



InTIME: Cyber-Threat Intelligence Extraction & Sharing

- An entirely Open Source framework for Cyber-Threat Intelligence (CTI) Extraction & Sharing
 - manages the **complete CTI lifecycle**
 - enhances security **preparedness & awareness**
- Machine learning-based web data gathering
 - efficient data discovery (e.g., 0-day vulnerabilities, exploits) from the **clear**, **social**, and **dark** web (e.g., forums, marketplaces, security-related websites, pastebins)
- Language model-based **content ranking** to assess its usefulness
 - statistical language modelling to represent content in low-dimensional space
 - ranks crawled content according to its potential to be useful



InTIME: Cyber-Threat Intelligence Extraction & Sharing

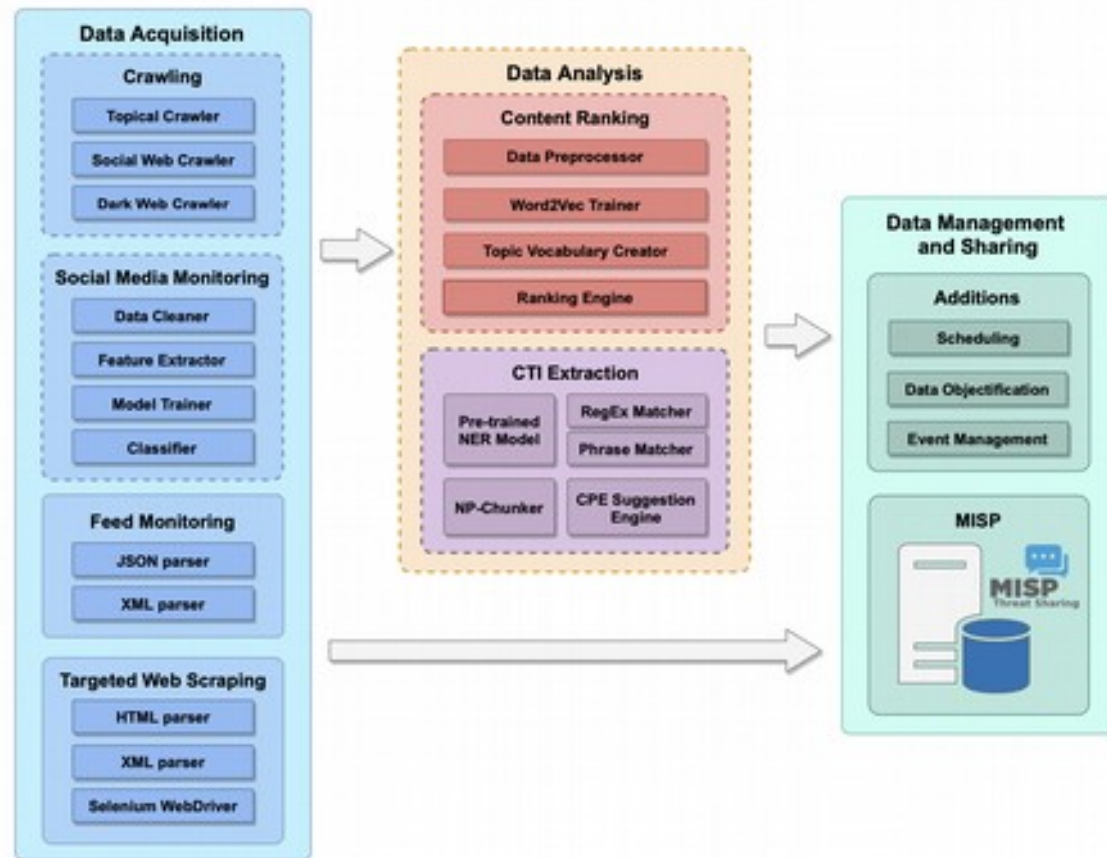
- **Natural language understanding and named entity recognition/disambiguation**
 - extracts CTI from crawled free text
 - domain specific entities
 - dependency parsing for more accurate extraction
 - specialised tools for semi-automated linking to known platform/vulnerability naming schemes (e.g., CPEs, CVEs)
- **Scalable CTI management and sharing**
 - extends MISP platform
 - supports CTI storage, consolidation, exploration, inspection and visualisation
 - supports CTI sharing in both human and machine-based formats



InTIME architecture

Three main modules:

- data acquisition – content harvesting (crawlers/scrapers)
- data analysis
 - content ranking
 - CTI extraction
- data management & sharing





Content harvesting

- A family of crawlers/scrapers
 - tunable & automated **clear/deep/social/dark** web traversal
 - thematic & in-depth crawling
 - machine-learning based (SVMs/random forests direct the crawls)
 - regex-support** (structured domains)
 - Integration with **public VDBs** (JVN, NVD, kbcert, vulldb, exploit db)

The screenshot displays the InTIME web interface, divided into several sections:

- Monitoring:**
 - Crawler ID:** default
 - General:**

Uncrawled Links in Frontier	1,451,815
Successful Requests	526,873
Failed Requests	154,866
Aborted Requests	20,132
 - Page Relevance:**

Total Pages	392,157
Relevant Pages	37,157
Irrelevant Pages	355,000
Harvest Rate	9.475%
 - HTTP Response:**

HTTP 200: Success	434,740
Error HTTP 401: Unauthorized	
Error HTTP 403: Forbidden	
Error HTTP 404: Not Found	
Error HTTP 500: Server Errors	
 - Page Fetcher Performance:**

Fetch Time (Mean)	27360.09 ms
-------------------	-------------
- CRAWLER:**
 - Data crawler:** Data harvested from AcheCrawler
 - Data crawler:** Includes a pie chart showing 'Success' (green) and 'Failed' (red) requests, and a line graph showing 'Success' and 'Failed' requests over time.
 - Number of requests:** A bar chart showing the number of requests over time, with red bars for failed requests and green bars for successful requests.
- Crawler settings:**
 - Crawler list:** A table with a 'Start crawler' button.
 - Crawler status:** Shows 'The crawler logical_crawler_example is Offline' with a 'Start crawler' button.
 - Seed Finder query:** A text input field for a query.
 - Seeds:** A table with a search bar and a 'No data available in table' message.
 - Training URLs:** A table with a search bar and a 'No data available in table' message.



Content ranking

A novel ranking/classification module:

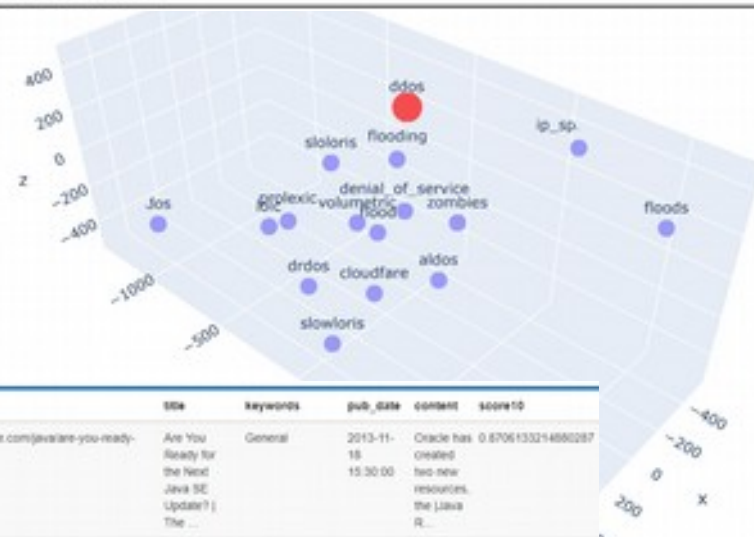
- represent harvested content in **latent low-dimensional space** (adaptive to vocabulary changes)
- statistical language modeling** to assess usefulness harvested content (captures salient concepts)
- machine-learning based (**word embeddings/2-layer NNs & CNNs**)
- noSQL** data store

Excerpt from: www.iotforall.com/5-worst-iot-hacking-vulnerabilities

The **Mirai Botnet** (aka **Dyn Attack**) Back in October of 2016, the largest **DDoS attack** ever was launched on **service** provider Dyn using an **IoT botnet**. This led to huge portions of the **internet** going down, including **Twitter**, the Guardian, **Netflix**, Reddit, and CNN.

This **IoT botnet** was made possible by **malware** called **Mirai**. Once **infected** with **Mirai**, computers continually search the **internet** for vulnerable **IoT devices** and then use known default **usernames** and **passwords** to **log** in, infecting them with **malware**. These **devices** were things like digital cameras and **DVR players**.

Relevance Score: 0.8563855440900794



_id	source_url	site	keywords	pub_date	content	score10
500518f19ad3e0d749e4977	https://blogs.oracle.com/java/are-you-ready-for-l...	Are You Ready for the Next Java SE Update? The ...	General	2013-11-15 15:30:00	Oracle has created two new resources, the Java R...	0.8706133214880287
500518f19ad3e0d749e4978	https://blogs.oracle.com/security/security-alert-...	Security Alert CVE-2017-9605 Released Oracle Se...	Critical Patch Updates, Security Updates	2017-09-22 20:10:00	Last week, Equifax identified an Apache Struts 2 ...	0.8216238044289645
500518f19ad3e0d749e4979	https://blogs.oracle.com/security/security-alert-...	Security Alert CVE-2019-2723 Released Oracle Se...	Critical Patch Updates	2019-04-26 17:43:00	Oracle has just released Security Alert CVE-2019-...	0.7925260253604896



CTI extraction

Information extraction module:

- natural language understanding (machine-learning & rule based)
- named entity recognition & disambiguation (e.g., CPE suggestion)
- introduced CTI-specific entities (e.g., shell_cmd, version, file, cpe/cve/cwe) alongside standard ones (e.g., org, date, ordinal)
- REST API w/ extracted CTI (JSON objects w/ entities & metadata)

```
{
  "docId": 915240,
  "rawText": "Symantec Web Security Group (WSG) products",
  "matches": [
    {
      "text": "Symantec Web Security",
      "entity": "PRODUCT",
      "start": 0,
      "end": 3
    },
    {
      "text": "WSG",
      "entity": "ORG",
      "start": 5,
      "end": 6
    },
    {
      "text": "OpenSSL",
      "entity": "PRODUCT",
      "start": 12,
      "end": 13
    }
  ]
}
```

Symantec Web Security Group (WSG org) products using affected versions of `openssl` may be susceptible to multiple vulnerabilities. A local or remote attacker can obtain private key or other secret key information. A remote attacker can also cause denial of service.

CVE-2017-15511 is exploitable in ASG (versions 6.7 VERSION , 7.1 VERSION , 7.2 VERSION), CA (versions 2.3 VERSION , 2.4 VERSION , 3.0 VERSION), `bluecoat:proxysg` (cpe:/a:bluecoat:proxysg cpe : 6.5.8.7 VERSION), `bluecoat:ssl` (versions 10.4 VERSION , 10.5 VERSION), and SSLV org (versions 4.5 VERSION , 5.0 VERSION) only when customers configure the products' SSL/TLS org interfaces with 1024-bit RSA org keys.

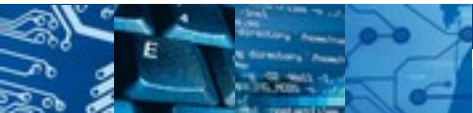


CTI management & sharing

CTI management & sharing module:

- efficient CTI storage & indexing
- easy-to-use CTI consolidation & correlation
- intuitive CTI inspection & visualisation
- supports CTI sharing in both human and machine-based formats
- extends MISP platform

The screenshot displays the InTIME interface. At the top, there is a network graph showing relationships between various events and attributes. Below the graph is a table of events with columns for 'Published', 'Org', 'Owner org', 'Id', 'Clusters', 'Tags', 'Addr', 'Email', 'Date', 'Info', 'Distribution', and 'Actions'. The table lists several events from 'ORGNAME' with IDs ranging from 152629 to 152648. Below the table, a detailed view of an event is shown, including its 'id', 'url', 'description', and 'references'.





InTime & Open Source

- InTime has been built using solely open source tools
- Very positive experience from the Open Source Ecosystem
 - Good documentation
 - Clear installation instructions, descriptive APIs
 - More details could be provided for DB design
 - Good scalability
 - The MISP system scaled up well to hundreds of thousands of entries
 - The storage and management of correlations was suboptimal, leading to performance bottlenecks and manageability issues (e.g. transferring of DB)
 - A more efficient database design could serve well here
- InTime is in the process of being published as open source
 - Code cleanup and documentation improvements

