Impact of Carrier-Grade NAT on web browsing

Impact of Carrier-Grade NAT on Web Browsing

Enrico Bocchi[1], Ali Safari Khatouni[1], Stefano Traverso[1],
Alessandro Finamore[1,2], Valeria Di Gennaro[1,3], Marco Mellia[1], Maurizio Munafò[1], Dario Rossi[3]

[1]Politecnico di Torino - first.last@polito.it
[2]Telefonica Research - first.last@telefonica.com
[3]Telecom ParisTech - first.last@enst.fr

***Abstract*—Public IPv4 addresses are a scarce resource. While IPv6 adoption is lagging, Network Address Translation (NAT) technologies have been deployed over the last years to alleviate IPv4 exiguity and their high rental cost. In particular, Carrier-Grade NAT (CGN) is a well known solution to mask a whole ISP network behind a limited amount of public IP addresses, significantly reducing expenses.**

Despite its economical benefits, CGN can introduce connectivity issues which have sprouted a considerable effort in research, development and standardization. However, to the best of our knowledge, little effort has been dedicated to investigate the impact that CGN deployment may have on users' traffic. This paper fills the gap. We leverage passive measurements from an ISP network deploying CGN and, by means of the Jensen-Shannon divergence, we contrast several performance metrics considering customers being offered public or private addresses. In particular, we gauge the impact of CGN presence on users' web browsing experience.

Our results testify that CGN is a mature and stable technology as, if properly deployed, it does not harm users' web browsing experience. Indeed, while our analysis lets emerge expected stochastic differences of certain indexes (e.g., the difference in the path hop count), the measurements related to the quality of users' browsing are otherwise unperturbed. Interestingly, we also observe that CGN protects customers from unsolicited, often malicious, traffic.

Keywords*—*IP networks; Computer network management; Network address translation; Web services; Performance

I. Introduction and Motivation

The long anticipated exhaustion of public IPv4 addresses is finally here. Given their scarcity, IPv4 addresses have been the subject of an ongrowing market. In 2014, such market increased by 220% and 600% according to recent estimates of APNIC and RIPE, respectively, while the cost of a single IP address reached 10$/year.[1] This clearly translates into non negligible economical investments for Internet Service Providers (ISPs) and organizations running large IP networks. The most natural solution would be adopting IPv6, but, unfortunately, its deployment is still lagging.[2]

Network Address Translation (NAT) techniques have become a viable, reasonably cheap, solution to alleviate public IPv4 exhaustion. In a nutshell, a router implementing NAT functionality remaps the IP address space of a network into one (or more) public IP address by modifying the network address information in the packet header. For instance, in the typical access network scenario, the home router runs a NAT to "mask" users' private network behind a single public IP address. Hence, multiple user devices can connect to the Internet at the cost of a single public IP for the ISP.

Carrier-Grade NAT (CGN) technologies further extend this concept allowing to mask a whole ISP network using NAT [11]. In this scenario, the home routers are assigned private IP addresses to communicate within the ISP network, while CGN deployed at the peering routers interconnect the ISP network with the Internet. This approach allows the ISP to transform its network in a large private network, significantly reducing the total amount of public IP addresses to use, and thus the costs for maintaining them.

Despite the economical benefits, NAT techniques are more or less successful depending on their configurations [6], [8] and compliance to standards [4], [5], [10], [12]. Moreover, devices connected to the network through NAT cannot be accessed from the Internet unless specific actions are taken, and can be subject to delays due to NAT mapping operations. Hence, some questions naturally emerge: (i) Can CGN significantly impact users' experience when accessing Internet services? (ii) What are the benefits for residential users in having a public IP address at their home routers to access the Internet? (iii) Apart cost savings, what are the incentives for the ISP in assigning users' home routers private IP addresses?

While a significant effort has been put in standardizing CGN [11], [12], to the best of our knowledge little work has been dedicated to study the impact on the traffic exchanged with the Internet [4]. To fill this gap, we take advantage of a dataset of traffic traces collected in the operational network of a real ISP providing its customers with ADSL home routers, which are assigned either public or private IP addresses depending on the kind of subscription. In this scenario, all home routers with a private IP access the Internet through CGN technology. We split the monitored customers in two groups based on the type of IP address (private or public) of their home routers, and we assess the impact of CGN on the traffic they generate when browsing the web. Specifically, we use several metrics to evaluate the impact of CGN on the quality of the users' browsing experience, and rely on the Jensen-Shannon divergence to pinpoint differences between the two populations of customers.

Our findings show that no sharp differences can be observed between the two populations, testifying that CGN is a mature and reliable technology. Secondly, only 2% of users with public IP addresses run services which actually need to be reached from outside the ISP. Third, we observe a positive side-effect of accessing the Internet through CGN: home routers are more protected against unsolicited connection attempts (e.g., netscans, portscans, etc.), and malicious activities (e.g., DDoS, intrusion attempts, etc.).

[1]http://www.ipaddressnews.com/2014/04/07/343
[2]https://labs.ripe.net/Members/gih/counting-ipv6-in-the-dns

II. Related work

In the last years, CGN has been deployed in several ISP networks to limit the utilization of the IP address space [2]. Given its strategic importance, the research community and standardization authorities have made a great effort in understanding the impact of these technologies on the QoS and end-users experience. For instance, IETF RFCs standardize NAT requirements, implementations and behaviors [4], [5], [10], [12]. In particular [4] describes a case study conducted in a controlled testbed where multiple CGN configurations are tested to identify possible impact on DSL residential customers. Unfortunately, results are only qualitative and lack of generalization due to the artificial scenario.

In [9] authors collect aggregate traffic traces from a real ISP network to study ports allocation and mapping retention in CGN. The analysis shows that recommended timeout values in [5], [10] might be too long, resulting in suboptimal retention policies, especially for UDP traffic. Similarly to [9], we collect and analyze traffic traces from a real ISP network, but ours is the first work, to the best of our knowledge, to specifically target the problem of quantifying the impact of CGN on web browsing experience.

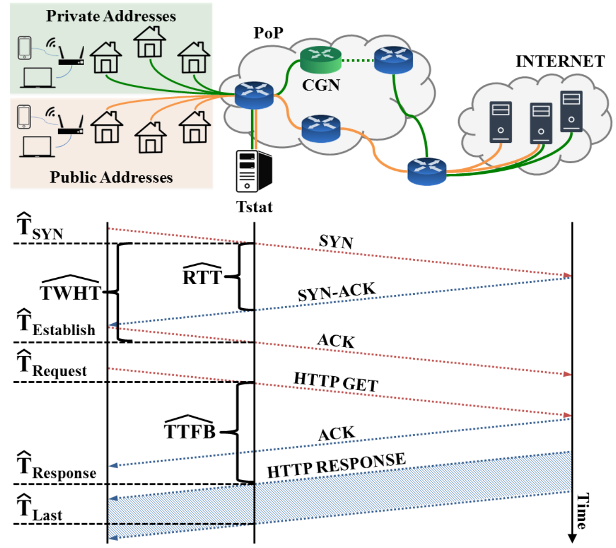


Fig. 1. The monitoring scenario we consider in this study (upper part) and an example of time line of a Web transaction.

III. Monitoring Setup and Dataset

We rely on passive measurements obtained by instrumenting a passive monitoring probe in the operational network of an European country-wide ISP. Fig. 1 (top) depicts the monitoring scenario. Each customer device accesses the Internet via an ADSL home router. The ISP assigns either a public or private IP address to each home router according to the user's contract. Home routers with a public IP address (public home routers) access the Internet directly, while the traffic of customers behind home routers with a private address (private home routers) reaches the Internet through a CGN device.

The CGN used by the monitored ISP is based on the NAT444 standard [3], which relies on *sessions* to translate the private IP address of a home router into a public one. When the CGN receives the first packet from a private home router, it starts a new session, temporarily mapping the private address to the first available in a pool of public addresses. It then converts the address of all subsequent packets according to the mapping.[3] After a given inactivity time of the private home router, the session expires and the public address is put back in the pool of free addresses.

In our monitoring setup, we install a passive probe at one Point of Presence (PoP) of the ISP to monitor the traffic generated by home routers having either a public or a private IP address. The probe runs Tstat [1], a passive monitoring tool that observes all packets flowing on the link connecting the PoP to the ISP backbone network. Tstat rebuilds each TCP flow, tracks it, and, when the connection is torn down, logs more than 100 statistics in a simple text file. For instance, Tstat logs the client[4] and server IP addresses, the application (L7) protocol type, the amount of bytes and packets sent and received, etc. Finally, Tstat implements DN-Hunter [7], a plugin that annotates each TCP flow with the server Fully Qualified Domain Name (FQDN) retrieved via DNS queries. This is particularly useful for unveiling services running on HTTP and HTTPS. Tstat separately logs TCP connections for which the Three-Way Handshake is not completed (e.g., when the sole SYN message is observed). In the remaining, we refer to this log type as *failed-TCP*, and we focus on such traffic to investigate on possibly unsolicited traffic reaching the ISP customers (see Sec. VII).

For this study we leverage a dataset collected during the month of October 2014. It consists of TCP and failed-TCP logs carrying 1,757M and 648M records respectively, for a total of more than 50 TB of network traffic. As we target the performance assessment for web browsing, we specifically focus on flows carrying either HTTP or HTTPS transactions. Overall, we process more than 400M TCP flows containing 688M HTTP requests. We split each of our logs in two subsets according to the IP address type of the customer's home router. We find more than 10,000 home routers active over the month, out of which 60% (40%) are assigned a private (public) IP address.[5] Similarly, 238M (59%) TCP flows are generated by private home routers, and 162M (41%) by public ones.

IV. Methodology

Among the many measurements provided by Tstat, we consider for each TCP flow: (i) The Round-Trip-Time (RTT) between client and server; (ii) the Time-To-Live (TTL) of packets sent by the server; (iii) the amount of bytes sent and received by the client; (iv) the application layer protocol (e.g., HTTP and HTTPS); and (v) the timestamps of packets that are instrumental to obtain further indices.[6] These metrics are

[3]The amount of public addresses available at the NAT is smaller than the number of customers provided with a private IP. Consequently, the pool size of public addresses must be carefully set to minimize allocation costs, while guaranteeing satisfactory connectivity.

[4]We take care of obfuscating any privacy sensitive information in the logs (e.g., customer IP addresses are anonymized using irreversible hashing functions with the advantage of the Crypto-PAn library). Private IP addresses are labeled as such by Tstat before anonymization.

[5]The home router IP address can be considered as an identifier of the household. It may hide several devices connected to the Internet.

[6]Notice that the probe measures the timestamps at a vantage point close to the customers. Therefore, for some metric X we can only gauge its estimated measure $\widehat{X}$.

straightforward to monitor, and details can be found in [1]. We also consider the FQDN, and we leverage it to split the traffic according to the service generating it.

We combine basic metrics provided by Tstat to build indices we use to compare the impact of the CGN on users' traffic at network and transport level (Sec. IV-A and Sec. IV-B, respectively). Plus, in Sec. IV-C we present some indices defined on purpose to measure the potential impact of the CGN on users' browsing experience.

A. Network Metrics

1) Number of Hops – $\#Hops$: The minimum number of hops being traversed by packets transmitted from the server to the client. Given the maximum server-to-client TTL in a flow (TTL), we choose x as the exponent minimizing $\#Hops = 2^x - TTL$.[7] The resulting $\widehat{\#Hops}$ is the minimum number of hops that packets in flow have traversed before reaching their destination. In our scenario we expect packets received by private home routers to traverse a higher number of hops due to the presence of the CGN.

2) Round Trip Time – RTT: The average RTT Tstat measures in a flow ($\widehat{RTT}$) on packets transmitted from the client to the server (as depicted in the lower part of Fig. 1). We expect packets transmitted by private home routers to experience a higher latency because of the CGN packet processing.

B. TCP Metrics

1) Three-Way Handshake Time – $TWHT$: The amount of time measured by Tstat ($\widehat{TWHT}$) the client takes to successfully establish a TCP connection using the standard Three-Way Handshake (TWH). Referring to the lower part of Fig. 1, let $\hat{T}_{SYN}$ be the timestamp of the SYN packet sent by the client to start the connection establishment procedure, and let $\hat{T}_{Establish}$ be the timestamp of the packet carrying the ACK message ending the TWH. We define the $\widehat{TWHT}$ as

$$\widehat{TWHT} = \hat{T}_{Establish} - \hat{T}_{SYN}$$

In our scenario we expect the $\widehat{TWHT}$ to be higher for private home routers due to the time needed by the CGN to allocate the resources for the new communication session.

For the sake of completeness, we also consider some specific TCP metrics: (i) The number of SYN messages needed to open a connection, SYN; (ii) the number of out of sequence segments, OoS; (iii) the number of duplicated segments Dup. These are measurements that we expect to be altered in case of connectivity issues introduced by the CGN. A large value of SYN, for instance, indicates that the client experienced some difficulties trying to establish the connection.

C. Performance Metrics

1) Time to first byte – $TTFB$: The amount of time that elapses between the first segment containing the HTTP request sent by the client to the first segment with payload sent by the server. Again referring to Fig. 1, let $\hat{T}_{Request}$ be the timestamp of the first segment the client sends carrying application data, and $\hat{T}_{Response}$ the timestamp of server first response with payload. We define the TTFB as

$$\widehat{TTFB} = \hat{T}_{Response} - \hat{T}_{Request}$$

In HTTP flows, it represents a measure of the time span between the application request issued by the client and the consequent response by the server.

2) Goodput – G: The average rate at which the server delivers information to the client. Let $\hat{T}_{Response}$ and $\hat{T}_{Last}$ (see Fig. 1) be the timestamps of the first and the last data packet sent by the server, and let D be the size of the application payload carried by the flow. We define the server goodput as

$$\hat{G} = \frac{D}{\hat{T}_{Last} - \hat{T}_{Response}}$$

It is similarly possible to evaluate the goodput in the upload direction by considering the amount of bytes sent by the client to the server and referring to the timestamps relative to the client traffic. To avoid the bias of short-lived flows, we evaluate the download goodput only on flows for which $D \geq 1$ MB, and the upload goodput for flows where $D \geq 500$ kB.

For each of the above metrics, we build empirical distributions, i.e., Probability Density Functions (PDFs), separating the traffic involving private and public home routers. Hence, to pinpoint the metrics affected by the CGN, we adopt a tool that allows the comparison of the collected empirical distributions. Our choice falls on the Jensen-Shannon divergence, a popular statistical index based on the Kullback-Leibler divergence. Among its relevant properties, the the Jensen-Shannon divergence is bounded to finite values and symmetric.

D. Jensen-Shannon divergence

To compactly represent the difference between a PDF p and a PDF q we use the Jensen-Shannon divergence [13], which varies in the range $[0, \ln(2)]$, and is defined as

$$JS_{div} = \sum_i \left\{ \frac{1}{2} p_i \ln\left(\frac{p_i}{\frac{1}{2}p_i + \frac{1}{2}q_i} \right) + \frac{1}{2} q_i \ln\left(\frac{q_i}{\frac{1}{2}q_i + \frac{1}{2}p_i} \right) \right\}$$

The two variables p_i and q_i are the probabilities composing the two distributions. To avoid statistical bias, which may lead to wrong conclusions, we need to put ourselves in conditions to properly evaluate the JS_{div}, and to discriminate between notable and negligible differences in the distributions. Aside the requirement for statistically relevant population sizes, the JS_{div} may be affected by more sneaky sources of bias, for instance, tied to heterogeneity in the population size, as well as to the binning strategy to compute p_i and q_i. Intuitively, the population size must to be large enough to prevent border effects tied to the finitude of the dataset. The type of samples included in the population should also be akin, and the tool used to measure the statistics should be well calibrated to avoid arising artifacts. In our case, the population samples consists of TCP flows. As described in Sec. III, our dataset is large enough to avoid biases due to border effects.

We focus on the selection of a threshold to discriminate among notable and negligible differences. We remind that for two completely disjoint statistics, the JS_{div} saturates to $\ln(2)$.

To visually tie the JS_{div} to some examples, we resort to negative exponential distributions. We generate a reference sample from distribution of parameter λ_0=1. A second set

[7] Depending on the OS of the device generating the packets, the initial TTL may be set to different values. Common choices are 32, 64, 128, 255.

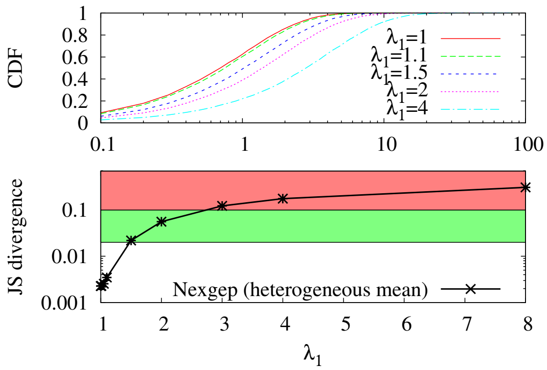


Fig. 2. Jensen-Shannon divergence computed on negative exponential distributions with mean rates λ_1 versus reference mean $\lambda = 1$.

TABLE I. JENSEN-SHANNON DIVERGENCE FOR CONSIDERED METRICS AND DIFFERENT INTERNET SERVICES.

	All Flows	**www.google.com**	**TOP-50 Google**	**phobos.apple.com**
$\#\widehat{Hops}$	**0.223**	**0.666**	**0.682**	**0.689**
$\widehat{RTT}$	0.001	0.006	0.007	0.007
$\widehat{TWHT}$	0.002	0.010	0.011	0.016
$\#SYN$	<0.001	<0.001	<0.001	<0.001
OoS	<0.001	–	–	–
Dup	0.001	0.001	0.001	<0.001
$\widehat{TTFB}$	0.002	0.006	0.008	0.006

of samples is instead shaped according to a distribution of parameter λ_1. Then, we compute JS_{div} comparing the two PDFs of parameter λ_0 and λ_1. We set a very large population size (10^6) so that non null JS_{div} scores are only minimally tied to the population size. For our experiment we use $\lambda_1 \in [1, 8]$. Cumulative Distribution Functions (CDFs) of the negative exponential distributions are depicted in the top portion of Fig. 2, whereas the bottom plot reports the JS_{div}.

As shown, thresholds are set in such a way that clearly visible changes in the distribution space also raise alerts in the JS_{div} space. Intuitively, when $JS_{div} \in [1/10, \ln(2)]$, the difference between the two CDFs is significant (red area). When $JS_{div} \in [2/100, 1/10)$ the difference is noticeable (green area), and negligible if $JS_{div} \in [0, 2/100)$ (white area).

We also generate finite sequences with other known distributions of which we evaluate JS_{div}, and we observe that scores are similar across them (we do not report results for the lack of space). This means that JS_{div} is robust against the kind of distributions we analyze.

In the next sections we make use of the JS_{div} to contrast private home routers against public home routers over several empirical metric distributions we obtain from our dataset.

V. IMPACT OF CGN ON USERS' TRAFFIC

We start our analysis by gauging the impact of CGN on the network and the transport level metrics described in Sec. IV-A and Sec. IV-B, respectively. The goal is to check if home routers with private IP addresses experience worse performance than those with public addresses.

A. Impact of CGN on Network- and Transport-level metrics

For this analysis, we consider TCP flows in which the client IP address belongs to the set of monitored customers while the server IP address is external. Distinguishing between clients with private and public IP addresses, we compute the distributions for each metric described in Sec. IV-A and Sec. IV-B, and we evaluate the Jensen-Shannon divergence for them. We report the results in Tab. I, repeating the experiment selecting flows directed to (i) any remote server ("all flows"); (ii) "www.google.com" servers (i.e., *Google Search*); (iii) TOP-50 most popular Google servers in our dataset ("TOP-50 Google"); and (iv) "phobos.apple.com" servers providing *iTunes Store* contents.[8] As shown, the JS_{div} never overcomes the alarm threshold discussed in Sec. IV-D for all metrics but $\#\widehat{Hops}$, meaning that the CGN configuration of our scenario does not induce any significant bias.

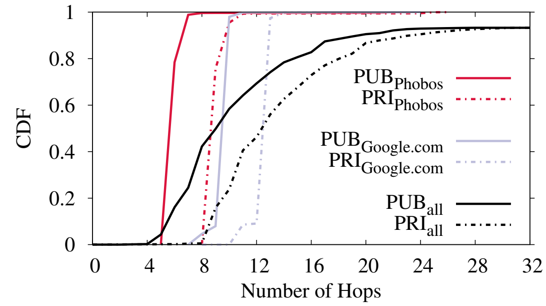


Fig. 3. CDF of the number of hops ($\#\widehat{Hops}$) measured from the server to the client for private and public home routers against different web services. Clear differences are visible.

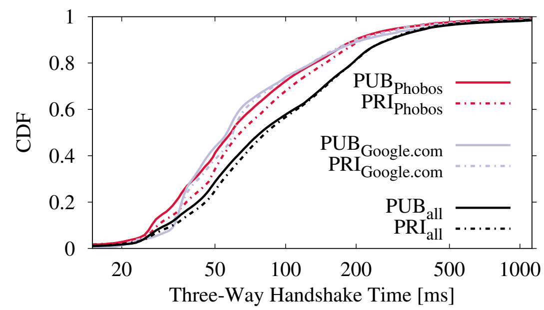


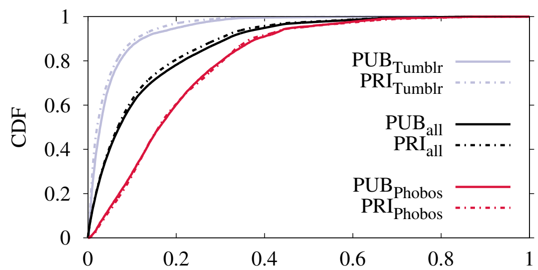
Fig. 4. CDF of time needed to complete the Three-Way Handshake ($\widehat{TWHT}$) for private and public home routers against different web services. No significant difference is visible.

The only metric that consistently overcomes the threshold across all the considered Internet services is the number of hops ($\#\widehat{Hops}$). To validate the above finding, we directly compare the distributions of $\#\widehat{Hops}$ for private and public home routers in Fig. 3. For the ease of visualization, we do not report the case of "TOP-50 Google" servers as we observe similar results to the "www.google.com" case. A clear offset between the $\#\widehat{Hops}$ of private and public home routers appears, showing that private ones have to traverse more hops to reach the Internet. Such offset is independent on the considered Internet service. We verified this outcome with the ISP network administrators, who confirmed that the difference is due to some extra routers that packets forged by private home routers have to go through to reach the CGN. However, such routers are well dimensioned and not congested, with little to no implication on the performance, as testified by other metrics considered in Tab. I.

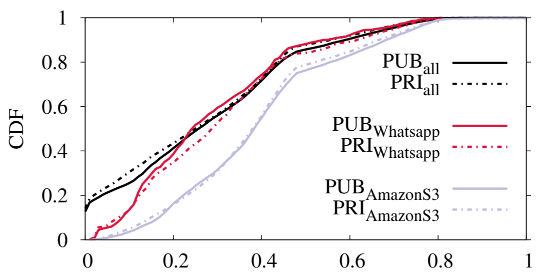
[8]We focus on this selection of services as they appear to be popular on the monitored network, and the amount of TCP flows for each of them satisfies the requirements for a proper use of the JS_{div}.

TABLE II. JENSEN-SHANNON DIVERGENCE FOR THROUGHPUT DISTRIBUTIONS IN DOWNLOAD AND UPLOAD DIRECTIONS.

	Service	FQDN	JS div
Download	All	*	0.001
	Facebook Video	fbcdn-video-*.akamaihd.net	0.004
	Tumblr	media.tumblr.com	0.021
	Phobos	phobos.apple.com	0.022
Upload	All	*	0.004
	Amazon S3	eu-irl-*.s3.amazonaws.com	0.007
	Whatsapp	mm*.whatsapp.net	0.033
	Dropbox	dl-*.dropbox.com	0.046



(a) Download. Only flows carrying $\geq$ 1 MB are considered.



(b) Upload. Only flows carrying $\geq$ 500 kB are considered.

Fig. 5. Normalized goodput CDFs for flows carrying Web traffic.

We also report the distributions for the connection establishment time $\widehat{TWHT}$. This is a typical metric one could expect to be affected by additional delay introduced by the CGN when private home routers try to establish new connections. Indeed, the CGN may require some time to initiate the session and translate addresses. Fig. 4 shows that distributions for private and public home routers with respect to the same Internet services are in practice identical, showing no considerable shift in the connection setup time. Such result is also confirmed by Tab. I, which reports low values of JS_{div} for this metric.

B. Impact of CGN on users' web browsing quality

We complement the above findings by applying the JS_{div} on the indices presented in Sec. IV-C. As reported in Tab. I, the JS_{div} of the Time to First Byte, $\widehat{TTFB}$, indicates that this metric is not affected by the presence of the CGN, and that users accessing the Internet from private or public home routers face similar delays.

Next, we perform the same analysis for the goodput $\hat{G}$. We consider several popular services that exchange large amount of data and for which $\hat{G}$ is thus relevant. In particular, we consider flows downloading content from Facebook Video, Tumblr and Phobos servers, and flows uploading user data to Amazon S3, Whatsapp and Dropbox. We report the results in Tab. II, and draw the CDFs in Fig. 5. Observe that the JS_{div} does not overcome the alarm threshold, meaning that the CGN does not significantly harm the download/upload speed of private home routers. Fig. 5(a) depicts the distribution of the normalized download $\hat{G}$ for the services reported in Tab. II (we omit Facebook Video to ease the visualization). Note that the differences between each pair of curves is negligible. Fig. 5(b) reports results for the normalized upload $\hat{G}$. Also in this case the curves referring to private and public home routers show very similar trends.[9]

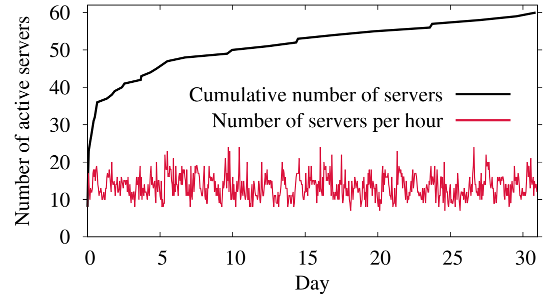


Fig. 6. Number of active servers in the PoP with per-hour granularity (red curve), and cumulative number of active servers over one month (black curve).

Interestingly, a relatively large amount of flows (13.98%) in Fig. 5(b) show almost zero throughput. By double-checking, we realize that those are long-lived flows with a duration of 10 min (or more), and showing a number of uploaded bytes that slightly exceeds the 500 kB threshold. For some services, indeed, clients establish a single TCP connection with the remote server and keep sending tiny portions of data intermittently, de facto zeroing the upload throughput.

VI. ACTIVE SERVERS IN THE POP

The per-year cost to rent a public IP address (around 10$) is a non-negligible expense when multiplied for the number of subscribers. Thus, considering the results presented in Sec. V, we can conclude that the ISP has no actual incentive to provide users' home routers with public IP addresses. However, one may argue that some customers may be interested in having a public IP address to host servers they want to maintain accessible from outside the ISP network. In this scenario, the only way to guarantee the server reachability is to assign a public IP address to the customer's home router. We thus perform a further investigation to gauge the number of home routers with public address behind which some kinds of servers are running. To count the number of active servers in our dataset we consider all public home routers that generate at least one HTTP, HTTPS, IMAP, POP or SMTP connection on a daily basis.

Fig. 6 shows the cumulative number of distinct active servers we observe over one month's time, together with the per-hour number of active servers. This result is boggling: among the approximately 4,000 public home routers we monitor in the PoP, 60 of them are actually running services being accessed from the Internet. This enforces our claim that the users have no effective need to ask for home routers with public IP addresses.

VII. UNSOLICITED TRAFFIC

In this last section, we quantify how many home routers interfacing the Internet by means of public/private IP address

[9] We normalize the measured throughput to not show the actual bandwidth provided by the monitored ISP.

TABLE III. PERCENTAGE OF PUBLIC AND PRIVATE HOME ROUTERS TARGETED BY UNSOLICITED TRAFFIC. TOP-20 DESTINATION PORTS ARE SHOWN AND SORTED ACCORDING TO PUBLIC ADDRESSES POPULATION.

Port	Description	Private	Public
0	Illegal –OS fingerprinting–	0.1	79.3
135	*Multiple*†, MS Remote Procedure Call⋆	<0.1	79.3
143	ADM†⋆ IMAP	<0.1	79.3
1433	*Multiple*†⋆, MS SQL Server	0.1	79.3
2222	*Multiple*†, Rockwell CSP2	1.0	79.3
3306	Nemog†, W32.Spybot†, MySQL Server	0.1	79.3
3389	Windows Remote Desktop Protocol⋆	0.1	79.3
5900	Evivinc†, Virtual Network Computing	<0.1	79.3
32764	Cisco Access Point⋆, Cisco Routers⋆	<0.1	79.3
3128	*Multiple*†, Proxy servers	<0.1	79.3
22	*Multiple*†, SSH	3.2	79.2
445	*Multiple*†⋆, MS Active Directory	<0.1	79.2
995	POP3 over SSL	<0.1	79.2
8080	*Multiple*†, HTTP Alternate	<0.1	79.1
25	*Multiple*†, SMTP	<0.1	79.0
443	*Multiple*†, HTTPS / SSL	0.1	78.9
80	*Multiple*†, HTTP	1.8	78.5
21	*Multiple*†, FTP	4.6	78.1
23	*Multiple*†, Telnet	<0.1	77.9
139	*Multiple*†⋆, NetBIOS	<0.1	56.6

†: Worm or Threat, ⋆: Known vulnerability.

are exposed to unsolicited incoming traffic. We perform an analysis based on the destination port used, which assesses the number of connection attempts we observe in our failed-TCP logs. First, we compile a list of IP addresses corresponding to potential attackers by counting the number of SYN messages they generate. In particular, we label as attacker every IP address that forges SYN messages directed to 50 (or more) distinct home routers in our PoP. Second, we check the port list, and we focus on those that are associated with known services or worms/threats. Hence, for each destination port, we compute: (i) The number of distinct attackers; (ii) the number of home routers contacted; and (iii) the number of connection attempts.

Tab. III reports, for the top-20 most contacted ports, the percentages of private and public home routers inside the PoP being targets of connection attempts. As clearly shown, the number of potential victims in the public home router set is close to 80% for the vast majority of the considered ports. Conversely, these percentages are minimal for private home routers (below 5% in the worst case), as private addresses can be reached only if the counterpart is within the borders of the ISP network. We observe similar results for the amount of connection attempts and the number of distinct attackers. Considering Port 22, for instance, the number of connection attempts peaks at 2 Millions against public home routers, and stops at only 6,500 against private home routers. Similarly, 10,000 attackers are found in the global Internet, while less than 200 are detected inside the ISP. We thus can validate our hypothesis: Public home routers are definitely more exposed to attacks than private ones, and CGN represents a first line of defense to limit unsolicited traffic. For instance, the CGN has the potential of curbing the spread of those bots whose goal is to exploit vulnerabilities at the home routers.

VIII. CONCLUSION AND FUTURE WORK

In this work, we leveraged passive measurements to gauge the impact of CGN deployment on the web browsing experience of users. To this end, we considered a large dataset of traffic traces that we split according to the type of IP address assigned to users' home routers, i.e., public or private. Then, we compared the two obtained populations leveraging different performance metrics. We relied on the Jensen-Shannon divergence to quickly pinpoint those showing stochastically significant difference.

Our results show that the CGN technology is stable and mature. If properly engineered and configured, the CGN does not harm users' web browsing activity. Moreover, we showed that the CGN presence brings some positive side effects, e.g., it protects home routers from unsolicited and possibly malicious traffic. We complemented such findings by analyzing the subset of users accessing services running on home servers from the Internet. We observed that only a marginal share of them actually exploits such setup. Hence, we conclude that the ISP may have no actual need to provide users with public IP addresses, when not specifically required.

In our ongoing efforts, we are planning to expand the list of metrics we considered in this paper. For instance, this work mostly relies on *per-flow* metrics to build its conclusions. It may be worth extending our focus to include *per-session* metrics. Finally, we are interested in performing the same analysis to gauge the impact of CGN on activities other than web browsing such as, e.g., P2P.

ACKNOWLEDGEMENTS

This work has been carried out at LINCS http://www.lincs.fr and funded by the *mPlane* project (grant agreement no. 318627) in the 7th European Framework Programme.

REFERENCES

[1] A. Finamore, M. Mellia, M. Meo, M. Munafo, and D. Rossi. Experiences of Internet traffic monitoring with tstat. *Network, IEEE*, 25:8–14, 2011.

[2] A. M uller, F. Wohlfart, and G. Carle. Analysis and Topology-based Traversal of Cascaded Large Scale NATs. In *HotMiddlebox*, 2013.

[3] C. Donley, L. Howard, V. Kuarsingh, A. Chandrasekaran, and V. Ganti. Assessing the Impact of NAT444 on Network Applications. Technical report, 2011.

[4] C. Donley, L. Howard, V. Kuarsingh, J. Berg, and J. Doshi. RFC 7021 - Assessing the Impact of Carrier-Grade NAT on Network Applications. Technical report, 2013.

[5] F. Audet, and C. Jennings. RFC 4787 - Network Address Translation (NAT) Behavioral Requirements for Unicast UDP. Technical report, 2007.

[6] G. Detal, B. Hesmans, O. Bonaventure, Y. Vanaubel, and B. Donnet. Revealing Middlebox Interference with Tracebox. ACM IMC, 2013.

[7] I. Bermudez, M. Mellia, M. Munafo, R. Keralapura, and A. Nucci. DNS to the Rescue: Discerning Content and Services in a Tangled Web. ACM IMC, 2012.

[8] L. Eggert, V. Stirbu, and H. Tschofenig. A Survey of Protocols to Control Network Address Translators and Firewalls. Technical report, 2007.

[9] S. Alcock, R. Nelson, and M. David. Investigating the Impact of Service Provider NAT on Residential Broadband Users. Technical report, University of Waikato, 2010.

[10] S. Guha, K. Biswas, B. Ford, S. Sivakumar, and P. Srisuresh. RFC 5382 - NAT Behavioral Requirements for TCP. Technical report, 2008.

[11] S. Jiang, D. Guo, and B. Carpenter. RFC 6264 - An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition. Technical report, 2011.

[12] S. Perreault, I. Yamagata, S. Miyakawa, A. Nakagawa, and H. Ashida. RFC 6888 - Common Requirements for Carrier-Grade NATs (CGNs). Technical report, 2013.

[13] H. Schutze and C. D. Manning. In *Foundations of Statistical Natural Language Processing*, Cambridge, MA, USA, 1999. MIT Press.