The SEPSES knowledge graph: An integrated resource for cybersecurity

Elmar Kiesling\textsuperscript{1,2}, Andreas Ekelhart\textsuperscript{1,3}, Kabul Kurniawan\textsuperscript{1}, Fajar Ekaputra\textsuperscript{1}

\textsuperscript{1} TU Wien, \textsuperscript{2} Vienna University of Economics and Business, \textsuperscript{3} SBA Research
Security in the news..

https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/
Security Data Sharing and Analytics: the bad..

A lot of security information is text..

Log data volume and heterogeneity

Integration? Interpretation? Contextualization?

Limitations of “analytic” tools

Image credits: https://twiki.org/cgi-bin/view/Codev/TWikiPresentation2013x03x07; https://www.ptonline.com/articles/finding-root-causes-of-a-material-problem; https://mti.com/blog/2019/01/24/mti-security-alert-subscription-770-million-passwords-made-available/alert.png.1/
Security Data Sharing and Analytics: ..and the ugly
Enter: The SEPSES Cybersecurity Knowledge Graph

Goals:
• Make cybersecurity knowledge more actionable*
• Establish a semantic foundation for security analytics and reasoning
• Improve situational awareness and enhance defensive capabilities
• Facilitate integration of global and local cybersecurity knowledge

How?
• Integrate dispersed cybersecurity information into a regularly updated public knowledge graph
• Make the available cybersecurity information queryable
• Facilitate contextualization with local information

Use cases:
• Security monitoring
• Risk assessment and decision support
• Incident detection and response
• Threat intelligence exchange
• Forensics
• etc.

Criteria (ENISA)[1]
1. Relevance
2. Timeliness
3. Accuracy
4. Completeness
5. Ingestibility

[1] Actionable information for security incident response — ENISA
Sharing Standards for Cybersecurity Information

Dandurand, L., Kaplan, A., Kácha, P., Kadobayashi, Y., Kompanek, A. and Lima, T., 2015. Standards and tools for exchange and processing of actionable information. European Union Agency for Cybersecurity (ENISA)
Vocabularies: High-level Overview
Vocabularies: CAPEC

- “Common **Attack Pattern** Enumeration and Classification”[1]
- Governing body: MITRE
- Publicly available catalog of attack patterns
- CAPEC entries are descriptions of particular attack patterns, i.e., the techniques and procedures used to carry out the sequence of steps that makes up the pattern.

[1] https://capec.mitre.org

https://sepses ifs.tuwien.ac.at/vocab/ref/capec/index-en.html#
Vocabularies: CVE

- “Common Vulnerabilities and Exposure”[1]
- Governing body: MITRE, published through National Vulnerability Database (NVD)[2]
- Various feeds available
- List of known security vulnerabilities
- Defines identifiers (CVE-IDs) that can be used to reference publicly-known vulnerabilities.
- Assigned by CVE Numbering Authorities (CNAs), i.e., software vendors and other organizations that have met requirements specified by MITRE.

[1] https://cve.mitre.org  [2] https://nvd.nist.gov
Vocabularies: CVSS

- “Common Vulnerability Scoring System”[1]
- Governing body: FIRST
- Scoring system for describing and rating IT vulnerabilities from 0 (least) to 10 (most critical)

Example: MySQL Stored SQL Injection (CVE-2013-0375)

| Metric       | Value | Comments |
|--------------|-------|----------|
| Attack Vector| Network The attacker connects to the exploitable MySQL database over a network. |
| Attack Complexity | Low Replication must be enabled on the target database. Following the guidance in Section 3.1.2 of the Specification Document that was added in CVSS v3.1, we assume the system is configured in this way. |
| Privileges Required | Low The attacker requires an account with the ability to change user-supplied identifiers, such as table names. Basicaccès to the database can get this privilege by default, but it is not considered a sufficiently trusted privilege to warrant this metric being high. |
| User Interaction | None No user interaction is required as replication happens automatically. |
| Scope | Changed The vulnerable component is the MySQL server database that the attacker logs into to perform the attack. The impacted components are a remote MySQL server database (or databases) that the database replicates to. |
| Confidentiality | Low The injected SQL runs with high privilege and can access information the attacker should not have access to. Although this runs on a remote database (or databases), it may be possible to utilize the information as part of the SQL statement. The malicious SQL is injected into SQL statements that are part of the replication functionality, preventing the attacker from executing arbitrary SQL statements. |
| Integrity | Low The injected SQL runs with high privilege and can modify information the attacker should not have access to. The malicious SQL is injected into SQL statements that are part of the replication functionality, preventing the attacker from executing arbitrary SQL statements. |
| Availability | None Although injected code runs with high privilege, the nature of this attack prevents arbitrary SQL statements being run that could affect the availability of MySQL databases. |

[1] https://www.first.org/cvss/

https://sepses ifs.tuwien.ac.at/vocab/ref/cvss/index-en.html#
Vocabularies: CPE

• “Common Platform Enumeration”\(^1\)
• Governing body: NIST\(^2\)

• Consistent and structured naming scheme for operating systems, software packages and classes of hardware devices.
• Includes a method by which vendors can validate that their product names are accurately represented in the CPE system.

Example: Entry for Protégé 4.0.2:

cpe:2.3:a:stanford:protege:4.0.2:*:*:*:*:*:*:*:*
Vocabularies: CWE

• “Common Weakness Enumeration”[1]
• Governing body: MITRE

• List of commonly occurring software weaknesses and vulnerabilities.
• Primary goal: avoid introducing vulnerabilities in the first place by educating software developers.

[1] https://cwe.mitre.org
ETL Process

Online Resources
(CVE, CVSS, CPE, CWE, CAPEC)

Continuous Update Engine

Data Acquisition, Steps:
- Downloading the resource (.ZIP File)
- Unzip downloaded zip file

Resource Extraction:
Straightforward extraction using RML Rules

Entity Linking & Validation:
Checking for term linking (e.g., CVE to CWE, CVE to CPE, CWE to CAPEC etc.)

Data storage:
Storing extracted data to the triplestore

SPARQL Endpoint
TPF Interface
Linked Data Interface
DataDump

Triple Store
## Breakdown of resulting CPSS Knowledge Graph

|                           | CVE | CVSS | CPE | CWE | CAPEC | SnortRules |
|---------------------------|-----|------|-----|-----|-------|------------|
| Axioms                    | 68  | 248  | 111 | 256 | 149   | 486        |
| Class count               | 7   | 9    | 5   | 10  | 8     | 10         |
| Object property Count     | 6   | 8    | 4   | 9   | 6     | 10         |
| Data property count       | 8   | 37   | 18  | 40  | 22    | 103        |
| Individual count          | 123,005 | 123,220 | 393,695 | 808  | 516   | 3,488      |

SEPSES knowledge graph statistics (As per July 2, 2019.)
Individuals over time (from June 2019)

- **CVE**
  - Graph showing a steady increase over time from June to October.
  - Y-axis: $x \times 10000$

- **CVSS**
  - Graph showing a steady increase over time from June to October.
  - Y-axis: $x \times 10000$

- **CPE**
  - Graph showing a steady increase over time from June to October.
  - Y-axis: $x \times 10000$

- **CWE and CAPEC**
  - Graph showing a steady increase in CWE and a flat line for CAPEC.
  - Y-axis: $x \times 10000$

SEPPES Knowledge Graph Statistics (as per Oct 10, 2019)
[UC 1] Vulnerability Assessment

Polystructured security information

Cybersecurity knowledge graph

Local asset inventory

Vulnerable Assets

Informatics, SBA Research
[UC 1] Vulnerability Assessment: Example Setting

- **Host name:** PC-1
  - **IP Address:** 192.168.1.1
  - **Host Type:** Client
  - **Has Data:** Financial Data
  - **Has CPE:**
    - cpeamicrosoftinternetexplorer11
    - cpeamicrosoftoffice2019macos
    - cpeomicrosoftwindows10x86

- **Host Name:** PC-2
  - **Host Type:** Client
  - **IP Address:** 192.168.1.2
  - **Has Data:** Employee Data
  - **Has CPE:**
    - cpecanonicalubuntulinux1604lts

- **Database Server**
  - **Host Name:** DB-1
  - **Host Type:** Database Server
  - **IP Address:** 192.168.1.3
  - **Has CPE:**
    - cpecanonicalubuntulinux1604lts

**Local asset inventory**
[UC 1] Vulnerability Assessment: Vulnerable Assets

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX asset: <http://w3id.org/sepses/vocab/bgk/assetKnowledge#>
PREFIX cve: <http://w3id.org/sepses/vocab/ref/cve#>
PREFIX cpe: <http://w3id.org/sepses/vocab/ref/cpe#>
PREFIX cvss: <http://w3id.org/sepses/vocab/ref/cvss#>
PREFIX cwe: <http://w3id.org/sepses/vocab/ref/cwe#>

SELECT  distinct ?hostName str(?ip) as ?IP ?product
        (group_concat(?cveId) as ?cveIds) from
        <http://localhost:8890/localdata2>
WHERE {
    ?s a asset:Host.
    ?s rdfs:label ?hostName.
    ?s asset:ipAddress ?ip.
    ?s asset:hasProduct ?p.
    SERVICE <http://sepses.ifis.tuwien.ac.at/sparql> {
        ?cve cve:hasCPE ?p .
        ?cve cve:hasCVSS ?s .
        ?p cpe:cpe:2.3 $cpeNumber
    }
} group by ?hostName

| hostName    | IP       | product               | cveIds                           |
|-------------|----------|-----------------------|----------------------------------|
| DBServer1   | 192.168.1.3 | Windows Server 2016   | CVE-2016-3332, ..., CVE-2017-8746 |
| Workstation1| 192.168.1.1 | Windows 10            | CVE-2016-3302, ..., CVE-2015-2554 |
[UC 1] Vulnerability Assessment: Critical Vulnerabilities

```
SELECT DISTINCT ?hostName ?cveId ?conf ?cvssScore AS ?score ?dataAsset ?classification AS ?class
FROM <http://localhost:8890/localdata>
WHERE {
  ?s a asset:Host.
  ?s rdfs:label ?hostName.
  ?s asset:hasProduct ?product.
  ?s asset:hasDataAsset ?dt.
  ?dt rdfs:label ?dataAsset.
  ?dt asset:hasClassification ?c.
  ?c rdfs:label ?classification.
  ?c asset:dataClassificationValue ?cv
FILTER (?confidentiality = "COMPLETE")
FILTER (?cv = 1)
```

| hostName       | cveId       | conf        | score | dataAsset | class     | consequence                         |
|----------------|-------------|-------------|-------|-----------|----------|-------------------------------------|
| Workstation2   | 2016-1646   | COMPLETE    | 9.3   | EmpData   | Private  | Read Memory                         |
| Workstation2   | 2016-1653   | COMPLETE    | 9.3   | EmpData   | Private  | DoS: Crash, Exit...                  |
| Workstation2   | 2016-1583   | COMPLETE    | 7.2   | EmpData   | Private  | DoS: Resource Cons...                |
| Workstation2   | 2016-1583   | COMPLETE    | 9.3   | EmpData   | Private  | Execute Unauthorized...              |
[UC 2] Intrusion Detection (NIDS)

Goals:
- Improve situational awareness
- Contextualize alerts
- Reduce false positives

Polystructured security information

Cybersecurity knowledge graph

Log graph

Local asset inventory

Enriched Alerts

Log graph

Polystructured Log Data
[UC 2] Intrusion Detection: Example references in log stream

IDS Snort Log

[**] [1:2129:2] WEB-IIS nsislog.dll access [**]
[Classification: access to a potentially vulnerable web application] [Priority: 2]
11/08-13:01:58.880651 10.2.190.254:50559 -> 154.241.88.201:80

IDS Snort Rule

alert tcp $EXTERNAL_NET any -> $HTTP_SERVERS $HTTP_PORTS ( msg:"SERVER-IIS nsislog.dll access"; flow:to_server,established; http_uri: content:"/nsislog.dll", fast_pattern,nocase; metadata:ruleset community; service:http; sid:2129; rev:25; reference:cve,2003-0227; reference:bugtraq,8035; classtype:web-application-activity)

CVE resource

CVE-2003-0227

Description

The logging capability for unicast and multicast transmissions in the ISAPI extension for Microsoft Windows Media Services in Microsoft Windows NT 4.0 and 2000, nsislog.dll, allows remote attackers to cause a denial of service in Internet Information Server (IIS) and execute arbitrary code via a certain network request.

Source: MITRE
Description Last Modified: 06/09/2003

Evaluation Ruleset

Snort Community rules V.3
(2,100 rules, custom format) [1]

[1] https://www.snort.org/
[2] https://w3id.org/seses/vocab/rule/snort
[UC 2] Intrusion Detection: Evaluation Setting

Evaluation Dataset
MACCDC 2012 cybersecurity competition[1,2] (11,000 alerts)

1. https://maccdc.org/2012-agenda/
2. https://www.secrepo.com/maccdc2012/maccdc2012_full_alert.7z (24MB)
3. https://w3id.org/sepses/vocab/log/snort-alert
[UC 2] Intrusion Detection: IDS Alert query

```sparql
PREFIX cve: <http://w3id.org/sepses/vocab/ref/cve#>
PREFIX cpe: <http://w3id.org/sepses/vocab/ref/cpe#>
PREFIX snort: <http://w3id.org/sepses/vocab/ref/snort#>
PREFIX snort-rule: <http://w3id.org/sepses/vocab/rule/snort#>
PREFIX snort-alert: <http://w3id.org/sepses/vocab/log/snort-alert#>

SELECT DISTINCT ?alert ?message ?sid ?sourceIp ?destinationIp ?cveId ?cpeId
FROM <http://localhost:8890/snortalert>
WHERE {
  ?alert a snort-alert:IDSSnortAlertLogEntry ;
  snort:signatureId ?sid ;
  snort:message ?message ;
  snort:sourceIp ?sourceIp ;
  snort:destinationIp ?destinationIp .
}
```

| alert       | message          | sid          | sourceIP          | targetIP          | cveId      | cpeId          |
|-------------|------------------|--------------|-------------------|-------------------|------------|----------------|
| Alert001    | WEB-MISC Chunked | 1807         | 10.2.190.254      | 154.241.88.201    | 2002-0392  | cpe:/a:apa:... |
| Alert002    | WEB-MISC WebDAV  | 1070         | 10.2.190.254      | 154.241.88.201    | 2000-0951  | cpe:/a:micr:... |
| Alert003    | WEB-MISC TRACE   | 2056         | 10.2.197.241      | 154.241.88.201    | 2004-2320  | cpe:/a:bea:w:... |
| Alert004    | WEB-FRONTPAGE    | 1248         | 10.2.190.254      | 154.241.88.201    | 2001-0341  | cpe:/a:micr:... |
| Alert005    | WEB-MISC Netscape| 1048         | 10.2.197.241      | 154.241.88.201    | 2001-0250  | cpe:/a:netsc:... |
```
Published Resources

Website

- **Documentation** and status information
- Vocabularies in various serializations: CVE, CPE, CVSS, CWE, CAPEC, Snort rules
- RDF dumps of the knowledge graph

Software

- **ETL Workflow**: implemented in Java

Services

- SPARQL endpoint
- TPF interface
- LD interface
Where do we go from here?

- Disseminate the resource in the security domain
- Grow the knowledge graph
- Link with higher-level conceptualizations (upper security ontology)
- Extend the scope – e.g., indicators of compromise, reporting formats (STIX) etc.
Questions?

Elmar Kiesling, PhD.
Department of Information Systems and Operations
Vienna University of Economics and Business

+43-1-313 36-6366
elmar.kiesling@wu.ac.at