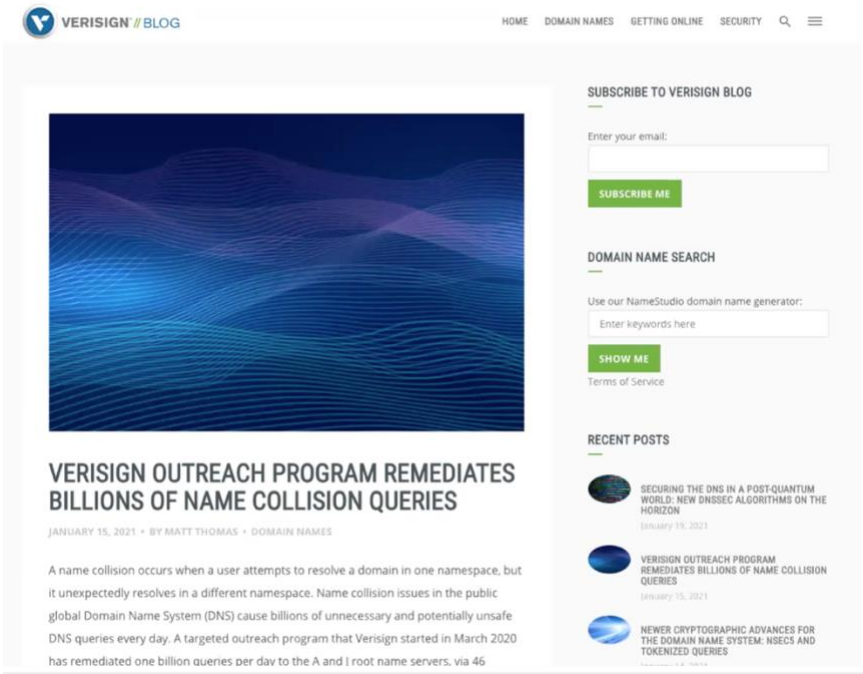
NCAP Discussion Group | 20 January 2021
Agenda:

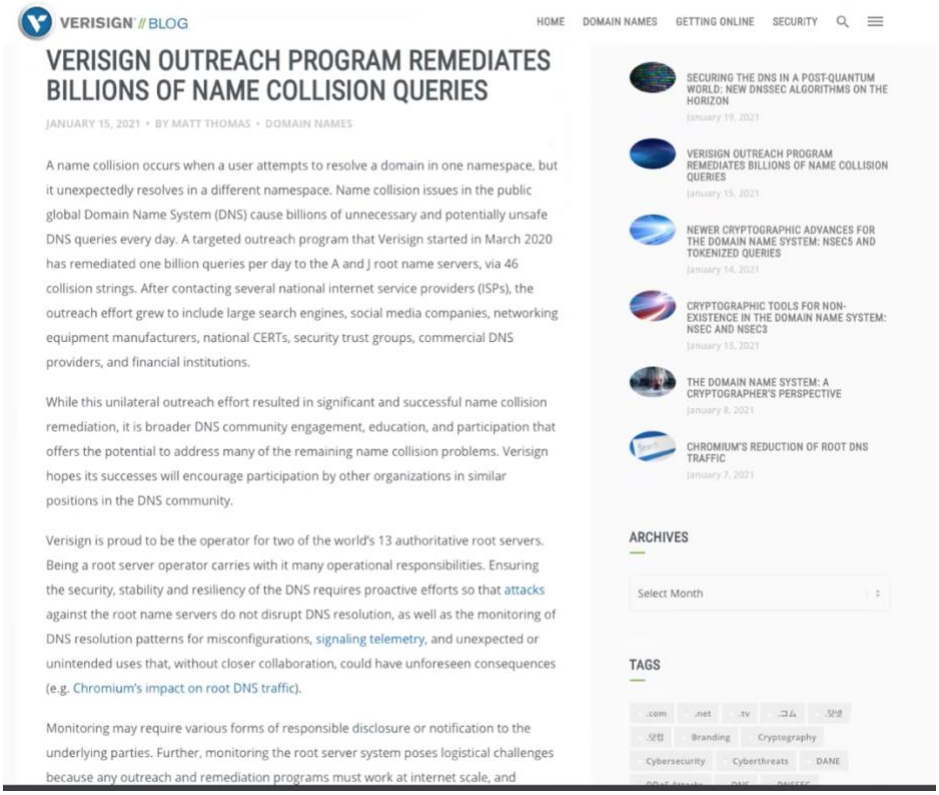
1. Welcome and roll call
2. Update to SOI
3. Update on Study 2
4. Name Collision Outreach Efforts
5. .INTERNAL Case Study Review: https://docs.google.com/presentation/d/1qJB2GuYvEKfCdGwJ7wRPfVR4Ji8yODxFS09FzvmOTsg/edit#slide=id.p
6. JAS Refresher
7. AOB

Table of Contents

Name Collision Outreach Efforts:

started inclusion outreach endeavor looking to identify and remediate high query volume strings.





Verisign: took A and J data and ranked top leaking tld strings by 2 factors: first was # of queries coming out of it; 2nd was looking at cumulative distribution function of # of ASNS for that string, look at what percentile the top 3 ASNS would be at becomes a gating mechanism to identify high query high concentrated strings of particular vendors. Look at Q names and IP addresses identify what kinds of systems were causing queries. Then did outreach to see if these entities had direct contests. Used DNS OARK. Sent emails to these people and people were open to working with us and they helped make many issues disappear.
But some strings were associated with home networking equipment. Some tied to vendors. In these cases too time consuming to fix quickly.

Common causes: Suffix search list where sting was applied in a separate search list and appended on all queries.

Global Advisors Study

Jeff Schmidt: refresher on analysis work done 5 yrs ago, on Jazz (??) horizontal and vertical study on DNS OARC data.

We did outreach to 200 firms. Meaningful engagement with 20 that resulted in improvements. Horizontal study effort to understand strings and query types across all name servers based on DNS OARK digital data. Across all queries that was in top level domain that was requested

Some interesting things that stood out in the data:

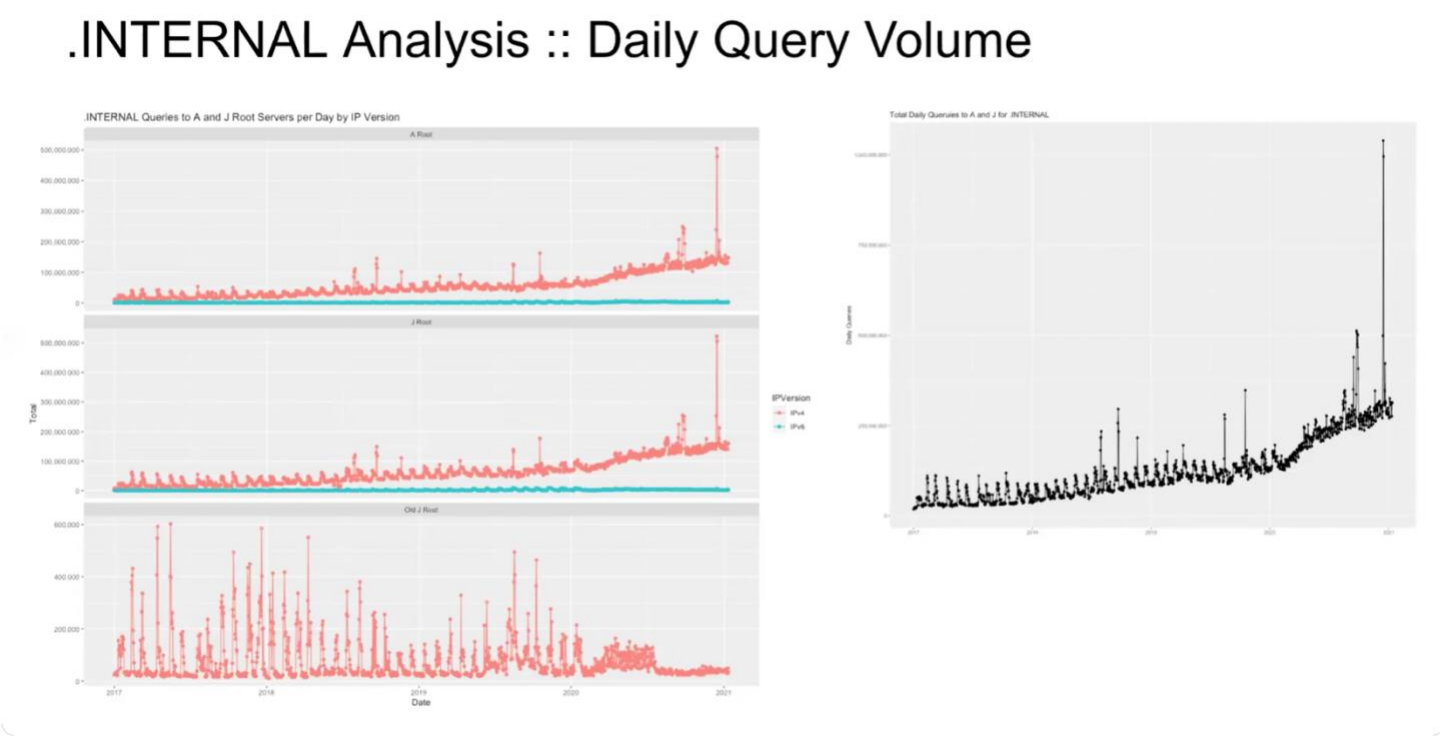
What kind of software generates a query that starts with AD Route (strings that start with AD route and have some stuff at the end accout for 1.3% of queires). Or software that generates a query that starts with compatibility additions or add ons……..this lead us to discovering that Microsoft Active Directory has a big role- its queries start with them as DCS

.internal Case Study

.internal case study

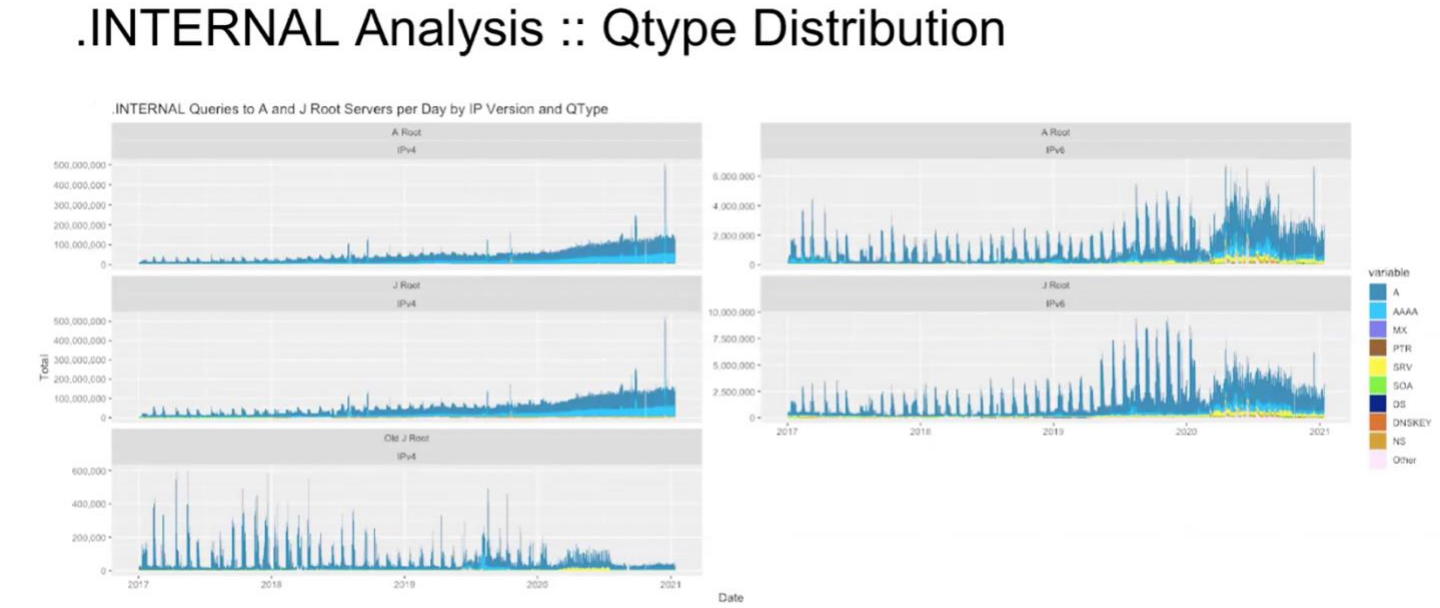
https://docs.google.com/presentation/d/1qJB2GuYvEKfCdGwJ7wRPfVR4Ji8yODxFS09FzvmOTsg/edit#slide=id.gb644be4d01_0_7

Slide 1: Daily Query Volume



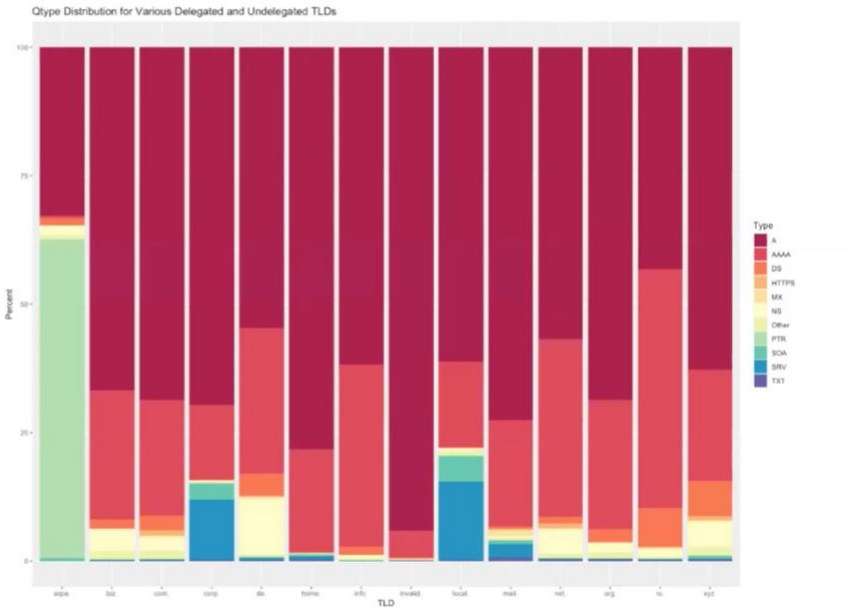
Graph on J, old J an dA route

Slide 2: Qtype Distribution



Slide 3: Qtype Distribution of Strings

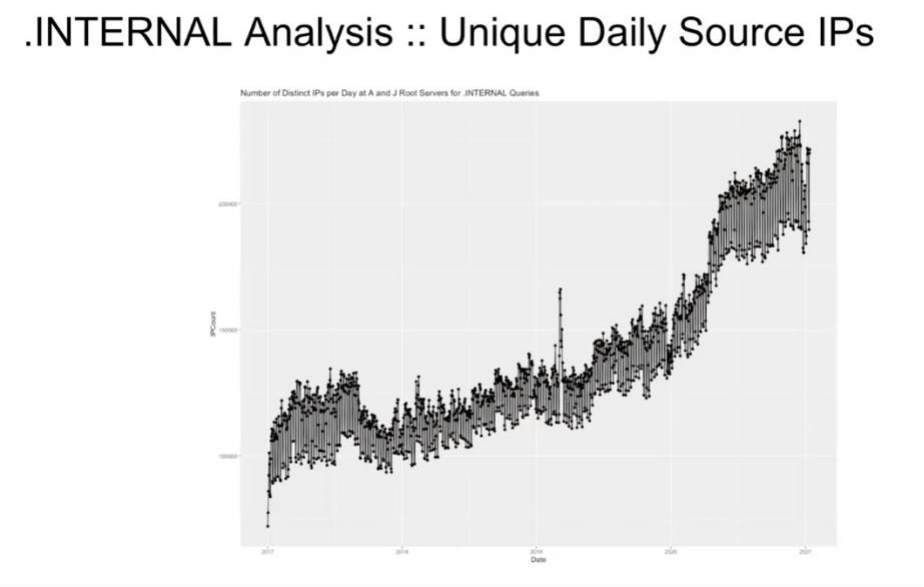
.INTERNAL Analysis :: Qtype Dist. of Strings



various delegated and non delegated strings on comparing their que types.

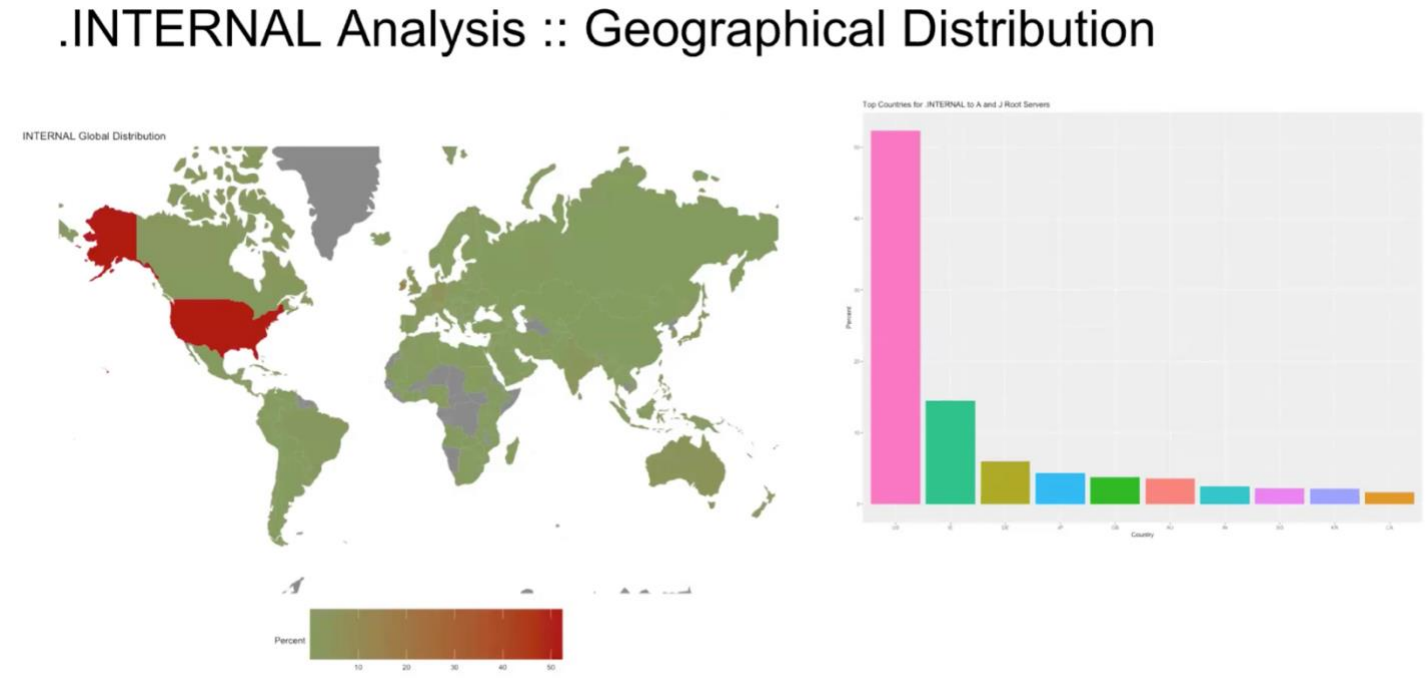
- Regular delegated seem to have a standardized mixture, but you will notice with .Corp, .home, .local and .mail as more of a Prevalence of SRT your service records, so maybe there is a little bit of a bias in some of these strings.
- Why as to why those strings are leaking in the first place. It's because they are DNS service discovery oriented type queries and just automatically getting sent out to find these new services and you have things like something search list depending to them and they're coming out.

Slide 4: Unique Daily Source IPs



Increase in distinct iP addresses - transient devices have been taken home due to quarantine, so .internal no longer resolving in enterprise resolution systems but going to residential ISPs

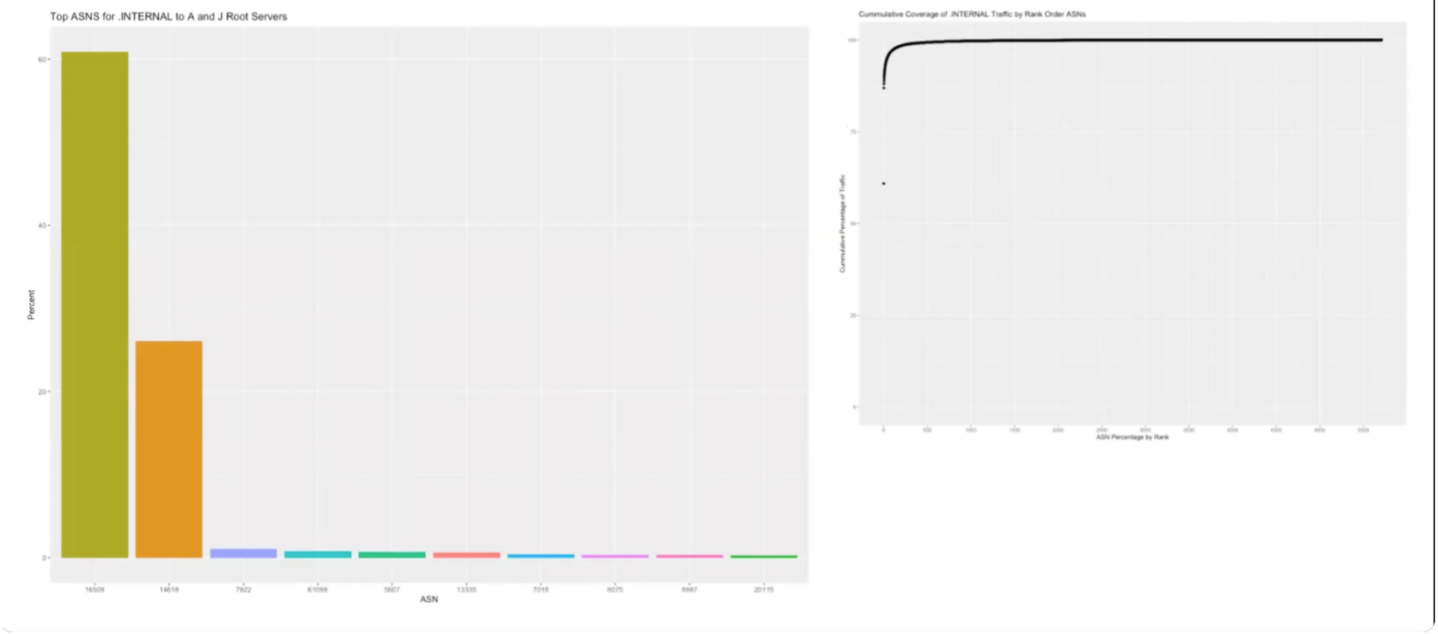
Slide 5: Geographical Distribution



US, then Ireland for .internal Different geo distribution then .mail

Slide 6: ASN Distribution

.INTERNAL Analysis :: ASN Distribution



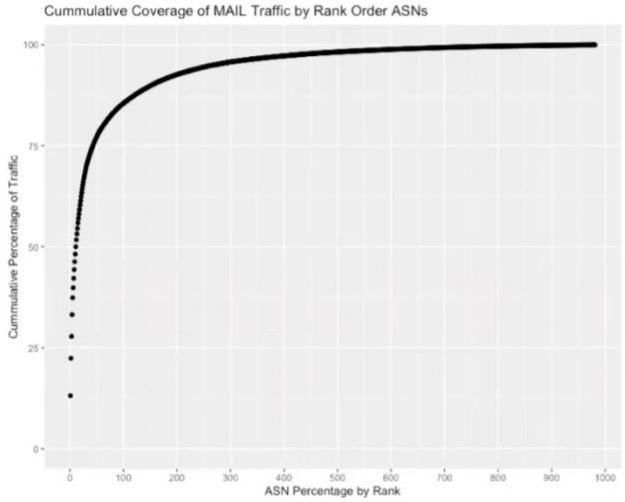
85% or 86% of all of the queries for .internal are coming from 2 autonomous systems that are actually the same company. It's just they have two different ASes. So there is one company out there that is responsible for upwards of 80 some percent of the leakage of .internal

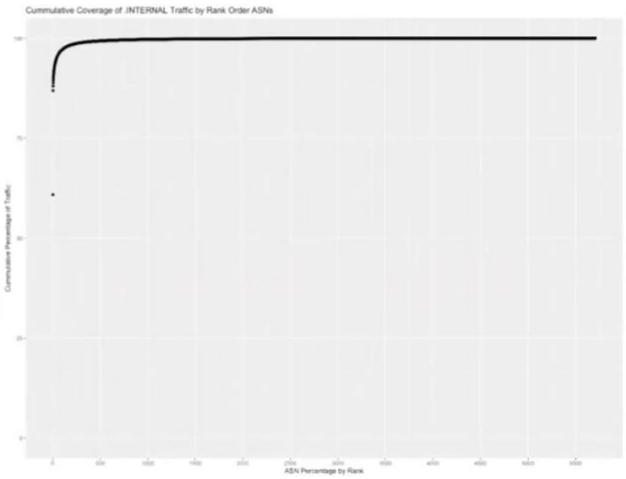
graph on the right is the cumulative distribution of those as isn't the percentage of traffic that they're leaking out.

Have been in contact with the “company” mentioned and working with them to fix it.

Slide 7: ASN Distribution .mail/.internal

.INTERNAL Analysis :: ASN Distribution MAIL/INTERNAL

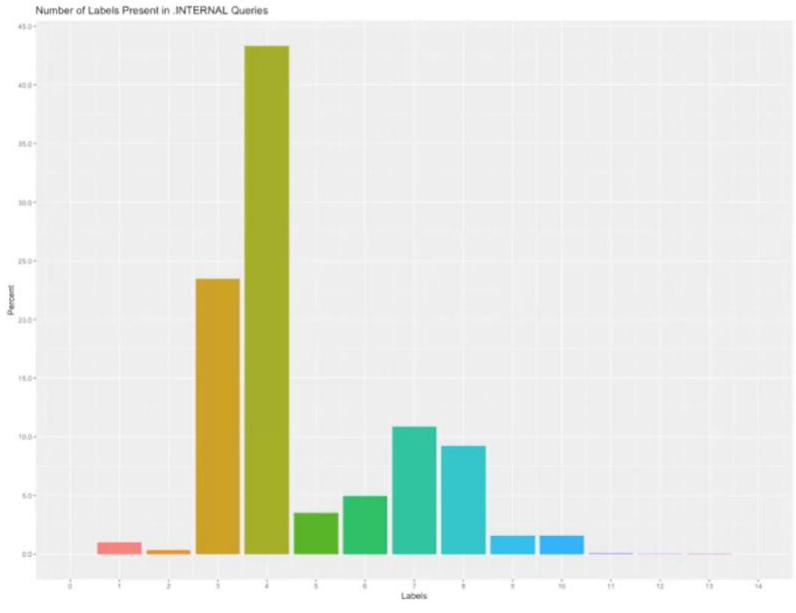




Fewer # of sources responsible for much more of the traffic

Slide 8: Label Analysis

.INTERNAL Analysis :: Label Analysis

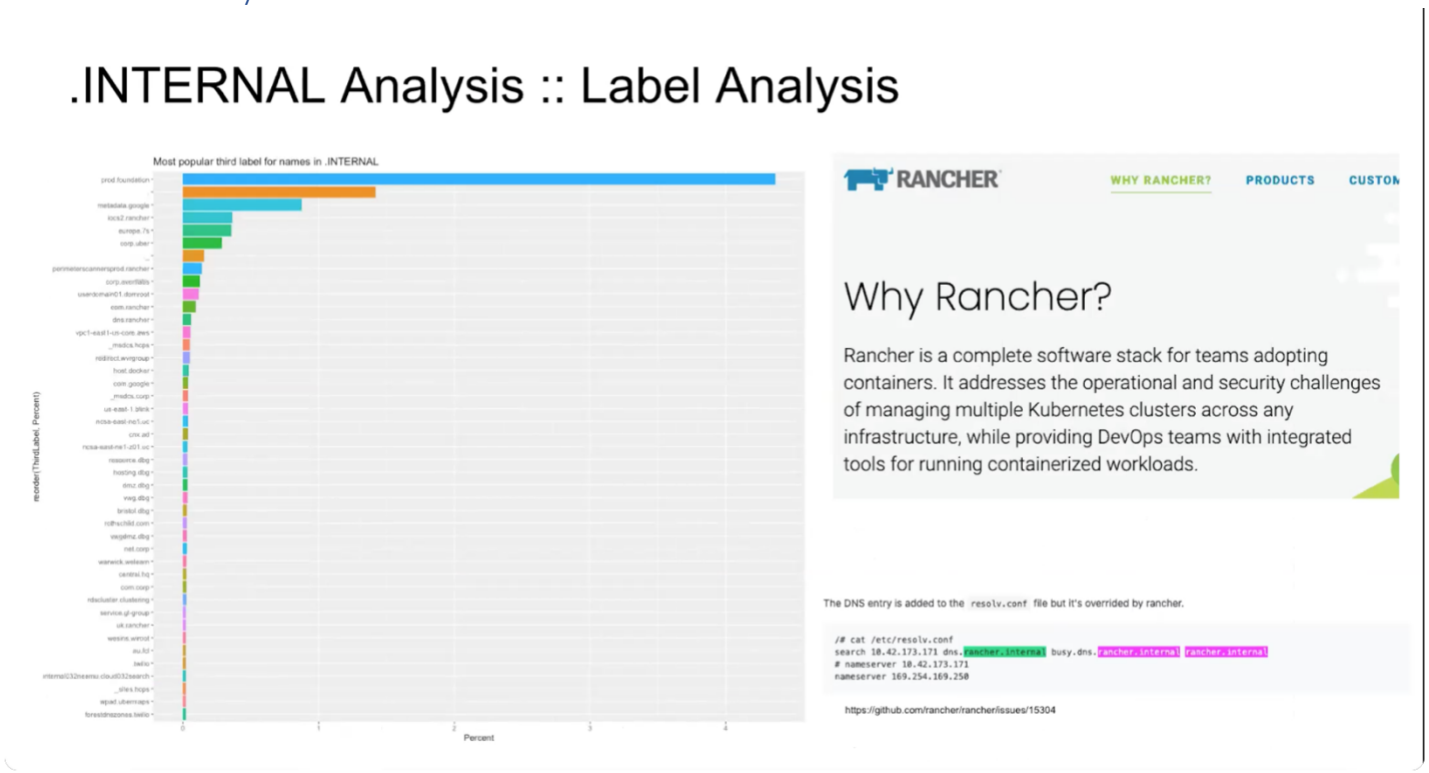


	SLD	Percent
1:	compute	64.06637157
2:	ec2	24.25419474
3:	foundation	3.29748196
4:		1.07270648
5:	compute-1	0.87439469
6:	google	0.75785944
7:	rancher	0.71297776
8:	clustering	0.55616635
9:	hcps	0.39810000
10:	corp	0.27232804
11:	7s	0.26998527
12:	uber	0.23709058
13:	ubermaps	0.18868883
14:	dbg	0.16952733
15:	ghdnet	0.15295878
16:	domroot	0.13477335
17:	_	0.11673546
18:	wvrgroup	0.10461611
19:	westada	0.10071690
20:	avertlabs	0.09441954
21:	cmw	0.06675903
22:	aws	0.06358604
23:	uc	0.06308703
24:	ad	0.06271632
25:	maxim-ic	0.05971856
26:	efi	0.05687003
27:	fcl	0.05675254
28:	wge	0.05359981
29:	domain	0.05233912
30:	blink	0.05020264

	ThirdLabel	Percent
1:	us-west-2.compute	17.1589977
2:	us-east-2.compute	16.4330502
3:	eu-west-1.compute	12.5938315
4:	eu-central-1.compute	6.3346503
5:	ap-northeast-1.compute	5.6828674
6:	prod.foundation	4.3639820
7:	ap-southeast-2.compute	3.6556092
8:	eu-west-2.compute	2.9668113
9:	ap-southeast-1.compute	2.8879647
10:	ap-south-1.compute	2.5913326
11:	ap-northeast-2.compute	2.5346497
12:	com.ec2	2.4392407
13:	sa-east-1.compute	2.2049954
14:	internal.ec2	1.7763651
15:	us-east-1.compute	1.7595327
16:	ca-central-1.compute	1.4508265
17:	.	1.4198201
18:	us-west-1.compute	1.4166473
19:	eu-west-3.compute	1.3051468
20:	metadata.google	0.8742713
21:	eu-north-1.compute	0.8657445
22:	ap-east-1.compute	0.6080187
23:	us-gov-west-1.compute	0.4586415
24:	iocs2.rancher	0.3656085
25:	europe.7s	0.3569946
26:	net.ec2	0.3532283
27:	me-south-1.compute	0.3512589
28:	af-south-1.compute	0.3455081
29:	de.ec2	0.3376713
30:	eu-south-1.compute	0.3026175

the graph on the left is looking at the percentage of the number of labels contained within the key name. vs .mail with 60% having only 1 label, .internal has more context in terms of labels and content within query name

Slide 9: Label Analysis 2



web page for Cooper Nettie technology that is Called rancher and it suggests putting your DNS under a suffix search list for rancher .internal ee're seeing a change that the rancher software is probably very likely the the root cause of many of these others. So I have made it a item to submit a ticket to GitHub and disclose this

Warren: believe this issue was with an older version of .rancher; ended in 2018 with new version. The fact that this is still seen gives us a metric for expectations at how quickly a software related issue ends even after software is changed.

Slide 10: .svc Segway and Kubernetes

.INTERNAL Analysis :: SVC Segway and Kubernetes

Table 6: Kubernetes References to URLs containing SVC

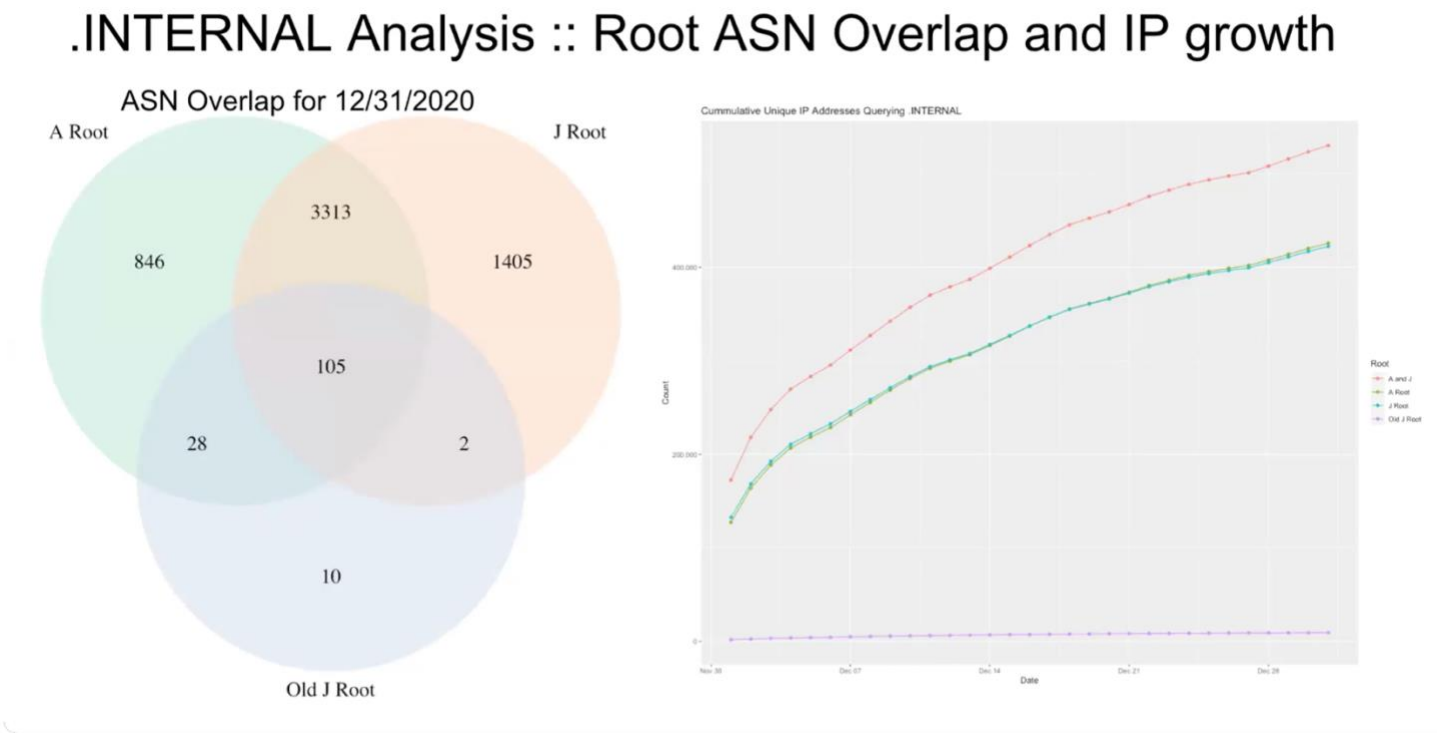
File	Line	Text
kubernetes/cmd/kubeadm/app/phases/controlplane/manifests.go	141	"service-account-issuer": fmt.Sprintf("https://kubernetes.default.svc.%s", cfg.Networking.DNSDomain),
kubernetes/staging/src/k8s.io/kube-aggregator/artifacts/self-contained/etcd-pod.yaml	22	- "--advertise-client-urls=https://etcd.kube-public.svc:4001"
kubernetes/staging/src/k8s.io/kube-aggregator/artifacts/self-contained/etcd-pod.yaml	28	- "--initial-advertise-peer-urls=https://etcd.kube-public.svc:7001"
kubernetes/staging/src/k8s.io/kube-aggregator/artifacts/self-contained/etcd-pod.yaml	33	- "--initial-cluster=default=https://etcd.kube-public.svc:7001"
kubernetes/staging/src/k8s.io/kube-aggregator/artifacts/self-contained/kubernetes-discover-pod.yaml	43	- "--etcd-servers=https://etcd.kube-public.svc:4001"

This is for .svc, which started getting 127million queries per day. Looks similar to .internal…..85% came from a company whose cloud infrastructure was using a Kubernetes configuration leaking out svc queries.

Data Sensitivity Analysis

Slide 11:

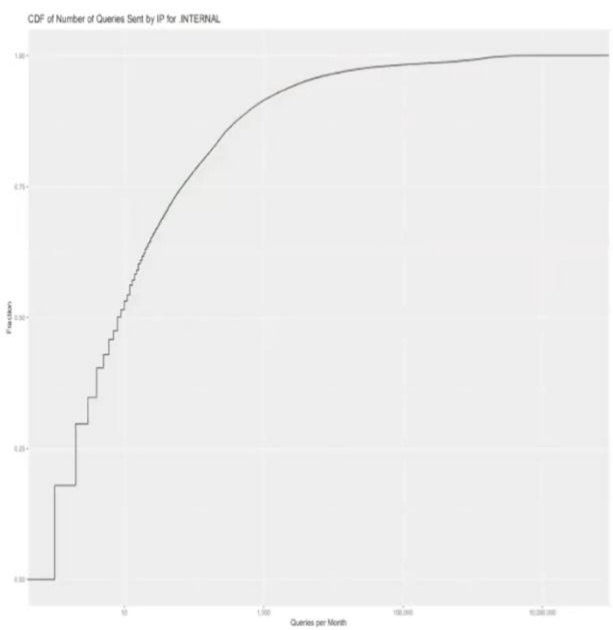
Slide 12: Root ASN Overlap and IP Growth

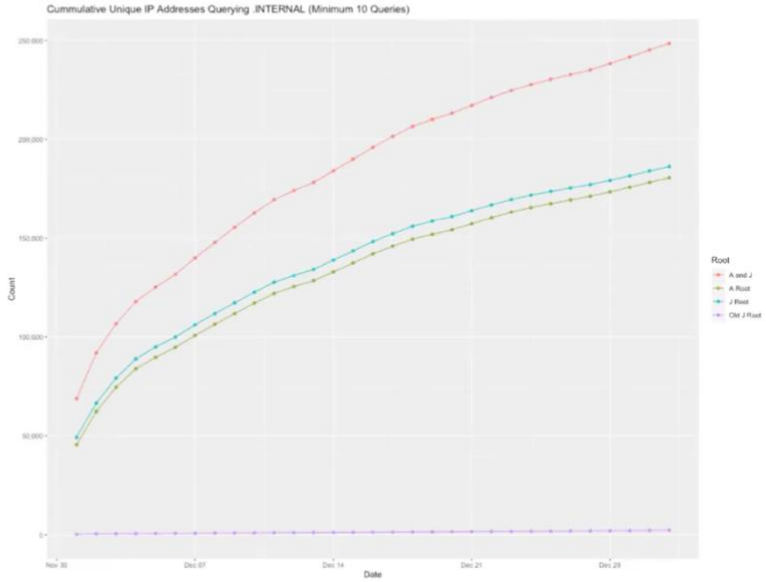


Different than .mail.

Slide 13: Root ASN Overlap and IP Growth 2

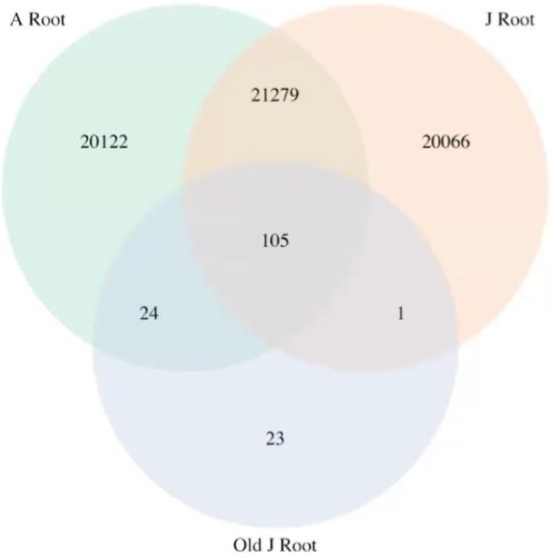
.INTERNAL Analysis :: Root ASN Overlap and IP growth

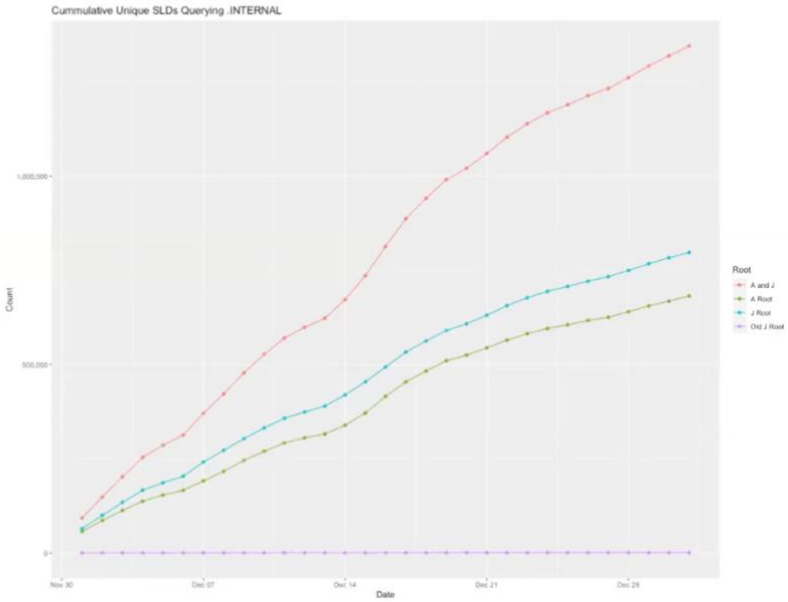




Slide 14: SLD Overlap Analysis

.INTERNAL Analysis :: SLD Overlap Analysis





The catchment of A and J differ. 80% of SLDs only seen once in the month.