NCAP Discussion Group teleconference | 13 January

Agenda:

1. Welcome and roll call
2. Update to SOI
3. Update on Study 2
4. Outstanding questions on previous NCAP DG call: Slides here (WIP): https://docs.google.com/presentation/d/1msCT0aZJ6fBuB7Xq5N7HUzQhLFGMCSlHnHchmvtQxFo/edit#slide=id.p [docs.google.com]
    - .MAIL Qtype
    - .INTERNAL Breakout
5. JAS / Interisle Review
6. AOB

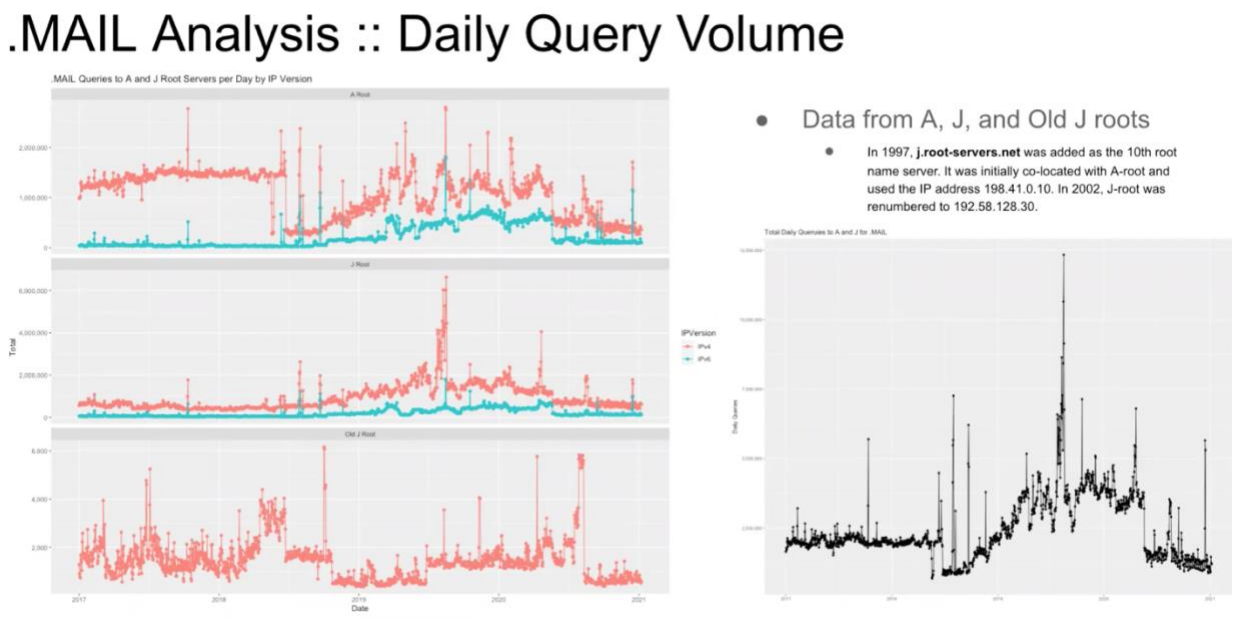
Table of Contents

Key NCAP Discussion Question: Where is the harm and how do we assess it?

Name Collision Analysis
.MAIL

this presentation is going to look at data captured from AJ root servers with an extra special old J route there.
J route was added as the 10th name server and it was initially co located with a route but it used the IP address.
in 2002 it was remembered to a new IP address and it's been that since then
since 2002 Verisign is continued to run instance on that IP address, It's still receives a fair amount of traffic.

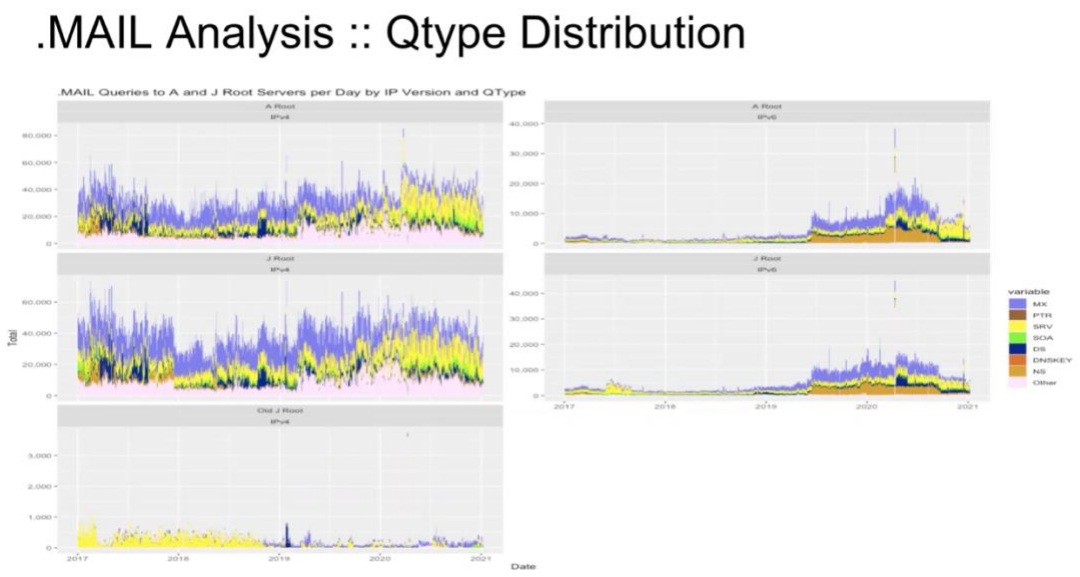
Slide 1: Daily Query Volume



Graph on right is total daily query vol per .mail over last 4 years. 2017 – 2018 stable then sudden drop on A route. Then ramped up and in April 2020 another drop (corona virus related??) and has stayed low.
3 graphs on right show breakout of each individual route by query volume.

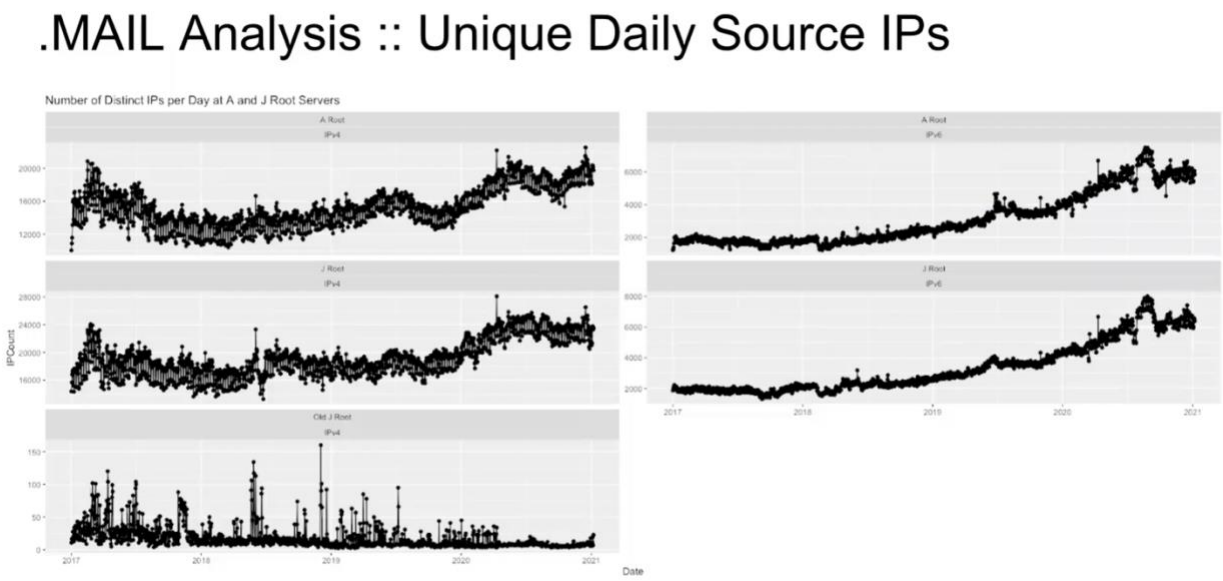
Wanted to see Q type distribution for .mail. Not for MX Q types, no special affinity for a particular Q type.

Slide 2: Qtype Distribution



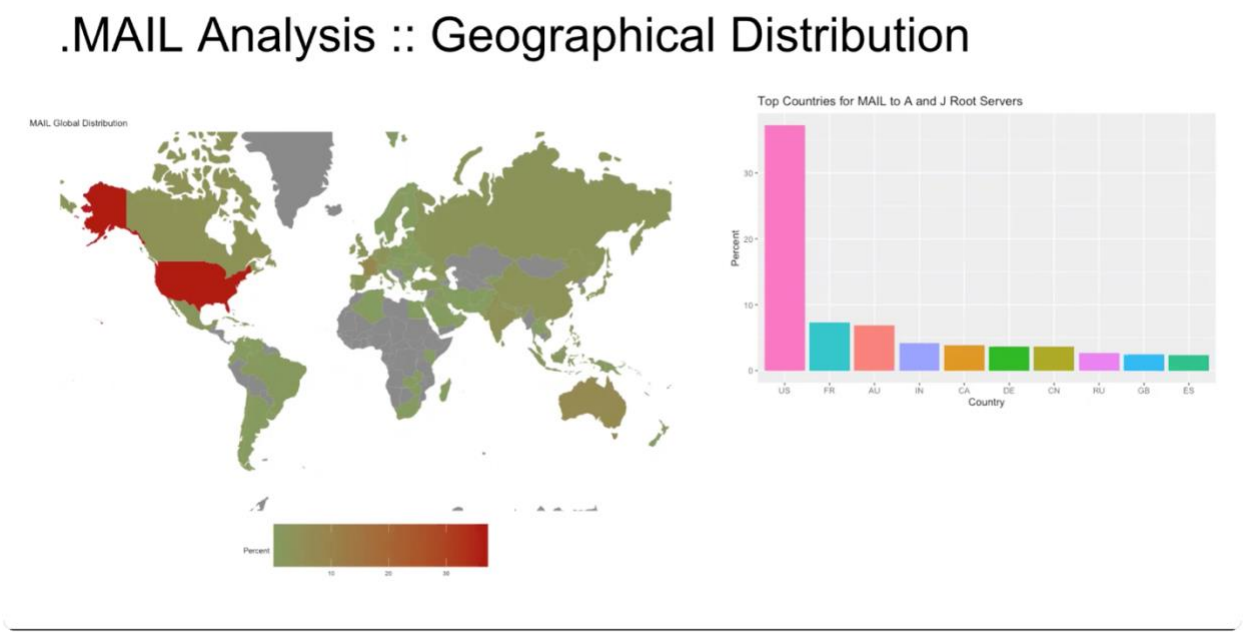
What is the affinity in other tlds, compared to .mail?

Slide 3: Unique Daily Source IPs



More and more source Ips sending more queries to A & J for .mail. Indicates that .mail is requested from a larger set of Ips out there

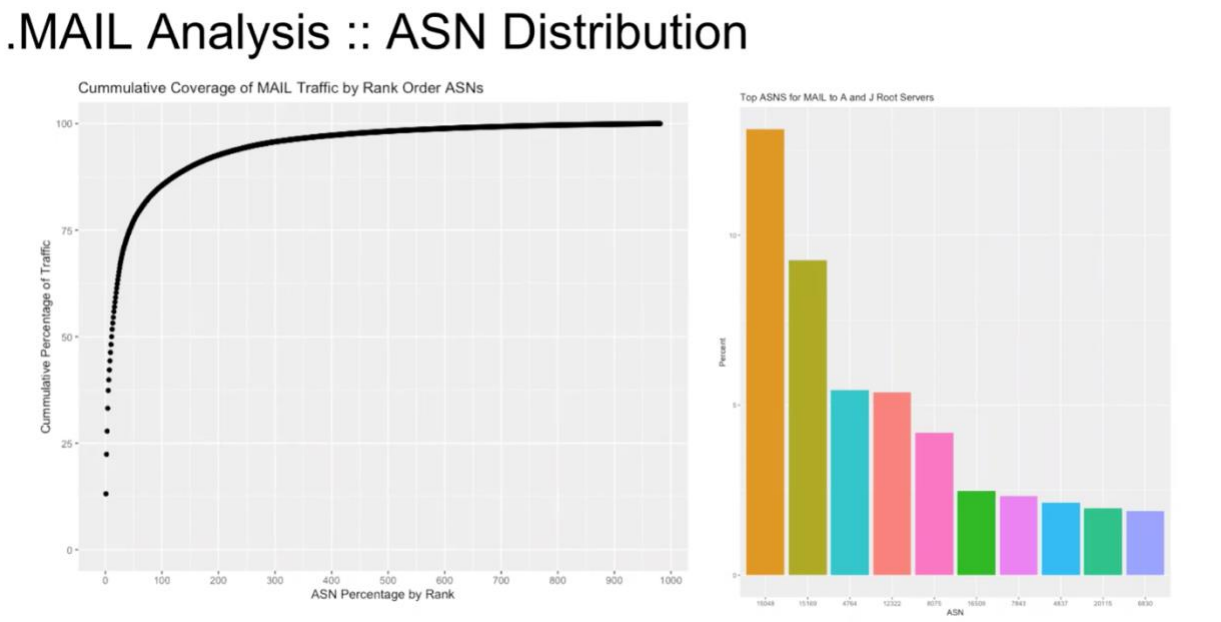
Slide 4: Geographical Distribution



Traffic from a disperse set of sources

Has anyone Looked at what is making the dot mail queries?
jeff Schmidt: found we j s long ago when we looked at this arm we found a set of sample send mail configuration files.
Set of XML configuration files published in one of the O'Reilly books **That had mail in them hard coded. if anybody copies and pastes it…..** we suspected that that that was responsible for at least some of the behavior..

Slide 5: ASN Distribution

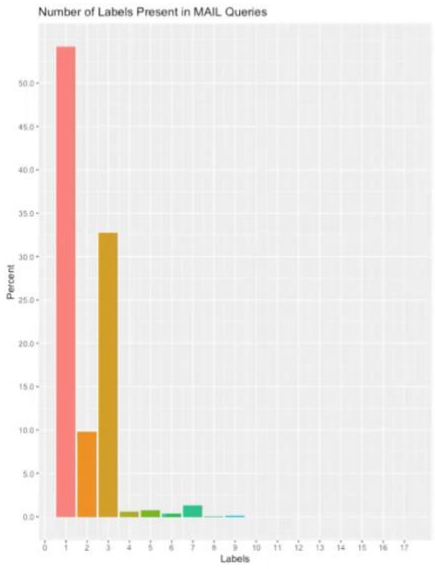


distribution and growth on the graph on the left is taken from the last day of 2020 so December 31, here we aggregated the IPS up two distinct autonomous systems or ANS for .mail at A&J on that day. We they received approximately 980 distinct ANS requesting various different .mail Strings. roughly 100 ASN makeup 85 or 87% of the traffic which is fairly diverse and suggests you will need a large outreach effort to remediate this traffic.

A lot of ISPs behind this traffic. Some using .mail internally, like an American Insurance company here.

Slide 6: Label Analysis

.MAIL Analysis :: Label Analysis



	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

	ThirdLabel	Percent
1:	wpad	19.8957231
2:	winhexbemig15	6.4311622
3:	winhexbemig16	6.3125283
4:	_ldap	3.6859604
5:	winhexbemig13	3.2862324
6:	inblrdmzftdp01	3.0225178
7:	winhexbeus105	2.9159740
8:	winhexbemig14	2.9076621
9:	winhexbeus103	2.2427082
10:	inblrprdbxpx01	2.0938492
11:	_dmarc	1.8112438
12:	inblrprdbxpx02	1.7039444
13:	isdcf01	1.6125132
14:	inblrepop01	1.4462747
15:	msoid	1.3027052
16:	winhexbeus101	1.2316760
17:	mail	1.1833157
18:	isdav	1.0813057
19:	winhexbeus106	1.0125434
20:	pop3	1.0012090
21:	plwawepo5	0.9989421
22:	ep	0.9656944
23:	imap	0.8772858
24:	adeca-sav2	0.7352274
25:	winhexbeus104	0.7261599
26:	ms1ma43	0.6770440
27:	smtp	0.6082817
28:	winhexfeus5	0.5999698
29:	oit-tanium-p01	0.5750340
30:	doc-pluto	0.5606770

ThirdLabel Percent

Month of Dec, all queries, looking at # of labels present at queries received. 57% of queries only had label mail. Middle column ranks most popular SLDs, 2nd row underscore, is another q name implementation where they've changed unnecessary labels to just a single underscore

Slide 7: 2017 vs 20121 SLD Ranking

.MAIL Analysis :: 2017 vs 2021 SLD Ranking

2020 MAIL List from A and J

	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
	SLD	Percent

2.6 MAIL, reverse order by volume (2017)

Rank	Requested string	Volume Observed	Average Daily Sources
1	.mail	8,496,910	10,941
2	system.mail	361,694	2,265
3	win.mail	357,709	1,417
4	alico.mail	350,051	796
5	al.mail	187,074	367
6	g.mail	173,779	1,054
7	yahoo.mail	145,612	1,094
8	com.mail	105,209	334
9	hot.mail	84,580	450
10	mail.mail	80,488	451
11	google.mail	55,151	263
12	company.mail	54,168	432
13	gmail.mail	53,229	238
14	navy.mail	50,687	193
15	army.mail	44,085	192
16	_tcp.mail	42,754	280
17	_sites.mail	41,395	261
18	infra.mail	38,370	61
19	net.mail	34,954	160
20	ct.mail	34,833	38
21	af.mail	34,639	133
22	aol.mail	34,228	211
23	www.mail	34,068	325
24	hotmail.mail	31,627	152
25	winus.mail	29,264	182
26	sw.mail	27,441	10
27	e.mail	26,351	218
28	receive.mail	24,005	124
29	maillocal.mail	20,768	63
30	smtp.mail	20,234	149

https://www.icann.org/en/system/files/files/octo-007-en.pdf

---

jeff Schmidt: we identified an issue where some infrastructure that was dropping .mil from queries exiting their network and so that exposed what was previously the second level domain as the top level domain which then resulted in an internet query into that top level domain. There is harm associated with that. have a sneaking suspicion that this might be related to that issue which I also know now has been fixed for a couple years. So that would explain the change in behavior, but that way when you see things related to service branch has the SL D. That that was a very specific situation that has been fixed.

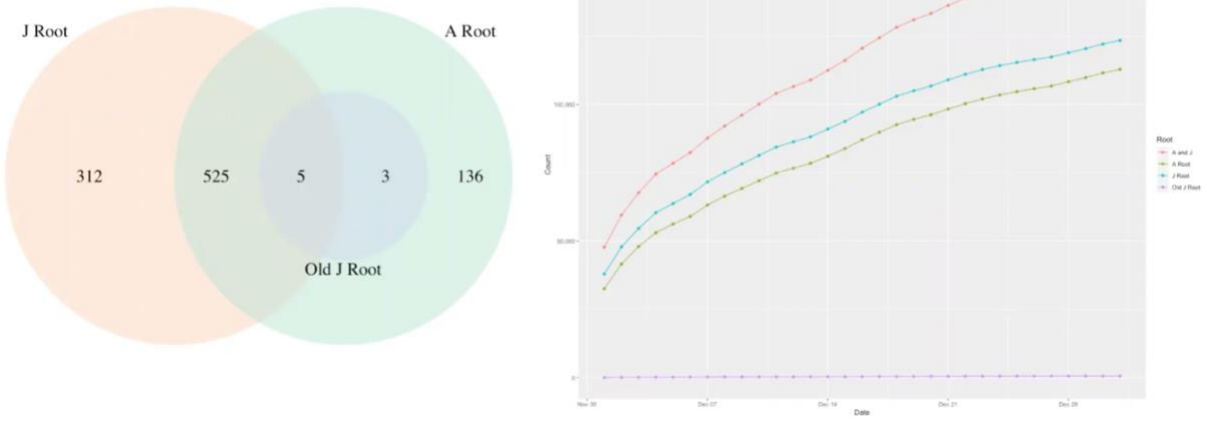
Matt: Moving forward if we make recommendations in terms of various different measurements that we should be doing to calculate and start to assess risk.**I would suggest that we include expanding the number of unique daily sources into various other network cuts either switch point for us or a sentence, specifically**

Slide 8:

Data Sensitivity Analysis

Slide 9: Root ASN Overlap and IP Growth

.MAIL Analysis :: Root ASN Overlap and IP growth

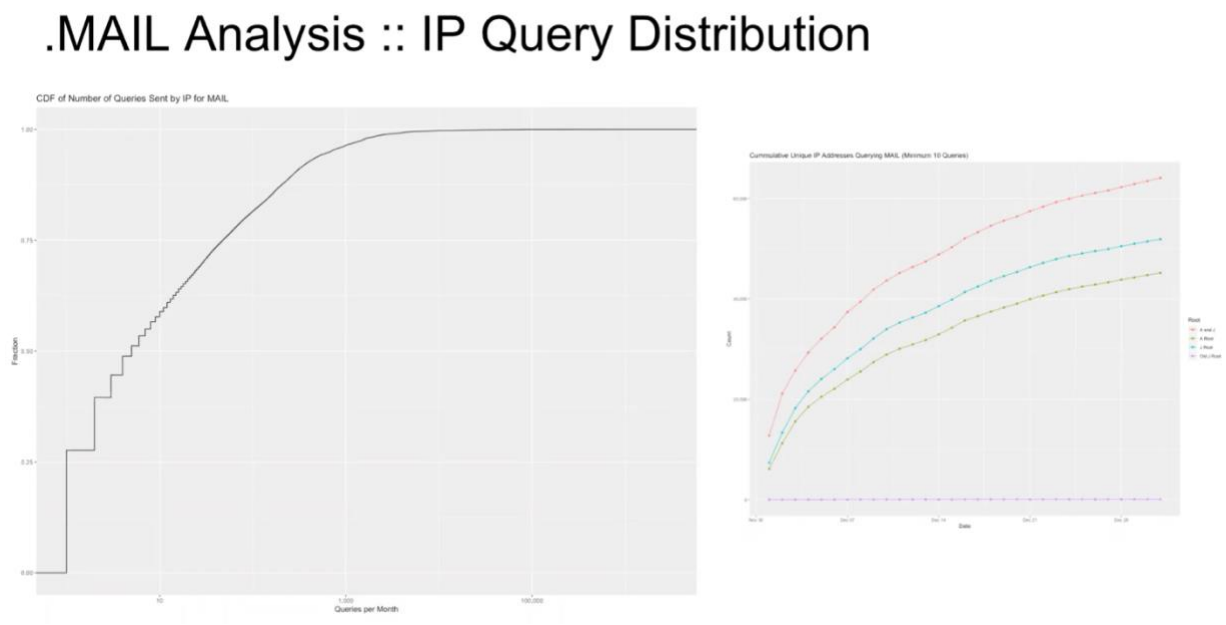


data sensitivity: How do we ensure that when you know risk assessments in the future being conducted that the data collected from whatever entity at that point in time is representative enough to show the actual or give confidence that we were actually measuring and conducting the correct risk assessment.

the graph on the left is only for the last day of the month from the 31st, and this is taking a look at which ASN sent the queries to which router. Significant specific collection point at each route

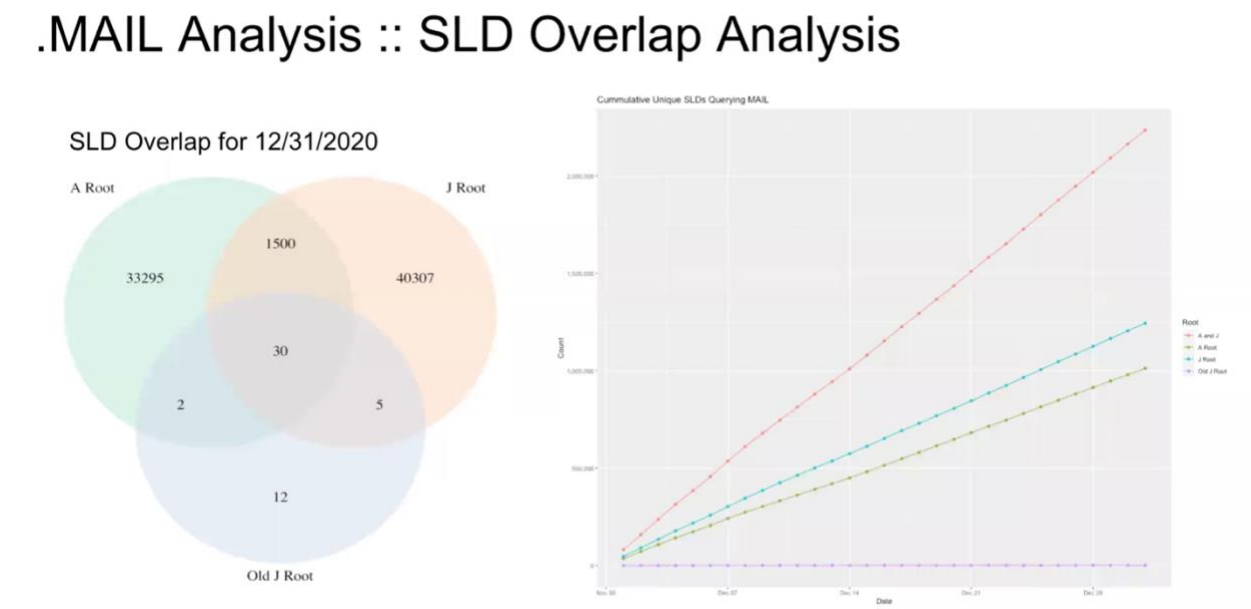
Right graph: how many unique Ips seen for .mail queries. Seeing more and more sources over time, which is a surprise.

Slide 10: IP Query Distribution



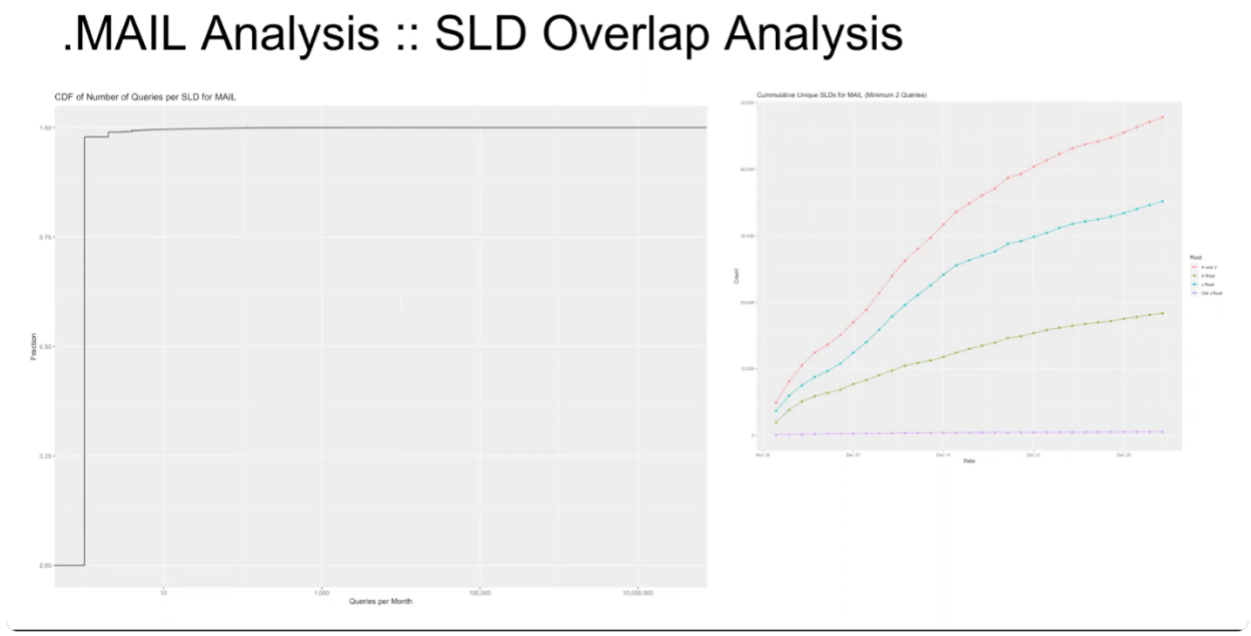
the graph on the left is looking at the cumulative distribution of traffic. So how many queries did in particular IP address sent over the course of the month. And it turns out that you know roughly 55 to 60% of them are sending less than 10 queries at four dot male domains over the entire month

Slide 11: SLD Overlap Analysis



Looking at second level domains. Min overlap on A and J….mostly unique per root. Catchment theory- each route has it's own vantage point. Graph on right is cumulative # of unique SLDs for .mail over time…straight lines, more and more unique over time

Slide 12: SLD Overlap Analysis 2



the graph on the left is looking at the number of queries a unique second level domain received over the entire month of December. 97% of these query a second level domains are only receiving one query to me that says that you're getting all of these random strings something random dot mail and you're never seeing it again.

this is possibly chromium queries that might be going through a suffix search list processing where doc mail is being attached to the random label being generated. And this is why you're seeing so many unique non overlapping domains going forward.

regarding chromium queries : in November the Chromium code base was actually modified and they've changed their behavior to how and when they push out the random domain queries to the root. And since the deployment and chromium 87 the total route server system traffic volume has decreased by 40% so you know that take that for what it is, but maybe that would change if we look at doc male again here in the next few weeks.