

20BCP055

Python script for email forensics

Extraction of Information like,

- Header Analysis
- Server investigation
- Network Device Investigation
- Sender Mailer Fingerprints

from EML files

```
import email
import os
import re

def extract_email_info(eml_file):
    with open(eml_file, 'r', encoding='utf-8') as eml:
        msg = email.message_from_file(eml)

        sender = msg['From']
        recipient = msg['To']
        subject = msg['Subject']

        body = ""
        if msg.is_multipart():
            for part in msg.walk():
                content_type = part.get_content_type()
                content_disposition = str(part.get("Content-Disposition"))
                try:
                    body += part.get_payload(decode=True).decode()
                except Exception as e:
                    pass
        else:
            body = msg.get_payload(decode=True).decode()

        headers = dict(msg.items())
        sender_ip = extract_sender_ip(headers)
        server_info = get_server_information(headers)
        mailer_fingerprint = extract_mailer_fingerprint(headers)

    return {
        'Sender': sender,
        'Recipient': recipient,
        'Subject': subject,
        'Body': body,
        'Headers': headers,
        'Sender_IP': sender_ip,
        'Server_Info': server_info,
        'Mailer_Fingerprint': mailer_fingerprint
    }

def extract_sender_ip(headers):
    received_headers = headers.get("Received", "")
    ip_pattern = r'\b\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}\b'
    match = re.search(ip_pattern, received_headers)
    if match:
        return match.group()
```

```

    return None

def get_server_information(headers):
    server_info = {}
    received_headers = headers.get("Received", "").split('\n')
    for header in received_headers:
        server_match = re.search(r'from\s+(\S+)', header)
        if server_match:
            server = server_match.group(1)
            if server not in server_info:
                server_info[server] = 1
            else:
                server_info[server] += 1
    return server_info

def extract_mailer_fingerprint(headers):
    user_agent = headers.get("User-Agent", "")
    x_mailer = headers.get("X-Mailer", "")
    return {
        "User-Agent": user_agent,
        "X-Mailer": x_mailer
    }

if __name__ == "__main__":
    eml_file_path = 'path/to/your/eml/file.eml'

    if os.path.exists(eml_file_path):
        email_info = extract_email_info(eml_file_path)
        print("Email Information:")
        for key, value in email_info.items():
            print(f'{key}: {value}')

        print("\nSender IP:", email_info['Sender_IP'])

        print("\nServer Information:")
        for server, count in email_info['Server_Info'].items():
            print(f'{server}: {count} times')

        print("\nMailer Fingerprint:")
        for key, value in email_info['Mailer_Fingerprint'].items():
            print(f'{key}: {value}')
    else:
        print(f"File not found: {eml_file_path}")

```

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Vipul\Desktop\College\Sem 7\DF\DF9> python code.py
Email Information:
Sender: Panthi Ganatra <Panthi.Ganatra@sot.pdpu.ac.in>
Recipient: CE20 <CE20@pdpu.ac.in>, ICT20 <ICT20@pdpu.ac.in>, MTDS22
<MTDS22@pdpu.ac.in>, MTCS22 <MTCS22@pdpu.ac.in>, MTAI22 <MTAI22@pdpu.ac.in>
Subject: Campus Hiring 2024 - Infoware
Body: Dear Students,

Hope you are doing well!

Infoware are looking to hire candidates from PDEU.

About the company:
Infoware, formed by the passionate entrepreneur Mr. Hemant Agarwal, is a dynamic software development firm based in Ahmedabad, India. We help global brands discover, design, build and launch their digital presence on mobile, web and connected platforms. We create, modify and analyses the user needs and develop software solutions or customized software applications for client use with the aim of optimizing operational efficiency. Our focus is to provide high-quality and effective software development services to our clients..

For more information about the company visit https://www.infowareindia.com/

Attached is the document for all the available positions.

CTC : 4 - 6 LPA

Type of opportunity: Internship + Placement

Selection Process: Interview + Assignment

Stipend during the internship: Rs. 6000 - 8000 , depending on candidate (Stipend will start from 2nd month of training)

Category of employment : Full time

Service Agreement / Bond: One year commitment , no bond

Location of Work: Ahmedabad
```

```
Windows PowerShell
bDLYLq5EHUFyJ8Kg/wQ4JM\n      67pHhJGwV2FwZNSMoBEOuxSNT0ezHJD58/Zk1FHMPojl0f0/LqL+Lm1Ge3LvZ2dbE\n      q7XRWdnqs3phN40AlCx8CFR9efW8sn+HLiNOJGtC929
14tEgXHFVD50JmZqWUoPMK\n      jHVg==', 'ARC-Message-Signature': 'i=1; a=rsa-sha256; c=relaxed/relaxed; d=google.com; s=arc-20160816;\n      h=mime-ve
rion:content-language:accept-language:message-id:date\n      :thread-index:thread-topic:subject:cc:to:from;\n      bh=nf6gApo8qpjDtAlm5+McBIpiM4HAZhcg
24ANddo4sWU;\n      fh=cXSfH2KRdM7hbUdtllkasfqsYs1C4FxlVgPoUMoyccgc;\n      b=lpRMV33Yi6mHXid1JcWwWfS1u0/YAsVoAXs5w/L5LoA/Y572zvNSD/cpRZQz6jYv/\n
fb1zXXtVwKzcl1XH1FmLTVrSt4mdDz5HWSJp4f2q7+oMgnwrmgVHof4ZP+P8TsnZBn6\n      pc7Erc1PiRDGIphUpAZJr+Cp4RYJwCaCWmWyyvdabOEa+yDalsDmd3eildXj9/7Prci\n
C70B1dCS2iEYVWmti8Zwa8U59zMDPYvu/iCSuEUUKW/yCICoqk402t+gPZ86PLXmNhCm\n      xg6CFsaFkjEsB8313B2aum1aOfFHHVjN59y4mFR2x1T9G6tev9ai+7CGZSWt4JaXYu\n
qoAA==', 'ARC-Authentication-Results': 'i=1; mx.google.com;\n      spf=pass (google.com: domain of panthi.ganatra@sot.pdpu.ac.in designates 209.222
.82.7 as permitted sender) smtp.mailfrom=Panthi.Ganatra@sot.pdpu.ac.in;\n      dmarc=pass (p=QUARANTINE sp=QUARANTINE dis=NONE) header.from=panthi.ganatra@sot.pdpu.ac.in;\n      'Return-Path': '<Panthi.Ganatra@sot.pdpu.ac.in>', 'Received-SPF': 'pass (google.com: domain of panthi.ganatra@sot.pdpu.ac.in designates 209.222.82.7 as permit
ted sender) client-ip=209.222.82.7;', 'Authentication-Results': 'mx.google.com;\n      spf=pass (google.com: domain of panthi.ganatra@sot.pdpu.ac.in design
ates 209.222.82.7 as permitted sender) smtp.mailfrom=Panthi.Ganatra@sot.pdpu.ac.in;\n      dmarc=pass (p=QUARANTINE sp=QUARANTINE dis=NONE) header.from=panthi.ganatra@sot.pdpu.ac.in;\n      'From': 'Panthi Ganatra <Panthi.Ganatra@sot.pdpu.ac.in>', 'To': 'CE20 <CE20@pdpu.ac.in>, ICT20 <ICT20@pdpu.ac.in>, MTDS22 <MTDS22@pdpu.ac.in>,
MTCS22 <MTCS22@pdpu.ac.in>, MTAI22 <MTAI22@pdpu.ac.in>', 'CC': 'CDCell <CDCell@pdpu.ac.in>, DirectorSOT <directorsot@pdpu.ac.in>, cshod\n\t<cshod@sot.pdpu.a
c.in>, ictchod <ictchod@pdpu.ac.in>, Kaushal Shah\n\t<Kaushal.Shah@sot.pdpu.ac.in>, Debabrata Swain\n\t<Debabrata.Swain@sot.pdpu.ac.in>, Pooja Shah <Pooja.Sha
h@sot.pdpu.ac.in>,\n\t<Sonam.Nahar <Sonam.Nahar@sot.pdpu.ac.in>, Santosh Satapathy\n\t<Santosh.Satapathy@sot.pdpu.ac.in>, Ritesh Vyas <Ritesh.Vyas@sot.pdpu.a
c.in>,\n\t<Soumyashree.Panda <Soumyashree.Panda@sot.pdpu.ac.in>', 'Subject': 'Campus Hiring 2024 - Infoware', 'Thread-Topic': 'Campus Hiring 2024 - Infoware'
, 'Thread-Index': 'AdoQmbZhenVw91vLrhWvUZAqjFkmhQ==', 'Date': 'Mon, 6 Nov 2023 11:42:55 +0000', 'Message-ID': '<3F778EB5CE8D1E468368B8C7090B4EE4FA6EF60C@mb>
', 'Accept-Language': 'en-US', 'Content-Language': 'en-US', 'X-MS-Has-Attach': 'yes', 'X-MS-TNEF-Correlator': '', 'x-originating-ip': '[10.30.6.69]', 'Conte
nt-Type': 'multipart/mixed;\n\tboundary="_.004_3F778EB5CE8D1E468368B8C7090B4EE4FA6EF60Cmb_"', 'MIME-Version': '1.0', 'X-Auto-Response-Suppress': 'DR, OOF, Au
toReply', 'X-BESS-ID': '1699271992-1023308-12445-15-6', 'X-BESS-VER': '2019.1.20231024.1900', 'X-BESS-Apparent-Source-IP': '180.211.113.73', 'X-BESS-Parts':
'H4sIAAAAAAACA03NMQrDMAwf8LtozhArkMxLKqWDLCSEstuhKRRK7L4PCXQRnw\n\t/v6/KF+Gwww9bVAM8XzAkVeIp7SbW60BdnJONQHJs5t6U2IcOssA+nX7f74UUVSPDynydn\t<KXs1kwiYqrUpK3p
oB/HPvx+38z8TLWNAMWvZLzVhCkUkn3yMJRsJFRhv/4A1Zmda8\n\ttIAAAA=', 'X-BESS-Outbound-Spam-Score': '0.74', 'X-BESS-Outbound-Spam-Report': 'Code version 3.2, rule
s version 3.2.2.251948 [from \n\t<cloudscan12-193.us-east-2a.ess.aws.cudaops.com>\n\tRule breakdown below\n\tpts rule name
description\n\t-----\n\t0.00 HTML MESSAGE BODY: HTML included in message\n\t0.20 BSF_BESS_OUTBOUND META:
BESS Outbound \n\t0.54 HTML_FONT_LOW_CONTRAST BODY: HTML font color similar to background\n\t0.20 MIME_DOC_AO_NAME4 META: Custom Rule MIME_DOC_AO_NAME
4', 'X-BESS-Outbound-Spam-Status': 'SCORE=0.74 using account:ESS71827 scores of KILL_LEVEL=7.0 tests=HTML_MESSAGE, BSF_BESS_OUTBOUND, HTML_FONT_LOW_CONTRAS
T, MIME_DOC_AO_NAME4', 'X-BESS-Status': '1'}
Sender_IP: 10.30.1.35
Server_Info: {'MB.pdpu.ac.in': 1}
Mailer_Fingerprint: {'User-Agent': '', 'X-Mailer': ''}

Sender IP: 10.30.1.35

Server Information:
MB.pdpu.ac.in: 1 times

Mailer Fingerprint:
User-Agent:
X-Mailer:
PS C:\Users\Vipul\Desktop\College\Sem 7\DF\DF9> |
```

Structuring MBOX files from Google Takeout using Python

```
import mailbox
import os
import email.utils

# Directory containing MBOX files from Google Takeout
mbox_dir = 'path/to/borts/mbox/directory'
```

```

# Directory to store structured emails
output_dir = 'path/to/output/directory'

def save_email_as_file(email_message, output_dir):
    # Create a unique filename based on the email's date and subject
    date_tuple = email.utils.parsedate(email_message['Date'])
    if date_tuple:
        email_date = email.utils.formatdate(email.utils.mktime_tz(date_tuple))
    else:
        email_date = 'unknown_date'

    subject = email_message['Subject'] or 'unknown_subject'

    email_filename = os.path.join(output_dir, f"{email_date}_{subject}.eml")

    # Save the email as a .eml file
    with open(email_filename, 'wb') as email_file:
        email_file.write(email_message.as_bytes())

def structure_mbox(mbox_dir, output_dir):
    if not os.path.exists(output_dir):
        os.makedirs(output_dir)

    mbox = mailbox.mbox(mbox_dir)

    for message in mbox:
        save_email_as_file(message, output_dir)

if __name__ == "__main__":
    structure_mbox(mbox_dir, output_dir)

```