Email Content Analysis

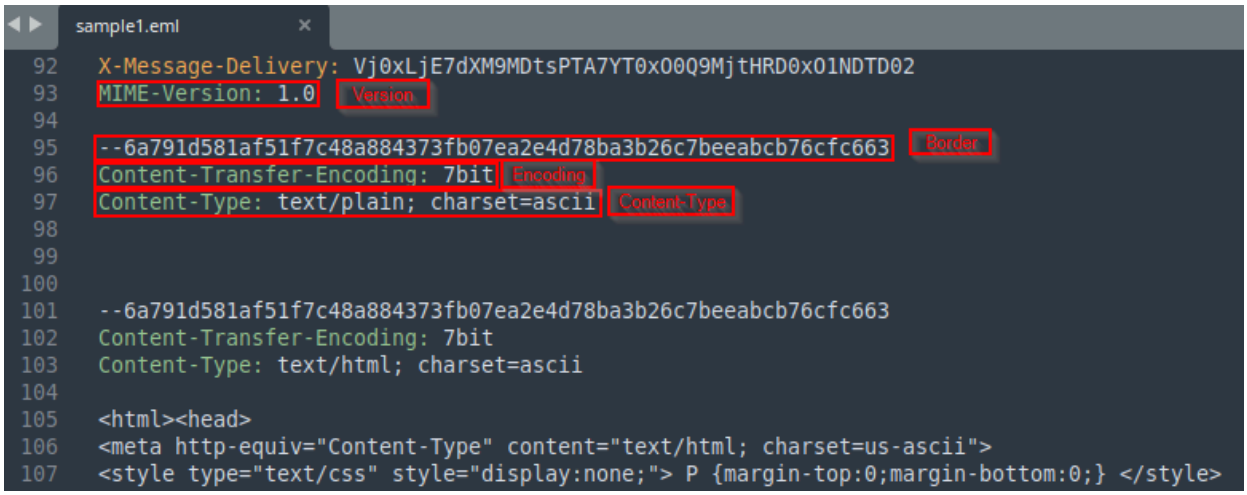
Introduction

While header analysis reveals technical information about an email's journey, content analysis focuses on the actual message body and social engineering tactics. This section examines MIME structure, encoding methods, and how attackers use various obfuscation techniques to bypass spam filters. Understanding these methods helps identify sophisticated phishing attempts that may look convincing in email clients but reveal suspicious patterns in their raw format.

What to look for

Are there signs of social engineering tactics?
How does the email appear when viewed in an email client?
What language is used, how is the grammar and formatting?

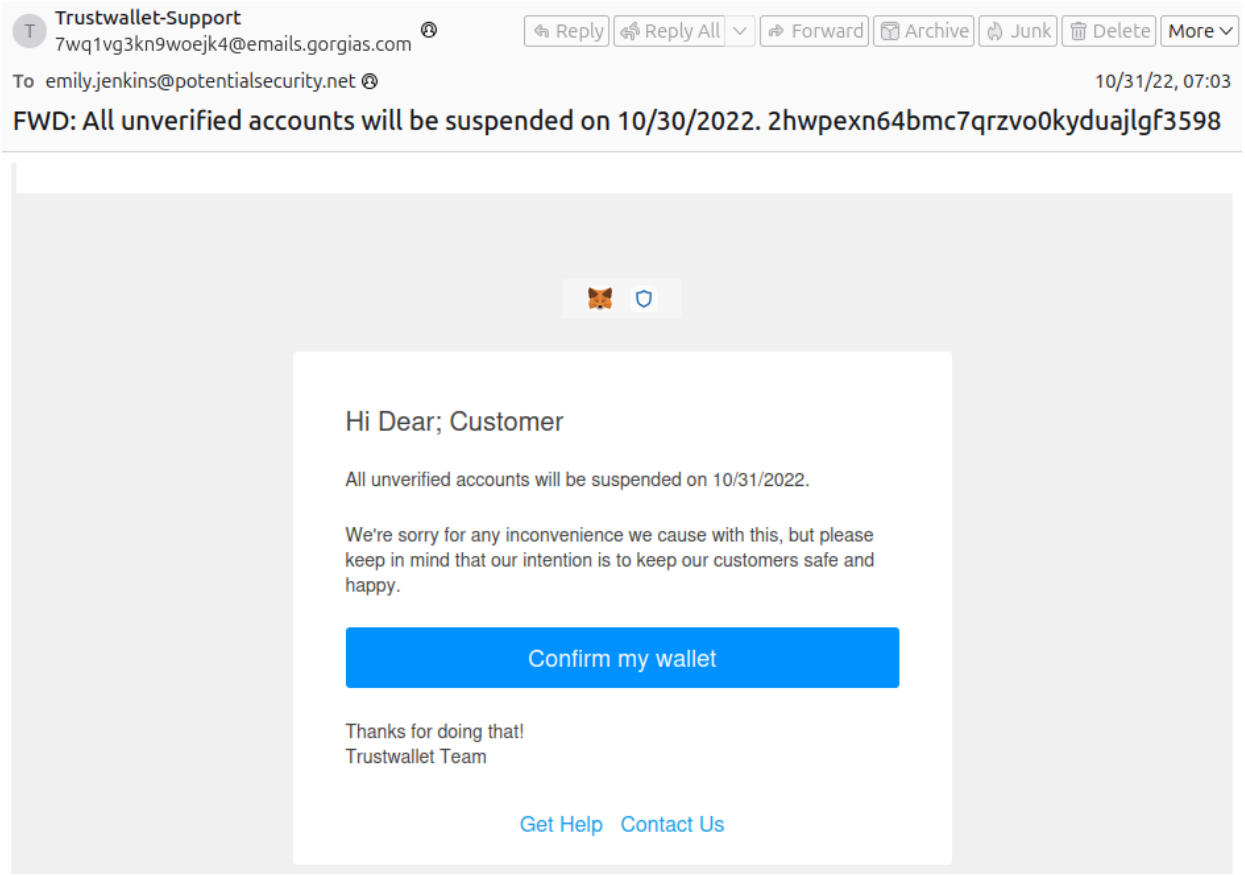
In an email file, the actual body of an email consists of the MIME (Multipurpose Internet Mail Extension) parts.



MIME-Version: Gives the version of MIME
Content-Transfer-Encoding: How was the content encoded for transmission? 7bit means no encoding was applied.
Content-Type: Specifies the kind of content, such as plain or html

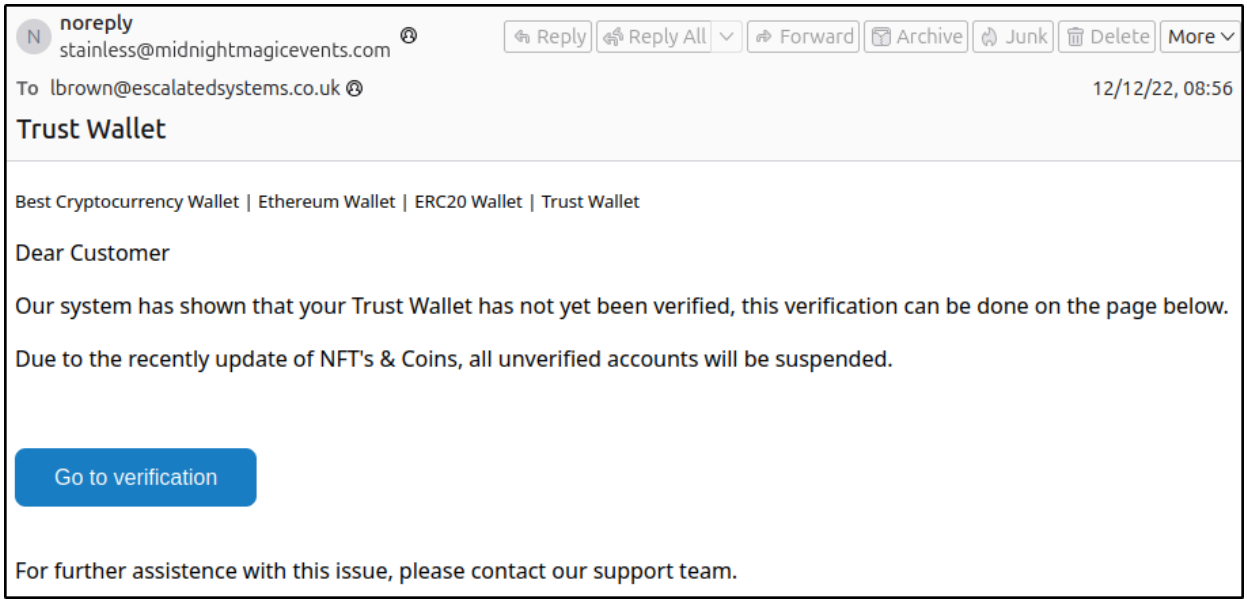
MIME boundary: Each MIME part is separated by a boundary, which starts with - -
Emails often have two versions, plain text or html, though in the above example, the plain text version is blank. Most email clients let you change the email version using the View menu.

Example 1:



- The deadline is on the day the email was sent, so less than 24 hours, which could be social engineering a sense of urgency.
- When Googling this company, we find that the company spells their name as Trust Wallet, not Trustwallet.
- And there are some formatting and grammar issues like “Hi Dear; Customer”.

Example 2:



N noreply
stainless@midnightmagicevents.com

Reply | Reply All | Forward | Archive | Junk | Delete | More

To lbrown@escalatedsystems.co.uk

12/12/22, 08:56

Trust Wallet

Best Cryptocurrency Wallet | Ethereum Wallet | ERC20 Wallet | Trust Wallet

Dear Customer

Our system has shown that your Trust Wallet has not yet been verified, this verification can be done on the page below.

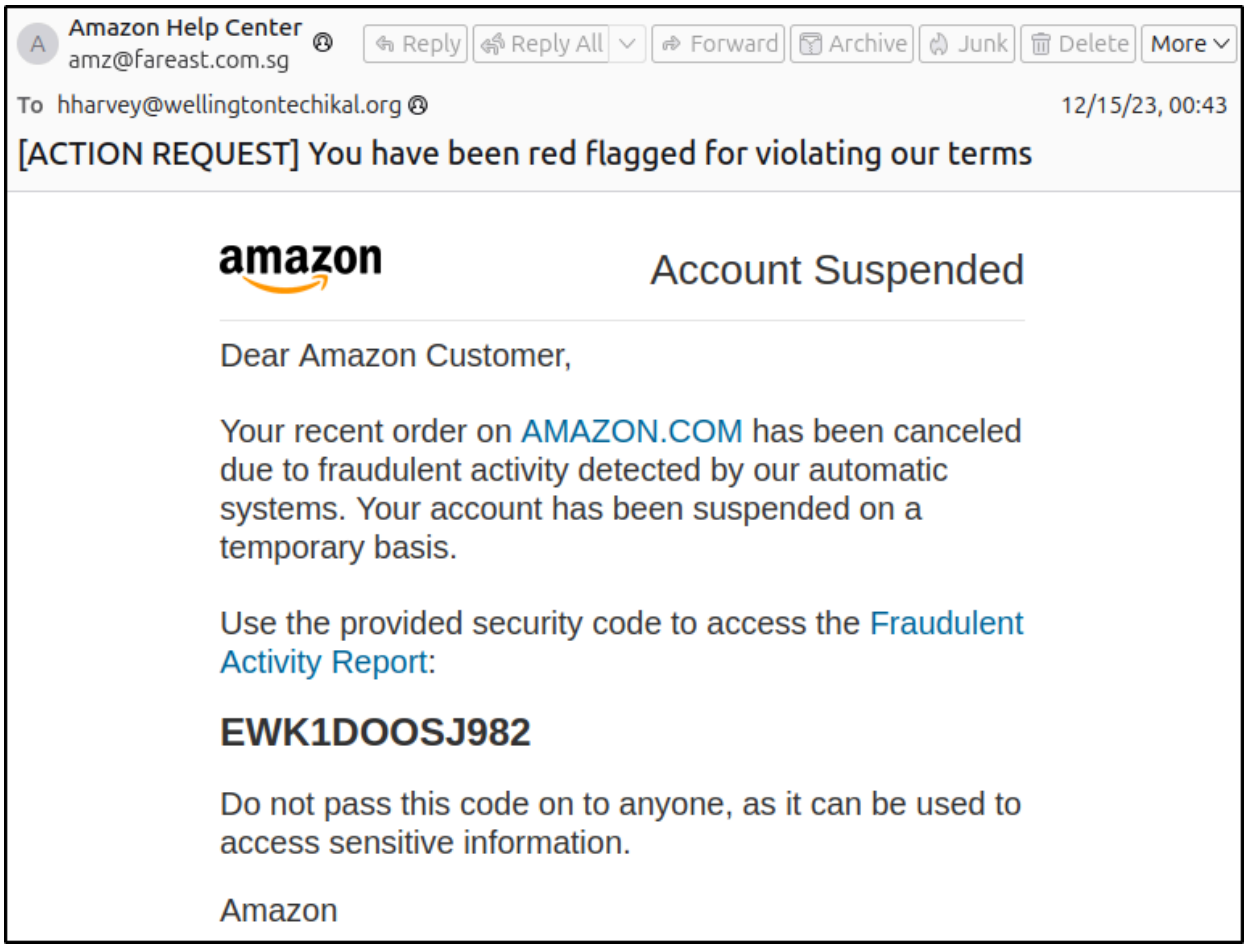
Due to the recently update of NFT's & Coins, all unverified accounts will be suspended.

Go to verification

For further assistence with this issue, please contact our support team.

- Generic greeting
- Grammar: NFT's instead of NFTs. Assistence instead of assistance. Recently instead of recent

Example 3:



A Amazon Help Center
amz@fareast.com.sg

Reply | Reply All | Forward | Archive | Junk | Delete | More

To hharvey@wellingtontechikal.org 12/15/23, 00:43

[ACTION REQUEST] You have been red flagged for violating our terms

amazon

Account Suspended

Dear Amazon Customer,

Your recent order on AMAZON.COM has been canceled due to fraudulent activity detected by our automatic systems. Your account has been suspended on a temporary basis.

Use the provided security code to access the Fraudulent Activity Report:

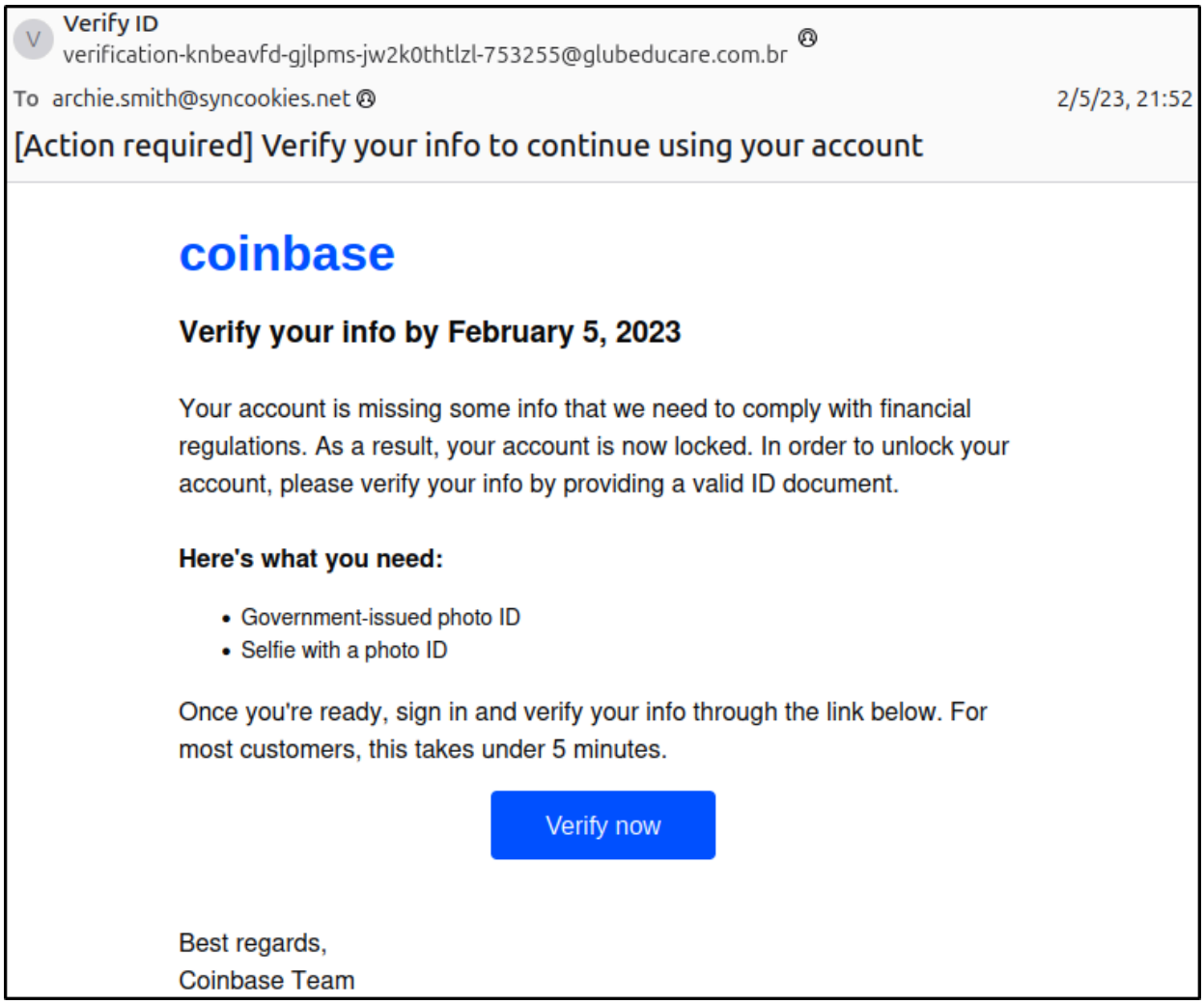
EWK1DOOSJ982

Do not pass this code on to anyone, as it can be used to access sensitive information.

Amazon

- All the URLs are non-Amazon URLs and a Non-Amazon email domain was used.
- However, the email body itself, which is what this section is about, is very convincing.

Example 4:



Verify ID
verification-knbeavfd-gjlpms-jw2k0thtlzl-753255@glubeducare.com.br

To archie.smith@syncookies.net 2/5/23, 21:52

[Action required] Verify your info to continue using your account

coinbase

Verify your info by February 5, 2023

Your account is missing some info that we need to comply with financial regulations. As a result, your account is now locked. In order to unlock your account, please verify your info by providing a valid ID document.

Here's what you need:

- Government-issued photo ID
- Selfie with a photo ID

Once you're ready, sign in and verify your info through the link below. For most customers, this takes under 5 minutes.

Verify now

Best regards,
Coinbase Team

Again, there are issues with the sender email domain and the URLS, but the email body looks convincing. When looking at the email text file there is only 1 MIME part, and it is in base64:

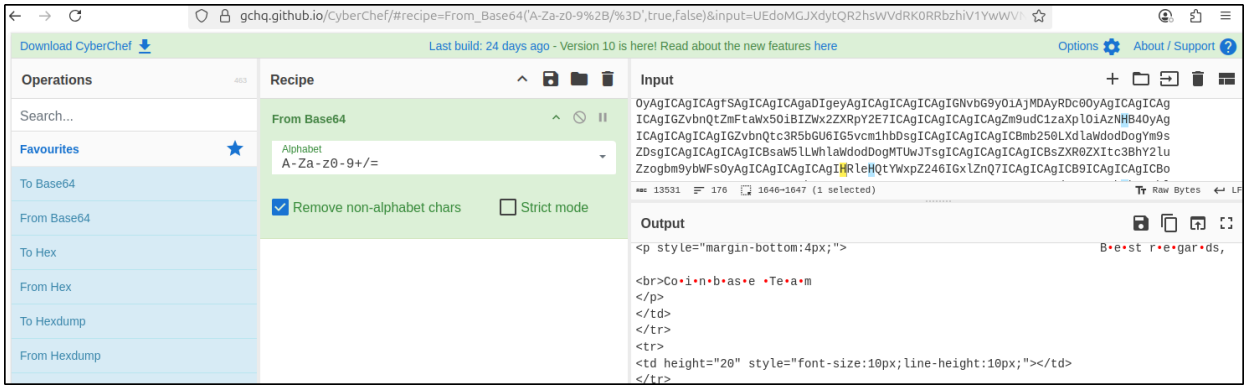
```
MIME-Version: 1.0

--alt-f15622523da4f45a4773829647be7e19
Content-Type: text/html; charset=utf-8
Content-Transfer-Encoding: base64

PGh0bWw+PGhlYWQ+DQo8bWV0YSBodHRwLWVxdWl2PSJDb250ZW50LVR5cGUiIGNvbnRlbnQ9InRl
eHQvaHRtbDsgY2hhcnNldD11dGYtOCI+CiAgICAgICAgPHRpdGxlPjwvdGl0bGU+ICAgICAgICAK
ICAgICAgICA8bWV0YSBuYW1lPSJ2aWV3cG9ydCIgY29udGVudD0id2lkdGjvv712aWNlLXdpZHRo
LCBpbml0aWFsLXNjYWxlPTEiPiAgICAKICAgICAgICA8bWV0YSBodHRwLWVxdWl2PSJYLVVBLUNv
bXBhdGlibGUiIGNvbnRlbnQ9IklF77+9Z2UiPiAgICAKICAgICAgICA8c3R5bGUgdHlwZT0idGV4
dC9jc3MiPiAgICAgICAgdGFibGUsICAgICAgICB0ZCwgICAgICAgIHRoLCAgICAgICAgdHIgeyAg
```

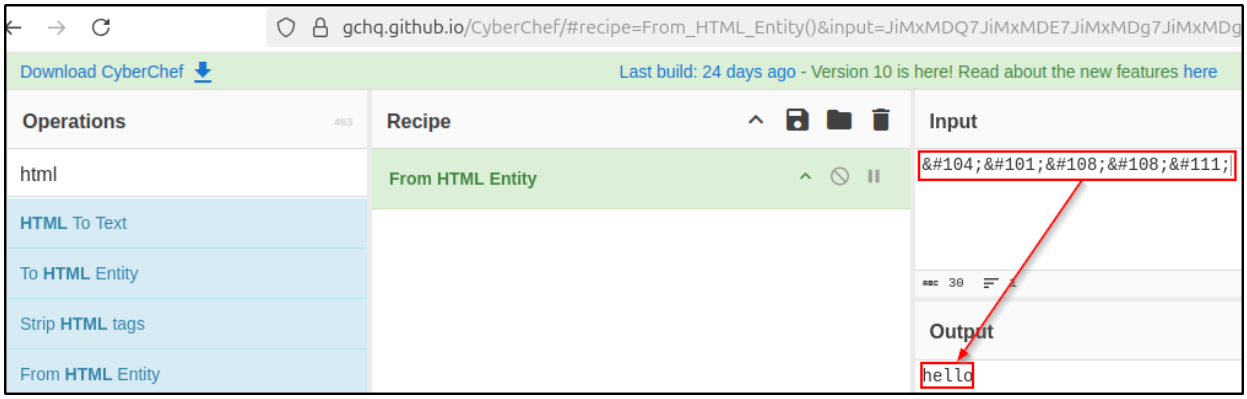
Base64 encoding

Base64 is not in itself unusual, but is often used to evade weaker spam filters. Base64 can be decoded with something like https://gchq.github.io/CyberChef



HTML entities

Another way attackers might hide something from a weak spam filter is by using html entities, which can also be decoded or encoded with CyberChef:

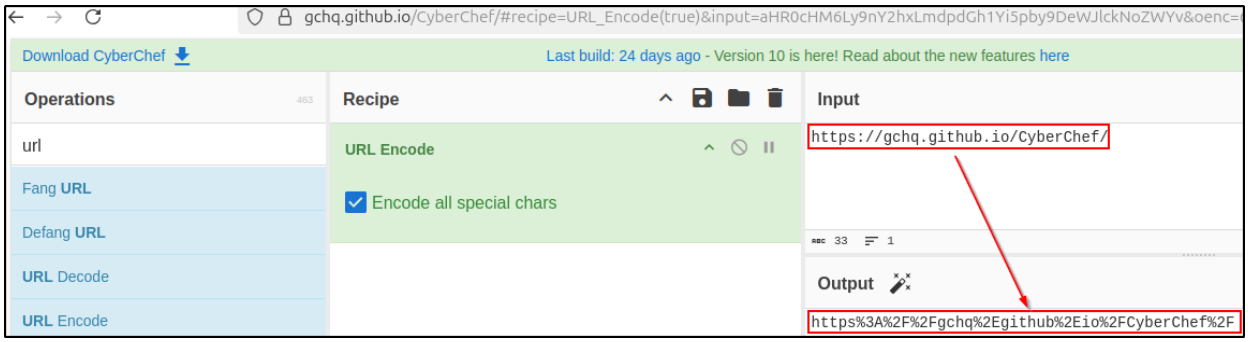


As you can see, HTML entities often have the format of &#{number};, like h being h in the above screenshot. But they can also have the format of &{entity name};, such as <, which is the lesser than symbol <

URL encoding

In the same way characters can be encoded into HTML entities, they can also be encoded as URL encoded entities. These start with a percentage sign. For example, %20 is the space character.

This can also be encoded and decoded using CyberChef:

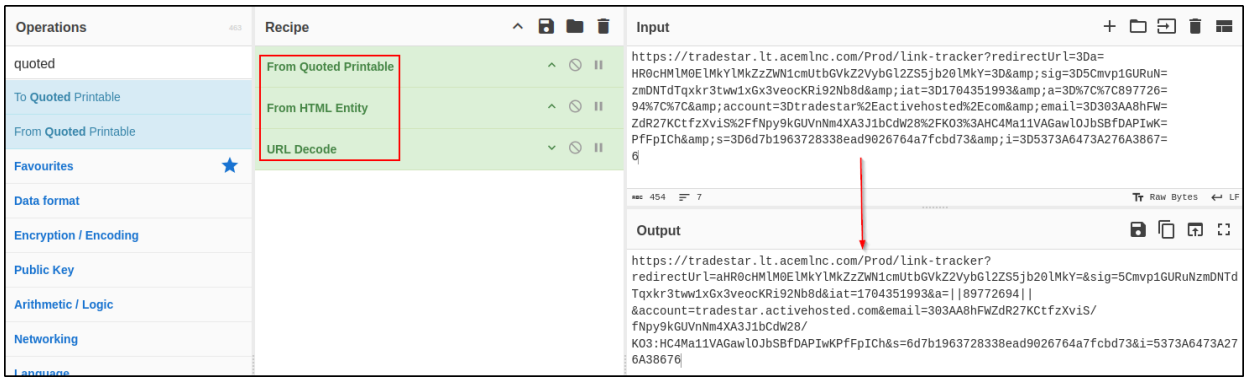


Example 5:

The fifth example email looks fine when looking at it from an email client, but when looking at the text in the eml file, the body of the fifth sample email has the URL encoded entity %2E, which is a period character, as well as HTML encoded entities such as &amp, which is the ampersand character. Also, the Content-Transfer-Encoding header specifies that this is Quoted Printable encoded, so there were 3 encoding techniques used at the same time.

```
                                                        <div class=3D"yiv15=
05425762button-container" style=3D"line-height:inherit;padding-top:10px;pad=
ding-right:10px;padding-bottom:10px;padding-left:10px;" align=3D"center">
                                                            <a style=3D"lin=
e-height:inherit;text-decoration:none;display:inline-block;color:#ffffff;ba=
ckground-color:#41CCB4;border-radius:4px;width:auto;width:auto;border-top:1=
px solid #41CCB4;border-right:1px solid #41CCB4;border-bottom:1px solid #41=
CCB4;border-left:1px solid #41CCB4;padding-top:10px;padding-bottom:10px;fon=
t-family:Arial, 'Helvetica Neue', Helvetica, sans-serif;text-align:center;"=
 href=3D"https://tradestar.lt.acemlnc.com/Prod/link-tracker?redirectUrl=3Da=
HR0cHMlM0ElMkYlMkZzZWN1cmUtbGVkZ2VybGl2ZS5jb20lMkY=3D&amp;sig=3D5Cmvp1GURuN=
zmDNTdTqxkr3tww1xGx3veocKRi92Nb8d&amp;iat=3D1704351993&amp;a=3D%7C%7C897726=
94%7C%7C&amp;account=3Dtradestar%2Eactivehosted%2Ecom&amp;email=3D303AA8hFW=
ZdR27KCtfzXviS%2FfNpy9kGUVnNm4XA3J1bCdW28%2FKO3%3AHC4Ma11VAGawlOJbSBfDAPIwK=
PfFpICh&amp;s=3D6d7b1963728338ead9026764a7fcbd73&amp;i=3D5373A6473A276A3867=
6" rel=3D"noopener noreferrer" target=3D"_blank"><span style=3D"line-height=
:inherit;padding-left:20px;padding-right:20px;font-size:16px;display:inline=
-block;">
                                                                <span s=
tyle=3D"font-size:16px;line-height:2;">Upgrade to latest version</span> </s=
pan></a>
```

You can add multiple decoders in CyberChef and see what the actual URL is:



Summary

Email content analysis involves examining the message body for social engineering tactics and technical obfuscation. Attackers often create urgency through tight deadlines, use generic greetings, and make grammar/spelling errors. They may employ encoding techniques like Base64, HTML entities, or URL encoding to evade spam filters. MIME structure analysis can reveal multiple encoding layers used simultaneously. The key is comparing how emails appear in clients versus their raw text format, as malicious emails often look convincing when rendered but show suspicious patterns in their source code.

Red Flags Checklist

- ☐ Urgent deadlines (especially same-day or within 24 hours)
- ☐ Generic greetings ("Dear Customer", "Hi Dear")
- ☐ Grammar and spelling errors in professional communications
- ☐ Company name misspellings or formatting inconsistencies
- ☐ Base64 encoding without clear legitimate purpose
- ☐ Multiple encoding layers (Base64 + HTML entities + URL encoding)
- ☐ Excessive use of HTML entities for common characters
- ☐ URL encoding in unexpected contexts
- ☐ Content-Transfer-Encoding set to unusual methods
- ☐ Blank or minimal plain text MIME parts when HTML version exists
- ☐ Suspicious formatting that differs between email client and raw view