🌓 Digisuraksha Cybersecurity Internship 2025

Intern Name: Dhairya Kumar Patel
Intern ID: 445
Date: July 2025

Homoglyph Detector Tool

🔍 What are Homoglyphs?

Homoglyphs are characters that appear visually identical or very similar but are actually different Unicode characters. Attackers exploit this by creating deceptive domains or text that look legitimate but use different character encodings.

Examples of Homoglyph Attacks:

```
apple.com (uses Cyrillic 'a' instead of Latin 'a')

google.com (uses zeros instead of 'o')

microsoft.com with Cyrillic characters
```

🚀 Features

- Domain Analysis: Detect suspicious characters in domain names
- Text Analysis: Analyze any text string for homoglyph characters
- Batch Processing: Analyze multiple domains or text strings from files
- Multiple Output Formats: Text and JSON reporting
- Unicode Analysis: Detailed character information including Unicode codes
- Risk Scoring: Percentage-based risk assessment

- Legitimate Version Generation: Automatically suggest correct versions
- Punycode Support: Display punycode representation of suspicious domains

---

📋 Requirements

Python 3.6 or higher

No external dependencies (uses only standard library)

🛠️ Installation

Clone or download the script:

git clone https://github.com/DhairyaKumarPatel/Homoglyph-Detector-Tool.git

cd Homoglyph-Detector

Make the script executable (Linux/macOS):

chmod +x homoglyph_detector.py

📖 Usage

Command Line Interface

python homoglyph_detector.py [OPTIONS]

Options

`-d, --domain`: Analyze a single domain

`-t, --text`: Analyze a text string

`-f, --file`: Analyze domains/text from a file (one per line)

`--format`: Output format (`text` or `json`)

`-o, --output`: Save results to a file

`--batch`: Enable batch processing mode for files

Examples

Analyze a Single Domain

```
python homoglyph_detector.py -d google.com

python homoglyph_detector.py -d apple.com    # Contains Cyrillic 'a'
```

Analyze Text

```
python homoglyph_detector.py -t "This is a test with homoglyphs"
```

Batch Analysis from File

```
python homoglyph_detector.py -f domains.txt --batch
```

JSON Output

```
python homoglyph_detector.py -d apple.com --format json
```

Save Results to File

```
python homoglyph_detector.py -f suspicious_domains.txt -o report.txt
```

📊 Sample Output

Text Format

```
============================================================
HOMOGLYPH DETECTION REPORT
============================================================
DOMAIN: APPLE.COM
TOTAL CHARACTERS: 9
SUSPICIOUS CHARACTERS: 1
RISK SCORE: 11.11%
STATUS: SUSPICIOUS

SUSPICIOUS CHARACTERS DETECTED:
----------------------------------------
POSITION 0: 'А'
  UNICODE: U+0430 (CYRILLIC SMALL LETTER A)
  SUGGESTED: 'A'

LEGITIMATE VERSION: APPLE.COM
PUNYCODE VERSION: XN--PPLE-43D.COM
============================================================
```