SHOWERTHOUGHTS: DDOSING AN IMPORTANT SOCIAL INSTITUTION (AND FIXING IT). PART2

So. In light of recent events (read part1) there is clearly a gap about what people think they know, and what they actually know about DDOS-attacks and how to take care of them.

First of all we need to do a reality check on what a DDOS means today, since it has slightly changed since we first got to know the phrase somewhere in the dawn of the Internet. To create a DDOS today you need three essential item.

1. **The Buyer:** A potential perpetrator that wants to cause harm is usually not the one that creates the actual bits and bytes to DDOS someone with. The perpetrator takes the easy way out, acquire a bag of money and hand it over to someone, preferably as bitcoins or other unregulated currency to a potential seller that you usually find in your not-so-regular webshop that's anonymized and hidden somewhere. Usually referenced in populistic media as ”the darknet”, while in reality it's somewhere in a overlay network.
2. **The Seller:** This is usually the one that has accumulated a botnet. Usually the seller has created (or bought) a malware or a virus that's been spread around the globe through any common form of transport, i.e email, trojans and such. Gathering computers to the botnet, usually trying to keep the infected host as incognito as possible to make sure whenever a potential buyer comes in with a big order, the botnet is ready to go and not spotted beforehand. These virgin botnets is usually the most expensive ones, but also the most effective.
3. **The Operator:** Last piece of the puzzle. Depending on what type of attack it is, the operator is more or less guilty as well. And in this case im referring to the operator that is closest to the participating computers in the botnet. If we are dealing with an amplification-attack where the botnet asks legitimate resources on the Internet, such as NTP-server, about certain type of data but uses a spoofed-source (the ddos-target) the network at hand is not complying to BCP38 and should fix their stuff after getting publicly shamed. The operator is also responsible to taking abuse seriously and should make an effort doing as much as possible to make sure the operator's subscribers is being dealt with accordingly when mis-using the network's resources. This usually works quite well in countries that has been working with Internet for a very long time and not so well in countries where Internetworking is fairly new, and also where the growth has been substantial. See China, Russia, South-America etc.

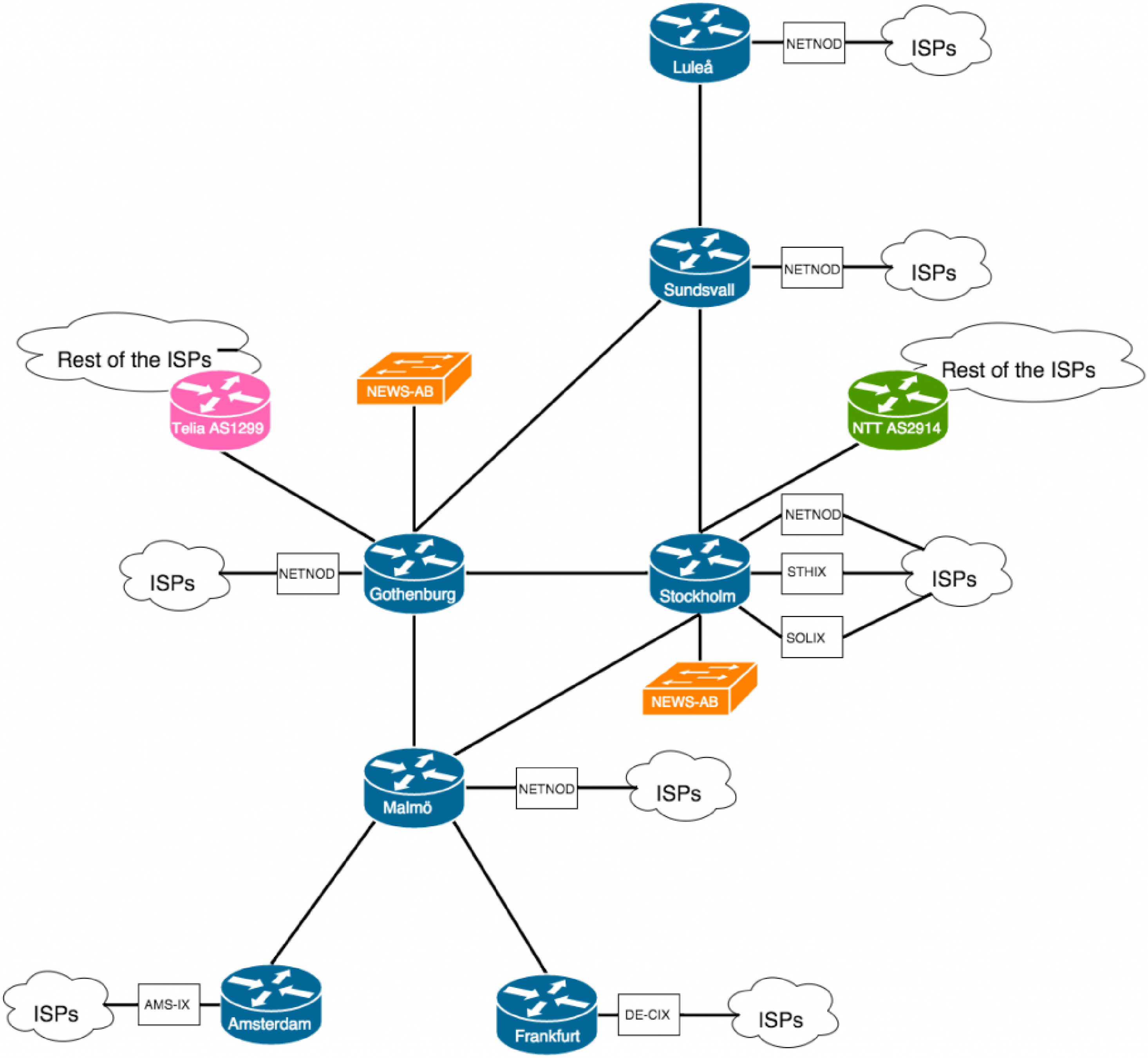
This is the standard framework, and as with everything there is a lot of exceptions to this but if looking at this fairly standard chain of events we can clearly see that when media cries wolf that the threat is coming from the east, it does not mean anything. Yes, the bits and bytes could be generated in the far-east but the seller of the product could be in Spain for all we know and the buyer could be a disgruntled teenager from Sveg in Härjedalen. Or the seller could be an oppressive regime and the buyer could be another oppressive regimes three-letter agency, depending on how thick the tinfoilhat is. We have no idea, and we probably never will in most cases.

Anyway. We know the DDOS can come from anywhere, at anytime and in any size, shape, color or form and be initated by anyone. How to prepare for this? time to talk dirty and get technical.

First lets set the mood here. Meet the fictional company of Den Schwenska IPbolaget AB, IP-AB. AS1337. One of the fictional customers to IP-AB is Den Schwenska Newstidningen AB, NEWS-AB.

IP-AB is a medium-sized ISP in Sweden that serves residential and business-customer as well as doing hosting for companies in their datacenters. They do local peering in the Swedish IXP's Netnod Stockholm/Gothenburg/Sundsvall/Luleå, COMIX, STHIX, SOLIX and they have also invested capacity down south to connect to the biggest IXPs in the world, AMS-IX in Amsterdam and DE-CIX in Frankfurt. IP-AB believes that local-peering will not only save money of the local peering but they will also increase diversity and increase quality due to getting shorter bit-miles to the customer. They will also lower the IP-AB bill with their two IP transit providers, NTT AS2914 and TeliaSonera AS1299 if they can exchange the traffic settlement-free with networks connected to the aforementioned IXP's.

A pretty classical setup that could probably be applied to most operators that might be interested in this in any country of the world, change the name of IXPs and and cities and its applicable anywhere.

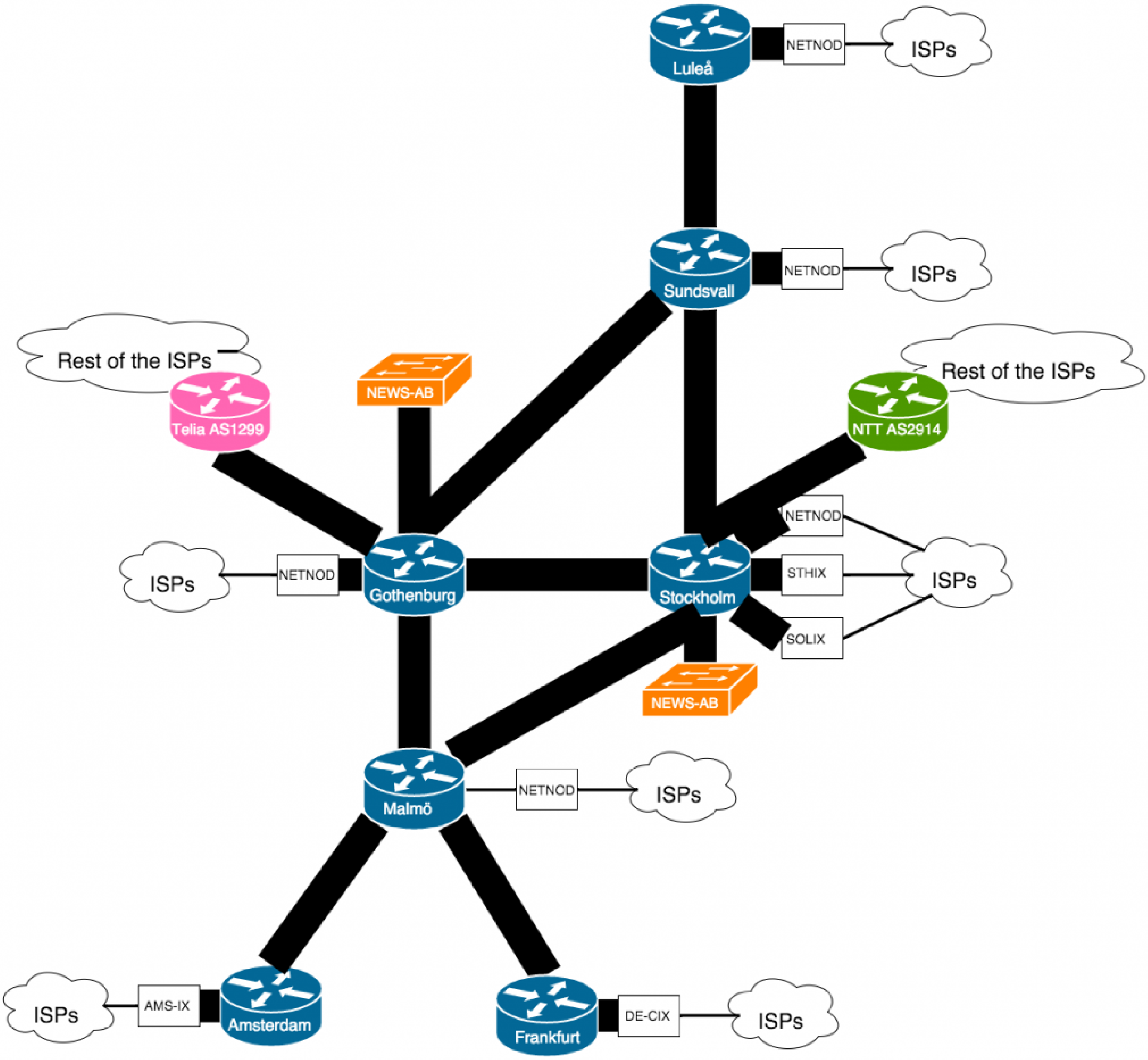


At any given time this operator could probably offload 60% of the traffic through the connected IXPs. Since this is modeled as a Swedish operator we can assume that most of the ”core business” is handled over the IXP´s. National-to-national flows of traffic is usually the most important and revenue-generating flows and what you usually care the most about. Hungry locals to the local pizza business , Swedish webshop to Swedish customers, Swedish newspaper to Swedish readers, etc. Here we are also assuming that most of streaming and content is also handled primarily through the IXPs, Netflix, Youtube, Akamai, Spotify and the such are usually present as close as customer as possible to lower costs and increase quality. The traffic that will not be handled over the IXP is traffic going or coming outside of this region. There is just a small handful of South-American networks present at these exchanges for example, so that traffic will need to hit the payed IP-transit eventually since IP-AB does not want to develop business in Brazil, in this case IP-transit is modeled with TSIC AS1299 or NTT AS2914 which is two global carrier-networks that amongst a lot of others can solve your connectivity needs.

This network is afraid of DDOS. What do we do? First let us walk through the bread & butter

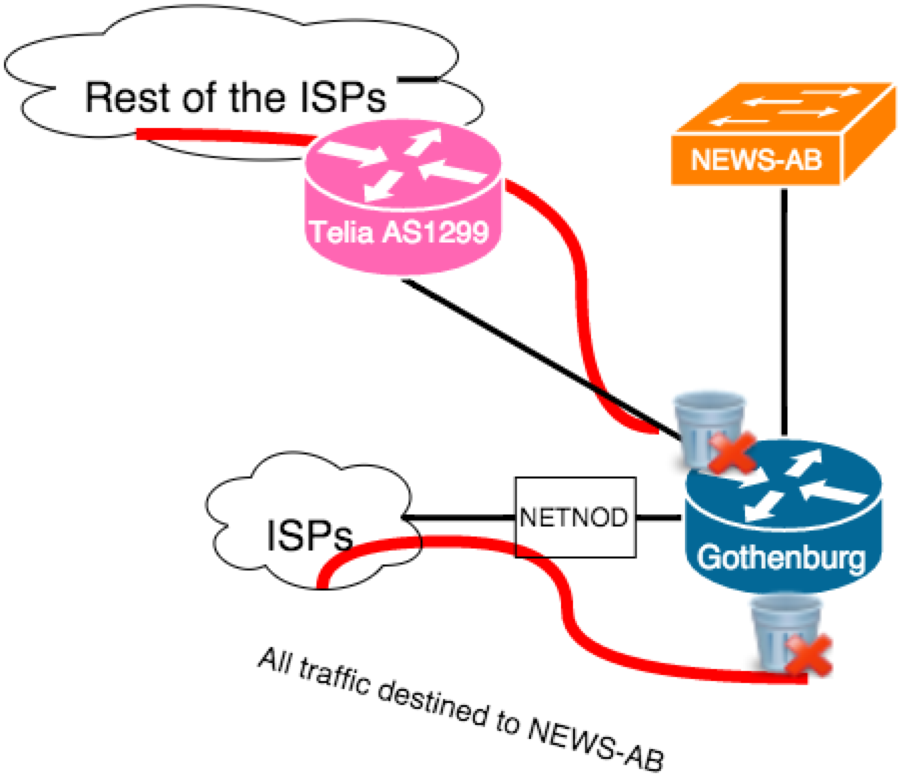
Solution 1. Beef up the network

Build a network that can withstand any type of attack by adding capacity and making the links so big that congestion will never occur is a perfectly fine way of solving the problem, but also a great way of going bankrupt. Unless you happen to be one of the biggest networks in the world this is probably not a feasible strategy even though making sure your network has sufficient capacity is important. But ridiculous overprovisioning is not a feasible way forward really.



Solution 2: Blackhole the prefix Internally.

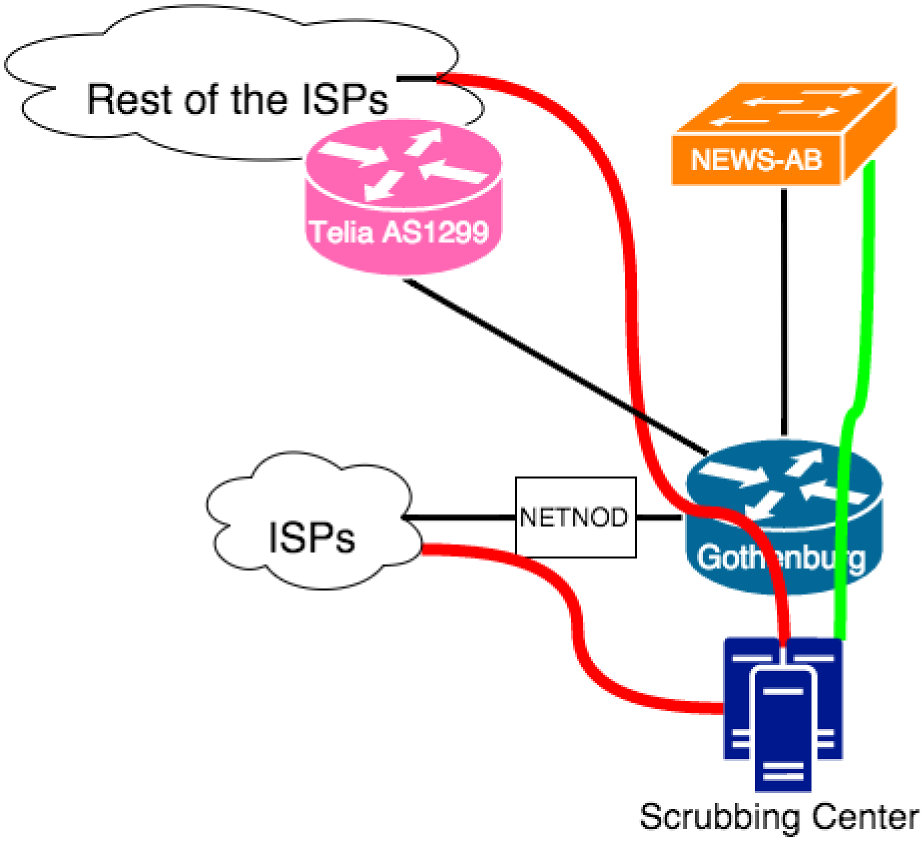
So the prefix of NEWS-AB is under attack, one way of solving is to route traffic to that destination to nothing. Every router sends the traffic to discard instead of sending it to the actual destination. Destination-based blackholing. This can be done with static configurations or a RTBH type of mechanism. This makes the customer unreachable since IP-AB discards all traffic destined for NEWS-AB as close as the originator as possible in IP-ABs network, in this case in all the edge-routers connecting to peers and exchanges. This protects the NEWS-AB internal infrastructure and IP-ABs underlaying infrastructure but the unwanted traffic is still entering the network, possibly causing congestion in the interconnect-points between IP-AB and the IXP and IP-AB and the two IP-transits. NEWS-AB is not happy either since their main product (the news website) is not reachable at all and from the customers point of view this is most likely equally bad as being ddosed, NEWS-AB can not collect revenue.



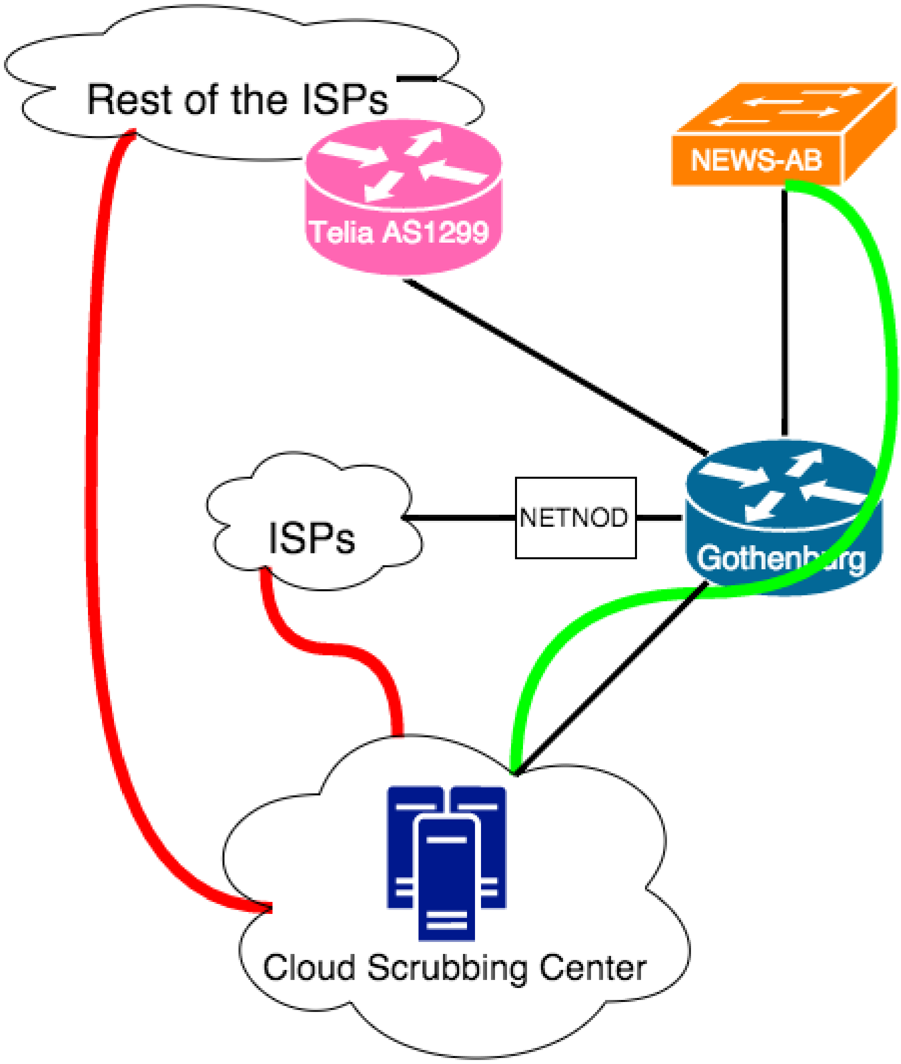
There is also versions of this where you are allowed to export your blackhole to the paid transits and they drop traffic to the destination of your choice before it enters your network, freeing up capacity and making the IP-transit ports unclogged again. But your IXP-ports is still fully congested since settlement-free peers usually does not accept blackhole announcements. And during all this the customer is not getting the revenue-traffic in so NEWS-AB = still not happy.

Solution 3: Scrub the traffic and clean it.

If there is many ways to skin a cat there is definitely more ways to scrub a flow of traffic. One common way of doing this is have a appliance in your network that do flow-analysis using NETFLOW/SFLOW from the edge-routers and try to find anomalies. When it finds anomalies it will through BGP send out advertisements internally in IP-ABs network to attract traffic to a certain destination that will have the traffic go through a scrubbing-center. Dirty-traffic comes in one ingress-side and the scrubbing-center takes away unwanted traffic and on the egress-side comes the clean traffic with people clicking on ads and reading articles which is exactly what NEWS-AB wants. These type of solutions is (usually) not for free. They require not only a vast amount of hardware to be able to analyze traffic in real-time but the cost of someone building a business and technical intelligence on what good or bad traffic can look like is not to be taken lightly, the whole business of ddos-mitigation intelligence is a multi-hundred-billion dollar industry.



A recent modified approach to this methodology is to have the scrubbing-center ”in the cloud”. How this essentially work is the same, through BGP you alter the path to the destination network and in this case you announce to the ”cloud-ddos-shop” that NEWS-AB is only reachable if going through the ”cloud-ddos-shops” AS-number before coming into IP-AB, essentially working as a third IP-transit provider in this case. What these people do is that they not only buy truckloads of servers to scrub the traffic, but they buy extremely big pipes to the big carriers in the world, to be able to receive flows in the ”big-ddos” range, which can be up to 600Gbps. The biggest shops brags about their total ingress-capacity adding up to about 3Tbps which may or may not be true. The idea here is that they front the traffic for you, let their serverfarm act as a proxy and just send the clean traffic backwards into IP-ABs network, hopefully clean as a whistle.



While this sounds all good there is two problems here.

1. Your giving away potentially very sensitive business-intelligence to an off-site company, they will see all of NEWS-ABs customers and what they do. Its their job to look inside the data destined for you and it's also their job to profile your flows so they can make fair assumptions on what an average day look like so they can mitigate whenever it's not a normal day. This may or may not be a problem integritywise depending on what type of business you run.
2. This is by far one of the more expensive solutions, since you will enjoy paying both for the big serverfarms and the big pipes of transit to the tier1 carriers to make sure all traffic reaches the scrubbing center.

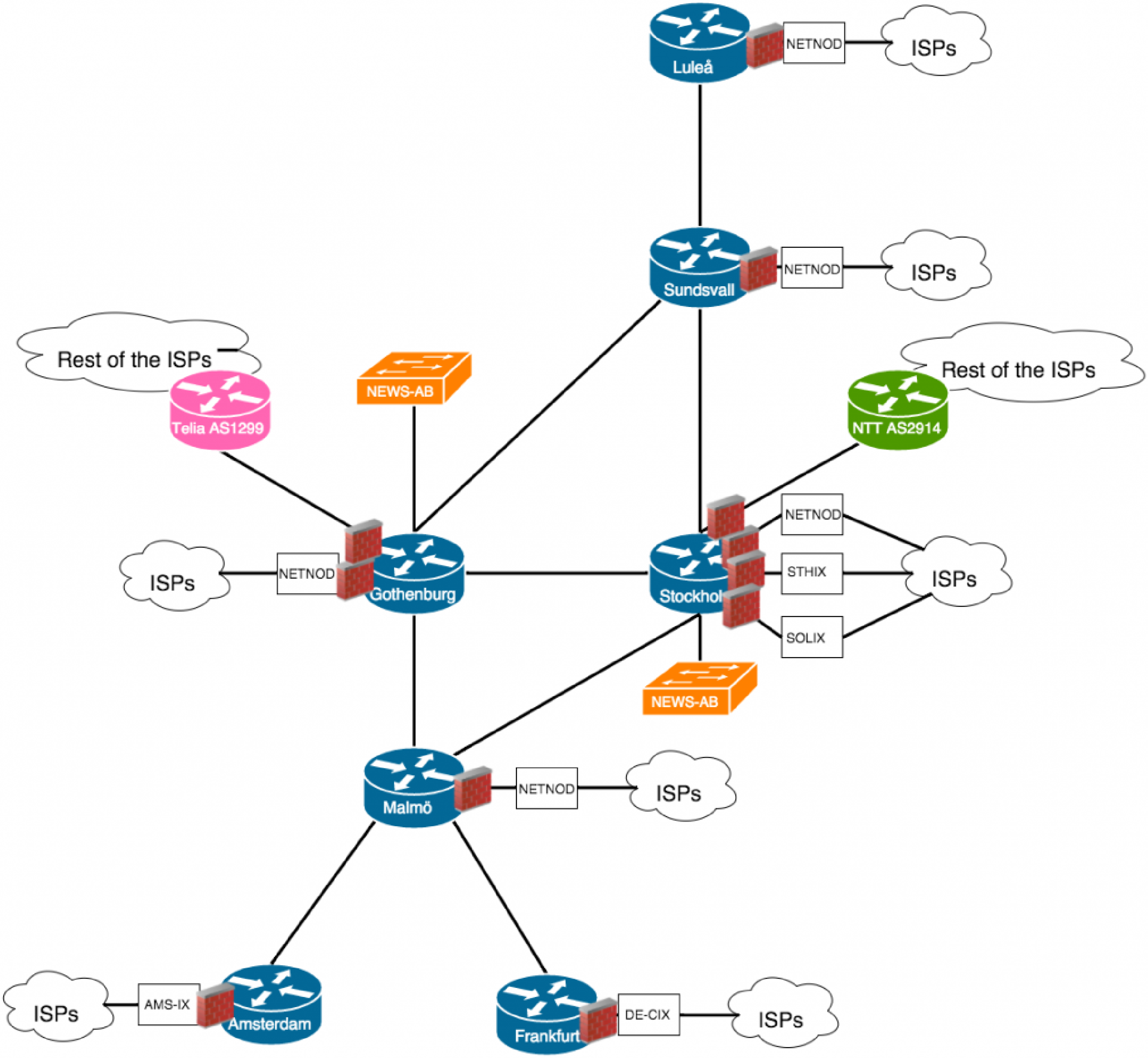
It should be noted that scrubbing-centers is usually something you need to pay for as stated above, its both resource intensive so you need a lot of hardware and the intelligence behind whats DDOS and whats not is something that companys has spent a fair share of R&D money into and need to get ROI on.

But what if all operators that has this problem (that is all operators in the world essentially) would collaborate and be open and tell the story like it is, provide profiling and intelligence on what's attacking them and what does it look like? This would probably be a better R&D department then whatever we currently pay for coming from Burlington in Massachusetts today.

Solution 4: Edge filtering

Also a common method to ease the problem. At the operator's edge-ports various amount of filtering and rate-limiting will be applied. A common example is to rate-limit certain type of protocols that is used very often in DDOS-attacks. Chargen and SSDP is common UDP-based protocols that has no real use on the Internet anymore but is easy exploitable and vulnerable to amplification-attacks. An operator can without much problem apply a static rate-limit on SSDP to lets say 5% of the the edgeports total capacity, if it trips that limit SSDP will just be dropped, and since no one missed it to begin with… no harm was done.

This is mainly for static configurations and general housekeeping. The last two years UDP Amplification attacks using NTP, DNS, SSDP and Chargen primarily has been very successful, so my guess is that a-lot of operators already has these type of static rate-limiters in place already. They might not talk about it but it's a guesswork (we do it in our network for example)



So how about we make this smarter? What if there were dynamic filtering? This is where BGP Flowspec comes into play. There has been a-lot of fuzz around BGP Flowspec what it can or cannot do and CISOs around the world using big words that flowspec is mandatory without really knowing what flowspec does for you. What it can do is to dynamically and automatically send, propagate, and delete edge-filters across IP-ABs network using the already established protocol BGP. While this is all fine and handy it's no good if no one tells Flowspec what to do. You need a analyzer, and this is where it can get tricky.

This analyzer is most commonly deployed by collecting sampled netflow-data from your network and having the data crunched centrally somewhere in a server and look for anomalies. Whenever a anomaly is detected in the flow-analyzer it will with the help of a new address-family in BGP announce a flow-rule, the mockup hypothetically looking something like this...

”rule 101 source 1.2.3.4 apply rate-limit 1Mbps to destination prefix-set NEWS-AB”

or

”rule 102 source 1.2.3.4 discard protocol UDP to destination prefix-set NEWS-AB”

This will then be sent out and applied to all ports where this traffic can be seen through and the accumulated amount of traffic if the perpetrator achieves a perfect split is just a few Mbps at best since each port will have its own set of rules. Hopefully saving the end-customer in this case. We are still not protected for when the DDOS is bigger than what the actual port is, since then its no use of filtering the traffic internally in the IP-AB network if you can't even take the traffic to begin with.

What we do here is to create regular stateless filter and they are therefore by definition dumb. There are ways to trick them and they are not very accurate, but it's not really about surgical precision here, we want to take the bulkload away so whatever trickles down to the application can be sorted out there.

There is companies out there that will happily sell you this analyzer, but seeing as i don't like paying for things let's talk about how we can do this without paying. First out is FastNetMon.

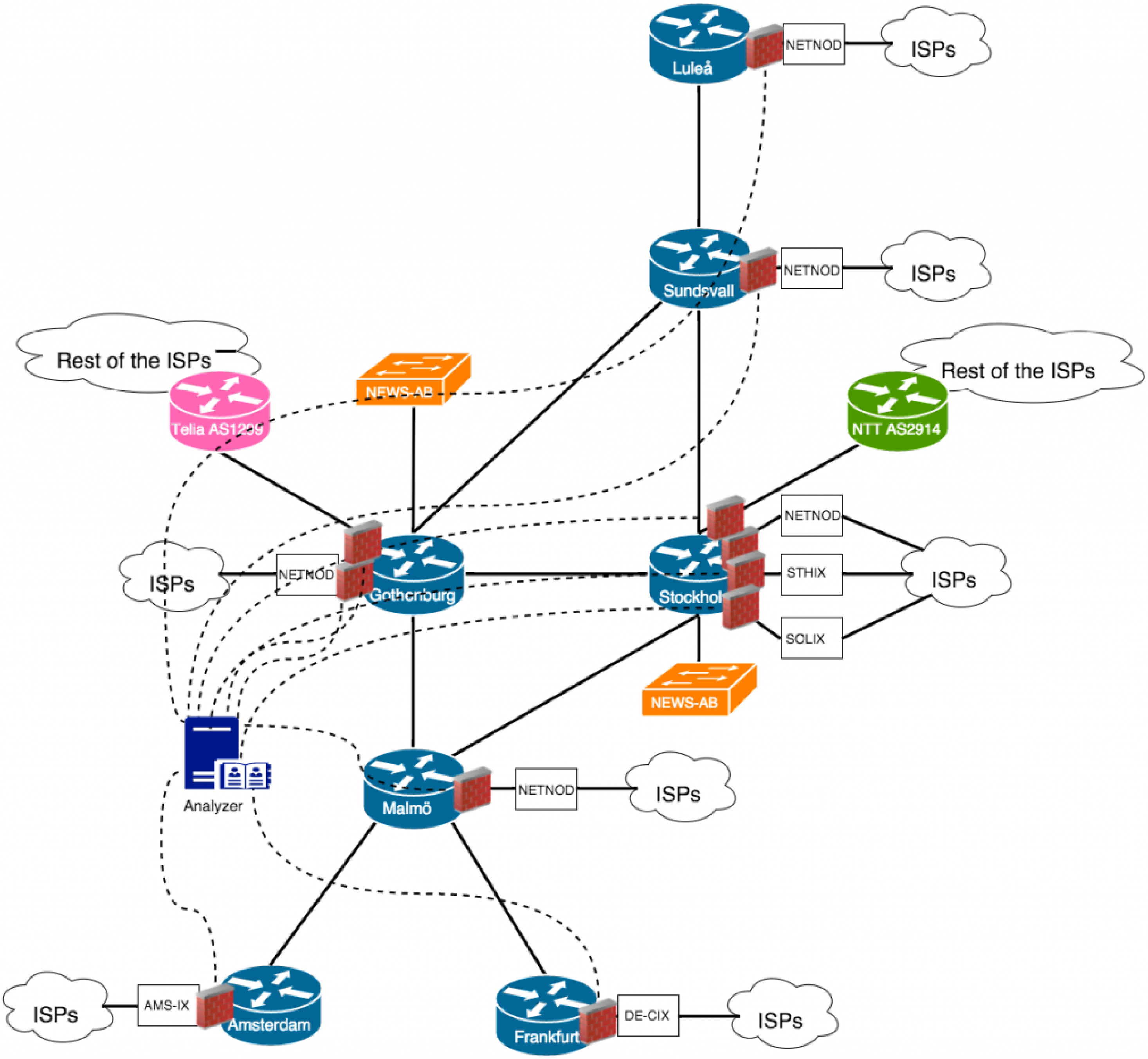
https://github.com/pavel-odintsov/fastnetmon

Fastnetmon is one of the new and shiniest stars in the FOSS-galaxy, the reception has been very good since it is what a-lot of people has been waiting for. A tool you can feed netflow-data of your choice from the network and the application will look in netflow trying to find anomalies. When it does, FastNetMon will produce flowspec-rules and sent out in the network. If filtering is not good enough, it can also trigger a various assortment of blackholes or inject new routes into the network steering the traffic away from its regular path and perhaps have the traffic enter a scrubbing-center. Unfortunately it can't be a scrubbing-center yet but my guess here is that it's just a (short) matter of time until someone figures that one out, the technology is essentially already there with SnabbSwitch.

ENOG 9: FastNetMon - Open Source Toolkit for DDo…

If someone reads this – that happen to be building a scrubbing-center on SnabbSwitch (or similar) i will happy provide my network(s) as testing-grounds for this and provide any type of lab-equipment one would typically need.

Fastnetmon is intended to be fully automatic, when its up and running it's your job to just make sure that it does its things correctly and forget it's there.



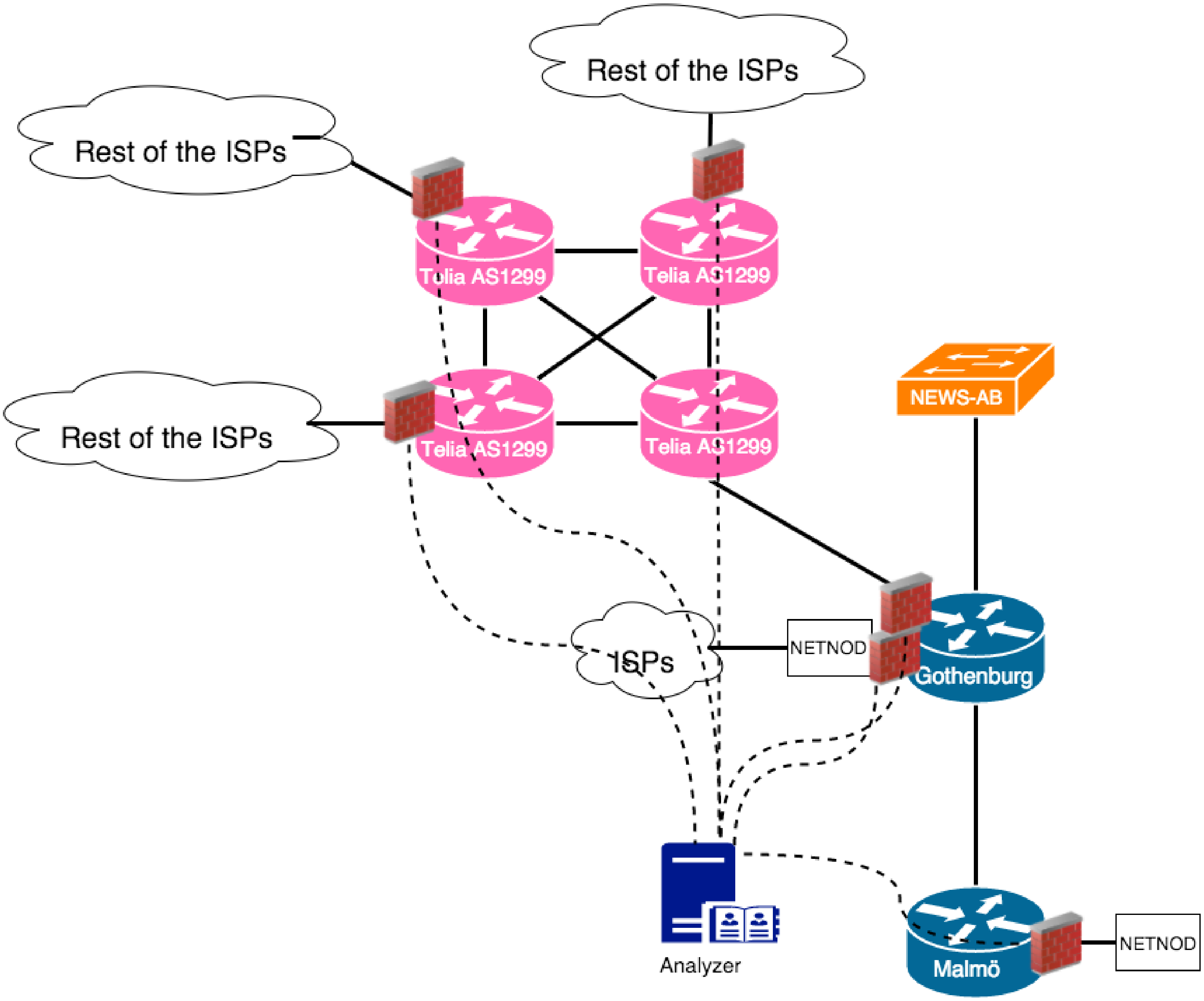
Another approach to this is to not automate it but instead do an abstraction that both IP-AB and NEWS-AB could use and control themselves. If NEWS-AB believes they do a better job and knows better that type of traffic they do not want reach their network, and want IP-AB to drop it as far from NEWS-AB as possible, this it not a bad way. Lets look at the project Flowspy…

https://flowspy.readthedocs.org/en/latest/
https://github.com/grnet/flowspy

This is project made by the GRNET, the R&E Network of Greece (academic fistbump!) and is usually called ”firewall on demand” when productified. What it essentially does is that it provides a user-friendly interface to both the operator and its customer to create rules by themselves which will then be applied throughout the network with Flowspec. What is neat here is that the customer/user decides its own fate, the customer will only be able to create rules concerning their own networks and if they fuck it up, it didn’t affect anyone else. Its also possible to feed Flowspy through the REST API if you still want to automate your fate, which you probably want.

This project is run at big-scale in the european R&E network of GEANT (AS21320) so that every participating country can create whatever acl-rules they feel for.

Alright, let’s turn it up to 11. Lets propagate this even further! Inter-domain BGP Flowspec to the rescue!



Now i will propagate the rules i have created in my Flowspec-device to my paid transit provider, in this case TeliaSonera, either automatically through the likes of FastNetMon or by hand using tools such as Flowspy (or just actually by hand, in a router directly). This looking freaking awesome right? putting a ACLs that says ”please no chinese traffic right now” all the way in Beijing in my transits global network instead of at the border to my transit-network, saving the transit-ISP a-lot of money since he doesn't need to carry crap in the backbone and IP-AB does not need to have the unwanted traffic enter the network. This unfortunately has not catched on at all, i don't know of anyone that does inter-domain flowspec successfully today since there a lot of security concerns that no one have figured out yet.

The operator accepting flowspec rules from a downstream network has a few caveats to consider. How do you make sure your infrastructure have enough computational capacity to make these firewall-filters? Eventough most of the modern platforms that a upstream-provider would typically use has plenty of capacity to drop a vast amount of filtered traffic directly in the dataplane is not endless. What if IP-AB announces 100 000 firewallrules (per second) to the upstream-network which then need to propagate it through its global network, can the infrastructure actually cope with this? I'm not sure. Or what if IP-AB makes a typo, or gets hacked, or a bug occurs in their automatic flowanalyzer-thing that suddenly decides to start announcing rules to AS1299 that the webpage of AS1299 should not be reachable from anyone anywhere. It's' a quite big task for AS1299 in this case to make sure that sanity is kept intact and as far as i know, the flowspec-protocol itself does not have methods for controlling this.

This is one of the primary reasons that inter-domain flowspec has not catched on unfortunately. While it's still a concept that is fairly new i'm confident to say this will probably catch up, soon enough. However the concept will probably never be accepted amongst settlement-free bilateral peers, and if the bulkload of unwanted traffic comes in from these peers, we need to make something smarter, something that is not relying on an opt-in/out type of protocol such as flowspec.

Solution 5: Selective Blackholing / Control Communites (with a distance vector)

This is a very broad subject so lets try to stick to the very ends of distance control communites. We can start with looking at NORDUnet where we have the bare minimum.

Control Communites

2603:664 Do not advertise to Commodity (transit)

2603:665 Do not advertise to Commodity and non-Nordic Peerings.

2603:666 Do not advertise to Commodity and Peerings.

2603:667 Do not advertise to non-NORDUnet-member R&E.

2603:668 Only advertise to GEANT R&E sessions and nothing else

Blackhole Communities

Prefixes tagged with community 2603:999 will be blackholed. All NORDUnet edge routers will discard traffic sent towards the tagged destination.

Prefixes will also be forwarded to upstream with the corresponding blackhole community.

You can only tag your own prefixes with this community.

You can only tag more specific / longer prefixes then what NORDUnet filters on. This means that if your aggregate for example consists of a /16 to NORDUnet, the maximum prefix length you can blackhole is a /17.
This constraint was put in place to avoid mayhem if someone accidentally tags their aggregate with the blackhole community.

Prefixes tagged with community 2603:998 will be blackholed outside of the Nordics. Traffic will still flow in routers in Iceland, Denmark, Norway, Sweden and Finland. Otherwise, the same constraints apply.

The downstream networks of NORDUnet has a few selections to do when tagging prefixes here. Lets presume IP-AB would have this exact same ruleset and is currently noticing a big influx of DDOS-traffic coming in from IP-Transit, NTT and TeliaSonera hitting NEWS-AB. What either IP-AB or NEWS-AB could do is to with BGP tag the prefix with 1337:664 (2603:664 but with the correct asn). This will tell the routers of IP-AB to not announce NEWS-ABs tagged IP-adresses to the payed transit, effectively cutting of complete reachability to NEWS-AB from anything that's not a direct peer with IP-AB. If this is not enough then you can hit harder and tag with 1337:665 which will kill non-national peers as well and if things is really hairy, you can tag 1337:666 which will withdraw the prefix from anyone that's not a direct customer to IP-AB. Now NEWS-AB is not happy anymore, while the infrastructure was partly saved no-one can reach the NEWS-AB websites and generate revenue anymore except the small portions of IP-AB customers that still sees the route.

In this case for IP-AB (and for NORDUnet) tagging a prefix with 1337:998 will most likely not solve the problem while it feels like it would have been a good scenario, keep full visibility in the core-markets but not reachable from the evil botnets from outside. The problem here is that both IP-AB and NORDUnet peers with for example the russian networks (which we learnt was a major source of botnets in part1) in the nordic region, these will still get the prefixes and problem is still not solved.

So. we need something more granular.

Job Snijders of NTT has an excellent idea on selective blackholing with a distance vector based on source of the origin. This is a more granular approach of selectively blackholing prefixes. In an essence it's about compiling a community-matrix in your network that makes it possible for any customer in the network to use communities to blackhole prefixes at X amount of kilometers away from the source of origin (the customers premises) since the theory is that your customers customers is most likely close by.

I'd suggest watching this presentation to get a good overlook what the concept is about.

This is a mostly relevant concept, but not perhaps for IP-AB. IP-AB is too small to get the full benefits of this, for a global carrier this makes perfect sense. So can we do something better? Now its concept-time!

Concept Solution 6: Selective Blackholing / Control Communites (with a revenue vector)

So we know by now that prefix withdrawals is really the only way of solving this problem in a cost-effective way that will benefit both you (since the bad traffic wont even touch your network) and it will benefit your peers and your ip-transit (since they dont need to transport you dont want in their backbone) but the problem is that there is a big risk that NEWS-AB will also be cut of from its revenue sources using traditional somewhat static blackholing.

Can we make it smarter?

In part1 i published the statistics about which networks usually attacks us. This was the 6 month average but needless to say most of these networks is present in the top50 ddos-sources on the things we get everyday, so they appear from day to day in traffic our datacruncher classifies as ”malicious ddos”. What if we can use what we already seem to know beforehand and generate communities based on this factor? If you monitor your flows and categorise traffic and sort data in a structured way you probably already know beforehand where the attack is coming. There is really no use to act surprised about ”oh damn the ddos is from china, russia, brazil and usa” since you already know that.

What im proposing as a addition to all aforementioned community-control we already have available is to provide control-communities based on a X-factor. The X-factor should be generated of a few things, the probability that the origin network will carry DDOS to your network, and this should be weighted against on how important the originators traffic is in a normal scenario.

Traffic from Google is most likely extremely important to any operator in the world, however the likelihood of google ddosing you is very very small. Hence the x-factor of google will be very low, let’s say the factor is 10. A few other networks that will get the factor 10 is other type of content-type networks such as Netflix, Akamai, Twitch, Dropbox, Microsoft etc. Next tier is the important local peers you have, a company such as IP-AB that has both residential, business and hosting type of customers is very likely to be hugely affected if similar companies in the area does not have connectivity to IP-AB, such as the company IP-AB1, IP-AB2 and IP-AB3. Hopefully your neighbour won’t DDOS you but the revenue-impact if the bigger local national networks does not have reachability to IP-ABs customer NEWS-AB is immense. These types of peer might get the factor of 20.

Then you can fill out with as many tiers of networks here as you think is viable, lets assume the magic number is about 6 tiers of DDOS-levels. The last tier, which we can give the factor of 100 i where we put our top25 month-to-month ddos-networks. This community should solve the bulkload of the customers problem by cutting out the common origins of DDOS but that has very little value to your network when under an attack. From a customer point of view the purpose of this is that you start tagging your prefixes with the highest possible x-factor value and stops whenever the traffic-levels is on a acceptable level, hopefully finding the balance between revenue and having workable services.

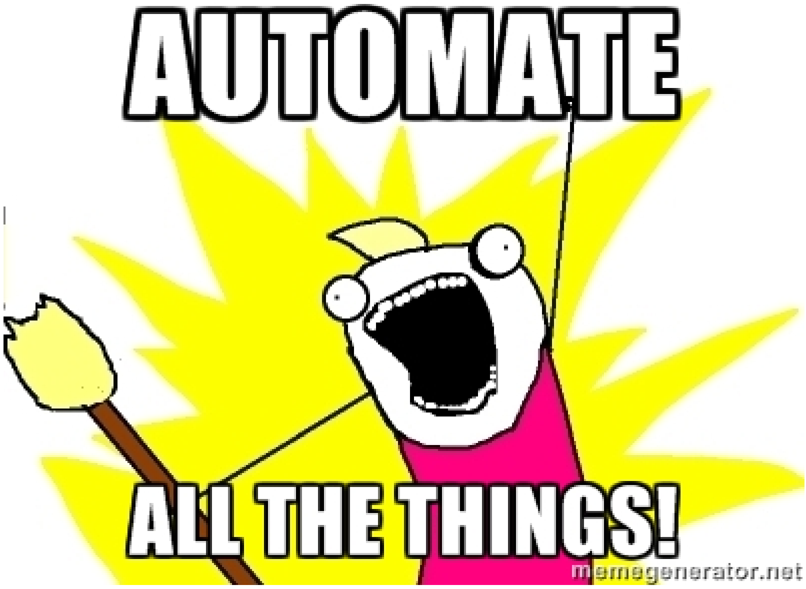
1337:1010 Blackhole towards high-value, low-risk origin
1337:1020 Blackhole towards high-value, medium risk origin
1337:1040 Blackhole towards high-value, high risk origin
1337:1060 Blackhole towards medium value, medium risk origin
1337:1080 Blackhole towards medium value, high risk origin
1337:1100 Blackhole towards low value, high risk origin

Keeping track of your netflow is essential in this type of setup. You need constant evaluation on which origins goes into which category. You also need to be aware off that these type of classifications will be public information and that may count as business intelligence. Which you may or may not want to protect, but in this case i think protecting your infrastructure is much more important than what you classify other networks as. With the current state of network-neutrality and the never-ending peering-wars this is a legitimate problem.

If information like this would not be kept secret all the time, but that we collaboratively try to figure out who the baddies of the Internet are we could as a community improve the overall health of the Internet. We already do this in the ”BGP-report” to find out networks that is de-aggregating their BGP-announcements creating unnecessary padding the global routing table.

Using control-communities instead of just shutting down BGP-peers is that the non-abusive traffic you still do with these networks will get through just fine, keeping the traffic you had with the network at hand intact which in the long run saves IP-AB money.

Alright – so with this we can handle our peers, having pre-defined classifications on a per-peer basis is really quite simple. A network such as IP-AB maybe has somewhere in the line of 400 unique peers. So what about traffic that does not come from your direct peers? In IP-ABs case most prefixes will actually be visible through the paid-transits. In this case NTT and TeliaSonera, traffic coming from far outside IP-ABs reach will need to come this way to the end-customer NEWS-AB through IP-AB. Now things gets tricky again…



So with this principal. How do we work with the transit-operators. If you choose to withdraw NEWS-ABs prefix from an assortment of direct-peers the originating networks will just receive the NEWS-ABs prefix from their transit instead, since IP-AB are still announcing the prefix to both NTT and TeliaSonera you will just move the problem to those ports instead. Possibly causing less harm than filling your IXP-ports, but it still hurting both IP-ABs business and NEWS-ABs business.

We need to find a way to propagate these type of rule upstream as well .Lets take a look on what our current choice of upstream-providers can help us with. Based on public information. This is what TeliaSonera have.

————————————————————

BGP COMMUNITY SUPPORT FOR AS1299 TRANSIT CUSTOMERS:

Community Action (default local pref 200)
———————————————
1299:50 Set local pref 50 within AS1299 (lowest possible)
1299:150 Set local pref 150 within AS1299 (equal to peer, backup)

European peers
Community Action
——— ——
1299:200x All peers Europe incl:

1299:250x Sprint/1239
1299:252x NTT/2914
1299:253x Zayo/Abovenet/6461
1299:254x FT/5511

1299:256x Level3/3356
1299:257x Verizon/702
1299:258x AT&T/2686
1299:259x Telefonica/12956
1299:261x Centurylink/Qwest/209
1299:263x TATA/6453
1299:264x DTAG/3320
1299:268x AOL/1668
1299:269x GTT/Inteliquent/3257
1299:273x Cogent/174
1299:274x Telecom Italia/6762
1299:275x Tele2/1257
1299:286x KPN/286
1299:287x China Unicom/4837
1299:288x China Telecom/4134

US peers
Community Action
——— ——
1299:500x All peers US incl:

1299:550x Sprint/1239
1299:552x NTT/2914
1299:553x Zayo/Abovenet/6461
1299:554x FT/5511
1299:556x Level3/3356
1299:557x Verizon/701
1299:558x AT&T/7018
1299:559x Telefonica/12956
1299:561x Centurylink/Qwest/209
1299:563x TATA/6453
1299:564x DTAG/3320
1299:568x AOL/1668
1299:569x GTT/Inteliquent/3257
1299:573x Cogent/174
1299:574x Telecom Italia/6762
1299:578x XO Comm/2828
1299:586x KPN/286
1299:587x China Unicom/4837
1299:588x China Telecom/4134

Asia peers
Community Action
——— ——
1299:700x All peers Asia incl:

1299:754x FT/5511
1299:758x AT&T/2687
1299:761x Centurylink/Qwest/209
1299:764x DTAG/3320
1299:769x GTT/Inteliquent/3257
1299:774x Telecom Italia/6762
1299:787x China Unicom/4837
1299:788x China Telecom/4134

Where x is number of prepends (x=0,1,2,3) or do NOT announce (x=9)

———————————————————

Alright. So there is control-communities that i can use so i can decide where 1299 is allowed to announce tagged prefixes. Seeing as TeliaSonera is a big player, they only peer with other big networks of similar size and status as themselves. Which makes sense. So what we want to look for here is the least common denominator to the ddos-origin. If the DDOS is primarily coming from the US there is a high chance that the originating networks is buying transits themselves from any of the big networks listed as US-peers. These gives us a decent chance to minimize the usefulness of the attack. Once again we rely on effectively mining our flowdata to get visibility on where traffic comes from. A good flow-analyzer will give you the full as-path on traffic it considers as DDOS. Working with the control communities we can essentially do the same as we did with IP-ABs own peers, we rank our upstreams peers the same way we rank our own peers. Where do we believe that malicious DDOS will primarily come from, based on empirical data our network has already collected? Here the distance-vector is more important due to the size of these carrier-networks, starting to blackhole as far away from IP-ABs primary revenue-streams as possible makes sense.

This is once again something that needs to be automated, and honestly it's not that hard if you have your data in place. Something needs to read your flowdata, aggregate facts and based on that you generate communities so these are well and prepared in your network for when the attack will eventually happen . There is no need for advanced artificial intelligence to make fair assumptions on which networks is evil to you. We are not aiming for surgical precision either in this case, we just want to blackhole as little as possible but still get rid of most of the DDOS. It just needs to be more accurate than ”blackhole the whole thing until it stops”.

It should be noted here that NTT offers even more TE-communities for their customers so the possibilities to optimize is essentially endless here.

http://www.us.ntt.net/support/policy/routing.cfm

If you want to play the game of toying with your transit, there is also the choice of having a dedicated port only used to announce prefixes under attack. This is the case of ”sacrificing one to protect the many”. If the network is not under attack, not a single prefix will be announced over this specific port. And when we detect through our magical systems we just described that we are under attack, the destination-prefix will get withdrawn from the regular IP-transits and just gets announced through this specific physical port. Achieving something smarter here would naturally be to tag the prefix with ”do not annonce to bad origins” and on the other port with another BGP-sessions you announce ”only announce me to the bad origins but nothing else”. Attracting the bad traffic to the small port which you don't care about, while still having good flows go over the regular IP-transit ports minimising damage.

As one can see there is a million ways of sorting these things out. My primary goal for writing this was to show that there is good ways of solving DDOS, long before it even reaches your network. And without you needing to spend a very vast amount of money. Making a good job with blackhole communities and prefix-tagging is a method that is essentially free of charge, it does cost a fair share of OPEX surely. But no hardware needs to be spent, since you already have it (your routers).

Investing time and/or money into something that takes cares of flow-statistics for you is crucial, in my opinion. This can be a FOSS-tool such as pmacct but it could also be something that's turn-key from any of the flow-intelligence vendors, there is only about 100 to choose from.

Skriven av

FREDRIK “HUGGE” KORSBÄCK

Network architect and chaosmonkey for AS1653 and AS2603. Fluent in BGP hugge@nordu.net