Homoglyph Detection Tool

Name: Prateek

Intern ID : 166

1. Description

This script-based tool is designed to identify suspicious URLs or domain names that contain homoglyphs—characters from different alphabets (like Cyrillic or Greek) that look similar to standard Latin letters. Homoglyphs are often used in phishing attacks to trick users into clicking on malicious links disguised as legitimate websites.

2. About the Tool

The Homoglyph Detection Tool scans a text file containing URLs/domains and uses the 'homoglyphs' Python library to determine whether any links use deceptive characters. It is primarily a command-line Python script meant for cybersecurity practitioners to analyze texts, logs, or email dumps for potential homoglyph attacks.

3. Characteristics / Features

- Detects suspicious homoglyph-based links
- Supports multiple alphabets (Latin, Cyrillic, Greek)
- Uses regex to extract URLs/domains
- Command-line based, lightweight
- Highlights suspicious links separately
- Modular and customizable for further integration

4. Type / Modules / Requirements

Type: CLI (Command-Line Interface) Python Script

Modules used:

- re (built-in)
- sys (built-in)
- homoglyphs (external; install using pip)

Requirements:

- Python 3.6+
- homoglyphs module (install via: pip install homoglyphs)
- A text file containing links to analyze

5. Usage

1. Install required package:

```
pip install homoglyphs
```

2. Create a text file (e.g., test.txt) containing the links you want to analyze.

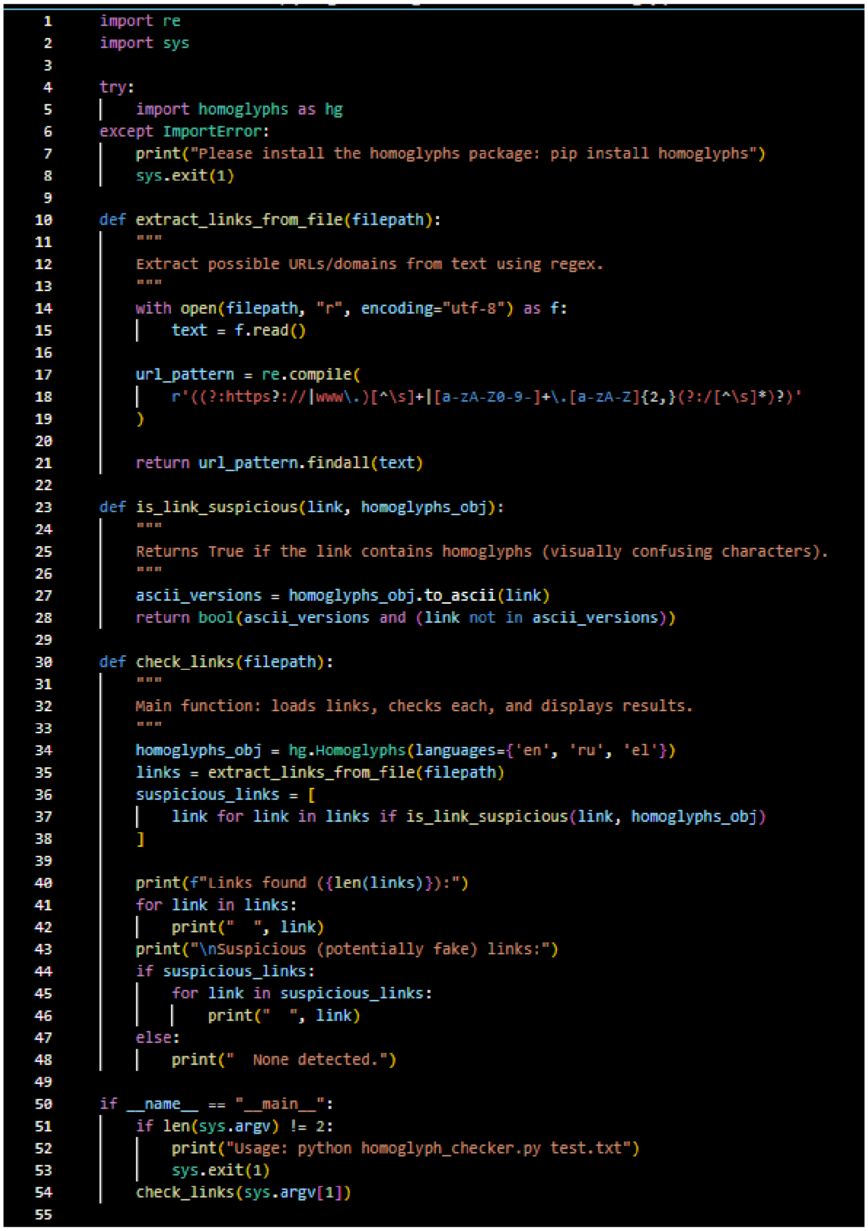
3. Run the script using:

```
python Homoglyph_Detector.py test.txt
```

4. The script will output all links found and highlight those that are suspicious.

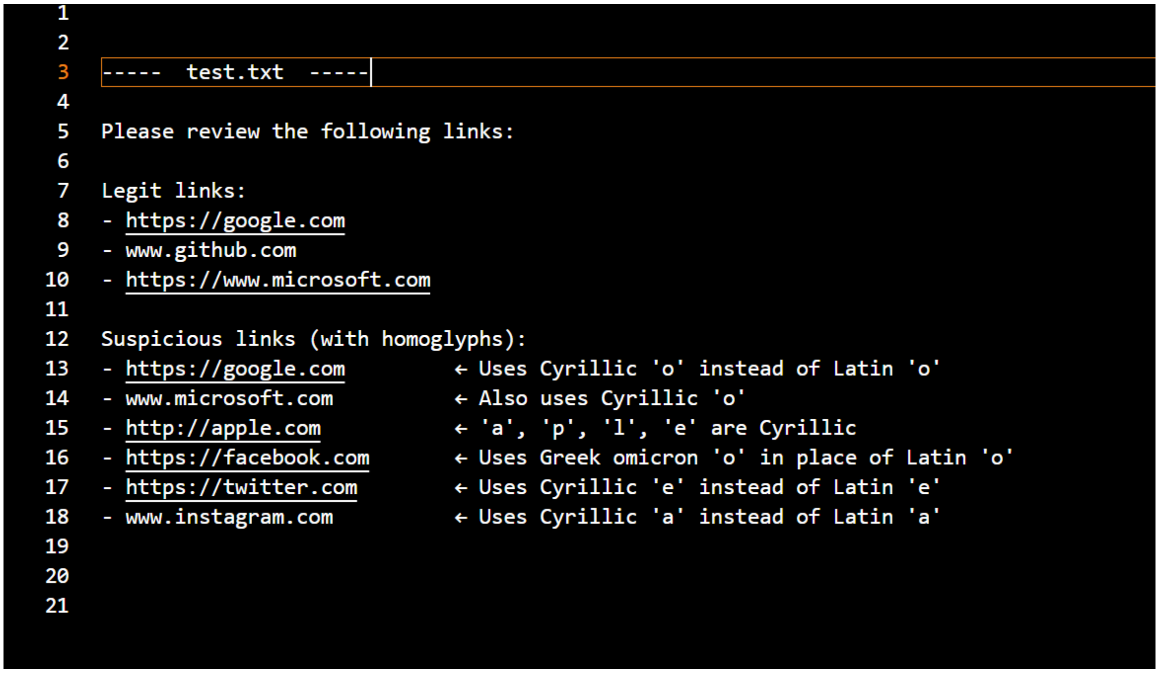
6. PoC (Proof of Concept)

Code: Homoglyph_Detector.py



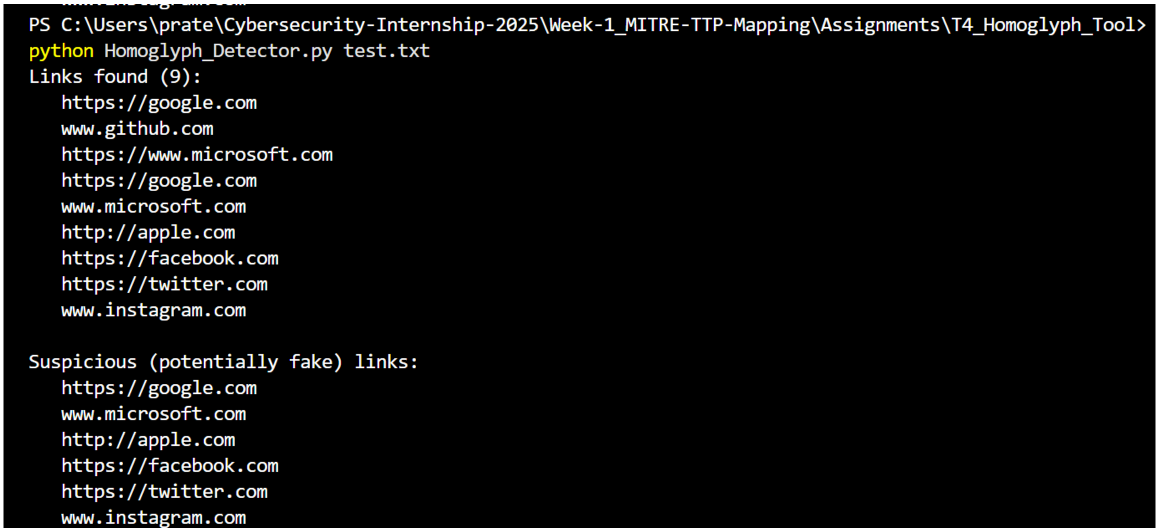
```python
import re
import sys

try:
    import homoglyphs as hg
except ImportError:
    print("Please install the homoglyphs package: pip install homoglyphs")
    sys.exit(1)

def extract_links_from_file(filepath):
    """
    Extract possible URLs/domains from text using regex.
    """
    with open(filepath, "r", encoding="utf-8") as f:
        text = f.read()

    url_pattern = re.compile(
        r'((?:https?://|www\.)[^\s]+|[a-zA-Z0-9-]+\.[a-zA-Z]{2,}(?:/[^\s]*)?)'
    )

    return url_pattern.findall(text)

def is_link_suspicious(link, homoglyphs_obj):
    """
    Returns True if the link contains homoglyphs (visually confusing characters).
    """
    ascii_versions = homoglyphs_obj.to_ascii(link)
    return bool(ascii_versions and (link not in ascii_versions))

def check_links(filepath):
    """
    Main function: loads links, checks each, and displays results.
    """
    homoglyphs_obj = hg.Homoglyphs(languages={'en', 'ru', 'el'})
    links = extract_links_from_file(filepath)
    suspicious_links = [
        link for link in links if is_link_suspicious(link, homoglyphs_obj)
    ]

    print(f"Links found ({len(links)}):")
    for link in links:
        print("  ", link)
    print("\nSuspicious (potentially fake) links:")
    if suspicious_links:
        for link in suspicious_links:
            print("  ", link)
    else:
        print("  None detected.")

if __name__ == "__main__":
    if len(sys.argv) != 2:
        print("Usage: python homoglyph_checker.py test.txt")
        sys.exit(1)
    check_links(sys.argv[1])
```

test.txt



```


-----  test.txt  -----

Please review the following links:

Legit links:
- https://google.com
- www.github.com
- https://www.microsoft.com

Suspicious links (with homoglyphs):
- https://gοogle.com          ← Uses Cyrillic 'о' instead of Latin 'o'
- www.micrоsoft.com           ← Also uses Cyrillic 'о'
- http://аррlе.com            ← 'а', 'р', 'l', 'е' are Cyrillic
- https://facebοok.com        ← Uses Greek omicron 'ο' in place of Latin 'o'
- https://twittеr.com         ← Uses Cyrillic 'е' instead of Latin 'e'
- www.instаgram.com           ← Uses Cyrillic 'а' instead of Latin 'a'



```

Output:



```
PS C:\Users\prate\Cybersecurity-Internship-2025\Week-1_MITRE-TTP-Mapping\Assignments\T4_Homoglyph_Tool>
python Homoglyph_Detector.py test.txt
Links found (9):
   https://google.com
   www.github.com
   https://www.microsoft.com
   https://google.com
   www.microsoft.com
   http://apple.com
   https://facebook.com
   https://twitter.com
   www.instagram.com

Suspicious (potentially fake) links:
   https://google.com
   www.microsoft.com
   http://apple.com
   https://facebook.com
   https://twitter.com
   www.instagram.com
```

7. Use Case Scenario

Imagine a security analyst reviewing logs or phishing email samples. The analyst can use this tool to quickly identify malicious links that may appear visually identical to trusted domains but actually redirect users to harmful websites.

8. Best Person to Use This Tool

- Cybersecurity Analysts
- Digital Forensics Investigators
- Red/Blue Team Members
- Security Researchers
- SOC Analysts

9. Advantages / Benefits of the Tool

- Easy to use and lightweight
- Accurate in detecting homoglyph-based phishing links
- Open-source and extendable
- Helpful for automating threat detection
- Supports multi-script homoglyph detection

10. Summary

The Homoglyph Detection Tool is a straightforward yet powerful utility for identifying deceptive links in textual data. By leveraging the 'homoglyphs' library, it enables early detection of phishing attempts based on visual character spoofing. With minimal setup, this tool can serve as a valuable asset in any security analyst’s toolkit.