. Their combination does yield better results than their individual application, as shown by the results on Fig. 10 and Table 7 (case 3). The average amplitude of load repercussions is reduced by 62% on average, compared to the base case. The maximum amplitude of load repercussions is reduced by 78% on average.

6.4. Discussion

Centralisation is often the result of performance-driven design decisions, and decentralisation is generally a non-profitable architectural evolution. The topology improvements presented here all require to invest in developing the backbone at both the global level (increasing the number of transit ASs) and local levels (increasing core-connectivities). In particular, an important conclusion of the experiments is that the two policies can be followed in parallel, as the two will increase the robustness of the backbone. A careful trade-off analysis of security improvements against costs would certainly be necessary to ensure the feasibility of topology improvements.

Decentralising the logical architecture should provide an opportunity to decentralise the physical architecture as well. Spreading critical BGP hubs over national territories would improve the robustness of the infrastructure against localised physical attacks, incidents and power outages.

In the longer term, it is essential that both physical and logical topologies of the Internet are thoroughly observed and understood so as to avoid such centralised risks. This requires continuous monitoring and analysis of the BGP infrastructure, not limited to the scope of a national territory. Additional investigation is also necessary to identify and better understand the hidden parts of the BGP infrastructure.

Global supervision of the Internet raises the question of authority and information confidentiality. As discussed before, it is unclear which institutions would be in charge of observing and controlling parts of the infrastructure's topology. Whatever governance model is chosen, it will probably rely on several layers of management partitioning the system into recursive, decoupled domains.

7. Limitations & Future Work

7.1. Partial Information is a Security Issue

This paper presents the results of an analysis of the BGP infrastructure in Europe in the light of various attack vectors. These results identify serious security issues regarding the topology of a critical infrastructure, and the shortcomings of available information sources are major limiting factors that require deep further research. First and foremost, the BGP map in section 5 the whole analysis relies upon has several limitations, already discussed. Its limited scope in space hides important dependencies of the different backbones between themselves and towards the rest of the worldwide BGP system. However, this partial data has revealed regular topological patterns that should be considered relevant for at least portions of the global system. Future work will focus on trying to complement existing data with a more exhaustive model of the infrastructure, drawing from different data sources. Several research initiatives have been recently started in this direction [85], [88], [89].

More knowledge is necessary to investigate escalation scenarios precisely. The simple estimates computed in this study only provide a rough order of magnitude of possible traffic re-routing during an attack – it is however the best guess one can make with the data that is currently available. Investigating AS disruptions beyond core ASs – e.g. looking at multiple tier-2 ASs, regional disruptions – and their consequences, both logical and physical, is necessary. Analysing the topology of the backbone and its robustness over the course of several years would be a interesting indicator. Actual traffic volumes and maximum capabilities per route and per AS are key information that would allow to model the network with sufficient accuracy.

Major vulnerabilities of the visible BGP infrastructure are due to its centralised topology and consequent vulnerable points – undersea cables, colocation points, transit providers – where an attack could disrupt the entire infrastructure significantly. The exact importance of these vulnerable points is difficult to ascertain, as it depends on possible invisible dependencies and opaque management policies. The lack of documentation and studies of BGP topology is again a significant danger, since it is unclear whether the relevant authorities are aware of the situation – or indeed, if there are well-defined relevant authorities to deal with the situation.

Several security scenarios are presented in section 5, each of them opening rich perspectives. Further comparison is necessary with the large corpus of traffic engineering and security investigations on BGP, in addition to experience reports from the field. In turn, such scenarios are a potential basis for deeper analysis and security exercises, or war games, that improve the reactivity of system administrators while improving the knowledge about the system.

7.2. Applicability to Concrete Backbone Enhancements

The importance of dynamic events and ripple effects in security scenarios and the difficulty of foreseeing their actual consequences is key to understand the risk of backbone disruptions. A major follow-up of this study would be the implementation of a large-scale BGP simulator refining the approach presented in this paper. Fuelled with past and live routing data – not necessarily limited to RIPE – such a simulator would allow to perform in-depth analysis of the system over extended periods of time, in terms of topology and traffic patterns. Large-scale events and cascading failures could be implemented and studied in details, which would provide a precious help for establishing precise dynamical models of this critical infrastructure.

In section 6, several robustness and resilience solutions are discussed and possible improvements are proposed. The applicability of these improvements with respect to field constraints is yet to be established, which is a matter of business and political decisions as much as security. A formal framework capturing the architecture of resilience and robustness solutions for critical infrastructures would support the activity of both technical actors and decision makers. Also, such a framework would help in generalising this security analysis to other kinds of critical infrastructures.

In every BGP incident in the literature, collective recovery was performed in an ad-hoc, non-formalised fashion. Clearly there is room for improvement in this area. Specifying an inter-AS collaboration protocol would improve the global efficiency of the system for security and non-security purposes such as traffic engineering, while formal procedures open the way for recovery automatisation. Again, this will raise the question of authority distribution to determine which institutions are in charge of managing such a process. Establishing collaborative recovery solutions could also help in framing what non-collaborative – or in other words, openly hostile – situations would look like. Here again, significant research and practical efforts – such as war games – have to be conducted.

BGP networks are not the only critical assets of the Internet. The Domain Name System and Certificate Authorities are examples of well-known infrastructures that prove essential to the functioning of the network of networks. The inter-dependencies between these different systems are still a question to be investigated, not mentioning physical infrastructures such as the power network. Again, developing a generic framework for modelling and improving the robustness and resilience of mixed critical infrastructures would be a significant step in this direction.

8. Conclusion

This paper presents the results of a security analysis of the BGP infrastructure in Europe. The first main conclusion is that despite good resilience against random failures, the organic topology of the backbone could prove vulnerable to a variety of targeted attacks. The core of the infrastructure, composed of a small number of highly-connected transit providers, is also its main weakness. Global disruption of the infrastructure via a successful attack on this core, either logical or physical, is to be considered possible and could lead to unforeseen consequences at a global scale. Even limited individual attacks on core networks may escalate into cascading failures, with seemingly unpredictable effects.

The second conclusion is that significant research efforts are necessary to better understand and predict the behaviour of the system. Most of the BGP infrastructure is still an uncharted territory, which severely hinders any attempt to analyse and improve its robustness and resilience. Interactions between the BGP infrastructure and other critical infrastructures, such as the DNS or the power grid, are mostly unexplored. This study proposes a series of improvements to enhance the ability of system administrators to resist and recover from major attacks. Yet, beyond the complex technical, political and business involvement that is required, knowledge about the infrastructure is by far the limiting resource in this domain.

Acknowledgements

The authors would like to thank Christian Rossow for his support as a shepherd for this paper, and the anonymous reviewers for their constructive comments.

References

[1] S. Goldberg, "Why is it taking so long to secure internet routing?" *Communications of ACM*, vol. 57, no. 10, pp. 56–63, Oct 2014.

[2] N. Hoque, M. H. Bhuyan, R. Baishya, D. Bhattacharyya, and J. Kalita, "Network attacks: Taxonomy, tools and systems," *Journal of Network and Computer Applications*, vol. 40, pp. 307–324, 2014.

[3] G. Huston, M. Rossi, and G. Armitage, "Securing BGP – a literature survey," *IEEE Communications Surveys Tutorials*, vol. 13, no. 2, pp. 199–222, 2011.

[4] R. Albert, H. Jeong, and A.-L. Barabsi, "The internet's achilles' heel: Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 200–0, 2000.

[5] V. Rosato and F. Tiriticco, "Growth mechanisms of the AS-level internet network," *Europhysics Letters*, vol. 66, no. 4, p. 471, 2004. [Online]. Available: http://stacks.iop.org/0295-5075/66/i=4/a=471

[6] A. Elmokashfi, A. Kvalbein, and C. Dovrolis, "On the scalability of BGP: The roles of topology growth and update rate-limiting," in *ACM CoNEXT Conf.* ACM, 2008, pp. 8:1–8:12.

[7] B.-F. Zhang, Y. Li, Y.-J. Liu, and J.-S. Su, "Analysis of prefix hijacking based on as hierarchical model," in *Conf. on Network and System Security (NSS)*, Sept 2011, pp. 322–326.

[8] J. Zhao and Y. Wen, "Analysis on the effect of prefix hijacking attack and internet hierarchy," in *IEEE Conf. on Computer and Information Technology (CIT)*, Oct 2012, pp. 375–382.

[9] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271 (Draft Standard), Internet Engineering Task Force, Jan 2006, updated by RFCs 6286, 6608, 6793, 7606, 7607. [Online]. Available: http://www.ietf.org/rfc/rfc4271.txt

[10] R. Oliveira, B. Zhang, D. Pei, and L. Zhang, "Quantifying path exploration in the internet," *IEEE/ACM Transactions on Networking*, vol. 17, no. 2, pp. 445–458, April 2009.

[11] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger, "Anatomy of a large european IXP," in *ACM SIGCOMM Conf.* ACM, 2012, pp. 163–174.

[12] P. Richter, G. Smaragdakis, A. Feldmann, N. Chatzis, J. Boettger, and W. Willinger, "Peering at peerings: On the role of IXP route servers," in *ACM SIGCOMM Conf. on Internet Measurement*. ACM, 2014, pp. 31–44.

[13] Telegeography, "Submarine Cable Map," http://www.submarinecablemap.com, 2015, [Online; accessed 29-Jan-2015].

[14] Tata Communications, "UK Cable Map," http://map.tatacommunications.com/, 2015, [Online; accessed 19-Feb-2015].

[15] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP misconfiguration," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4, pp. 3–16, Oct 2002.

[16] H. Ballani, P. Francis, and X. Zhang, "A study of prefix hijacking and interception in the internet," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 265–276, Oct 2007.

[17] R. Kuhn, S. Liu, and H. Rossman, "Practical interdomain routing security," *IT Professional*, vol. 11, no. 6, pp. 54–56, Nov 2009.

[18] K. Butler, T. Farley, P. McDaniel, and J. Rexford, "A survey of BGP security issues and solutions," *Proceedings of the IEEE*, vol. 98, no. 1, pp. 100–122, Jan 2010.

[19] A. Elmokashfi, A. Kvalbein, and C. Dovrolis, "BGP churn evolution: A perspective from the core," *IEEE/ACM Transactions on Networking*, vol. 20, no. 2, pp. 571–584, April 2012.

[20] S. Bakkali, H. Benaboud, and M. Ben Mamoun, "Security problems in BGP: An overview," in *National Security Days (JNS3)*, April 2013, pp. 1–5.

[21] R. Lychev, S. Goldberg, and M. Schapira, "BGP security in partial deployment: Is the juice worth the squeeze?" *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 171–182, Oct 2013.

[22] S. A. Misel, "Wow, AS7007!" http://www.merit.edu/mail.archives/nanog/1997-04/msg00340.html, APR 1997.

[23] J. A. Farrar, "C&W routing instability," http://www.merit.edu/mail.archives/nanog/2001-04/msg00209.html, APR 2001.

[24] S. Goldberg, M. Schapira, P. Hummon, and J. Rexford, "How secure are secure interdomain routing protocols," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 87–98, Oct 2010.

[25] R. Hiran, N. Carlsson, and P. Gill, "Characterizing large-scale routing anomalies: A case study of the china telecom incident," in *Passive and Active Measurement Conf.* Springer, 2013, pp. 229–238.

[26] J. Cowie, "The New Threat: Targeted Internet Traffic Misdirection," http://www.renesys.com/2013/11/mitm-internet-hijacking/, Nov 2013, [Online; accessed 12-Jan-2014].

[27] J. Schlamp, G. Carle, and E. W. Biersack, "A forensic case study on as hijacking: The attacker's perspective," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 2, pp. 5–12, Apr 2013.

[28] O. Nordström and C. Dovrolis, "Beware of BGP attacks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 1–8, Apr 2004.

[29] V. Khare, Q. Ju, and B. Zhang, "Concurrent prefix hijacks: Occurrence and impacts," in *ACM SIGCOMM Conf. on Internet Measurement*. ACM, 2012, pp. 29–36.

[30] X. Zhao, D. Pei, L. Wang, D. Massey, A. Mankin, S. F. Wu, and L. Zhang, "An analysis of BGP multiple origin AS (MOAS) conflicts," in *ACM SIGCOMM Conf. on Internet Measurement*. ACM, 2001, pp. 31–35.

[31] G. Goth, "Should we stop trusting trust?" *IEEE Internet Computing*, vol. 12, no. 3, pp. 6–9, May 2008.

[32] U. Bornhauser and P. Martini, "About prefix hijacking in the internet," in *Conf. on Local Computer Networks (LCN)*, Oct 2011, pp. 143–146.

[33] D. Turk, "Configuring BGP to Block Denial-of-Service Attacks," RFC 3882 (Informational), Internet Engineering Task Force, Sep 2004. [Online]. Available: http://www.ietf.org/rfc/rfc3882.txt

[34] T. Griffin and G. Huston, "BGP Wedgies," RFC 4264 (Informational), Internet Engineering Task Force, Nov 2005. [Online]. Available: http://www.ietf.org/rfc/rfc4264.txt

[35] D. McPherson, V. Gill, D. Walton, and A. Retana, "Border Gateway Protocol (BGP) Persistent Route Oscillation Condition," RFC 3345 (Informational), Internet Engineering Task Force, Aug 2002. [Online]. Available: http://www.ietf.org/rfc/rfc3345.txt

[36] Dell SecureWorks, "BGP Hijacking for Cryptocurrency Profit," http://www.secureworks.com/cyber-threat-intelligence/threats/bgp-hijacking-for-cryptocurrency-profit/, 2014, [Online; accessed 23-Sep-2014].

[37] Renesys, "China's 18-Minute Mystery," http://research.dyn.com/2010/11/chinas-18-minute-mystery/, 2010, [Online; accessed 17-Feb-2015].

[38] ——, "Chinese Routing Errors Redirect Russian Traffic," http://research.dyn.com/2014/11/chinese-routing-errors-redirect-russian-traffic/, 2014, [Online; accessed 17-Feb-2015].

[39] ——, "The New Threat: Targeted Internet Traffic Misdirection," http://research.dyn.com/2013/11/mitm-internet-hijacking/, 2013, [Online; accessed 17-Feb-2015].

[40] G. Huston and R. Bush, "Securing BGP with BGPSEC," in *The Internet Protocol Forum*, vol. 14, no. 2, 2011.

[41] M. Lepinski, "BGPSEC protocol specification," http://tools.ietf.org/html/draft-ietf-sidr-bgpsec-protocol-05, IETF Network Working Group, Mar 2014, [Online; accessed 23-Sep-2014].

[42] S. Kent, C. Lynn, and K. Seo, "Secure Border Gateway Protocol (S-BGP)," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 4, pp. 582–592, Apr 2000.

[43] M. G. Gouda, E. N. (Mootaz) Elnozahy, C.-T. Huang, and T. M. McGuire, "Hop integrity in computer networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 308–319, Jun 2002.

[44] R. White, "Securing BGP through secure origin BGP (soBGP)," *Internet Protocol Journal*, vol. 6, no. 3, Sep 2003.

[45] T. Wan, E. Kranakis, and P. C. van Oorschot, "Pretty Secure BGP (psBGP)," Carleton University, Tech. Rep. TR-04-07, Sep 2004.

[46] G. Goodell, W. Aiello, T. Griffin, J. Ioannidis, P. D. McDaniel, and A. D. Rubin, "Working around BGP: An incremental approach to improving security and accuracy in interdomain routing," in *Network and Distributed System Security Symposium (NDSS)*, Feb 2003.

[47] W. Aiello, J. Ioannidis, and P. McDaniel, "Origin authentication in interdomain routing," in *ACM Conf. on Computer and Communications Security*. ACM, 2003, pp. 165–178.

[48] Y.-C. Hu, A. Perrig, and M. Sirbu, "SPV: Secure path vector routing for securing BGP," in *ACM SIGCOMM Conf.* ACM, 2004, pp. 179–192.

[49] G. Huston, "Measures of self-similarity of BGP updates and implications for securing BGP," in *Passive and Active Measurement Conf.* Springer, 2007, pp. 1–10.

[50] Y. Song, A. Venkataramani, and L. Gao, "Identifying and addressing protocol manipulation attacks in "secure" BGP," *IEEE Conf. on Distributed Computing Systems (ICDCS)*, pp. 550–559, 2013.

[51] J. Gersch, D. Massey, and C. Papadopoulos, "Incremental deployment strategies for effective detection and prevention of bgp origin hijacks," in *Distributed Computing Systems (ICDCS), 2014 IEEE 34th International Conference on*, June 2014, pp. 670–679.

[52] P. Gill, M. Schapira, and S. Goldberg, "Let the market drive deployment: A strategy for transitioning to BGP security," in *ACM SIGCOMM Conf.* ACM, 2011, pp. 14–25.

[53] R. Bush, "BGPSEC operational considerations," http://tools.ietf.org/html/draft-ietf-sidr-bgpsec-ops-01, 2012.

[54] Q. Li, X. Zhang, X. Zhang, and P. Su, "Invalidating idealized BGP security proposals and countermeasures," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 3, pp. 298–311, May 2015.

[55] P. Gill, M. Schapira, and S. Goldberg, "A survey of interdomain routing policies," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 1, pp. 28–34, Jan 2014.

[56] Wikipedia, "512k day," http://en.wikipedia.org/wiki/512k_day, 2014, [Online; accessed 24-Sep-2014].

[57] K. Kielhofner, "Packets of Death," http://blog.krisk.org/2013/02/packets-of-death.html, 2013, [Online; accessed 25-Feb-2014].

[58] Common Vulnerabilities and Exposures, "CVE-2014-6271," https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-6271, 2014, [Online; accessed 18-Oct-2014].

[59] D.-F. Chang, R. Govindan, and J. Heidemann, "An empirical study of router response to large BGP routing table load," in *ACM SIGCOMM Conf. on Internet Measurement*. ACM, 2002, pp. 203–208.

[60] L. Cavedon, C. Kruegel, and G. Vigna, "Are BGP routers open to attack? an experiment," in *Open Research Problems in Network Security*, ser. Lecture Notes in Computer Science, vol. 6555. Springer Berlin Heidelberg, 2011, pp. 88–103.

[61] A. Pilosov and T. Kapela, "Stealing the internet: An internet-scale man in the middle attack," *NANOG-44*, pp. 12–15, Oct 2008.

[62] M. Schuchard, C. Thompson, N. Hopper, and Y. Kim, "Peer pressure: Exerting malicious influence on routers at a distance," in *IEEE Conf. on Distributed Computing Systems (ICDCS)*, Jul 2013, pp. 571–580.

[63] N. Feamster, J. Jung, and H. Balakrishnan, "An empirical study of "bogon" route advertisements," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 1, pp. 63–70, Jan 2005.

[64] Y. Zhang, Z. M. Mao, and J. Wang, "Low-rate TCP-targeted DoS attack disrupts internet routing," in *Internet Society's Network and Distributed System Security Symp. (NDSS)*, 2007.

[65] M. Schuchard, A. Mohaisen, D. Foo Kune, N. Hopper, Y. Kim, and E. Y. Vasserman, "Losing control of the internet: Using the data plane to attack the control plane," in *ACM Conf. on Computer and Communications Security*, 2010, pp. 726–728.

[66] M. Schuchard, C. Thompson, N. Hopper, and Y. Kim, "Taking routers off their meds: Why assumptions of router stability are dangerous," in *Internet Society's Network and Distributed System Security Symp. (NDSS)*, 2012.

[67] W. Eddy, "TCP SYN Flooding Attacks and Common Mitigations," RFC 4987 (Informational), Internet Engineering Task Force, Aug 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4987.txt

[68] J. Nazario, "Politically motivated denial of service attacks," *The Virtual Battlefield: Perspectives on Cyber Warfare*, pp. 163–181, 2009.

[69] BBC, "Web slows after Jackson's death," http://news.bbc.co.uk/1/hi/8120324.stm, 2009, [Online; accessed 25-Feb-2014].

[70] US-CERT, "UDP-based amplification attacks," https://www.us-cert.gov/ncas/alerts/TA14-017A, 2014, [Online; accessed 25-Feb-2014].

[71] M. Prince, "The DDoS That Almost Broke the Internet," http://blog.cloudflare.com/the-ddos-that-almost-broke-the-internet, Mar 2013, [Online; accessed 25-Feb-2014].

[72] ——, "The DDoS That Knocked Spamhaus Offline (And How We Mitigated It)," http://blog.cloudflare.com/the-ddos-that-knocked-spamhaus-offline-and-ho, 2013, [Online; accessed 25-Feb-2014].

[73] S. Adler, "The Slashdot effect: an analysis of three Internet publications," *Linux Gazette*, vol. 38, p. 2, 1999.

[74] N. Chatzis, "Motivation for behaviour-based DNS security: A taxonomy of DNS-related internet threats," in *Conf. on Emerging Security Information, Systems, and Technologies (SecureWare)*, Oct 2007, pp. 36–41.

[75] T. Deshpande, P. Katsaros, S. Basagiannis, and S. Smolka, "Formal analysis of the DNS bandwidth amplification attack and its countermeasures using probabilistic model checking," in *IEEE Symposium on High-Assurance Systems Engineering (HASE)*, Nov 2011, pp. 360–367.

[76] J. Damas and F. Neves, "Preventing Use of Recursive Nameservers in Reflector Attacks," RFC 5358 (Best Current Practice), Internet Engineering Task Force, Oct 2008. [Online]. Available: http://www.ietf.org/rfc/rfc5358.txt

[77] M. S. Kang, S. B. Lee, and V. Gligor, "The crossfire attack," in *IEEE Symposium on Security and Privacy*, May 2013, pp. 127–141.

[78] C. Rossow, "Amplification Hell: Revisiting Network Protocols for DDoS Abuse," in *Network and Distributed System Security (NDSS) Symposium*, February 2014.

[79] Renesys, "Beware the Ides of March: Subsea Cable Cut Trend Continues," http://research.dyn.com/2014/03/beware-the-ides-of-march/, 2014, [Online; accessed 17-Feb-2015].

[80] "RIS Raw Data," http://www.ripe.net/data-tools/stats/ris/ris-raw-data, 2014, [Online; accessed 30-Apr-2014].

[81] S. Frey, A. Rashid, Y. Elkhatib, K. Follis, J. Vidler, N. Race, and C. Edwards, "Resilience of the Internet: the case of the BGP backbone," in *IEEE Symposium on Security and Privacy*, May 2015.

[82] T. Bates, P. Smith, and G. Huston, "CIDR report," http://www.cidr-report.org/, 2014, [Online; accessed on 12 March, 2015.].

[83] Geoff Huston, "Potaroo," http://bgp.potaroo.net/tools/asn32/, 2015, [Online; accessed 26-Nov-2015].

[84] CAIDA, "AS Rank," http://as-rank.caida.org/, 2015, [Online; accessed 04-Feb-2015].

[85] "RIPE Atlas," https://atlas.ripe.net/, 2015, [Online; accessed 06-Aug-2015].

[86] J. Gersch and D. Massey, "Characterizing vulnerability to IP hijack attempts," in *IEEE Conf. on Technologies for Homeland Security (HST)*, Nov 2013, pp. 328–333.

[87] CAIDA, "BGP traceroute," https://www.caida.org/workshops/bgp-traceroute/, 2015, [Online; accessed 26-Nov-2015].

[88] A. Faggiani, E. Gregori, A. Improta, L. Lenzini, V. Luconi, and L. Sani, "A study on traceroute potentiality in revealing the internet AS-level topology," in *IFIP Networking Conference*, June 2014, pp. 1–9.

[89] R. Durairajan, J. Sommers, and P. Barford, "Layer 1-informed internet topology measurement," in *ACM SIGCOMM Conf. on Internet Measurement*. ACM, 2014, pp. 381–394.