Signs of Triviality

Opinions, mostly my own, on the importance of being and other things.

[homepage] [blog] [jschauma@netmeister.org] [@jschauma] [RSS]

Who reads your email?

March 9th, 2023

This is the second blog post on the topic of the centralization of the internet. The first post, discussing diversity of authoritative name servers, can be found here.

According to various statistics, there are somewhere around 330 billion emails being sent every day, approximately 3.82 million per second. Who reads all these emails?

Ok, ok, *nobody* does. Who would want to? Most of it is spam anyway. But, given how personal email is, how much we rely on email for business, how useful email can be in legal discovery, and, most importantly, how -- over 40 years after RFC821 was published -- we *still* use a clear text protocol and have no realistic solution for end-to-end encryption of this private content... given all that, who *could* read that email if they wanted to? Ah, well, that's another question altogether.

The *Simple Mail Transfer Protocol* (SMTP) uses `MX` records in the DNS to identify which server(s) it should hand the mail off to. It used to be common for domain owners to run their own mail server, but it turns out that doing that well while efficiently combating spam (both incoming and outgoing), email abuse, and the ever increasing traffic volume is not that easy. And what do we do when things aren't easy? We pay somebody else to do it for us. To the cloud!

In 2023, chances are that, regardless of the domain in question, your personal and/or business email is actually handled by e.g., Google, Microsoft, Yahoo, Apple, Yandex, or, say, GMX. But even if those are your email service provider, it's also quite likely that your domain uses another layer in front of *that*, which provides spam-, malware-, filtering, and data-loss prevention (DLP) features. Popular service providers here include Proofpoint, Barracuda, Sophos, Trustwave, and some other offerings from big name companies as well as ones you likely have never heard of.

So let's take a look at which of these various companies are fronting the most domains and could thus, in theory, anyway, read your email!

Methodology

Much like I did when I looked at the NS record diversity, I went through all the gTLD zone files (again leaving out ccTLDs), extracted all second-level domains, and then went to work with nothing but my little, trusty bind9 caching resolver running on my personal VPS.[1]

For each gTLD zone file, I extracted the full list of domains within that TLD, defined as any unique label in the zone file with an NS record. This yielded a grand total of approximately 203 million domain names: > 164 million in `.com` alone, with all other gTLDs adding up to roughly 39 million domain names. For each of those domains, I then performed DNS lookups for its MX records, and a few million queries later I ended up with a whole bunch of mail server FQDNs.

A single domain may of course have multiple MX records which may or may not be in the same domain (which itself may or may not be within the original domain):

```
$ dig +short mx netmeister.org           # <--+ 1 MX
50 panix.netmeister.org.                 # <--+ within the same domain

$ dig +short mx akamai.com               # <--+ 4 MX
20 mx0b-00190b01.pphosted.com.           #    |
10 mxa-00190b01.gslb.pphosted.com.       #    | all in a different domain
10 mxb-00190b01.gslb.pphosted.com.       #    |
20 mx0a-00190b01.pphosted.com.           # <--+

$ dig +short mx twitter.com              # <--+ 5 MX
30 ASPMX3.GOOGLEMAIL.com.                #    + in two different domains
20 alt1.aspmx.l.google.com.              #    | (owned by the same org)
10 aspmx.l.google.com.                   #    |
30 ASPMX2.GOOGLEMAIL.com.                #    +
20 alt2.aspmx.l.google.com.              # <--+

$ dig +short mx whynot.coffee            # <--+ 4 MX
10 mailin.mx-hub.cz.                     #    | in four different domains
10 mailin.mx-hub.eu.                     #    | in four different TLDs
10 mailin.mx-hub.sk.                     #    |
10 mailin.mx-hub.net.                    # <--+
$
```

So we need to flatten the data a bit and reduce the individual MX servers to their second-level domain. With the help of some perl and the Public Suffix List, I mapped the approximately 30 million unique MX servers listed for the 203 million domains into around 21 million second-level domains.

So... who *does* ~~read~~ host everybody's email?

Stats by MX

No MX

As noted above, I found approximately 30 million unique mail servers, but of course not every domain *has* an MX record. In that case, SMTP assumes an "implicit MX" and attempts to deliver the mail to the IP address (if any) of the bare domain name.

As it turns out, no explicit MX record is indeed the most widely found configuration: almost 119 million domains (58% of all domains) are lacking any such resource record. Of those, 76 million (64%) do have an IP address and thus *could* at least theoretically receive mail; reversing those IP addresses again, we note that 28.8 million are AWS IPs (in the `amazonaws.com.`, `awsglobalaccelerator.com.`, and `cloudfront.net.` domains),

18 million Google's (`1e100.net.` and `googleusercontent.com.`; `34.102.136.180` is used by 12.8 million domains alone), and 7.3 million Wix's (`wixsite.com`).

That leaves around 42 million domains that do not have any means of accepting mail simply by not having either an MX record, nor an IP address. However, there are other ways that a domain owner may signal that it does not accept mail: 1.5 million (or 0.7% of all) domains have their MX set to `localhost` (and 425 to `localhost.localdomain`), which of course is a bit janky a way of telling folks not to bother you. Because this isn't quite ideal, we now have a much better way of expressing the fact that a domain does not want any mail: the *"Null MX" No Service Resource Record*, specified in RFC7505. That is, simply set an MX record with a preference number of `0` and a zero-length label (i.e., `.`):

```
$ host -t mx livemediastreaming.com
livemediastreaming.com mail is handled by 0 .
$
```

This approach appears to be marginally more popular than using `localhost`: around 2 million or just about 1% of all domains have a Null MX record set. (That approach also has the advantage that it can help in combating impersonation without having to specify an SPF policy: a receiving mail server can reject mail upon encountering an undeliverable `MailFrom`/`From` address.)

So all in all, just about 46 million domains or around 23% of all domains do not have any way of getting mail.

Number of MX Records

Now let's take a look at the ~40% (approximately 81 million) of domains *with* MX records. Most domains have between one and five mail exchange records, but of course there are outliers: 464 domains have more than ten MX records, 28 more than 20, and four domains have over 100! For example, the ever so aptly named `everymailbox.com` domain has 398 MX records, `whiteinbox.net` has 253, and `rm02.net` has 235. All of these MX records have the same priority, suggesting they are trying to aim for some DNS round-robin load balancing here.

`gaodong.com` is another outlier: 123 MX records with 117 distinct priorities, similar to `connectingdonors.net` with 59 records with unique priorities from 1 to 58.

And then there are domains that spread their MX records across multiple second-level domains, although some of them are clearly misconfigured and including what appear to be non-fqdn names as well as some that simply don't resolve at all:

```
$ host -t mx trustedomain.com
trustedomain.com mail is handled by 10 imtat4.          # these appear to be
trustedomain.com mail is handled by 5 imta6.            # non-fqdn names under
trustedomain.com mail is handled by 5 imta21.           # the trustedomain.com domain
[... 40 more records like that ...]
$ host -t mx dabafunk.xyz
dabafunk.xyz mail is handled by 10 mail.dabafunk.xyz.
dabafunk.xyz mail is handled by 0 smtp.dabafunk.xyz.
dabafunk.xyz mail is handled by 2 mail.bhargo.          # similarly, some are non-fqdn
dabafunk.xyz mail is handled by 1 smtp.wesak.
dabafunk.xyz mail is handled by 1 smtp.maitreya.
dabafunk.xyz mail is handled by 1 smtp.shamballa.
dabafunk.xyz mail is handled by 2 mail.wesak.           # but others don't resolve
dabafunk.xyz mail is handled by 1 smtp.bhargo.
dabafunk.xyz mail is handled by 2 mail.maitreya.
$
```

And my favorite: `moshelasky.net`, which set MX records for a number of completely unrelated and necessarily mutually exclusive big name domains, basically saying "go give my mail to Cisco, and if that doesn't work out, try Microsoft, Intel, Google, Yahoo... whatever":

```
$ host -t mx moshelasky.net
moshelasky.net mail is handled by 70 mail.facebook.com.
moshelasky.net mail is handled by 100 mail.thunderbird.com.
moshelasky.net mail is handled by 100 mail.yahoo.com.
moshelasky.net mail is handled by 90 mail.pirisoft.com.
moshelasky.net mail is handled by 30 mail.moshelasky.com.
moshelasky.net mail is handled by 40 mail.moshelasky.net.
moshelasky.net mail is handled by 100 mail.walla.co.il.
moshelasky.net mail is handled by 20 mail.outlook.com.
moshelasky.net mail is handled by 50 mail.intel.com.
moshelasky.net mail is handled by 80 mail.grc.com.
moshelasky.net mail is handled by 100 mail.mailchimp.com.
moshelasky.net mail is handled by 100 mail.digicert.com.
moshelasky.net mail is handled by 100 mail.noip.com.
moshelasky.net mail is handled by 100 mail.google.com.
moshelasky.net mail is handled by 60 mail.microsoft.com.
moshelasky.net mail is handled by 100 mail.windows.com.
moshelasky.net mail is handled by 10 mail.cisco.com.
$
```

Valid MX Records

But ok, let's look at the domains with reasonable MX records: In the 30 million unique servers listed, we expect to see several of the popular email and hosting providers' mail servers, but of course less popular domains will have their own MX records that are likely to be unique. In fact, almost 98% of all domains have a globally unique mail server, making only a single appearance. Of the other 380K mail servers, around 2K appear more than 1,000 times. The top 20 most frequently used mail servers here are:

Rank	# of instances	hostname	company / organization
01.	10.3 M	mailstore1.secureserver.net.	GoDaddy Hosted Mail
02.	10.3 M	smtp.secureserver.net.	
03.	9.6 M	aspmx.l.google.com.	Google
04.	9.5 M	alt1.aspmx.l.google.com.	
05.	9.5 M	alt2.aspmx.l.google.com.	
06.	6.7 M	alt3.aspmx.l.google.com.	
07.	6.7 M	alt4.aspmx.l.google.com.	
08.	3.9 M	eforward1.registrar-servers.com.	Namecheap

09.	3.9 M	eforward5.registrar-servers.com.	
10.	3.9 M	eforward4.registrar-servers.com.	
11.	3.9 M	eforward2.registrar-servers.com.	
12.	3.9 M	eforward3.registrar-servers.com.	
13.	2.7 M	aspmx2.googlemail.com.	Google[2]
14.	2.7 M	aspmx3.googlemail.com.	
15.	1.1 M	mx3.mail.ovh.net.	OVH / OVH Groupe SAS
16.	804 K	mx01.1and1.com.	IONOS / United Internet AG
17.	802 K	mx00.1and1.com.	
18.	793 K	mx4.mail.ovh.net.	OVH / OVH Groupe SAS
19.	784 K	mail.h-email.net.	Unknown / parked domains?[3]
20.	784 K	smtpin.rzone.de.	Strato AG / United Internet AG

You can see an obvious trend here: Google's mail servers are rather popular (although not the *most* popular), and of course chances are that domains that have e.g., `alt1.aspmx.l.google.com.` as one `MX` will likely have also `alt2.aspmx.l.google.com.` as a second record. This suggests that we can gain more insights by reducing them to their domain name:

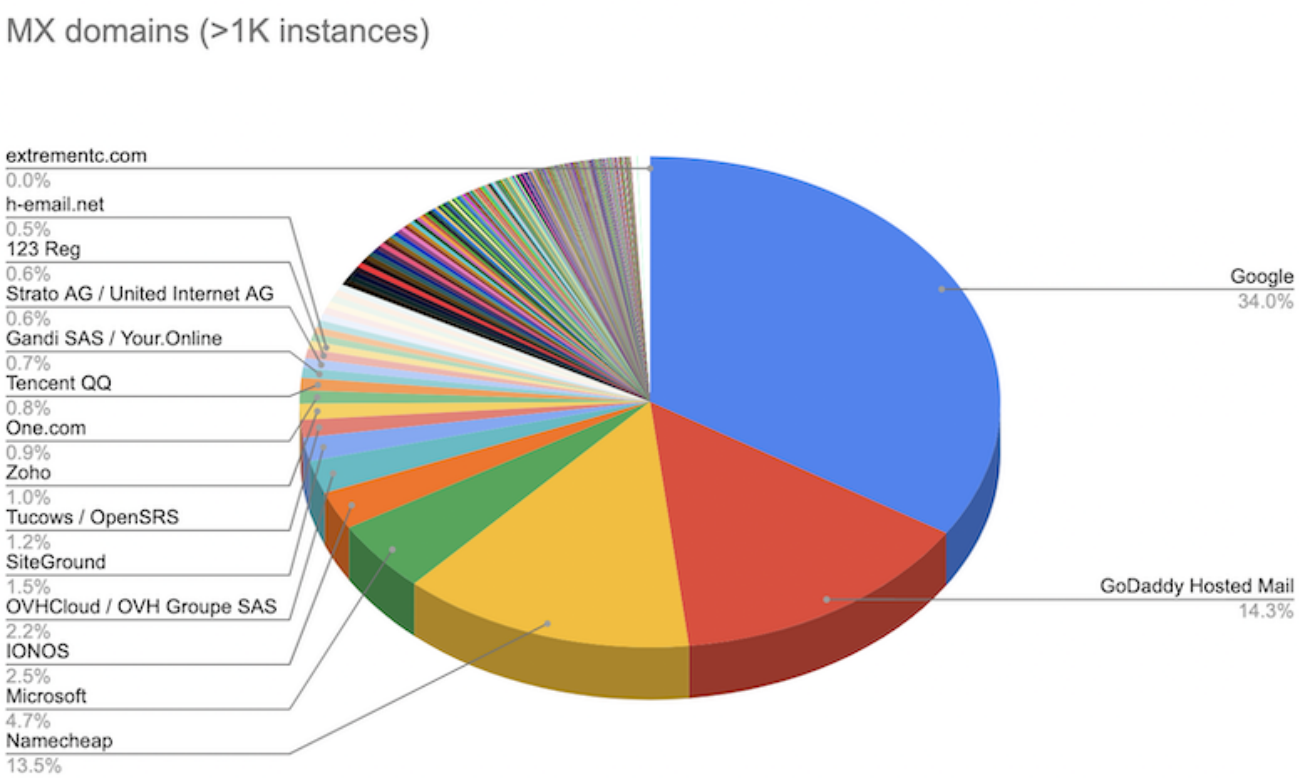
MX Record Domains

To better understand who the operators of these mail servers are, I flattened the data such that a domain that contains `MX` records pointing to, say, `aspmx.l.google.com.`, `alt1.aspmx.l.google.com.`, and `smtp.secureserver.net.` would be counted once each for the domains `google.com` and `secureserver.net.`.

This breaks down our data set to 21 million unique domains, and the top 20 domains in which we find most `MX` records are:

Rank	# of instances	domain	company / organization
01.	46.7 M	google.com.	Google
02.	22.5 M	secureserver.net.	GoDaddy Hosted Mail
03.	19.7 M	registrar-servers.com.	Namecheap
04.	7.4 M	outlook.com.	Microsoft
05.	6.9 M	googlemail.com.	Google[2]
06.	3.4 M	ovh.net.	OVH / OVH Groupe SAS
07.	2.4 M	mailspamprotection.com.	SiteGround
08.	1.8 M	hostedemail.com.	Tucows / OpenSRS
09.	1.7 M	1and1.com.	IONOS / United Internet AG
10.	1.6 M	zoho.com.	Zoho Corporation
11.	1.6 M	jellyfish.systems.	Namecheap
12.	1.3 M	one.com.	One.com
13.	1.3 M	qq.com.	Tencent QQ
14.	1.2 M	ionos.com.	IONOS / United Internet AG
15.	1 M	gandi.net.	Gandi SAS / Your.Online
16.	1 M	rzone.de.	Strato AG / United Internet AG
17.	992 K	kundenserver.de.	IONOS / United Internet AG
18.	973 K	123-reg.co.uk.	123 Reg
19.	835 K	h-email.net	Unknown / parked domains?[3]
20.	765 K	oxcs.net	EuroDNS / Datacenter Group

Obviously we can combine some of the domains by company or organization to better reflect the concentration of the mail servers. With that, we find that Google takes the lion's share of domains with about 34%, GoDaddy around 14%, Namecheap 13.5%, and Microsoft trailing behind with about 4.7%[4]:



To note: All of this is for *all* generic second-level domains but *excluding* country-code TLDs. Necessarily, this skews the findings a bit, as we'd expect e.g., European countries to use non-American service providers.

Spot-checking 100,000 domains each from `.ch`, `.fr`, and `.se` -- three of the only 17 ccTLD zone files / domain name listings I was able to access -- shows OVH and Gandi ahead of Google in `.fr`, Hostpoint AG and Infomaniak in the top 3 in `.ch`, and the Swedish One.com not surprisingly taking the top spot in `.se`, but a full analysis of all ccTLD zones would obviously be needed to get a complete view.

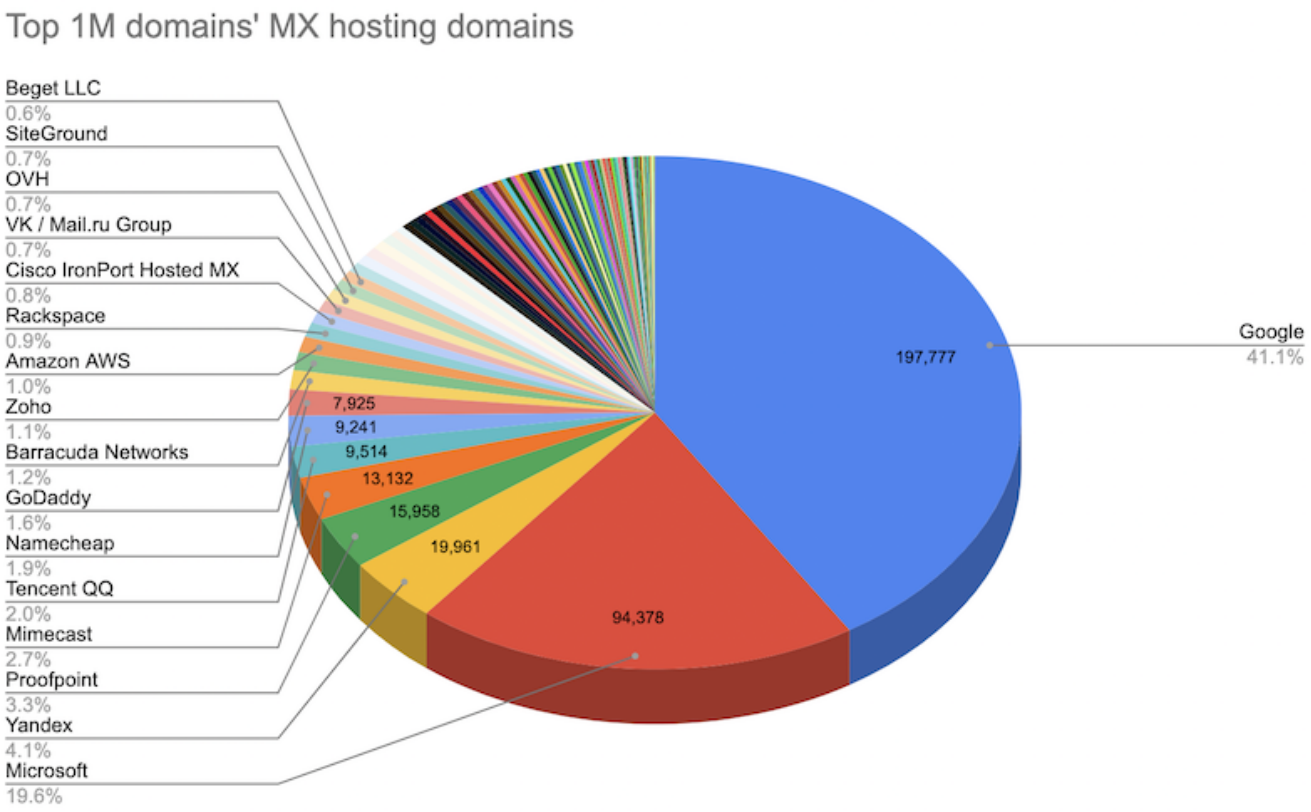
Stats for Top 1M Domains

Looking at *all* domains tells us which mail servers are listed most frequently, but that of course includes hundreds of thousands if not millions of parked domains, spam domains, one-time or dormant domains etc. So let's instead look at the Tranco Top 1 Million list and see if our distribution changes.

For those 1 million domains, we find around 433K distinct `MX` servers in 230K domains. The top 20 mail server domains there are slightly different from those for *all* domains:

Rank	# of domains out of top 1M	domain	company / organization
01.	138 K	google.com.	Google
02.	94 K	outlook.com.	Microsoft

03.	59 K	googlemail.com.	Google[2]
04.	15.8 K	yandex.net.	Yandex LLC
05.	13 K	mimecast.com.	Mimecast Limited
06.	12.8 K	pphosted.com.	Proofpoint, Inc.
07.	9.5 K	qq.com.	Tencent QQ
08.	9.2 K	registrar-servers.com.	Namecheap
09.	8 K	secureserver.net.	GoDaddy Hosted Mail
10.	5.7 K	barracudanetworks.com.	Barracuda Networks
11.	5.5K K	zoho.com.	Zoho Corporation
12.	4.7 K	amazonaws.com.	Amazon Web Services, Inc.
13.	4.4 K	emailsrvr.com.	Rackspace Technology
14.	4.1 K	yandex.ru.	Yandex LLC
15.	3.7 K	iphmx.com.	Cisco IronPort Hosted MX
16.	3.5 K	mail.ru.	VK / Mail.ru Group
17.	3.5 K	ovh.net.	OVH / OVH Groupe SAS
18.	3.4 K	mailspamprotection.com.	SiteGround
19.	3 K	ppe-hosted.com.	Proofpoint, Inc.
20.	3 K	beget.com.	Beget LLC



We observe that amongst the top 1M domains, many outsource mail not just to the big providers (Google and Microsoft together account for 60% of all!), but often add another layer of email protection via different, more specialized service provides such as Proofpoint, Barracuda Networks, or Cisco / IronPort. Those may then well also hand the mail to e.g., Google or Microsoft, further increasing their share, but that remains opaque to us from the outside.

Summary

In summary, some of the information were we able to pull out of our MX data collection includes:

- 58% of all domains (119 million) have no `MX` record (42 million of those have no IP)
- 1% of all domains (~2 million) use a RFC7505 "Null MX" (`0 .`)
- 0.7% of all domains (~1.5 million) use `localhost`
- 40% of all domains (81 million) have an `MX` record, yielding around 30 million unique records in 21 million unique domains
- 98% of those are unique, and around 380K mail servers a used by more than one domain
- ~2,000 mail servers are used by >1,000 domains each; the most frequently used `MX` records are GoDaddy's `mailstore1.secureserver.net.` and `smtp.secureserver.net.` (used by 10.6 million domains each) and Google's `aspmx.l.google.com.` (used by 9.6 million domains)
- 34% of all domains (53.7 million) use one of Google's mail servers, 14% (22.5 million) one of GoDaddy's, 13.5% (~21.3 million) one of Namecheap's
- for the Top 1M domains, over 60% use Google's (41%) and Microsoft's (20%) mail servers
- many mail protection services dominate the remainder

So all in all, the answer to the question of who can read your email pretty much boils down to -- yep -- "Google and Microsoft". Even if *your* domain doesn't use one of their mail servers, chances are that whoever you are sending mail *to* does.

To be fair: these companies are going to be doing a *much* better job at running and securing your email than you are, and outsourcing this critical functionality often makes good sense. And yet, this is another example of the continuously increasing centralization of the internet. Our businesses just like our personal online lives are concentrated in the hands of just a few companies.

March 9th, 2023

Footnotes:

[1] Performing millions of parallel DNS lookups leads to some interesting problems in different areas, which are probably worth a separate blog post all on their own.

[2] In countries where "gmail" was already trademarked, Google uses the `googlemail.com` domain. This includes e.g., the UK, Germany, Russia, and Poland.

[3] `h-email.net` appears to be a domain used primarily or exclusively for parked domains by e.g., ParkingCrew. A peculiarity of the domain is it's SPF record (`ip6:fd96:1c8a:43ad::/48 -all`), which allows only traffic on an IPv6 Unique Local Address (ULA), despite `mail.h-email.net` having only IPv4 addresses that belong to Digital Ocean and Hetzner Online GmbH.

[4] The percentages here are not quite accurate, since they are over only those mail servers that are used by 1,000 or more domains. Over all 21 million mail servers, they are reduced somewhat, but the proportional dominance of the

top domains remains.

Links:

- This blog post as a Mastodon thread
- This blog post as a Twitter thread
- Who controls the internet?
- TLDs -- Putting the '.fun' in the top of the DNS
- (All) DNS Resource Records
- Your E-Mail Validation Logic is Wrong

Elsewhere:

- Discussion on HackerNews
- Discussion on Lobste.rs
- Not that Simple: Email Delivery in the 21st Century

Previous: [AWS IAM and Cost Explorer CLI Setup]

[homepage] [blog] [jschauma@netmeister.org] [@jschauma] [RSS]