CTI & OSINT

Cyber Threat Intelligence & Open Source Intelligence

Gianluca Peco

Mini WS CCR sulla Sicurezza Informatica

13-15/2/2023 - Padova

Public

Introduzione CTI e OSINT

Finding malicious activity and other traces of adversaries is extremely challenging in today's complex environments, especially since it is easy for an adversary to look like a legitimate user. Cyber threat intelligence (CTI) is valuable way to augment the SOC's ability to identify adversaries and discern their movements from that of authorized users'.

It moves the SOC from a per-incident approach to an adversary-focused paradigm. In SOC environments, CTI can augment defenses by informing the following:

- Identifying unwanted actors in networks
- Tuning sensors and analytic systems/frameworks for better monitoring
- Prioritizing resources
- Providing context to incidents
- Anticipating adversary activities in more advanced SOCs
- Preventing or slowing down imminent attacks

CTI is increasing in importance due to extended boundaries and interconnectedness of modern organizations.

Definizione

- CTI comes from internal and external sources, and both are important.
  - **Internal sources** include analysts and security researchers who curate, correlate, and analyze information about adversaries, based on sources from within the constituency. This includes incident data combined with other information and data about an adversary
  - **External sources** include the dedicated commercial threat feeds and inter-SOC or constituency threat reporting
- Combining government, industry, and academia working definitions:

CTI refers to the collection, processing, organizing, and interpreting of data into actionable information or products that relate to capabilities, opportunities, actions, and intent of adversaries in the cyber domain to meet a specific requirement determined by and informing decision-makers

Indicator of Compromise

In cybersecurity defense, SOC analysts typically rely on the structured analytic technique of indicator analysis.

This involves analyzing and identifying those facts around a cyber compromise such as malware, connections to malicious websites, IP addresses, and other artifacts that could change with (not-so) sophisticated adversaries.

This is just one technique, and it has a downside when applied to asymmetric threats, such as APTs, which is that it relies on historical (past) data which usually does not predict future activities.

By definition, indicators of compromise have already occurred.

What is a CTI ?

Because CTI is both produced and consumed across various fields, there tends to be misunderstanding about what is and is not CTI. SOC data sources are often interchanged and referred to as CTI by some; level-setting if a data source is generally considered to be CTI goes a long way towards creating a shared understanding

CTI Examples	Not CTI*
Finished unstructured threat reporting **Structured threat reporting** **Open-source intelligence (OSINT)** **Curated subscriber reports and feedback**	**IP Addresses** **Domain names** **Email addresses** **Malware samples** **Virus signatures** **PCAP captures** **DNS logs** **Intrusion detection alerts** **System logs** **Social media** * These are not considered CTI unless they are associated with adversary context.

Finished unstructured threat reporting: Includes intelligence reports and analysis, sometimes lengthy, describing adversaries, observables & context analysis, e.g., multi-page CTI reports produced by many SOCs, often distributed in portable document format (PDF) and HTML format.
Structured threat reporting: Includes contextualized TTPs: collected TTPs associated with threat actors; also includes prescribed formats including who, what, when, where, and/or why. Examples: MITRE ATT&CK®, STIX-formatted CTI feeds
Curated subscriber reports and feedback: Anonymized threat information from subscribers or customers. Example: Mandiant Advantage Free

Evaluating CTI Characteristics

Having a river of CTI flowing can be a great help to the SOC, but discerning which CTI to use and prioritize is important. Continual updating of instrumentation and analytics adds to the SOC analyst workload and if not prioritized, can get out of control; this leads to a lot of effort without sufficient return on the time invested.

To decide what cyber threat intelligence to use, consider the following criteria for evaluation :

- **Actionable:** Can the SOC do something constructive with the information, such as correlate with other data, create threat hunting scenarios or actions, or enact preventative protections? Is the CTI specific enough for the SOC to operationalize? Does it come in a format that is consumable and enrich a decision, while not complicating it ?
- **Timely:** Are events recent (in days, hours, minutes for streams, or weeks for analysis)? Are there stale data?
- **Relevant:** Does it apply to the organization and reveal unknown and possible threats? Does it come from a reputable source? Is the data volume manageable? How is the CTI ingested or analyzed? Are there application program interfaces (APIs) for feeds and platforms?
- **Accurate:** Does the content correctly describe what happened? Did the CTI include spurious or wrong data about the original attack?

All these criteria together provide means of comparison among CTI subscriptions and tools and are indicators of a CTI source that can be trusted and is likely to be of value to the organization

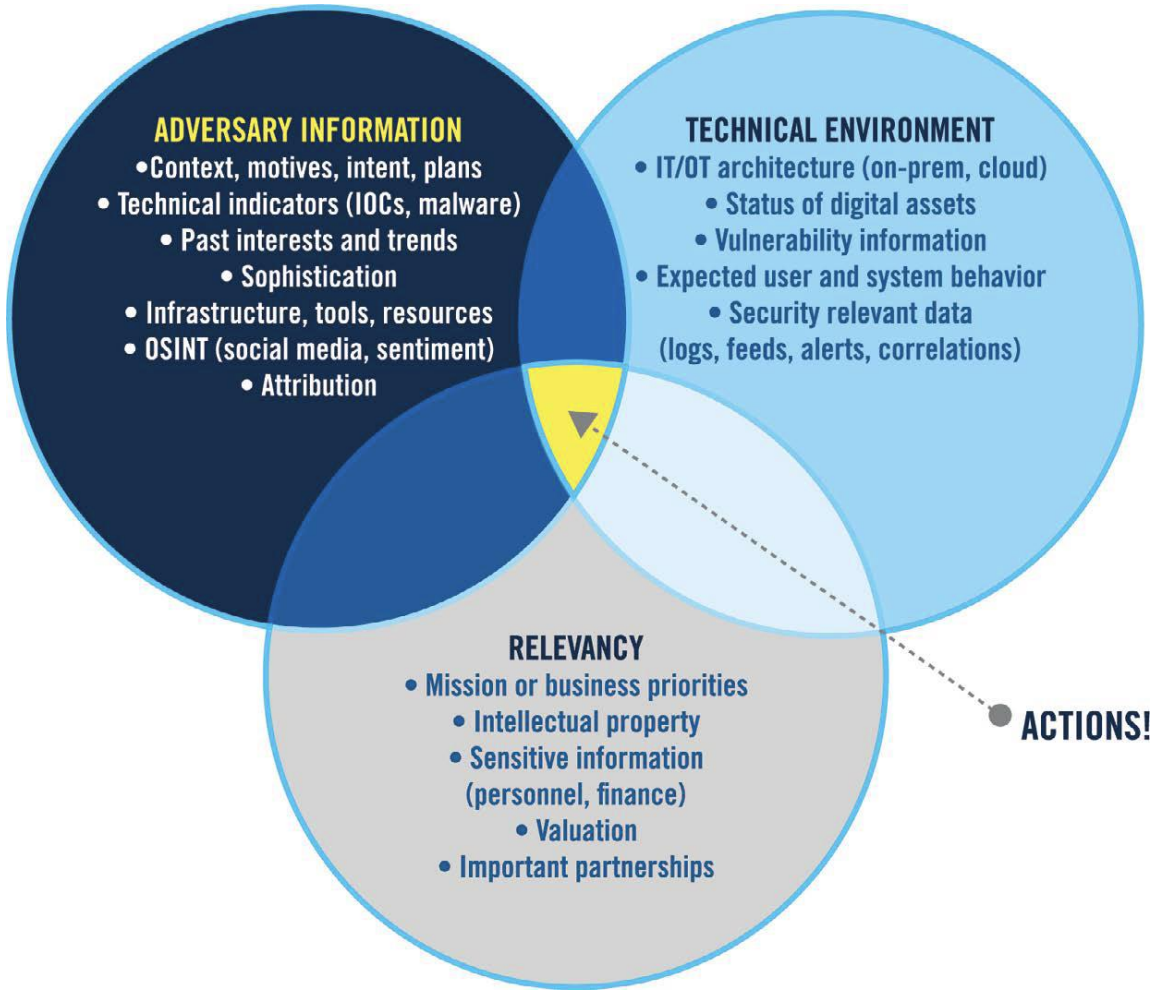
Effective CTI Correlation

Analyzing all three informative aspects of CTI in context of the constituency's business priorities, enables analysts to tailor CTI and choose CTI commercial and open-source services and platforms.

Importantly, it provides the SOC with the actions needed to take against an adversary

Effective CTI actions are the result of analyzing adversary information for relevancy as applied to the technical environment.

Without analyzing the three aspects together, the SOC will not be as effective



Adversary information

- This includes CTI as well as any other information and data about the adversary that might provide context and enable further correlation, anticipation, or other analysis.
- It includes IOCs and TTPs, intrusion detection alerts, interests, intent, resources, geopolitical context, past, present targets (intellectual property, organizations, etc.), and cultural norms to anticipate or predict adversary movements or targets and other information and analysis informing who and what the adversaries are targeting the constituency.
- This also includes impact or potential effects of the adversary on a constituency (or nation), and information on cyber actors

Adversary information

Many sources provide both technology CTI feeds as well as written, unstructured reporting.

Adversary information, with context, can assist in anticipation of adversaries.

Free Sources on adversaries include:

- Organizational incident data (IOCs, TTPS, etc.) generated by the SOC and its partners
- Various Opensource Threat Feeds identified (p.e https://abuse.ch/ https://firehol.org/ etc)
- AlienVault OTX: Adversary: Open Threat Exchange, AT&T AlienVault community
- GitHub: Awesome-threat-intelligence, A curated list of Awesome Threat Intelligence resources
- Dragos: Threat Activity Groups – Commercial
- MISP Open-Source Threat Intelligence Platform Open Standards for Threat Information Sharing (formerly known as Malware Information Sharing Platform)
- GitHub MITRE ATT&CK - Cyber Threat Intelligence Repository expressed in STIX 2.0
- Proofpoint feeds
- SANS Internet Storm Center – also other interesting tool
- IBM X-Force Exchange – Free registration required
- CISCO TALOS intelligence – Free registration required

Technical environment

Digital assets and connectedness.This is comprised of understanding what types of data are in the enterprise, what the SOC is interested in monitoring, and how it can leverage the CTI it has. To get started, information in this area includes:

- IT/OT architecture showing design of networks, clouds, and perimeters
- User access and account information
- System logs
- Vulnerability information, including patch status
- Endpoint data
- Network sensor data and alerting such as traffic metadata
- The intel correlation capabilities of its sensors and analytic platforms

Leveraging a Cyber Threat Intelligence Platform

These platforms, also called threat intelligence platforms (TIPs), ingest, organize, connect, correlate, and use high volumes of adversary-related data, including IOCs, to enable the SOC to perform more effective threat knowledge management beyond incident-focused tracking.

CTI platforms are generally most powerful when they are integrated with the SOC's other high-scale data processing environments such as a SIEM or big data platform, thereby supporting correlation, data enrichment, and workflow. A CTI platform can enable the SOC to better answer questions including:

- Has this adversary been seen before, and when?
- What adversary activities were exploited in the past in the enterprise?
- Who is reporting on this TTP?
- What reports might be related to the current activity?

Public

Evaluating a CTI Platform

In considering these platforms, the SOC should evaluate the following criteria and tips to select and leverage the right platform for the enterprise:

Workflow and organization

The CTI platform should support collaboration and repeatability around CTI artifact handling, correlation, and organization, particularly around adversary campaign tracking and indicator linkages. This means that the tool supports functionality similar to a case tracking tool (case open/close, analyst notes, and other structured knowledge capture)

Data integration

The CTI platform should be able to ingest, persist, correlate, and interface with many other CTI and other relevant tools, including open-source CTI feeds, commercial CTI feeds, and the SOC's analytic architecture(s) (SIEM, SOAR, big data, etc.).

This means the CTI tool should support both open CTI standards (STIX/TAXII) as well as the APIs of the tools the SOC favors, such as their SIEM/SOAR. This also means that the threat intel management tool supports both batched and NRT data automation in and out of the tool.

Public

Evaluating a CTI Platform

Off-the-shelf feeds

A good CTI platform will offer a “menu” of integration and ingestion from paid and open-source CTI feeds. At the same time, it should also enable and automate deduplication across the feeds ingested.

Feedback and confidence scoring

Analysts should be able to tag, vote, or otherwise score various CTI according to their assessment of the quality and usefulness of the CTI in question. This will support not only an internal feedback loop of the SOC’s own products, but more importantly, scoring against the feeds it receives from others. This should also enable the binning and filtering of IOCs and other intel so that the SIEM/SOAR can be set to only alert on the top-quality CTI sources, according to SOC’s assessment of CTI pedigree

Public

Evaluating a CTI Platform

Access management and identity integration

The CTI platform should support identity integration for the SOC, such that the CTI platform is insulated from risk by general constituency compromise and yet enables analysts to correlate user identity mapping to other workflow activities (email, chat, ticketing, and others).

Confidentiality and source tagging

SOCs that ingest CTI data from many different sources often have data sharing agreements and data handling caveats to preserve and maintain. A good CTI platform should support various tagging and metadata for handled artifacts (IOCs and otherwise) that enable the SOC to be clear on the pedigree and handling caveats of source data, and label data sharing for intel product.

This should include sharing protocols, such as Traffic Light Protocol

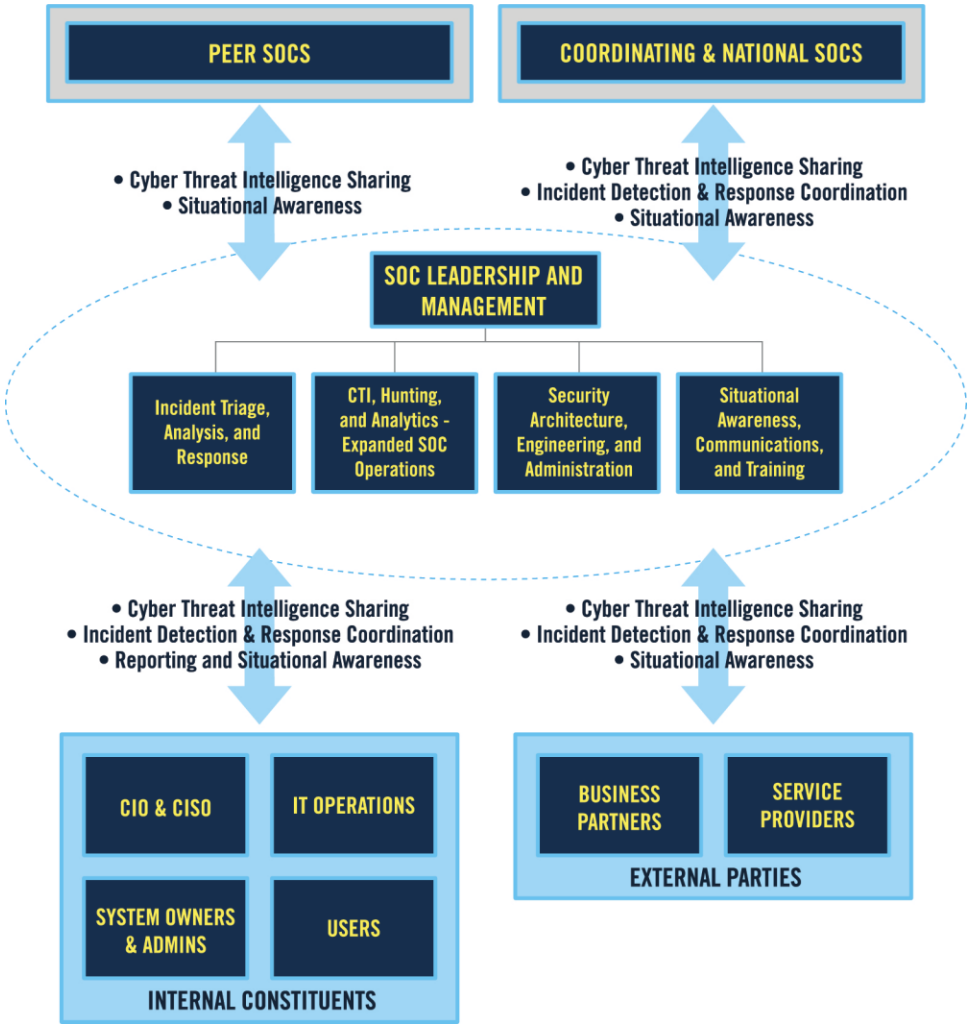
Organizational Relationships

Cyber threat analysts' relationships with personnel outside the SOC will vary depending on the nature of the relationship.

Most users and the IT help desk need to know that potential cybersecurity incidents should be referred to the SOC; they see the SOC as one unit and have no visibility into specific functions like CTI.

Other parties, however, may recognize and interface with cyber threat analysts directly, due to their special role in operations.

Intel analysts are the SOCs early warning system, when they get reports of incoming attacks from threat sharing partners, they relay that information in real time to the IR analysts to take defensive actions.



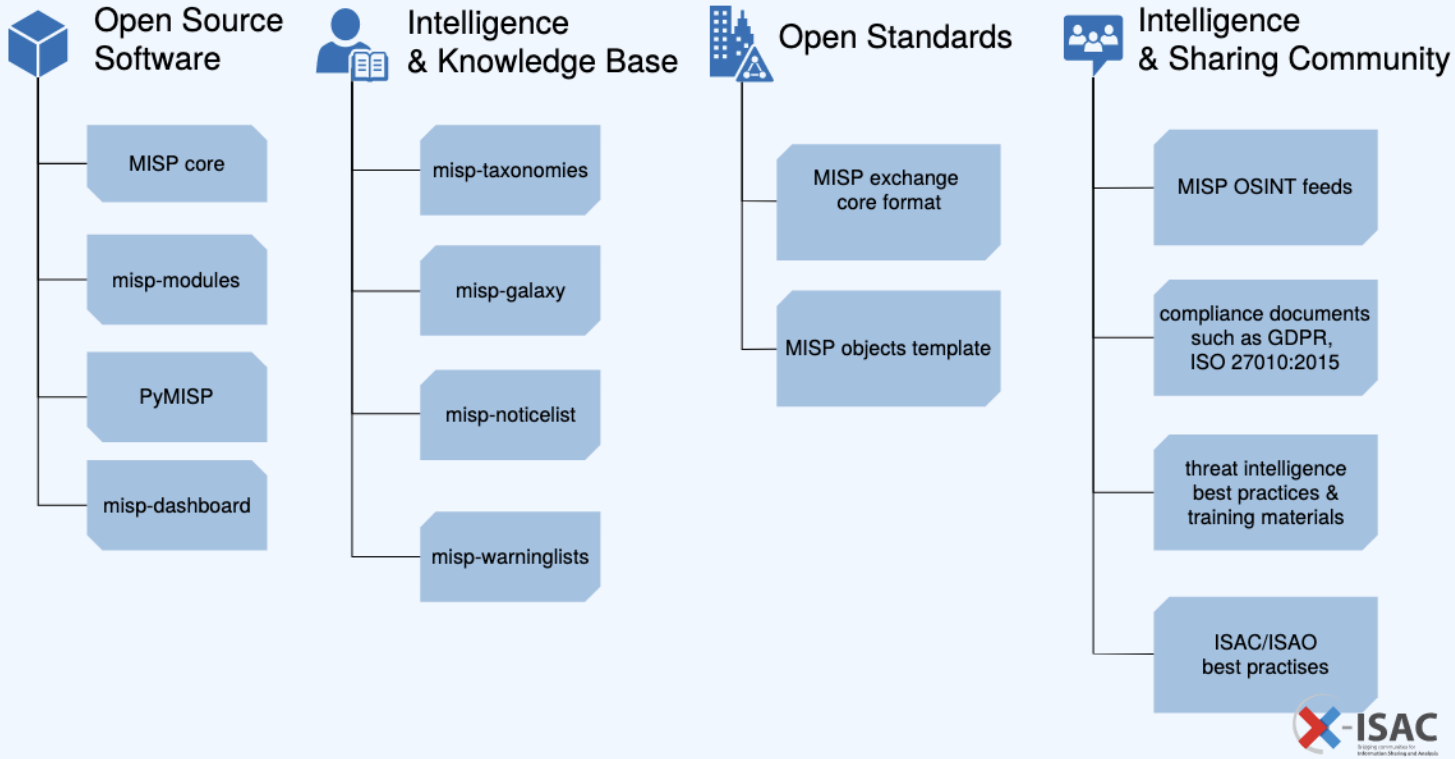
MISP

What is MISP?

- MISP is a **threat information sharing** platform that is free & open source software
- A tool that **collects** information from partners, your analysts, your tools, feeds
- Normalises, **correlates**, **enriches** the data
- Allows teams and communities to **collaborate**
- **Feeds** automated protective tools and analyst tools with the output

MISP

MISP Project Overview



MISP

Sharing in MISP

- Sharing via distribution lists - **Sharing groups**
- **Delegation** for pseudo-anonymised information sharing
- **Proposals** and **Extended events** for collaborated information sharing
- Synchronisation, Feed system, air-gapped sharing
- User defined **filtered sharing** for all the above mentioned methods
- Cross-instance information **caching** for quick lookups of large data-sets
- Support for multi-MISP internal enclaves

MISP

Viper - Main ideas

Viper is a **binary analysis and management framework.** *Its fundamental objective is to provide a solution to* **easily organize** *your collection of* **malware** *and* **exploit samples** *as well as your collection of* **scripts** *you created or found over the time to facilitate your daily research. Think of it as a* **Metasploit for malware researchers**: *it provides a terminal interface that you can use to* **store, search** *and* **analyze** *arbitrary files with and a framework to* **easily create plugins** *of any sort.*

MISP

Viper

- **Solid CLI**
- Plenty of modules (PE files, *office, ELF, APK, ...)
- Connection to **3rd party services** (MISP, VirusTotal, cuckoo)
- Connectors to **3rd party tools** (IDA, radare)
- **Locale storage** of your own zoo
- Django interface is available (I've been told)

MISP funzionalita’

- An **efficient IOC and indicators** database, allowing to store technical and non-technical information about malware samples, incidents, attackers and intelligence.
- Automatic **correlation** finding relationships between attributes and indicators from malware, attack campaigns or analysis. The correlation engine includes correlation between attributes and more advanced correlations like Fuzzy hashing correlation (e.g. ssdeep) or CIDR block matching. Correlation can also be enabled or event disabled per attribute.
- A **flexible data model** where complex objects can be expressed and **linked together to express threat intelligence, incidents or connected elements**.
- Built-in **sharing functionality** to ease data sharing using different model of distributions. MISP can automatically synchronize events and attributes among different MISP instances. Advanced filtering functionalities can be used to meet each organization's sharing policy including a **flexible sharing group** capacity and an attribute level distribution mechanisms.

MISP funzionalita’

- An **intuitive user-interface** for end-users to create, update and collaborate on events and attributes/indicators. A **graphical interface** to navigate seamlessly between events and their correlations. An **event graph** functionality to create and view relationships between objects and attributes. Advanced filtering functionalities and warning lists to help the analysts to contribute events and attributes and limit the risk of false-positives.

- **storing data** in a structured format (allowing automated use of the database for various purposes) with an extensive support of cyber security indicators along fraud indicators as in the financial sector.

- **export**: generating IDS, OpenIOC, plain text, CSV, MISP XML or JSON output to integrate with other systems (network IDS, host IDS, custom tools), Cache format (used for forensic tools), STIX (XML and JSON) 1 and 2, NIDS export (Suricata, Snort and Bro/Zeek) or RPZ zone. Many other formats can be easily added via the misp-modules.

- **import**: bulk-import, batch-import, import from OpenIOC, GFI sandbox, ThreatConnect CSV, MISP standard format or STIX 1.1/2.0. Many other formats easily added via the misp-modules.

- Flexible **free text import** tool to ease the integration of unstructured reports into MISP.

- A user-friendly system to **collaborate** on events and attributes allowing MISP users to propose changes or updates to attributes/indicators.

MISP funzionalita’

- **data-sharing**: automatically exchange and synchronize with other parties and trust-groups using MISP.
- **delegating of sharing**: allows for a simple, pseudo-anonymous mechanism to delegate publication of event/indicators to another organization.
- Flexible **API** to integrate MISP with your own solutions. MISP is bundled with PyMISP which is a flexible Python Library to fetch, add or update events attributes, handle malware samples or search for attributes. An exhaustive restSearch API to easily search for indicators in MISP and exports those in all the format supported by MISP.
- **Adjustable taxonomy** to classify and tag events following your own classification schemes or existing classification. The taxonomy can be local to your MISP but also shareable among MISP instances.
- **Intelligence vocabularies** called MISP galaxy and bundled with existing threat actors, malware, RAT, ransomware or MITRE ATT&CK which can be easily linked with events and attributes in MISP.

MISP funzionalita’

- **Expansion modules in Python** to expand MISP with your own services or activate already available misp-modules.
- **Sighting support** to get observations from organizations concerning shared indicators and attributes. Sighting can be contributed via MISP user-interface, API as MISP document or STIX sighting documents.
- **STIX support**: import and export data in the STIX version 1 and version 2 format.
- **Integrated encryption and signing of the notifications** via GnuPG and/or S/MIME depending on the user's preferences.
- **Real-time** publish-subscribe channel within MISP to automatically get all changes (e.g. new events, indicators, sightings or tagging) in ZMQ (e.g. misp-dashboard) or Kafka publishing.

Public

Fonti

Https://www.misp-project.org/feeds/

- CIRCL OSINT Feed - CIRCL - feed format: misp
- The Botvrij.eu Data - Botvrij.eu - feed format: misp
- blockrules of rules.emergingthreats.net - rules.emergingthreats.net - feed format: csv
- malwaredomainlist - malwaredomainlist - feed format: csv
- Tor exit nodes - TOR Node List from dan.me.uk - feed format: csv
- Tor ALL nodes - TOR Node List from dan.me.uk - feed format: csv
- cybercrime-tracker.net - all - cybercrime-tracker.net - feed format: freetext
- Phishtank online valid phishing - Phishtank - feed format: csv
- listdynamic dns providers - http://dns-bh.sagadc.org - feed format: csv
- ip-filter.blf - labs.snort.org - https://labs.snort.org - feed format: freetext
- longtail.it.marist.edu - longtail.it.marist.edu - feed format: freetext
- longtail.it.marist.edu 7 days - longtail.it.marist.edu - feed format: freetext
- diamondfox_panels - pan-unit42 - feed format: freetext
- pop3gropers - home.nuug.no - feed format: csv
- Feodo IP Blocklist - abuse.ch - feed format: csv
- hosts-file.net - hphost - malwarebytes - hosts-file.net - feed format: csv
- hosts-file.net - hphost - malwarebytes - EMD classification ONLY - hosts-file.net - feed format: csv
- OpenPhish url list - openphish.com - feed format: freetext
- firehol_level1 - iplists.firehol.org - feed format: freetext
- IPs from High-Confidence DGA-Based C&Cs Actively Resolving - osint.bambenekconsulting.com -

INTELMQ

IntelMQ is a solution for IT security teams (CERTs & CSIRTs, SOCs abuse departments, etc.) for collecting and processing security feeds (such as log files) using a message queuing protocol. It's a community driven initiative called **IHAP** (Incident Handling Automation Project) which was conceptually designed by European CERTs/CSIRTs during several InfoSec events. Its main goal is to give to incident responders an easy way to collect & process threat intelligence thus improving the incident handling processes of CERTs.

MISP integrations in IntelMQ

MISP API Collector

The MISP API Collector fetches data from MISP via the MISP API.

Look at the Bots' documentation for more information.

MISP Expert

The MISP Expert searches MISP by API for attributes/events matching the *source.ip* of the event. The MISP Attribute UUID and MISP Event ID of the newest attribute are added to the event.

Look at the Bots' documentation for more information.

MISP Feed Output

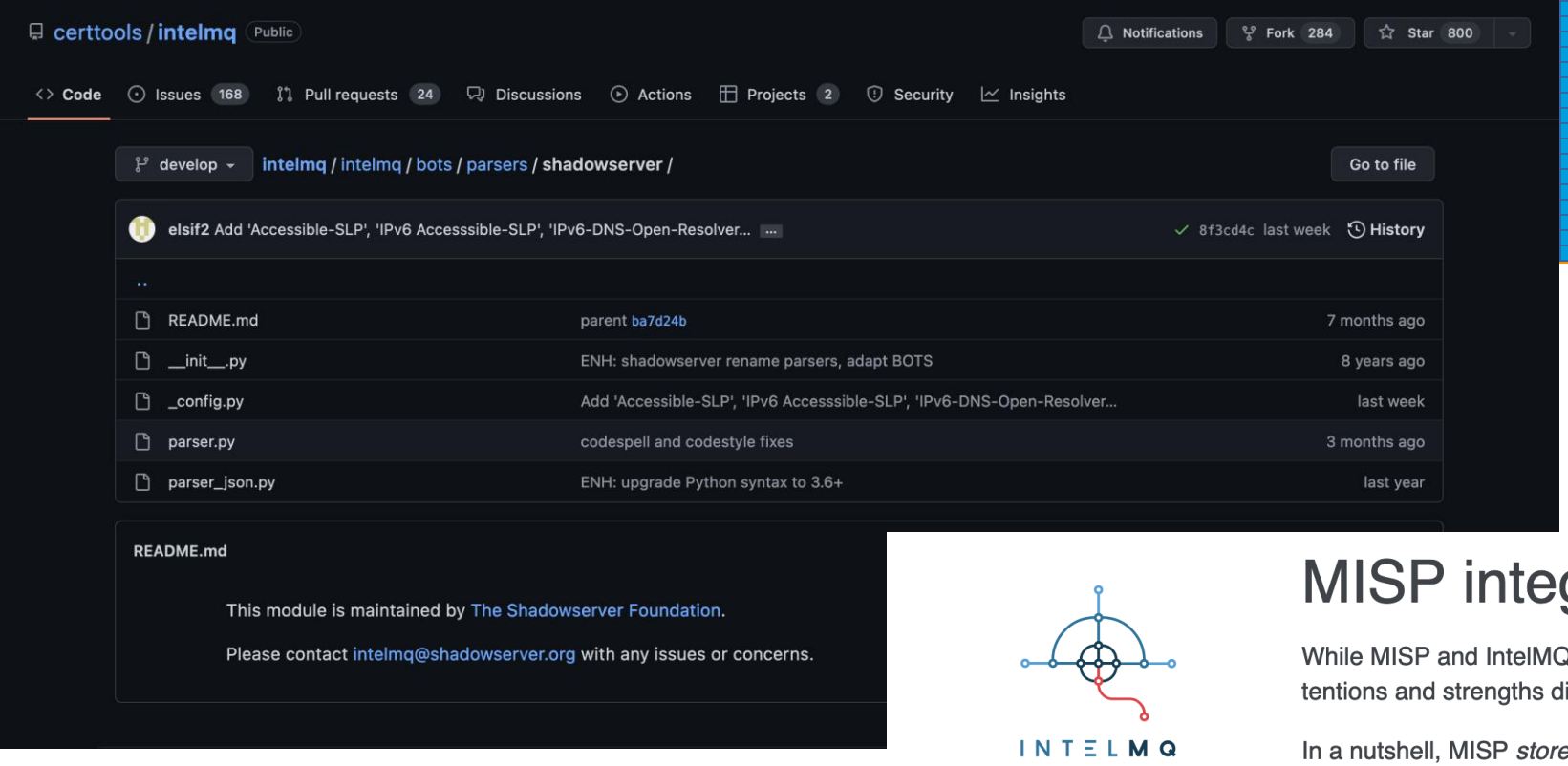
This bot creates a complete "MISP feed" ready to be configured in MISP as incoming data source.

Look at the Bots' documentation for more information.

MISP API Output

Can be used to directly create MISP events in a MISP instance.

Look at the Bots' documentation for more information.



Il problema Shadowserver , possible soluzione con IntelMQ

IntelMQ potrebbe aiutare anche altri feed ingestion e consente l'analisi indipendente dei feed

Misp e' uno strumento estremamente complesso da utilizzare e visto l'alto grado di integrazione e' soggetto possible gli effetti negativi involontari

INTELMQ

IntelMQ is a solution for IT security teams for collecting and processing security feeds using a message queuing protocol.

Navigation

Introduction
IntelMQ Organizational Structure
Getting support
Development
Hardware Requirements
Installation
Upgrade instructions
Configuration and

MISP integrations in IntelMQ

While MISP and IntelMQ seem to solve similar problems in the first hindsight, their intentions and strengths differ significantly.

In a nutshell, MISP *stores* manually curated indicators (called *attributes*) grouped in *events*. An event can have an arbitrary number of attributes. MISP correlates these indicators with each other and can synchronize the data between multiple MISP instances.

On the other side, IntelMQ in it's essence (not considering the EventDB) has no state or database, but is stream-oriented. IntelMQ acts as a toolbox which can be configured as needed to automate processes of mass data with little or no human interaction At the end of the processing the data may land in some database or be sent to other systems.

Both systems do not intend to replace each other or do compete. They integrate seamless and combine each other enabling more use-cases and

MISP API Collector

The MISP API Collector fetches data from MISP via the MISP API.

Look at the Bots' documentation for more information.

Proposte

- MISP dispiegamento
  - Abbiamo un paio di istanze operative una pro e una preprod. Occorre “rimettere in funzione” la prod e fare un po’ di prove di integrazione con la preprod
  - Integrare alcuni OSINT feeds e alcune peer collaboration (CERN)
  - Prime integrazioni raccogliere correlare e generare rules personalizzate per NIDS/DNSFw/ACL (CRITICO!!)
  - Test di IntelMQ come predigestore e normalizzatore di feed
  - Integrazioni con altri strumenti (EDR)