Qi Points: Placing Decoy Routers in the Internet

Devashish Gosain, Anshika Agarwal, H. B. Acharya and Sambuddho Chakravarty
Indraprastha Institute of Information Technology Delhi (IIITD), New Delhi, India
{devashishg,anshika1448,acharya,sambuddho}@iiitd.ac.in

***Abstract*—Decoy Routing, the use of routers (rather than end hosts) as proxies, is a new direction in anti-censorship research. However, existing proposals require control of hundreds of Autonomous Systems (AS) to provide Decoy Routing to Internet users in a single censorious country (*e.g.* China). This is considered necessary, as the adversary - in this case the Chinese Government - has connections to many Autonomous Systems (ASes), and we want to make sure it cannot simply route around those ASes which have decoy routers.**

In this paper, we present a new approach to the question of placing decoy routers. In decoy routing, the router intercepts messages en route to an overt destination and proxies them to covert destinations. Instead of trying to capture flows from an entire country, as proposed, we stipulate that the overt destination be a well known site (such as Alexa top-100), and concentrate on the AS-level paths to these sites. We construct a map of the structure of the Internet, as a graph of such AS-level paths and present a new way to identify key points - *those few ASes which appear on a large fraction of paths leading to these popular websites*. Our method yields results an order of magnitude cheaper than earlier proposals, and needs to be run only once, rather than for each censorious country. (We also identify the key routers inside a few key ASes.) Our results indicate that decoy routing is much more powerful than previously believed: using our new approach to place decoy routers, we need very few (less than 0.1% of Internet AS) to force an adversary to route through them. However, while the number of key ASes is small, the number of key routers in these ASes may be quite large – a new challenge for decoy routing.

I. Introduction

Anti-censorship, the circumvention of censorship by governments or their ISPs, uses a few fundamental techniques. One of the most important anti-censorship techniques for Internet traffic, is *proxying* - the use of a "go-between" to communicate when direct connections are forbidden. A user, who wishes to access a censored resource, connects to the innocuous-looking proxy server instead; in turn, the proxy connects to the target resource, and forwards messages from the user. However, all proxies - simple HTTP and CGI proxies, tunneling over VPN to a VPN host, rerouting over `Tor` [1] - share a weakness. To be used, any proxy service must be discoverable by users. The mechanism used by a legitimate user to find the proxy can also be employed by the adversary (censorious ISP or government). In time, the adversary finds the proxy, and adds it to the blacklist so the user can no longer connect to it. The current state-of-the-art is the Chinese government's attempt to find and subvert Tor traffic [2].

A recent attempt to disrupt this dynamic is *Decoy Routing* [3, 4, 5, 6, 7]. Decoy Routing is an anti-censorship scheme that employs routers (rather than end hosts) as proxies. Packets, sent from the client to an apparent "overt" destination, are intercepted en route by the "decoy" router. The router identifies the packets using a secret handshake, extracts their contents, and diverts the message towards the true "covert" destination. The idea, as per Karlin *et al.* [3], is that routers are very hard for the adversary to blacklist. Firstly, Internet routing is federated, rather than decided at the source; more importantly, a government that blocks known decoy routers will pay an unacceptable price - it becomes disconnected from a large fraction of the Internet. (It is still important to remain undetected, to protect the *users* from repercussions. But the router itself is too important to block.) Houmansadr [4] reports that it is possible to provide decoy routing services to 100% of clients around the world, if we can convert the routers of two major Autonomous Systems (ASes) into Decoy Routers.

However, the fundamental idea, that "routers are too important to blacklist", has come under attack. Schuchard *et al.* [8] propose a new attack - Routing Around Decoys (RAD) - where the adversary redirects network flows to avoid decoy routers. Mapping the Internet at AS level (Autonomous Systems and their connections), they argue that government-level adversaries (China, Iran, etc.) have connections to many ASes, and thus enough alternative paths to route around a particular AS. In other words, there are no essential ASes where we could place decoy routers:

- Avoiding the top 100 ASes (by degree in the CAIDA connectivity graph) only cuts China off from 2.3% of Internet destinations.

Houmansadr *et al.* [9] reply that AS connections are directional, reflecting business relationships (*provider*, *customer*, and *peer*) [10], and a RAD attack is very costly:

- The latency of China's Internet routes increase by a factor of 8
- 44 ASes have to become "transit" ASes *i.e.* carry traffic to other ASes (China only has 30 transit ASes to begin with)
- The load on some transit ASes increases dramatically (by up to 2800 times)

They also demonstrate that, if we choose ASes intelligently, we need only place decoy routers in around 2% of ASes. If China avoids these ASes, it is cut off from 30% of the ASes in the Internet.

There is a huge gap between requiring only two ASes to provide decoy routing to all clients, and requiring some 900 ASes (i.e. 2%) for only clients in China. (This sharp contrast arises because, in the first case, we only say there exists a path with decoy routers on it - that a client may discover by probing - whereas in the second case, we cover all paths, so the government cannot avoid decoy routers.) It seems quite

impractical to use hundreds of ASes to provide decoy routing to users in one country. Is it possible to do better?

In this paper, we present a fresh approach to the problem of placing decoy routers.

- Our first contribution is a new way to find "key ASes" in the Internet. We apply the approach of Gao and Qiu [11] to real BGP routing tables [12], and map the actual path of traffic from each AS in the Internet to the ASes hosting the most important WWW destinations (Alexa top 100 sites).
  Using this map, we select a small number of ASes that appear in a large fraction of the paths (*i.e.* we use the metric named *path frequency*).
  Our map is not complete - it only deals with paths to the top 100 web destinations - so we also perform cross-validation. We construct the paths from all ASes to some more destinations (Alexa-101 to 225, and also the top-50 destinations from each of nine known censorious countries), and measure how well our chosen ASes cover these new paths. (They achieve over 90% coverage.)
  We name this new algorithm Overt Destination-based Sorted Placement (ODSP). [The name reflects that, while previous approaches select ASes that cover the most BGP routes from a target censorious country, we select the ASes that cover the most routes to target destinations. Our thinking is that these popular websites - search engines, social media, online shopping sites, or cloud providers, for instance - see a lot of traffic, and are thus also good candidate "overt destinations" for decoy routing. It also makes the algorithm efficient - we need to run it only once, and not country by country!]
- Various metrics have been used to identify key ASes. Schuchard et al use AS degree (the number of neighbors of an AS), and Houmansadr, path frequency and customer-cone size (total count of customer ASes, customers of customers, etc.) Our second contribution is to show that customer cone size is not in fact a good predictor of importance of an AS. We also provide an explanation for why path frequency is a better metric. We add in passing that our results also demonstrate that common heuristics, such as selecting Tier-1 ASes (those with no providers), are not correct in practice.
- Most importantly, we are the first to investigate the internal structure of an AS to identify "key points" for Decoy Routing.
  Previous studies of the practicality of Decoy Routing have focused only on the AS-level map of the Internet. But a large AS typically has *thousands* of routing elements. Which of these should be replaced with Decoy Routers?
  In order to answer this question, we map the internal structure of some ASes using `traceroute`, using the approach of Mahajan et al [13]. We demonstrate how to identify the routers in an AS that intercept a large fraction of the AS traffic, and are therefore suitable for use as Decoy Routers.
  Our results show that the fraction of significant routers (i.e. those that intercept a large fraction of paths) is highly variable across ASes. For example, for the ASes we mapped, the number of routers intercepting over 90% of the paths varies between 70 and 200. This shows that though the number of candidate ASes may be small, the actual number of key routers may be significantly greater. We also show that the common approach, of replacing the edge routers of the AS with Decoy Routers, is not a good heuristic.

Our results imply that the Internet is much less able to "route around decoys" than previously thought. 30 ASes, i.e. 0.068% of the world ASes, intercept about 4.5 million paths, i.e. over 90% of paths to the most popular WWW destinations, as per Alexa [14]. If the adversary does indeed perform a RAD attack *et al.* [8] against these ASes, it would disconnect most network paths - not only to the Alexa top 100, but to most other destinations. (In our tests, about 91% of the paths originating from Iran and about 99% of the paths originating from China were disconnected). This has consequences beyond the inconvenience of citizens of these countries: a censor also (inadvertently) blocks the other users of the paths that transit through ASes in the country. For instance, a whopping 92.25% of network paths connecting Chinese ASes to these popular destinations also serve other ASes, that are customers of Chinese networks, but located outside China. It is hard to justify subjecting so many users to "collateral damage", especially as the censorship policies in their home countries might not be compatible with those enforced by China.

However, while we demonstrate that control of a few key ASes allows us to intercept most flows in the Internet (and thus, provide Decoy Routing services to clients), we also show that the cost of such Decoy Routing may still be very high. An ISP that decides to collaborate in implementing Decoy Routing may have to switch out hundreds of switches and routers. In our experience, a single element costs around 2500 USD to 50000 USD [15], so the cost is of the order of tens of millions of dollars, plus implementation and downtime costs. Convincing a major commercial ISP to do this (especially considering that many must adopt these changes at once - a single ISP AS with Decoy Routers can easily be Routed Around) remains a challenge for future work.

We begin with a discussion of background and related research, in the next section.

II. Background and Related Research

This section presents the relevant background for our work, and a brief discussion of how it fits into the existing literature.

A. Background

Our work in this paper involves mapping the Internet and finding the best places to put decoy routers. The context for this work comes from two areas of research: decoy routing and anti-censorship, and network tomography i.e. mapping the structure of the Internet.

*1) **Network anti-censorship**:* There exists a large body of work studying the use of proxy servers to circumvent censorship. A good survey is provided by Leberknight [16]. The current state of the art is to use Tor [1]. (Onion routing was originally designed to ensure anonymity over the Internet, but

its use of encryption to protect privacy also makes it suitable for evading censorship, and it has extensive infrastructure.) Unfortunately, there exist techniques to detect TLS flows carrying Tor [17]. More generally, traffic for most proxy based solutions can be detected and censored [18, 19], even if camouflaged [20].

Decoy Routing [3, 5, 4, 6] takes a new direction where proxying is performed by special network routers called *Decoy Routers (DRs)*.

The user of decoy routing is hosted within a censorious ISP network, but wishes to communicate with network destinations censored by its ISP. To achieve this, it sends packets addressed to an innocuous-looking website, known as the *overt destination (OD)*. (The packets are encrypted using TLS, so the ISP cannot see the contents, and the header shows that they are meant for an unfiltered destination.) As they seem harmless, these packets are allowed out of the censoring ISP. However, the packets are not what they seem: they carry a small, secret message, using bits in the TLS header (e.g. the TLS nonce field). On their way to the overt destination, if the packets pass through a Decoy Router, this message acts as a secret handshake. Instead of forwarding them, the DR identifies them, decrypts their payload (the key, the TLS shared secret, is also sent as part of the secret message), and establishes a new connection to the filtered site - the true, *covert destination (CD)*. This procedure, end-to-middle censorship circumvention, is shown in Figure 1.

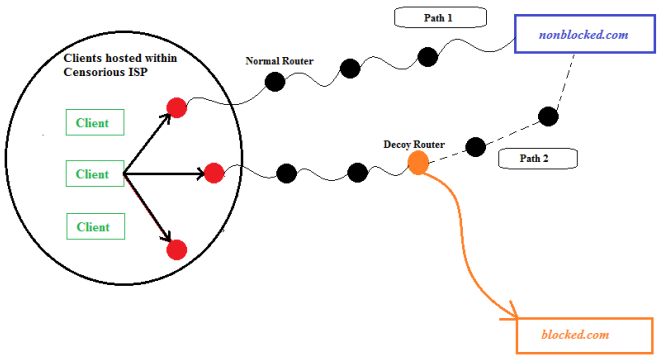

Fig. 1. Clients residing within censorious ISP send packets, apparently to the OD `unblocked.com`. They traverse a DR en route, which identifies them as special packets when it sees the secret message. The packets are decrypted and proxied to the CD `blocked.com`

Several systems have been built to implement decoy routing: Telex [5], Cirripede [4], TapDance [6] and Rebound [7]. These systems differ in features (message replay protection, tolerance of asymmetry in routing, inline blocking of traffic to/from overt destination, secret handshake, *etc.*, but share the basic design given above: all use routers, rather than end hosts, as proxies. This design decision stems from the realization that it is much harder for the censor to prevent the packets passing through a router, than it is to block an end host.

*2) **Network tomography**:* Network tomography [21], the mapping of the structure of the Internet, involves several kinds of map. The Internet consists of routers and hosts, organized into networks called Autonomous Systems (ASes) that decide how to route traffic among themselves. Besides physical connections, there must be an acceptable business relationship before an AS will route traffic over a link. There is also the question of the geographic locations of networks and hosts.

For our research, we require:

- Maps representing the connections between ASes
- IP-level connectivity within these ASes.

AS-level mapping.: The approach taken in earlier work, such as the CAIDA Ark project [22], involves mapping routes with `traceroute`. Traceroute returns router-level paths from a source to a destination, hop by hop; the map is built by running traceroute from distributed volunteer nodes to various /24 prefixes. This data is consolidated into a graph where the nodes represent ASes, and edges represent links between them.

However, the CAIDA approach simply finds all available paths, and makes no distinction between heavily-used paths and ones that only exist in theory. *The actual path between ASes may be decided by factors other than path length* [1] We use a different approach, following Gao and Qiu [11]. This algorithm builds paths from a given IP prefix to every AS of the Internet, using publicly-available BGP routes (which we obtain from Internet Exchange Points across the globe [12]). It also allows us to infer inter-AS relationships.

Router-level mapping. A large AS, such as an ISP, generally has several thousand routers. Just as we mapped the inter-AS graph using BGP information, it is possible to map their internal structure using their SNMP Management Information Bases (MIBs) [23]. However, we have no access to this data. Instead, we map the routers and connections in ASes of interest using the Rocketfuel algorithm [13].

In brief, Rocketfuel runs `traceroute` probes from looking glass servers [24] to prefixes inside a chosen AS. We also resolve IP aliases [2] using `Midar` [25], and perform reverse DNS lookups. Our mapping is similar to that of CAIDA, but has much fresher data.

B. Related Research

Our paper contributes to the study of how best to place decoy routers in the Internet, a question first raised by Houmansadr *et al.* [4]. The authors claim that if DRs are placed in two tier-1 ASes, they can serve all clients. (The client sequentially probes the Alexa top-30 sites. On at least one such attempt, its packets pass through a Decoy Router and are proxied to the true, covert destination.)

Schuchard *et al.* [8] seem less positive. According to their paper - Routing Around Decoys (RAD), a powerful adversary like China can detect the locations of decoy routers (and the

[1] ASes pay other ASes to route their traffic. A customer AS will, of course, route traffic through its providers - but providers do not route "through" traffic through their customers. The only traffic a provider sends a customer, is meant for that customer, or *its* customers, and so on. The AS-level path between two hosts on the Internet is said to follow a "valley free" path, as the path first rises - an AS, then its provider, then a provider of the provider, etc.; peaks - or plateaus, as it crosses through several peering links - and then descends, through provider-to-customer links, until it reaches the destination. There are no "valleys" in the path; no provider-to-customer links between two customer-to-provider links, or vice versa.

[2] Different interfaces of the same router, with different IP addresses, are called IP aliases

clients using them), *and can also evade decoy routing*. In their Internet topology, they show that there are many paths between two hosts in the Internet. The "warden" adversary (here, the Chinese Government) loses very little connectivity if it blocks all traffic through decoy router ASes and insists on using some other route.

Houmansadr et al. [9] argue that this statement is far too strong. RAD interprets Internet topology as simple graph of AS connectivity, and ignores how business relationships affect routing decisions. In practice, AS connections are chosen for routing using the *valley free* model [10].

Also, the authors admit, random choice of ASes is likely to perform poorly. 86.2% of the ASes on the Internet are edge ASes, and see only their own traffic. Instead, they propose the following strategies.

1) *Sorted placement*: Decoy routers are placed in the ASes the most often appear in the adversarys routing tables, but are not controlled by the adversary directly (or indirectly through business relationships).
2) *Strategic random*: ASes are chosen randomly, as in Schuchard [8]. However, a selected AS must have a large enough *customer-cone*. (Customer cone refers to customers, customers of customers, etc. In other words, a selected AS must be a significant provider to other ASes.) Also, the AS should not be controlled by the adversary.

They demonstrate, that using their AS selection strategy, it is possible to cut China off from 30% of all ASes, with only 2% of ASes (≈ 900 ASes).

It may be noted that their approach requires a fresh estimation of ASes for every adversary. Moreover, the use of customer-cone size as a metric to choose ASes, tacitly implies that it is a good predictor of how many flows they carry and how frequently they appear in routing paths. Our results show otherwise; we observed a weak correlation between the customer-cone sizes of ASes and the fraction of network paths that transit them (see Section VI.

Our approach - ODSP, an improved approach for applying the path frequency metric on AS paths leading towards popular websites (potential ODs), does not require such re-estimations for individual nations.

We inspect our data and infer that clients probing both popular (Alexa top-100) and not so popular (Alexa top-225 locations and sites that are popular in nine censorious nations) sites can encounter DR hosting ASes en route $\approx 99\%$ of the times in atmost two attempts. Attempts to avoid these ASes, cuts-off most of these censorious nations from a very large fraction ($>90\%$ for most of these nations) of paths to popular websites (which includes popular search engines, social-media sites and cloud services).

III. Motivation

The problem in this paper is to determine where in the Internet we should place DRs, in order to have the maximum impact. This impact is measured by the fraction of network paths passing through an AS (and the routers within it). (An AS present on many paths may able to provide decoy routing to a large number of users, and conversely, a censor who refuses to use these paths is cut-off from much of the important Internet based services *e.g.* search engines, could services and social-media.)

In order to find the *key* routers, i.e. a small number of routers that carry most of the Internet traffic, we proceed in two phases. First, we identify the key ASes in the Internet; next, we identify the key routers in these ASes.

The current state of the art [9] proposes two criteria for choosing ASes:

1) They are well linked with the networks of censorious regimes.
2) Their customer-cone size exceeds some threshold.

However, this approach has several significant limitations, as follows.

1) The approach requires new ASes to be identified for each censorious nation, without yielding a small set of ASes to positioning DRs. For *e.g.*, their approach requires placing DRs in nearly 900 ASes to circumvent censorship by a single country (China).
2) *Customer-cone size, by itself, is not a good metric for selecting candidate ASes.* An AS with a large customer cone provides Internet access to a large number of other ASes, and may intuitively be a backbone AS, carrying a large volume of traffic. Unfortunately, this appealing hypothesis does not hold up well in practice. *As we see in section V, the ASes with large customer-cone sizes do not necessarily transport most of the flows (customer cone size is, therefore, a rather sub-optimal metric for choosing an AS to place DRs).* We describe why this happens, with evidence, in Section VI.
3) A DR is one single router, whereas a large AS consists of thousands of routers[3]. It is not enough to determine which AS to place a router in. We also need to know the internal structure of the AS, and which specific routers transport all or most of the network traffic. (These key routers are the ones that may be replaced by DRs.)

In order to address these limitations, we construct a map of the Internet, and select the ASes that occur most frequently in our paths (taken from BGP routing tables). Next, we map four ASes of particular interest, to identify their key routers; this allows us to estimate the number of DRs we need to be able to intercept a large fraction of Internet traffic. In the following sections, we explain our approach in detail (SectionIV), and provide our experimental results (Section V) and analysis (Section VI).

IV. Data Collection Methodology and Decoy Router Placement Strategy

A. Network mapping process

Our network mapping process consists of two phases. First, we build an AS-level Internet map, using the paths connecting popular WWW destinations and the various ASes

[3]Sometimes placed across several countries

of the Internet. From these paths, we identify ASes of high path frequency (those that appear in a large number of paths) as "key ASes" (for hosting DRs).

In the second phase, we estimate the router-level topology of key ASes to identify key routers - the actual routers inside the ASes that transport the majority of traffic. As most traffic flows pass through these routers, replacing a relatively small number of key routers with DRs will suffice to build a viable Decoy Routing infrastructure.

Generating AS level maps: For the first phase of network mapping, we use the approach presented by Gao *et al.* [11]. Their approach uses existing AS paths appearing in BGP tables collected from a number of Internet Exchange Points(IXes) [12], and infers paths that do not explicitly appear. Existing BGP paths are augmented by appending other ASes that frequently appear adjacent to path ASes, and which do not invalidate the path's *Valley-Free* property. The aim is to build paths connecting a given IP prefix to all ASes in the Internet.

For our analysis, we used a snapshot of BGP routes corresponding to March 1, 2016, derived by merging the Routing Information Bases (RIBs) obtained from 15 Internet Exchange Points. Taking the IP address prefixes corresponding to the top-100 most-popular sites, we generate paths connecting all ASes to these prefixes.

Finally, we ranked the ASes according to the number of times they appear in different paths. (These ASes, and the fraction of paths they intercept, are presented in Section V).

Creating router level maps: After identifying key ASes in the Internet, as above, we are still left with the problem of where in the AS to put DRs. An AS involves a complex topology of routers and hosts. Even the AS administrator, who may know the internal topology, may not know how frequently a router appears in actual network paths. [For instance, our tests show that not every edge router appears frequently; replacing the edge routers with DRs is a naive strategy.] When approaching the AS admin to ask them to implement Decoy Routing, it is helpful to estimate how many (and which) routers they will need to replace. To that end, we map the topology of routers inside some key ASes, so as to identify the actual routers that potentially transport large fraction of the ASes' traffic.

For this, we employed Rocketfuel [13]. The original authors, Mahajan et al., executed `traceroute` probes from about 200 looking glass servers (presented in `www.traceroute.org`) to all the IP prefixes in a chosen AS. Many of the servers are now unavailable; instead, we used 390 `planetlab` [26] nodes, hosted in educational institutions across the globe.

For each chosen AS, we identified the prefixes it advertised (on `cidr-report.org`). We then ran `traceroute` probes targeted to one representative IP address in every prefix. This gave us a router-level path ending inside the AS.

Next, using Whois [27], we inspected each `traceroute` trace to identify the first IP address belonging to the target AS. These are potentially the edge routers of the AS. We trim the traces up to these addresses, as we only need paths inside the AS.

The router IPs (belonging to the target AS), discovered through the above process, are quite noisy and suffer from problems such as anonymity and aliasing [21]. To clean them, we used the state-of-the-art alias resolution tool, `Midar` [25].

Finally, we combined the paths into a map of the AS, and selected routers appearing in a large number of `traceroute` probes as candidates for DR replacement.

B. ODSP: Overt Destination based Sorted Placement

Our placement strategy is to select the key ASes that appear in a large number of paths leading to potential overt destinations. We now present it formally.

Algorithm 1 ODSP: AS and Router selection

procedure FINDAS(List of Websites $WList$, Routing Tables RT, Threshold $Thold = 0.9$) ▷ Routing Tables are from RouteViews. Target Websites are from Alexa.
  PathList = (empty list), ASList = (empty list)
  for every website w_i in $WList$ **do**
    Find $w_i.p$, the prefix corresponding to w_i
    PathList.append (ExtendPath(RT, $w_i.p$))
    Add all new AS found to ASList
  end for ▷ ExtendPath uses Gao's algorithm. We compute the paths to prefix $w_i.p$ using the routing information in RT.
  Sort ASList by frequency with which ASes appear in paths of PathList
  Return list of most common k ASes, excluding ASes in censorious countries, where k is the smallest number s.t. k ASes cover a fraction of at least $Thold$ of paths (in PathList) ▷ We find that the list of ASes, that cover 90% of all paths, is quite small - about 30 ASes.
end procedure

procedure FINDROUTER(AS A, Threshold t)
  Map internal structure of AS A with Rocketfuel
  Remove aliases with MIDAR
  Sort routers by frequency with which they appear in traceroute traces.
  return list of most common k routers, where k is the smallest number s.t. k routers cover a fraction of at least t of traces through the AS.
end procedure

C. Advantages

The ODSP algorithm, as presented in the previous subsection, promises the following advantages.

1) The placement of Decoy Routers is global; they serve paths from ASes all over the Internet, rather than only paths from censorious countries (as seen in [9]).
2) The ASes selected are located far away from the adversary nations, and thus outside their geo-political and economic sphere of control. This makes it more difficult to bring pressure to bear on them. [Note: This is the only part of ODSP that is sensitive to the identity of adversary nations. If new nations become censorious, the algorithm may have to be run again, to select trustworthy ASes.]

3) The ASes selected through ODSP lie on a very large fraction of paths. This makes it hard for adversaries like RAD [8] to bypass them without risking disconnection from a correspondingly large fraction of the most popular Internet based services. Indeed, the overt destinations are themselves popular WWW destinations in most censorious nations (as per [14]).

Caveat: It may be argued that our the map of the Internet is incomplete. Why did we stop at the 100 most popular websites, and What about paths to the all other IP prefixes? In fact, our original motivation was to take the most famous websites as overt destinations, but we found that the ASes we chose (being large, multinational network providers) appear frequently in network paths in general. Besides our target sites, they also cover the paths to "less-popular" sites (ranked >100), as well as the sites popular in censorious nations. We give details and figures in Section VI.

An additional advantage of our approach, is that paths derived from BGP RIBs better represent the true paths connecting ASes to various network destinations. Moreover, unlike previous approaches involving CBGP [28] simulator, that requires actual router level configuration files, Gao *et al.*'s approach places no such constraints. In contrast, the paths obtained from other network tomography projects such as CAIDA Ark [22] may or may not be the actual paths chosen by packets, at a given point in time.

V. Data and Evaluations

A. Identification of Key ASes

As described in the previous section, our first step is to build a map of the Internet, using AS paths connecting the prefixes of Alexa top-100 most visited sites to all the ASes of the Internet. We obtained a total of $4,497,547$ paths.

Thereafter, the ASes in these paths are sorted (and ranked) in descending order of *path frequency*, i.e., the number of paths that traverse them. Figure 2 presents the cumulative fraction of paths that pass through these ASes.

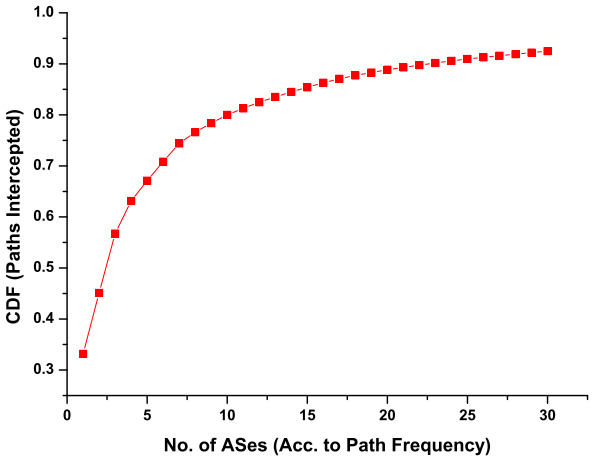

Fig. 2. CDF of top-30 ASes ranked according to fraction of paths that they intercept (for top-100 websites from alexa). These may collaborate for DR placement.

The X-axis represents the top-30 ASes ranked by their path frequencies; the Y-axis represents the actual fraction of paths. The details of these top-30 ASes are summarized in Table I.

In our results, the highest-ranked AS, AS3356 (Level 3 Communications), intercepts $1,492,079$ paths ($\approx 33\%$ of total paths). The next highest, AS174 (Cogent Communications), intercepts $536,752$ more paths (not counting overlaps, i.e. paths intercepted by both); in other words, the top 2 ASes intercept $2,028,831 (= 1,492,079 + 536,752)$ unique IP-prefix-to-AS paths, about $\approx 45\%$ of all the paths. The top 30-ASes by path frequency together intercept 92.4% of all paths.

ASN	Country	Rank (P_{freq})	Rank (C_{size})
3356	US	1	1
174	US	2	2
2914	US	3	5
1299	SE	4	4
3257	DE	5	3
6939	US	6	13
6461	US	7	8
6453	US	8	52
7018	US	9	17
10310	US	10	6
4134*	CN	11	10
3549	US	12	79
4837*	CN	13	85
209	US	14	19
9002	UA	15	97
6762*	IT	16	7
8359*	RU	17	22
2828	US	18	30
20485*	RU	19	21
16509	US	20	9
9498*	IN	21	18
4323	US	22	16
3216*	RU	23	99
2497	JP	24	15
701	US	25	12
12956	ES	26	65
37100	MU	27	23
4826*	AU	28	26
12389*	RU	29	67
1335	US	30	92

TABLE I. Top 30 ASes. These intercept more than 90% of paths. ASes headquartered in potentially censorious nations are starred and highlighted.

The table presents the corresponding AS numbers (ASNs), along with their hosting country and their ranks based on path frequency and customer-cone sizes. ASes highlighted in red and marked with a $\star$ correspond to countries that are known to censor Internet traffic, such as Russia and China. (This assessment is based on the censorship ratings by Freedom House Report [29] and the Open Net Initiative (ONI) [30].) Though some of these ASes are traversed by a large fraction of paths, we do not consider it safe to choose them as DR hosts.

Number of prefix-to-AS paths intercepted by individual ASes: Figure 3 presents the number of network paths that pass through each of the individual top-50 ASes. (Note: These figures are not individuated. For example, $52,333$ of $536,752$ (9.75%) paths traversing AS174 also pass through AS3556.)

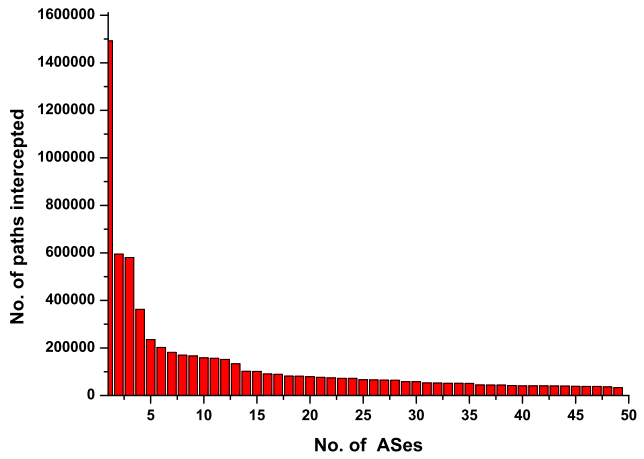

Fig. 3. No. of paths intercepted by each of the top-50 ASes sorted by path frequencies.

This could be used to determine which individual ASes could be selected to maximize decoy traffic interception. In situation where some ASes don't collaborate, the numbers may give insight into which other ASes can be individually selected for DR placement.

Our first (naive) approach to capture decoy routing traffic suggested that decoy routers be placed in the home ASes of (the prefixes of) potential Overt Destinations. However, we realized that the censor can easily block specific Overt Destinations. Instead, we now select ASes that appear in most of the prefix to AS paths. This has two main advantages:

- The solution works for many different Overt Destinations (as the routing "backbone" of the Internet is common to most routes).
- We need to control a much smaller number of ASes (30 instead of over 100).

Key ASes excluding censorious regimes

As described in Table I, several key ASes are hosted in censorious regimes. We excluded them from the list of top-30 AS and included the next best ASes (by path frequency), as per Table II. The AS path frequency ranks come from the original list (which included the potentially censorious ASNs). Thereafter, we re-estimated the proportion of paths covered by ASes in non-censorious nations alone (excluding the ASes in censorious nations). The results of these re-estimations are presented in Figure 4.

We see that the path frequencies follow a similar trend to those seen in the top-30 ASes (including potentially hostile ones). As per our results the top-30 ASes *hosted in non-censorious regimes* are sufficient to intercept a very large fraction, $\approx 90\%$, of the prefix-to-AS paths (not substantially different from the results presented previously in figure 2.)

Key ASes for traffic to specific popular destinations:

We now present our analysis for the paths to (the prefixes corresponding to) some specific popular destinations on the web (as opposed to the aggregate of top-100 popular destinations). Our case studies include popular social media sites, search engines, commercial media software products sites and a very popular `Wiki` site. We determined the ASes which intercept a large fraction of paths. The results of these computations are presented in Figure 5.

ASN	Country	Rank (P_{freq})	Rank (C_{score})
13030	SW	31	84
1273	UK	32	83
16735	BZ	33	98
6830	EU	34	91
18881	BZ	35	95
3491	US	36	42
10026	HK	37	87
32787	US	39	93
1239	US	46	45

TABLE II. ASES OWNED BY NON-CENSORIOUS NATIONS RANKED BY PATH FREQUENCY (RANKS >30 AND <50)

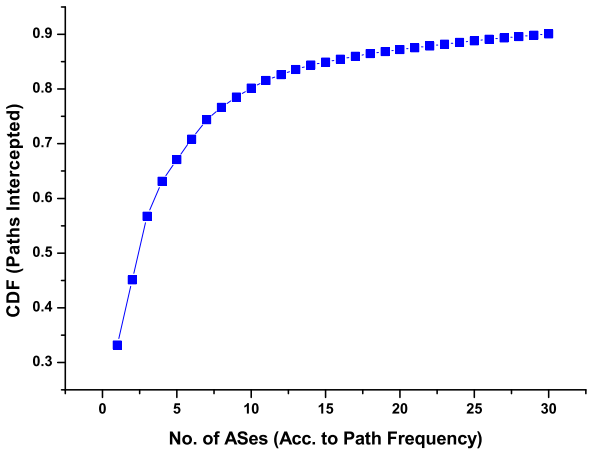

Fig. 4. CDF of top-30 ASes (hosted in non-censorious ASes) ranked according to fraction of paths that they intercept.

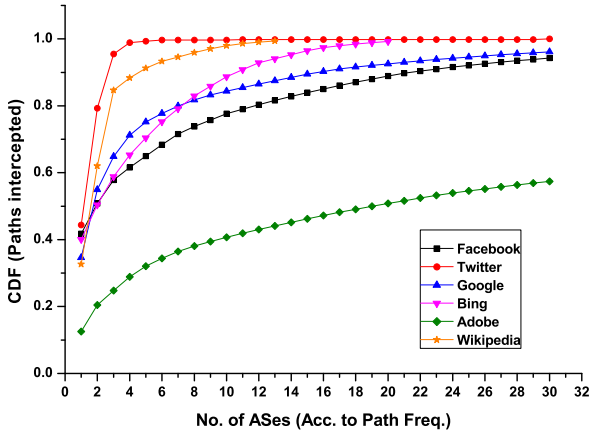

Fig. 5. CDF of paths to popular destinations, intercepted by individual ASes.

We observe a trend similar to the one presented in Figure 2. About 15 ASes collectively transport paths to about over 80% of the AS-paths leading to these destinations (except for

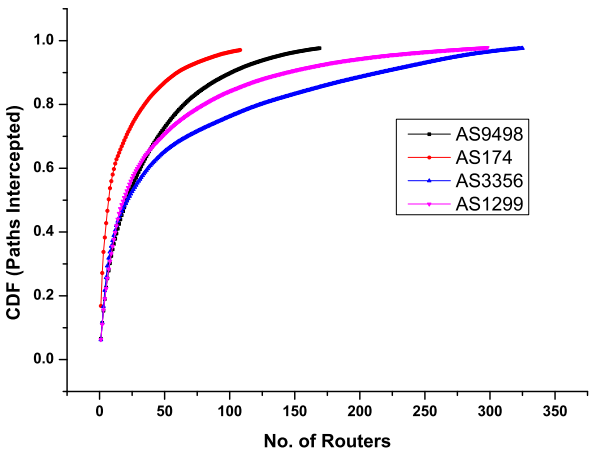

Fig. 6. CDF of routers appearing in `traceroute` paths

Adobe[4]) These 15 ASes are among the original top-50 ASes ranked as per (prefix-to-AS) path frequency.

Only 5 ASes collectively transport all the paths to the prefix corresponding to `twitter.com`, while about 18 ASes intercept all paths carrying traffic from `bing.com`. Finally, about 30 ASes cumulatively transport about 98% of the paths corresponding to `google.com` and `facebook.com`.

B. Identifying key routers inside key ASes

The second part of our research involves identifying the key routers inside important ASes. These are the routers that transport a large fraction of network traffic; replacing these routers with DRs would allow us to offer Decoy Routing services to all traffic through the AS.

As described in the previous section, we run `traceroute` probes from about 390 `planetlab` hosts to IP prefixes inside the target ASes. For our experiments, we chose 4 of the top-30 ASes identified above – AS3556, AS174, AS9498 and AS1299. While AS3556 (Level-3 Communications) and AS174 (Cogent Communications) are headquartered in the US, AS1299 (TeliaSonera) is headquartered in Sweden, and AS9498 (Airtel) in India.

Our probe results were cleaned using the `midar` tool to resolve aliases. From the final results, in figure 6, we obtained statistics regarding the number of routers which appeared most in our `traceroute` paths.

In some ASes (AS174), a few (≈ 70) routers appear in over 80% of the traced paths. In others, such as AS3356, many more routers (>250) are required to get good path coverage (90% of the paths). Table III presents the statistics.

The table not only presents the number of edge and core routers, but also those that appear in the most paths, covering a large fraction of the network topology. In case of AS3556, `traceroute` probes to its 674 advertised prefixes revealed about 1010 routers (707 edge + 303 core). We need more than half of these, 574 (513 edge + 61 core) of the most frequently seen routers, to cover (most of) the `traceroute` paths. In contrast, for AS174, we need only 288 (113 edge + 175 core) out of 1737 routers (165 edge + 1572 core).

[4]The trend appears bit different for `Adobe.com` because most of the BGP paths terminate at its provider AS3556, which aggregates the routes for `Adobe.com`. A very small fraction terminate at its own AS, hence not considered for path frequency computations.

ASN	No. of Edge Routers	No. of Core Routers	No. of prefixes advertised
3356	513/707	61/303	674
174	113/165	175/1572	706
1299	252/493	265/1989	167
9498	235/320	34/199	120

TABLE III. NO. OF ROUTERS (EDGE AND CORE) INSIDE THE AS THAT APPEAR IN MOST OF THE PATHS

We see that we can do considerably better than the naive solution of replacing the edge routers of the AS. While replacing edge routers would indeed intercept all the traffic entering and leaving an AS, several of the edge routers rarely intercept traffic, while some ("backbone") core routers are much more significant. Our mapping approach allows us to get good coverage with a smaller number of routers. However, in some ASes (AS3556 for instance), the number of required decoy routers is still quite large.

We mention in passing that, in contrast to the results of Mahajan *et al.* [13], reverse DNS lookup was unsuccessful for most the routers that we discovered. Thus it was hard to determine information such as the type of router, location *etc.* which we had hoped to discover from the DNS name corresponding to the router.

Hardware and software resources used: Our AS-level map uses BGP routes. Accordingly, we collected a snapshot of Routing Information Base data (dated 1st March 2016) from 15 Internet Exchange points, using routeviews [12]. We also used the inferred AS relationships from CAIDA [31].

To construct the AS-level map, we used virtual machines with a total of 10 CPU cores (x64) and 24 GB RAM, running Ubuntu Linux (14.04LTS server). Our multi-threaded implementation of Gao's [11] algorithm took $\approx 3-4$ hours to compute paths to 10 prefixes.

To identify key routers in an AS, we ran `traceroute` probes from about 390 `planetlab` machines to a randomly selected IP in each prefix advertised by the AS. Depending upon the number of prefixes advertised, and network latency of the paths, it took approximately $18-36$ hours to complete probing an AS; alias resolution took a further $5-8$ hours on our local VMs.

VI. DATA ANALYSIS AND DISCUSSION

A. Path frequency vs customer-cone size

As mentioned previously in Section III, we chose ASes having high frequency of appearance in prefix-to-AS path, rather than customer-cone size. Our claim is that customer-cone size is not a reliable metric to identify the ASes that transport a large fraction of traffic. We explain why, with an example. Consider the AS graph in Figure 7.

The figure represents a hypothetical AS graph where node A represents an AS with the highest customer-cone size of 6,

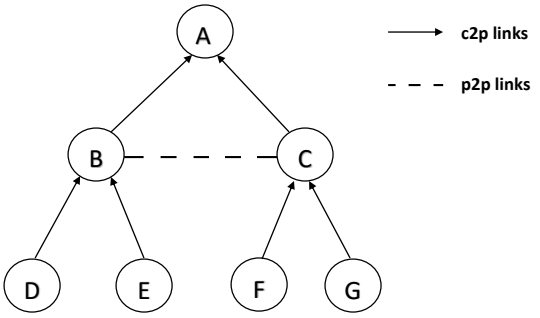

Fig. 7. Schematic AS graph with multiple valid valley-free paths: $D-B-E$, $D-B-C-F$, $D-B-C-G$, $D-B-A-C-F$, $D-B-A-C-G$, $E-B-A-C-F$ and $E-B-A-C-G$. Some of these do not traverse A, the AS with the highest customer-cone size.

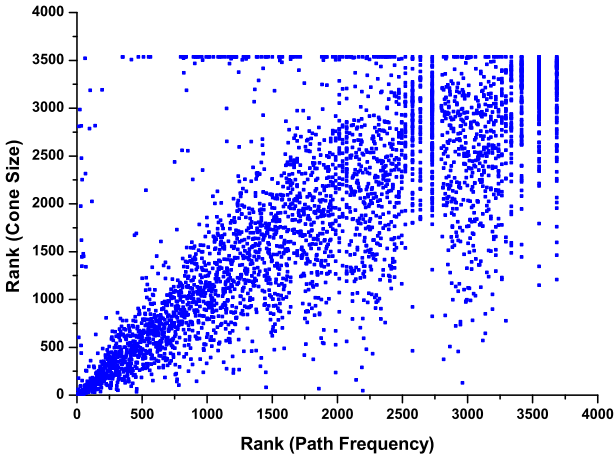

Fig. 8. AS Rank variation based on path frequency and cone size for all transit ASes (corresponding to the non-terminal nodes of the prefix-to-AS paths)

the total number of ASes that A can reach via its customers and their customers (D, B, E, F, C, G). ASes B and C have customer cones of size 2 (for each of the individual nodes).

There are several valid valley free paths in this hypothetical AS graph: $D-B-E$, $D-B-C-F$, $D-B-C-G$, $D-B-A-C-F$, $D-B-A-C-G$, $E-B-A-C-F$ and $E-B-A-C-G$. However, as evident from the example, not all of them pass through the *root AS*, *i.e.* the node with the highest customer-cone size.

We present the top-30 ASes by path frequency, along with their ranks by customer-cone size (obtained from [32]), in Table I. As is clearly evident, customer-cone sizes and AS path frequencies are not well correlated. More formally, Spearman's Rank Correlation metric for AS ranks based on prefix-to-AS path frequency and based on customer-cone size (as shown in Figure 8) is only 0.2.

We also identified several AS paths that traversed the customers of ASes with very large customer cones, without traversing the said root AS. We show some examples in Table IV. The second column shows the fraction of paths that do not traverse the root AS, but do traverse its immediate customers. The third column, the fraction of paths that do actually traverse the root AS. As we can see, for example, 34.16% of the paths to top-100 IP prefixes traverse the AS with the largest customer cone, AS3356 (cone size = $24,553$). But nearly as many paths, 33.17%, pass through its 1-hop (immediate) customers, and DRs placed in AS3356 would not be able to intercept the traffic on these paths. For example - as we see in Figure 9, the traffic through AS9002 to AS2818 (www.bbc.co.uk) does not pass through AS3356, though it is the provider to both these ASes. This analysis can be further extended to n-hop customers of AS3356.

ASN	% of path not reaching the AS	% of path reaching the AS
3356	34.16	33.17
174	29.05	13.13
2914	28.16	12.90
1299	36.50	8.05
3257	21.00	5.23
6939	7.46	4.40
6461	5.13	4.03
6453	26.00	3.76
7018	7.40	3.70
10310	0.07	3.52

TABLE IV. FRACTION OF PREFIX-TO-AS PATHS THAT DO NOT TRAVERSE ASES WITH LARGE CUSTOMER-CONES BUT THEIR 1-HOP IMMEDIATE CUSTOMERS ALONG WITH THE FRACTION OF PATHS INDIVIDUALLY TRANSPORTED BY THE AS

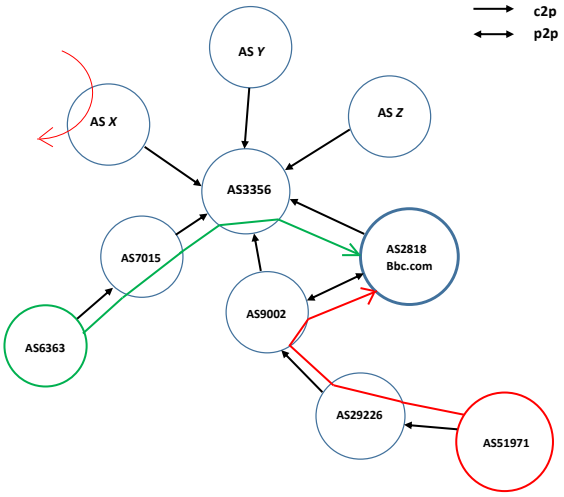

Fig. 9. Valley free paths not reaching AS3356 with largest cone size. Green lines indicate network paths that traverses AS3356 to reach AS2818 directly. Red lines correspond to network paths that traverse the one-hop customers of AS3356, without traversing AS3356 itself.

We conclude that, given the considerable fraction of paths which do not transit "root" ASes with large cone sizes (preferring to transit through their customer ASes instead), customer-cone size is not a *sufficient* parameter to choose key ASes (for DR placement, etc.)

B. Stability of the results

Our data shows that a small fraction of ASes ($\approx$ 30) together intercept over 90% of the total paths to popular web destinations, or more precisely, to the prefixes corresponding to them. However it is difficult to judge the universality of our AS selection heuristic and the ASes identified. This raises several questions.

1) What if a censorious nation simply blocks the Alexa top-100 destinations?
2) What if it produces alternatives that are popular among local users?

Would we need to re-estimate the important ASes, and would the results be dramatically different?

The Top-30 key ASes intercept a large fraction of paths leading to *other* destinations also. To cross-check the importance of the identified ASes (in terms of paths intercepted), we estimated AS paths to sites which were globally ranked 101–225 by Alexa. The 30 key ASes we identified, intercepted over 90% of the paths to these sites as well (see Figure 10).

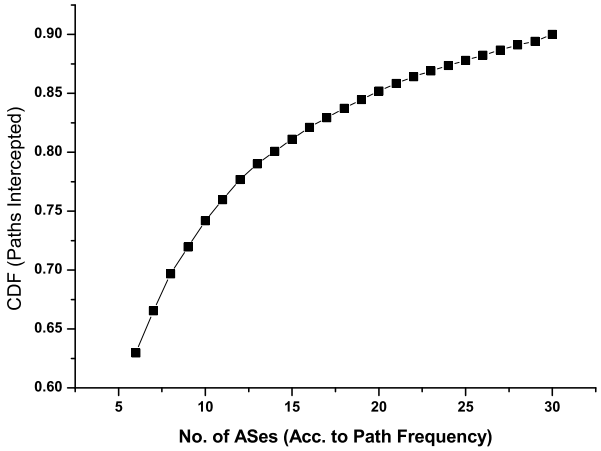

Fig. 10. CDF of top-30 ASes ranked according to fraction of paths that they intercept (for top 101 to 225 websites from Alexa).

The second question is more pernicious. One may argue, with examples of Iran or China, that loss of these paths means little to a censorious regime, as they have their own homegrown substitutes (`facenama.com` and `renren.com` respectively for `facebook.com`).

However, as is evident from the Alexa website, popular web destinations in censorious countries include several of the top-100 globally popular sites (search engines, social-media sites, cloud services, e-commerce sites *etc.*) alongside local websites. While it is true that the choice of network destinations may vary across nations (*e.g.* based on user's choice of language), web access may not be as "insular" as one might fear.

To check, we selected nine censorious nations – China, Venezuela, Russia, Syria, Bahrain, Pakistan, Saudi Arabia, Egypt and Iran – and constructed paths to the top-50 popular websites in each of these nations (a total of around 400 destinations). We observed that our chosen ASes intercept $\approx 93\%$ of the paths originating or transiting these censorious nations (see Figure 11). In other words, our chosen ASes - some of which are multinational top-tier network providers - together intercept a very large fraction of the paths to the WWW destinations popular in these countries.

Caveat: Inspecting the data, we infer that clients probing both popular (Alexa top-100) and other (Alexa top-225, sites popular in censorious nations) sites, encounter our chosen

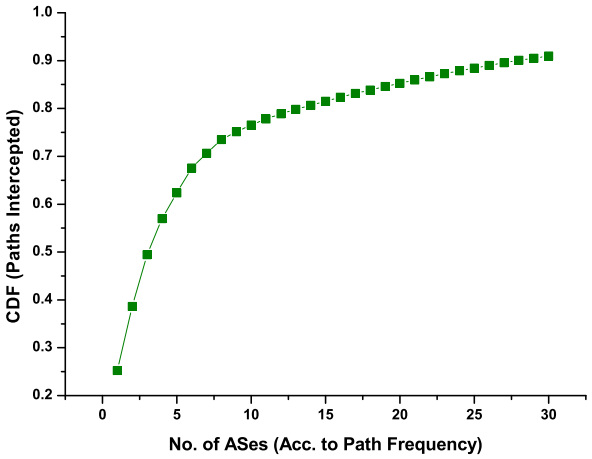

Fig. 11. CDF of top-30 ASes ranked according to fraction of paths that they intercept (for popular websites of potentially censorious nations).

ASes en route $\approx 99\%$ of the times in at most two probe attempts. In other words, we are not only able to provide decoy routing services to all clients, but 99% of the time they can find a decoy router in two probes (as opposed to up to 30 probes required when using only two ASes [4]).

1) Impact of avoiding ASes selected though ODSP.: **Fraction of paths disconnected for censorious nations:** In previous work [9], the authors describe how to place decoy routers, emphasizing that avoiding them (as seen in RAD [8]) would disconnect the adversary about from about $30-40\%$ of most other ASes of the Internet. For lot of users (and their hosting ISPs and nations), disconnection from $30-40\%$ of most other ASes might be less damaging compared to being disconnected from most of the popular Internet based services (*e.g.* Alexa top-200 sites), that includes most popular search engines, social media sites, streaming media sharing sites *etc.* As mentioned above, several of these appear among the top-100 sites accessed by users in most even in censorious nations.

We analyzed the fraction of prefix-to-AS paths in various censorious nations that would be disconnected if an adversary (like the RAD adversary by [8]) forces user traffic to bypass the ASes selected using ODSP strategy.

Avoiding the top-30 ASes (ranked by path frequency), would disconnect about 98% of all paths from Chinese ASes to prefixes of most popular destinations. Our complete results, showing the fraction of paths disconnected across 11 censorious nations, are is presented in Figure 12. The horizontal axis represents the country names (using 2-letter initials), and the vertical axis shows the fraction of the paths disconnected if the nation decides to bypass the ASes selected with ODSP.

As in the case of China, we observe very high fractions of actual paths being disconnected, in most cases above 80%.

In passing, we note that over 85% of the ASes of the Internet are customer ASes, while transit ASes make up for less than 15% of the ASes (≈ 4500 ASes). A successful RAD [8] attack involves poisoning the BGP updates for several thousand BGP routers simultaneously, an impractical challenge for most censorious regimes.

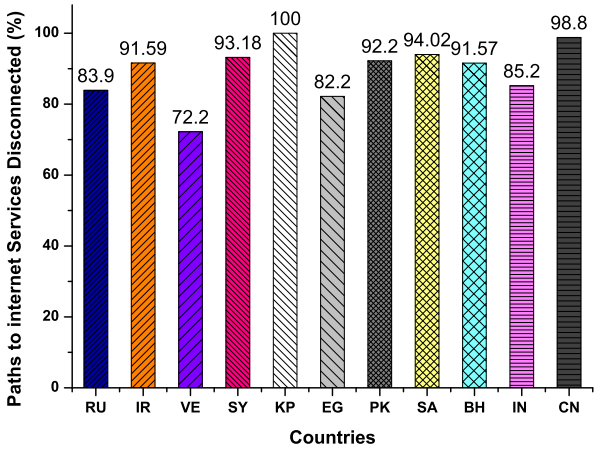

Fig. 12. Fraction of paths disconnected from major WWW destinations (for 11 censorious nations)

Finally, filtering traffic to popular WWW destinations (potential ODs) may not only affect the users of the AS - it also causes the inadvertent disconnection of users in customer ASes which route traffic through the censorious AS [33].

Collateral damage from censoring popular WWW destinations: For some of the known censorious ASes we computed the number of customer AS paths which would be disconnected if the said ASes filtered sites going to the popular WWW destinations. The fraction of traffic disconnected for each individual censorious nation is presented in Figure 13.

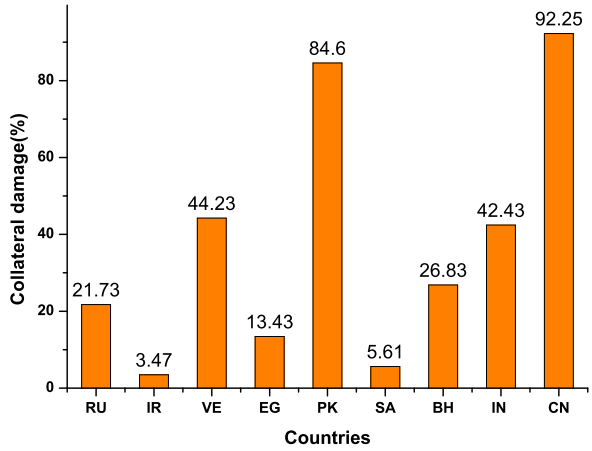

Fig. 13. Fraction of paths that would be disconnected if censorious nations filtered traffic to potential ODs (top-100 Alexa sites).

The histograms represent the fraction of paths that transit or originate in ASes outside the censorious nations, before traversing ASes hosted within the said nations. For example, in case of China, $306,874$ AS paths visited or originated from an AS outside China[4] This constitutes 92.25% of the $332,742$ paths connecting Chinese ASes to the popular destinations.

Filtering traffic to potential Overt Destinations, besides inconveniencing the AS' own customers, may potentially disrupt connectivity of customer ASes, over which the censorship policies of the censorious AS do not have jurisdiction.

[4]362 particularly interesting paths originated at a Chinese AS, passed through non-Chinese ASes, then re-entered China and passed through one or more Chinese ASes, before finally leaving for its destination.

C. Limitations and future efforts

AS path estimation: Previous effort considered using CBGP [28] to build models of the BGP routing decisions. The approach, apart from BGP tables, also relies on router configuration files. In contrast we use Gao *et al.*'s approach to determine the IP prefix-to-AS path for all ASes, using BGP tables derived from various IXes [12]. Gao *et al.*'s approach yields a single optimized actual paths connecting customer ASes to various IP prefixes at any given point of time, without requiring router level configuration data.

Further, our approach, involving real BGP paths connecting users to actual destinations, is more accurate than the topologies of the CAIDA-Ark project [22], which relies on `traceroute` probes executed from different vantage hosts. The CAIDA topology presents an AS graph, it might not be easy to estimate the actual path of traffic from a source AS to a prefix (on-demand).

However, our path estimation strategy is limited on by the quality of BGP routes we receive from the Routeviews [12] project. Publicly available BGP routes are not free of artifacts of misconfiguration and bogus advertisements [34].

Router level topology estimation: Our router-level mapping of intra-AS topology is performed using `traceroute` probes from various `planetlab` hosts to IP addresses in the AS.

In several situations, the routers filter the probes. Thus we are limited by the fraction of routers discoverable through the `traceroute` probes. Moreover, in most cases, reverse DNS look-ups for most router IPs did not succeed. Thus it is hard to guess details regarding the physical location of routers (like what was shown by Mahajan *et al.*).

Moreover, we have we determined the router level topology of a small number of ASes. As part of our future efforts, we plan to map and measure more ASes, to get a better picture of the overall fraction of routers that need to be replaced with Decoy Routers to create a worldwide Decoy Routing infrastructure.

Finally, since router level topology is far more dynamic than AS level peering information, which is based on business relationships, for successfully mapping the routers it would be best to determine the backbone routers and routers whose connectivity does not change considerably over time. These maybe good sites to place decoy routers.

VII. Concluding Remarks

In this paper, we have made several contributions towards answering the question of placing decoy routers on the Internet.

1) Our first contribution, ODSP, is an improved heuristic to identify ASes to place DRs. It involves applying path frequency metric to AS paths leading to popular websites (potential ODS). The process yields a small set of candidate ASes (≈ 30) which intercepts a very large fraction of AS paths ($>90\%$) to both popular sites, and, as far as we see, most others.
2) It is evident from our data, that DRs hosted in such ASes are very hard to "route around". Adversaries attempting this, lose connectivity to most of the highly-popular websites (in some cases as $>98\%$

of the paths are disconnected). Moreover, filtering traffic destined to such highly-popular sites may also collaterally cut-off non-censorious nations from such sites.

3) We describe, through our data analysis, why customer-cone size may not be a good metric to choose candidate ASes. We show that is it is poorly correlated to AS path frequency metric.
4) An AS is not, in practice, a simple entity. We consider the question of which routers in an AS should be replaced with Decoy Routers; for this purpose, we map the internal structure of some key ASes. We find that, in practice, the key routers in an AS, that potentially intercept large fraction of actual network paths are quite distributed - both edge and core routers.

Thus, to conclude, while it is feasible to choose a small set of ASes (≈ 30) that cover a large fraction of AS paths, it is quite a different matter to choose routers in an AS. In fact, to intercept all the flows in a single key AS, we need of the order of 500 routers, where all of which are not edge routers as often common intuition suggests. In other words, building a worldwide decoy routing infrastructure is clearly *possible*, but not trivial. It will require very strong incentives to get a key AS (which is usually a commercial ISP) to deploy the required decoy routing infrastructure.

While if it seems somewhat expensive to intercept flows on the Internet (*i.e.* requires control over a large number of routers), it may also be expensive for an adversary to filter all the traffic on the Internet. We intend to explore this question in our future work.

REFERENCES

[1] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," DTIC Document, Tech. Rep., 2004.

[2] P. Winter, R. Köwer, M. Mulazzani, M. Huber, S. Schrittwieser, S. Lindskog, and E. Weippl, "Spoiled onions: Exposing malicious tor exit relays," in *Privacy Enhancing Technologies*. Springer, 2014, pp. 304–331.

[3] J. Karlin, D. Ellard, A. W. Jackson, C. E. Jones, G. Lauer, D. Mankins, and W. T. Strayer, "Decoy routing: Toward unblockable internet communication." in *FOCI*, 2011.

[4] A. Houmansadr, G. T. Nguyen, M. Caesar, and N. Borisov, "Cirripede: Circumvention infrastructure using router redirection with plausible deniability," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, ser. CCS '11. New York, NY, USA: ACM, 2011, pp. 187–200.

[5] E. Wustrow, S. Wolchok, I. Goldberg, and J. A. Halderman, "Telex: Anticensorship in the network infrastructure." in *USENIX Security Symposium*, 2011.

[6] E. Wustrow, C. M. Swanson, and J. A. Halderman, "Tapdance: End-to-middle anticensorship without flow blocking," in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 159–174.

[7] D. Ellard, C. Jones, V. Manfredi, W. T. Strayer, B. Thapa, M. V. Welie, and A. Jackson, "Rebound: Decoy routing on asymmetric routes via error messages," in *Local Computer Networks (LCN), 2015 IEEE 40th Conference on*, Oct 2015, pp. 91–99.

[8] M. Schuchard, J. Geddes, C. Thompson, and N. Hopper, "Routing around decoys," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 85–96.

[9] A. Houmansadr, E. L. Wong, and V. Shmatikov, "No direction home: The true cost of routing around decoys." in *NDSS*, 2014.

[10] L. Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM Transactions on Networking (ToN)*, vol. 9, no. 6, pp. 733–745, 2001.

[11] J. Qiu and L. Gao, "As path inference by exploiting known as paths," in *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE*. IEEE, 2006, pp. 1–5.

[12] "Route views project," http://archive.routeviews.org/.

[13] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring isp topologies with rocketfuel," *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, Feb. 2004.

[14] "Alexa - actionable analytics for the web," http://www.alexa.com/topsites.

[15] H. Y. Technologies. (2016) Switch and router models and prices. [Online]. Available: www.router-switch.com

[16] C. S. Leberknight, M. Chiang, H. V. Poor, and F. Wong, "A taxonomy of internet censorship and anti-censorship," in *Fifth International Conference on Fun with Algorithms*, 2010.

[17] "Deep Packet Inspection," https://en.wikipedia.org/wiki/Deep_packet_inspection.

[18] H. M. Moghaddam, B. Li, M. Derakhshani, and I. Goldberg, "Skypemorph: Protocol obfuscation for Tor bridges," in *Proceedings of the 19th ACM conference on Computer and Communications Security (CCS 2012)*, October 2012.

[19] Z. Weinberg, J. Wang, V. Yegneswaran, L. Briesemeister, S. Cheung, F. Wang, and D. Boneh, "StegoTorus: A camouflage proxy for the Tor anonymity system," in *Proceedings of the 19th ACM conference on Computer and Communications Security (CCS 2012)*, October 2012.

[20] A. Houmansadr, C. Brubaker, and V. Shmatikov, "The parrot is dead: Observing unobservable network communications," in *Proceedings of the 2013 IEEE Symposium on Security and Privacy*, May 2013.

[21] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier, "Network topologies: inference, modeling, and generation," *IEEE Communications Surveys Tutorials*, vol. 10, no. 2, pp. 48–69, Second 2008.

[22] "Archipelago (ark) measurement infrastructure," http://www.caida.org/projects/ark/.

[23] S. Pandey, M.-J. Choi, S.-J. Lee, and J. W. Hong, "Ip network topology discovery using snmp," in *Proceedings of the 23rd International Conference on Information Networking*, ser. ICOIN'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 33–37.

[24] "Traceroute Looking Glass."

[25] "Midar," http://www.caida.org/tools/measurement/midar/.

[26] "Planetlab – an open platform for developing, deploying and accessing planetary-scale services," https://www.planet-lab.org/.

[27] "Ip to asn mapping," http://www.team-cymru.org/IP-ASN-mapping.html.

[28] B. Quoitin and S. Uhlig, "Modeling the routing of an autonomous system with c-bgp," *Netwrk. Mag. of Global Internetwkg.*, vol. 19, no. 6, pp. 12–19, Nov. 2005.
[29] "Freedom house - freedom of press," https://freedomhouse.org/.
[30] "Open net initiative," https://opennet.net/.
[31] "As relationships," http://www.caida.org/data/as-relationships/.
[32] "AS Rank: AS Ranking," http://as-rank.caida.org/.
[33] Anonymous, "The collateral damage of internet censorship by dns injection," *SIGCOMM Comput. Commun. Rev.*, vol. 42, no. 3, pp. 21–27, Jun. 2012.
[34] "Cidr report," http://www.cidr-report.org/as2.0/.