

TEU00311

What is the Internet doing to me?
(witidtm)

Stephen Farrell
stephen.farrell@cs.tcd.ie

<https://github.com/sftcd/witidtm>
<https://down.dsg.cs.tcd.ie/witidtm>

Overview

- Email Architecture
- Spam
- End-to-end email security
- A few lessons to learn along the way

Mail pre-history

- There used to be non-Internet email systems
 - uucp, X.400, other proprietary
- Internet email “won” - largely because it could connect all the others
- Fundamental basis of email is that anyone can (try) send mail to anyone, without needing permission first
 - Contrast with Twitter, FB etc. walled gardens
- Prime requirement: deliver the message, at all costs!
 - There really is no marginal cost:-)
 - That allows mail to be the basis for most other Internet/web services (e.g. so you can sign-up or do a password reset)
 - That also allows spam and other nasties
- Though there is no real per-message cost, running an effective mail server does have increasing costs
 - Often imposed by anti-spam measures and their ecosystem effects

Mail Addresses

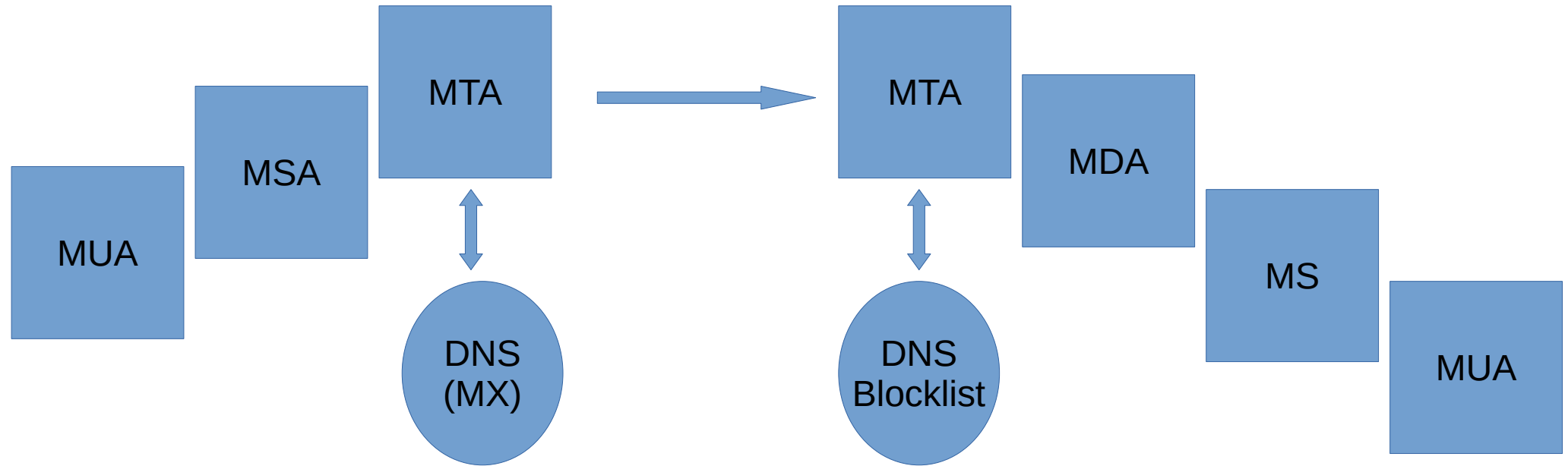
- Mail address: [user@example.com](#)
 - “user” is left-hand-side (LHS) of the “@”
 - Right Hand Side of the “@” is the email domain and is almost always a DNS domain
- RHS (plus DNS) is basically how mail gets routed to recipients – at example.com in the above
- LHS is (or should) only interpreted at the recipient’s server(s)
- Internationalisation (I18N) for canonical mail addresses can be done but support isn’t so good, so we often see “decorated” forms of the email address:
 - “provost@tcd.ie” <notthat@example.com>
 - What you see in your mail user agent will vary
 - But it’s the bit in “<>” that counts – the decoration can be forged

A helpful trick: LHS “+”

- For some mail services you can customise the LHS, e.g. if you have to give your email to some web site
- Instead of `user@example.com` you give them `user+place@example.com`
 - Other characters (not just “+”) can work sometimes
- Mail to `user+place@example.com` will still arrive in the mailbox for `user@example.com`
- That allows you to spam-folder them later and see who else got your address from them
- Can also avoid username collisions in services where email address is username and you want some form of separation, e.g. between work and personal accounts when forced to setup a web account somewhere
- Suggestion: try it out – that works for `cs.tcd.ie` but IIRC doesn't work for `tcd.ie` though that could change, I'm told it does work for many mail providers though

Email Architecture

- RFC5598 - <https://tools.ietf.org/html/rfc5598>



Component Terminology

- As a prompt for me, so I cover the basics, we'll chat about various parts of the email ecosystem
- Mail User Agent (MUA) – that's the “client” you use, which could be outlook, thunderbird, a phone-app, or a Web User Agent, where you access your mail via a normal web browser
- Mail Submission Agent (MSA) – the thing to which email is submitted by an MUA, often via the SMTP or SMTP/SUBMIT protocol
- Mail Transmission Agent (MTA) – a server that receives mail routes it to the next MTA on it's way to the recipient, and that can do other things, e.g. scan for spam/malware using the SMTP protocol, hopefully, and often, over TLS so the mail is encrypted **in-transit**
- Mail Delivery Agent (MDA) – mail delivery agents delivery of email into a message store or generating a bounce message e.g. if no such mailbox exists
- Message Store (MS) – is where your email messages live before (and after) you read 'em and is how you can see the same messages from different MUAs (e.g. laptop/phone)

Protocol Terminology

- Protocols for fetching messages from MS to MUA:
 - Internet Message Access Protocol (IMAP)
 - Earlier similar thing: Post office Protocol (POP/POP3)
- Protocol for sending messages from MUA to MTA, or between MTAs: Simple Mail Transfer Protocol (SMTP)
- Today: all of those should really be run over TLS, which is true for about 99% of messages
 - If not, you're using the wrong service!
- Not all email interactions are standardised (actually a significant %):
 - Web MUA<->MS protocol generally proprietary stuff done via Javascript
 - Outlook/MS-Exchange also does proprietary stuff

Mailing Lists

- Mailing lists are (mostly) handled via special MUAs – you send to the list address and then a list server creates new email messages (with the same content) and sends those to a set of list subscribers
- Might seem v. basic, but that is how most Internet standardisation still happens!

Message Terminology

- Message = envelope + body
- Envelope = sequence of header fields:
 - E.g. From:, Date:, Subject:, Resent-From: ...
- Body = set of MIME body-parts
 - MIME = Multimedia Internet Media Extensions is how cat images are added to emails, which were originally just ASCII bodies
- ... shows example messages...

Spam

- A “first principles” description



<https://www.theguardian.com/environment/2017/feb/13/extraordinary-levels-of-toxic-pollution-found-in-10km-deep-mariana-trench#img-1>

What is spam?

- Various acronyms:
 - Unsolicited bulk email (UBE)
 - Unsolicited commercial email (UCE)
- Spam is bad:
 - Would overwhelm utility of mail if not countered
 - Resource consumption
 - Filters, scanners etc. cost time & money
 - Malware
 - Phishing attempts

Original spam tricks

- Just send email!
 - Ahh...the naivety of it all!
- Email to list:
 - List servers got better, e.g. subscriber only with controlled subscription
- Forge headers and submit to your own server
 - Mostly outbound messages are sanity checked now
- Send via open relay
 - Used to be a lot of these, very few now

Originating Spam

- Rent a botnet for 48 hours, send some millions of spam
 - Ideally using the mail credentials of the botted device's true owner
- Can be done via single compromised hosts but those'll quickly be closed down
- Spend a few quid, setup a supposedly “legitimate” domain, wait a month or two, then send millions of spams before you're shut down
 - Cost: maybe $O(10k\text{€})$ @ high-end, but might still make a profit
- State of the art in originating spam will always be ahead of defenders so long as there's money to be made

Sometimes hard to know...

- Is this spam?

“HILARY TERM GREETINGS FROM THE COLLEGE CHAPLAINS The College Chaplains send best wishes to all, and would like to bring the following upcoming events to your attention. They are open to any students or staff members who wish to join us. ...”

- There are “legitimate” bulk email senders – for announcements, marketing, support, newsletters,...
- Those can easily start to look like spam (esp. marketing)

More spam tricks...

- Confusion:
 - accounts@paypa1.com
 - support@eboy.com
 - postmaster@boi-support.com
 - About to get worse thanks to l18N
 - security@bigbank.com
 - ^ Unicode 0430 is cryllic small 'a'
- Throwaway domains/addresses
- Zombie hosts
- Trojans
- Fake ISPs

HTML messing

- Originally mail was text based, then we added MIME, one kind of MIME body is “text/html” which turns your MUA into a dangerous web browser
- Vulnerabilities created by HTML rendering in your MUA:
 - Colour-related trickery
 - Font size 0: break words with zero width spaces
 - Presentation of nice-looking link-text not clearly-dodgy URL
 - Use of image URLs to track user
- Solution: turn off HTML rendering in your MUA!
 - Can be tricky on phones – in that case turn it off on your laptop and wait to read possibly dodgy mails there
- You do NOT have to buy into other people’s idea of what is urgent!

How much spam is there?

- Lots
 - Hard to get good figures, these are ones I've overheard
- ISP backbones:
 - 70% + of email traffic
- Delivered mail:
 - 40% + delivered
- Increasing or not?
 - Harder to tell if MTAs silently filter

Some Anti-spam techniques

- Content filtering (Bayesian, etc.)
- DNS Block Lists (SORBS, DNSBL)
- Register of known spam operators (ROKSO)
- Greylisting
- Sender Policy Framework (SPF)
- Domain Keys Identified Mail (DKIM)
- Domain Message Authentication Reporting and Conformance (DMARC)

End-to-end (e2e) email security

- That's where you use cryptography to protect messages between the sender and recipient
 - Rather than just between each “hop” on the path the mail follows
- There are two different standards for how to do that: S/MIME and PGP
 - S/MIME is mostly used inside enterprises/govt
 - PGP is mostly used by nerds or smallish groups

Deployment

- Most MUAs support s/mime or PGP either built-in or as an option
 - There are also “plug-in” products
- And mostly then *can* work together
 - I’ve used both, PGP more usable (via Thunderbird/Enigmail)
- But secure mail is not ubiquitous
 - Why?

e2e email security barriers

- Designs pre-date web user agent which changes trust model (where's the private key kept? Needs new infrastructure)
- Needs all major email service providers (yahoo, hotmail, gmail) to deploy the same thing which also needs to be implemented by all major user agent developers (microsoft, mozilla, apple, google)
- Public key retrieval needs to be fixed (doable if the above done, but a killer if not done), likely with some new PKI (doable but who's gonna pay?)
- Mail headers need to be protected as users don't get that S/MIME and PGP only protect body and not e.g. Subject, From (new enveloping protocol needed, can be done but kludgy)
- We need to unify S/MIME and PGP or pick one or we'll lose interop (it's ok if the other soldiers on for some niches)
- Users don't care much, so it has to be entirely transparent for them (needs significant UI work, co-ordinated across MUAs and significant web-UAs)

What can you do?

- Target diversity – don't all use the same service(s)
- Use LHS “+” trick if it works for your service
- Don't react immediately
- Take **much** longer to write mails than to read mails
 - Always re-read before you send
- Don't render mails as HTML if your MUA allows that
 - It's ok to wait until you're using a laptop to process mail
- Don't assume any names/links displayed to you are real
- Treat all mails and especially attachments with caution