The hidden mailman and his mailbag: Routing path analysis from a European perspective

Josef Gustafsson, Rahul Hiran, Vengatanathan Krishnamoorthi and Niklas Carlsson

The Hidden Mailman and His Mailbag: Routing Path Analysis from a European Perspective

Josef Gustafsson, Rahul Hiran, Vengatanathan Krishnamoorthi, Niklas Carlsson
Linköping University, Sweden

***Abstract*—The postal system is often used as an analogy when describing Internet routing. However, in addition to similarities, there are some significant differences. First, and most importantly, the Autonomous Systems (ASes) that operate the routers along the end-to-end path of a packet can often inspect and manipulate the packet and its content. Second, due to lack of secure routing mechanisms, packet paths can be diverted through additional non-trusted ASes. Although we often know the first network we connect through and the service that we access, we seldom know the networks that forward our packets. We can think of these networks as *hidden mailmen*. To better understand these networks and their potential access to information, we characterize the ASes along the paths of typical Internet packets between European example clients and the most popular web domains. We also identify ASes and countries with higher path coverage and investigate if there are differences in the HTTPS usage among paths that may take additional detours. Our results highlight the role played by North American (typically US-based) ASes and glean insights into how vulnerable the detoured traffic is to man-in-the-middle attacks compared to regular traffic.**

I. Introduction

We are increasingly relying on the information and services delivered over the Internet. Many existing services are moving online and new online services are being introduced. As a consequence, the e-commerce industry (worth 1,500 trillion dollars in 2014 [1]) is setting new records every year. As the value of these services and the Internet traffic they generate increases, it becomes increasingly important to understand who forwards all this data and information, especially as forwarding entities often can read or modify the data in transit.

Today, a typical Internet packet traverses many routers, operated by different Autonomous Systems (ASes), each owned and operated by organizations registered in different countries. The inter-AS paths of these packets are primarily determined by the Boarder Gateway Protocol (BGP) and the routing rules and business agreements that each AS sets up with its neighboring ASes [2]. Due to BGP's distributed nature and many hidden and non-transparent business agreements, global Internet paths are often non-optimal and can quickly change.

Any AS along the end-to-end path is in a position to make copies or manipulate the packets that their routers forward. To protect against wiretapping and manipulation of network traffic, many services are starting to use HTTPS [3], [4]. With HTTPS, regular HTTP requests/responses are transferred over an end-to-end connection encrypted using Transport Layer Security (TLS). Although the use of HTTPS promises secure end-to-end transfers of sensitive information, HTTPS has known weaknesses (including common use of weak certificates and keys) making connections susceptible to man-in-the-middle (MITM) attacks [5]–[8].

Although BGP is highly susceptible to routing attacks and many routing attacks already have been reported [9], [10], we focus primarily on the ASes along typical Internet paths and do not attempt to detect actual route hijacks. Of course any compromised AS (or router belonging to an AS) along a "typical" path may take part in a third-party MITM attack.

Recent controversies involving government agencies such as the National Security Agency (NSA), for example, may also raise concerns for nation state based monitoring in routers operated by organizations based in foreign countries. For example, already in 2013, it was reported that some countries (e.g., Brazil and Germany) encourage local Internet traffic to be routed locally, rather through US-based ASes [11]. Whereas many countries have laws that protect against monitoring of the Internet traffic within a country's boundary, the exact boundaries for what and where such traffic may be inspected is much more complex [12]. Furthermore, ASes increasingly have many vantage points across the globe, often are not restricted to a single country or region, and may be conflicted by multiple different laws, policies, and other agreements.

In this paper, we study the ASes and origin countries of the ASes that operate the forwarding routers along example paths on the Internet. We focus our analysis on the paths taken between example locations in Europe and carefully selected sample domains from the top-1M most popular website domains according to `alexa.com`. Focusing on the top-1M is motivated by the high popularity skew of web domains, but also allows us to compare the paths to domains hosted in different regions of the world. When analyzing the routes taken, we observe a significant number of detours through non-European ASes. Such detours raise questions regarding the integrity of the Internet paths. One way to protect the transferred data is with the use of HTTPS. Motivated by the impact that both detours and HTTPS (and its implementation with certificates, hashes, and keys) can have on the security of the data transferred, we also characterize differences and similarities in the adoption of HTTPS, and the use of weak certificates and keys, for both direct paths and detours.

Although our study takes a European perspective on network traffic to the most popular domains, it is easy to see that foreign ASes, primarily US-based, play a major role in forwarding this traffic. Not only are the North American ASes responsible for the majority of the end-host networks (50.0-69.4%, depending on popularity segment) hosting these

This is the authors' version of a paper accepted to *Proc. IEEE International Conference on Communications (IEEE ICC)*, Paris, France, May. 2017.

domains, they are also often at some point relaying the traffic originating in Europe destined to end-host networks in Europe and Asia. These detours typically have significantly longer hop-counts and round-trip times than paths not going through external ASes, and sometimes even physically leave the continent before later returning.

We have not observed (or tried to obtain) any evidence suggesting nation state based monitoring or ASes trying to adapt their forwarding rules to attract weaker encrypted traffic. However, we have observed that clients' non-encrypted HTTP connections to North American domains typically see a much smaller fraction of detours than the corresponding HTTPS paths. These differences may suggest that greater care is taken to protect the routes of these North American based domains. Interestingly, no such biases are observed when comparing the paths destined to domains hosted in networks associated with other regions. In general, North American domains also typically appear to have higher HTTPS adoption and are more up-to-date with the current recommendations, such as the use of strong certificates and keys, for example.

The remainder of this paper is organized as follows. Section II presents our measurement methodology. The following two sections present our results. Section III characterizes the ASes along sample paths and Section IV presents an HTTPS-based analysis of the differences and similarities observed between different path types. Finally, Section V presents related work, before Section VI concludes the paper.

II. Data Collection Methodology

For our data collection, we performed targeted traceroute experiments between 74 geographically dispersed PlanetLab [13] nodes in Europe and 10,000 logarithmically sampled domains from the top-1M globally most popular website domains according to `alexa.com` in June 2015. The choice of logarithmic sampling is motivated by a heavy popularity skew towards the most popular domains [14]. The sampling was performed by placing points x_i uniformly on the interval $0 \leq x_i \leq 6$ and then selecting all unique domain ranks $\lfloor 10^{x_i} \rfloor$, with the point density selected such that 10,000 unique sample domains are obtained [15].

The path between each PlanetLab node and sample domain was traced using Scamper [16], a flexible traceroute tool from CAIDA. For each traceroute, we collected the router names, IP addresses, and the round trip times (RTT) between the PlanetLab node, each router, and the destination. We then used the Cymru (whois.cymru.com) whois database (which includes IP-to-AS mappings) and basic lookups to obtain the AS associated with each (incoming) router interface along the path and the country in which each AS was registered.

Finally, we performed an HTTPS-based analysis of each domain. Here, we first tried to connect to the identified IP address (of each domain) on port 443 using `openSSL`. If no certificate was returned as part of the TLS/SSL handshake, we listed the site as not supporting HTTPS, and if it did return a certificate we listed it as supporting HTTPS. To extract the offered and selected (after negotiation) cipher suites, we used `nmap` scans. Since most PlanetLab nodes are not up-to-date and do not provide full access to the network interface, the cipher suite experiments were performed from a local, but up-to-date, Kali Linux machine. For the analysis, we assume that the agreed upon cipher suites would be similar for the other hosts (with the same configurations) regardless of location, and instead focus on differences in types of certificates, keys, and hashes that takes detours to their final destination, if any.

A. Limitations

Our measurements are performed from 74 European PlanetLab nodes. As PlanetLab nodes typically are hosted in academic/research institutions, the use of PlanetLab nodes introduces a bias, and the full paths may therefore not be representative of regular Internet users. To discuss the potential effect of these edge-based biases we have analyzed results both for when we include and when we exclude the first observed AS along the end-to-end path. The choice to use the top-1M global list, rather than top lists for the individual countries associated with the individual PlanetLab nodes, reduces the impact of which PlaneLab nodes are selected and ensures a more fair head-to-head comparison across client locations.

When discussing our results, it is also important to note that we focus on the country in which the ASes along the paths are registered, rather than the location (or country) of the routers themselves. This has the advantage that we do not have to rely on unreliable geo-location techniques to map IP addresses to locations. Since ASes typically are in control of the routers within an AS, it also provides some insight into the information that foreign networks could potentially gain, regardless of the location of the forwarding routers. While the accuracy of the absolute values presented still depends on the accuracy of the IP-to-AS mappings provided by the Cymru database and public IP-to-AS mappings have their shortcomings [17], [18], we note that similar mappings have successfully been used to answer many AS topology related questions [19], [20], identify complex AS relationships [21], and even interdomain routing questions [22].

Finally, we note that we primarily are interested in whether a path goes through an AS or not. We do not consider ownership of individual routers or the exact location of AS-to-AS links. For such analysis more advanced techniques are needed which leverage multiple paths through the same router or other information to identify AS boundaries [23], [24].

III. Path-based Analysis

A. Server Location

Figure 1 shows the geographic region of the networks that hosted the top-1M web domains, when visited from European locations. Results are broken down based on the popularity of the different web domains and end-host networks are mapped to one of the following regions/categories: North America (NA), Europe (EU), Asia, and Other. In Asia we include Oceania (a few instances of Australia and New Zealand). The category Other is dominated by South American networks, with a few instances of networks registered in Africa.

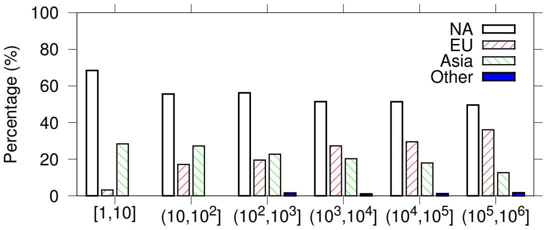


Fig. 1. Network ownership of networks hosting the services for top-1M web domains when visited from European locations.

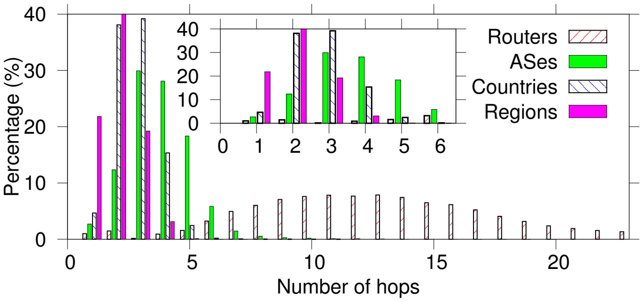


Fig. 2. Percent (%) of paths from European (EU) clients that are at different distances from the server.

Despite significant replication across geographically distributed data centers, the majority of end-host networks are owned by NA-based organizations, with the fraction of EU originated paths ending in NA-hosted ASes decreasing (monotonically) from 69.4% in the top-10 domains to 50.0% for the domains in the popularity range 100K-to-1M. Also, the percent of paths ending in Asian ASes decreases monotonically from 28.4% to 12.6%, as the popularity of the domains decreases, whereas the percent EU-terminated paths increases from 3.2% to 36.1%. The Other category is responsible for much fewer domains (1.1-1.7%); all outside the top-100.

With servers located across the globe, path lengths will differ significantly from service to service. Figure 2 shows the distribution of path lengths for European clients that are accessing the sampled websites. Here, we include results for (i) the number of routers on the end-to-end paths, (ii) the number of traversed ASes, (iii) the number of countries associated with the traversed ASes, and (iv) the number of regions (or continents) associated with these ASes. For countries and regions we consider the sets of all ASes associated with a country/region, and count a country/region set twice if the path leaves the country/region set to another, and then re-enters it again. In general, the distributions are relatively symmetric, resulting in relatively similar mean (12.9, 3.8, 2.7, 2), median (12, 4, 3, 2), and mode (13, 3, 3, 2) values for all four measures. Also the path length differences (results omitted) for domains of different popularity are small.

B. Routes Taken

Although the Internet topology is known to have flattened over the last decade [20], there is still a large skew in the ASes responsible for forwarding the majority of the Internet traffic. Figure 3 shows the percent of paths that each individual AS observed. Here, ASes are ordered from the AS that sees the most paths (rank 1) to the AS in our dataset that sees the least paths (rank 2378). We also include a CDF of the total number of observations of the ASes as an aggregate. Although a few ASes have high coverage, we note that there is a sharp drop in the number of paths observed per AS. For example, whereas the top-3 ASes (Level 3, NORDUNET, COGENT) observe 22.1%, 18.1% and 13.4% of the EU-originating paths, respectively, only 55 ASes observe more than 1% of the paths and 250 ASes more than 0.1% of the paths.

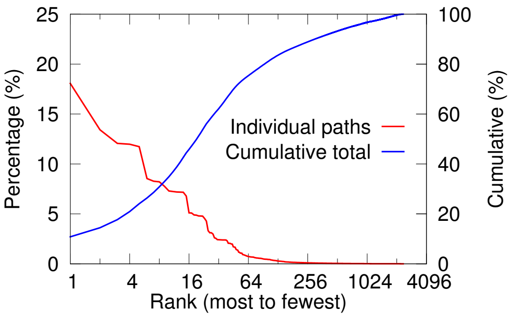


Fig. 3. Percent (%) of paths individual ASes sees and the cumulative number of observations, respectively, for European (EU) client paths.

Table I shows the top-5 most frequently forwarding ASes, conditioned on the region of the destination network. To better understand the impact of the university/research network bias in PlanetLab measurement locations, we show results both when including and when excluding the first AS along the path. As the results are similar, in the following, we discuss the results with all ASes included. We note that NORDUNET is the most observed AS in the EU destined traffic, whereas Level 3 (US-based AS) and CHINANET (China Telecom) observe most paths when the traffic is destined for NA and Asia, respectively. Interestingly, US-based COGENT is on the top-5 list for both EU-to-EU paths and EU-to-Asia paths (with rank 4 and 3, respectively). Also Level 3 ranks high, with a top-7 spot for the EU-to-EU paths and top-4 spot for the EU-to-Asia paths. Although these observations in many cases can be explained by these ASes having global peering locations and links all over the globe, the results highlight the high coverage some of these ASes may gain, including of paths that would not need to pass through external regions, and hence also not through ASes operated by external organizations.

While we emphasize that we do not make any claims of any form of wrong doings, it is important to note that ASes with access to rich information may result in significant information leakage if compromised and/or equipped to extract network intelligence. The high coverage ASes can hence be considered high risk, if compromised.

C. AS Origins

We next consider the origin countries of the ASes along the paths. Figure 4 shows a breakdown of the country-based ownership of ASes observed along the paths when European (EU) clients access all domains (bars) or EU-hosted domains (markers). For clarity, we use different colors on the bars, to distinguish between countries (of the forwarding ASes)

TABLE I
SUMMARY OF THE TOP-5 MOST FREQUENTLY FORWARDING ASES WHEN EU CLIENTS ACCESS WEBSITES IN EU, NA, AND ASIA, RESPECTIVELY. RESULTS SHOWN BOTH WHEN INCLUDING AND WHEN EXCLUDING THE FIRST AS ALONG THE PATH. DOUBLE ASTERISK (**) SHOWS EXTERNAL ASES.

		EU to EU			**EU to NA**			**EU to Asia**		
	Rank	AS	AS#	Perc.	AS	AS#	Perc.	AS	AS#	Perc.
Including	1	NORDUNET	2603	17.8%	Level 3	3356	31.7%	CHINANET	4134	28.0%
	2	FR-RENATER	2200	12.1%	NORDUNET	2603	17.0%	NORDUNET	2603	21.6%
	3	RedIRIS	766	11.0%	CLOUDFLARENET	13335	15.2%	** COGENT	174	21.2%
	4	** COGENT	174	10.9%	ASGARR	137	13.1%	** Level 3	3356	17.3%
	5	ASGARR	137	10.0%	RedIRIS	766	12.2%	FR-RENATER	2200	13.2%
Excluding	1	NORDUNET	2603	17.5%	Level 3	3356	31.7%	CHINANET	4134	28.0%
	2	FR-RENATER	2200	12.0%	NORDUNET	2603	16.7%	** COGENT	174	21.2%
	3	** COGENT	174	10.9%	CLOUDFLARENET	13335	15.2%	NORDUNET	2603	20.8%
	4	GEANT	20965	9.7%	COGENT	174	11.8%	** Level 3	3356	17.3%
	5	** Level 3	3356	8.7%	FR-RENATER	2200	11.6%	CHINANET	23724	13.1%

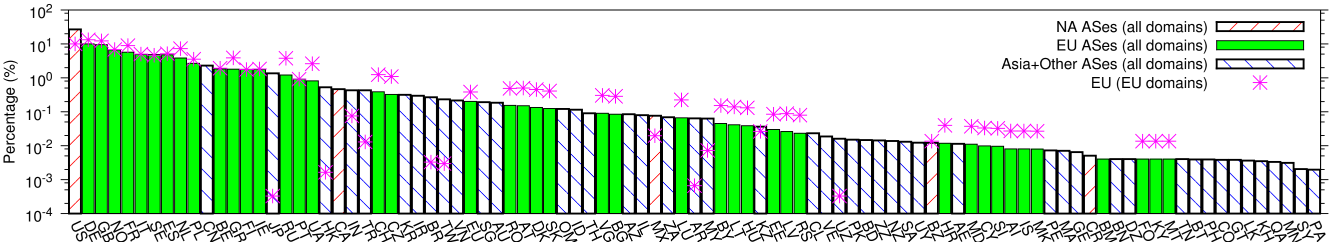


Fig. 4. Country-based breakdown of the ownership of the networks observed along the paths between clients and servers. European (EU) clients accessing Global (total) and EU websites from the top-1M.

belonging to different regions (with Asia and Other grouped together). Interestingly, US-based networks' exposure significantly dominates all other countries, regardless of the clients' regions. Only when considering EU-to-EU paths (markers) does two European countries (Germany (DE) and Great Britain (GB)) have higher exposure than the US-based networks.

We have also evaluated the fraction of paths the different countries would observe if all ASes registered in those countries would cooperate. Table II summarizes the top-5 countries with the most coverage of the paths starting in EU and the fraction of paths that the ASes registered in these countries cover. We include results both across all paths, as well as broken down per destination region. In total, US-based ASes (71.3%) covers almost three times as many paths as ASes from the top-European countries (Germany (26.6%) and Great Britain (25.6%)). In fact, even for EU-to-EU paths US-based ASes have a 26.9% coverage, only behind German (34.3%) and British (31.8%) ASes. For the EU-to-Asia paths, US-based ASes covers the majority (66.7%) of the paths. These results show that many paths take detours through US-based ASes although it is not justified from a geographical point of view. The following sections investigate whether this is a concern.

D. Intra-continental Traffic and their Detours

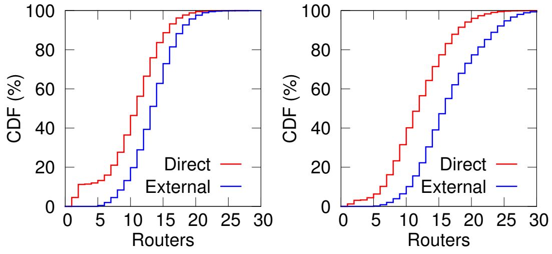
Let us look closer at the case in which EU-based clients access domains with end-servers in networks registered with EU-based countries. For this case we have already seen that US-based ASes have the third highest coverage (26.9%), but we also see the presence (Figure 4) of ASes from Japan (JP), Hong Kong (HK), India (IN), Turkey (TR), Brazil (BR), Taiwan (TW), Argentina (AR), Malaysia (MY), Kazakhstan (KZ), Uzbekistan (UZ), and Belize (BZ) among these paths.

To put these paths in perspective, we next take a closer look at the path lengths when going through at least one AS that does not belong to the source or destination region compared to when the path does not go through such external ASes. Figure 5 shows a CDF of the number of routers for both EU-to-EU paths and across all paths starting in EU. In the EU-to-EU case, we say that a path is "direct" if none of the ASes along the path is registered in countries outside Europe, and "external" otherwise. For the case of a different destination region, also ASes registered in the destination region are allowed for the paths to be considered "direct". Figure 6 shows the corresponding CDFs for the RTT differences.

TABLE II
SUMMARY OF THE TOP-5 MOST FREQUENTLY FORWARDING ASES WHEN EU CLIENTS ACCESS WEBSITES IN EU, NA, AND ASIA, RESPECTIVELY.

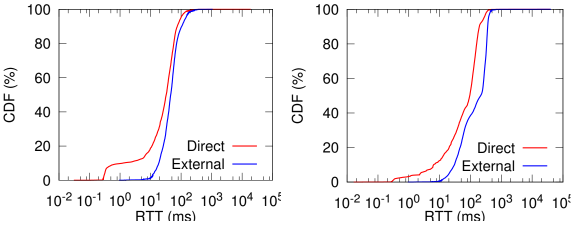
Rank	**EU to ∗**		**EU to EU**		**EU to NA**		**EU to Asia**	
1	US	71.3%	DE	34.3%	US	99.0%	US	66.7%
2	DE	26.6%	GB	31.8%	DE	22.1%	CN	35.8%
3	GB	25.6%	US	26.9%	GB	21.3%	GB	27.5%
4	NO	18.3%	FR	24.0%	NO	17.0%	DE	26.3%
5	FR	15.6%	NL	18.6%	IT	13.2%	NO	21.6%

As expected, the "external" paths add substantially to both the number of routers (Figure 5) and the RTTs (Figure 6). Looking closer at the "external" paths, we have even observed paths where the data path physically appear to leaves the geographic region (e.g., based on router names and RTTs) before returning. Although it is well known that BGP does not provide optimal paths, such non-optimal paths raise additional concerns, since both the ASes and the physical routers (which now are located outside the region) now potentially are governed by different laws. As noted earlier, different countries and regions may have different policies about what and where different traffic is allowed to be inspected, and what other actions are permissible on the data. Mutual legal assistance treaties and other surveillance agreements may further complicate this situation [25], [26].



(a) EU to EU (b) EU to ∗

Fig. 5. CDF of the number of routers along the paths.



(a) EU to EU breakdown (b) EU to ∗ breakdown

Fig. 6. CDF of the RTTs for external and direct paths.

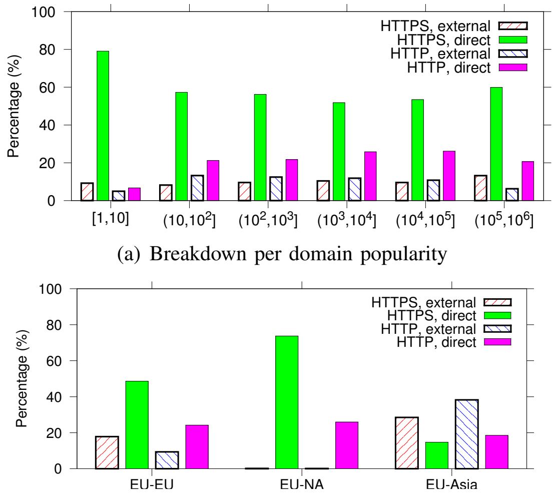
IV. HTTPS-BASED ANALYSIS

One way to protect the information as it flows along the end-to-end forwarding path, regardless of the route, is to use encryption. For example, with HTTPS, regular HTTP requests/responses are transferred over an end-to-end connection encrypted using Transport Layer Security (TLS) or its predecessor Secure Sockets Layer (SSL). To investigate if the detours is a concern, we first compare and contrast the paths taken by connections that allow the use of HTTPS and the paths taken to the HTTP (only) domains. We also looked closer at the paths of HTTPS enabled domains and potential differences in the paths taken when using weak ciphers, including short keys and weak algorithms, for example.

A. *HTTP (only) Domains*

We first classified all paths based on (i) if the end-domain allows use of HTTPS, and (ii) if the path includes any external regions beyond the source and destination region. If the domain allows HTTPS and the path only involve ASes registered in countries belonging to the region(s) where the source and destination ASes are located, then we classify it as a “direct HTTPS” path. On the other hand, if the domain did not allow the use of HTTPS we instead classify that (same) path as “direct HTTP”. The same classification system is applied for the “external” paths. Figure 7(a) shows a breakdown into the four resulting path types, based on the domain popularity.

Interestingly, the HTTP (only) domains include a much larger relative ratio of external detours (blue stripes) compared to direct paths (solid purple) than the corresponding external/direct ratio when doing the same comparison for HTTPS domains (i.e., solid green vs red stripes in the figure). These differences are most apparent for the top-10K domains. In all these cases, the ratio of the first two (blue stripes divided by solid purple) is larger than 0.46, whereas the second ratio (red stripes divided by solid green) is much smaller (0.12-0.22).



(a) Breakdown per domain popularity

(b) Regional breakdowns

Fig. 7. HTTPS availability for services based on the fraction of paths that are “direct” and those that go through “external” regions.

This can partially be explained by most EU-to-NA paths being direct and more frequently allowing HTTPS. The most popular services have also been found to be more likely to use HTTPS and third-party trackers [27], suggesting that they may be more aware of the information value in the traffic.

As the payload of (non-encrypted) HTTP traffic is much simpler to inspect, the larger observed fraction of detours of HTTP paths is a concern for both clients and servers. To better understand potential geographic differences, Figure 7(b) shows breakdowns for the different destination regions. Interestingly, the NA-destined paths almost always are direct in our dataset, regardless if considering the HTTPS or HTTP paths. This may suggest that the policies of the NA-based ASes provide higher preference for local routes.

We also note that the Asia-based domains (EU-to-Asia) use much smaller fraction HTTPS (e.g., compare red stripes + solid green with blue stripes + solid purple) than NA-based and EU-based domains. The highest HTTPS usage is seen among the NA-based domains.

B. *Weak Certificates*

With HTTPS, a TLS handshake is used to establish a new TLS/SSL connection. During this handshake, the client and server agree which cryptographic algorithms (i.e., cipher suite) to use and determine the session key. At a high level, the server first presents an X.509 digital certificate to the client, which checks that the identity matches the target servers, that the certificate has not expired, and that the digital signature is valid, before a session key finally is determined and an end-to-end encrypted communications channel is opened.

Weak cryptography in the certificates can significantly impact the security of the connection and can enable potential MITM attacks. It is therefore concerning that many sessions still use weak encryption in the certificates. For example, of the observed certificates 35.9% used SHA1 and 0.41% used MD5, rather than SHA256 or better, which currently is recommended. Although this is a significant improvement

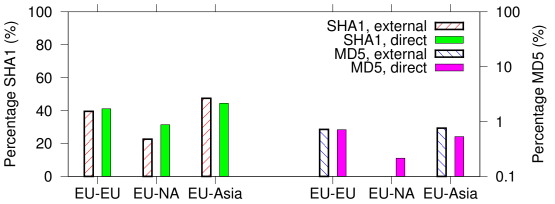
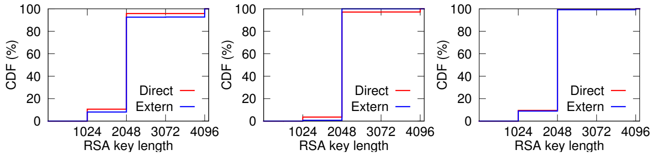


Fig. 8. Usage of weak certificates.



(a) EU-EU (b) EU-NA (c) EU-Asia

Fig. 9. RSA key length distributions.

compared to the 98.7% share of SHA1 and 0.54% share MD5 that Durumeric et al. [28] observed in 2013, there is still a long way to go, especially given that attacks against MD5 were demonstrated already in 2008 [7].

Figure 8 shows the usage of SHA1 and MD5 for domains with direct and external paths, broken down per source-destination region. Here, the SHA1 usage is shown on linear scale (left y-axis) using the red-striped and green bars, and the (relatively smaller) MD5 usage is shown on logarithmic scale (right y-axis) using blue-striped and purple bars.

The use of both SHA1 and MD5 is significantly smaller among the NA-based domains (EU-to-NA). This suggests that many of these sites (with a larger relative portion in the top-10K, for example) may be more up-to-date and better follow current security recommendations. We do not observe any significant differences in the usage of SHA1 between the direct and external paths. For MD5, on the other hand, the EU-to-NA paths again sticks out, as the domains with external paths never use MD5, whereas there is a non-negligible number of domains with direct paths for which we observe usage of MD5. The smaller fraction of external paths associated with weaker algorithms indicates that these paths either endures higher protection or that they are not intentionally diverted more than other domains. For the other regions we observe no significant differences between direct and external paths. While it is encouraging that we do not observe any apparent path biases towards weaker certificates taking external (longer) paths, the overall statistics highlights that many domains (especially non-NA-based) still use weak certificates.

C. Weak Keys

Almost all observed keys used RSA encryption. However, despite the National Institute of Standards and Technology (NIST) recommending to stop using 1,024-bit RSA keys already in 2012 [29], we have observed a non-negligible number of certificates using 1,024-bit RSA keys. Figure 9 shows the key lengths for all RSA keys, broken down per direct/external paths, for each destination region. Again, the NA-based domains (EU-to-NA) appear more up-to-date, and use a significantly smaller fraction weak (1,024 bit) keys, especially for sites with external paths. However, in none of the cases do we observe any significant differences in the usage of weak keys between when paths are direct and external, respectively. Perhaps the most significant differences are in the usage of very strong (4,096 bit) keys. Most noticeable is the somewhat higher usage of very strong (4,096 bit) keys among domains with external EU-to-EU paths. While the use of very strong keys is an interesting observation and the use of strong keys potentially can help compensate for the lack of control of the Internet paths, there is no evidence that these domains would make a different choice if the paths were direct.

V. Related Work

HTTPS and routing has typically been studied separately. Various traceroute- and BGP-based tools and methods have been used to understand Internet routes [22], [30]. While there are many legitimate reasons why announced AS paths (observed from BGP data) may differ from the actual data paths (observed using traceroute) [30], differences can also be used to understand interception attacks [31]. In this paper, we focus on the data paths observed trough traceroutes and use public IP-to-AS mappings to estimate the AS paths of example traceroutes. Of course these types of databases are not perfect, so it is perhaps not surprising that there is rich literature that studies the inaccuracy of different IP-to-AS mapping approaches [18], how to best infer complex AS relationships [21], or how to build models of the AS topology [17] from traceroute measurements. Recently, Anwar et al. [22] used IP-to-AS mappings to capture the interdomain routing policies used in practice.

The increasing adoption of HTTPS is well established [3], [4]. For example, Holz et al. [4] characterize the X.509 public key infrastructure using HTTPS scans of popular HTTPS servers from nine different locations, a third-party scan of the entire IPv4 space, and by monitoring the 10 Gbps uplink of a large research network. Similar to our results, they find that the top-domains have come further in their deployment of HTTPS and is more up-to-date. Others have examined attacks targeting particular aspects of the connection establishment of HTTPS connections, including the key exchange [5], [6], specific ciphers [8] and the MACs [7]. These studies demonstrate that users are sensitive to MITM attacks. To make things worse, common use of TLS warnings reduces users' attention to actual MITM attack warnings [32].

Wählisch et al. [33] use domain-to-IP mappings and IP-to-AS mappings of the top-1M domains according to `alexa.com` to show that popular CDNs typically are hesitant to deploy RPKI. Liang et al. [34] propose an interesting solution (based on DANE [35]), which helps alleviate the delegation problem in the CDN setting. In a recent poster, Edmundson et al. [36] use geo-mapping of IP addresses along traceroutes to study transnational detours from five different countries (Brazil, Netherlands, Kenya, India and US). While their focus is on the physical location of the routers, we focus on the forwarding networks. Furthermore, in contrast to the above works, we characterize the Internet paths taken from different European vantage points, while also taking into account the use of HTTPS, weak keys, and weak certificates.

Although this paper does not explicitly consider routing attacks, we note that there is no universally deployed solution that prevents routing hijacks, and there has been an increasing number of observed routing attack occurrences [10], [12], [31], [37]. Part of the lack of globally deployed solution is due to the difficulty incentivizing operators to invest in existing solutions or share information [9], [10]. For example, crypto-based efforts [38], [39] often come with high deployment costs and monitoring-based efforts [40] are typically limited by the amount of data that AS operators are willing to share with other ASes. Others have demonstrated the risks with authoritative structures, as used with RPKI, for example [41].

VI. Conclusions

In this paper we have characterized the ASes and the origin countries of the organizations that operate the ASes along the paths taken by typical Internet packets between European clients and some of the most popular domains. It is noteworthy that NA-based ASes play a major role in both hosting and forwarding much of the traffic, including often forwarding traffic originating in Europe destined to end-host networks in Europe and Asia. In general, we have observed a significant number of detours through non-European ASes, whose presence (even though the traffic often may not physically leave Europe) raises questions regarding the integrity of the Internet paths, as these ASes may be conflicted by multiple (potentially conflicting) laws, policies, and other agreements. We have also compared differences and similarities in the adoption of HTTPS, the use of weak certificates, and the use of weak keys, for direct paths and external detours. While we have not observed any results suggesting that ASes would try to attract weakly encrypted traffic, we have observed that the paths to non-encrypted HTTP domains hosted in NA typically see a much smaller fraction of detours than the corresponding HTTPS paths. This may suggest that greater care is taken to protect the routes of these NA-based domains. The NA-based domains also appear to have higher HTTPS adoption rate, and when using HTTPS use stronger certificates and keys than EU- and Asia-based domains. The use of stronger algorithms and larger keys may be a sign that companies in these regions try to thwart US government snooping of data. However, there is no way to confirm whether ASes (along the path) cooperate with the government.

References

[1] eMarketer, "Global B2C ecommerce sales to hit 1.5 trillion this year driven by growth in emerging markets," 2014.

[2] P. Gill, M. Schapira, and S. Goldberg, "A survey of interdomain routing policies," *SIGCOMM CCR*, vol. 44, no. 1, pp. 28–34, Dec. 2013.

[3] D. Naylor *et al.*, "The cost of the "S" in HTTPS," in *Proc. ACM CoNEXT*, 2014.

[4] R. Holz *et al.*, "The SSL landscape: A thorough analysis of the X.509 PKI using active and passive measurements," in *Proc. IMC*, 2011.

[5] B. Beurdouche et al., "A messy state of the union: Taming the composite state machines of TLS," in *Proc. IEEE S&P*, 2015.

[6] D. Adrian et al., "Imperfect forward secrecy: How Diffie-Hellman fails in practice," in *Proc. ACM CCS*, 2015.

[7] A. Sotirov *et al.*, "MD5 considered harmful today," http://www.win.tue.nl/hashclash/rogue-ca/, 2008.

[8] N. AlFardan, D. Bernstein, K. Paterson, B. Poettering, and J. Schuldt, "On the security of RC4 in TLS," in *Proc. USENIX Security*, 2013.

[9] K. Butler *et al.*, "A survey of BGP security issues and solutions," *Proc. IEEE*, vol. 98, no. 1, pp. 100–122, Jan. 2010.

[10] S. Goldberg, "Why is it taking so long to secure Internet routing?" *ACM Queue*, vol. 12, no. 8, pp. 327–338, Oct. 2014.

[11] M. Taylor, N. Hopkins, and J. Kiss, "NSA surveillance may cause breakup of internet, warn experts," *The Guardian*, Nov. 2013.

[12] A. Arnbak and S. Goldberg, "Loopholes for circumventing the constitution: Unrestrained bulk surveillance on americans by collecting network traffic abroad," in *Proc. HotPETs*, Jul. 2014.

[13] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman, "PlanetLab: An overlay testbed for broad-coverage services," *ACM CCR*, vol. 33, no. 3, pp. 3–12, Jul. 2003.

[14] P. Gill *et al.*, "Characterizing organizational use of web-based services: Methodology, challenges, observations, and insights," *ACM TWEB*, vol. 5, no. 4, pp. 19:1–19:23, Oct. 2011.

[15] A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, "Third-party identity management usage on the Web," in *Proc. PAM*, 2014.

[16] M. Luckie, "Scamper: A scalable and extensible packet prober for active measurement of the Internet," in *Proc. IMC*, 2010.

[17] H. Chang, S. Jamin, and W. Willinger, "Inferring AS-level internet topology from router-level path traces," in *Proc. ITCom*, 2001.

[18] Y. Zhang, R. Oliveira, H. Zhang, and L. Zhang, "Quantifying the pitfalls of traceroute in AS connectivity inference," in *Proc. PAM*, 2010.

[19] K. Chen *et al.*, "Where the sidewalk ends: Extending the internet AS graph using traceroutes from P2P users," in *Proc. ACM CoNEXT*, 2009.

[20] P. Gill *et al.*, "The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse?" in *Proc. PAM*, 2008.

[21] V. Giotsas, M. Luckie, B. Huffier, and K. Claffy, "Inferring complex AS relationships," in *Proc. ACM IMC*, 2014.

[22] R. Anwar *et al.*, "Investigating interdomain routing policies in the wild," in *Proc. ACM IMC*, 2015.

[23] A. Marder and J. M. Smith, "MAP-IT: Multipass accurate passive inferences from traceroute," in *Proc. IMC*, 2016.

[24] M. Luckie *et al.*, "Bdrmap: Inference of borders between IP networks," in *Proc. IMC*, 2016.

[25] S. Cortes, "MLAT Jiu-Jitsu and Tor: Mutual legal assistance treaties in surveillance," *Journal of Law and Technology*, Dec. 2015.

[26] A. Jaggard *et al.*, "20,000 in league under the sea: Anonymous communication, trust, MLATs, and undersea cables," in *Proc. PETS*, 2015.

[27] J. Purra and N. Carlsson, "Third-party tracking on the web: A Swedish perspective," in *Proc. IEEE LCN*, Nov. 2016.

[28] Z. Durumeric, J. Kasten, M. Bailey, and J. A. Halderman, "Analysis of the HTTPS certificate ecosystem," in *Proc. IMC*, 2013.

[29] E. Barker *et al.*, "Recommendation for key management, part 1: General (revision 3)," *NIST Special Publication 800-57*, Jul. 2012.

[30] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an accurate AS-level traceroute tool," in *Proc. ACM SIGCOMM*, 2003.

[31] R. Hiran *et al.*, "Characterizing large-scale routing anomalies: A case study of the China telecom incident," in *Proc. PAM*, Mar. 2013.

[32] D. Akhawe *et al.*, "Here's my cert, so trust me, maybe?: Understanding TLS errors on the web," in *Proc. WWW*, 2013.

[33] M. Wählisch *et al.*, "RiPKI: The tragic story of RPKI deployment in the web ecosystem," in *Proc. ACM HotNets*, 2015.

[34] J. Liang *et al.*, "When HTTPS meets CDN: A case of authentication in delegated service," in *Proc. IEEE S&P*, 2014.

[35] P. Hoffman and J. Schlyter, "The DNS-based authentication of named entities (DANE) transport layer security (TLS) protocol: TLSA," RFC 6698 (Proposed Standard), Aug. 2012.

[36] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford, "A first look into transnational routing detours," in *Proc.ACM SIGCOMM(poster)*, 2016.

[37] Dyn Research, "Pakistan hijacks YouTube," 2008. [Online]. Available: http://research.dyn.com/2008/02/pakistan-hijacks-youtube-1/

[38] M. Lepinski and S. Kent, "An Infrastructure to Support Secure Internet Routing," RFC 6480 (Informational), IETF, Feb. 2012.

[39] S. Kent, C. Lynn, and K. Seo, "Secure border gateway protocol (S-BGP)," *IEEE JSAC*, vol. 18, no. 4, pp. 582–592, Apr. 2000.

[40] M. Lad, , D. Pei, Y. Wu, B. Zhang, and L. Zhang, "PHAS: A prefix hijack alert system," in *Proc. USENIX Security*, Jul/Aug. 2006.

[41] D. Cooper, E. Heilman, K. Brogle, L. Reyzin, and S. Goldberg, "On the risk of misbehaving RPKI authorities," in *Proc. ACM HotNets*, 2013.