URL Checker-Tools for edu.lu (NGSOTI project)

Sam KAFAI
Student of the BTS CyberSecurity at LGK
Trainee at Restena
Cédric RENZI
DevSecOps Engineer at Restena
Denim LATIC
Security Analyst at Restena

restena
réseau · sécurité · .lu

OPEN SOURCE CONFERENCE 2025 LUXEMBOURG

1.

Introduction and some context

What is NGSOTI ? Why EdTech ?

NGSOTI: an EC-funded project to build an educative Training Infrastructure for the Next Generation (cyber)Security Operator(s)

Partners: CIRCL-LHC-Restena, uni.lu, Tenzir GmbH

Resources:

https://www.restena.lu/fr/project/ngsoti

https://github.com/ngsoti

NGSOTI

EdTech - to facilitate online resources sharing, a secured shortener service was created

edu.lu service hosted by Fondation Restena

Resources:

https://www.restena.lu/fr/service/raccourcisseur-URL

restena
réseau · sécurité · .lu

OPEN SOURCE CONFERENCE 2025 LUXEMBOURG

What is edu.lu ?

edu.lu
- an URL shortener Service hosted by Fondation Restena
- at the open disposal of the .lu educationnal community
- based on a shortener utility in the backend (SURFshort from surf.nl)

Resources:

https://github.com/SURFnet/short

Motivation for this resource choice:
- an **OpenSource** tool (Apache 2.0)
- already present in the European Education Community
- fully compatible with our Software Stack

Why an URL Checker Webservice ?

But ... **Users may** (unintentionnaly ?) **shorten URL to malicious** (or debatable ...) **content ...**

Image designed by Freepik

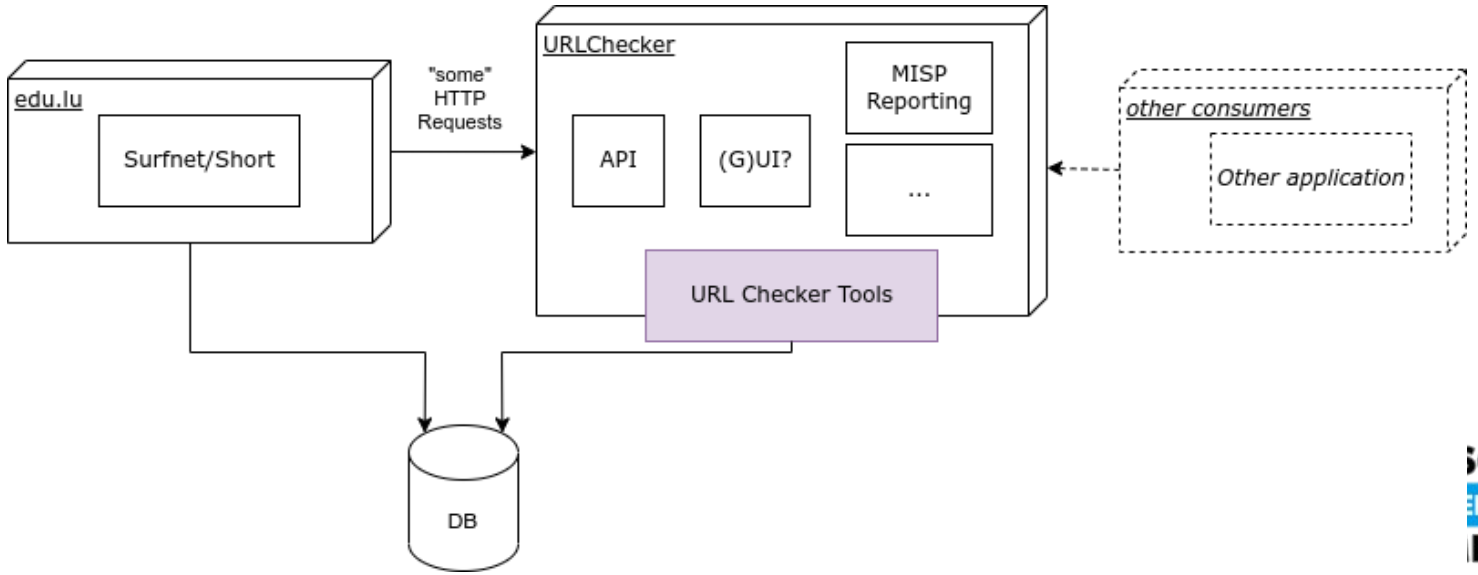
Malicious URLs = primary vectors for

Malware, Phishing, Command & Control threats

Need for some Security Scanning in the path

analyse and report any suspicious/malicious URL

Need for a flexible and modular implementation:



restena
réseau · sécurité · .lu

SOURCE
ENCE 2025
BOURG

Why an URL Checker Webservice ?

But ... **Users may** (unintentionnaly ?) **shorten URL to malicious** (or debatable ...) **content ...**

Image designed by Freepik

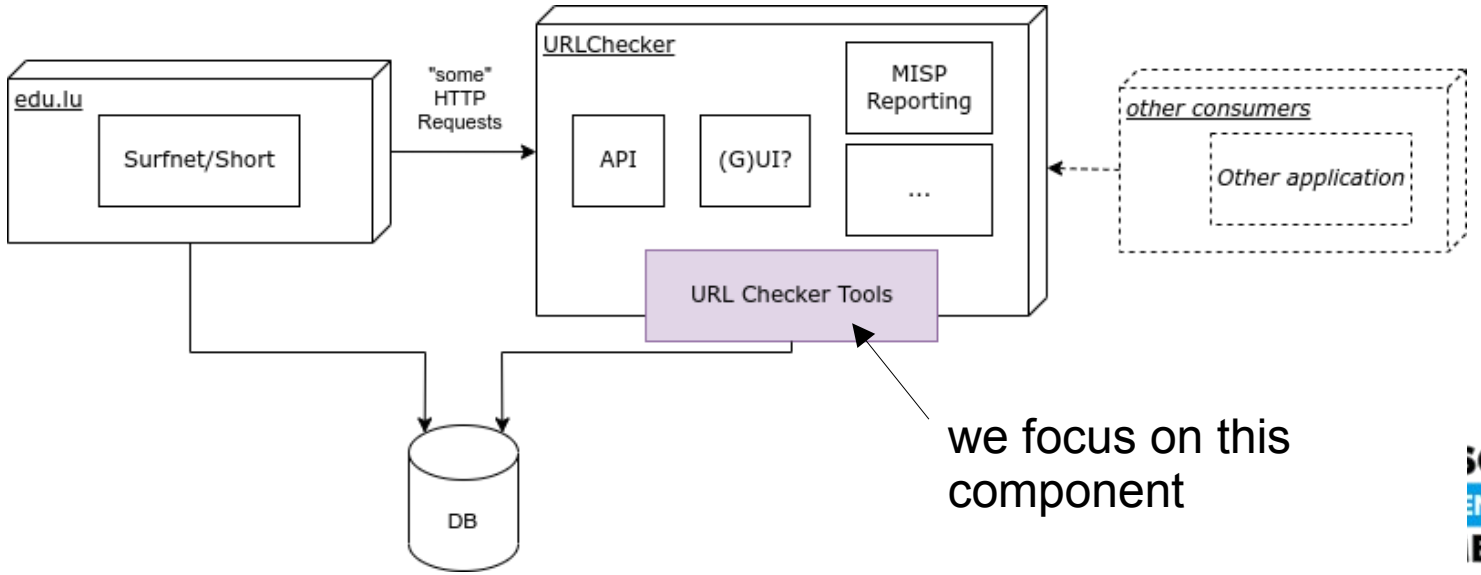
Malicious URLs = primary vectors for

Malware, Phishing, Command & Control threats

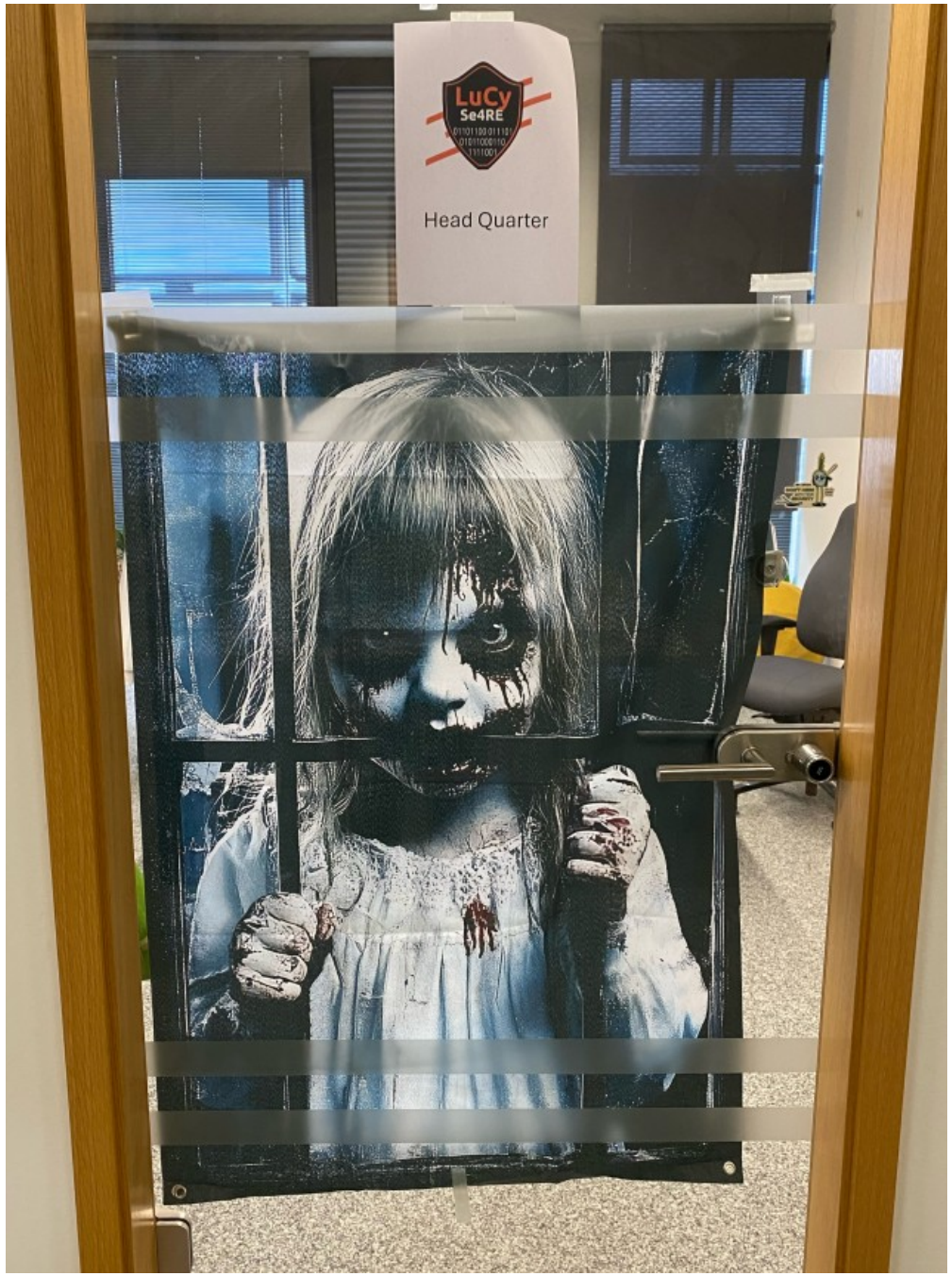
Need for some Security Scanning in the path

analyse and report any suspicious/malicious URL

Need for a flexible and modular implementation:



Threats ?



!!! DO NOT CLICK ON RANDOM URLs !!!

*If you did, do not be shy
and contact your CSIRT or us (institutions)
for support :)

2.

URL-CHECKER TOOLS - Technical details and Analysis

Problem & Motivation

Need for automated, multi-source URL analysis

Thus the URL Checker Tools was created to:

- run independantly (release source code when/where appropriate)
- be compatible with our Software Stack
- be modular and configurable
- tested and deployed On-Premises at Restena
- provide interfaces to various reknown providers
(free and/or open-source and/or commercial)
- allow the compilation of results on top of several criteria and scan outputs

Architecture Overview Workflow

Modular Python CLI tool

Inheritance-based provider system (BaseProvider)

Supports manual analysis & automation pipelines

Automated

1)Input target (URL / domain / IP) & SID

2)Provider queries (reputation checks)

3)Content scanning (YARA, downloads, redirects)

4)Unified scoring & confidence calculation

5)POC: MISP reporting

6)Session logs & JSON synthesis

CLI help text with usage instructions (1)

```
(url_checker) student@Aquarium ~/url_checker (Inheritance-based-Architecture) % uv run main.py -h
usage: main.py [-h] [--providers PROVIDERS | --all] [--format {human,json,synthesis}] [--raw] [--yara-rules [YARA_RULES ...]] [--yara-timeout YARA_TIMEOUT] [--yara-max-bytes YARA_MAX_BY
               [--workflow {fast,complete,reputation}] [--list-providers] [--providers-status] [--verbose]
               [target]

Comprehensive URL and domain threat intelligence checker

positional arguments:
  target                URL or domain to check

options:
  -h, --help            show this help message and exit
  --providers PROVIDERS
                        Comma-separated list of additional providers to run alongside baseline (whois,link_analyzer)
  --all                 Use all available providers
  --format {human,json,synthesis}
                        Output format
  --raw                 Output raw JSON (deprecated: use --format json)
  --yara-rules [YARA_RULES ...]
                        Specific YARA rule files or directories to use
  --yara-timeout YARA_TIMEOUT
                        Timeout for fetching content (default: 30s)
  --yara-max-bytes YARA_MAX_BYTES
                        Maximum bytes to scan (default: 10MB)
  --auto-scan-files     Automatically scan detected download files
  --download            Enable file download scanning (placeholder)
  --sid SESSION_ID, --session-id SESSION_ID
                        Session ID for tracking and logging
  --log                 Save output to structured log file
  --robot               Automation mode: minimal output, dual logging (.log/.dlog files). Requires --sid.
  --score               Include basic threat score result
  --score-detail        Include detailed scoring calculation breakdown (supersedes --score)
  --misp-report         Enable MISP threat intelligence reporting (create new MISP events from scan results)
  --workflow {fast,complete,reputation}
```

CLI help text with usage instructions (2)

```
  --raw                 Output raw JSON (deprecated: use --format json)
  --yara-rules [YARA_RULES ...]
                        Specific YARA rule files or directories to use
  --yara-timeout YARA_TIMEOUT
                        Timeout for fetching content (default: 30s)
  --yara-max-bytes YARA_MAX_BYTES
                        Maximum bytes to scan (default: 10MB)
  --auto-scan-files     Automatically scan detected download files
  --download            Enable file download scanning (placeholder)
  --sid SESSION_ID, --session-id SESSION_ID
                        Session ID for tracking and logging
  --log                 Save output to structured log file
  --robot               Automation mode: minimal output, dual logging (.log/.dlog files). Requires --sid.
  --score               Include basic threat score result
  --score-detail        Include detailed scoring calculation breakdown (supersedes --score)
  --misp-report         Enable MISP threat intelligence reporting (create new MISP events from scan results)
  --workflow {fast,complete,reputation}
                        Use Celery workflow for distributed scanning
  --list-providers      List all available providers and exit
  --providers-status    Show detailed provider status and exit
  --verbose             Enable verbose output

Examples:
  main.py example.com                        # Baseline scan (whois + link_analyzer)
  main.py --providers virustotal,google_sb example.com  # Baseline + additional providers
  main.py --all example.com                  # All available providers
  main.py --list-providers                   # Show available providers

  # Robot mode for automation
  main.py --robot --sid session123 example.com
  main.py --robot --misp --sid session123 example.com       # Query MISP for existing events
  main.py --robot --misp-report --sid session123 example.com # Create new MISP events
```

CLI example of automation mode

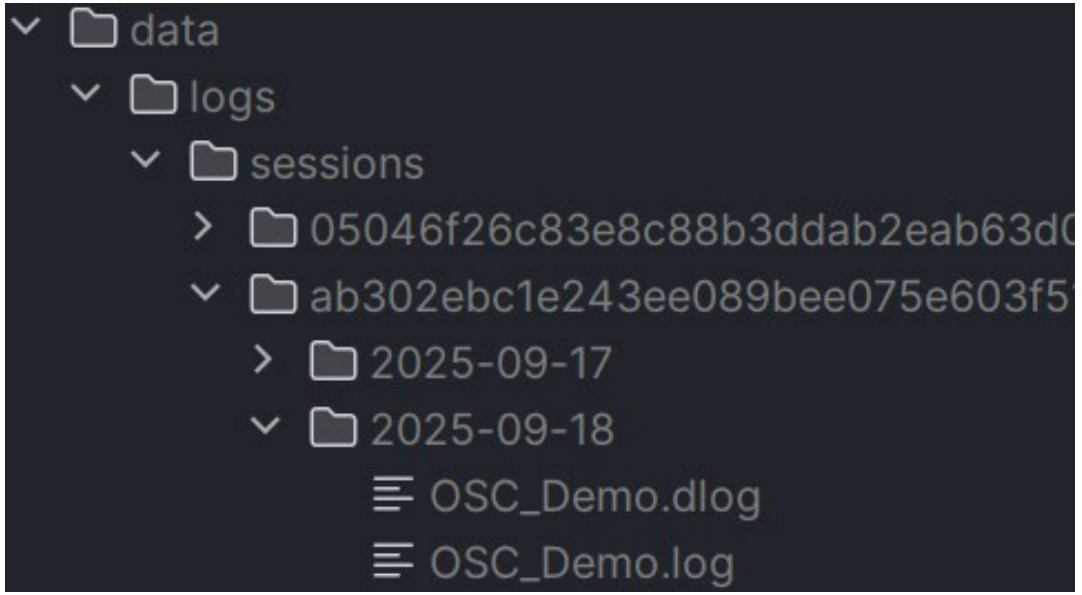
```
% uv run main.py 'http://malware.wicar.org/data/eicar.com' --robot --sid OSC_Demo --misp-report
SID: OSC_Demo
[SCAN] Checking: whois, misp, link_analyzer, whalebone, virustotal, google_sb, yara, abuseipdb
RESULT: MALICIOUS (79/100)
MISP event created: 192
[INFO] Robot logs created:
  Synthesis: data/logs/sessions/ab302ebc1e243ee089bee075e603f5146d08462e4a59a7b27270611c9f1ad82d/2025-09-18/OSC_Demo.log
  Detailed: data/logs/sessions/ab302ebc1e243ee089bee075e603f5146d08462e4a59a7b27270611c9f1ad82d/2025-09-18/OSC_Demo.dlog
% uv run main.py 'http://malware.wicar.org/data/eicar.com' --robot --sid OSC_Demo --misp-report --verbose
SID: OSC_Demo
[INFO] Target: http://malware.wicar.org/data/eicar.com
[INFO] Enabled providers: whois, misp, link_analyzer, whalebone, virustotal, google_sb, yara, abuseipdb
RESULT: MALICIOUS (79/100)

============================================================
HUMAN-READABLE SUMMARY
============================================================
✓ whois: safe
▲ misp: threat detected (39 events found)
   └─ Threat Level: malicious
✓ link_analyzer: safe
▲ whalebone: threat detected (threats: c&c, malware (max 100%)) (categories: porn) [classification: policy:1]
   └─ Threat Level: malicious
▲ virustotal: malicious (14/98 vendors) (categories: misc, known infection source, command and control, malware sites, computer security, malware, phishing)
   └─ Threat Level: malicious
▲ google_sb: threat detected
   └─ Threat Level: critical
✓ yara: safe
✓ abuseipdb: safe
[INFO] MISP event created: Event ID 193 (UUID: 20297480-7e47-4f4c-aa92-d98a950e641a)
[INFO] Robot logs created:
  Synthesis: data/logs/sessions/ab302ebc1e243ee089bee075e603f5146d08462e4a59a7b27270611c9f1ad82d/2025-09-18/OSC_Demo.log
  Detailed: data/logs/sessions/ab302ebc1e243ee089bee075e603f5146d08462e4a59a7b27270611c9f1ad82d/2025-09-18/OSC_Demo.dlog
```

Logging & Audit Trail

- Session-based logs with SHA256 hash hierarchy
- Dual-mode logs:

  .log = synthesis output

  .dlog = detailed provider-level results
- MISP event IDs embedded for traceability
- Ensures reproducibility for compliance



.log


```
{
  "session_metadata": {
    "session_id": "OSC_Demo",
    "timestamp": "2025-09-18T09:33:29.408973+00:00",
    "target_info": {
      "original": "http://malware.wicar.org/data/eicar.com",
      "normalized": "http://malware.wicar.org/data/eicar.com",
      "type": "url",
      "hash": "ab302ebc1e243ee089bee075e603f5146d08462e4a59a7b2
      "scheme": "http",
      "domain": "malware.wicar.org",
      "path": "/data/eicar.com",
      "query": ""
    },
    "misp_event": {
      "event_id": 196,
      "status": "success",
      "error": null,
      "uuid": "075b4ef2-da6d-4b1b-aae7-ae0cca759828"
    }
  },
  "synthesis": {
    "whois": {
      "is_active": true,
      "age": 14870,
      "domain": "",
      "status": "domain_found"
    },
    "providers": {
      "whois": "Age: 14870 days; Registrar: None",
      "misp": "Threats Found (45 events)",
      "link_analyzer": "Safe | 0 redirect(s)",
      "whalebone": "Malicious | Threats: c&c, malware (max 100%) | Categories: porn",
      "virustotal": "Malicious (14/98) | Categories: misc, known infection source, com
      "google_sb": "Malicious",
      "yara": "No pattern matches",
      "abuseipdb": "Low Risk (0 reports)"
    },
    "metadata_analysis": {
      "behavioral_anomalies": [],
      "risk_indicators": [],
      "confidence_score": 0.7,
      "cross_validation_status": "medium_consistency"
    },
    "link_analysis": {
      "redirects": 0,
      "domain_change": false,
      "is_blocked": false,
      "final_url": "http://malware.wicar.org/data/eicar.com",
      "dns_resolved": true,
      "security_downgrades": 0,
      "contains_shorteners": false,
      "suspicious_patterns": []
    },
    "result": {
      "verdict": "MALICIOUS",
      "threat_score": 79,
      "metadata_confidence": 0.7,
      "cross_validation_status": "medium_consistency"
    }
  }
}
```

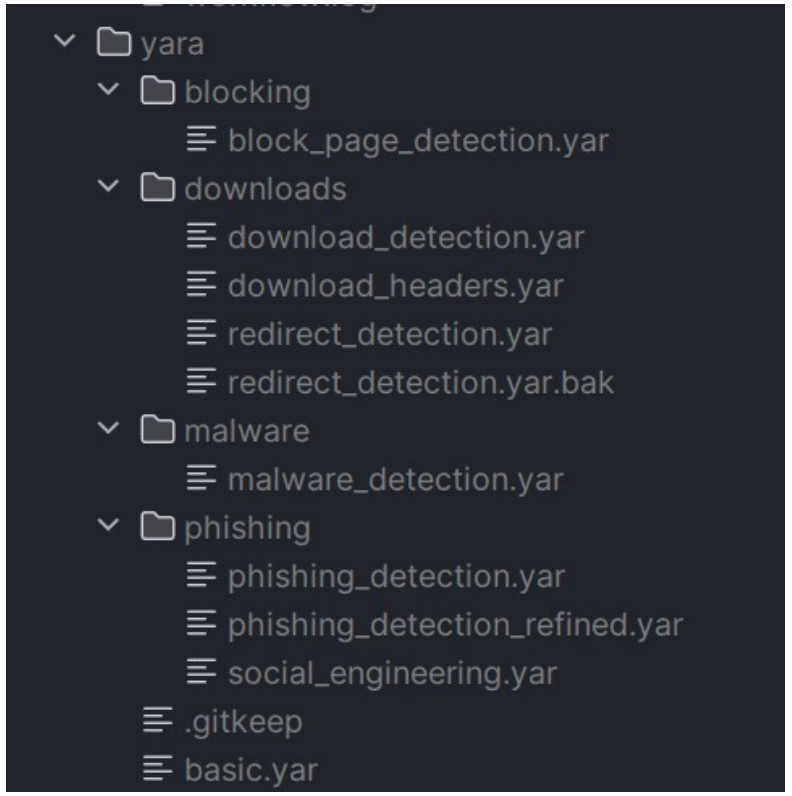
.dlog


```
{
  "session_metadata": {
    "session_id": "OSC_Demo",
    "timestamp": "2025-09-18T09:33:29.409169+00:00",
    "target_info": {
      "original": "http://malware.wicar.org/data/eicar.com",
      "normalized": "http://malware.wicar.org/data/eicar.com",
      "type": "url",
      "hash": "ab302ebc1e243ee089bee075e603f5146d08462e4a59a7b27270611c9f1ad82d",
      "scheme": "http",
      "domain": "malware.wicar.org",
      "path": "/data/eicar.com",
      "query": ""
    },
    "misp_event": {
      "event_id": 196,
      "status": "success",
      "error": null,
      "uuid": "075b4ef2-da6d-4b1b-aae7-ae0cca759828"
    }
  },
  "results": {
    "target": "http://malware.wicar.org/data/eicar.com",
    "scan_timestamp": "2025-09-18T09:33:29.409175+00:00",
    "provider_results": [
      {
        "provider": "whois",
        "status": "clean",
        "threat_detected": false,
        "threat_type": "safe",
        "confidence": "0.80",
        "tags": [],
        "error_message": null,
        "raw_response": {
          "status": "domain_found",
          "creation_date": "1985-01-01",
          "registrar": null,
          "domain_age_days": 14870,
          "raw_whois": "% IANA WHOIS server\n% for more information on IANA, visi
        }
      {
        "provider": "misp",
        "status": "threat",
        "threat_detected": true,
        "threat_type": "malicious",
        "confidence": "0.90",
        "tags": [],
        "error_message": null,
        "raw_response": {
          "events_found": 45,
          "recent_events": 45,
          "search_terms": [
            "malware.wicar.org",
            "http://malware.wicar.org/data/eicar.com",
            "/data/eicar.com"
          ],
          "categories": [
            "Threat Analysis: http://malware.wicar.org/data/js_crypto_miner.html [CRITICAL] - SID: Final_test_both_fixed",
            "Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID: Test_eicar",
            "Threat Analysis: http://malware.wicar.org/data/js_crypto_miner.html [CRITICAL] - SID: Test_crypto_miner",
            "Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID: OSC_Demo"
          ],
          "threat_levels": [
            "high"
          ],
          "misp_instance": "https://localhost",
          "sample_events": [
            {
              "info": "Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID: Test_eicar",
              "date": "2025-09-16",
              "threat_level": "1"
            }
```

```
    {
      "provider": "whalebone",
      "status": "threat",
      "threat_detected": true,
      "threat_type": "malicious",
      "confidence": "0.95",
      "tags": [],
      "error_message": null,
      "raw_response": {
        "threats": [
          {
            "threat_type": "c&c",
            "accuracy": 79,
            "first_detection": "2025-02-13T17:16:23+0000"
          },
          {
            "threat_type": "malware",
            "accuracy": 78,
            "first_detection": "2024-11-30T01:58:20+0000"
          },
          {
            "threat_type": "malware",
            "accuracy": 73,
            "first_detection": "2025-04-28T09:22:15+0000"
          },
          {
            "threat_type": "malware",
            "accuracy": 100,
            "first_detection": "2024-11-30T12:13:23+0000"
```

```
      "provider": "virustotal",
      "status": "threat",
      "threat_detected": true,
      "threat_type": "malicious",
      "confidence": "0.14",
      "tags": [],
      "error_message": null,
      "raw_response": {
        "stats": {
          "malicious": 14,
          "suspicious": 1,
          "undetected": 29,
          "harmless": 54,
          "timeout": 0
        },
        "malicious_count": 14,
        "suspicious_count": 1,
        "total_engines": 98,
        "scan_date": 1758117674,
        "categories_vt": [
          "misc",
          "known infection source",
          "command and control",
          "malware sites",
          "computer security",
          "malware",
          "phishing"
        ],
        "categories": [
          "misc",
          "known infection source",
          "command and control",
          "malware sites",
          "computer security",
          "malware",
          "phishing"
        ],
        "raw_response": {
```

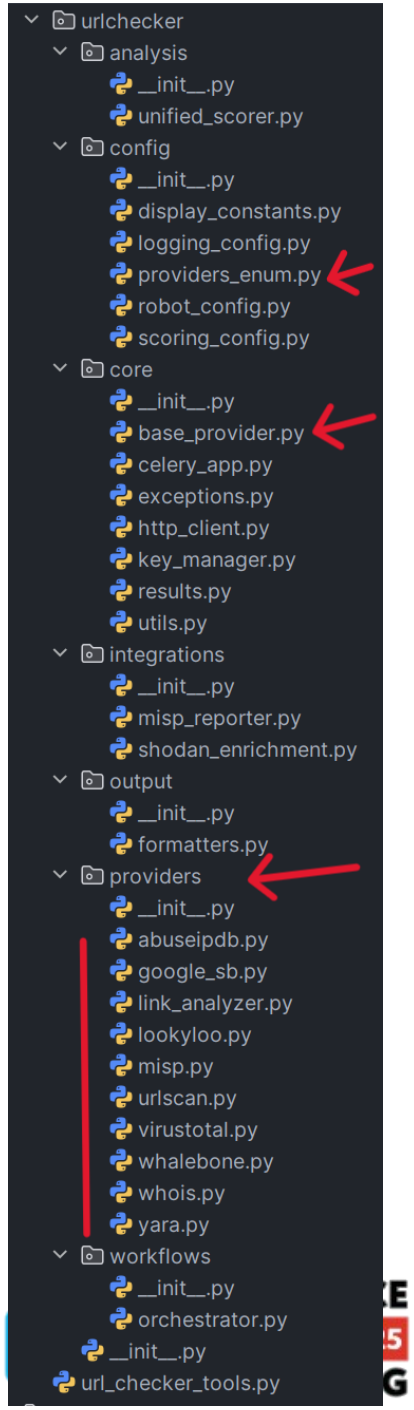
Modularity



- Supports custom YARA rules
- Integration-ready for SOC workflows or webservices

- New providers
  → add class inheriting BaseProvider
- Register in config enum
  → auto-loaded by CLI

```python
class BaseProvider(ABC):  Sam Kafai
    """Base class that all providers inherit from. Contains everything a provider needs."""

    def __init__(self, provider_name: str, config: Optional[Dict] = None):  Sam Kafai
        self.provider_name = provider_name

        # Auto-load configuration if not provided
        raw_config = config or ProviderConfigTemplate.get_all_provider_configs().get(
            provider_name, {}
        )

        # Wrap dictionary config for attribute access
        if isinstance(raw_config, dict):
            self.config = ConfigDict(raw_config)
        else:
            self.config = raw_config

        # Set up logger
        self.logger = logging_config.get_logger()

        # Set up HTTP client
        self.http = HTTPClient(provider_name, self.config, self.logger)

        # Log configuration (sanitized)
        self._log_configuration()
```



restena
réseau · sécurité · .lu

Supported Threat Intelligence Providers

- **VirusTotal** – multi-engine analysis
- **Google Safe Browsing** – phishing/malware lists
- **URLScan.io** – web snapshot & reputation analysis
- **Lookyloo** – forensic crawling
- **WHOIS** – domain age & reputation
- **Pandora** – file/malware detection
- **Whalebone** – DNS security and threat categories
- **MISP** – threat intelligence integration
- **(Local) YARA** – pattern-based content scanning

Why Those Providers ?

- **VirusTotal** – 1st choice, aggregation of sources, casts a “wide net”
- **Google Safe Browsing** – more limited in scope but reliable
- **URLScan.io** – Initially part of the tool, but slow and API gives confusing results
- **Lookyloo** – good for in depth analysis, but takes long. Huge info dump.
- **WHOIS** – perfect as a “basic check”
- **Pandora** – Good addition, but files scanning is not the focus of this tool.
- **Whalebone** – “Premium” provider, helps with categorisation and blacklisting
- **MISP** – Cross-referencing and sharing of reports
- **(Local) YARA** – Provides fallback in case providers are unreachable. Can help catch yet unreported malicious websites/domains

Challenges and other results

- Finding a balanced range of relevant (free) providers
- Handling contradictory provider reports
- Unreported malicious websited/domains
- Finding/Creating precise YARA rules
- Minimizing false-positives
- Calibrating the synthesised scoring
- Download links & encrypted files
- Ambiguity of some potential Use Cases wrt to commercial/cloud providers

...

Side notes – POC Misp Report (Event)

Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID: OSC_Demo

Event ID	193
UUID	20297480-7e47-4f4c-aa92-d98a950e641a
Creator org	ADMIN
Owner org	ADMIN
Creator user	admin@admin.test
Protected Event (experimental)	Event is in unprotected mode. Switch to protected mode
Tags	urlchecker:verdict=critical x, urlchecker:score=79 x, urlchecker:providers=8 x, urlchecker:threats=4 x, tlp:white x, verdict:clean:whois x, verdict:malicious:misp x, verdict:clean:link_analyzer x, verdict:malicious:whalebone x, verdict:malicious:virustotal x, verdict:malicious:google_sb x, verdict:clean:yara x, verdict:clean:abuseipdb x, confidence:multi-source x, coverage:comprehensive x
Date	2025-09-18
Threat Level	High
Analysis	Ongoing
Distribution	Your organisation only
Warnings	**Distribution**: The event is tagged as tlp:white, yet the distribution is not set to all. Change the distribution setting to something more lax if you wish for the event to propagate further.
Published	No
#Attributes	8 (0 Objects)
First recorded change	2025-09-18 09:24:55
Last change	2025-09-18 09:24:56
Modification map	
Sightings	0 (0) - restricted to own organisation only.

Related Events

- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-18 4
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-18 4
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-18 4
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-18 4
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-18 4
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-17 2
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-17 2
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-17 2
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-17 2
- Threat Analysis: http://malware.wicar.org/data/eicar.com [CRITICAL] - SID:... 2025-09-16 2

Show (1 more)

restena
réseau · sécurité · .lu

OPEN SOURCE CONFERENCE 2025 LUXEMBOURG

Side notes – POC Misp Report (Attributes) (1)

y	Type	Value	Tags	Galaxies	Comment	Correlate
activity	url	http://malware.wicar.org/data/eicar.com			Primary target URL (Score: 79/100)	☑
activity	domain	malware.wicar.org			Domain extracted from target URL	☑
	text	WHOIS: - Domain age: 14870 days - Registrar: None - Created: 1985-01-01			WHOIS registration details	☑
activity	text	Network Analysis: - Summary: DNS resolved 2 IPs - Resolved IPs (2): - 208.94.116.246 - 2607:ff18:80:6::6a08			Comprehensive network analysis and DNS/IP resolution	☑
activity	text	Link Analysis: - Verdict: Safe - Original URL: http://malware.wicar.org/data/eicar.com - Final URL: http://malware.wicar.org/data/eicar.com - Redirects: 0 - Domain Change: no - Status Code: 200 - Resolved IPs: 208.94.116.246, 2607:ff18:80:6::6a08 - Shortener: no - Blocked Page: no			HTTP redirect/destination and DNS resolution analysis	☑
	text	Threat Intel: - WHOIS: Safe - MISP: Malicious - WHALEBONE: Malicious (c&c, malware; Max accuracy: 100%) [Categories: porn]			Consolidated threat intelligence vendor verdicts	☑

restena
réseau · sécurité · .lu

OPEN SOURCE CONFERENCE 2025 LUXEMBOURG

Side notes – POC Misp Report (Attributes) (2)

		- Verdict: Safe - Original URL: http://malware.wicar.org/data/eicar.com - Final URL: http://malware.wicar.org/data/eicar.com - Redirects: 0 - Domain Change: no - Status Code: 200 - Resolved IPs: 208.94.116.246, 2607:ff18:80:6::6a08 - Shortener: no - Blocked Page: no				
	text	Threat Intel: - WHOIS: Safe - MISP: Malicious - WHALEBONE: Malicious (c&c, malware; Max accuracy: 100%) [Categories: porn] - VT: Malicious (14/98) [Categories: misc, known infection source, command and control, malware sites, computer security, malware, phishing] - GOOGLE_SB: Malicious (MALWARE, MALWARE) - YARA: Safe - ABUSEIPDB: Safe			Consolidated threat intelligence vendor verdicts	☑
delivery	text	YARA Analysis: - Summary: No pattern matches			YARA behavioral analysis and pattern detection	☑
	comment	Threat Assessment: 4/8 providers flagged as malicious | Clean: 4/8 | Threat Score: 79/100			Comprehensive threat assessment summary across all providers	☐

restena
réseau · sécurité · .lu

OPEN SOURCE CONFERENCE 2025 LUXEMBOURG

3.

A Conclusion ?

- Having an OpenSource basic brick for the URL Shortener was essential to us
- This is why we aimed at complementing it with the URL Checker Tools that could be useful to others
- We also provided our own analyses results in this talk in the hope to share openly our results to the community
- Happy to get your feedback on strategies on how to release such sourcecode (Licensing)

Merci.
Thank you !
Questions ???

This is Hex :)

restena
réseau · sécurité · .lu

www.restena.lu

OPEN SOURCE CONFERENCE 2025 LUXEMBOURG