NCAP Discussion Group | 27 January

Agenda:

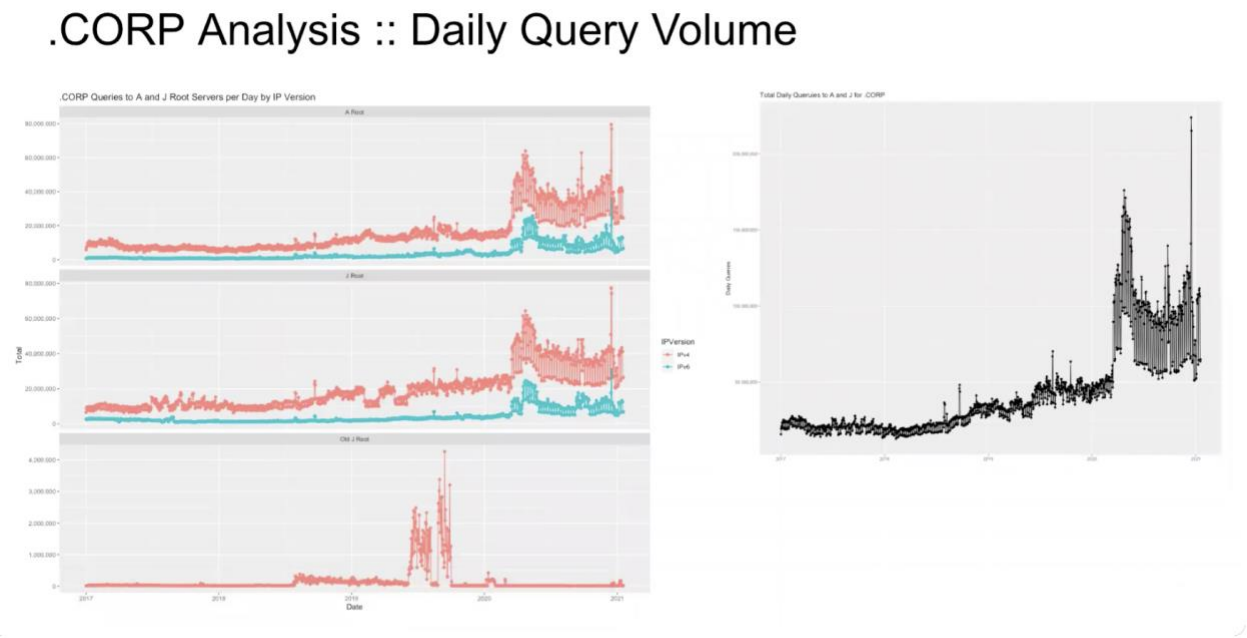
1. Welcome and roll call
2. Update to SOI
3. .CORP Case Study: https://docs.google.com/presentation/d/1mcOpf-4bugrc_aqVQCn5LaC5CwF4QHdOT1MClxSzcY4/edit#slide=id.p [docs.google.com].
4. HOME Case Study: https://docs.google.com/presentation/d/1A8u1acNf85PMCKiAEC_inzfOIm3cQUkGzpSsKyI33FQ/edit#slide=id.p
5. Update on Study 2
6. AOB

Table of Contents

Name Collision Analysis
.CORP

.CORP CASE STUDY

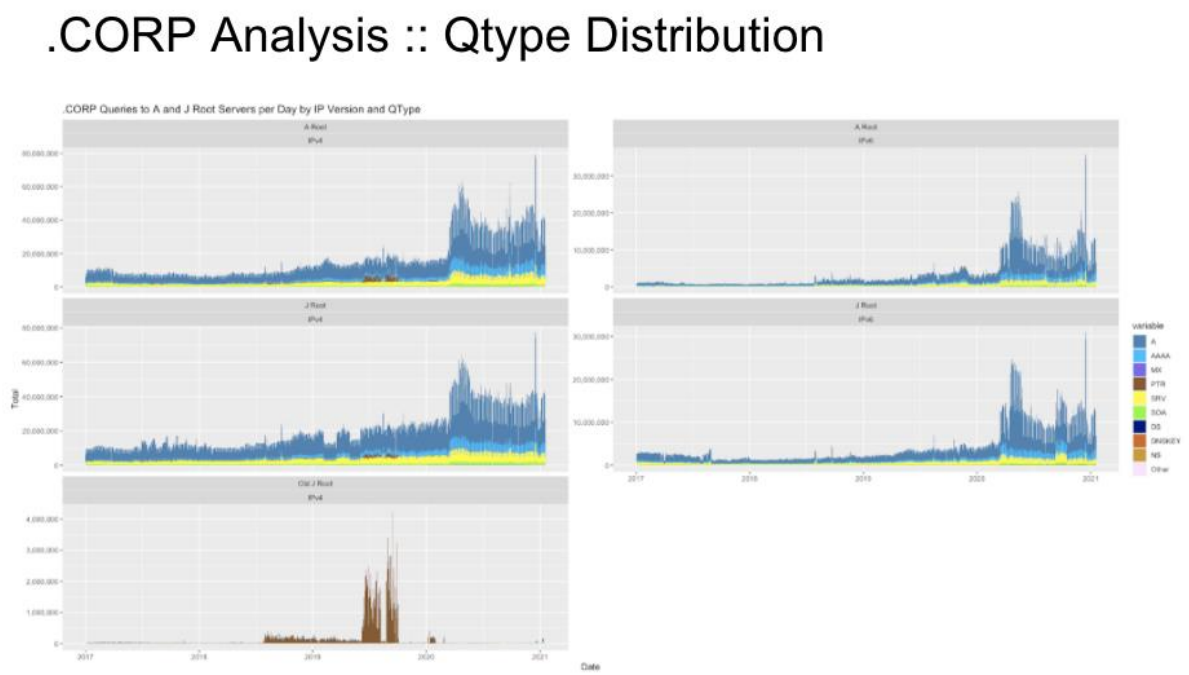
Slide 2: Daily Query Volume



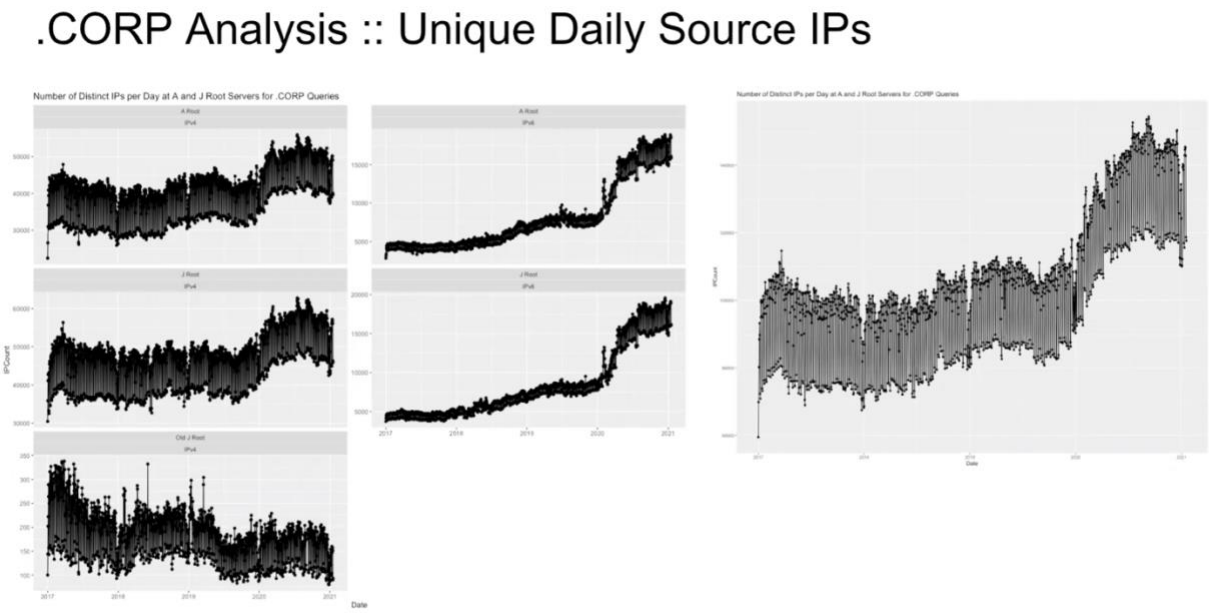
Similar to what was seen on .mail and .internal.

A larger than normal % delegated tlds for srv records inside of .corp

Slide 3: Qtype Distribution

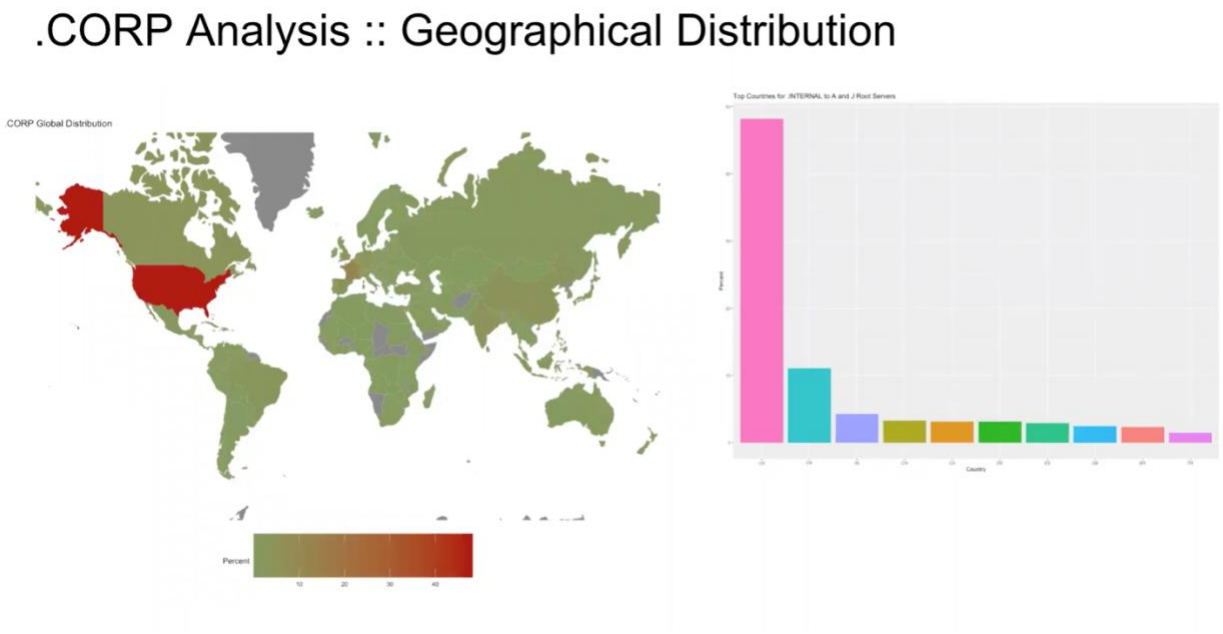


Slide 4: Unique Daily Source IPs



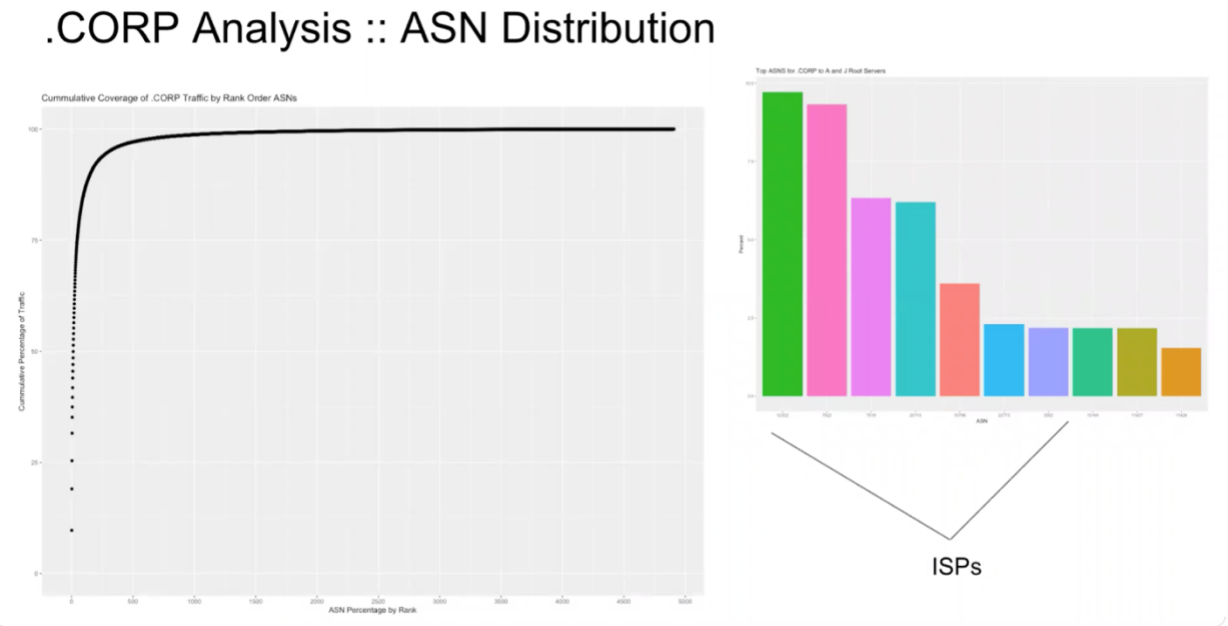
# of unique IP address requesting .corp. Unique addresses hitting A & J goes up 40%. Suddenly corp is coming from a much larger, much wider # of sources out there on the Internet - this is likely from transient devices that are leaving the corporate networks and being used in more residential places. (work from home)

Slide 5: Geographical Distribution



½ of traffic from US

Slide 6: ASN Distribution

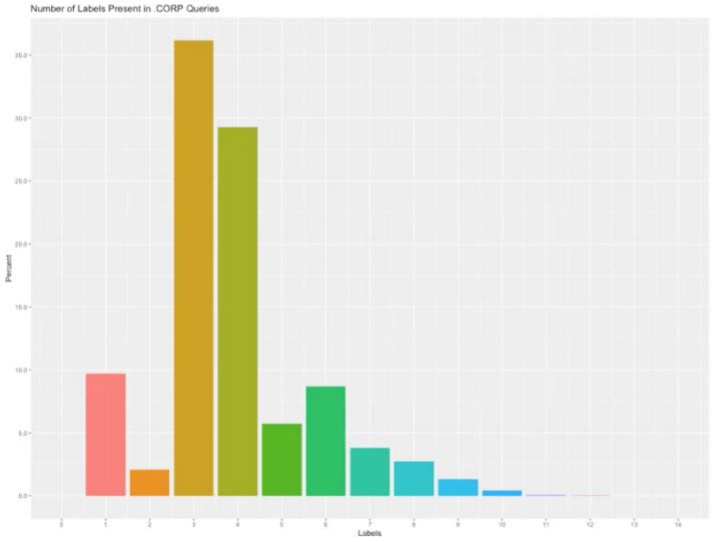


Shows what networks are sending this traffic:
.mail had 900ish ASNs, .corp is spread out over 5000 ASNs
95th percentile of traffic you need to interact with 250 ASN to fix this .corp query leak
Graph to the right looks at top ASNs sending .corp traffic and is mostly residential ISPs in US or France

Now is residential; was previously corporate.

Slide 7: Label Analysis

.CORP Analysis :: Label Analysis



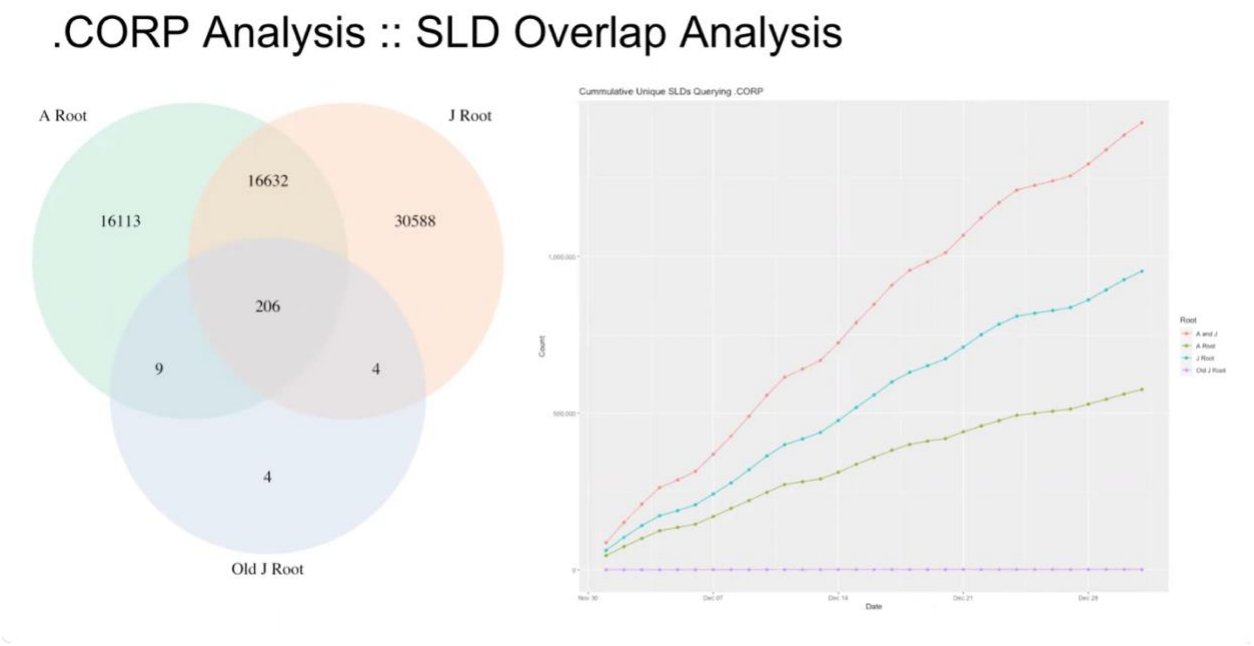
	SLD	Percent
1:	airbus	8.0896220
2:	sap	6.4356841
3:	bank	4.3002450
4:	zurich	2.1306262
5:	teva	1.6235524
6:	parker	1.5906210
7:	bvcorp	1.5695795
8:	ecolab	1.3010942
9:	root	1.2664888
10:	stream	1.1201463
11:	_	0.8882908
12:	info	0.7628606
13:	hospira	0.7359782
14:	bmw	0.7028646
15:	alico	0.6924749
16:	global	0.6794046
17:	sdl	0.6703310
18:	logistics	0.6678828
19:	davita	0.6549482
20:	concentra	0.6327179
21:	sungard	0.5835661
22:	us	0.5488492
23:	bi	0.5419070
24:	ad	0.5335444
25:	atd	0.5097991
26:	t2	0.4871012
27:	casa	0.4660969
28:	eurocopter	0.4254081
29:	ares	0.4045543
30:	internal	0.4021990

	ThirdLabel	Percent
1:	eu.airbus	6.4472254
2:	wfb.bank	3.5449216
3:	wdf.sap	1.8544951
4:	global.ecolab	0.8849001
5:	amer.zurich	0.8116516
6:	phl.sap	0.7746473
7:	emea.zurich	0.6780281
8:	us.parker	0.6369637
9:	chs.concentra	0.6003767
10:	internal.sungard	0.5391947
11:	eua.bvcorp	0.5154707
12:	prod.atd	0.5082877
13:	uk.parker	0.4730215
14:	accounts.root	0.4417752
15:	pal.sap	0.4389217
16:	americas.stream	0.3029020
17:	sin.sap	0.2951222
18:	res.airbus	0.2910688
19:	as.airbus	0.2887339
20:	asi.bvcorp	0.2758672
21:	blrl.sap	0.2757036
22:	il.teva	0.2604276
23:	emea.stream	0.2556482
24:	apac.stream	0.2555943
25:	dyson.global	0.2534640
26:	na.airbus	0.2515864
27:	ame.bvcorp	0.2493928
28:	dtc.dish	0.2477579
29:	ihs.internal	0.2461601
30:	eadscasa.casa	0.2415583

What is going on in .corp? What % of names contain how many labels?
10% of names coming in for .corp contain ONLY 1 label. 60% have 3 or 4 labels- better Q names to give us better understanding of the source of those queries.

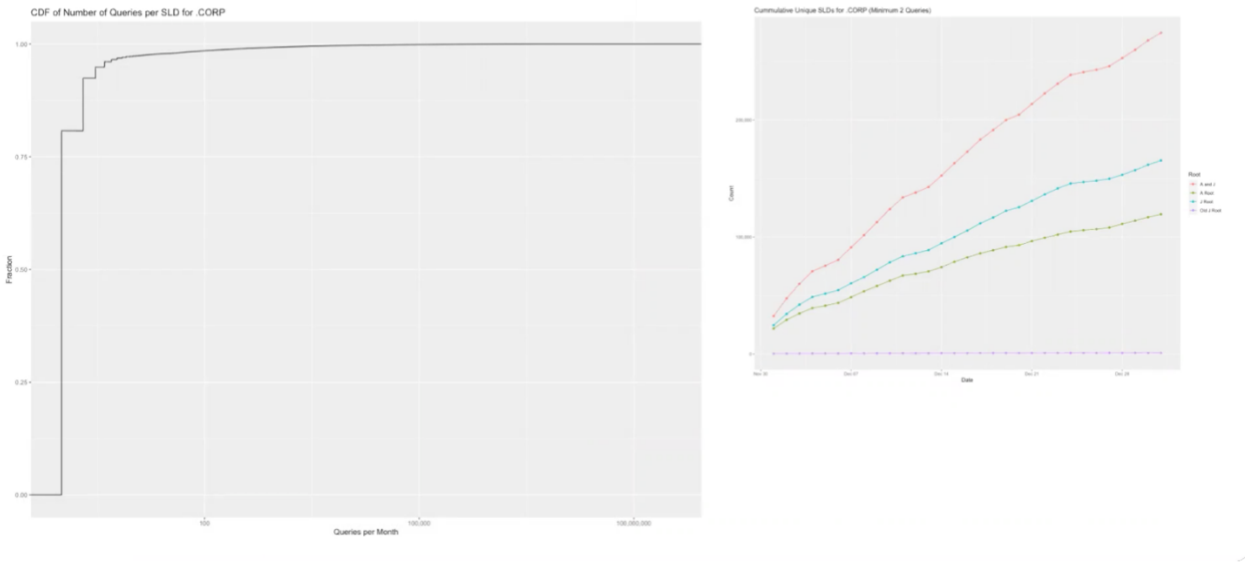
two lists on the right side here, the first list in the middle, is looking at the top 30 SLDs ranked by the total percentage of traffic. Apparent than many are big corporations (elected to use .corp string within their corporate network) but then you see .internal
maybe some of this is still the byproduct of suffix search list appendage and .corp is getting appended to things like .internal, so you have X .internal .corp ending up being on there

Slide 6: SLD Overlap Analysis



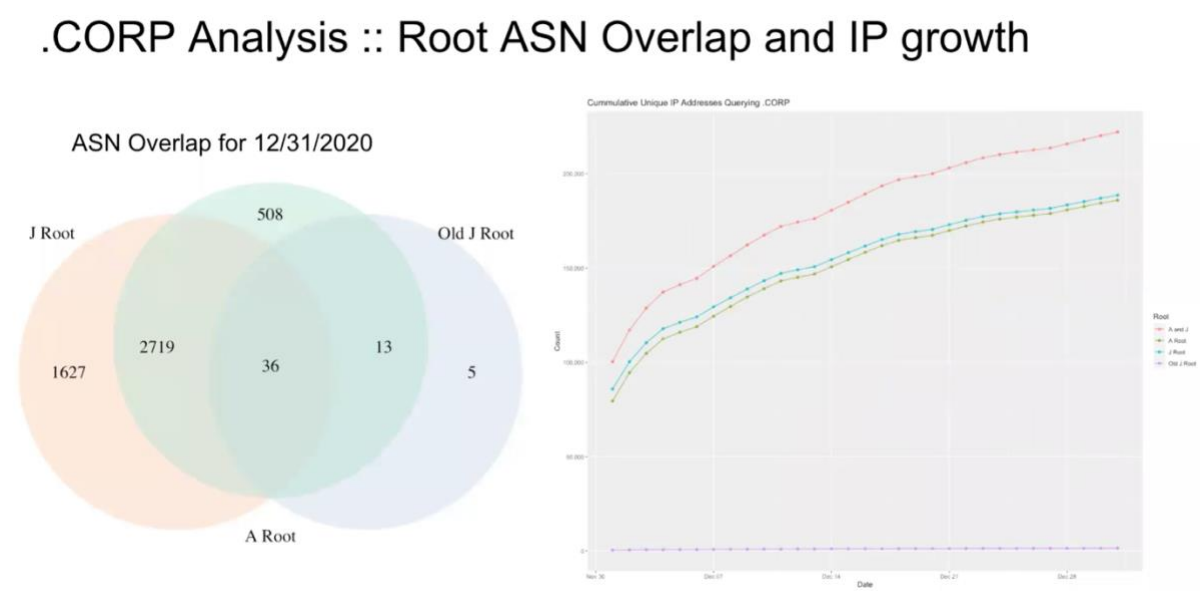
If we look at 1, 2 or 3 roots how muc of overall picture do we have. Left graph broken by A and J Root – clearly each root has it's own unique capture point. Graph at right moving at linear growth rate.

Slide 7: SLD Overlap Analysis 2



This is a cumulative distribution plot of how many times, a specific SLD was queried. In the entire month of December, roughly 80% of these strings are SLDs only be inquired 1 time. Indicative of a random label, or chromium,

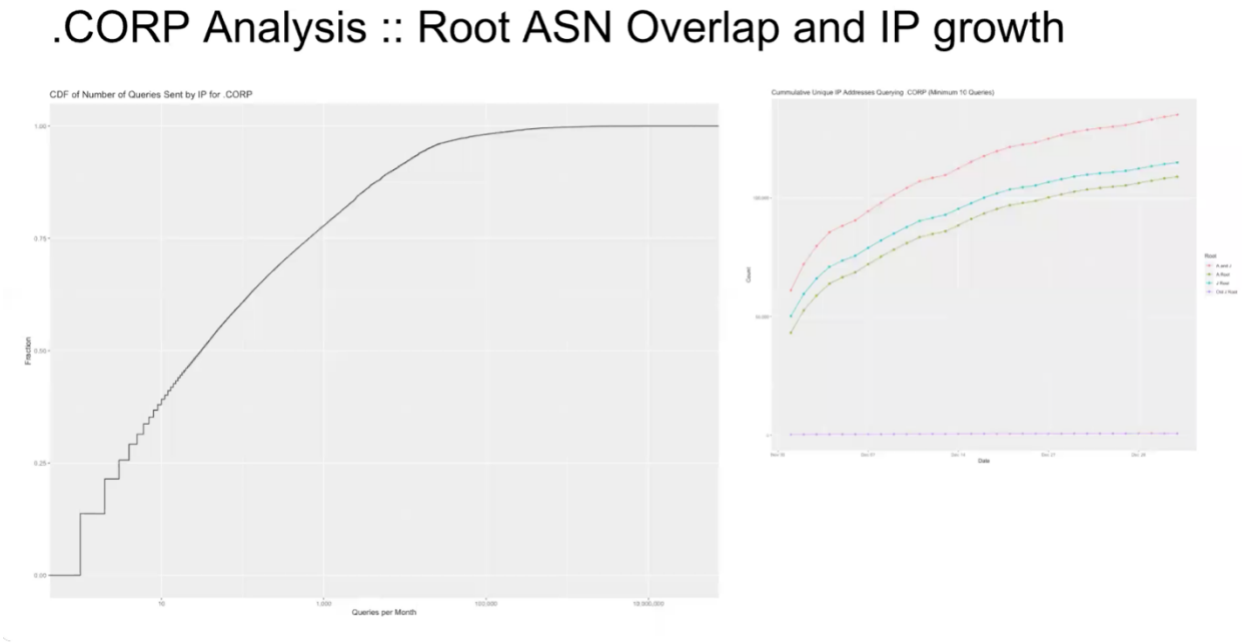
SLIDE 8: Root ASN Overlap and IP growth



This repeats that analysis on the amount of ASNs and different various recursive resolve network operators

venn diagram on the left, is showing that a and j having fairly significant overlap, but J group has 1600 different ASNs. So again, that catchment of a particular route and its contribution to the data analysis, is telling. Maybe this is one of those questions that we're going to want to further study when we do the data sensitivity analysis by either asking additional routes to help fit fill out this venn diagram or we conduct that kind of exercise on something like the dital (?) data (which limits us to 2 days).

Slide 9: Root ASN Overlap and IP growth 2

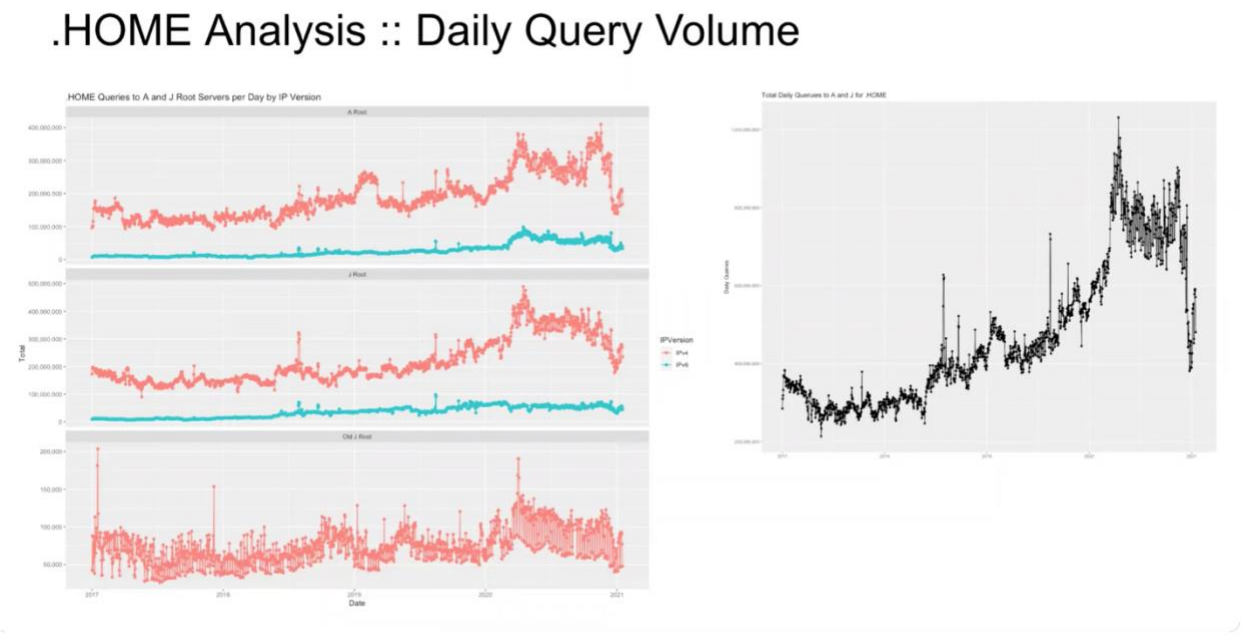


END OF .CORP SLIDES

.HOME CASE STUDY

Name Collision Analysis .HOME

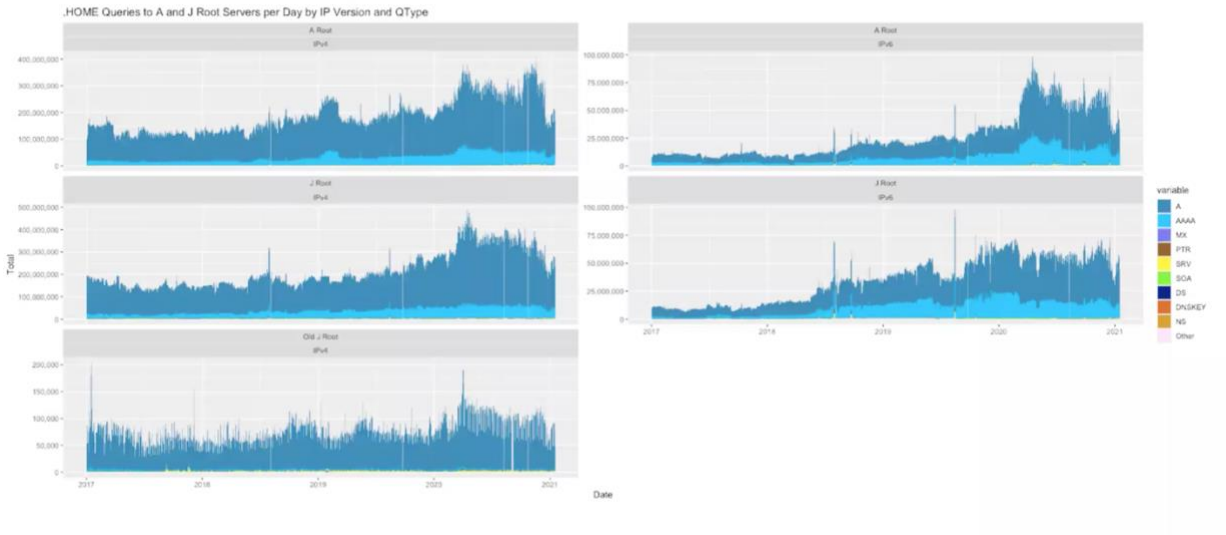
Slide 1: Daily Query Volume



Looking at A & J: query volume rose until Nov/Dec 2020 rapid decline. Decline due to change in chromium and change in a way it sends out a probe request on the android platform to detect an Internet redirecter

SLIDE 2: Qtype Distribution

.HOME Analysis :: Qtype Distribution



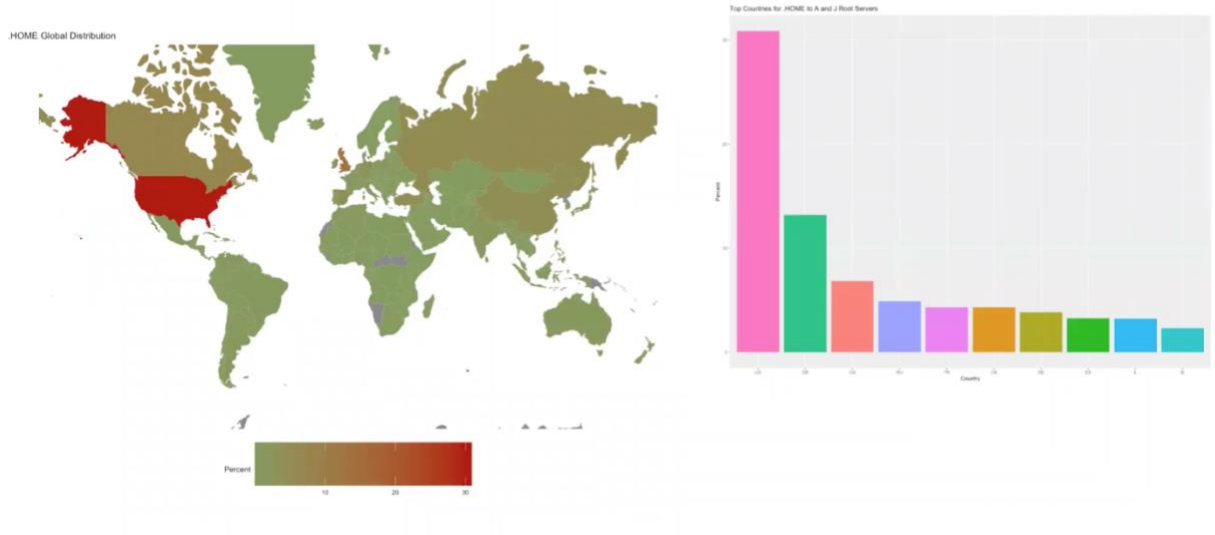
SRP records make up significantly less than .corp -an indicator of how the .home string is actually being used. .home string is caused from suffolk search list appendage where .copr was intentionally anchored by those various corporations.

Slide 3: Unique Daily Source IPs

.HOME Analysis :: Unique Daily Source IPs

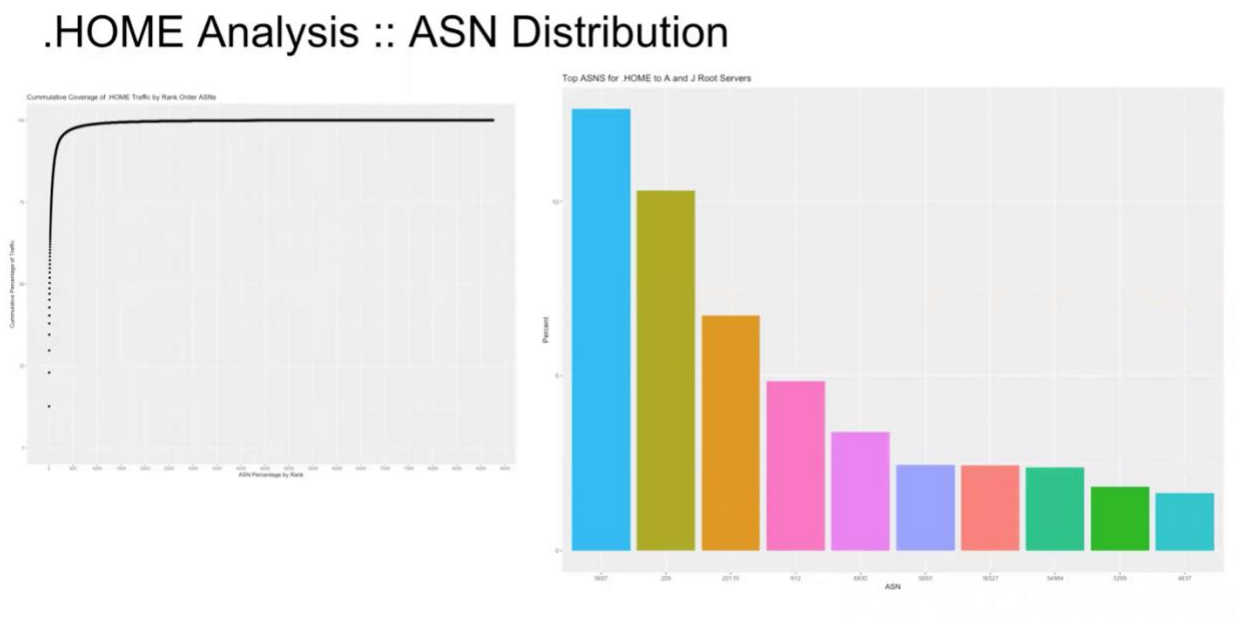
Slide 4: Geographical Distribution

.HOME Analysis :: Geographical Distribution



.home is not as US centric as .corp

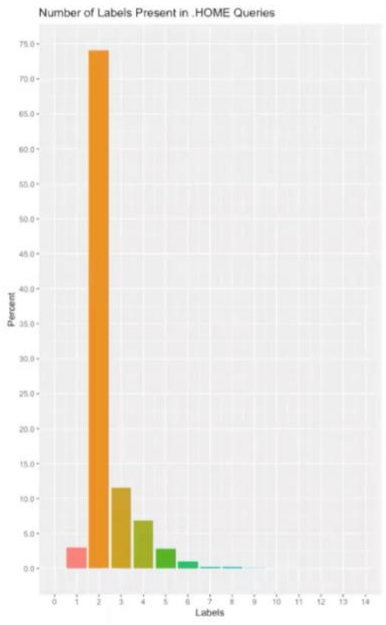
Slide 5: ASN Distribution



9000+ on highest bar: 9000 ASNS that A& J were receiving traffic from .home which was significantly higher than anything coming out of it. to get to the 95th percentile you're going to probably have to talk to upwards of 500 or so a sentence.

Slide 6: Label Analysis

.HOME Analysis :: Label Analysis



	SLD	Percent
1:	hitronhub.home.	9.75412714
2:	com.home.	6.61782804
3:	home.	3.03802436
4:	_.home.	2.68200505
5:	net.home.	1.19227441
6:	ht.home.	0.57062557
7:	fios-router.home.	0.52528746
8:	_tcp.home.	0.30401179
9:	wpad.home.	0.29160084
10:	org.home.	0.28520555
11:	cn.home.	0.26442140
12:	_udp.home.	0.23789951
13:	ch.home.	0.21296067
14:	ru.home.	0.19720468
15:	arpa.home.	0.08625248
16:	io.home.	0.08591727
17:	tv.home.	0.08381899
18:	isatap.home.	0.07388558
19:	me.home.	0.06444656
20:	biz.home.	0.05839011
21:	unifi.home.	0.05735203
22:	workgroup.home.	0.05627388
23:	in.home.	0.05361050
24:	home.home.	0.05328508
25:	info.home.	0.04995074
26:	uk.home.	0.04905044
27:	co.home.	0.04637510
28:	xyz.home.	0.04604986
29:	jpg.home.	0.03915547
30:	local.home.	0.03533617

```
                SLD    Percent
> sum(x$Percent)
[1] 27.11263
```

	ThirdLabel	Percent
1:	com.hitronhub.home.	0.45201229
2:	googleapis.com.home.	0.25313555
3:	google.com.home.	0.23203541
4:	_dns-sd._udp.home.	0.23035555
5:	infomaniak.ch.home.	0.20446238
6:	ksmobile.com.home.	0.19954698
7:	facebook.com.home.	0.15047152
8:	qq.com.home.	0.13119609
9:	tiktokv.com.home.	0.12342135
10:	googlevideo.com.home.	0.11478654
11:	fbcdn.net.home.	0.11288061
12:	hicloud.com.home.	0.10143395
13:	amazon.com.home.	0.09411570
14:	ntp.org.home.	0.08098048
15:	amazonaws.com.home.	0.08032620
16:	net.hitronhub.home.	0.07344758
17:	cloudfront.net.home.	0.06952440
18:	com.cn.home.	0.06615579
19:	in-addr.arpa.home.	0.06239578
20:	ovh.net.home.	0.06166118
21:	nwtelecom.ru.home.	0.05785983
22:	crashlytics.com.home.	0.05344172
23:	xiaomi.com.home.	0.05337555
24:	ksmobile.net.home.	0.05265146
25:	gstatic.com.home.	0.05202927
26:	_aaplcache3._tcp.home.	0.04942245
27:	_aaplcache._tcp.home.	0.04940269
28:	_aaplcache1._tcp.home.	0.04905751
29:	_aaplcache4._tcp.home.	0.04900385
30:	_aaplcache2._tcp.home.	0.04895816

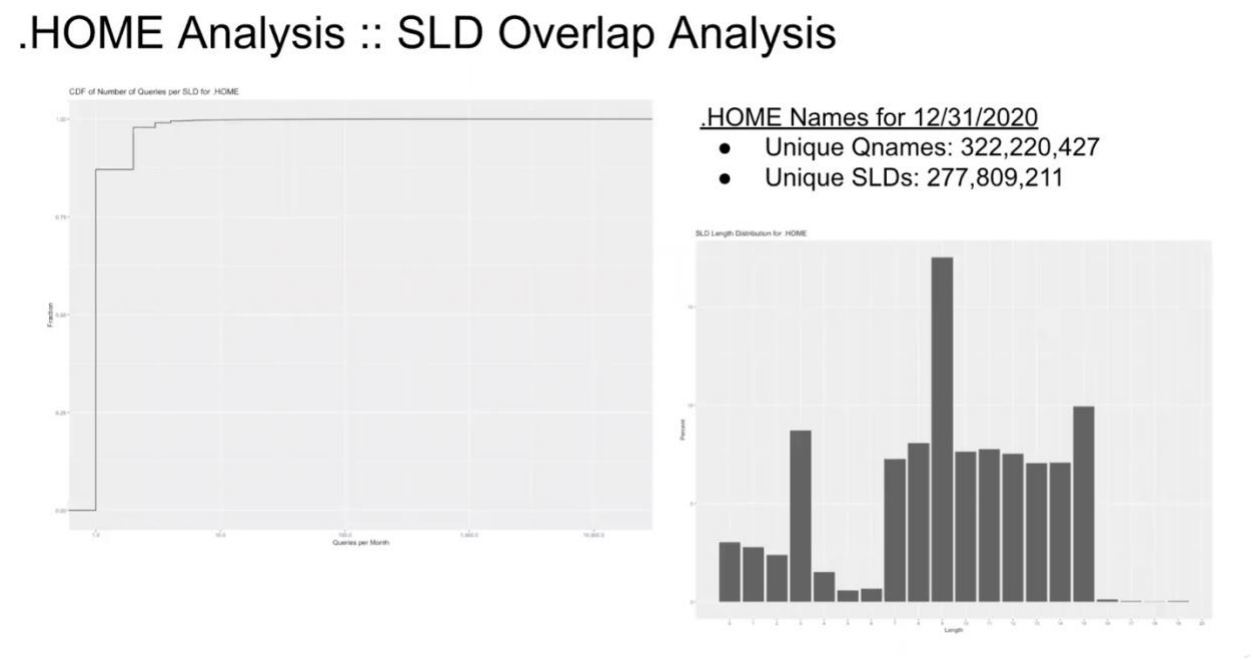
Left is # of labels present in actual strings; almost 75% of all queries coming out with only 2 labels, seems strange. The middle column list, 2nd label domains, many of these are delegated top level domains, that all have .home randomly appended to it. Very different than .corp where those strings looked purposely anchored on.

Warren: 2nd level domains – some CPE integrated software doing Q&A minimization badly, followed by search. DNS Mask version where someone tried Q&A minimization and did badly

Suggested Action: what would be useful to be to be able to hunt down somebody who's got a piece of CPE who's doing these look ups and sort of beat them and see if we can run it in a lab.

Column on right: what is underneath 2nd label domains we see most common domains

Slide 7: SLD Overlap Analysis



This is just looking at the properties of the 2nd level domains coming in
85% receive 1 query
1 day of Q names comin in for .home was over 322 million that broke down to 277 million unique SLDs
Takes a look at SLDSs based off of the length of characters of that string. See flat bar between 7 – 15 chars long, spike at 10 chars; probably chromium queries.

Slide 8: First Label Analysis

.HOME Analysis :: First Label Analysis

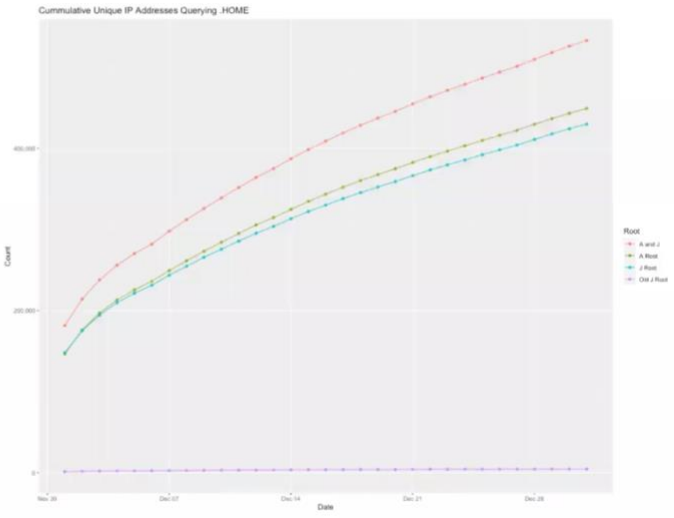
Column1	Column2
home	16913452
_	15263738
wpad	2076216
www	1545338
api	1418605
_ldap	1164569
swift01-prx	1127655
lb	843759
isatap	459277
tracker	445200
clock	401891
unifi	348209
android	325169
lb1040	322647
b	310944
graph	291564
_aaplcache3	288944
_aaplcache	288750
_aaplcache1	286837
_aaplcache4	286487
_aaplcache2	286414
connectivitycheck	281549
ntp1	278197
cdn	273304
helpcmsecurity1	263168
time	259139
db	258949
static	253269
i	253109
cmdts	247335

Column1	Column2
_	15263738
_ldap	1164569
_aaplcache3	288944
_aaplcache	288750
_aaplcache1	286837
_aaplcache4	286487
_aaplcache2	286414
_vlmcs	206050
_kerberos	187617
_bradfordagent	42964
_goverlanserver	36979
_goverlan	18195
_sip	15645
_msdcs	15151
_tcp	14138
_autodiscover	13062
__bridge_loaded__	12578
_udp	11433
__hola__	9637
_pcoip-bootstrap	7340
_tzmgr_discovery	6772
_https	5599
_hpdm-gateway	5399
_dns-sd	4907
_sips	4781
_sipinternaltls	4583
__sw	4454
_capwap-control	3019
_elfws	2812
_wdmserver	2605

First table is most popular labels contained under .home, so this is first label in the queue name that is specified. Shows what is being used under .home

Slide 9: Root ASN Overlap and IP Growth

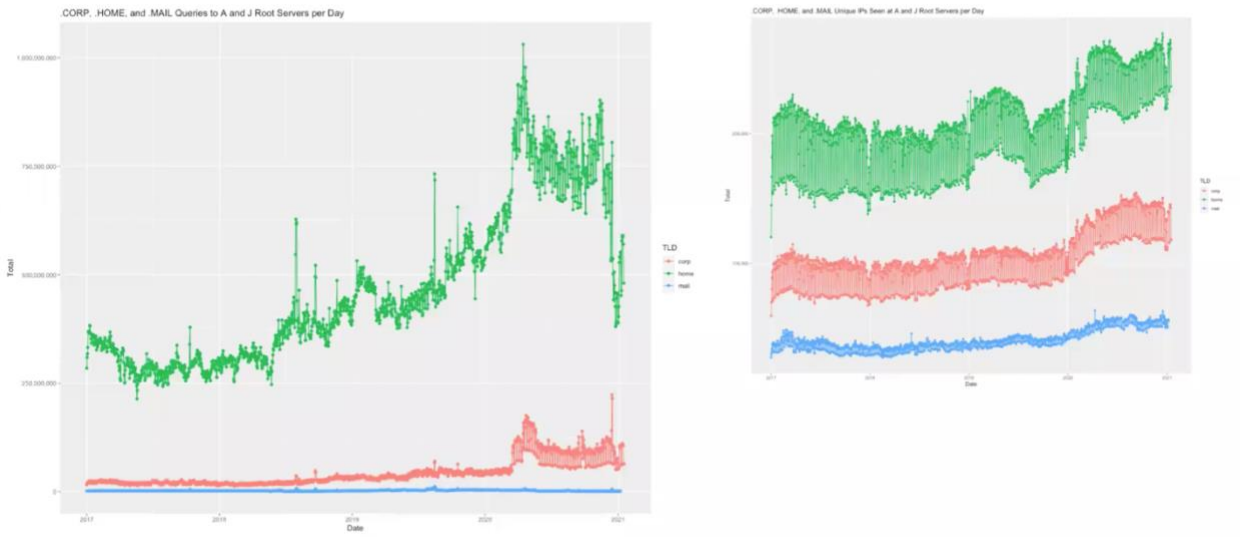
.HOME Analysis :: Root ASN Overlap and IP growth



.CORP, .HOME, and .MAIL Comparison

Slide 10: .corp, .home and .mail comparison

.CORP, .HOME, and .MAIL Comparison



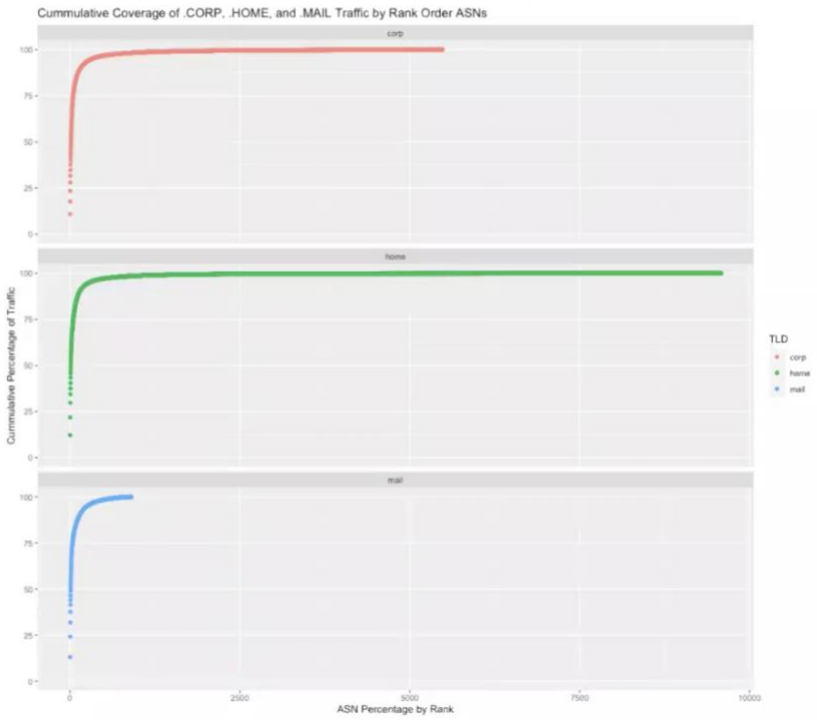
Left: query volume, .home is several magnitudes larger than .mail, and a magnitude or more than.corp. Dip in .home was from chromium

Query vol isn't directly an indicator of risk, but is a reflection of it

Right: # of unique Ips

SLIDE 11: .corp, .home and .mail 2

.CORP, .HOME, and .MAIL Comparison

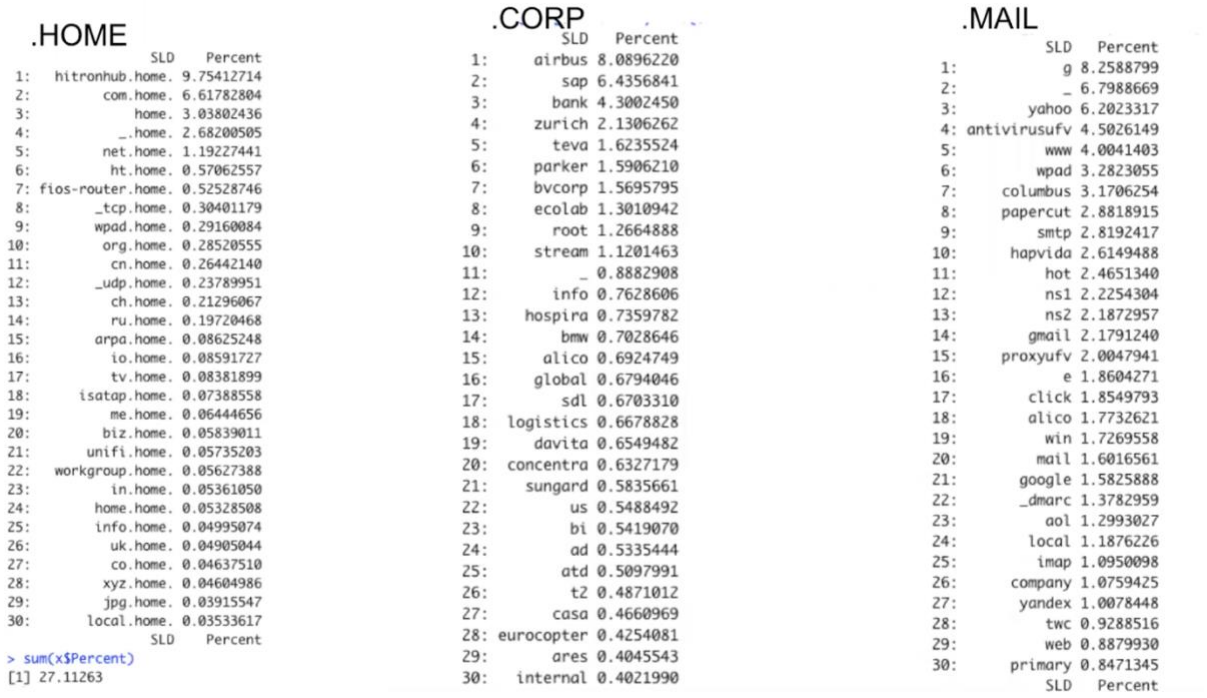


Is the traffic from all over different places, or is it relatively confined to a set of particular networks? Importan to inform our decisions in regards to risk assessment and remediation possibilities right.
.mail has less than 900n different Ass if you did the top 10 it was going to remediate more than 50% of traffic.

.home I spread out over larger network operator bases, to get that kind of remediation means you have to interacgt with a lot of operators unless you can use clues within the queue names

SLIDE 12: .corp, .home and .mail 3

.CORP, .HOME, and .MAIL Comparison



.HOME

	SLD	Percent
1:	hitronhub.home.	9.75412714
2:	com.home.	6.61782804
3:	home.	3.03802436
4:	_.home.	2.68200505
5:	net.home.	1.19227441
6:	ht.home.	0.57062557
7:	fios-router.home.	0.52528746
8:	_tcp.home.	0.30401179
9:	wpad.home.	0.29160084
10:	org.home.	0.28520555
11:	cn.home.	0.26442140
12:	_udp.home.	0.23789951
13:	ch.home.	0.21296067
14:	ru.home.	0.19720468
15:	arpa.home.	0.08625248
16:	io.home.	0.08591727
17:	tv.home.	0.08381899
18:	isatap.home.	0.07388558
19:	me.home.	0.06444656
20:	biz.home.	0.05839011
21:	unifi.home.	0.05735203
22:	workgroup.home.	0.05627388
23:	in.home.	0.05361050
24:	home.home.	0.05328508
25:	info.home.	0.04995074
26:	uk.home.	0.04905044
27:	co.home.	0.04637510
28:	xyz.home.	0.04604986
29:	jpg.home.	0.03915547
30:	local.home.	0.03533617

```
> sum(x$Percent)
[1] 27.11263
```

.CORP

	SLD	Percent
1:	airbus	8.0896220
2:	sap	6.4356841
3:	bank	4.3002450
4:	zurich	2.1306262
5:	teva	1.6235524
6:	parker	1.5906210
7:	bvcorp	1.5695795
8:	ecolab	1.3010942
9:	root	1.2664888
10:	stream	1.1201463
11:	_	0.8882908
12:	info	0.7628606
13:	hospira	0.7359782
14:	bmw	0.7028646
15:	alico	0.6924749
16:	global	0.6794046
17:	sdl	0.6703310
18:	logistics	0.6678828
19:	davita	0.6549482
20:	concentra	0.6327179
21:	sungard	0.5835661
22:	us	0.5488492
23:	bi	0.5419070
24:	ad	0.5335444
25:	atd	0.5097991
26:	t2	0.4871012
27:	casa	0.4660969
28:	eurocopter	0.4254081
29:	ares	0.4045543
30:	internal	0.4021990

.MAIL

	SLD	Percent
1:	g	8.2588799
2:	_	6.7988669
3:	yahoo	6.2023317
4:	antivirusufv	4.5026149
5:	www	4.0041403
6:	wpad	3.2823055
7:	columbus	3.1706254
8:	papercut	2.8818915
9:	smtp	2.8192417
10:	hapvida	2.6149488
11:	hot	2.4651340
12:	ns1	2.2254304
13:	ns2	2.1872957
14:	gmail	2.1791240
15:	proxyufv	2.0047941
16:	e	1.8604271
17:	click	1.8549793
18:	alico	1.7732621
19:	win	1.7269558
20:	mail	1.6016561
21:	google	1.5825888
22:	_dmarc	1.3782959
23:	aol	1.2993027
24:	local	1.1876226
25:	imap	1.0950098
26:	company	1.0759425
27:	yandex	1.0078448
28:	twc	0.9288516
29:	web	0.8879930
30:	primary	0.8471345

Looking at SLDs like this gives insight into what it's being used for and possibility to identify the underlying cause and remediate it if you know what caused the CPE devices and you c an figure that out- can you work with CPE devices to push that out and remediate.