An analysis on spoofable IPs and Source Address Validation using CAIDA's Spoofer datasets

Amit Kumar Gupta
Masters Student, Security and Privacy, University of Twente, Netherlands
a.k.gupta@student.utwente.nl

ABSTRACT
This research paper brings forward the importance of protection of IP addresses from being spoofed. Based on the vast utilization of internet and the inter connected devices over the networks, the impact of cybersecurity is strongly reflected on the economies of the world, specially how of the DDoS attacks lead to massive losses to e-businesses. The information from CAIDA's Spoofer application has been used as a basis to study the impact of Source Address validation techniques on control over the uncontrolled DDoS attacks in present day Internet infrastructures. In the later section of the document, the existent problem has been devised to a research question. The probable solutions have been hypothesized and using statistical measures, the conclusion has been deduced. We conclude that the network nodes should implement ingress filtering in order to contribute for a better Worldwide network system. Analyzing the Spoofer data for the Netherlands, it appears that majority of the Dutch ASes are protected however, due to the lack of sufficient data we cannot extrapolate the result as a reflection of 100 percent of the ASes.

Keywords
Source address validation, IP spoofing, filtering, Autonomous Systems

1. INTRODUCTION
Today, we call the world as a global village. We, the humans have been really innovative and tremendously successful in connecting the people and machines and facilitate their communications. From developing ARPANET in 1960's [1] to today's light fast fiber optics internet connections, we have progressed a long way in the right direction. But it goes unsaid that with great power, comes great responsibilities to make systems safe and secure. No doubt that the billions of connected devices over IoT [2] give us immense empowerment, but the question is how sure are we that the generation old Internet protocols are robust to handle the abuses.

WORLD INTERNET USAGE AND POPULATION STATISTICS JUNE 30, 2016 - Update						
World Regions	Population (2016 Est.)	Population % of World	Internet Users 30 June 2016	Penetration Rate (% Pop.)	Growth 2000-2016	Table % Users
Asia	4,052,652,889	55.2 %	1,846,212,654	45.6 %	1,515.2%	50.2 %
Europe	832,073,224	11.3 %	614,979,903	73.9 %	485.2%	16.7 %
Latin America / Caribbean	626,119,788	8.5 %	384,751,302	61.5 %	2,029.4%	10.5 %
Africa	1,185,529,578	16.2 %	340,783,342	28.7 %	7,448.8%	9.3 %
North America	359,492,293	4.9 %	320,067,193	89.0 %	196.1%	8.7 %
Middle East	246,700,900	3.4 %	141,489,765	57.4 %	4,207.4%	3.8 %
Oceania / Australia	37,590,820	0.5 %	27,540,654	73.3 %	261.4%	0.8 %
WORLD TOTAL	7,340,159,492	100.0 %	3,675,824,813	50.1 %	918.3%	100.0 %

Fig 1: "Billions of devices are connected to the Internet by 2016"

One such gigantic problem today is the Distributed Denial of Service Attack (DDoS) or Denial of Service, as you may call it. We understand one of the core issues how attackers are able to exploit the current Internet infrastructure is by spoofing data request packets from innocent IP Addresses, just with the intention to overload the target servers bandwidth capacity and lead to its downtime. Forcing all IP packets to carry correct source addresses can greatly help network security, attack tracing, and network problem debugging. [3]

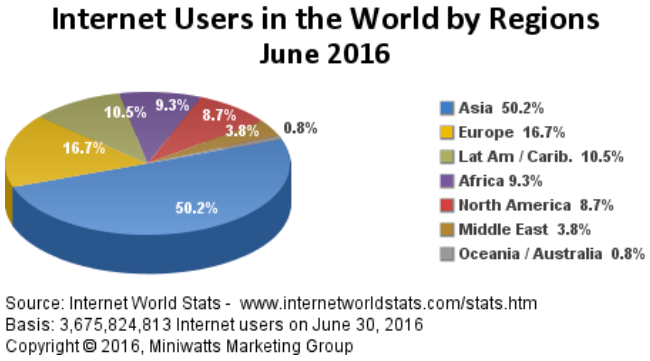


Fig 2: Geographical spread of Internet users in the world

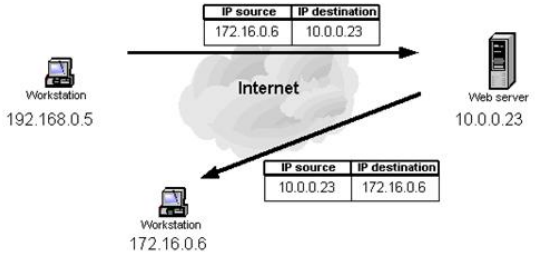


Fig 3: IP Spoofing: A simplified example.

In this paper, we will analyze the study of Center for Applied Internet Data Analysis, CAIDA's "Spoofer" [4] project. Since the problem of spoofable IPs is diverse and geographically spread, in particular, we will be focusing on analyzing the status of one of the countries – in this case the Netherlands, and the data around it.

2. LITERATURE REVIEW

Securing the network of networks its susceptibility to malicious attacks has been quite a concern of the Internet Service Providers and the Infrastructure architects.

Research and development in this domain has been a hot plate since its existence. There have been many research papers, which reflect intensity on the same domain of problem. In one of the papers [5] presented at the ACM [6], the authors say how IP source address forgery, or "spoofing," is a long-recognized consequence of the Internet's lack of packet-level authenticity. Despite historical precedent and filtering and tracing efforts, attackers continue to utilize spoofing for anonymity, indirection, and amplification. Using a distributed infrastructure and approximately 12,000 active measurement clients, they collected data on the prevalence and efficacy of current best practices source address validation (SAV) techniques. They quote that, of clients able to test their provider's source-address filtering rules, they find 31% were able to successfully spoof an arbitrary, routable source address, while 77% of clients otherwise were unable to spoof can forge an address within their own /24 sub-network.

Seeking to minimize Internet's susceptibility to spoofed DDoS attacks, CAIDA has been developing and supporting open-source software tools to assess and report on the deployment of source address validation (SAV) best anti-spoofing practices. CAIDA has developed and support a new client-server system that periodically test a network's ability to both send and receive packets with forged source IP addresses (spoofed packets).

One logical solution that came out from the observed gravity of the problem was to use ingress filters on the end of Internet nodes and serving houses. In ingress filtering, packets coming into the network are filtered if the network sending it should not send packets from the originating IP address (es). If the end host is a stub network or host, the router needs to filter all IP packets that have, as the source IP, private addresses (RFC 1918), bogon addresses or addresses that do not have the same network address as the interface. [7]

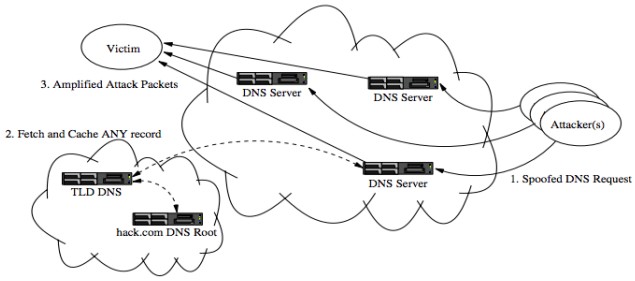


Fig 3:DNS Amplifier: 1) Attacker spoofs DNS request with victim's source for large TXT record. 2) Third-party DNS servers fetch and cache record. 3) Server responds to victim. Attacker's small query packets are amplified and victim cannot identify attack source. [5]

In one of the informational publications [8], Dr. Ferguson and Dr. Senie say, that Ingress traffic filtering at the periphery of Internet connected networks will reduce the effectiveness of source address spoofing denial of service attacks. They believe that though many Network service providers and administrators have already begun implementing this type of filtering on periphery routers, it is highly recommended that all service providers do so as soon as possible. In the also report that, in addition to aiding the Internet community as a whole to defeat this attack method, it can also assist service providers in locating the source of the attack if service providers can categorically demonstrate that their network already has ingress filtering in place on customer links.

In one of the research studies [9] where authors Beverly and Bauer study the origin and nature of source address filtering claim an interesting conclusion that a significant fraction, approximately one-quarter, of the net blocks, IP addresses and ASes observed permit spoofing. This suggests that a large portion of the Internet is still vulnerable to spoofing and concerted spoofing attacks remain a serious concern. Projecting our results to the entire Internet yields over 360M spoofable addresses and 4,600 ASes from which spoofing is possible. This is particularly significant if next generation attack farms intelligently probe the network and adaptively change behavior based on the ability to spoof.

3. OBJECTIVE

Given our understandings about the harsh effects because of DDoS attacks on standing economies, this paper brings forward a perspective about the gravity of the issues pertaining around IP Spoofing in the present day internet

protocol system and how source address validation can mean some help. Keeping in mind the results from CAIDS's Spoofer tools, this brings in to the postulation of a strong question in the domain of economics of cyber security - ***given the negative impacts of DDoS attacks, it is suggested that all the ASes implement SAV on their network nodes, then what is that one corrective action that decision makers can take?***

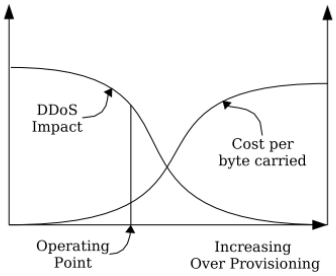


Fig 4: Cost per-byte-carried and DDoS Impact as a function of network over-provisioning [10]

The cost of a Distributed Denial of Service (DDoS) attack can continue to impact on the targeted organization long after the event has been dealt with. It is not just the disruption to the public interface, which is damaging enough to any organization that conducts a substantial volume of its business online. Loss of revenues while services and systems are unavailable to customers are compounded by the cost of rectifying the crisis and long-term damage to the business's reputation In some cases an organization might even submit to extortion from the hackers, effectively paying a ransom to rid itself of the problem – until the next strike from another hacker source. [11]. Then what is it that's stopping the empowered people to take charge on the situation and take decisions to mitigate the arising problems. The reason behind the situation can be hypothesized into two basic provisions:

H0: ***As a result of the SAV ingress filtering, there will be no change on the security of Servers against DDoS, or there will be significant improvement in security of the systems against DDoS.***

H1: ***With advents in SAV ingress filtering, there will be no visible change on the security of systems exposed to Internet against DDoS. ; (NULL Hypothesis)***

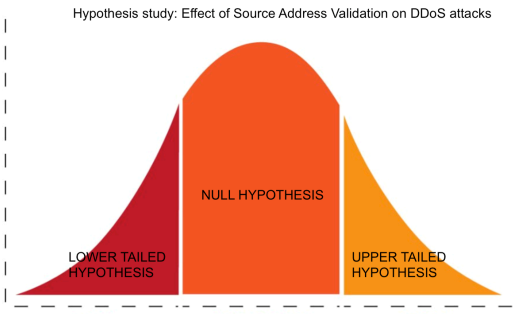


Fig 5: Two tailed hypothesis study for H0 and H1.

4. METHODOLOGIES

To resolve the legitimacy of the hypothesis, I will first try to analyze the options different actors have in hand to take a step forth in implementing ingress filtering on their network nodes and do quantitative and statistical analysis on the data available on Spoofer application on CAIDA's portal.

Having multiple actors involved in a security issue like network operators, country IT security board and ASes, always makes it difficult to find a common concrete countermeasure that all of them can take. Not to mention in SAV the one who bears the cost doesn't fully reap the benefits. However, considering a counter measure where all the actors can at least benefit the mitigation of the security issue opens the window for many risk strategies that can be deployed mainly at Network Providers end.

The one we will focus in this paper is Ingress Filtering, which can solve the root cause and give solution as well to other actors. Now it can be debated that ingress filtering is a very old technique to do source address validation. But, literature [8] shows that this method has been found to be highly effective in diminishing DDoS. New techniques like Unicast Reverse Path Forwarding, or uRPF which are preferred to use today or also the automation of Ingress Filtering implementation. Which are further explained below.

The prime decision makers, Network Providers will need to implement ingress-filtering rules, which check the source IP field of the IP packets it receives. If the source IP address is not within a range of legitimately advertised prefixes, a router will drop the packet.

There are at least five ways to implement *ingress filtering* technique in network operator:

4.1 Ingress Access List

An Ingress Access List will filter and checks the source address of every message received on a network interface against a list of acceptable prefixes, then dropping any packet that does not match the filter.

4.2 Strict Reverse Path Forwarding

It is conceptually identical to using access lists for ingress filtering, with the exception that the access list is dynamic.

4.3 Feasible Path Reverse Path Forwarding

Feasible Path Reverse Path Forwarding (Feasible RPF) is an extension of Strict RPF. The source address is still looked up in the RPF-specific table but instead of just inserting one best route there, the alternative paths (if any) have been added as well, and are valid for consideration.

4.4 Loose Reverse Path Forwarding

Loose Reverse Path Forwarding (Loose RPF) is algorithmically similar to strict RPF, but differs in that it checks only for the existence of a route (even a default route, if applicable), not where the route points to. Practically, this could be considered as a "route presence check" (loose RPF is a misnomer in a sense because there is no "reverse path" check in the first place).

4.5 Loose Reverse Path Forwarding ignoring default routes

The fifth implementation technique may be characterized as Loose RPF ignoring default routes, i.e., an "explicit route presence check". In this approach, the router looks up the source address in the route table, and preserves the packet if a route is found. However, in the lookup, default routes are excluded. Therefore, the technique is mostly usable in scenarios where default routes are used only to catch traffic with bogus source addresses, with an extensive (or even full) list of explicit routes to cover legitimate traffic.

By implemented ingress filtering it will not only mitigate the security issue and provide benefits to the problem owner (network provider) but as well as all the rest of the actors mentioned above.

5. RESULTS

In this section, lets perform statistical analysis to several factors that can explain that the variances are assessed, i.e. the percentage of spoofable and non-spoofable ASNs, the type of these ASNs, the prefixes and the degree and degree of their correlation.

Correlations

		Prefixes	ASN
Prefixes	Pearson Correlation	1	-0.131
	Sig. (2-tailed)		0.604
	N	18	18
ASN	Pearson Correlation	-0.131	1
	Sig. (2-tailed)	0.604	
	N	18	29

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	0.875	0.541
	Cramer's V	0.875	0.541
	Contingency Coefficient	0.659	0.541
N of Valid Cases		18	

The Pearson's r for the correlation between the prefixes and ASN (spoofable or not spoofable) is -0.131. When Pearson's r is close to 0 means that there is a weak relationship between the two variables. And when the Pearson's r is negative means that as one variable increases in value, the second variable decreases in value. In our case there is a weak relationship between ASNs and the number of prefixes and also a negative correlation.

Sig (2-Tailed) value is the value, which tells us if there is a statistically significant correlation between our two variables. In our example, our Sig. (2-tailed) value is 0.604 and it is bigger than the value 0.05 (the acceptable probability of error). We can conclude that there is no statistically significant correlation between our two variables. That means, increases or decreases in one variable do not significantly relate to increases or decreases in the second variable.

Phi and Cramer's V are both tests of the strength of association. We can see that the strength of association between the variables is very strong (0.875), as it is closer to 1.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	1.165	0.659
	Cramer's V	0.824	0.659
	Contingency Coefficient	0.759	0.659
N of Valid Cases		18	

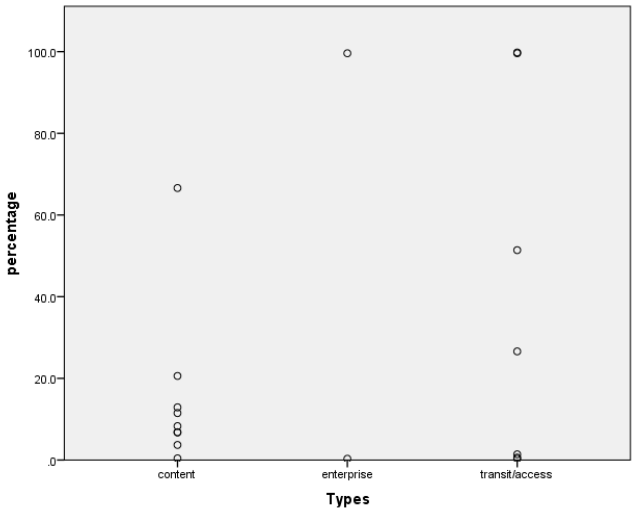
Correlations

		type	percentage
type	Pearson Correlation	1	-0.327
	Sig. (2-tailed)		0.186
	N	18	18
percentage	Pearson Correlation	-0.327	1
	Sig. (2-tailed)	0.186	
	N	18	18

The Pearson's r for the correlation between the type of the ASNs (content, enterprise and transit/access) and the percentage of spoofable and non-spoofable ASNs is -0.327. That means that there is a weak relationship between these two variables, as the value of the Pearson's r is closer to 0. Also, the Pearson's r value is negative which means that as the one variable increases in value, the second variable decreases in value, i.e. the more enterprise ASNs we have, the percentage of non spoofable ASNs decreases. In our case there is a weak relationship between type of ASNs and the percentage of spoofable IPs, as well as a negative correlation.

In this case, the Sig. (2-tailed) value is 0.186, which is bigger than the value 0.05. We can conclude that there is no statistically significant correlation between our two variables. That means, increases or decreases in one variable do not significantly relate to increases or decreases in the second variable.

From Phi and Cramer's V tests we can see that the strength of association between the variables is very strong. The Cramer's V value is 0.824 and the Phi is 1.165.



Correlations

		ASN	DEGREE
ASN	Pearson Correlation	1	0.529**
	Sig. (2-tailed)		0.003
	N	29	29
DEGREE	Pearson Correlation	0.529**	1
	Sig. (2-tailed)	0.003	
	N	29	29

** Correlation is significant at the 0.01 level (2-tailed).

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	0.908	0.021
	Cramer's V	0.908	0.021
	Contingency Coefficient	0.672	0.021
N of Valid Cases		29	

The Pearson's r for the correlation between the ASNs (spoofable and un-spoofable) and the degree of them is 0.529. That means that there is a relative strong relationship between these two variables, as the value of the Pearson's r is closer to 1. Furthermore, the Pearson's r value is positive which means that as the one variable increases in value, the second variable increases also in value, i.e. increasing the degree of ASNs, we also increase the spoofable IPs connecting to them.

In this case, the Sig. (2-tailed) value is 0.003, which is smaller than the value 0.05. We can conclude that there is statistically significant correlation between our two variables. That means, increases or decreases in one variable, significantly relate to increases or decreases in the second variable.

From Phi and Cramer's V tests we can see that the strength of association between the variables is very strong. The Cramer's V and the Phi value is 0.908, very close to 1.

6. LIMITATIONS

6.1 The research produced in this paper has primarily focused on the trends in the data from the Spoofer application. Since the data is not very consistent, hence data cleaning has been done as in where required.

6.2 Due to having access to a limited dataset, the calculations in the section results have been constrained to focus on the limited high quality datasets coming from

6.3 The samples for the ASN information for the country Netherlands were quite small. For Prefixes, we had 18 and ASNs we have 29 samples. Though this can be seen as a reflection of the wholesome information, but it would definitely be more impactful to have more samples.

7. CONCLUSIONS

Substantiated with a statically valid argument, to answer the research question mentioned in section 3, it appears that Ingress filtering is one counter measure that can be adopted my the decision makers to minimize the DDoS attacks.as a valid countermeasure that can be implemented to all actors

Network operators, the country, and victims to mitigate the security issue. Obviously, there are positive and negative incentives; to each actor in the chain who can implement the countermeasure in the network nodes, and the promise of returns on investment can inspire actions.

Ingress traffic filtering at the periphery of Internet connected networks will reduce the effectiveness of source address spoofing denial of service attacks. Internet Providers and corporate network administrators implement ingress filtering, the opportunity for an attacker to use forged source addresses, as an attack methodology will significantly lessen. Tracking the source of an attack is simplified when the source is more likely to be "valid." [8]

From the statistical analysis given above in section 5, to explore the impact of these factors on the metric, we can verify the hypothesis (section 3) and make a conclusion. It is clear that as a result of the SAV ingress filtering, there will be significant improvement in security of the systems against DDoS; Hence, ***Hypothesis H0 is true***.

8. REFERENCES

[1] History of Internet, https://en.wikipedia.org/wiki/History_of_the_Internet.

[2] “6.4 Billion Connected "Things" Will Be in Use in 2016” http://www.gartner.com/newsroom/id/3165317

[3] "SAVE: source address validity enforcement protocol" 07 November 2002: http://ieeexplore.ieee.org/document/1019407/?arnumber=1019407

[4] Spoofer Project:https://www.caida.org/projects/spoofer/

[5] "Understanding the Efficacy of Deployed Internet Source Address Validation Filtering" : ACM, Chicago, Illinois, USA — November 04 - 06, 2009

[6] ACM Conference: http://dl.acm.org/

[7] Robert Gezelter (1995) Security on the Internet Chapter 23 in Hutt, Bosworth, and Hoytt (1995) "Computer Security Handbook, Third Edition", Wiley, section 23.6(b), pp 23-12, et seq.

[8] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", RFC 2267,DOI 10.17487/RFC2267, January 1998, http://www.rfc-editor.org/info/rfc2267

[9] The Spoofer Project: Inferring the Extent of Source Address Filtering on the Internet (SRUTI '05): http://static.usenix.org/legacy/events/sruti05/tech/full_papers/beverly/beverly_html/

[10] “MIDAS: An Impact Scale for DDoS attacks” - Rangarajan Vasudevan Z. Morley Mao Oliver Spatscheck Jacobus Van der Merwe: web.eecs.umich.edu/~zmao/Papers/midas.pdf : EECS

[11] “DDoS attacks: The impact”: http://www.itp.net/591175-ddos-attacks-the-impact.