Anycast Agility: Network Playbooks to Fight DDoS *

A S M Rizvi† Leandro Bertholdo† João Ceron John Heidemann
USC/ISI University of Twente SIDN Labs USC/ISI

Abstract
IP anycast is used for services such as DNS and Content Delivery Networks (CDN) to provide the capacity to handle Distributed Denial-of-Service (DDoS) attacks. During a DDoS attack service operators redistribute traffic between anycast sites to take advantage of sites with unused or greater capacity. Depending on site traffic and attack size, operators may instead concentrate attackers in a few sites to preserve operation in others. Operators use these actions during attacks, but how to do so has not been described systematically or publicly. This paper describes several methods to use BGP to shift traffic when under DDoS, and shows that a response playbook can provide a menu of responses that are options during an attack. To choose an appropriate response from this playbook, we also describe a new method to estimate true attack size, even though the operator’s view during the attack is incomplete. Finally, operator choices are constrained by distributed routing policies, and not all are helpful. We explore how specific anycast deployment can constrain options in this playbook, and are the first to measure how generally applicable they are across multiple anycast networks.

1 Introduction
Anycast routing is used by services like DNS or CDN where multiple sites announce the same prefix from geographically distributed locations. Defined in 1993 [50] anycast was widely deployed by DNS roots in the early-2000s [4, 29, 64], and today it is used by many DNS providers and Content Delivery Networks [16, 17, 24, 26, 80].

In IP anycast, BGP routes each network to a particular anycast site, dividing the world into catchments. BGP usually associates networks with nearby anycast sites, providing generally good latency [62]. Anycast also helps during Distributed-Denial-of-Services (DDoS) attacks, with each site adds to the aggregate capacity at lower cost than a single very large site. Each site is independent, so should DDoS overwhelm one site, sites that are not overloaded are unaffected.

DDoS attacks are getting larger and more common. Different root servers and anycast services frequently report DDoS events [18, 43, 48, 49]. Different automated tools make it easier
to generate attacks [81], and some offer DDoS-as-a-Service, allowing attacks from unsophisticated users for as little as US$10 [68]. DDoS intensity is still growing, with the 2020 CLDAP attack exceeding 2.3 Tbps in size [66], and the 2021 VoIP.sms attack lasting for over 5 days [51, 65]. The reservoir of attack sources grow with millions of Internet-of-Things devices whose vulnerabilities fuel botnets [36].

Operators depend on anycast during DDoS attacks to provide capacity to handle the attack and to isolate attackers in catchments. Service operators would like to adapt to an ongoing attack, perhaps shifting load from overloaded sites to other sites with excess capacity. Prior studies of DDoS events have shown that operators take these actions but suggested that the best action to take depends on attack size and location compared to anycast site capacity [46]. While prior work suggested countermeasures, and we know that operators alter routing during attacks, to date there has been only limited evaluation of how routing choices change traffic [4, 27, 37, 53]. Only very recent work examined path poisoning to avoid congested paths [70]; there is no specific public guidance on how to use routing during an attack.

The goal of this paper is to guide defenders in traffic engineering (TE) to balance traffic across anycast during DDoS.

Our first contribution is a system with novel mechanism to estimate true attack rate and plan responses. First, we propose a new mechanism to estimate the true offered load, even when loss happens upstream of the defender. Estimating the relative load on each site (§3.3) is the first step of defense, so that the defender can match load to the capacities of different sites, or decide that some sites should absorb as much of the attack as possible. Second, we develop a BGP playbook: a guide that allows operators to anticipate how TE actions rebalance load across a multi-site anycast system. Together, these two elements provide a system that can automate response to DDoS attacks by adjusting anycast routing according to the playbook, or recommend actions to a human operator.

The second contribution is to understand how well routing options for multi-hop TE work: AS prepending, community strings and path poisoning. While well known, it is not widely understood how available and effective these mechanisms are. In §6 we show that while AS prepending is available almost anywhere, community strings and path poisoning support varies widely. We also show that their effectiveness varies greatly, in part because today’s “flatter Internet” [15] means

---

* A shorter version of this paper will appear in the proceedings of the 31st USENIX Security’22 Symposium. This is the full version.
† Shared first author
AS prepending often shifts either nearly all or nearly no traffic. Community strings provide finer granularity control, but we show their support is uneven. Path poisoning may provide control multiple hops away, but like community strings it is often filtered, particularly for Tier-1 ASes. When these factors combine with the interplay between multiple sites and an anycast system, a BGP playbook is important to guide defenders. Since the effects of TE are often specific to the peers and locations of a particular anycast deployment, we explore how sensitive our results are to different locations and numbers of anycast sites ($\S7$).

Our final contribution is to demonstrate successful defenses in practice. We replay real-world attacks in a testbed and show TE can defend ($\S8$). Of course no single defense can protect against all attacks, these examples show our approach provides a successful defense to many volumetric and polymorphic DDoS attacks. They show that our algorithm and process contributions (attack size estimation and playbook construction) have practical application.

Our work uses publicly available datasets. Datasets for the input and results from our experiments are available at no charge. Because our data concerns services but not individuals, we see no privacy concerns.

2 Related Work

Anycast routing has been studied for a long time from the perspective of routing, performance, and DDoS-prevention.

**BGP to steer traffic:** Prior work showed BGP is effective to steer traffic to balance load on links [8, 27, 54]. However, Ballani et al. showed that anycast requires planning and care for effective load balancing [4]. Others proposed to manipulate BGP based on packet loss, latency and jitter [47, 53]. We build on Ballani’s recommendation to plan anycast, proposing a BGP playbook, and studying how well it can work.

Chang et al. [14] suggested using BGP Communities for traffic engineering [10, 13, 74]. Recent work has examined BGP communities for blackhole routing in IXPs and ISPs [21, 28]. Smith and Glenn examined path poisoning to address link congestion [70]. While each of these are important options in routing for defense, we show a system that guides the operator to select between them. A system with multiple choices is necessary because no single method works against all attacks. For example, we show path poisoning does not work when we poison a Tier-1 AS.

**Anycast performance:** Most anycast research focused on efficient delivery and stability [11, 39, 41, 59, 79]. Later studies explicitly investigate the proximity of the clients [4, 11, 39].

Some studies try to improve anycast through topology changes [45, 62]. Anycast services for DDoS is already used in commercial solutions e.g., Amazon [63], Akamai [75] and AT&T [72]. However, none of them address how to use routing manipulations as a DDoS defense mechanism.

**Anycast catchment control as a DDoS mitigation tool:** To our knowledge, the idea of handling DDoS attacks by absorb-
response from the playbook. Together, these tools help to understand not only how to reduce traffic at a given site, but also the sites where that traffic will go.

These components come together in our automated response system (§3.4) that iterates between measurement and attack size estimation, defense selection, then deployment. Defense uses the playbook built during mapping; we provide an example playbook in §6.4. We show how these defenses operate in testbed experiments in §8.

Our system is designed for services that operate with a fixed amount of infrastructure on specific anycast IP addresses and do not employ a third-party scrubbing service. Operators of CDNs with multiple anycast services, DNS redirection, or scrubbing services may use our approach, but also have those other tools. However, many operators cannot or prefer not to use scrubbing and DNS redirection: all operators of single-IP services (all DNS root servers), many ccTLDs who value national autonomy, and scrubbing services themselves. Our approach defends against volumetric attacks where we have spare capacities in other sites. Since DDoS causes unavailability of services, suboptimal site selection during an attack is not a concern.

3.2 Measurement: Mapping Anycast

We map the catchments of anycast service before an attack so that the defender can make an informed choice quickly during an attack, building a BGP playbook (§6.4).

To map anycast catchments we used Verfploeter [20]. As an active prober (ICMP echo request), Verfploeter observes the responses of all ping-responsive IPv4 /24s and maps which anycast site receives the responses. We provide a detailed description of anycast, BGP, and Verfploeter in Appendix A and Appendix B. Since mapping happens before the attack, mapping speed is not an issue.

Alternatively, we can map traffic by observing which customers are seen at each site over time, or measuring from distributed vantage points such as RIPE Atlas [3, 73]. (Operators may already collect this information for optimization.)

Mapping should consider not only the current catchments but also potential shifts we might make during the attack. This full mapping is easy to do with Verfploeter, which can be continuously running in an adjacent BGP prefix to map the possible shifts. This mapping process is important to anticipate how traffic may shift. We will show later that BGP control is limited by the granularity of routing policy (§6) and by the deployment of the anycast sites (§7).

A challenge in pre-computed maps with routing alternatives is that routing is influenced by all ASes. Thus, the maps may shift over time due to changes in the routing policies of other ASes. Fortunately, prior work shows that anycast catchments are relatively slow to change [79]. We also show that our BGP playbook is stable over time (§6.4 and Appendix F).

3.3 Estimation of the Attack Size

After the detection of an attack, the first step in DDoS defense is to estimate the attack size, so we can then select a defense strategy of how much traffic to shift. Our goal is to measure offered load, the traffic that is sent to (offered to) each site. During DDoS offered load balloons with a mix of attack and legitimate traffic, and loss upstream of the service means we cannot directly observe true offered load. We later evaluate our approach with real-world DDoS events (§4).

Idea: Our insight is that we can estimate true offered load based on changes in some known traffic that actually does arrive at the service, even when there is upstream loss.

To know how much offered load actually arrives at the service, we need to estimate some fraction of legitimate traffic. We can then observe how much this traffic drops during the attack, inferring upstream loss. Unfortunately, there is no general way to determine all legitimate traffic, since legitimate senders change their traffic rates, and attackers often make their traffic legitimate-appearing. Our goal is to reliably estimate some specific legitimate traffic; we describe several sources next.

Traffic sources: There are several possible sources of known legitimate traffic—we consider known measurement traffic and regular traffic sources that are heavy hitters [5].

For DNS, our demonstration application, RIPE Atlas provides a regular source of known-good traffic, sent from many places. RIPE makes continuous traffic from around 10k publicly available vantage points [56]. Each RIPE vantage point queries every 240 s, and there is enough traffic (about 2500 queries/minute) to provide a good estimate of offered load. (Although RIPE Atlas is specific to DNS, other commercial services often have similar types of known monitoring traffic.)

To find the known-good traffic at each site, we use the catchments of RIPE vantage points with pre-deployed RIPE DNS CHAOS queries (one exists for each root DNS IP, such as measurement ID 11309 for A-root). We can also use Verfploeter or captured traces in the anycast sites. An advantage of using RIPE traffic is that it does not place any new load on the service.

Heavy hitters can provide an additional source of known-good traffic. Many services have a few consistently large-volume users with regular traffic patterns, and while they vary over time, many are often stable. For DNS, we find that most heavy hitters have a strong diurnal variation in rate; we model senders change their traffic rates, and attackers often make their traffic legitimate-appearing. Our goal is to reliably estimate some specific legitimate traffic; we describe several sources next.

Idea: Our insight is that we can estimate true offered load based on changes in some known traffic that actually arrives at the service, even when there is upstream loss.

To know how much offered load actually arrives at the service, we need to estimate some fraction of legitimate traffic. We can then observe how much this traffic drops during the attack, inferring upstream loss. Unfortunately, there is no general way to determine all legitimate traffic, since legitimate senders change their traffic rates, and attackers often make their traffic legitimate-appearing. Our goal is to reliably estimate some specific legitimate traffic; we describe several sources next.

Traffic sources: There are several possible sources of known legitimate traffic—we consider known measurement traffic and regular traffic sources that are heavy hitters [5].

For DNS, our demonstration application, RIPE Atlas provides a regular source of known-good traffic, sent from many places. RIPE makes continuous traffic from around 10k publicly available vantage points [56]. Each RIPE vantage point queries every 240 s, and there is enough traffic (about 2500 queries/minute) to provide a good estimate of offered load. (Although RIPE Atlas is specific to DNS, other commercial services often have similar types of known monitoring traffic.)

To find the known-good traffic at each site, we use the catchments of RIPE vantage points with pre-deployed RIPE DNS CHAOS queries (one exists for each root DNS IP, such as measurement ID 11309 for A-root). We can also use Verfploeter or captured traces in the anycast sites. An advantage of using RIPE traffic is that it does not place any new load on the service.

Heavy hitters can provide an additional source of known-good traffic. Many services have a few consistently large-volume users with regular traffic patterns, and while they vary over time, many are often stable. For DNS, we find that most heavy hitters have a strong diurnal variation in rate; we model senders change their traffic rates, and attackers often make their traffic legitimate-appearing. Our goal is to reliably estimate some specific legitimate traffic; we describe several sources next.

Estimation: Our goal is to estimate offered load, $T_{\text{offered}}$. We can measure the observed traffic rate, $T_{\text{observed}}$, at the access link. We define $\alpha$ as the access fraction—the fraction of traffic that is not dropped. Therefore $T_{\text{observed}} = \alpha \cdot T_{\text{offered}}$. 
To estimate the access fraction ($\alpha$), we observe that known good traffic has the same loss on incoming links as does other good traffic and attack traffic. We estimate the known traffic rate (from RIPE Atlas measurement traffic, or from heavy hitters, or both), as $T_{\text{known}}$. Then $\alpha \cdot T_{\text{known,offered}} = T_{\text{known,observed}}$, and our estimate of offered load is $T_{\text{offered}} = T_{\text{observed}} \cdot T_{\text{known,offered}} / T_{\text{known,observed}}$.

3.4 Traffic Engineering as a Defense Strategy

With knowledge of the offered load, the defender can select an overall defense strategy that will drive traffic engineering decisions. The defender first must determine if the attack exceeds overall capacity or not.

For attacks that exceed overall capacity, the defender’s goal is to preserve successful service at some sites, while allowing other sites to operate in degraded mode as absorbers [46]. The defender may also choose to shift traffic away from some degraded sites to ease their pain. Unloading the overloaded sites is recognized as breakwaters [37].

For moderate-size attacks, the defender should try to serve all traffic, rebalancing to shift traffic from overloaded sites to less busy sites. In heterogeneous anycast networks, where some sites have more capacity than others, the defense approach can be different. In these cases, larger, “super”-sites can attract traffic from smaller sites. For moderate-size attacks, it may even be best for smaller sites to shut down if the super-sites can handle the traffic.

Regardless of attack size, traffic engineering allows the defender to shift attack traffic to absorber or breakwater sites. We next describe traffic engineering options, and then how one can automate response. For operators unwilling to fully automate response, our system can still provide recommendations for possible actions and their consequences.

3.4.1 Traffic Engineering to Manage an Attack

Given an overall defense strategy (absorb or rebalance), the defender will use traffic engineering to shift traffic, either automatically (§3.4.2) or as advice under operator supervision. For anycast deployments connected by the public Internet, BGP [8] will be the tool of choice to control routing and influence anycast catchments. Organizations that operate their own wide-area networks may also be able to use SDN to manage traffic on their internal WAN [31, 61]. Fortunately, BGP has well established mechanisms to manage routing policy. We use three BGP mechanisms in the paper: AS-Path prepending, BGP communities and Path Poisoning.

**AS-Path Prepending** is a way to de-prefer a routing path, send traffic to other catchments. BGP’s AS-Path is the list of ASes back to the route originator. The AS-Path both prevents routing loops and also serves as a rough estimate for distance, with BGP preferring routes with shorter AS-Paths. By artificially inserting extra ASes into the AS-Path, the route originator can de-prefer one site in favor of others. Path prepending is known to be a coarse routing technique for traffic engineer-

![Figure 2: TE techniques to shift traffic from Site-1 to Site-2.](image-url)

We define **Negative Prepending** as the use of AS-Path prepending to draw traffic towards a site, preferring one site over others. Prepending can only increase path lengths, but an anycast operator in control of all anycast sites can prepend at all sites except one, in effect giving that site a shorter AS-Path (relative to the other sites) than it had before. “Negative prepending by one at site S” is, therefore, shorthand for prepending by one at all sites other than S.

Long AS-Paths due to prepending can make prefixes more vulnerable to route hijacking [42]. However, this issue has a small impact on anycast prefix, as always there is a site announcing without any prepend, keeping the path length limited. We suggest that formal defenses to hijacking such as RPKI are needed even without prepending, and when they are in place, prepending can be an even more valuable tool for TE.

**BGP Communities** (or community strings) label specific BGP routes with 32 or 64 bits of information. How this information is interpreted is up to the ASes. While not officially standardized, a number of conventions exist where part of the information identifies an AS and the other part a policy such as blackholing, prepend, or set local-preference. Community strings are widely supported to allow ISPs to delegate some control over routing policy to their customers [1, 74].

**Path Poisoning** is another way to control the incoming traffic. This technique consists of adding the AS of another carrier to the AS PATH. Paths that repeat ASes in different parts of the AS PATH indicate routing loops and must be discarded by BGP.

When using path poisoning we announce a path with both the poisoned AS and own AS (otherwise neighbors may filter our announcement as not from us). We must therefore also prepend twice at all other anycast sites, otherwise poisoning also results in a longer AS path.

Figure 2 shows how traffic engineering can be applied to anycast systems in order to modify the catchment. In this example, site-1 is overwhelmed by an attack. Aiming to shift bins of traffic to site-2 with spare capacity, we can make BGP announcements. site-1 poisons AS3, prepends (only showing to AS4), and prevents announcement to AS5 using not-export BGP community. These changes decrease load in site-1, shifting the traffic to site-2.
3.4.2 Automatic Defense Selection

To automate defense we use a centralized controller. The controller collects observations for all sites (from external measurements, or assuming the site is saturated if it cannot reach the site), then takes action, if required ((4) of Figure 1): (1) The controller identifies sites that are overcapacity by comparing estimated load to expected capacity and observed resources at each site. (2) The controller identifies all playbook options that will reduce load at any impacted sites without overloading currently acceptable sites. (3) It selects from any viable options, favoring a uniform distribution and smallest change (or selecting arbitrarily if necessary). If all changes leave some sites overwhelmed, it can choose the “least bad” scenario, or request operator intervention.

After deploying a new routing policy, the decision machine continues to evaluate the traffic level at each site (5 of Figure 1). If any site is still overwhelmed after 5 minutes, we try again, repeating size estimation, decision, and action. In the subsequent iterations, the controller only considers the routing options that were considered in the previous iteration (from step (2) of this decision process). We allow time between attempts so announcements can propagate [38]. To avoid oscillation or interference with route flap dampening, after three attempts we escalate the problem to the human operator. We choose these values for timer duration and number of retries from recommendations of operators to avoid oscillation, other options are possible. Explore other options is possible as future work.

Return to service: After a period with no overloaded sites, we can automatically revert any interventions, on the assumption that default routing provides users best service. Leaving interventions in place for some time can help with polymorphic attacks (§8).

3.4.3 Operator Assistance System

We discussed our approach with operators of root DNS and cloud services to get feedback on the approach. While they were enthusiastic about automated defenses to deal with common DDoS events, and to handle events during non-business hours, some operators prefer human-supervised (non-automated) response, and all expected human supervision of response during initial deployment to build trust before full automation.

To support human-supervised response, we design an ope-
Table 1: Estimating sizes of offered load (second from right) based on known-good traffic (second from left) with real-world attacks at B-root and testbed experiment. Traffic rates are in queries/second (reporting only the peaks).

| Date       | Dur. | known-good traffic | offered load during attack | estimated/ | key takeaways                      |
|------------|------|--------------------|---------------------------|------------|-----------------------------------|
|            |      | normal observed α  | normal reported estimated | α          |                                   |
| 2015-11-30 | 3h   | 33.08 1.85 0.0559  | 0.03 M 0.37 M 5.1 M 6.6 M | 0.07 1.3   | Path prepending works everywhere to de-prefer a site ($6.1.2$), but shifts traffic in large amounts ($6.1.3$), and has few traffic levels (Figure 6). |
| 2016-06-25 | 3h   | 36.58 0.33 0.0091  | 0.03 M 0.10 M 10 M 11 M  | 0.01 1.1   | Neg. Prepending works everywhere to prefer a site ($6.1.2$). |
| Testbed    | 5min | 425.2 207.0 0.4900 | 8.5 k 16.3 k 29.2 k 33.2 k | 0.56 1.1   | BGP community strings are not universal ($6.2.1$). When supported, they provide finer-granularity control than prepending ($6.2.2$). |

Table 2: Testbed and respective sites used in our experiments. Transit providers (*) and IXP (†).

| Testbed | Used Sites                                                                 | # |
|---------|---------------------------------------------------------------------------|---|
| Peering | Amsterdam*(AMS), Boston*(BOS), Belo Horizonte*(CNF), Seattle*(SEA)        | 8 |
|         | Athens*(ATH), Atlanta*(ATL), Salt Lake City*(SLC), Wisconsin*(MSN)        |   |
| Tangled | Miami*(MIA), London*(LHR), Sydney*(SYD), Paris*(CDG)                      | 8 |
|         | Los Angeles*(LAX), Enschede*(ENS)                                       |   |
|         | Washington*(IAD), Porto Alegre*(POA)†                                   |   |

However, in practice, even a rough estimation allows a far better response than using directly observed load.

We conclude that attack size estimation is close enough to help plan response to DDoS events.

5 Evaluation Approach

We next describe how we will evaluate the effectiveness of TE ($6$) and that results generalize to different deployments ($7$). Traffic engineering in response to DDoS depends on the anycast deployment—where sites are and with whom they peer. We evaluate on two different testbeds. Our approach (estimation, TE, and playbook construction) can be applied anywhere with different anycast setups. We expect network operators will execute our approaches on a test prefix (in parallel with their operational network) prior to an event so that no service interruption happens.

5.1 Anycast Testbeds

We evaluate our ideas on testbeds to see the constraints of real-world peering and deployments. We use two independent testbeds: Peering [60] and Tangled [7]. Table 2 summarizes information about each testbed with their own set of geographically distributed sites along with their locations (Peering supports more sites but we used 8 sites). These sites show different connectivity, and have one or more transits and IXP peers. Most Peering sites have academic transits while Tangled has more commercial providers. Our testbed is about the same size as many operational networks, since nearly half of real-world networks have five or fewer sites [16].

5.2 Measuring Routing Changes

To measure the effect of a BGP change, we first change the routing announcement at a site, give some time to propagate, confirm that the announcement is accepted, and finally start the anycast measurement.

Table 3: Experiment summarization and findings.

| Experiment | Key Takeaways                      |
|------------|-----------------------------------|
| Path prepending | Works everywhere to effectively de-prefer a site ($6.1.2$), but shifts traffic in large amounts ($6.1.3$), and has few traffic levels (Figure 6). |
| Neg. Prepending | Works everywhere to prefer a site ($6.1.2$). |
| BGP community strings | Although widely implemented, well-known communities are not universal ($6.2.1$). When supported, they provide finer-granularity control than prepending ($6.2.2$). |
| BGP path poisoning | Many Tier-1 ASes drop the announcements when it sees Tier-1 ASes in the paths. ($6.3.1$) Control over traffic is limited by the filters from other ASes. ($6.3.2$). |

Route convergence: After a change, we allow some time for BGP route propagation. We know that routing and forwarding tables can be inconsistent (resulting in loops or black holes) while prefix is updating [38, 67, 76]. Although routing updates are usually stable within 5 minutes [67], we wait 15 minutes for routing to settle when building our playbook since it is a non-attack period. When the attack is not mitigated after deploying a routing policy, our system moves to a different approach after 5 minutes.

Propagation of BGP policies: Policy filtering could limit the acceptance of announced routes, although in practice these limits do not affect our traffic engineering. Best practices for networks at the edge to filter out AS-Paths longer than 10 hops, and ASes in the middle often accept up to 50 hops, both more prepends than we need. Based on routing observations from multiple global locations using RIPE RIS, we confirm that configurations in our experiments are never blocked due to route filtering in multi-hops away from our anycast sites.

6 Traffic Engineering Coverage and Control

From an estimate of attack load, operators use BGP to shift traffic. We next evaluate three TE mechanisms: AS-Path prepending, community strings and path poisoning. For each we consider when it works and what degree of control it provides. Table 3 summarizes our key results from tests on two testbeds ($5.1$); in §$7$ we evaluate generalizability.

6.1 Control With Path Prepending

First we consider AS-Path prepending as a defense strategy.

6.1.1 Prepending coverage

Support for AS-Path prepending is quite complete—it requires no explicit support from the upstream provider, so we found prepending worked at all sites in both of our testbeds.
In Peering, we are allowed to use a maximum of three prepaends, and in Tangled we use up to five prepaends. Previous study [14] shows a maximum of 5 prepaends is sufficient because 90% of active ASes are located less than six AS hops away. We use RIPE RIS [57] to check the routing visibility when prepaends are in place, and we do not observe changes in the routing propagation for both testbeds. Otherwise, this might reveal the existence of AS path length filters [32, 33].

6.1.2 Does prepending work?
Since AS-Path prepending is widely supported, we next evaluate this attractive TE method.

We explore this question for a representative scenario using Peering using three sites from three continents—Europe (Amsterdam-AMS), North America (Boston-BOS) and South America (Brazil-CNF). In §7 we generalize to other configurations. We estimate load by counting /24 blocks in catchments, then compare the baseline with TE options. (We also explored traffic weighted by traffic loads instead of blocks, getting the same qualitative results and shapes with different constants, Appendix G.)

Figure 5 shows the traffic from each site under different conditions. The middle bar in each graph is the baseline, the default condition with no prepending. We then add prepending at each site, with one, two or three prepaends in each bar going to the right of center. We also consider negative prepending (§3.4.1) in one to three steps, with bars going left of center. We first consider the baseline (the middle bar) of all three graphs in Figure 5. Amsterdam (AMS, the bottom, maroon part of each bar) gets about 68% of the traffic. AMS receives more traffic than BOS and CNF because that site has two transit providers and several peers, and Amsterdam is very well connected with the rest of the world.

We next consider prepending at each site (the bars to the right of center). In each case, prepending succeeds at pushing traffic away from the site, as expected. For AMS, each prepend shifts more traffic away, with the first prepend cutting traffic from 68% to 37%, then to 29%, then to about 16%. BOS and CNF start with less traffic and prepending has a stronger effect, with one prepend sending most traffic away (at BOS, from 15% to 7%) and additional prepaends showing little further change. These non-linear changes are because changing BGP routing with prepending is based on path length, and the internet’s AS-graph is relatively flat [2, 15].

The bar graphs also show that when prepending pushed traffic away from a site, it all goes to some other site. Where it goes depends on routing and is not necessarily proportional to the split in other configurations. For example, after one prepend to AMS, more traffic goes to CNF (the top sky blue bar) than to BOS (the middle yellowish bar). These unexpected shifts are why we suggest pre-computing a “playbook” of routing options before an attack (§3.2) to guide decisions during an attack and anticipate the consequences of a change.

We also see that negative prepending succeeds at drawing traffic towards the site—in each case the bars to the left of center see more traffic in the site that is not prepending while the others prepend. AMS sees relatively little change (68% to 89%) since it already has most traffic, while BOS and CNF each gain up to 68% of traffic. All three sites show some networks that are “stuck” on that site, regardless of prepending. One reason for this stickiness is when some networks are only routable through one site because they are downstream of that exchange. We confirm this by taking traceroute to two randomly chosen blocks that are stuck at BOS. Traceroutes and geolocation (with Maxmind) confirm they are in Boston, at MIT and a Comcast network (based on the penultimate traceroute hop). We have used the local-preference BGP attribute to move such stuck blocks, but a systematic exploration of that option is future work.

In summary, the experiment shows that AS prepend does work and can shift traffic among sites, however, this traffic shift is not uniform.

6.1.3 What granularity does prepending provide?
Having established that prepending can shift traffic, we next ask: how much control does it provide? This question has two facets: how much traffic can we push away from a site or attract to it, and how many different levels are there between minimum and maximum.

Limits: Figure 5 suggested that in Peering, with those three sites, there is a limit to the traffic that can shift. AMS, BOS, and CNF always get about 16%, 7% and 3% of blocks, regardless of prepending.

Figure 6 confirms this result with a 5-site deployment (two from Europe, one from North America, one from South America and one from Australia) in our other testbed (Tangled). X axis is presented with the number of prepaends applied to each site. The number zero (0) represents the baseline, the positive numbers (1-5) are the number of prepending applied and the negative numbers represent negative prepaends. As depicted, each site can capture at most 55-65% of blocks, and can shed at most 95% of blocks, even with up to 5 prepaends. We can also see that we do not get a granular control as only three points are between the minimum and maximum.

We conclude that while prepending can be a useful tool to shift traffic, it provides relatively limited control.

6.2 Control with BGP Communities
We next show that BGP community strings have the opposite trade-off: what options they support vary from site to site, but when available, they provide more granular control over traffic. We use whatever community strings that can be supported at each site. Specific values for the same concept often vary.

6.2.1 Community string coverage
ASes must opt-in to exchange community strings with peers, as opposed to prepending’s near-universal support (since AS paths are used for loop detection, prepending works unless it is explicitly filtered out). Explicit support is required because
communities are only a tagging mechanism; the actions they trigger are at the discretion of peering AS. Prior work has studied the diverse options supported by community strings [28].

To evaluate coverage, we review support for BGP communities in the testbeds we use. The testbeds provide information about two dozen locations with diverse peers. Each one of these peers has been evaluated about its support to this feature.

In Table 4 we describe path prepending and poisoning support and what types of community strings are supported at each site. We group communities by class: advertisement options (no-peer, no-export to customers, and no export to anyone), selective prepending, and peers and transits that support selective advertisement. We also show the number of non-transit peers and transits.

Peering allows selective announcement to the transits and peers at each site, although the number of peers and transits varies. Many sites with one transit provide no alternatives. We considered selective announcement options at AMS, with 854 peers (106 bilateral peers including 2 route servers with 748 peers), and 2 transit providers [60]. CNF has one transit provider and 129 peers (with only 6 bilateral peers, other peers are connected through 2 route servers). For our Verfploeter measurement, we consider the peers and route servers with bilateral BGP sessions. A single peer covers a small fraction of the address space in our Verfploeter measurement. For some peers, we observed no coverage at all which requires further investigation with the peers to confirm our observation. Hence, all the selective announcement options do not make difference in the catchment distribution (see the catchment in AMS with 12 peers compared to the transit-1 in Figure 7). The options column of Table 4 summarizes these results, showing how many routing options we have using community strings.

We evaluate Tangled to provide a second deployment with different peers. Tangled built its anycast network over cloud providers, crowdsourced transit providers and IXPs. All transit providers and IXPs sites support communities as described in Table 4. With Tangled, the POA site has 250 peers and most of them support communities strings.

We conclude that the number of options at each anycast site may vary depending on the number of connections with peers and transits. This uncertainty shows the need for a playbook that shows the possible options.

### 6.2.2 At what granularity do community strings work?

We next examine how well community strings work and what granularity of control they provide. We use community strings to make BGP selective announcements, where we propagate our route only to specific transit providers or IXP peers.

For our experiment, we use Peering, varying announcements at AMS and observing traffic when anycast is provided from AMS, BOS, and CNF (the same topology as §6.1.2). As described in §6.2.1 selective announcement community strings are provided only at AMS and CNF, and they affect our Verfploeter measurement only at AMS with several peers together, two transits one by one, and route servers.

To select the target ASes for selective announcement, we sort all the working peers of AMS site, based on the size of their customer cone using CAIDA’s AS rank list [9]. We then choose the 6 largest IXP peers and the 12 largest, as the left two bars in Figure 7. We then examine the route server, announced separately (the next bar), and then all IXP peers including route servers. Finally, we see the coverage with
each of the two transit providers, announced separately.

First, we see that selective announcement provides more control than prepending, as AMS shifts from baseline 68% of blocks to other configurations from 53 to 6% of blocks.

Second, we see that there is some overlap in some combinations. For example, each transit reaches more than half of all blocks reachable from AMS, so we know some blocks are reachable from both transit providers. Thus, while there is some control over how many blocks to route to AMS, some peers are very “strong” and will pick up many blocks if they are allowed to announce our prefix.

Third, we see the important role of route servers. While direct coordination with 12 IXP peers brings only 7% blocks at AMS, a route server lets AMS reach more ASes and 14% of the blocks alone.

Finally, we see that transit providers play an important role. AMS site has two transit providers—BIT BV (AS12859) and Netwerkvereniging Coloclue (AS8283). Announcing to AS8283 attracts more traffic to AMS than announcing to AS12859. Different AS relationship of these two transits with AS8283 attracts more traffic to AMS than announcing to AS2914 (NTT cing to IXP peers reduces traffic from 69% to 64%. But using selective prepending and selective announcement mostly provides more “flexibility” depending on how many IXP peers and transits are connected. We also find that some sites do not provide the support that we expect which means community strings require an extra step like contacting the transit provider for an explicit agreement.

6.3 Control with Path Poisoning
We next turn to path poisoning, and show that like community strings, coverage and granularity are limited by routing filters deployed in upstream peers.

6.3.1 Poisoning coverage
Support for path poisoning is dependent on the ASes we are poisoning and on route filters deployed by our upstream ASes.

We find that many ISPs, especially Tier-1 ASes, filter out AS paths that poison any Tier-1 AS. Tier-1 ASes deploy these filters to block BGP announcements from customers that contain other Tier-1 ASes in the path to prevent route leaks [44,71]. This filtering often makes path poisoning ineffective to control traffic.

To verify that poisoning Tier-1 ASes is often ineffective from filtering, we poison Tier-1 ASes announcing only from AMS in Peering, a unicast set-up blocking the impacts of other sites, and make traceroutes from 1000 RIPE vantage points to our prefix. Our measurement shows the evidence of filters when we poison Tier-1 ASes—AS7018 (AT&T), AS6453 (Tata Communications America), and AS1299 (Telia Company). We observe many vantage points fail to reach our prefix as they are dependent on Tier-1 ASes for their routes. Some others change their paths avoiding Tier-1 ASes. We also validate route disappearance via most Tier-1 ASes using RouteViews telescopes [77].

Although poisoning Tier-1 ASes is often ineffective, poisoning is effective with most non-Tier-1 ASes. Unfortunately, these ASes carry little traffic when they are not immediate upstreams. Poisoning these small ASes only has little impact on traffic. We again traceroute after poisoning a non-Tier-1 AS (AS57866), and observe that Tier-1 ASes propagate the poisoned path. This proves poisoned paths with Tier-1 and non-Tier-1 ASes are treated differently by other ASes.

6.3.2 What granularity does poisoning provide?
Path poisoning coverage is limited because one cannot usually poison a Tier-1 AS. This same filtering limits the granular-

| Site: | AMS | BOS | CNF | SEA | ATH | ATL | SLC | MSN | MIA | LHR | IAD | CDG | LAX | ENS | SYD | POA |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Peering | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Selective prepending | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Selective announcement | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Path poisoning | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| # non-transit peers | 854 | 0 | 129 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 250 |
| # transit | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| # options | 856 | 1 | 130 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 252 |

Table 4: Traffic engineering options on each testbed sites. ✓: supported, -: not supported, △: not tested.
we conclude that path poisoning is not generally an effective approach. Table 5 identifies which combinations result in specific traffic ratios at each site. Each letter in this table refers back to a specific configuration from Table 5. During an attack, if the poisoner attacks multiple hops away because they represent little traffic. Poisoning immediate neighbors may shift traffic, but is more complex than just not announcing to them. We confirm these observations with detailed experiments in Appendix E, but we conclude that path poisoning is not generally an effective tool for traffic engineering.

6.4 Playbook Construction

Based on our understanding of prepending, communities and poisoning, we can now build a playbook of possible traffic configurations for this anycast network. In practice, we build the playbook using scripts that connect to our testbed, with the baseline of 65% blocks to a site shown in white. We group different levels of prepending (positive or negative) at each site, and show selected community string and poisoning configurations.

To summarize the many configurations from Table 5, Table 6 identifies which combinations result in specific traffic ratios at each site. Each letter in this table refers back to a specific configuration from Table 5. During an attack, if the anycast system begins at the baseline configuration (q), if AMS is overloaded, the operator could select a TE configuration higher in the table (perhaps ‘e’, ‘g’, or ‘j’). The operator can then see the implications of that TE choice on other sites (for example, ‘e’ increases load on both other sites, with ‘g’ increases load on BOS but decreases it at CNF).

An operator may also use a playbook with traffic load for two reasons. First, loads in most interesting services have diurnal pattern. Second, loads from each /24 prefix may vary because of the number of clients behind each prefix (more on Appendix G). Building the playbook with load is computationally simple; an operator can just use the same catchment mapping along with the per prefix load.

Even with attack size estimation, attacks are accompanied by uncertainty, and attacker locations may be uneven. However, the playbook provides a much better response than “just relying on informal prior experience” in two ways: the defender can anticipate the consequences of the TE action (that traffic will go somewhere!), and the defender can choose between different possible outcomes if the first is incomplete.

Playbook flexibility and completeness: Table 6 helps quantify the “flexibility” that traffic engineering allows us in this anycast deployment. Using these 10% traffic bins, we see that AMS has 9 options, CNF 7, and BOS only 6. Because AMS and CNF mostly swap traffic after TE changes, and because BOS is less well connected, no configuration with three sites allows BOS to take traffic within 50-60% range, and no 3-site configuration can drive BOS or CNF over 70%.

This analysis shows more central sites like AMS, and it may suggest the need for topology changes (perhaps adding another site in Europe or Asia to share AMS’ load).

7 Deployment Stability and Constraints

In §6 we showed BGP-based TE provides considerable flexibility. Building playbooks supports defenders by allowing them to explore how transit providers, prepending, community strings, and poisoning affect their specific deployment. We next look at how stable the results are depending on choice of sites and the number of sites. While the details of the playbook vary for each deployment, and we do not claim our testbeds represent all possible deployments, we show our approach is
flexible and can respond to attacks in different deployments—our approach generalizes.

7.1 Effects of Choice of Anycast Sites

First we see how sites affect our playbook. New sites change catchments because they depend on location and peering.

In §6.1, we studied catchments with three specific Peering sites on three continents: AMS, at a large, commercial IXP in Europe; CNF with an academic backbone transit in Brazil; and BOS, an academic site in the U.S. We now switch to three educational sites all in the United States: SEA, at University of Washington on the west coast; SLC, at the University of Utah in the Rockies; and BOS, at Northeastern University in Boston on the east coast.

More important than just geographic location, site connectivity is the most important factor in choosing sites. Multiple transit providers increase the chance of having more BGP options to affect traffic control and granularity. While a poorly connected site inside a university network tends to provide less traffic control options.

Prepending baseline: Figure 9 shows catchment sizes for the three North American sites with positive and negative prepending. Now the baseline distribution is unbalanced, but less so than before, with SEA capturing 50% of blocks. We discussed SEA’s heavy traffic with the Peering operators. They suspect that SEA is near to the Seattle IXP, making its paths one hop from many commercial providers. Which site has the greatest visibility depends on its peering and will vary from deployment to deployment.

Prepending coverage and granularity: As with our prior experiments, we can adjust prepending to see how traffic shifts. With these three sites, traffic shifts very quickly for BOS and SEA after one positive or negative prepend. SLC has more flexibility, perhaps because it has the smallest catchment at the baseline, and gains more coverage with each step of negative prepending, to 42%, 63%, and 91% of blocks. Often (but not always), we see that academic sites exhibit less granularity because either they have few peers, or their peers are academic networks with similar connectivity. As a result, minor changes in AS-Path length place one site further from the others. In addition, this less granular control shows the importance of building a playbook that is specific to a given deployment, or when the anycast topology changes.

7.2 Effects of Number of Anycast Sites

Next, we vary the number of sites and see how that changes control traffic. We select 3, 5 and 7 sites from each testbed, and build a playbook to evaluate defense options. Figure 10 shows selected configurations, grouped by number of sites.

Baseline: With more sites, overall capacity increases and baseline load at each site falls. For example, in Figure 10, the baselines (with an asterisk*) at the largest site (AMS) shifts from 70% of blocks with three sites to 61% and 56% with 5 and 7 sites. Smaller sites shift less (BOS goes from 14% to 6% and 6%, and CNF from 15% to 8% and 6%). Greater capacity and distribution requires a larger and distributed attacker to exhaust the overall service. We see similar results on our
A playbook has a limited use if routing changes immediately. Comparing baseline to one prepending in Figure 10, AMS shifts from 70% to 37% with three sites, from 61% to 29% with five, and from 56% to 23% with seven, always dropping the traffic in our system. We show that we can successfully predict traffic shifts for a new site is difficult, but experimenting with a test prefix can build a playbook pre-deployment.

Traffic flexibility: With more sites, the largest site usually shows the largest changes and has the fewest catchment sizes. Comparing baseline to one prepending in Figure 10, AMS stability over hours to days [\( \text{§H.3} \)]. We checked the stability of B-root, with its own set of “stuck blocks” that are captive to it and will not move with traffic engineering.

With more sites, the fine control of BGP communities becomes more important because path-prepending becomes less sensitive. For example, selective announcements with communities are need for AMS with 5 or 7 sites; prepending three times shifts all traffic.

New sites: Adding more sites also shows how our playbook can help guide deployment of new sites. Predicting traffic shifts for a new site is difficult, but experimenting with a test prefix can build a playbook pre-deployment.

### 7.3 Playbook Stability Over Time

A playbook has a limited use if routing changes immediately. We know routing changes when links fail, or when ISPs begin new peering or purchase new transit. For how long is a playbook applicable?

To answer this question, Table 7 shows the fraction of /24 blocks going to each catchment over time for the baseline configuration. We see that the fraction of blocks is generally quite stable, with only about 5% of blocks shifting in or out of a site. In addition, prior work has shown very strong anycast stability over hours to days [\( ?, \text{§79} \)]. We checked the stability of B-root. We found that after two weeks 0.35% prefixes, and after one month only 0.65% prefixes changed their catchment (more on Appendix F). While catchments are relatively stable, we expect operators will refresh playbooks periodically (perhaps weekly or monthly).

### 8 Defenses at Work

In this section we describe four real-world attacks processing the traffic in our system. We show that we can successfully respond to a different types of attacks in different ways.

Methodology: We use real-world attacks from B-root server operator, the Dutch National Scrubbing Center, and from an anonymized enterprise network. These events include polymorphic, adversarial, and a volumetric attack.

We evaluate these events by simulating traffic rates against a three-site anycast network. The first two events use Peering with our AMS, BOS, CNF configuration from §6. We vary this topology, using BOS, SEA, SLC from §7.1 in the last event. We replay the traffic in simulation, assigning traffic to each anycast site based on catchments measured in our experiments. We do not simulate the gradual route propagation, but instead have routing take effect 300 s after a change (a conservative bound, most routing changes happen in half that time). We then evaluate traffic levels at each site and compare that to a target capacity.

For each attack we run our system in defense, estimating the attack size and selecting a pre-computed playbook response. Since our playbook allows different responses: when we have choices we select different methods of defense: prepending, negative prepending, or community strings (Figure 11).

A 2017 polymorphic attack: Our first event is a DNS flood from 2017-03-06 in B-root [\( \text{§52} \)] (Figure 11a). This event was a volumetric polymorphic attack where the attack queries have common formats like \( \text{RANDOM}.\text{qycl520.com\032} \) (from 0 s) and \( \text{RANDOM}.\text{calling168.com\032032032} \) (changed at 4750 s, so polymorphic in nature). We assume 60k packets/s (30 Mb/s) capacity at each anycast site. The event was small enough that B-root was able to fully capture it across all active anycast sites at the time. The event lasted about 5 hours, but we show only the first 2.25 hours. Services and attacks capacity today will both be much larger; we use a small attack, scaling the attack and capacity up would show similar results.

In Figure 11a we can identity AMS site receives 100k packets/s traffic that is more than the capacity (shown as the maroon striped area). Our system notices the attack from bitrate alerts. It then estimates the AMS overload by computing the offered load using observed load and access fraction. The system maps networks to number of packets to each site using the pre-computed playbook (Table 6). Using this mapping our system/operator can then select a response. From Figure 11a, we can see the impact of the selected routing approach—announcing only to Transit-I using community string. After 300 s, we can see no striped area which indicates the attack is mitigated.

The attacker changes the query names at 4750 s, making this attack polymorphic. Filtering on query names would need to react, but our routing changes can still mitigate the attack regardless of this type of change.

A 2021 variable-length polymorphic attack: We next examine an HTTP-attack launched on an enterprise network on 2021-09-05 in Figure 11b. This polymorphic attack changes after each of three pauses. The initial attack consists of millions of HTTP GETs (15k packets/s) launched from an IoT botnet; it terminates when the enterprise’s operator deploys IP-based filtering. About 1000 s later, a different botnet launched a multi-vector attack combining HTTP GETs using random paths (to avoid caching) and spoofed TCP ACKs. We then

| Months       | AMS(%) | BOS(%) | CNF(%) |
|--------------|--------|--------|--------|
| 2020-02      | 68.1   | 14.6   | 17.3   |
| 2020-04      | 70.4   | 14.2   | 15.4   |
| 2020-06      | 65.3   | 14.1   | 20.6   |

Table 7: Percent blocks in each catchment over time.

We use real-world attacks from B-root server operator, the Dutch National Scrubbing Center, and from an anonymized enterprise network. These events include polymorphic, adversarial, and a volumetric attack.

We evaluate these events by simulating traffic rates against a three-site anycast network. The first two events use Peering with our AMS, BOS, CNF configuration from §6. We vary this topology, using BOS, SEA, SLC from §7.1 in the last event. We replay the traffic in simulation, assigning traffic to each anycast site based on catchments measured in our experiments. We do not simulate the gradual route propagation, but instead have routing take effect 300 s after a change (a conservative bound, most routing changes happen in half that time). We then evaluate traffic levels at each site and compare that to a target capacity.

For each attack we run our system in defense, estimating the attack size and selecting a pre-computed playbook response. Since our playbook allows different responses: when we have choices we select different methods of defense: prepending, negative prepending, or community strings (Figure 11).

A 2017 polymorphic attack: Our first event is a DNS flood from 2017-03-06 in B-root [\( \text{§52} \)] (Figure 11a). This event was a volumetric polymorphic attack where the attack queries have common formats like \( \text{RANDOM}.\text{qycl520.com\032} \) (from 0 s) and \( \text{RANDOM}.\text{calling168.com\032032032} \) (changed at 4750 s, so polymorphic in nature). We assume 60k packets/s (30 Mb/s) capacity at each anycast site. The event was small enough that B-root was able to fully capture it across all active anycast sites at the time. The event lasted about 5 hours, but we show only the first 2.25 hours. Services and attacks capacity today will both be much larger; we use a small attack, scaling the attack and capacity up would show similar results.

In Figure 11a we can identity AMS site receives 100k packets/s traffic that is more than the capacity (shown as the maroon striped area). Our system notices the attack from bitrate alerts. It then estimates the AMS overload by computing the offered load using observed load and access fraction. The system maps networks to number of packets to each site using the pre-computed playbook (Table 6). Using this mapping our system/operator can then select a response. From Figure 11a, we can see the impact of the selected routing approach—announcing only to Transit-I using community string. After 300 s, we can see no striped area which indicates the attack is mitigated.

The attacker changes the query names at 4750 s, making this attack polymorphic. Filtering on query names would need to react, but our routing changes can still mitigate the attack regardless of this type of change.

A 2021 variable-length polymorphic attack: We next examine an HTTP-attack launched on an enterprise network on 2021-09-05 in Figure 11b. This polymorphic attack changes after each of three pauses. The initial attack consists of millions of HTTP GETs (15k packets/s) launched from an IoT botnet; it terminates when the enterprise’s operator deploys IP-based filtering. About 1000 s later, a different botnet launched a multi-vector attack combining HTTP GETs using random paths (to avoid caching) and spoofed TCP ACKs. We then
We respond with negative prepending, with the purple cross-hatched area showing how much the traffic will overwhelm SEA, a smaller site, but can be handled at the super-site. We respond with negative prepending, with the traffic shift to BOS visible at 300 s. This response mitigates the attack (no striped area).

Other attacks: We have assessed additional attacks, and describe them in Appendix I. The additional polymorphic and volumetric attacks show that routing can successfully address attacks after routes propagate.

9 Limitations and Future Work

Our playbook of routing options (§3) is effective against many attacks (§8 and Appendix I). However, like any defense, it is not impervious. We next describe known limitations and areas of future work.

First, Internet routing is distributed, requiring time to converge. The effects of routing defenses cannot be seen until convergence. We do not make changes faster than 5 minutes.

Routing convergence time implies that routing changes will have limited applicability to short-lived attacks (less than 5 minutes). Although routing changes will not hurt the service, their benefits may not occur until routing shifts.

In addition, routing convergence means that polymorphic attacks that shift traffic sources quickly will be more effective. Routing changes are robust to polymorphic attacks that change method but take effect by traffic volume, they will spread load regardless of what it is, as we show in events in §8. However, when defending an attack where traffic shifts locations faster than routing converges, one must provision for the worst case volume to any site under the heaviest traffic it sees. Rapid shifts make defense harder, but not impossible.

Finally, we assume the anycast catchments of the underlying service change slowly (over days). We showed in §7.3 that this assumption generally holds.

Although we change routing during an attack to balance load across catchments, we do not explicitly attempt to locate attack origins. As future work, we could use such information to improve defense selection.

Attack response depends on human factors in service operators and attackers. Explicitly studying such human factors is potential future research. Our current work focused on the technical feasibility of our defenses.

10 Conclusions

This paper provides the first public evaluation of multiple anycast methods for DDoS defense. Our system estimates attack size, selects a strategy from a pre-computed playbook, and automatically performs traffic engineering (TE) to rebalance load or to advise the operator. Our contributions are attack-size estimation and playbook construction. We experimentally evaluate TE mechanisms, showing that prepending is widely available but offers limited control, while BGP communities...
and path poisoning are the opposite.

Acknowledgments: ASM Rizvi and John Heidemann’s work on this paper is supported, in part, by the DHS HSARPA Cyber Security Division via contract number IISHQDC-17-R-B0004-TTA.02-0006-L. Joao Ceron and Leandro Bertholdo’s work on this paper is supported by Netherlands Organisation for scientific research (4019020199), and European Union’s Horizon 2020 research and innovation program (830927). We would like to thank our anonymous reviewers for their valuable feedback. We are also grateful to the Peering and Tangled admins who allowed us to run measurements. We thank Dutch National Scrubbing Center for sharing DDoS data with us.

References

[1] AMPATH. Bgp resources. https://ampath.net/AMPATH_BGP_Policies.php. [Online; accessed 12-Oct-2021].
[2] APNIC. BGP-stats routing table report—Japan view. https://mailman.apnic.net/mailing-lists/bgp-stats/archive/2020/05/msg00001.html, May 1 2020.
[3] Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwälder. Lessons learned from using the ripe atlas platform for measurement research. ACM SIGCOMM Computer Communication Review, 45(3):35–42, 2015.
[4] Hitesh Ballani, Paul Francis, and Sylvia Ratnasamy. A measurement-based deployment proposal for IP anycast. In Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, pages 231–244, 2006.
[5] Ran Ben-Basat, Gil Einziger, Roy Friedman, and Yaron Kassner. Heavy hitters in streams and sliding windows. In IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, pages 1–9, 2016.
[6] Terry Benzel, Robert Braden, Dongho Kim, Clifford Neuman, Anthony Joseph, Keith Sklower, Ron Ostrenga, and Stephen Schwab. Experience with deter: a testbed for security research. In 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, 2006. TRIDENTCOM 2006., pages 10–pp. IEEE, 2006.
[7] Leandro M. Bertholdo, João M. Ceron, Wouter B. de Vries, Ricardo de Oliveira Schmidt, Lisandro Zambenedetti Granville, Roland van Rijswijk-Deij, and Aiko Pras. Tangled: A cooperative anycast testbed. In 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), pages 766–771, 2021.
[8] Matthew Caesar and Jennifer Rexford. BGP routing policies in ISP networks. IEEE Network Magazine, 19(6):5–11, November 2005.
[9] CAIDA. AS rank. https://asrank.caida.org/, 2020. [Online; accessed 12-Oct-2021].
[10] CAIDA. CAIDA UCSD BGP community dictionary. https://www.caida.org/data/bgp-communities/, 2020. [Online; accessed 12-Oct-2021].
[11] Matt Calder, Ashley Flavel, Ethan Katz-Bassett, Ratul Mahajan, and Jitendra Padhye. Analyzing the performance of an anycast CDN. In Proceedings of the 2015 Internet Measurement Conference, pages 531–537, 2015.
[12] Mark D Carney, Jeffrey A Jackson, Andrew L Bates, and Dante J Pacella. Method and apparatus for mitigating distributed denial of service attacks, November 24 2015. US Patent 9,197,666.
[13] R. Chandra, P. Traina, and T. Li. BGP communities attribute. Technical Report 1997, RFC Editor, 1996.
[14] Rocky KC Chang and Michael Lo. Inbound traffic engineering for multihomed ASs using AS path preping. IEEE network, 19(2):18–25, 2005.
[15] Yi-Ching Chiu, Brandon Schlinker, Abhishek Balaji Radhakrishnan, Ethan Katz-Bassett, and Ramesh Govindan. Are we one hop away from a better Internet? In Proceedings of the ACM Internet Measurement Conference, pages 523–529, Tokyo, Japan, October 2015. ACM.
[16] Danilo Cicalese, Jordan Augé, Diana Joumblatt, Timur Friedman, and Dario Rossi. Characterizing ipv4 anycast adoption and deployment. In Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies, pages 1–13, 2015.
[17] Danilo Cicalese and Dario Rossi. A longitudinal study of IP anycast. ACM SIGCOMM Computer Communication Review, 48(1):10–18, 2018.
[18] Cloudflare. Famous DDoS attacks | the largest DDoS attacks of all time. https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/, [Online; accessed 12-Oct-2021].
[19] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association, 106(496):1513–1527, 2011.
[20] Wouter B. de Vries, Ricardo de O. Schmidt, Wes Hardaker, John Heidemann, Pieter-Tjerk de Boer, and Aiko Pras. Verifloater: Broad and load-aware anycast mapping. In Proceedings of the ACM Internet Measurement Conference, London, UK, 2017.
[21] Christoph Dietzel, Anja Feldmann, and Thomas King. Blackholing at IXPs: On the effectiveness of DDoS mitigation in the wild. In International Conference on Passive and Active Network Measurement, pages 319–332. Springer, 2016.
[22] Ramin Ali Dousti, Frank Scalzo, and Suresh Bhogavilli. Automated ddos attack mitigation via bgp messaging, March 22 2018. US Patent App. 15/273,510.
[23] Xun Fan and John Heidemann. Selecting representative ip addresses for internet topology studies. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, pages 411–423. ACM, 2010.
[24] Xun Fan, John Heidemann, and Ramesh Govindan. Evaluating anycast in the domain name system. In 2013 Proceedings IEEE INFOCOM, pages 1681–1689. IEEE, 2013.
[25] Seyed K Fayaz, Yoshiaki Tobioka, Vyas Sekar, and Michael Bailey. Bohatei: Flexible and elastic ddos defense. In 24th USENIX Security Symposium, pages 817–832, 2015.
[26] Ashley Flavel, Pradeepkumar Mani, David Malzt, Nick Holt, Jie Liu, Yingying Chen, and Oleg Surnachev. Fastroute: A scalable load-aware anycast routing architecture for modern CDN. In 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15), pages 381–394, 2015.
[27] Ruomei Gao, Constantinos Dovrolis, and Ellen W Zegura. Interdomain ingress traffic engineering through optimized AS-path preping. In International Conference on Research in Networking, pages 647–658. Springer, 2005.
[28] Vasileios Giotssas, Georgios Smaragdakis, Christoph Dietzel, Philipp Richter, Anja Feldmann, and Arthur Berger. Inferring BGP blackholing activity in the internet. In Proceedings of the
Internet Measurement Conference, pages 1–14. ACM, 2017.

[29] T. Hardie. Distributing authoritative name servers via shared unicast addresses. Technical Report 3258, RFC Editor, 2002.

[30] Lee Hahn Holloway, Srikanth N Rao, Matthew Browning Prince, Matthieu Philippe François Tourne, Ian Gerald Pye, Ray Raymond Bejani, and Terry Paul Rodery Jr. Mitigating a denial-of-service attack in a cloud-based proxy service, October 7 2014. US Patent 8,856,924.

[31] Chi-Yao Hong, Subhasree Mandal, Mohammad Al-Fares, Min Zhu, Richard Alimi, Kondapa Naidu B., Chandan Bhagat, Sourabh Jain, Jay Kaimal, Shiay Liang, Kirill Mendelev, Steve Padgett, Faro Rabe, Saikat Ray, Malveeka Tewari, Matt Tierney, Monika Zahn, Jonathan Zolla, Joon Ong, and Amin Vahdat. B4 and after: Managing hierarchy, partitioning, and asymmetry for availability and scale in Google’s software-defined WAN. In Proceedings of the ACM SIGCOMM Conference, Budapest, Hungary, August 2018. ACM.

[32] Geoff Huston. BGP in 2017. https://labs.apnic.net/?p=1102, Jan 8 2018. [Online; accessed 12-Oct-2021].

[33] Team Cymru Inc. Secure Cisco IOS BGP template. https://www.team-cymru.com/secure-bgp-template.html. [Online; accessed 12-Oct-2021].

[34] Quan Jia, Huangxin Wang, Dan Fleck, Fei Li, Angelos Stavrou, team Cymru Inc. Secure Cisco IOS BGP template. www.team-cymru.com/secure-bgp-template.html. [Online; accessed 12-Oct-2021].

[35] Thomas Koch, Ke Li, Calvin Ardi, Ethan Katz-Bassett, Matt Calder, and John Heidemann. Anycast in context: A tale of two systems. In Proceedings of the ACM SIGCOMM Conference, Virtual, August 2021. ACM.

[36] Brian Krebs. Krebssecurity hit with record DDoS. KrebsOnSecurity, Sept, 21, 2016.

[37] Jan Harm Kuipers. Anycast for DDoS. https://essay.utwente.nl/73395/1/Kuipers_MA_EWI.pdf. 2017. [Online; accessed 12-Oct-2021].

[38] Craig Labovitz, Abha Ahuja, Abhijit Bose, and Farnam Jahanian. Delayed Internet routing convergence. ACM SIGCOMM Computer Communication Review, 30(4):175–187, 2000.

[39] Zhihao Li, Dave Levin, Neil Spring, and Bobby Bhattacharjee. Internet anycast: Performance, problems, & potential. In Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, pages 59–73, 2018.

[40] Zhihao Li, Dave Levin, Neil Spring, and Bobby Bhattacharjee. Internet anycast: Performance, problems, and potential. In Proceedings of the ACM SIGCOMM Conference, pages 59–73, Budapest, Hungary, August 2018. ACM.

[41] Qizian Liu, Bradley Huffaker, Marina Fomenkov, Nevil Brownlee, et al. Two days in the life of the DNS anycast root servers. In International Conference on Passive and Active Network Measurement, pages 125–134. Springer, 2007.

[42] Doug Madory and Matt Prosser. Excessive BGP AS path prepondering is a self-inflicted vulnerability. Presentation at RIPE 79, October 2019.

[43] Marek Majkowski. Memcrashed - major amplification attacks from UDP port 11211. https://blog.cloudflare.com/memcrashed-major-amplification-attacks-from-port-11211/, 2018. [Online; accessed 12-Oct-2021].

[44] Tyler McDaniel, Jared M Smith, and Max Schuchard. Flexseal-
Saif Hasan, Petr Lapukhov, and Hongyi Zeng. Engineering egress with Edge Fabric: Steering oceans of content to the world. In Proceedings of the ACM SIGCOMM Conference, pages 418–431, Los Angeles, CA, USA, August 2017. ACM.

[62] Ricardo de O. Schmidt, John Heidemann, and Jan Harm Kuipers. Anycast latency: How many sites are enough? In International Conference on Passive and Active Network Measurement, pages 188–200, Sydney, Australia, March 2017.

[63] Thomas Bradley Scholl. Methods and apparatus for distributed backbone internet ddos mitigation via transit providers, February 3 2015. US Patent 8,949,459.

[64] A. Shaikh, R. Tewari, and M. Agrawal. On the effectiveness of DNS-based server selection. In Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213), volume 3, pages 1801–1810 vol.3, 2001.

[65] AX Sharma. Phone calls disrupted by ongoing ddos cyber attack on voip.ms. https://arstechnica.com/gadgets/2021/09/canadian-voip-provider-hit-by-ddos-attack-voice-calls-disrupted/, 09 2021.

[66] AWS Shield. Aws shield - threat landscape report – q1 2020. https://aws.amazon.com/security/2017/09/growth-of-ddos-as-a-service-stresser-services/, 2017.

[67] R. B. da Silva and E. Souza Mota. A survey on approaches to reduce BGP interdomain routing convergence delay on the Internet. IEEE Communications Surveys & Tutorials, 19(4):2949–2984, 2017.

[68] Daniel Smith. The growth of DDoS-as-a-service: Stresser services. https://blog.radware.com/security/2017/09/growth-of-ddos-as-a-service-stresser-services/, 2017. [Online; accessed 12-Oct-2021].

[69] Donald J Smith, Michael Glenn, John A Schiel, and Christopher L Garner. Network traffic data scrubbing with services offered via anycasted addresses, May 24 2016. US Patent 9,350,706.

[70] Jared M. Smith and Max Schuchard. Routing around congestion: Defeating DDoS attacks and adverse network conditions via reactive BGP routing. In 2018 IEEE Symposium on Security and Privacy (SP), pages 599–617. IEEE, 2018.

[71] Job Snijders. Practical everyday bgp filtering with as_path filters: Peer locking. NANOG-67, Chicago, June, 2016.

[72] Oliver Spatscheck, Zakaria Al-Qudah, Seunjoon Lee, Michael Rabinovich, and Jacobus Van Der Merwe. Multi-autonomous system anycast content delivery network, December 10 2013. US Patent 8,607,014.

[73] RIPE NCC Staff. Ripe atlas: A global internet measurement network. Internet Protocol Journal, 18(3), 2015.

[74] One Step. BGP community guides. https://onestep.net/communities/, [Online; accessed 12-Oct-2021].

[75] Eric Sven-Johan Swildens, Zaide Liu, and Richard David Day. Global traffic management system using IP anycast routing and dynamic load-balancing, March 8 2011. US Patent 7,904,541.

[76] Renata Teixeira, Steve Uhlig, and Christophe Diot. BGP route propagation between neighboring domains. In International Conference on Passive and Active Network Measurement, pages 11–21, Springer, 2007.

[77] University of Oregon. Route Views Project. http://www.routeviews.org/routeviews/, 2021.

[78] USC/ISI. Usc/isi ant datasets. https://ant.isi.edu/datasets/all.html, 2019. [Online; accessed 12-Oct-2021].

[79] Lan Wei and John Heidemann. Does anycast hang up on you? In 2017 Network Traffic Measurement and Analysis Conference (TIM), pages 1–9, Dublin, Ireland, July 2017. IEEE.

[80] Fernanda Weiden and Peter Frost. Anycast as a load balancing feature. In Proceedings of the 24th international conference on Large installation system administration, pages 1–6. USENIX Association, 2010.

[81] Curt Wilson. Attack of the Shuriken: Many hands, many weapons. https://www.arbornetworks.com/blog/asser/ddos-tools/, 2012. [Online; accessed 12-Oct-2021].

Appendix A Anycast and BGP Background

IP anycast is a routing method used to route incoming requests to different locations (sites). Each site uses the same IP address, but at different geographic locations. Anycast then uses Internet routing with BGP to determine how to associate users to different sites—that is known as site’s anycast catchment. BGP has a standard path selection algorithm that considers routing policy and approximate distance [8].

Figure 12 shows a conceptual version of one of our three-site anycast deployments. Clients are split into three sites from three continents. Although we illustrate catchments by continent here, in practice they follow BGP routing rules and not geography, with users intermixed.

Although BGP is not perfect, in practice it often does a reasonably good job at associating users to nearby sites and thereby minimizing service latency [35, 40]. Moreover, anycast increases the service resiliency since it spreads traffic over multiple sites. If one site goes offline, perhaps due to maintenance, that site withdraws its BGP route and routing automatically redistributes users previously going to that site to other sites. Thus, anycast avoids service interruptions in addition to capacity expansion.

Operators can influence the routing decisions process using different traffic engineering techniques (TE) to manipulate BGP. We describe TE techniques in §3.4.1 and how they can be used to rebalance the load during a DDoS attack.

Appendix B Verfploeter Mapping

We use Verfploeter [20] to find out the client to anycast site mapping. Using Verfploeter we build our BGP playbook with various BGP changes (§6.4).

The main intuition behind Verfploeter is to send pings using an anycast prefix as source address. The replies to these pings will be routed to the nearest anycast site by the inter-domain routing system. Figure 13 shows how this works. One of the sites (green) of the anycast service runs a packet generator that sends pings (ICMP Echo Requests) to a hitlist of IP addresses. The replies (ICMP Echo Replies) from these IPs are then automatically sent to anycast prefix as source address. The replies to these pings will be routed to the nearest anycast site by the inter-domain routing system. Figure 13 shows how this works. One of the sites (green) of the anycast service runs a packet generator that sends pings (ICMP Echo Requests) to a hitlist of IP addresses. The replies (ICMP Echo Replies) from these IPs are then routed to the “closest” (in terms of routing distance) anycast site. The catchment of the site is thus determined by the IP prefixes from which ping replies arrive at that particular site.
While in principle Verfploeter can work with any type of IP hitlist, for our measurement we use a publicly available hitlist [78] based on the Fan et al. [23] methodology. This hitlist includes the IP addresses that are most likely to respond to pings for each /24 prefix in the IPv4 address space.

Appendix C Operator Assistance System

To assist operators (§3.4.3), we provide an interface for defense. To react and reconfigure the anycast network, the operators can use a web interface similar to an equalizer, choosing the percentage of load to be increased or dropped at an anycast site. The possible ranges of slider positions are based on the playbook alternatives or presets of routing policies. This process hides the playbook complexity from the operator, making the process less error-prone and more intuitive, but still giving the operator full control of the BGP routing.

In Figure 3 we can visualize a snapshot of this interface. Each slider represents an anycast site and each site has predetermined settings indicated by “notches”. The positions of the “notches” are the results of all the measurements obtained to create our playbook. The bar graph shows the results of the measurement process, indicating how many networks will be attracted to each anycast site. The operators can visualize the forecasted traffic to each position and then apply the configuration on the production network.

Appendix D Detailed Attack Size Estimation

We validate our attack estimation with both testbed experiments and real-world events.

D.1 Testbed Experiment

We validate our model with experiments in a testbed (DE-TER [6]) where we can control all factors, where actual offered load is estimated and topology is fixed.

We consider a simple topology (Figure 15) where two access links from R1 and R2 towards R3 has a capacity of 100Mb/s each. We assume the link from the service router (R3) to the servers has 1 Gb/s capacity, so the internal network is never a bottleneck.

Here we use a slightly unequal legitimate traffic—40 Mb/s from R1 to R3 and 60 Mb/s from R2 to R3. As part of legitimate traffic we generate known-good traffic that we use for estimation (§3.3): 2 Mb/s on R1-R3 link and 3 Mb/s on R2-R3.

The attack consists of 250 Mb/s of traffic following the distribution of 100 Mb/s on R1-R3 and 150 Mb/s on R2-R3. Offered load is therefore 140 Mb/s (1.4 × link capacity) and 210 Mb/s (2.1 × link capacity) on the two links, for a total offered load of 350 Mb/s (29.2 k queries/second).

We show our estimation works well with the testbed experiment in Table 1. Our estimation needs to know the access fraction or $\alpha$ (how much traffic arrives at the system during an attack). We observe the rate of known-good traffic at the server to find the $\alpha$. Table 1 shows the expected typical
Table 8: Load distribution with Peering catchment and B-root load. Catchment: 2020-02-24, Load: 2020-02-25 and 2020-02-26 (only showing selected policies). Catchment distribution remains similar over the course of the day showing by a single value.

Figure 16: Peering: Impact of path poisoning (from AMS on 2021-04-09).

Figure 17: Tangled: Impact of path poisoning (from MIA on 2021-04-11).

Figure 18: One month of catchment stability in B-root.

Figure 19: Peering: Impact of path prepending in catchment distribution with ATH, BOS and CNF sites on 2020-05-30.
With poisoning coverage limited by filters (§6.3.1), we next show that our approach works for another event from 2016-06-25.

Using the observed offered load (“observed”) of 16.3 kq/s at the server, and $\alpha$ of 0.49, we estimate 33.2 kq/s offered load (“estimated”), which is close to the actual reported 29.2 kq/s (“reported”). Normal offered load (“normal”) is 8.5 kq/s, which is lower than what we can observe during an attack. Our observation is significantly lower than the reported or estimated rate, which tells us the importance of estimation.

We carried out other tested experiments, varying topology, traffic ratios, and distribution of attackers. In general, we find our approach works well unless attack traffic is highly unbalanced (one or few sources, not a distributed DoS).

D.2 Case Studies: 2016-06-25 Event

We showed real-world case studies in §4.1. Here we show that our approach works for another event from 2016-06-25 (Figure 14). We observe our estimation (varying orange line) is close to the reported line (dashed purple line). We can also see that our observation is only a tiny fraction of the true offered load (bottom blue line).

Both these results show the effectiveness of our approach with both testbeds and real-world events.

Appendix E BGP Poisoning Granularity

With poisoning coverage limited by filters (§6.3.1), we next examine what granularity control it provides. We expect to see limited range since we cannot poison Tier-1 ASes, and small ASes carry little traffic.

We test path poisoning in both Peering and Tangled using three sites from each testbed. As expected, we observe the same traffic distribution when we poison any Tier-1 AS—30-35% load at AMS (Peering in Figure 16) and 1-3% load at MIA (Tangled in Figure 17).

When we poison a non-Tier-1 AS that is more than one hop away, we observe a small change in the traffic distribution. In Peering, we can see that poisoning AS57866 reduces a small fraction of traffic from AMS (Figure 16). We observe a similar outcome in Tangled (Figure 17).

Our results prove that poisoning Tier-1 ASes is limited by the filters, and poisoning non-Tier-1 ASes that are multi-hops away can change only a small fraction of traffic.

Poisoning an immediate upstream is equivalent to not announcing to them, so we do not consider that case here.

Appendix F Catchment Stability

Our insight is that we can use a playbook for several days or weeks since the catchment remains stable over the time (§7.3). To test this we use one month of B-root catchment mapping with test and production prefixes. We observe the stability in B-root catchment.

From Figure 18, we can see that the catchment remains stable over time. In two weeks, only 0.35% prefixes, and in one month, only 0.65% prefixes changed their catchment when we compare the catchment with day 1 considering ∼2 million prefixes. This shows only a tiny fraction of prefixes changes the catchment even after a month irrespective of the changes made by the ASes. Hence, building the playbook once every week/month should be sufficient.

We also make the catchment mapping at different times of the day. We found catchment distribution remains similar at different times of the day.

Appendix G Load Distribution

Our playbook with catchment (§6.4) distribution gives an adequate prediction of traffic distribution which we successfully apply in §8. Since services care about load, we want to see how the load is distributed in different routing changes. An operator can simply make the load playbook based on
the already computed catchment mapping without making additional BGP announcements.

In Table 8, we can see different routing changes and their impacts over load distribution in different times of the day. Load changes over the day—fewer load at 00 GMT in AMS site since most Europe sleeps at that time. BOS and CNF receive more load at 00 GMT as that is a busy hour for these two regions. We can also observe that some prefixes contribute more load due to the difference in number of clients behind each prefix. For this reason, BOS prefixes (mostly North American prefixes) contribute less load compared to the prefixes at other two sites. We can also see that load remains stable at the same time of different days (varies within 5% most of the time).

We can also see that the relative catchment distribution follows the load distribution, however, it is not exactly the same. Decisions will be even better when an operator considers different load playbooks at different times of the day. Building multiple load playbooks is simple since we can just use the same catchment mapping (catchment mapping remains stable (Appendix F)).

Appendix H BGP in Other Anycast Set-up

In this section, we present the complementary results from the experiments performed in both testbeds. Here we present data used to generalize our findings.

H.1 Peering: A Small Site in Europe

AMS in Peering is well-connected with two transits, and several IXP peers. Next, instead of AMS, we take ATH in Europe which is connected through a research network in Greece. Our goal is to see whether the findings from §6 are still valid in this anycast setup. Like the previous setup, we also take BOS and CNF.

Figure 19 shows the catchment distribution. In the baseline case, as ATH is not connected like AMS, it gets only 20% traffic. CNF serves almost 50% traffic in this anycast setup. So, the traffic distribution is still skewed in this setup. Even if both AMS and ATH are in Europe, we see a different catchment control which indicates the importance of site’s connectivity.

Prepending works similarly in this setup. BOS and CNF can cut most of their traffic after first prepend. However, ATH can only shift 7% traffic after first prepend, and for more preends it does not show any effect. In all these sites, we can also see that some blocks are always “stuck” to a particular site. Using negative prependning, we can push most of the traffic to BOS and CNF. However, we can only push 40% traffic to ATH site.

H.2 Peering: Sites in Nearby Location

Next, we take three sites that are in nearby location and have similar connectivity. We select sites from Boston (BOS), Atlanta (ATL) and Wisconsin (MSN) in Peering. All these sites are located within the eastern half of the U.S., and they are connected through education network—BOS with Northeastern University, ATL with Georgia Institute of Technology, and MSN with University of Wisconsin - Madison.

When the sites are located in nearby geo-location, and connected by similar network, path prependning can result in an “all or none” outcome. When we prepend from ATL or
BOS, most traffic goes away from these sites. Prepending one time from ATL leaves no traffic in ATL, and prepending one time from BOS cuts traffic from 42% to 14%. Even with negative prepending, BOS can get over 90% traffic, and ATL can get nearly 90% traffic. So, BOS and ATL can cut or gain almost all the traffic with positive and negative prepending.

MSN receives a small fraction of traffic in the baseline, and some blocks are always “stuck” at MSN. With two negative prepends, MSN receives only 27% catchment, however, with the third negative prepend, MSN receives almost 80% catchment. This slow and sudden increase in the catchment suggests us why we need a BGP “playbook” for anycast setup.

H.3 More Sites in Tangled

We want to confirm that increasing the number of sites in Tangled shows the similar results that we get in §7.2.

Like Peering, we take 3, 5 and 7 sites in Tangled testbed (Figure 21). As the overall capacity increases with more number of sites, in Tangled also we can see the baseline traffic is reduced in each site as the traffic spreads out. LHR gets 55%, 31% and 25% traffic when there are 3, 5 and 7 sites respectively.

With more sites, as there are more capacity in other sites, one site can cut almost all of its traffic. For example, LHR can cut all of its traffic when there are 3 or 5 sites.

We can see a “shadowing” instance when we have a 7-site testbed. After adding ENS and POA, we can see that all traffic from LAX site disappears (Figure 21). We believe Cogent traffic now shifts from LAX to POA, and academic traffic shifts from LAX to ENS. We explain this issue in §7.2 when IAD shadows LAX.

Like Peering, adding more sites can create new options where the shifted traffic goes. For example, with 3 sites, LHR traffic goes to MIA when we make prepending. But when we have 7 sites, a significant amount of LHR traffic goes to POA—POA traffic increases from 26% to 40% when we prepend from LHR. Hence, it is necessary to keep a “playbook” to see the traffic distribution after a BGP change.

Appendix I More Attacks And Mitigation

We evaluate more attack events captured at B-root and at the Dutch National Scrubbing Center. We follow the same methodology mentioned in §8. We use the same playbook built with AMS, BOS and CNF sites ($6$) from Peering.

A polymorphic event at B-root: We observed a polymorphic event at B-root on 2017-02-21 where the attackers used three different query names—RANDOM.phone.tianxinov.cn(032, RANDOM-clgc88.com(032, and RANDOM.jiang.com(032. The total offered load at AMS site exceeds the capacity of 60k packets/s (striped area).

Our system announces only to Transit-1 using community strings to mitigate this attack (Figure 22a). We can see that there is no striped area after the deployment of the new routing policy. Also, when the attackers change pattern, our system does not need to make any routing changes. This proves the applicability of TE approaches in polymorphic events.

A 2020 volumetric attack at B-root: We observed an ephemeral volumetric event at B-root on 2020-02-14 where the attackers used a single query name—peacecorps.gov. This event lasted very briefly for 3 minutes. In practice, no routing approach can work against such short-lived attacks due to the propagation delay of BGP. We stretched the event with similar traffic rate so that we can see the impact if the attack continues for more time.

In this event also, AMS is overloaded with 60k packets/s when the assumed capacity is 40k packets/s (Figure 22b). We prepend AMS by 1 so that the traffic shifts away from AMS. After 300 s, we can see no overloaded striped area in AMS.

These volumetric attacks are common at root servers. Routing based approaches can defend against such attacks.

A DNS amplification attack: We evaluate another DNS amplification attack collected at the Dutch National Scrubbing Center on 2021-08-22. In this event too, AMS site receives a huge traffic exceeding its capacity (large striped area). Announcing only to IXP peers, our system can mitigate this attack event (Figure 22c). This is another example where a community string helps us to mitigate the attack.

A 2021 B-root event where our system iterates: We evaluate another event at B-root occurred on 2021-05-28. In this event, the queries were IP fragmented (large packet size), and the common query name was pizzaseo.com (we stretched the event since it was short-lived). When the attack started, our system finds AMS site overloaded (Figure 22d). Our system finds prepending from AMS is the best approach to reduce traffic from AMS. However, after prepending AMS by 1, CNF site gets the most redirected traffic, and becomes overloaded. Redirected attack sources prefer CNF over BOS. When our system finds CNF site overloaded, it deploys an approach that will reduce traffic from CNF since it is now overloaded. Our system deploys negative prepending to push more traffic towards BOS site. After 900 s, we can see there is no overloaded site. This event shows how our system can gradually find out the best routing approach.

Defending with community strings: We next consider an attack observed at the Dutch National Scrubbing Center on 2021-08-27. This attack was a volumetric DNS amplification.

In this attack, AMS is overloaded. Consulting the playbook, we select a response using community strings to shift traffic, retaining six IXP peers at AMS, while dropping all other peers and transits. The impact of this change is visible at 300 s in Figure 22e, as the attack is successfully spread across all sites.

This example shows how different community strings provide control over traffic distribution.