Threat Intelligence Platforms

Giuseppe Manco

Giuseppe Manco

- Research Manager at Institute for high performance computing and networking of the National Research Council of Italy
- Head of the BMSA group
  - Behavioral Modeling and Scalable Analytics
  - 6 Researchers, 4 fellows, 2 associates

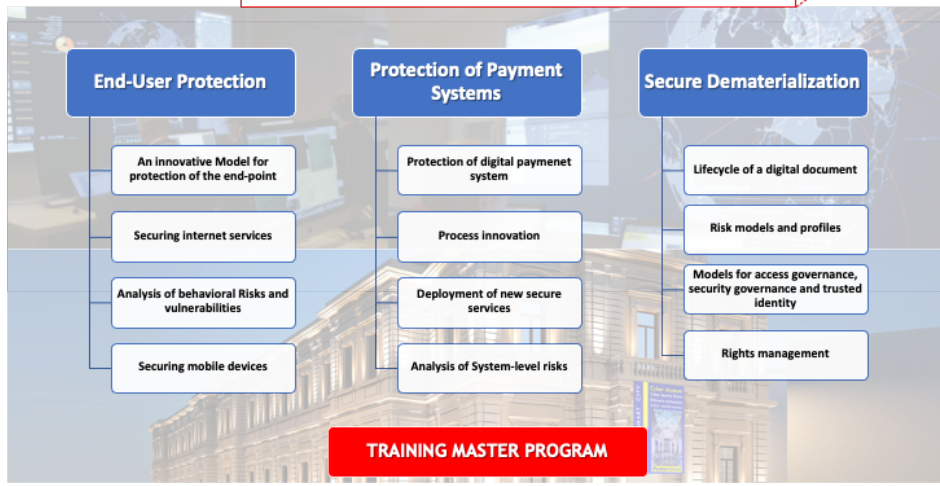
BMSA
BEHAVIORAL MODELING & SCALABLE ANALYTICS
ICAR

Cyber Security for Europe

Cyber Security District - Cosenza

PON Ricerca e Competitività 2007-2013

A SERVICE CENTER FOR
- INDUSTRIAL RESEARCH
- DEVELOPMENT OF INDUSTRIAL PROTOTYPES
- TRAINING SECURITY SPECIALISTS



Agenda

- CTI: What and Why
- Threats, Sources, Intellignce
- Standards & Platforms
- Issues and Challenges
- The CS4E experience

What is Cyber Threat Intelligence?

- A concise definition:

evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard.

What is Cyber Threat Intelligence?

- The collection and analysis of information about threats and adversaries and drawing patterns that provide an ability to make knowledgeable decisions for the preparedness, prevention and response actions against various cyber attacks.
- Involves collecting, researching and analyzing trends and technical developments in the area of cyber threats and if often presented in the form of Indicators of Compromise (IoCs) or threat feeds, provides evidence-base knowledge regarding an organization's unique threat landscape.
- Analysis if performed based on the intent, capability and opportunity. Experts can evaluate and make informed, forward-learning strategic, operational and tactical decisions on existing or emerging threats to the organization.

Motivations

- The static approach of traditional security based on heuristic and signature does not match the dynamic nature of new generation of threats that are known to be evasive, resilient and complex.

Why is it important?

- The number of data breaches is increasing each year
  - Reported breaches was up 54% in 2019 w.r.t 2018
  - Average cost of a data breach is expected to surpass $150 million in 2020
- Sustaining cybersecurity is getting more and more difficult
  - Cyber threats are getting more sophisticated
  - Number of threats and types of threats are increasing
  - Organizations face a shortage of sufficient skilled professionals
- With CTI, organizations gain a deeper understanding of threats and respond to the concerns of the business more effectively

https://research.aimultiple.com/cti/

Threat Intelligence: How?

- *Strategic* - provides high-level information regarding cyber security posture, threats and its impact on business.
- *Operational* - provides information about specific threats against the organization.
- *Tactical* - provides information related to threat actor's Tactics, Techniques and Procedures (TTPs) used to perform attacks.
- Technical - Actionable defense to reduce the gap between advanced attacks and organization defenses means.

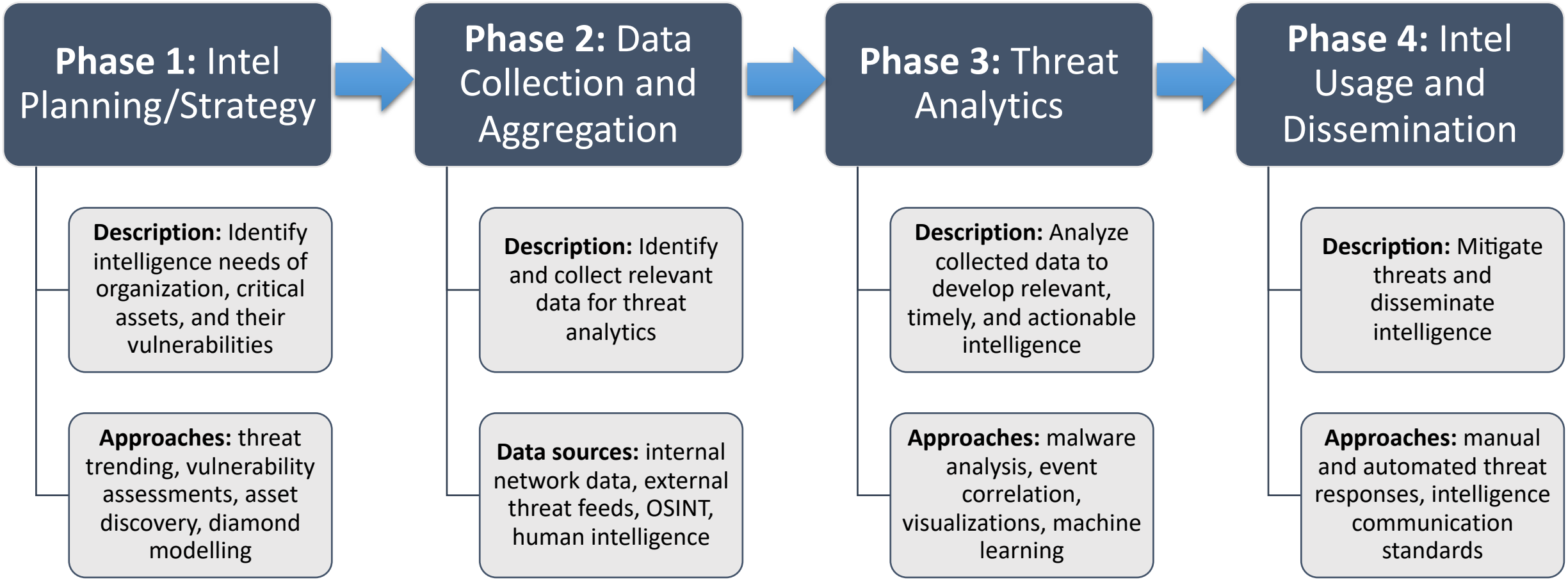
- **Strategic threat intelligence**
  - high-level information consumed by decision-makers
  - Help strategists understand current risks and identify further risks of which they are yet unaware
  - Generally in the form of reports, briefings or conversations
- **Operational threat intelligence**
  - Information about specific impending attacks against the organization. focuses on details of these attacks found in open source intelligence or providers with access to closed chat forums.

- **Tactical threat intelligence**
  - Tactics, Techniques, and Procedures and information about how threat actors are conducting attacks
  - Consumed by incident responders to ensure that their defenses and investigation are prepared for current tactics
  - Gained by reading technical press, white papers, communicating with peers in other organizations to know what they are seeing attackers do, or by purchasing from a provider of such intelligence.
- **Technical threat intelligence (TTI)**
  - Information that is consumed through technical resources
  - Feeds the investigative or monitoring functions of an organization
    - e.g., firewalls and mail filtering devices.
  - Also serves for analytic tools, or just for visualization and dashboards

	Strategic	Operational	Tactical	Technical
Level	High	High	Low	Low
Audience	The board	Defenders	Senior security management; architects	Security Operation Center staff; incident response team
Content	High level information on changing risks	Details of specific incoming attacks	Attackers' tactics, techniques and procedures	Indicators of compromise
Time frame	Long term	Short term	Long term	Immediate

[Tounsi, 2019]

CTI process

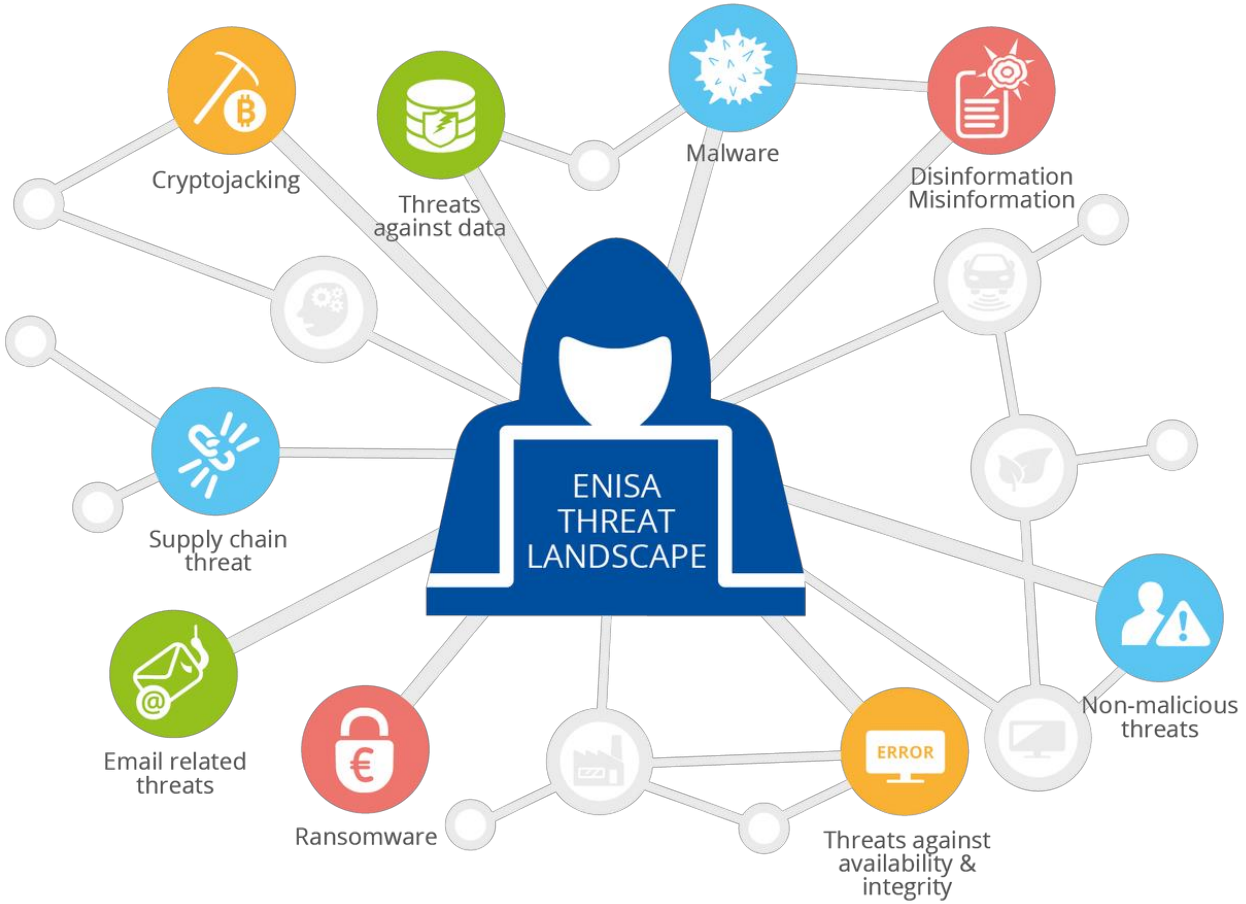

Phase 1: Intel Planning/Strategy
Description: Identify intelligence needs of organization, critical assets, and their vulnerabilities
Approaches: threat trending, vulnerability assessments, asset discovery, diamond modelling
Phase 2: Data Collection and Aggregation
Description: Identify and collect relevant data for threat analytics
Data sources: internal network data, external threat feeds, OSINT, human intelligence
Phase 3: Threat Analytics
Description: Analyze collected data to develop relevant, timely, and actionable intelligence
Approaches: malware analysis, event correlation, visualizations, machine learning
Phase 4: Intel Usage and Dissemination
Description: Mitigate threats and disseminate intelligence
Approaches: manual and automated threat responses, intelligence communication standards

Threats

A (simplified) taxonomy of threats

- multi-vectored
  - attacks can use multiple means of propagation (e.g., web, email, applications)
- multi-staged
  - attacks can infiltrate networks, spread, and ultimately exfiltrate the valuable data

Prime threats in 2021



[ENISA 2021]

Prime threats in 2021

- **Ransomware**
  - A type of malicious attack where attackers encrypt an organisation's data and demand payment to restore access
- **Malware**
  - Software or firmware intended to perform an unauthorised process that will have an adverse impact on the confidentiality, integrity, or availability of a system
- **Cryptojacking**
  - A type of cybercrime where a criminal secretly uses a victim's computing power to generate cryptocurrency
- **E-mail related threats**
- A bundle of threats that exploit weaknesses in the human psyche and in everyday habits, rather than technical vulnerabilities in information systems
- **Threats against data**
  - Data breaches/leaks. A data breach or data leak is the release of sensitive, confidential or protected data to an untrusted environment
- **Threats against availability and integrity**
  - Denial of Service (DoS), Web Attacks. DDoS is one of the most critical threats to IT systems, targeting their availability by exhausting resources, causing decreases in performance, loss of data, and service outages
- **Disinformation – misinformation**
  - Disinformation and misinformation campaigns are on the rise, spurred by the increased use of social media platforms and online media, as well as a result of the increase of people's online presence due to the COVID-19 pandemic
- **Non-malicious threats**
  - Threats where malicious intent is not apparent. Mostly based on human errors and system misconfigurations

Top Trends

- **Ransomware** has been assessed as the **prime threat for 2020-2021.**
- **Cybercriminals are increasingly motivated by monetisation** of their activities, e.g. ransomware. **Cryptocurrency** remains the most common pay-out method for threat actors.
- **Malware decline** that was observed in 2020 continues during 2021.
- The volume of **cryptojacking infections** attained a **record high** in the first quarter of 2021
- **COVID-19 is still the dominant lure in campaigns** for e-mail attacks
- There was a **surge in healthcare sector related data breaches**
- **Traditional DDoS (Distributed Denial of Service) campaigns** in 2021 are more targeted, more persistent and increasingly multivector.
  - The **IoT (Internet of Things)** in conjunction with **mobile networks** is resulting in a new wave of DDoS attacks.
- In 2020 and 2021 there has been a **spike in non-malicious incidents**, as the COVID-19 pandemic became a multiplier for **human errors** and **system misconfigurations**

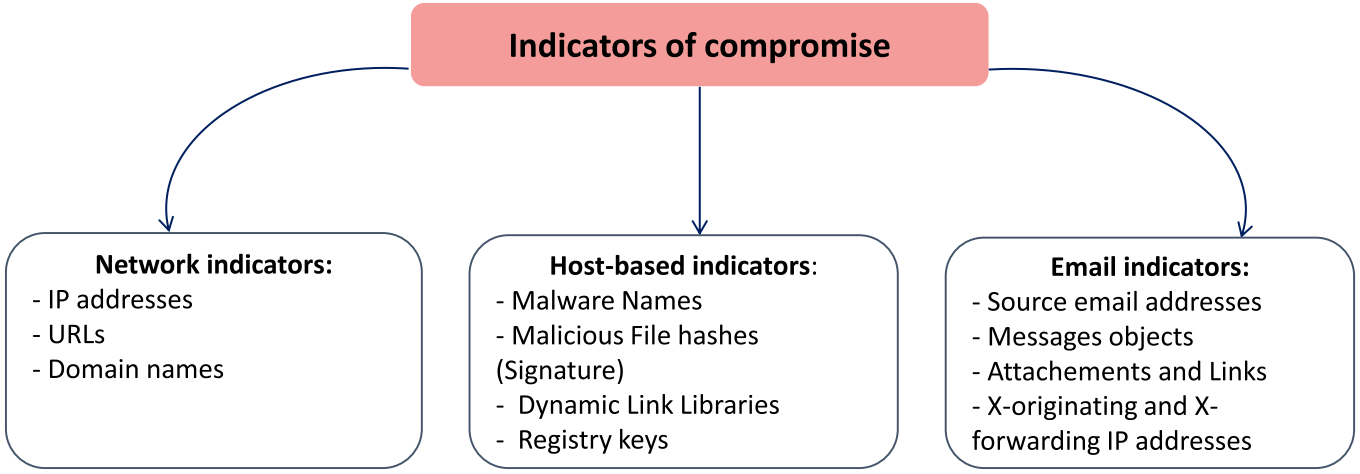
[ENISA 2021]

Challenges

- Advanced persistent threats (APT)
  - Sophisticated network attacks in which an attacker keeps trying until he gains access to a network
    - multi-vectored and multi-staged
- Polymorphic threats
  - cyber attacks, such as viruses, worms or Trojans that constantly change
    - filename changes, file compression, …
- Zero-day threats
  - cyber threats on a publicly unknown vulnerability
- Composite threats
  - exploit technical vulnerabilities in software and/or hardware
  - exploit social vulnerabilities to gain personal information
  - Phishing

Indicators of Compromise (IoC)

- Data fundamentals associated with cyber attacks



IoC: Network Indicators

- Found in URLs and Domain names used for Command & Control (C&C) and link-based malware delivery
  - IP addresses used in detecting attacks from known compromised servers, botnets and systems conducting DDoS attacks
  - Characterized by short lifetime
  - Cloud-based hosting services
    - It is no longer just compromised servers that are used, but also legitimate IP addresses belonging to large corporations.

Network indicators:

- IP addresses
- URLs
- Domain names

IoC: Host-based indicators

- Obtained through analysis of an infected device
- Malware names, decoy documents, file hashes of the malware
  - MD5 or SHA-1 hashes of binaries
- Dynamic Link Libraries (DLLs) are also often targeted
  - E.g., attackers replace Windows system files to ensure that their payload executes each time Windows starts.
- Registry keys added by malicious code
  - Common technique with Trojans

Host-based indicators:
- Malware Names
- Malicious File hashes (Signature)
- Dynamic Link Libraries
- Registry keys

IoC: email indicators

- Created typically when attackers use free email services to send socially engineered emails to targeted organizations and individuals
  - Created from addresses that appear to belong to recognizable individuals
  - Containing intriguing email subject lines
  - Often with attachments and links
  - X-originating and X-forwarding IP addresses
    - email headers identifying the originating IP address of:
      - a client connecting to a mail server
      - a client connecting to a web server through a HTTP proxy or load balancer
  - Monitoring these IP addresses when available provide additional insight into attackers

Email indicators:

- Source email addresses
- Messages objects
- Attachements and Links
- X-originating and X-forwarding IP addresses

Data Sources

IoC sources

- Commonly internal sources
  - crowdsourcing, log and network data, honeynets
- Government-sponsored sources
  - law enforcement, national security organizations
- industry sources
- Open Source INTelligence OSINT
  - Public threat feeds
  - Dshild, ZeuS Tracker, in-house intelligence collection such as attacker forums, social media)
- commercial sources
  - threat feeds, Software- as-a-Service (SaaS) threat alerting, security intelligence providers.

Data Sources

	Internal sources	**External sources**	
	Structured (mainly)	***Structured***	***Unstructured***
Example	Firewall and router logs, honeynets	Vulnerabilities databases, IP blacklists and whitelists, threat data feeds	Forums, news sites, social media, dark web
Technologies for collecting and processing	Feed parser	Feed/web scraper, parser	Collection: crawlers, feed/web parsers Processing: Natural Language Processing (NLP), machine learning

- Open source or public CTI feeds (DNS, MalwareDomainList.com, …)
- Community or industry groups
- Security data gathered from IDS, firewall, endpoint and other security systems
- Media reports and news
- Incident response and live forencis
- SIEM platform
- Vulnerability data
- Network traffic analysis (packet and flow data)
- Forensics
- Application logs
- Closed or dark web sources
- Security analytics platforms
- User access and account information
- Honeypot data
- User behavior data
- Shared spreadsheeds or email

Internal sources

- Internal sources for threat data collected from within the organization specifically internal network and SIEM that being implemented in organization.
  - Threat data from internal network can be in the form of email log, alerts, incident response report, event logs, DNS logs, firewall log, etc.

CTI	Systems	Description
System logs and events	All systems	System activity, principally errors and security events
Network events	Network equipment, (switches, routers, firewalls)	devices connecting/disconnecting, ACL alert, login/failed login, etc.
Network utilisation and traffic profiles	Network equipment, (switches, routers, probes)	SNMP, NetFlow, RMON, etc. to Network management platform
Alerts from boundary devices	IDS/IPS, Firewall, WAF	Alerts/events collected and analysed by SIEM or vendor-specific management portal
AV, system alerts	Corporate AV software installed on host systems, (client and Server)	Corporate AV system alerts from host AV software
Human	All systems	Observed anomalies or events
Forensic	All systems	Artefacts and intelligence gathered after an event

[Ramsdale et al., 2020]

Internal sources

Source	Examples
Network Data Sources	
Router, firewall, Wi-Fi, remote services (such as remote login or remote command execution), and Dynamic Host Configuration Protocol (DHCP) server logs	Timestamp Source and destination IP address Domain name TCP/UDP port number Media Access Control (MAC) address Hostname Action (deny/allow) Status code Other protocol information
Diagnostic and monitoring tools (network intrusion detection and prevention system, packet capture & protocol analysis)	Timestamp IP address, port, and other protocol information Network flow data Packet payload Application-specific information Type of attack (e.g., SQL injection, buffer overflow) Targeted vulnerability Attack status (success/fail/blocked)

[NIST 2016]

Internal sources

Source	Examples
Host Data Sources	
Operating system and application configuration settings, states, and logs	Bound and established network connection and port Process and thread Registry setting Configuration file entry Software version and patch level information Hardware information User and group File attribute (e.g., name, hash value, permissions, timestamp, size) File access System event (e.g., startup, shutdown, failures) Command history
Antivirus products	Hostname IP address MAC address Malware name Malware type (e.g., virus, hacking tool, spyware, remote access) File name File location (i.e., path) File hash Action taken (e.g., quarantine, clean, rename, delete)
Web browsers	Browser history and cache including: • Site visited • Object downloaded • Object uploaded • Browser extension installed or enabled • Cookies

[NIST 2016]

Internal sources

Source	Examples
Other Data Sources	
Security Information and Event Management (SIEM)	Summary reports synthesized from a variety of data sources (e.g., operating system, application, and network logs)
Email systems	Email messages: Email header content • Sender/recipient email address • Subject line • Routing information Attachments URLs Embedded graphic
Help desk ticketing systems, incident management/tracking system, and people from within the organization	Analysis reports and observations regarding: • TTPs • Campaigns • Affiliations • Motives • Exploit code and tools • Response and mitigation strategies • Recommended courses of action User screen captures (e.g., error messages or dialog boxes)
Forensic toolkits and dynamic and/or virtual execution environments	Malware samples System artifacts (network, file system, memory)

[NIST 2016]

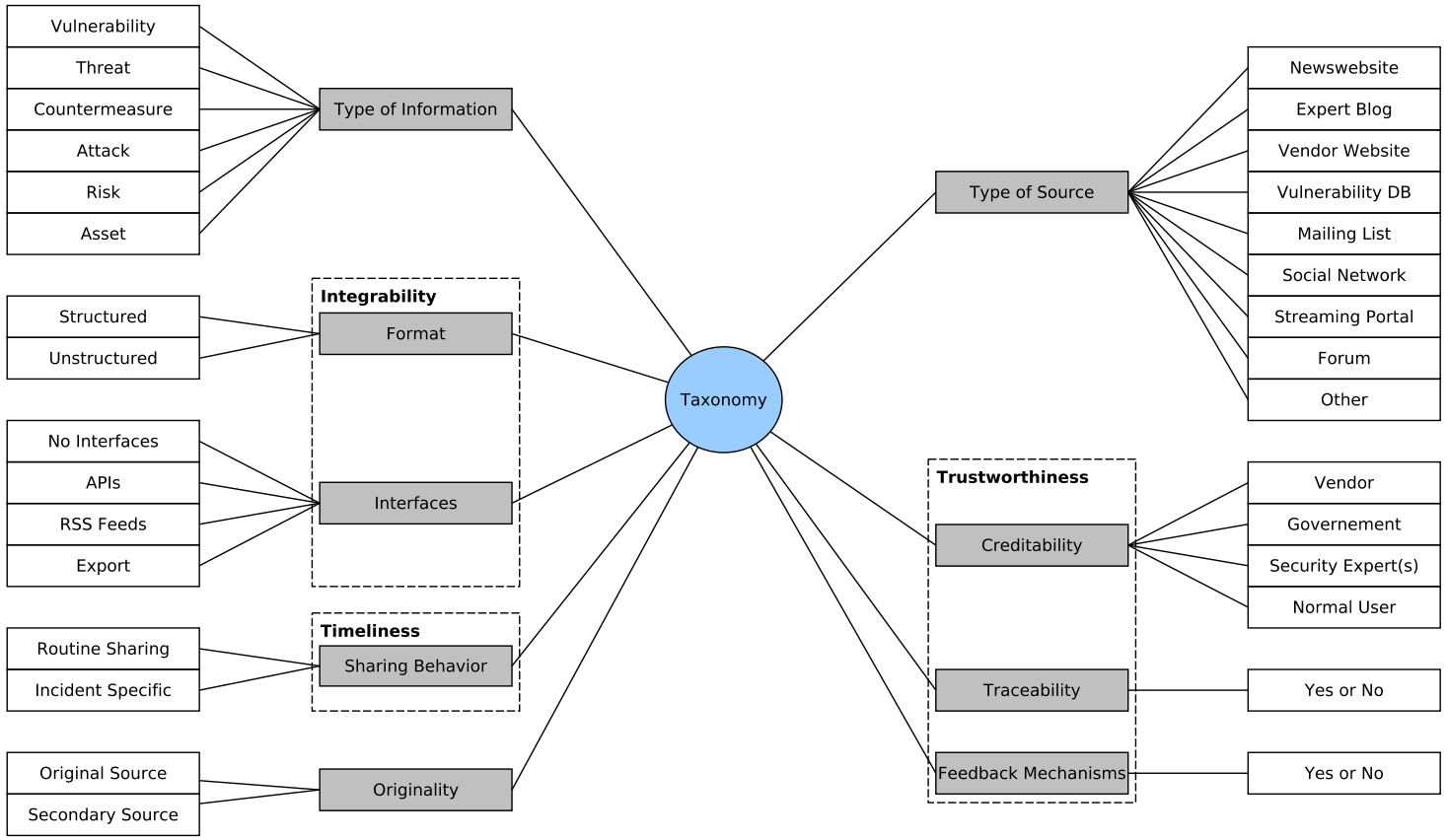
External sources

- External sources have a wide coverage
  - “Open source” intelligence
    - Security researcher, vendor blogs, publicly available reputation and block lists
  - Private or commercial sources
    - threat intelligence feeds, structured data reports, and unstructured reports (such as PDF and Word documents).

Source	Description
News feeds	News articles covering ongoing threats
Vulnerability	Alerts and advisories
Search automation	Using search technologies to find vulnerable systems: Google dorks, Shodan, etc.
Anti-virus vendors	Information, alerts, news feeds on malware activity and threats
Communications	Monitoring communication channels for intelligence: Slack, IRC, Twitter, etc.
Dark web	Intelligence available directly from the criminal underworld

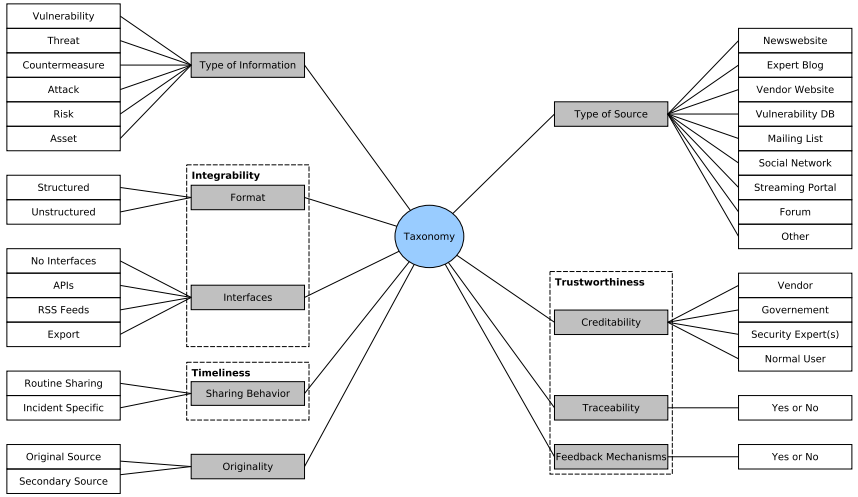
[Ramsdale et al., 2020]

Are external sources reliable?



[Sauerwein et al., 2019]

Are external sources reliable?



	Types of provided information						Integrability						Time-liness		Origi-nality		Trustworthiness					
	Vulnerabilities	Threats	Countermeasures	Attacks	Risk	Assets	Structured	Unstructured	No interfaces	APIs	Feeds	Export	Routine Information Sharing	Incident-Specific	Secondary source	Original source	Vendor	Government	Security Expert(s)	Normal User	Feedback Mechanism (Yes/No)	Traceability of Information (Yes/No)
Newswebsite (21%)	100	73	67	93	53	53	7	93	93	0	7	0	93	53	27	73	13	20	87	13	20	80
Blogs (20%)	92	46	38	77	15	38	0	100	100	0	0	0	69	62	0	100	46	0	54	23	38	85
Vendor Website (13%)	100	33	22	67	33	33	11	89	78	11	22	11	89	100	0	100	89	0	11	0	89	22
Vulnerability Databases (13%)	100	11	22	33	56	11	33	67	22	44	44	33	89	44	67	33	22	22	100	0	67	78
Mailinglists (4%)	100	100	67	100	33	33	0	100	100	0	0	0	67	67	67	33	0	67	100	67	0	33
Social Network (3%)	100	100	100	100	100	100	0	100	50	50	0	0	100	100	50	50	50	50	100	100	100	50
Streaming Portal (3%)	100	50	50	100	0	50	0	100	50	50	0	0	50	50	0	100	50	50	100	50	50	100
Forums (3%)	100	50	50	50	0	50	50	50	50	50	0	50	50	50	0	100	50	0	50	50	0	50
Other (20%)	31	31	54	31	15	8	85	15	23	31	15	31	54	46	15	85	38	8	85	38	54	46
Average percentage	90	53	50	70	32	40	22	78	59	30	10	16	71	65	25	75	43	25	75	41	50	58

[Sauerwein et al., 2019]

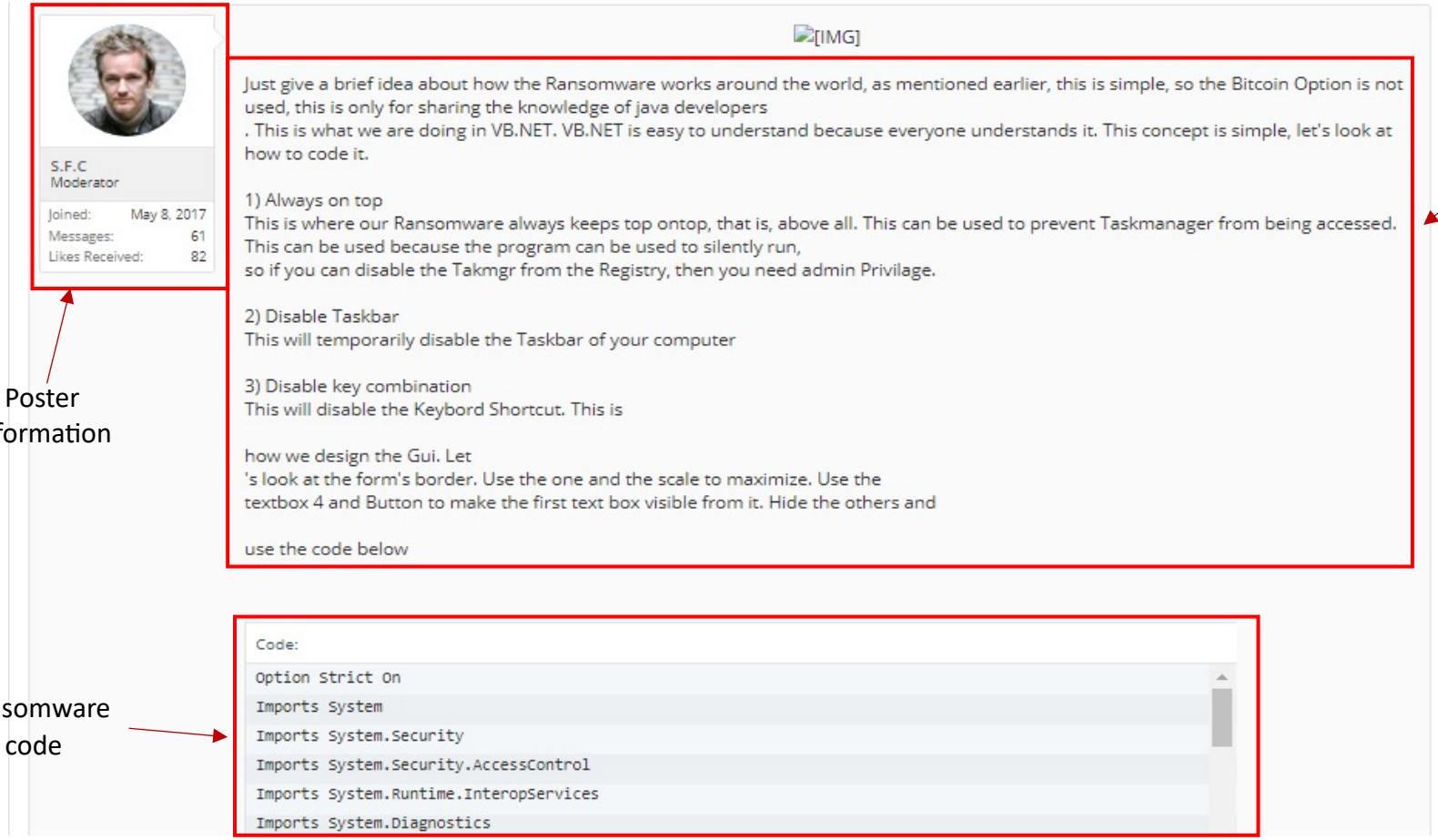
Smart Crawlers: Hacker Community Platforms

Platform	Data Sources	Description	Example Platforms	CTI Value
Hacker Forums	Leaked forums	Forums that have been leaked to the general public	Antichat, Blackhackerz, Blackhat World	-Discussions mentioning past and future attacks -Advertisements for hacking services (e.g., DDoS for hire) -Free hacking tutorials and exploits (e.g., SQLi, BlackPOS) -Identify key threat actors -Discover emerging hacking/threats
	Seized forums	Forums that have been shut down and seized by law enforcement	Darkode, shadowcrew, cardersmarket	
	Active forums	Active, accessible forums that have not been seized or are offline	OpenSC, Ashiyane, reverse4you, exelab	
Carding/Fullz Shops	Carding/Fullz shops	Shops selling stolen credit/debit cards and sensitive information (e.g., Social Security Numbers, drivers licenses, insurance cards)	cardershop, BESTVALID, rescatorccfullz, fullzshop	-Identify breached individuals and organizations -Discover trends of afflicted financial service industries
Internet-Relay-Chat	Active IRC Channels	Clear-text, instant messaging, communication that is not stored	Anonops, whyweprotest, anonet, opddosisis	-Preferred method of communication for hacktivist groups (e.g., Anonymous) -Since chats are not logged, hackers more freely share hacking knowledge and targets
DarkNet Markets	Grams	Search engine for identifying DNMs	–	-Identify markets to collect to generate CTI
	Active market website	Active marketplaces that have not been seized	Minerva, therealdeal, dream market	-Identify new, emerging exploits (0-days, ransomware) -Discover breached content (e.g., logins) -Early indicator for breached companies -Identify key sellers/buyers

[Samtani et al., 2021]

- Underlying Mechanism:
  - Hackers use forums and/or IRC to freely discuss and share Tools, Techniques, and Processes
  - Hackers download tools or navigate to DNMs to purchase exploits
  - These tools help hackers conduct cyber-attacks to attain sensitive data such as credit card and SSN
  - Finally, hackers load stolen data to DNMs and/or carding shops for financial gain

Hacker Forums



An example of a hacker forum member sharing ransomware code

[Du et al., 2018]

Data Collection Overview: IRC

```
02:41 < MaLi> https://forum.deathaddict.com/showthread.php?42-Mr-Hands
02:41 <+Meow> Title: Mr. Hands (deathaddict.com: encrypted)
02:42 < MaLi> Nice deflection snowman.
02:42 < MaLi> Stop being a cuck.
02:42 < Animosity> That link tho
02:43 < The_Snowman> .ud cuck
02:43 <+Effexor> Definition: Another name for the great Onision leader of cuckolds.In Cuck fashion, Onision is
02:43 <+Effexor> Example: There goes lord Onision again being the biggest Cuck that ever cucked in the history
02:43 <+Effexor> Tags: cuckold, cucked, cucks, cock, black cucked, cucking, cuckloaded, fuck, shit, agressor
02:43 < GrnMessiah> http://xfmro77i3lixucja.onion/
```

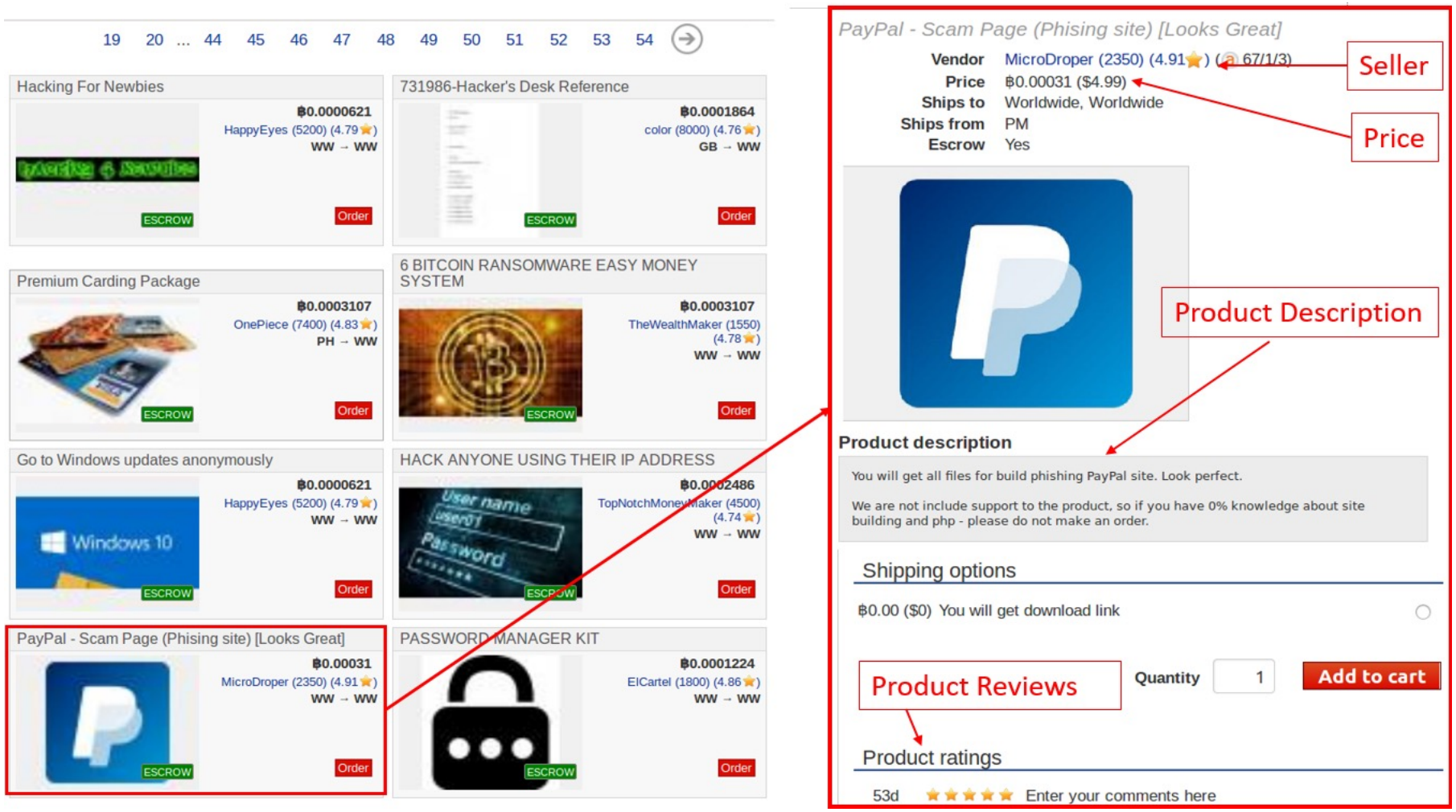
An example of hackers sharing links containing illegal contents

```
11:59 < Gustav> hack this ip 172.98.79.37
11:59 < Gustav> ddos it
11:59 < Gustav> it's my school
12:00 < Gustav> I love you
```

An example of an IRC user demanding hacker service

[Du et al., 2018]

Data Collection Overview: DNM



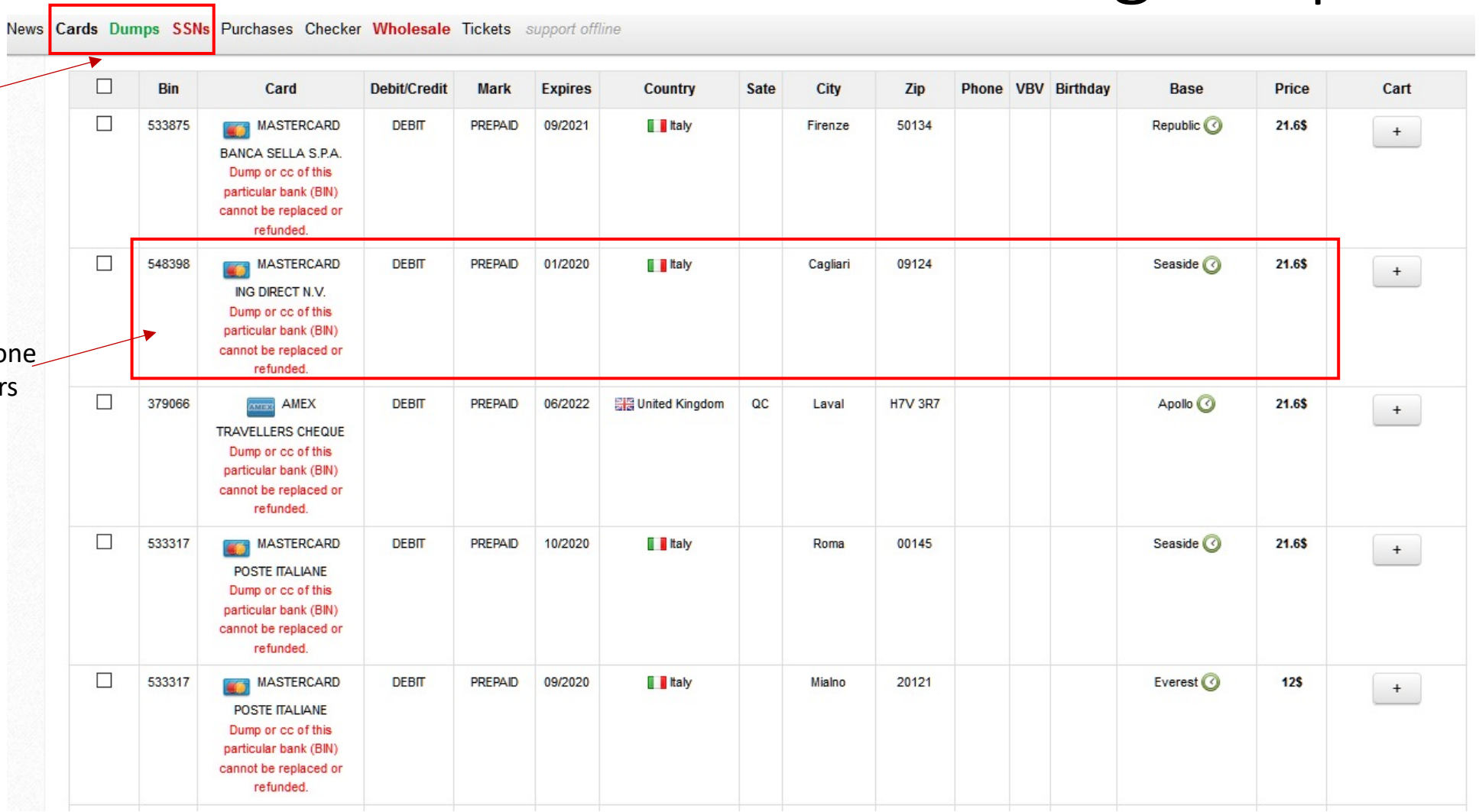
An example of a product listing page on DNM

[Du et al., 2018]

Data Collection Overview: Carding Shop

Card Type

Information of one card for carders



News | Cards | Dumps | SSNs | Purchases | Checker | Wholesale | Tickets | support offline

	Bin	Card	Debit/Credit	Mark	Expires	Country	Sate	City	Zip	Phone	VBV	Birthday	Base	Price	Cart
☐	533875	MASTERCARD BANCA SELLA S.P.A. Dump or cc of this particular bank (BIN) cannot be replaced or refunded.	DEBIT	PREPAID	09/2021	Italy		Firenze	50134				Republic	21.6$	+
☐	548398	MASTERCARD ING DIRECT N.V. Dump or cc of this particular bank (BIN) cannot be replaced or refunded.	DEBIT	PREPAID	01/2020	Italy		Cagliari	09124				Seaside	21.6$	+
☐	379066	AMEX TRAVELLERS CHEQUE Dump or cc of this particular bank (BIN) cannot be replaced or refunded.	DEBIT	PREPAID	06/2022	United Kingdom	QC	Laval	H7V 3R7				Apollo	21.6$	+
☐	533317	MASTERCARD POSTE ITALIANE Dump or cc of this particular bank (BIN) cannot be replaced or refunded.	DEBIT	PREPAID	10/2020	Italy		Roma	00145				Seaside	21.6$	+
☐	533317	MASTERCARD POSTE ITALIANE Dump or cc of this particular bank (BIN) cannot be replaced or refunded.	DEBIT	PREPAID	09/2020	Italy		Mialno	20121				Everest	12$	+

[Du et al., 2018]

Collection Challenges

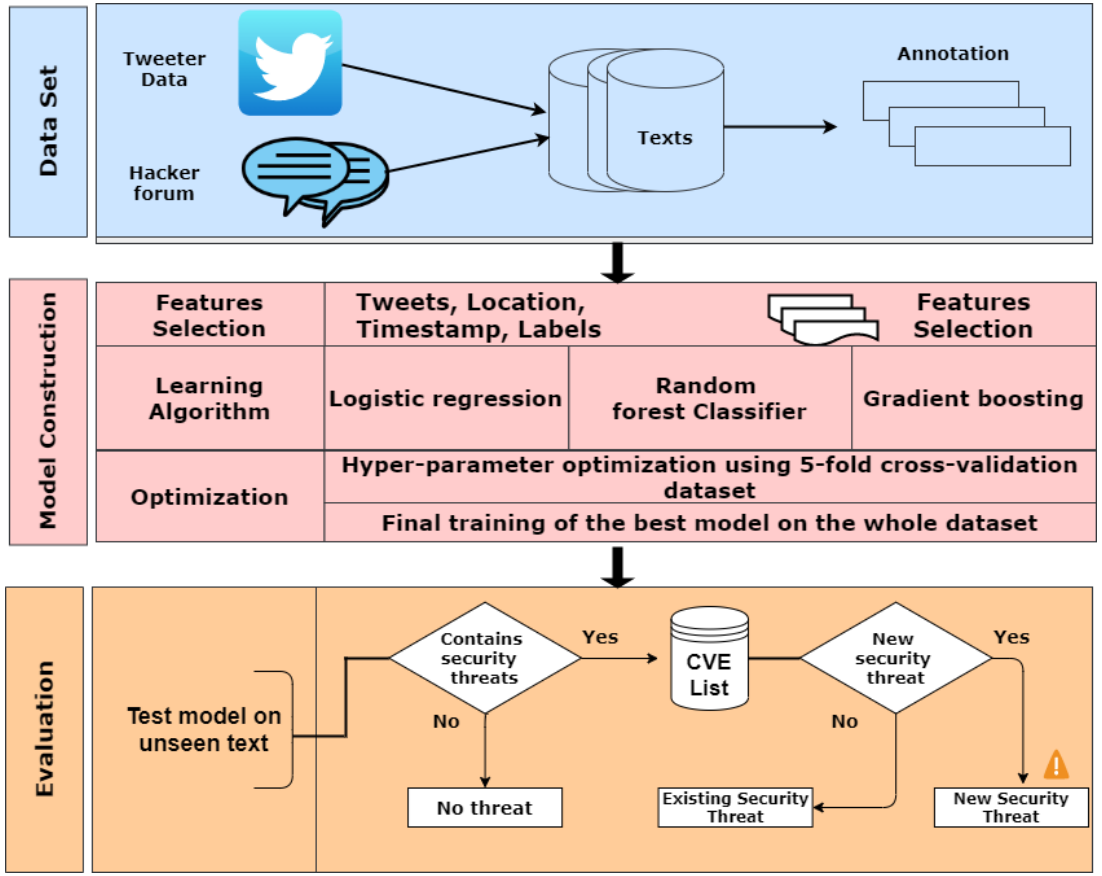
- Anti-crawling measures
  - IP address blacklisting
  - User-agent check
  - User/password authentication & CAPTCHA validation
  - Denial of service for too many requests
- Potential risks of retaliation
  - Constantly probing underground economy platforms may spook platform owners.
  - These owners can trace back to us based on network traffic log.
- Need for secure, intelligent automated collection capabilities

Identifying threats, actors and targets

- Artificial intelligence tools based on machine learning
  - Supervised learning (classification)
  - Unsupervised learning
    - NLP techniques (LDA, Named-Entity Recognition, …), Clustering, correlation analysis
    - Wrapping and information extraction

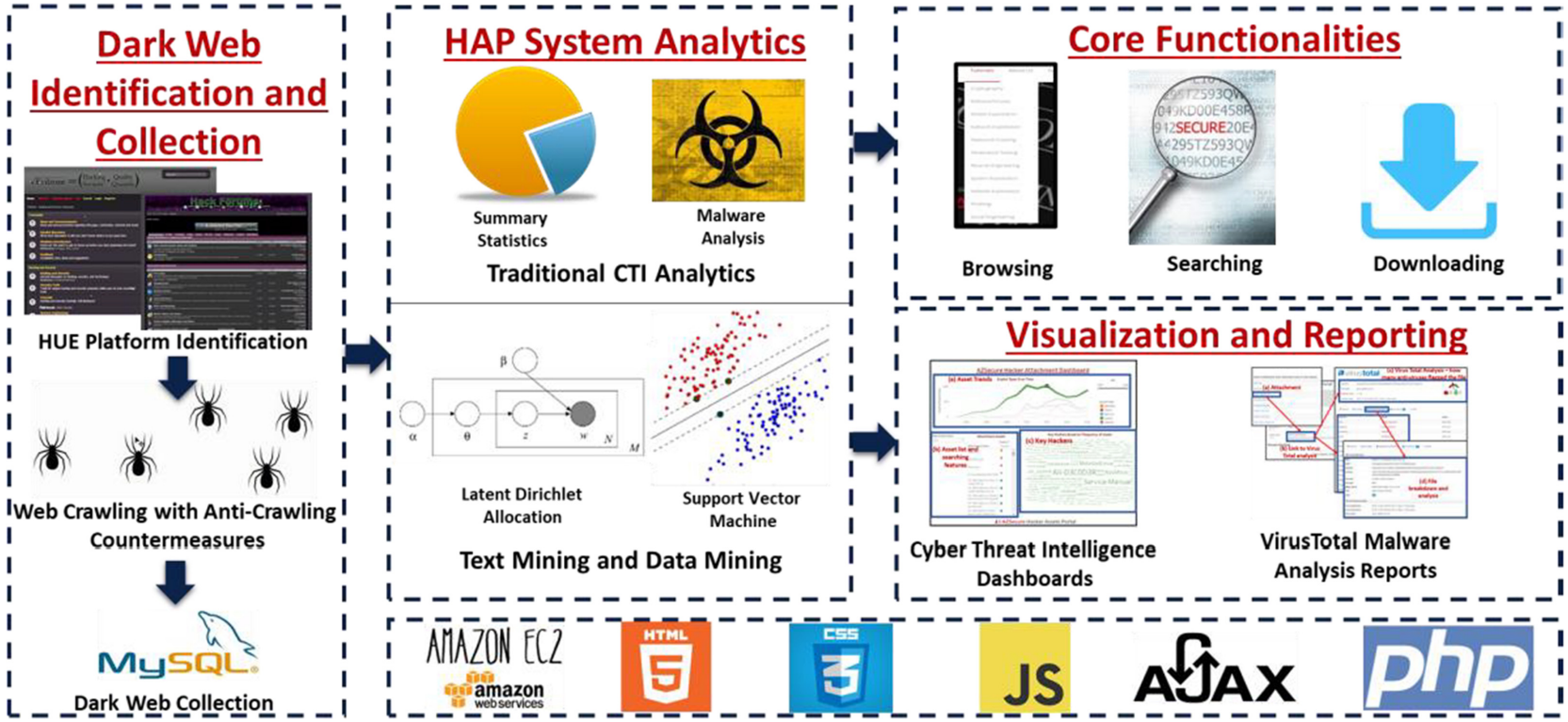
An example: identifying new threats

- An example architecture that analyzes twitter data and Darkweb hacker forums



[Adewopo et al., 2020]

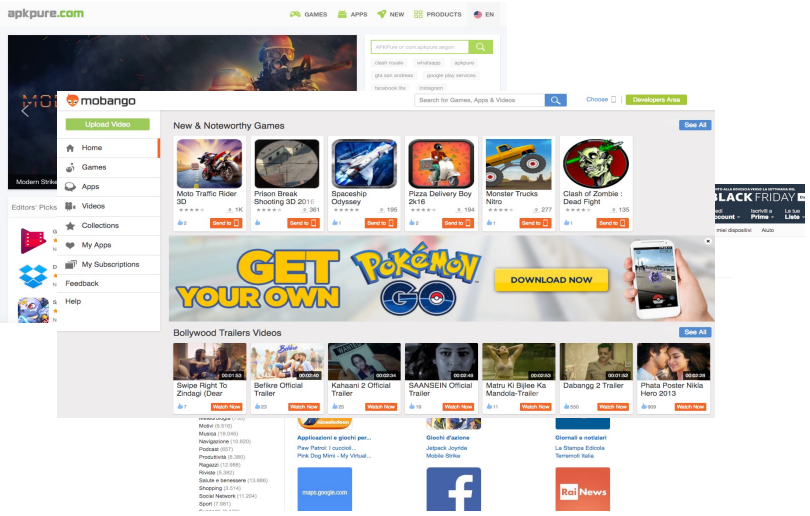
An example: AZSecure Hacker Asset Portal



[Samtani et al., 2021]

An example: Malware spreading in app stores

- The number of frauds perpetrated by means of mobile apps is continuously growing
- Several popular apps are cloned and modified with malicious code
- These apps are spread via alternative markets and app stores



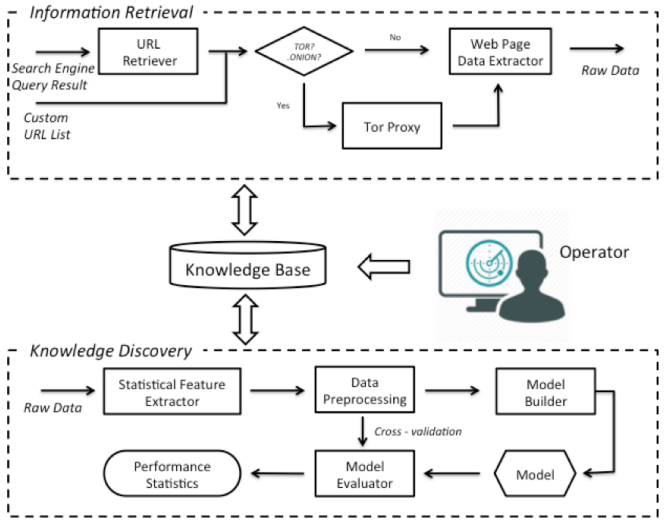
UASD - Unauthorized App Store Discovery

- **Goal**: *Discovering alternative app stores on the (dark) web*
- UASD is a ML-Based framework for the early detection of alternative markets advertised through social media (e.g., Twitter or Facebook) or hosted in the Dark Web
- UASD analyzes web pages extracted from Web pages and, by exploiting a classification model, allows for distinguishing between real app stores and similar pages (e.g., blogs, forums, etc.) which can be erroneously returned by a common search engine

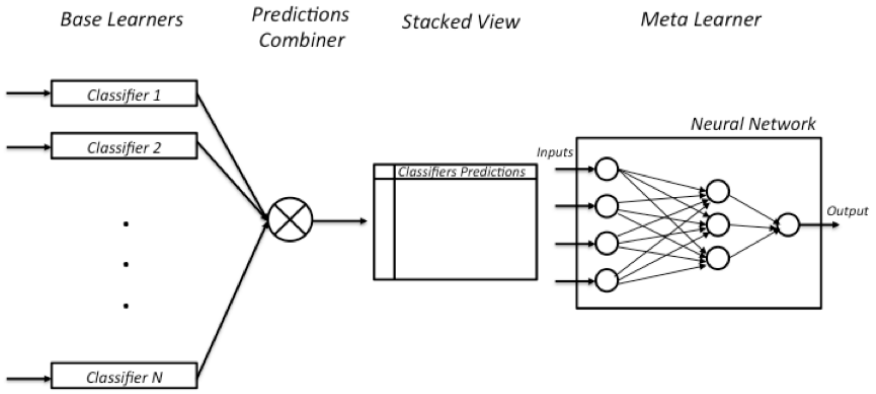
[Guarascio et al,. 2017]

UASD - Details

- Three main macro components (Information Retrieval, Knowledge Discovery and Interaction with the operator)
  - Raw data, extracted from Web and Dark Web, are preprocessed and stored in a Knowledge Base
  - An ensemble-based classification model exploiting a neural network to combine different methods provides a detection score
    - A set of Domain-Specific features are used to improve the classification performances
  - Detection score is used to rank the web pages and to provide a view for the operator in charge of evaluating the proposed links



UASD Framework Architecture

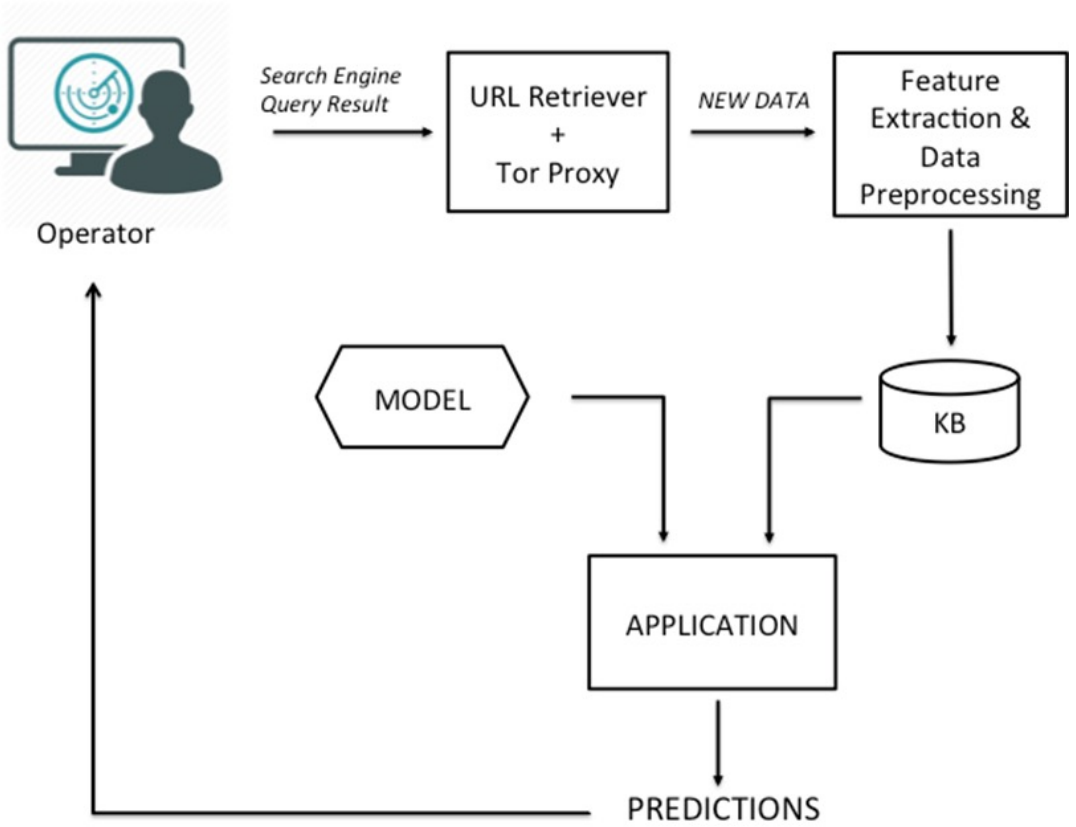


Ensemble-based classification/prediction model

Cyber Security for Europe

UASD – Human in the loop

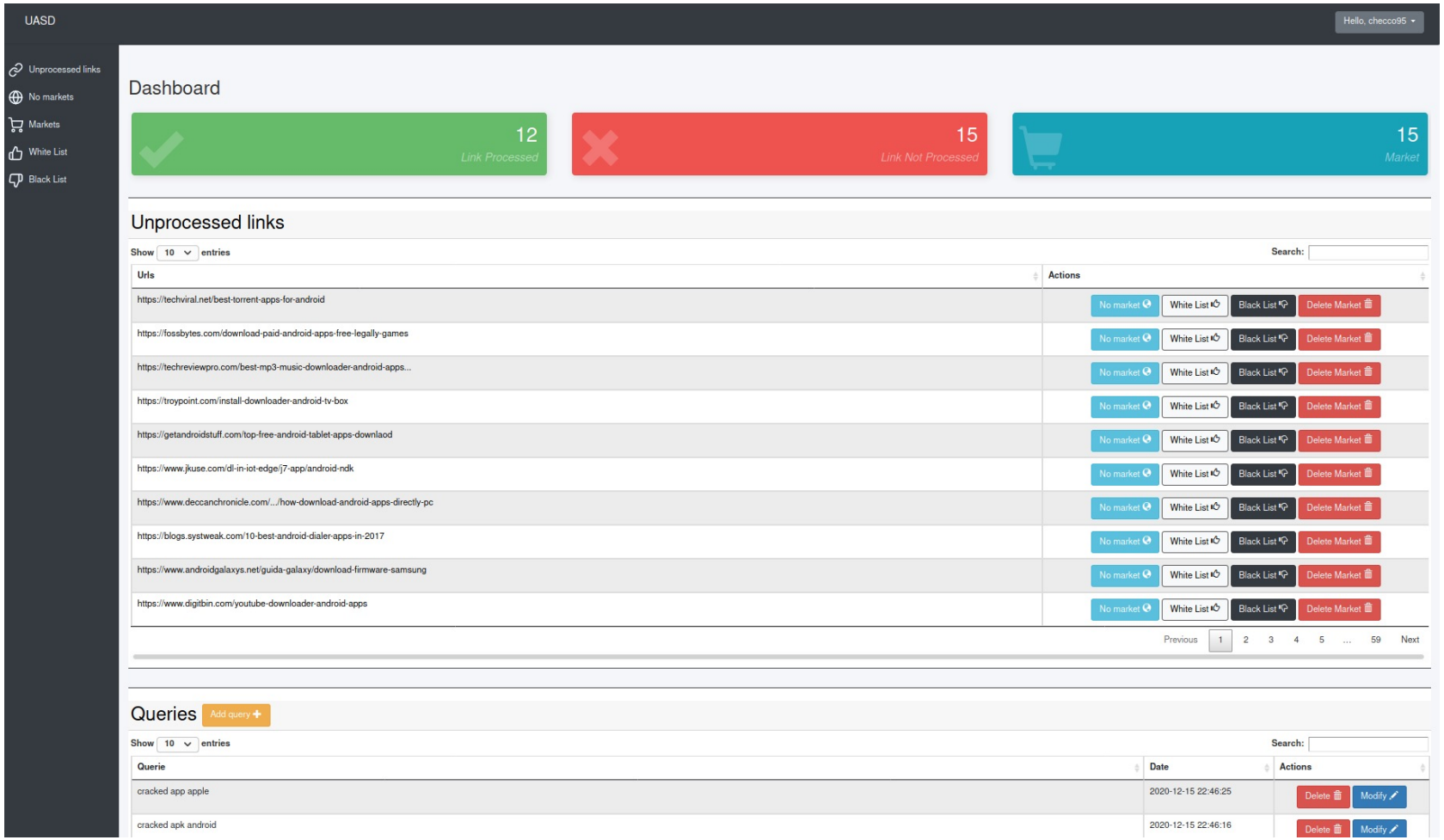
- UASD learns in a continuous fashion
- The operator is the origin of this loop
  - He/she asks a query to be performed and waits for the system response
  - UASD provides a ranked list on the basis of the computed probability scores
  - The domain expert analyzes the proposed web pages and chooses to accept/refuse them
  - The accepted sources are used to enrich the knowledge base (KB) with further positive examples for the learning phase



UASD – Dashboard

Link to be verified

Queries to be processed



Options for the operator

Cyber Security for Europe

Dark Web CTI platforms

Sector	Platform	Dark Web Data Source				Analytics*	Operational Intel*
		Forum	DNM	C. Shop	IRC		
Industry	Verint	√	√	NL	NL	Network/text	Portal, API
	Skybox Security	√	√	NL	NL	NL	Portal, Feeds
	LookingGlass	√	NL	NL	Yes	ML	Portal, API
	Recorded Future	√	√	√	NL	ML, NLP	Portal, Feeds
	Blueliv	NL	√	NL	NL	NL	Portal
	Digital Shadows	√	√	NL	NL	Basic search	Portal, API
	Flashpoint	√	NL	√	NL	Search, SME	API
	Surfwatch Labs	√	√	No	No	SME, search	Portal
	ZeroFox	NL	√	No	No	Search	Portal, API
	CYR3CON	√	√	NL	NL	Rule-based	Blogs, feeds
	DarkOwl	√	√	√	√	NL	Portal, feeds
	Experian	NL	√	√	NL	Search	Portal
Academic	AZSecure DIBBs	√	√	√	√	None	Newsletters
	Intl. CyberCrime Research	√	√	No	No	NL	Newsletters
	IARPA CAUSE	√	√	√	√	ML	Newsletters
	Cambridge Cybercrime Centre	√	No	No	No	None	Newsletters
	IMPACT	No	√	No	No	NL	Papers/data
	MEMEX	√	√	NL	√	NL	Papers/data

* *Note:* NL = Not Listed; ML=Machine Learning; API=Application Programming Interface; SME=Subject Matter Expert; NLP=Natural Language Processing.

[Samtani et al., 2021]

Standards and Platforms

Sharing is the key

Disjoint efforts to understand the complex nature of threats and the tactics and techniques of threat actors behind them give rise to insufficient and fragmented analysis

Benefits and barriers

Category	Benefits	Barriers
Operational	Reduces duplicate information handling Supports breach detection and damage Supports incident response Supports deterrence efforts	Lack of standardisation Capacity limits Accuracy and quality Ensuring timeliness Interoperability and automation Sensitive information
Organizational	Expands professional networks Validates intelligence derived from other sources Improves security posture and situational awareness Combats skills gap	Proliferation of redundant efforts Competition The risk of reputation damage Establishing trust among participants Lack of trained staff
Economic	Cost savings Allows subsidies provision by governments Lowers cyber insurance premiums Reduces uncertainty investment decisions	Resource draining Loss of clients confidence and satisfaction
Policy	Reinforces relationship with government agencies Offers liability protection	The risk of violating privacy or antitrust laws Government over-classification Upholding public values Different legal frameworks across jurisdictions

[Zibak & Simpson, 2019]

Incentives

High	Medium	Low
1. Economic incentives stemming from cost savings; 2. Incentives stemming from the quality, value and use of information shared;	3. The presence of trust among IE participants; 4. Incentives from receiving privileged information from government or security services; 5. Incentives deriving from the processes and structures for sharing; 6. Allowing IE participants' autonomy but ensuring company buy-in;	7. Economic incentives from the provision of subsidies; 8. Economic incentives stemming from gaining voice and influence; 9. Economic incentives stemming from the use of cyber insurance; 10. Incentives stemming from the reputational benefits of participation; 11. Incentives from receiving the benefits of expert analysis, advice, and knowledge; 12. Incentives stemming from participants' personal preferences, values, and attitudes.

[ENISA. 2010]

Challenges

Table 2 – Reasons for not to share.	
1	Fearing negative publicity
2	Legal rules, Privacy issues
3	Quality issues
4	Untrusted participants
5	Believing that the incident is not worth to share
6	Budgeting issues
7	Natural instinct to not to share
8	Changing nature of cyber attacks
9	Unawareness of the victimized organization about a cyber incident
10	Believing that there is a little chance of successful prosecution

[Tounsi, Rais, 2018]

Towards effective sharing

- Legal and regulatory landscape
- Regional and international implementation
- Standardization efforts
- Efficient cooperation and coordination
- Technology integration into organizations

TI sharing initiatives

- Computer Emergency Response Teams (CERTs)
  - Regional coverage
  - collect information on new threats, issue early warnings, provide help on request
- Forum for Incident Response and Security Teams (FIRST)
  - formed in 1990 with the goal of establishing better communication and coordination between incident response teams
- Task Force on Computer Security Incident Response Teams (TF-CSIRT)
  - Sharing statistical data about incidents in order to observe common trends, developing an European accreditation scheme, establishing education and training and assisting new teams
- European Government CSIRTs group (EGC)
  - informal group of governmental CERTs

TI Sharing initiatives

- Information Sharing and Analysis Centers (ISACs)
  - collect, analyze and disseminate private-sector threat information to industry and government and provide members with tools to mitigate risks and enhance resiliency
  - Financial, Oil&Gas, Aviation, Information Technologies, …

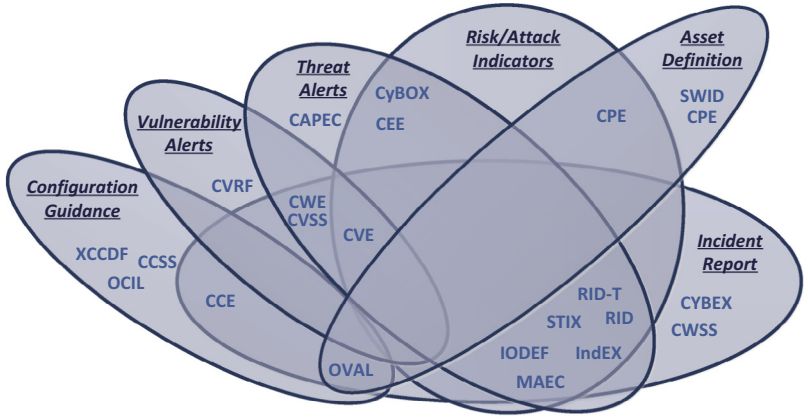
TI Sharing initiatives

- European Network and Information Security Agency (ENISA)
  - Convergence of efforts from the different European institutions and Member States by encouraging the exchange of network and information security threats, methods and results and avoiding duplication of work

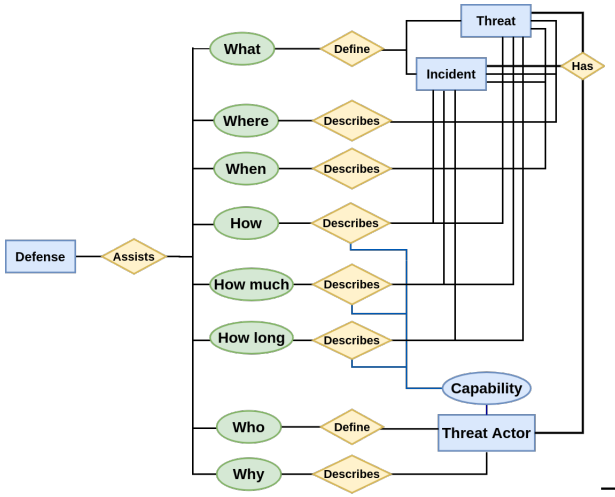
- National Institute of Standards and Technology (NIST)
  - supports the coordination of existing CSIRTs
  - identifies standards, methodologies, procedures, and processes related to Computer Security Incident Coordination (CSIC)
  - provides guidance and best practices on how to cooperate while handling computer security incidents

Standards and protocols

- Several attempts
  - IODEF/RID
  - STIX (Structured Threat Information eXpression), TAXII (Trusted Automated eXchange of Indicator Information),
    - CybOX (Cyber Observable Experssion),
  - OpenIOC (Open Incident of Compromise),
  - VERIS (Vocabulary for Event Recording and Incident Sharing)
  - CAPEC (Common Attack Pattern Enumeration and Classification)
  - MAEC (Malware Attribution and Enumeration Characterization)
  - ATT&CK (Adversarial Tactics, Techniques & Common Knowledge)



[Skopik et al., 2016]



	STIXv2 & TAXII	IODEFv2 & RID	OpenIOC
Holistc Architecture			
Threat	++++	++++	++++
Incident	++++	++++	+++
Threat Actor	++++	++++	++
Defense	++++	++	+
Intelligence Process			
Common formatting	++++	++++	++++
Structured format	++++	++++	++++
Low overhead	+++	+++	+++
Machine readability	++++	+++	++++
Unambiguous data model	++++	+++	++++
Relationship mechanisms	++++	++	+++
Interoperability	++++	+++	+++
Transport mechanism	++++	++++	+
Practical application	++++	++	+++

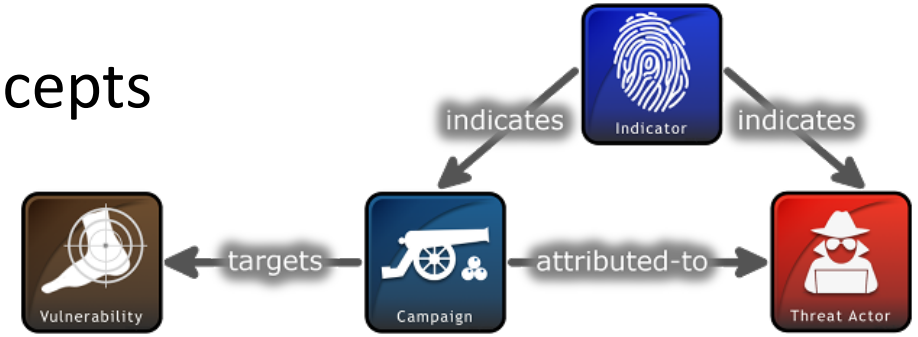
Legend: very high (++++) high (+++) medium (++) low (+).

Data Model Architecture	
Holistic Architecture	Threat
	Incident
	Threat Actor
	Defense
Intelligence Process	
Collection	Common formatting
Processing	Structured format
	Low overhead
	Machine readability
Analysis	Unambiguous data model
	Relationship mechanisms
Deploy	Interoperability
Dissemination	Transport mechanism
	Practical application

[de Melo et al, 2020]

STIX

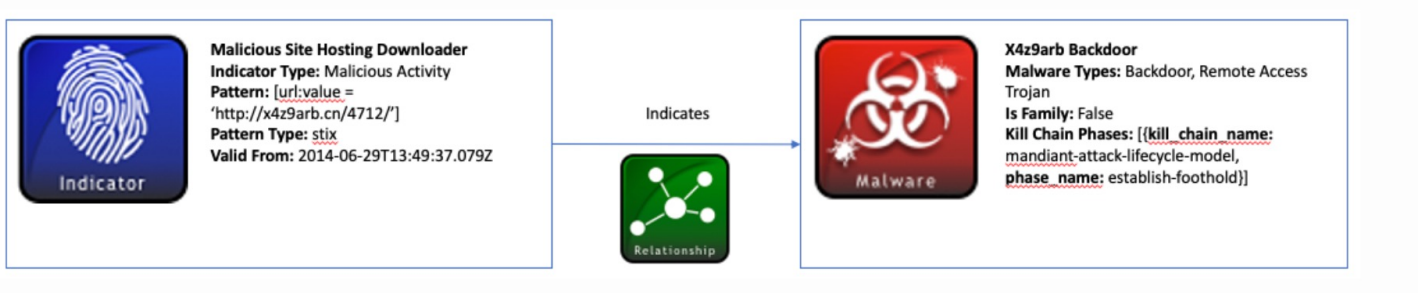
- A language and serialization format used to exchange cyber threat intelligence (CTI).
- Modular architecture
  - Can incorporate other standards efficiently
- Composed of a set of core cyber threat concepts
  - Campaigns
  - Indicators
  - ThreatActors
  - Vulnerabilities
  - …
- Can embed CybOX, IODEF and some OpenIOC extensions
- XML namespaces, extensions for YARA rules, Snort rules and non-XML bindings



Object	Name	Description
	Attack Pattern	A type of TTP that describe ways that adversaries attempt to compromise targets.
	Campaign	A grouping of adversarial behaviors that describes a set of malicious activities or attacks (sometimes called waves) that occur over a period of time against a specific set of targets.
	Course of Action	A recommendation from a producer of intelligence to a consumer on the actions that they might take in response to that intelligence.
	Grouping	Explicitly asserts that the referenced STIX Objects have a shared context, unlike a STIX Bundle (which explicitly conveys no context).
	Identity	Actual individuals, organizations, or groups (e.g., ACME, Inc.) as well as classes of individuals, organizations, systems or groups (e.g., the finance sector).
	Indicator	Contains a pattern that can be used to detect suspicious or malicious cyber activity.
	Infrastructure	Represents a type of TTP and describes any systems, software services and any associated physical or virtual resources intended to support some purpose (e.g., C2 servers used as part of an attack, device or server that are part of defence, database servers targeted by an attack, etc.).
	Intrusion Set	A grouped set of adversarial behaviors and resources with common properties that is believed to be orchestrated by a single organization.
	Location	Represents a geographic location.
	Malware	A type of TTP that represents malicious code.

Object	Name	Description
	Relationship	Used to link together two SDOs or SCOs in order to describe how they are related to each other.
	Sighting	Denotes the belief that something in CTI (e.g., an indicator, malware, tool, threat actor, etc.) was seen.
	Malware Analysis	The metadata and results of a particular static or dynamic analysis performed on a malware instance or family.
	Note	Conveys informative text to provide further context and/or to provide additional analysis not contained in the STIX Objects, Marking Definition objects, or Language Content objects which the Note relates to.
	Observed Data	Conveys information about cyber security related entities such as files, systems, and networks using the STIX Cyber-observable Objects (SCOs).
	Opinion	An assessment of the correctness of the information in a STIX Object produced by a different entity.
	Report	Collections of threat intelligence focused on one or more topics, such as a description of a threat actor, malware, or attack technique, including context and related details.
	Threat Actor	Actual individuals, groups, or organizations believed to be operating with malicious intent.
	Tool	Legitimate software that can be used by threat actors to perform attacks.
	Vulnerability	A mistake in software that can be directly used by a hacker to gain access to a system or network.

https://oasis-open.github.io/cti-documentation/stix/intro
https://oasis-open.github.io/cti-documentation/examples/visualized-sdo-relationships



A scenario consisting of an indicator for a URL and a backdoor piece of malware associated with it.

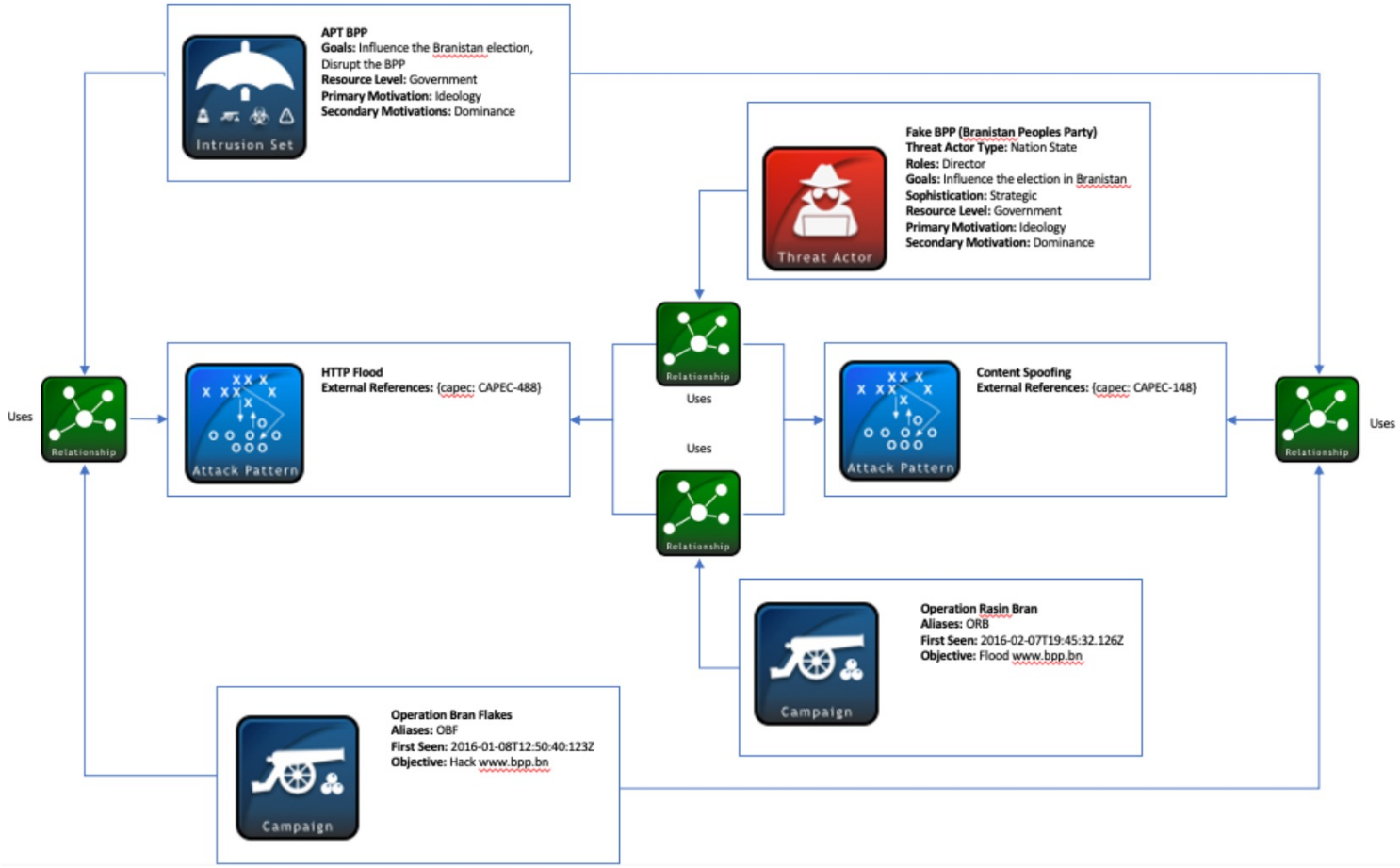
- The site has been shown to host this backdoor malware
- the malware has been known to download remote files.


```
{
    "type": "bundle",
    "id": "bundle--56be2a3b-1534-4bef-8fe9-602926274089",
    "objects": [
        {
            "type": "indicator",
            "spec_version": "2.1",
            "id": "indicator--d81f86b9-975b-4c0b-875e-810c5ad45a4f",
            "created": "2014-06-29T13:49:37.079Z",
            "modified": "2014-06-29T13:49:37.079Z",
            "name": "Malicious site hosting downloader",
            "description": "This organized threat actor group operates to create profit from all types of crime.",
            "indicator_types": [
                "malicious-activity"
            ],
            "pattern": "[url:value = 'http://x4z9arb.cn/4712/']",
            "pattern_type": "stix",
            "valid_from": "2014-06-29T13:49:37.079Z"
        },
        {
            "type": "malware",
            "spec_version": "2.1",
            "id": "malware--162d917e-766f-4611-b5d6-652791454fca",
            "created": "2014-06-30T09:15:17.182Z",
            "modified": "2014-06-30T09:15:17.182Z",
            "name": "x4z9arb backdoor",
            "description": "This malware attempts to download remote files after establishing a foothold as a backd
            "malware_types": [
                "backdoor",
                "remote-access-trojan"
            ],
            "is_family": false,
            "kill_chain_phases": [
                {
                    "kill_chain_name": "mandiant-attack-lifecycle-model",
                    "phase_name": "establish-foothold"
                }
            ]
        },
        {
            "type": "relationship",
            "spec_version": "2.1",
            "id": "relationship--864af2ea-46f9-4d23-b3a2-1c2adf81c265",
            "created": "2020-02-29T18:03:58.029Z",
            "modified": "2020-02-29T18:03:58.029Z",
            "relationship_type": "indicates",
            "source_ref": "indicator--d81f86b9-975b-4c0b-875e-810c5ad45a4f",
            "target_ref": "malware--162d917e-766f-4611-b5d6-652791454fca"
        }
    ]
}
```


https://oasis-open.github.io/cti-documentation/stix/intro

A scenario representing an advanced persistent threat (APT) intrusion set

- Suspected to be funded by the country “Franistan”.
- Target is the Branistan People’s Party (BPP),
- Two sophisticated campaigns and attack patterns
  - Insert false information into the BPP’s web pages,
  - DDoS effort against the BPP web servers.

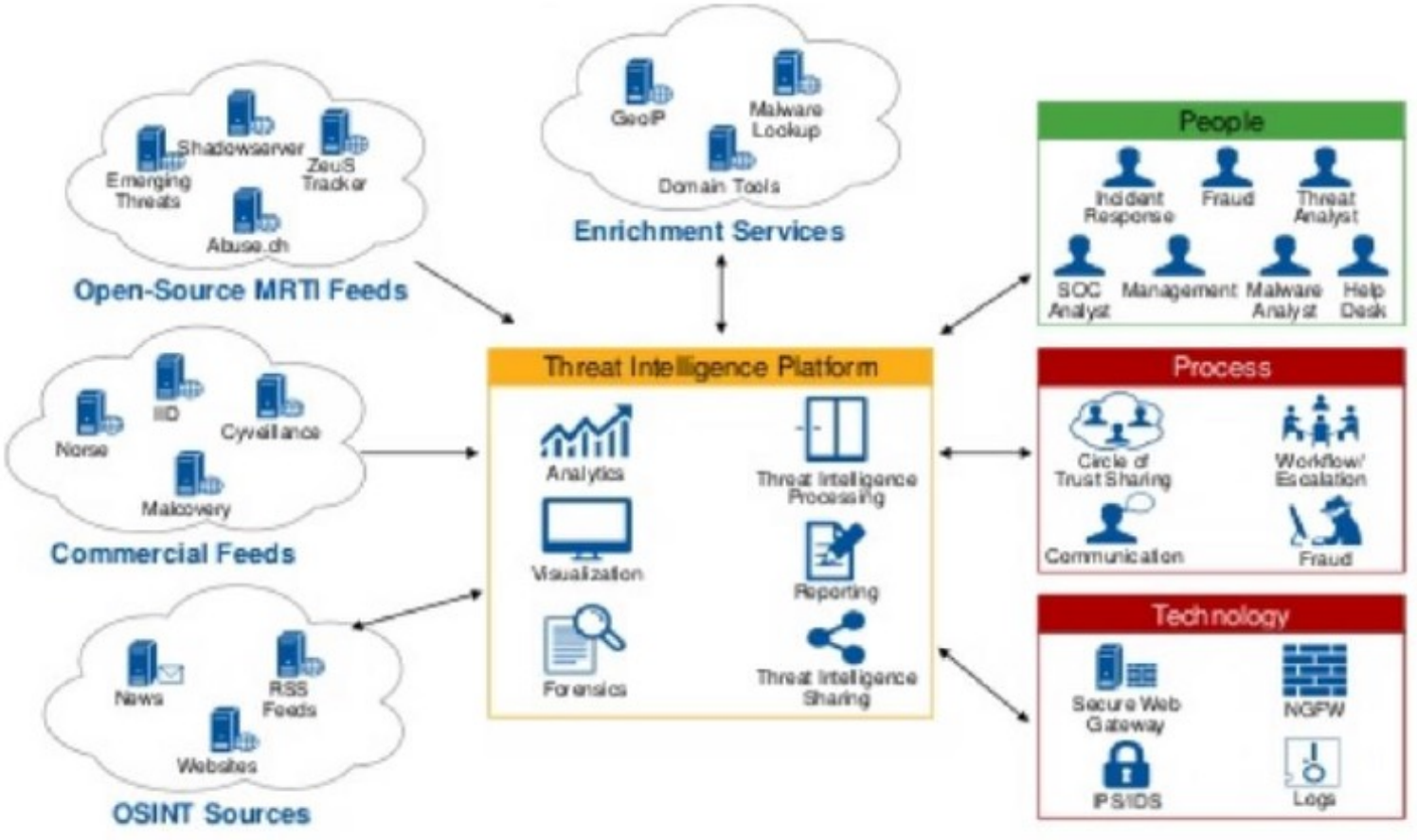


https://oasis-open.github.io/cti-documentation/stix/intro

Threat Intelligence Platforms

- Designed to solve the collection and storing problems of TTI and to facilitate sharing threat information with other organizations in the threat intelligence space
- An emerging technology discipline that supports organizations' threat intelligence programs and helps them to improve their cyber threat intelligence capabilities
  - TIPs enable organizations to easily bootstrap the core processes of collecting, normalizing, enriching, correlating, analyzing, disseminating and sharing of threat related information
  - Generally organized as large repositories that often use big data technologies (e.g. graph analysis and data warehousing) to draw links between types of TTI, allowing quicker response to detected threats, as well as a historical record of an IOC

TIP: Threat Intelligence Platforms



Who can use TIPs?

Role	Contributions	Needs and challenges
SOC Analysts	• provide feedback on indicators • annotate indicators based on observations, alerts and actions taken	• Enhanced context and low false positive rate • Automated data enrichment to reduce repetitive work. • Good integration with SIEM tools
Incident responders, cyber fraud analyss	• new indicators and malware samples coming from investigations	• need tailored and ad-hoc intelligence • need detailed context and enrichment over the indicators provided Lack of visibility into events across different systems or domains
CTI analysts	• Responsible for anything that goes in and out of the TIP • Enrich and analyse the data within TIP as well as linking intelligence Share intelligence with stakeholders	• centralised platform for managing TI • Too much threat intelligence information • Lack of threat intelligence best practices
Threat researchers	• High quality original research	• API support • Customization capabilities
Vulnerability analysis	• Provide insight on the vulnerability exposures	• Intelligence on high impact vulnerabilities
Decision makers	• Sharing policy • Security investment	• Need high level reports on exposures • Need to evidence of the ROI • Assurance that intelligence sharing does not expose the organisation.

[ENISA, 2017]

Commercial Threat Intelligence Information Systems

- TruSTAR: https://www.trustar.co/
- EclecticIQ: https://www.eclecticiq.com/platform
- LookingGlass Cyber: https://www.lookingglasscyber.com
- ThreatQ: https://www.threatq.com/
- IBM: https://www.ibm.com/security/solutions/stop-threats
- Kaspersky: https://www.kaspersky.com/enterprise-security/threat-intelligence
- FireEye: https://www.fireeye.com/solutions/cyber-threat-intelligence.html
- Cisco: https://www.cisco.com/c/en/us/products/security/threat-response.html
- …

Open Threat Intelligence Solutions

- MISP: https://www.misp-project.org/
  - Open source software solution for collecting technical and non-technical information about malware and attacks, storing data in a standardized format, and distributing and sharing cyber security indicators and malware analysis with trusted parties
- OpenCTI: https://www.opencti.io/
  - An open source framework with the main objective of aggregating, in a comprehensive way, general and technical information from the CTI context
- CRITs: https://crits.github.io/
  - Provides analysts with the means to conduct collaborative research into malware and threats. Employs a simple but very useful hierarchy to structure cyber threat information
- CIF: https://csirtgadgets.com/collective-intelligence-framework
  - Assists users in formatting, normalizing, processing, storing, sharing and building threat data sets
- OTX: https://www.alienvault.com/open-threat-exchange
  - Supports collection (via pulse), analysis and distribution of TI. Enables trust and privacy mechanisms
- Yeti: https://yeti-platform.github.io/
  - a platform meant to organize observables, indicators of compromise, TTPs, and knowledge on threats in a single, unified repository. Capable of automatially enriching observables.
- …

Desiderata

- Which software functions are required by cyber threat intelligence sharing platforms to support the processes of the intelligence cycle

Intelligence Processes	Functions
Planning & Direction	-
Collection	*Manual Data Creation, Manual File Upload, Feed Import, Import Connector, Import Agent, Web Collector*
Pre-Processing	*Data Cleaning, Data Normalization, Data Classification, Data Editing*
Analysis	*Expert Analysis, Collaborative Analysis, Data Investigation & Sandboxing, Search, Statistical Analysis, Correlation, Pattern Recognition, Rating & Prioritization, White- & Blacklisting, Monitoring, Prediction*
Dissemination	*Feed Export, Alerting & Notifications, Synchronization & Export Connector, Manual Download*
Evaluation & Feedback	*Dashboard, Standardized Reporting, Individual Reporting, Feedback*
Cross-Process Support	*Data Security, Communication Security, Platform Security, Access Control, Data Privacy, Group and Community Management, Communication & Messaging, Teamworking, Data Verification, Data Validation, Rating, Reputation, Traceability*

[Sauerwein at al., 2021]

The maturity level

Tool / Criteria	Import format[a]	Integration with/ export to standard security tools[b]	Support of collaboration	Data exchange standards	Analysis	Graph generation	License
MISP	bulk-import, batch-import, OpenIOC import, GFI sandbox, ThreatConnect CSV, JSON, OCR, VMRAY	(1) generating OpenIOC, plain text, CSV, MISP XML or JSON output to integrate with network IDS, host IDS. (2) generating network IDS data to export to Suricata, Snort and Bro or RPZ zone. (3) integration with SIEM using a restful API	Private instance or multiple instances interconnected with a selected community (many sharing options)	STIX, CybOX, TAXII[c]	(1) Analysis of the history records and displaying a trend (2) Correlation of analysis finding relationships between attributes and indicators (3) May include any other result from additional analysis of malware like tools output.	misp-graph to analyze a MISP XML, export and generate graphs from correlation between events and IOC. The export formats: Graphviz and gexf files	Open source (GNU General Public License)
CRITs	bulk-import via CSV file, blob, and spreadsheet, STIX CybOx, TAXII	(1) STIX CybOx, TAXII, CSV to export to network IDS and host IDS (2) a RESTful API for import/export/updates (3) Other services readily available that integrate with external sources and services[d]	Private instance or shared with a trusted community	STIX, TAXII, OpenIOC; Send/receive information through Facebook's ThreatExchange[d]	(1) Analysis of uploaded files with the possibility to link a Cuckoo sandbox (2) Upload threat data and automatically uncover critical information (3) Analysis of Samples, PCAPs, etc.	mcrits to visualize CRITs DB via local Maltego transforms.	Open source (GNU General Public License)
CIF v3	XML, JSON, Zip archives,[e]	Output into multiple formats (CSV, JSON, html, table) to integrate with various tools including Snort, Bro, Bind, TippingPoint, Elsa, PassiveDNS, FireEye	Private instance, or shared with a trusted community among different CIF instances via a centralized service.	STIX, CybOX[f], Feeds from a CIF instance can be added as a data source to another CIF instance	(1) Finding related threats e.g. different domains/URLs that point to IP addresses in the same autonomous system (2) Whitelist observations from entering a feed during the feed generation process (3) Setup filters for what kind of data to pull from the instance	Kibana to generate statistics, trends and maps	Open source (GNU General Public License)

[Tounsi, Rais, 2018]

The maturity level

	MISP	OpenCTI	CIF	CRITs
		Holistc Architecture		
Use case applicability	++++	++++	+++	+++
Adherence 5W3H method	++++	++++	+	++
		Intelligence Process		
Import formats	OpenIOC, STIX, CybOX, JSON, CSV, XML	STIX, CybOX, JSON, CSV, XML	XML, JSON, Zip	CSV, STIX, CybOX
Automatic gathering	Using MISP feeds	Using connectors with sources or other platforms	Automatic synchronization with different sources	Possible integration with gathering tools
Export format	MISP, OpenIOC, CSV, XML, JSON	CSV, STIX	CSV, JSON, HTML, XLS	CSV, STIX, CybOX
Graphic visualization	General and intuitive dashboard and relationship graphics	Diverse dashboards and STIXv2 based graphics	Command line interface with possible integration with visualization tool	Simple dashboard and an extension service for generating relationship graphics
Correlation	Automatic for every data in platform	Automatic for every data in platform	Not addressed	Necessary an extension service
Classification	Based on the type of the indicator	Based on STIXv2 objects	Based on the type of the indicator	Based on a proposed data model
Integration	IDS, SIEMs and other TI platforms	Other TI platforms	IDSs (Snort, Splunk, Bro, Bind)	Not addressed
Sharing method	Reliable group of instances using different models	Particular instance to share between users	Reliable group of instances using a centralized service	Reliable group of instances
		Additional		
Documentation	Extensive and well elaborated	Extensive and well elaborated	Limited detail with succinct descriptions	Satisfactory quantity and detailing
License model	Open Source (GNU General Public License)	Open Source (Apache License)	Open Source (GNU General Public License)	Open Source (GNU General Public License)

Legend: very high (++++) high (+++) medium (++) low (+).

[de Melo et al., 2020]

Some observations

- No common definition of threat intelligence sharing platforms
  - Sharing and aggregating data vs. intelligence
- STIX is the de facto standard
- Focus primarily on sharing IoC
- Data collection instead of analysis
  - Limited analysis and visualization capabilities
    - browsing, attribute based filtering and searching of information
- Trust issues are mostly neglected
- Too many manual tasks, lack of automation

An Example: MISP

By a group of developers from CIRCL, the Belgian Defense and NATO / NCIRC (Computer Incident Response Capability)

- https://www.misp-project.org
- https://github.com/misp/
- https://www.circl.lu

Co-financed by the European Union
Connecting Europe Facility

MISP: Open Source Threat Intelligence Platform

- MISP (Malware Information Sharing Platform) is an IoC and threat indicators sharing free software
- MISP has many functionalities e.g. flexible sharing groups, automatic correlation, free-text import helper, event distribution and collaboration
- Many export formats which support IDSes / IPSes, SIEMs, Host scanners, analysis tools, DNS policies

MISP: Main features

- MISP sharing is a distributed model where technical and non-technical information can be shared within closed, semi-private or open communities

- With the focus on automation and standards, MISP provides:
  - A powerful ReST API
  - Extensibility (via misp-modules)
  - Additional libraries such as PyMISP

MISP: Interfaces

Web interface
- Multiple users and groups
- Role based access

API access for automation
- Integration with other tools
- Synchronization with security controls
- Python library

PyMISP

MISP: Basic Concepts

- All the malware data entered into MISP are made up of event objects
- Events are containers of **contextually** linked information
  - From an incident, a security report or a threat actor analysis
- Contains attributes with **indicators**
- Indicators contain a pattern that can be used to detect suspicious or malicious cyber activity
  - IoCs are a subset of indicators

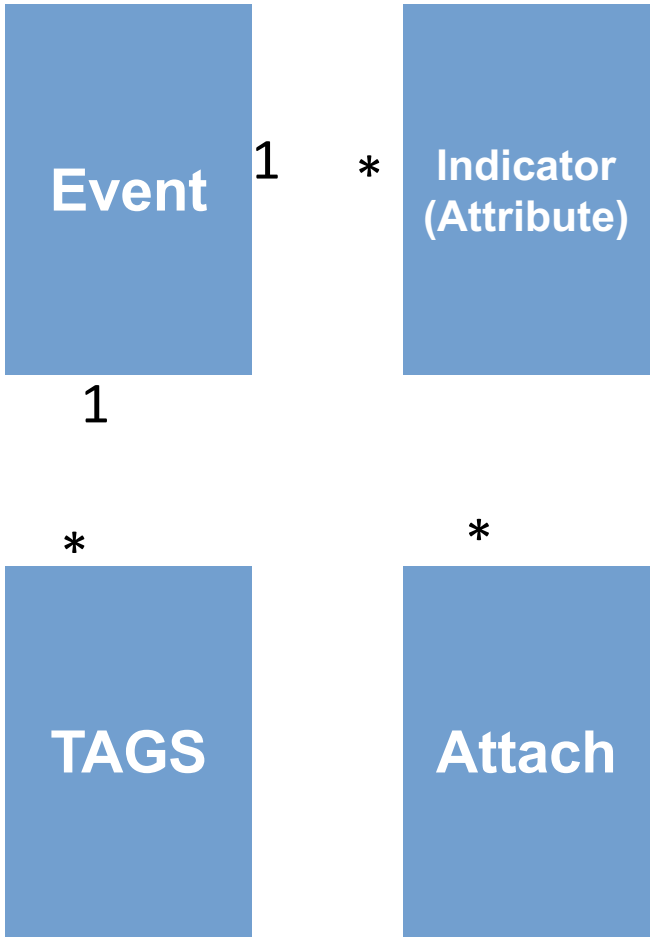
MISP: Basic Concepts: Proposals

- Each event can only be directly edited by users of the original creator organization
- However, if another organization would like to amend an event with extra information on an event, or if they'd like to correct a mistake in an attribute, they can create a Proposal
- Proposals can be accepted by the original creator
- Proposals can be pulled to another server, allowing users on connected instances to propose changes that, if accepted, can be subsequently pushed back

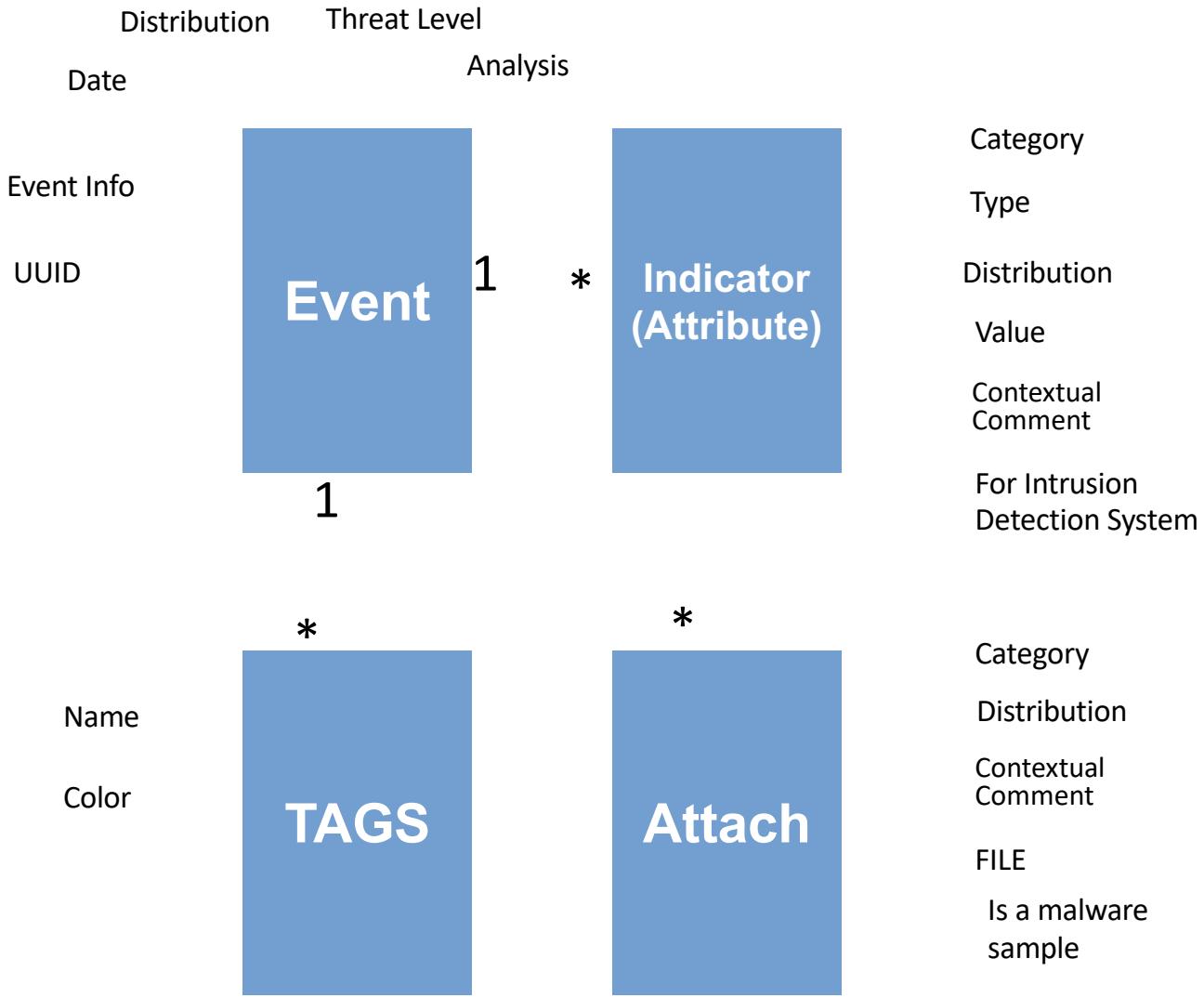
MISP: Basic Concepts: Delegation

- The privacy of the reporting organization can be established
  - to avoid the relation of an organization with the information shared
- MISP has a functionality to delegate the publication and completely remove the binding between the information shared and its organization
  - If you want to publish an event without you or your organization being tied to it, you can delegate the publication to an other organization
  - The other organization can take over the ownership of an event and provide pseudo-anonymity for the initial organization

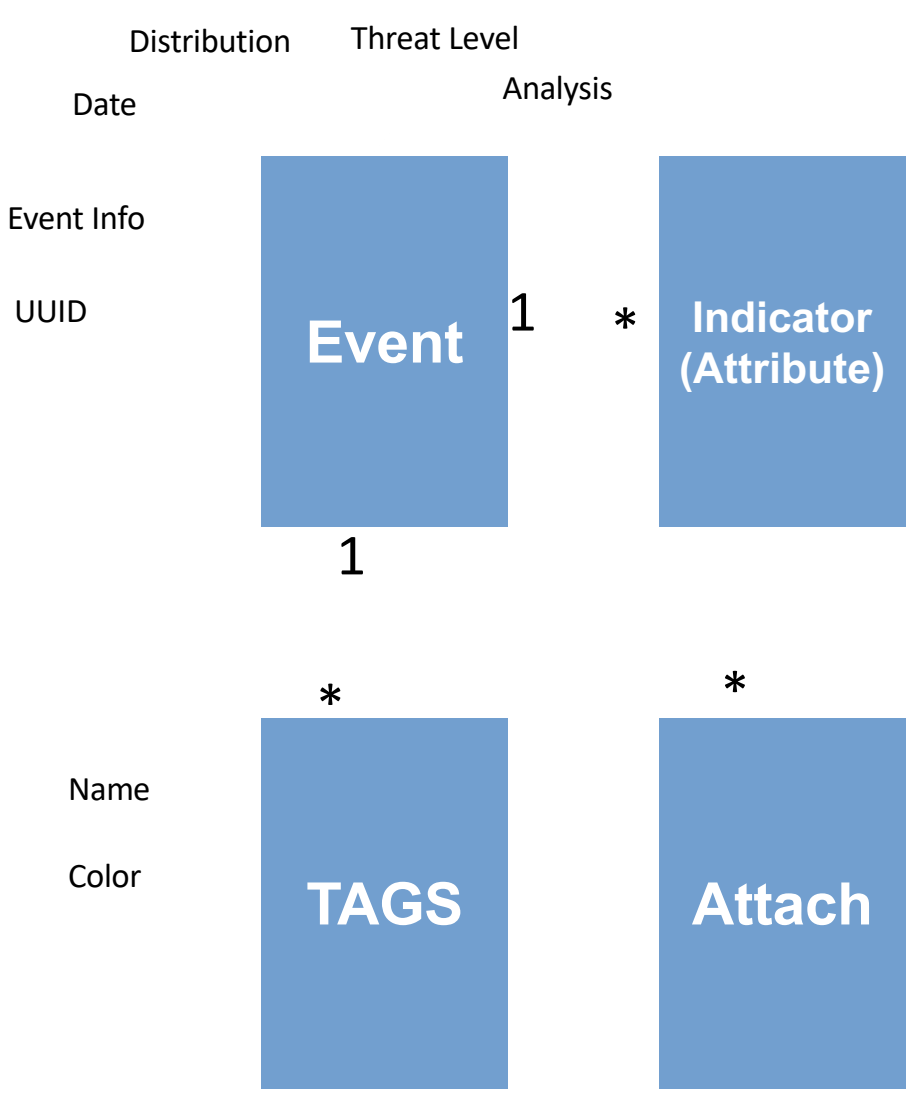
MISP DB Format (complete)



MISP DB Format (complete)



MISP DB Format (complete)



Network Activity

Payload Delivery

Category — Antivirus Detection

...

md5
hostname
domain
...
mac-address
regkey|value

Type

Distribution

Value

Contextual Comment

For Intrusion Detection System

Your Organization Only

This Community Only

Connected Communities

All Communities

Antivirus Detection
Payload Installation
...

Category

Distribution

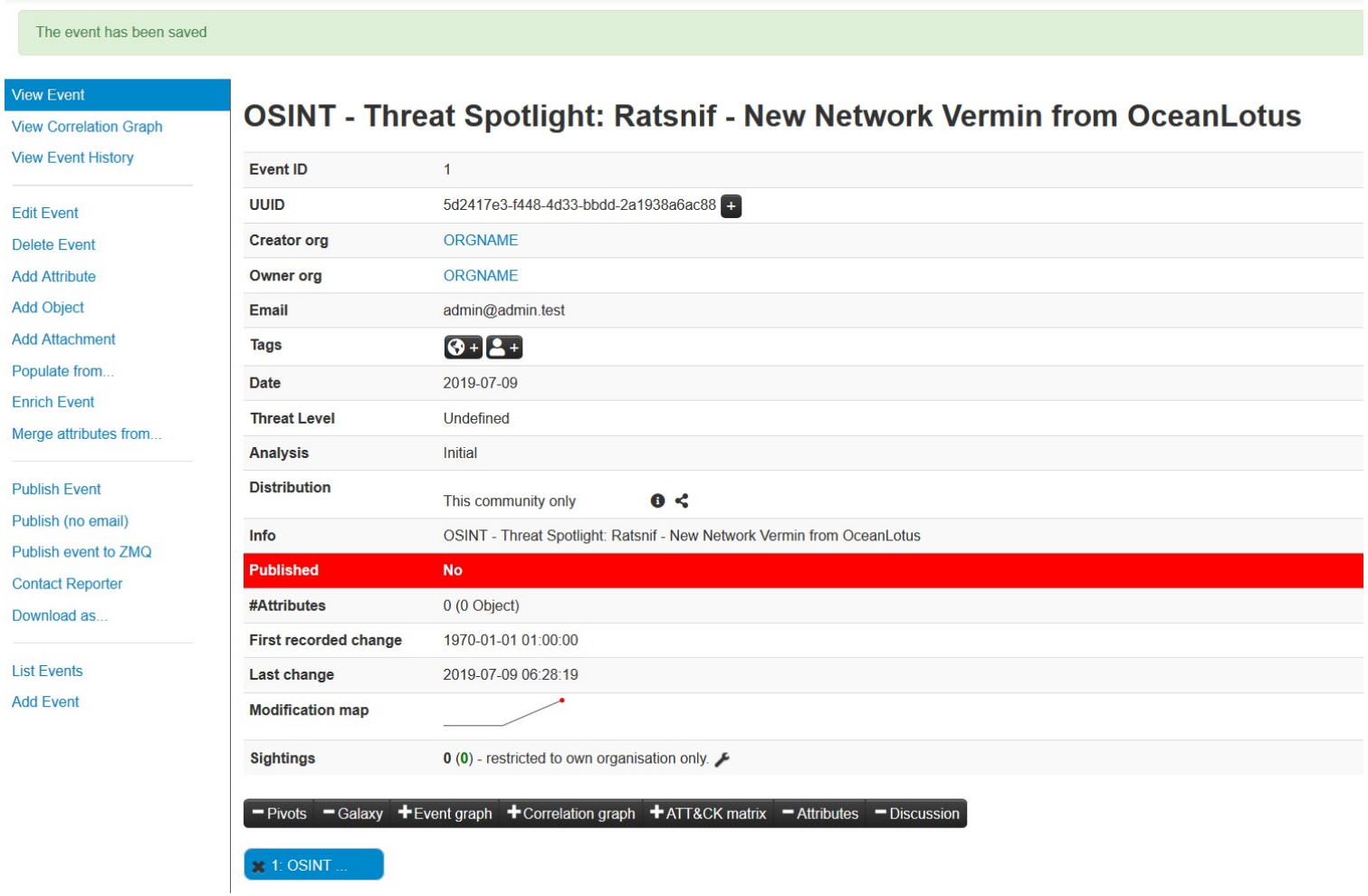
Contextual Comment

FILE

Is a malware sample

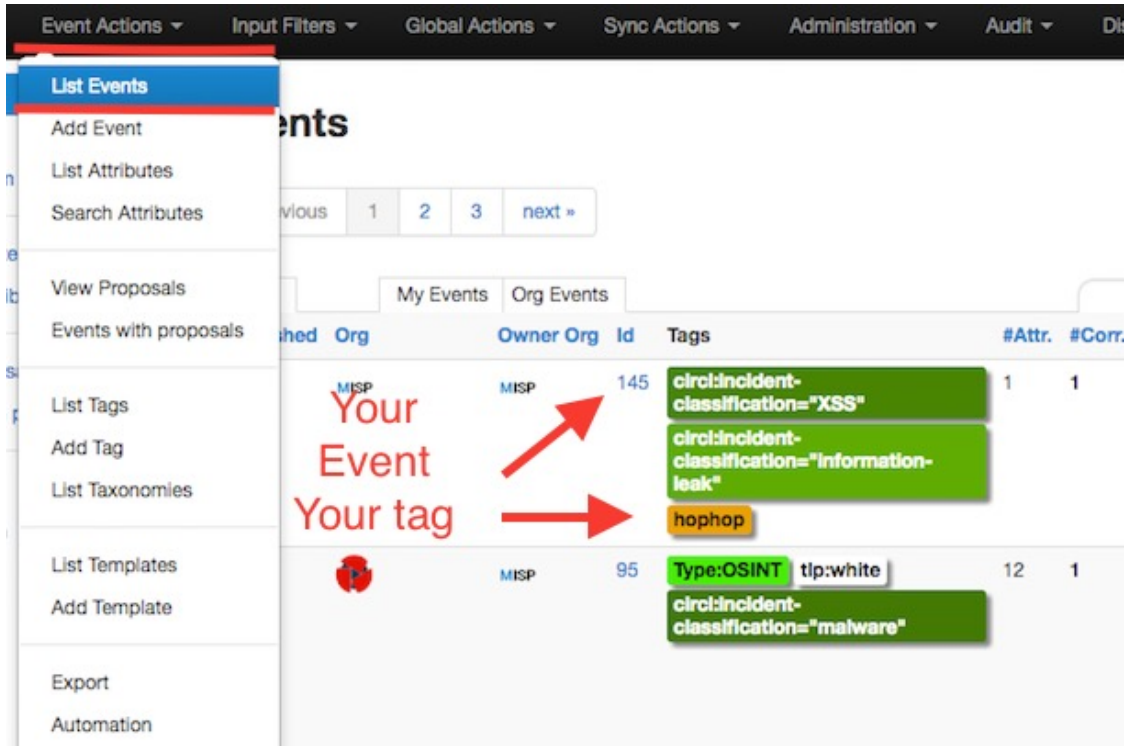
Network Activity

MISP: Event Example



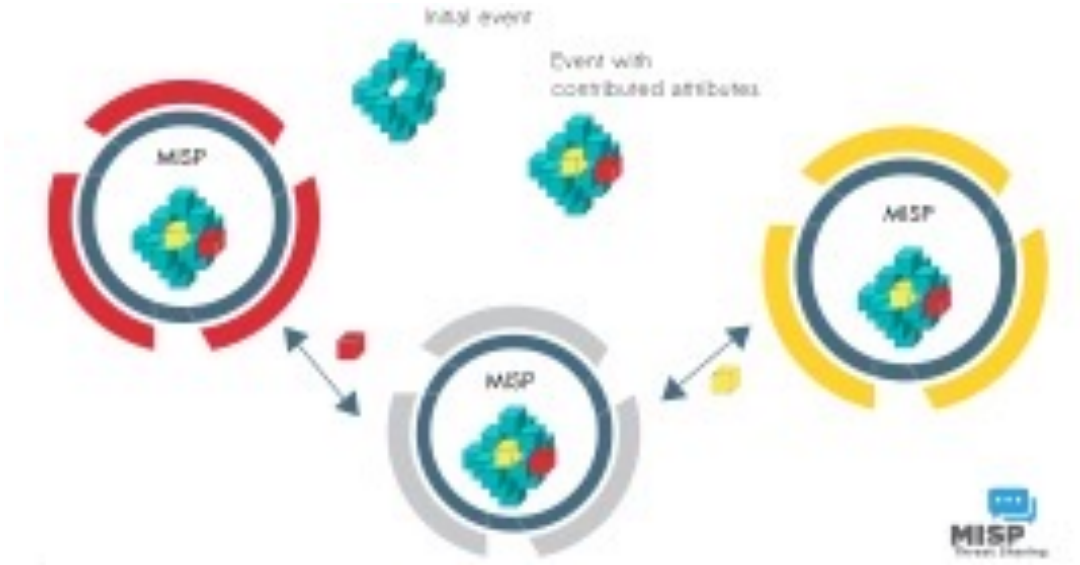
MISP: Event Browsing and Export

Export functionality is designed to automatically generate signatures for intrusion detection systems

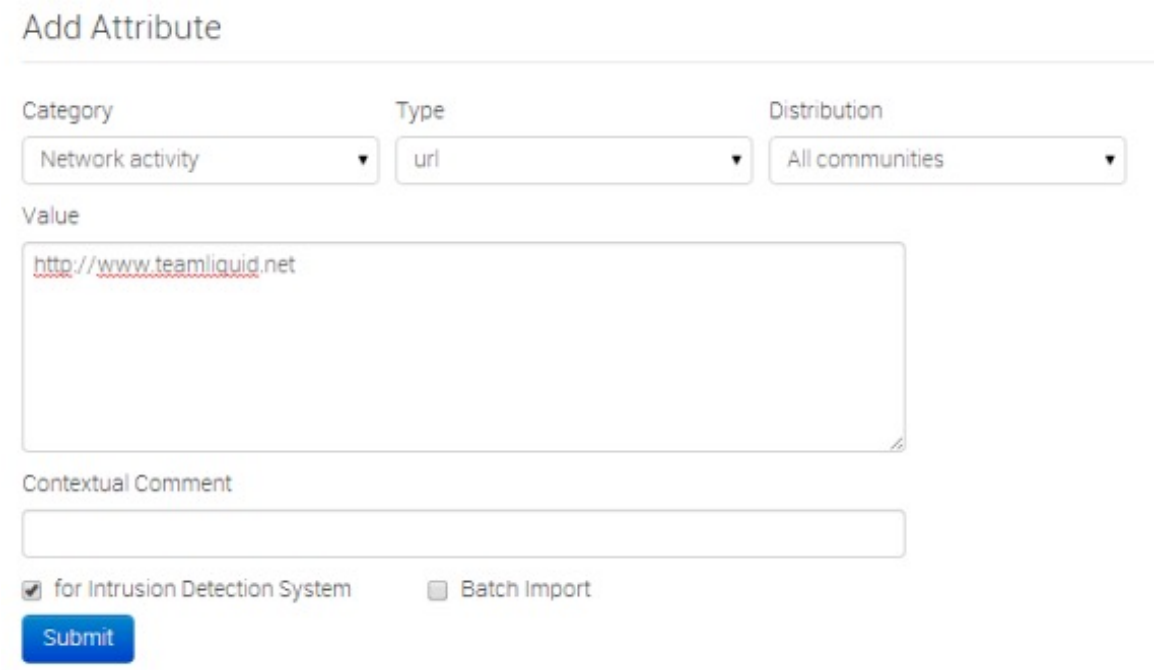


MISP: Remote Sync

- Two ways to get events from remote sources:
  - From another MISP server (also called MISP instance), by synchronizing two MISP servers
  - From a link, by using Feeds



MISP Attributes



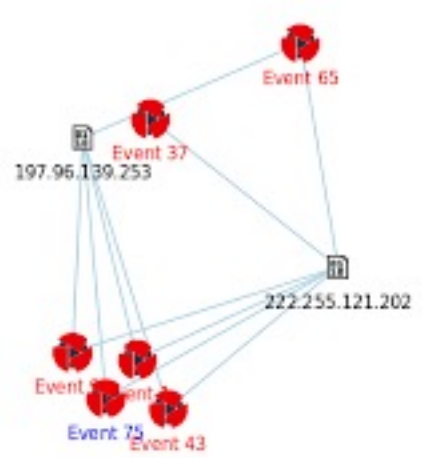
- For Intrusion Detection System: This option allows the attribute to be used as an IDS signature when exporting the NIDS data, unless it is being overruled by the white-list.
- If the IDS flag is not set, the attribute is considered as contextual information and not to be used for automatic detection.

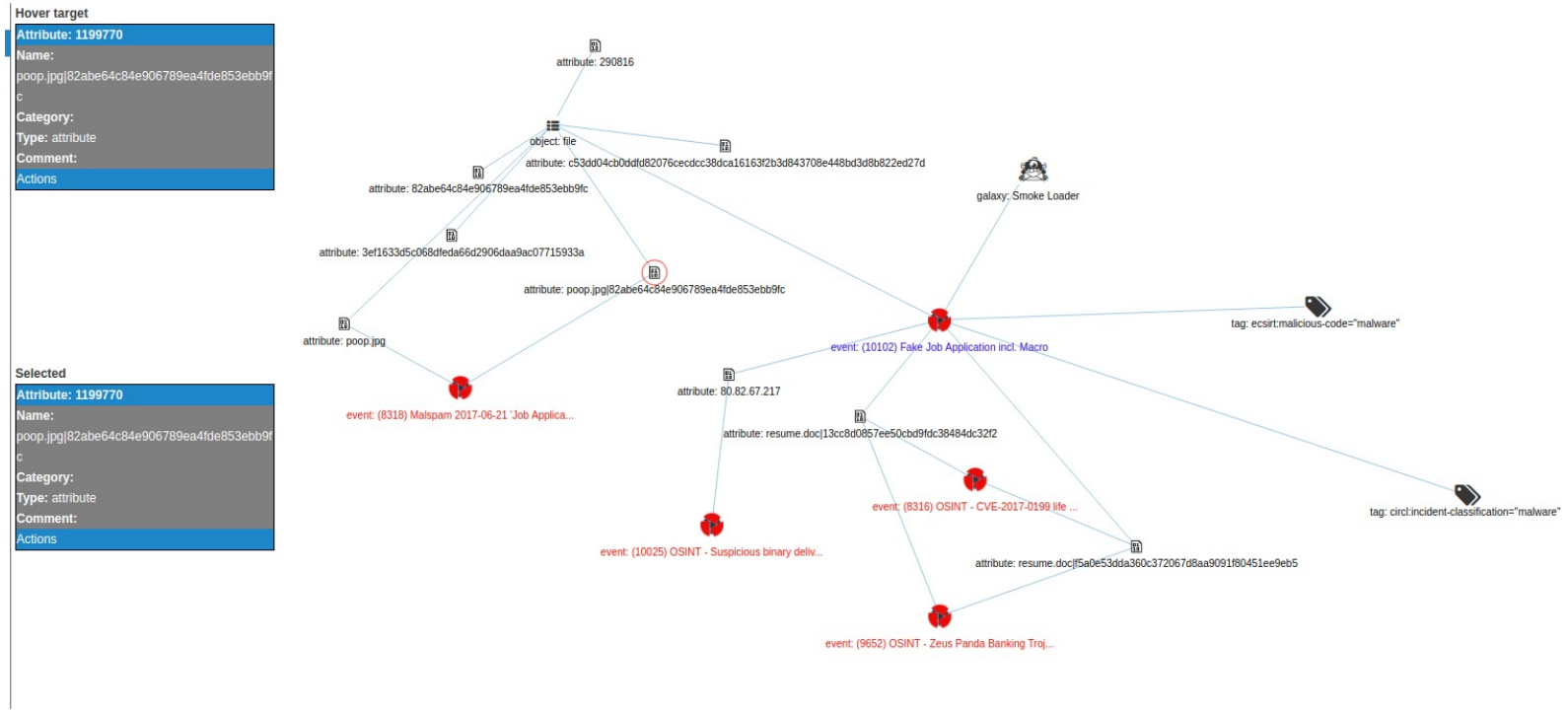
MISP: Event Indicator Examples

- Recommended IoCs for each Event (when possible)
  - ip-src: source IP of attacker
  - email-src: email used to send malware
  - md5/sha1/sha256: checksum
  - Hostname: full host/dnsname of attacker
  - Domain: domain name used in malware

Correlating data

- Correlate on indicators and context



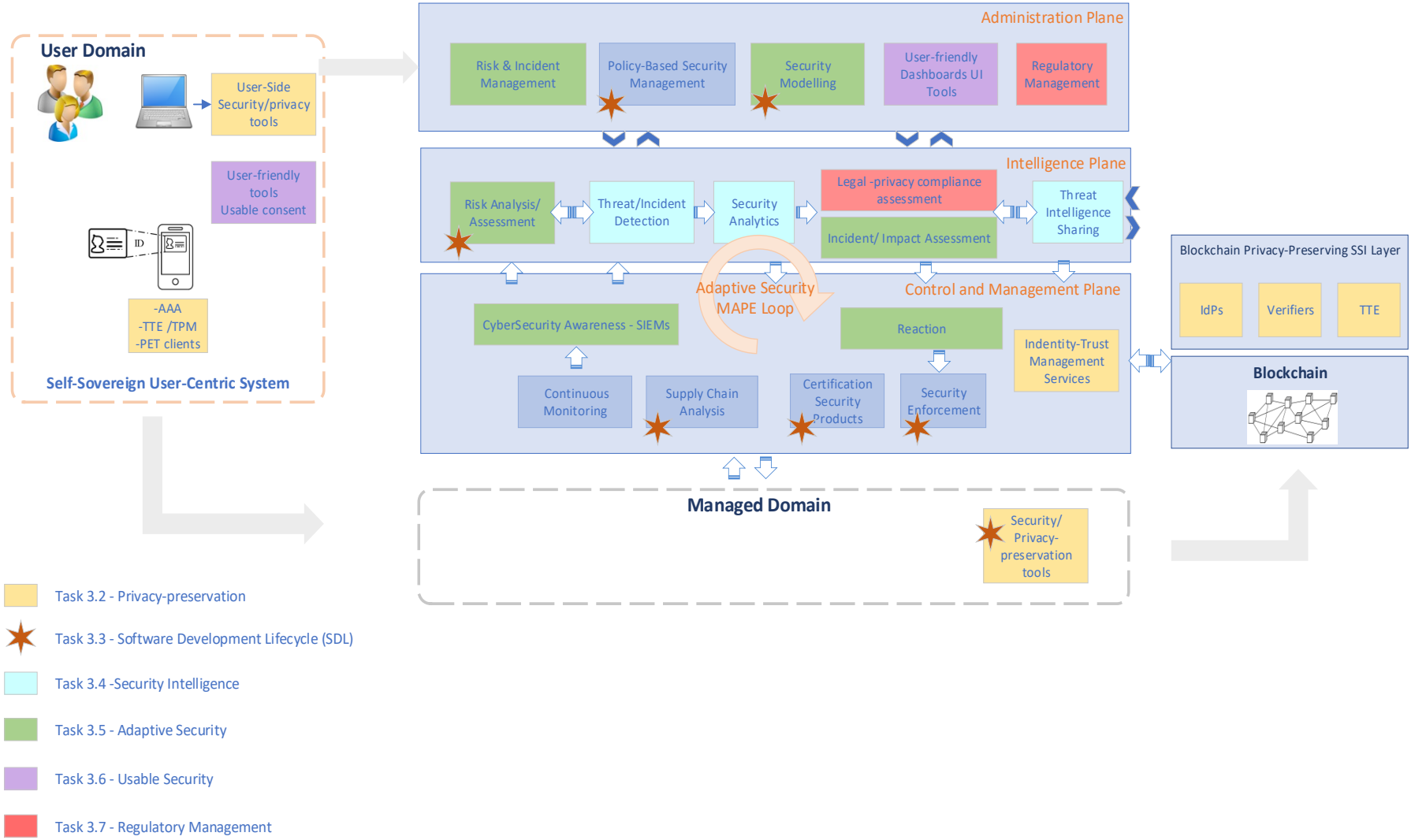


The CS4E Experience

Context: CyberSec4Europe

- A research-based consortium with 43 participants from 22 EU Member States
- The project addresses key EU Directives and Regulations, such as the GDPR, PSD2, eIDAS, and ePrivacy, and tries to implement the EU Cybersecurity Act including the development of the European skills base, the certification framework and ENISA role
- EU H2020-SU-ICT-03-2018

WP3
Global Architecture and Tasks Block



Task 3.4 Security Intelligence

“We will enhance the state of the art for reliability, safety and privacy guarantees of security intelligence techniques based on artificial intelligence, machine learning and data analytics.”

Objectives and scope

- Define requirements and mechanisms to **share digital evidence** between expert systems
- **Interoperability** through unification of language, format, interface, or trusted intermediaries with respect for privacy, business requirements and national regulations
- Interact with **Threat Intelligence Information Services** for early malware activity detection
- Log/event management, threat detection and security analytics with **privacy-respecting** big data analytics
- Fortify underpinning **security intelligence** in defensive systems

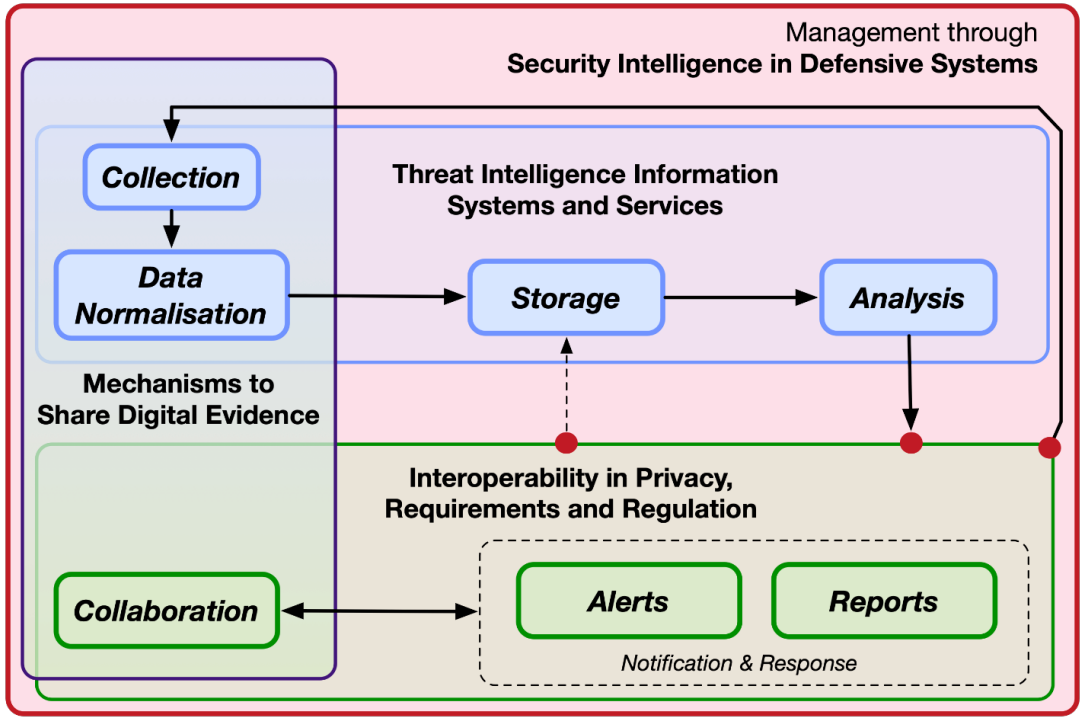
Starting observations

- Fast sharing of TI is not sufficient to avoid targeted attacks
- Choosing the best threat intelligence tool depends on the organization objectives
  - standardization and automatic analytics needs versus high speed requirements

A high level overview

- A collaborative security intelligence platform that aims to manage digital evidence
- The platform covers the whole life cycle of security related information

1. Raw data ingestion
2. Sharing data among trusted stakeholders
3. Covering all the levels of collaboration (technical and regulation)
4. Robustness with respect to the introduction of new components



Mechanisms to share digital evidence

- **Goal:** enabling the collaboration among organizations for defining defensive actions against complex attack vectors
  - **How**: Sharing information and knowledge about threats, sightings, indicators of compromise (IoC) and mitigation strategies

- **Challenges**:
  - Issues with IoC
    - Network indicators: “the faster you share, the more you theoretically will stop”
      - cumulative uniqueness, time of spread, time of validity
    - Malware indicators
      - Obfuscation techniques
      - Indicators such as created registry keys or file artifacts are less commonly changed by attackers but they can be given random or pseudorandom component in their names

  - the sharing of IoC (typically event-based) is incompatible with data-driven machine learning approaches incorporated in advanced monitoring and detection products

Threat intelligence information systems and services

- **Goal**: preventing the same incident from happening elsewhere
  - **How**: The usage of enabling technologies for managing digital evidence, i.e. tools to collect, examine, analyze and share digital evidence from heterogenous data sources

- **Challenges**:
  - Traditional solutions (e.g., SIEM and SOAR solutions) may lack the necessary capabilities to quickly adapt to new and/or evolving threats. They should integrate intelligent components to automatize the process.
  - Quality over quantity
    - The daily dump of indicators seen as suspicious in Internet, provides information approximating 250 to millions of indicators per day
    - A common standardized format for sharing TI minimizes the risk of losing the quality of threat data
      - Provides better automated analytics solutions on large volumes of TTI
    - customization, filtering, aggregation, search

Reducing the quantity of threat feeds

- Identifying the mutations of malware variants is essential in order to recognize those belonging to the same family
- Data science and machine-learning models are looking to deliver entirely new ways of searching malwares.
  - Analyzing a huge amount of threats, to learn shared patterns
  - Malware analysis, detection, classification, and clustering can help this automation

Examples: Malheur

- collects behavioral analysis data inside sandbox
  - malware binaries are collected in the wild and executed
  - The execution of each malware binary results in a report of recorded behavior
    - Extraction of prototypes from reports
    - Automatic identification of groups (clusters) of reports containing similar behavior
    - Classification of behavior based on a set of previously clustered reports
    - Incremental analysis, by processing reports in chunks

Interoperability in privacy, requirements and regulation

- **Goal**: Sharing trusted, reliable and privacy-preserving information
  - **How**: Enforcing appropriate security and privacy policies to enforce sharing requirements of threat intelligence and alerts

- **Challenges**:
  - ensuring that information collected within TIPs is reliable and accurate
    - **Example**: TIPs allow to export a subset of the data into Intrusion Detection System (IDS) rules that can be inserted in solutions like Snort or Suricata. Malicious or unreliable input may compromise such HIDS and NIDS
  - Enhance the privacy and trust capabilities to overcome concerns

- **Further requirements**: The procedures for handling sensitive data should be compliant with relevant regulations and directives e.g., the EU General Data Protection Regulation (GDPR) or the Payment Service Directive 2 (PSD2)

Security intelligence in defensive systems

- **Goal**: Preventing data exfiltration from TIP
  - Gathered threat data can be exploited for both, preventing or performing effective attacks

- **Requirement 1**: the security intelligence platform must implement appropriate measures to ensure that the platform itself does not increase the overall attack surface of the cybersecurity infrastructure

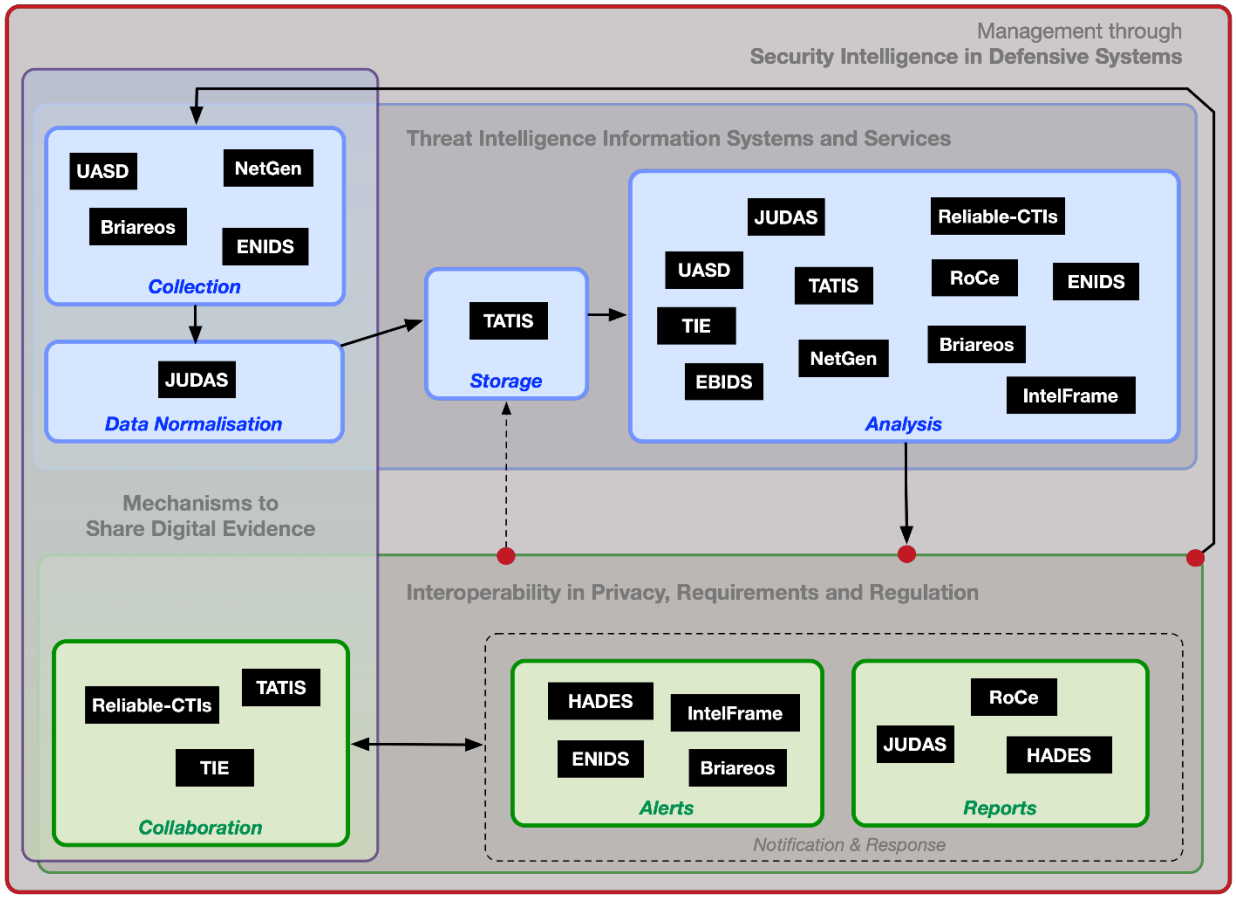
- **Requirement 2**: the security intelligence platform must be robust against adversarial attacks aiming at feeding the system with false information

Challenges – A summary

- Reducing the amount of false positive threat or attack alerts
- Lowering the time to threat detection amidst the growing amounts of data to analyze
- Contextualizing threat data to support analysis of disparate information sources
- Boosting trust among organization belonging to the sharing networks
- Defining flexible strategies, methodologies and data formats for collaborative TI
- Enhancing cyberthreat analysis and digital investigation techniques when privacy techniques are used
- Improving the notification mechanisms and automatization by introducing intelligent components
- Minizing the attack surface by strengthening the robustness of ML and DL models adopted by security applications

Assets and contributions

- CS4E has integrated several assets and mapped them within the overall scheme

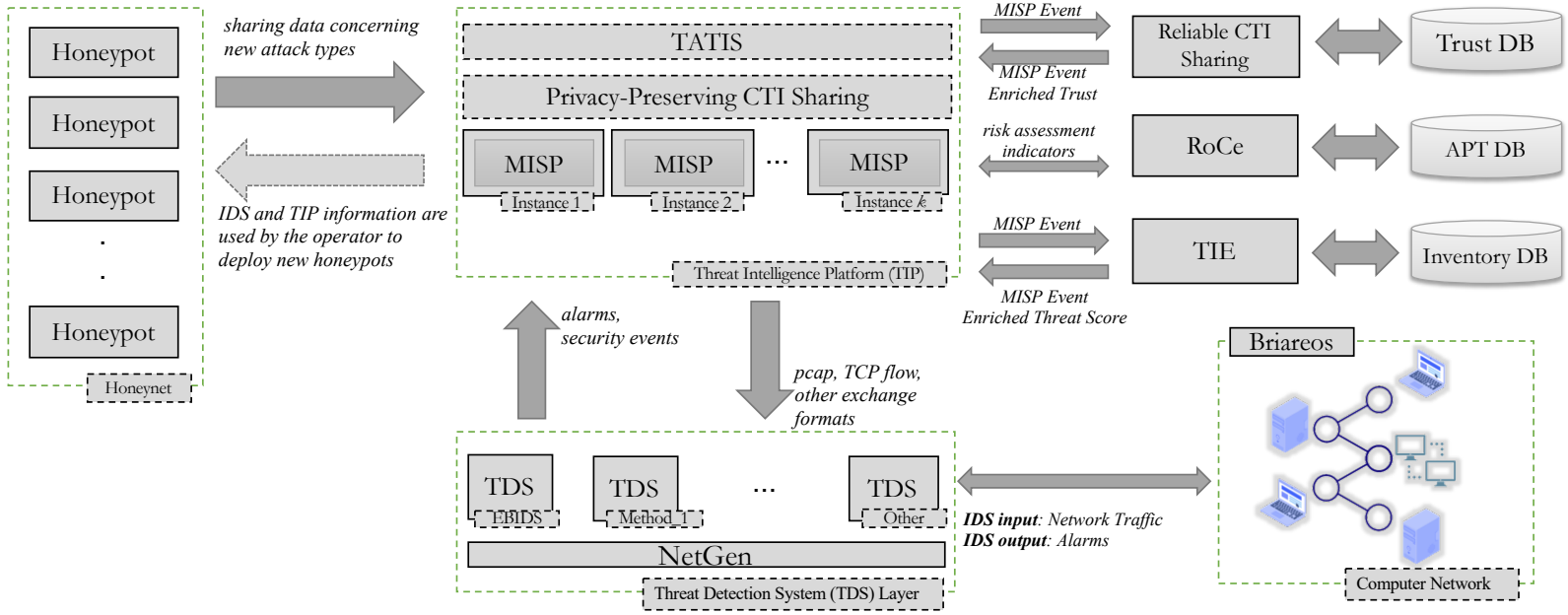


TIE: *Threat Intelligence intEgrator* **(ATOS)**
Briareos (C3P)
UASD: *Unautorized App Storage Discovery* **(CNR)**
EBIDS: *Ensemble Based Intrusion Detection System* **(CNR)**
IntelFrame: *Intelligent Machine Learning-based Intrusion Detection* **(DTU)**
TATIS: *Trustworthy APIs for Enhanced Threat Intelligence Sharing* **(KUL)**

NetGen (POLITO)
JUDAS: *JSON Users and Device Analysis Tool* **(UMA)**
HADES: *Automatic analysis of malware samples* **(UMA)**
Reliable-CTIs - *Reliable Cyber-Threat intelligent sharing* **(UMU)**
ENIDS: *Edge Network Intrusion Detection System* **(UNITN/FBK)**
RoCe: *Risk of Compromise estimation* **(UNITN)**

A Demonstration Platform

- Integrates different type of security services
  - E.g., risk indicators, enriched IoC, privacy-preserving utilities, etc.
- Aims at enriching TIP (MISP) events
- Three main scenarios
  - Sharing cyberthreat intelligence in a **confidential and privacy-preserving** manner
  - **Enriching the information** on detected threats via TDS **cooperation** and gathered by means of honeypot instances
  - **Adaptive deployment**
- https://github.com/cs4ewp3t4



Cooperation with Threat Intelligence Services

A case study

Focus

- **Scenario**: Timely sharing threat events and *indicators of compromise* (IoCs) among organizations is crucial in order to make quick decisions and set up effective countermeasures

- **Goal**: Designing a solution meant for gathering and managing threat information from different data sources

- **Main objectives**:
  - Improving the accuracy of Threat Detection Systems in detecting incoming attacks
  - Enabling the sharing of trusted, reliable and relevant threat information among organizations

Our proposal

- Defining a distributed platform enabling the sharing of reliable and privatized data

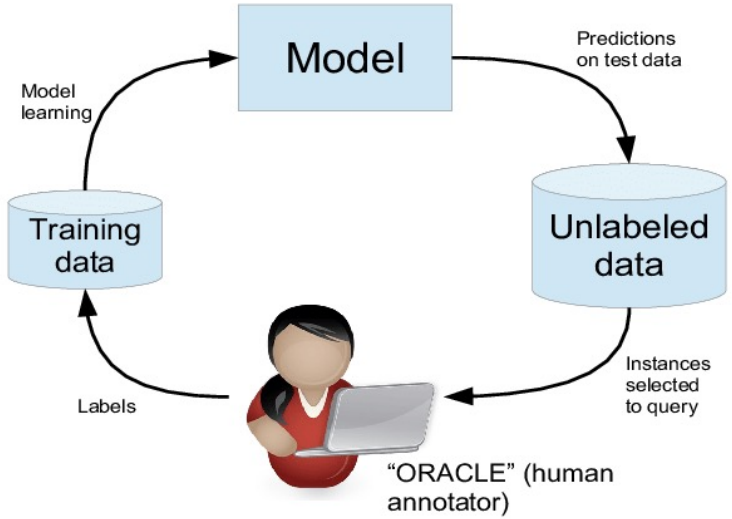
- Main capabilities
  - Threat Detection Systems cooperation
  - Human in the loop (*Active Learning*)
  - Data enrichment from different sources
    - E.g., *TDS, honeypots, etc*

Active Learning

- Active Learning (AL) refers a family of approaches and algorithms wherein new instances to be labelled are interactively chosen by means of a query
  - **Idea**: providing unknown examples (extracted with different strategies) to an *oracle* that will correctly label them

- **Usage Scenario**: AL can is used when data are hard to label or highly skewed and allows for making sense of data faster and more efficiently
  - E.g., *intrusion detection, fraud detection, fault detection, etc.*

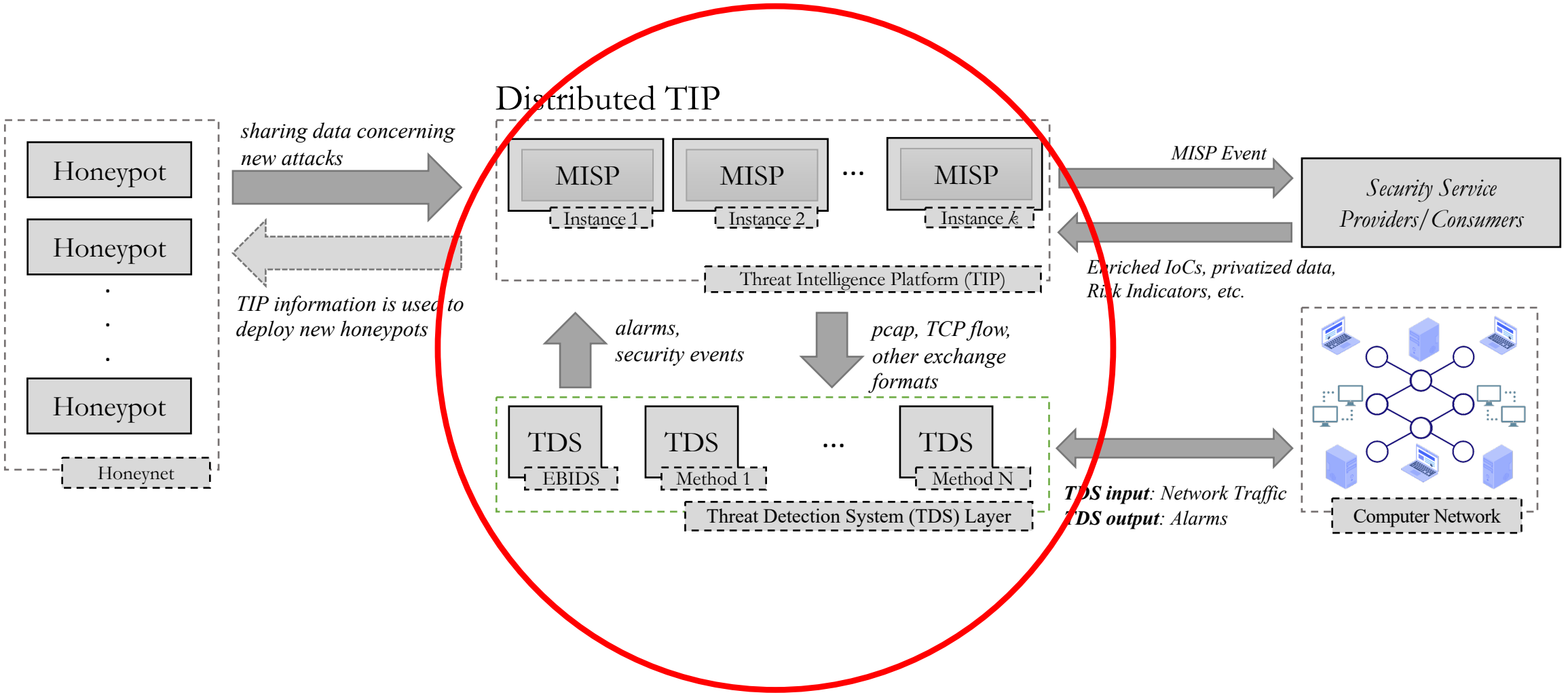
- **Strategies**:
  - *Uncertainty Sampling, Query-by-Committee, Expected Model changes, etc.*



Platform overview

- There are essentially three actors
  - **Distributed TIP** (*Threat Intelligence Platform*)
    - Core component
    - Two-folds role
      - Storing data coming from heterogeneous sources in an encrypted and distributed way
      - Delivering the gathered information to the other components
  - **TDS Layer**
    - Different types of Threat Detection Systems (e.g., *IDS, IPS, etc*) can interface with the TIP
      - TDSs provide information concerning incoming attacks
      - TDSs feed the TIP with new intrusion events/statistic
  - **Honeynet**
    - Honeypots are deployed with the aim to collect additional information concerning new attacks

Platform: main actors



TIP Details

- A network of MISP instances
- Motivation
  - Open source
  - Strong underlying community
  - Extensible (MISP Objects)
  - Good documentation
  - Support to different standards

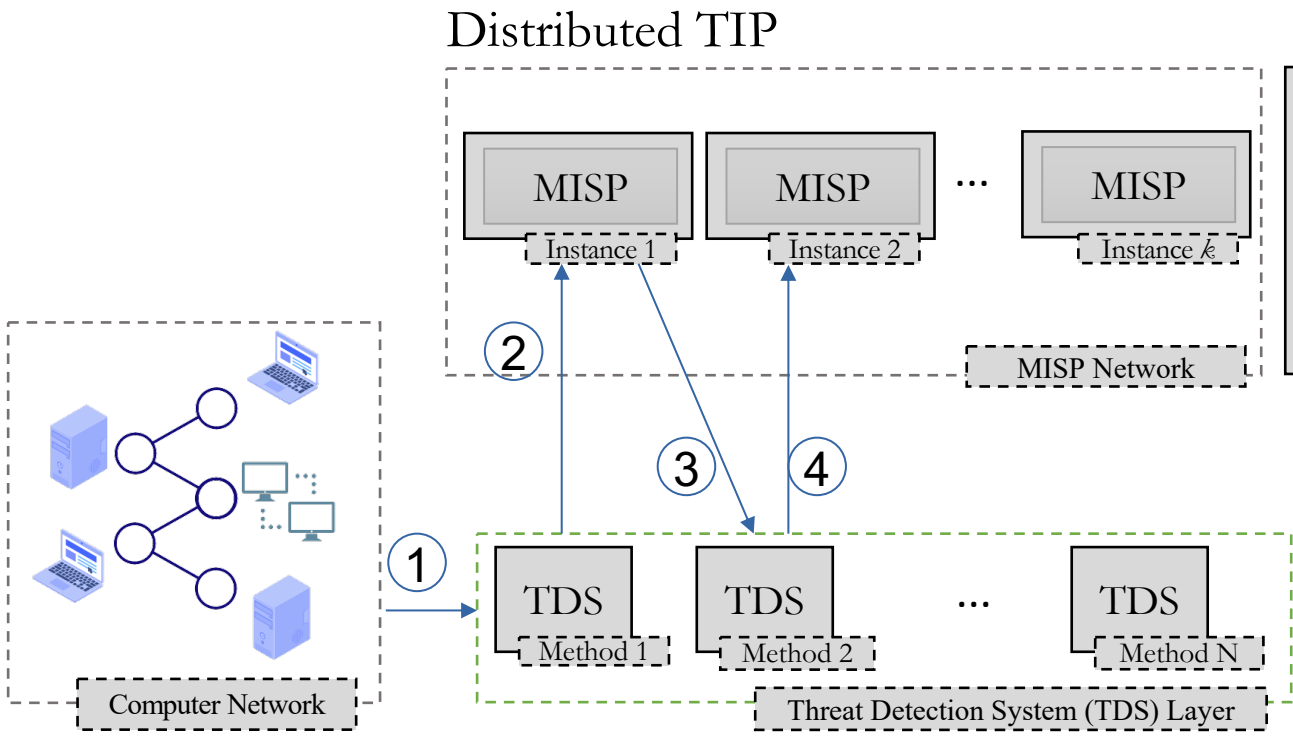
Data exchange format

- The assets interface among them by using a custom MISP Object in JSON format
  - The MISP object represents the data structure adopted by MISP to store shared threat events
  - The general template can be extended so as to include further relevant information on specific threat events

Object attribute	MISP attribute type	Description	Disable correlation	Multiple
creation-date	datetime	Threat Event Date	✔	–
ip_dst	ip-dst	Destination IP	–	✔
ip_dst_port	port	Destination Port	–	✔
ip_src	ip-src	Source IP	–	✔
pcap_file	attachment	PCAP file	✔	–
verified	boolean	It specifies if an operator verified the occurrence of the attack	✔	–
signature_type	text	Type of signature (md5, sha1,...)	✔	–
signature	text	Optional detected file signature	–	✔
attack_type	text	A JSON containing information on IDS classification	–	–
anomaly_details	attachment	Optional JSON file containing anomaly flow statistics	–	–
privatized	attachment	Privatized version of the attribute	✔	✔


```
{"Object": [{"id": "18919",
   "name": "security_event_object",
   "description": "CS4E Security Event Object",
   "uuid": "ba14028e-03a6-47d2-b35f-8147381493cf",
   "timestamp": "1615454674",
   "Attribute": [{"id": "262154",
     "type": "ip-src",
     "category": "Network activity",
     "object_relation": "ip_src",
     "value": "              "
     },
    {"id": "262155",
     "type": "ip-dst",
     "category": "Network activity",
     "object_relation": "ip_dst",
     "value": "              "
     },
    {"id": "262156",
     "type": "port",
     "category": "Network activity",
     "object_relation": "ip_dst_port",
     "value": "    "
     },
    {"id": "262157",
     "type": "datetime",
     "category": "Other",
     "object_relation": "creation-date",
     "value": "2021-03-11T10:24:34.194148+0000"
     },
    {"id": "262158",
     "type": "text",
     "category": "Other",
     "object_relation": "attack_type",
     "value": {"EBIDS":
       {"version": "0.2",
        "reference": "https://github.com/ebids/",
        "attacks": [{ "attack_type": "ANOMALY", "confidence": "0.98153543"}]}}
      },
    {"id": "262159",
     "type": "attachment",
     "category": "External analysis",
     "object_relation": "pcap_file",
     "value": "anomaly.pcap"
     },
    {"id": "262160",
     "type": "attachment",
     "category": "External analysis",
     "object_relation": "anomaly_details",
     "value": "anomaly.json"
     },
    {"id": "262161",
     "type": "boolean",
     "category": "Other",
     "object_relation": "verified",
     "value": "0"
     }]}]}
```

Platform in action: TDS Cooperation



1. Network flow (pcap) is sent to TDS 1
2. TDS 1 detects an anomaly and shares it with a MISP instance by sending a security event object (SEO)
3. TDS 2 gathers information from MISP to update its classifier
4. TDS 2 classifies the new threat and updates the SEO on MISP
5. An expert (either user or automated) checks the new threat via MISP Web Interface
6. The expert validates the threat event

Benefits

- The amount of false positive reduced
  - The sharing protocol allows different actors (either AI or humans) to validate threat evidence and mutually benefit from feedbacks provided by other peers
- time to threat detection lowered
  - Collaboration among automated predictive models allows for reducing the average time to detect an intrusion
- Threat information better contextualized with additional IoCs coming from other assets
- Privacy enhancement via cooperation with other assets in a seamless integration

Concluding remarks

- Security intelligence platforms and sharing mechanisms can substantially improve the security capabilities of cybersecurity applications in various vertical domains and use cases
- Current Threat Intelligence platforms can take advantage from the adoption of AI/ML tools
  - Knowledge extraction from different sources
  - Improving the quality of data via AI powered tools
- The need for strengthenining the collaborative mechanisms to include
  - data-driven and AI powered threat detection systems
  - Sophisticated refinements of IoCs
  - privacy enabling techniques and methods to guarantee trust and confidence

Concluding remarks

- The CS4E contribution
  - A research roadmap
  - Vertical demonstrations with measurable benefits
    - false positive alerts reduction
    - contextualizing threat data
    - boosting trust among producers and consumers of threat data
    - strengthening the robustness of ML models

References

- V. Adewopo, B. Gonen and F. Adewopo, "Exploring Open Source Information for Cyber Threat Intelligence," *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 2232-2241,
- S. Barnum. Standardizing cyber threat intelligence information with the structured threat information expression (stix). Mitre Corporation 11 (2012), 1–22.
- E.W. Burger, M.D. Goodman, P . Kampanakis, K. A. Zhu. Taxonomy model for cyber threat intelligence information exchange technologies, in: Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security, ACM, pp. 51–60; 2014.
- D . Chismon, M . Ruks. Threat intelligence: Collecting, analysing, evaluating, MWR Infosecurity, UK Cert, United Kingdom; 2015.
- A. de Melo e Silva, J.Costa Gondim, R. de Oliveira Al- buquerque, and L. J. García Villalba. 2020. A methodology to evaluate standards and platforms within cyber threat intelligence. Future Internet 12, 6 (2020), 1–23
- P. -Y. Du *et al.*, "Identifying, Collecting, and Presenting Hacker Community Data: Forums, IRC, Carding Shops, and DNMs," *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2018, pp. 70-75
- ENISA. 2010. Incentives and Challenges for Information Sharing in the Context of Network and Information Security. https://www.enisa.europa.eu/publications/incentives-and-barriers-to-information-sharing
- ENISA. 2018. Exploring the opportunities and limitations of current Threat Intelligence Platforms. https://www.enisa.europa.eu/publications/exploring-the-opportunities-and-limitations-of-current-threat-intelligence-platforms
- ENISA. 2021. Threat Landscape. https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends
- V . Ghanaei, C.S. Iliopoulos, R.E. Overill. Statistical approach towards malware classification and detection, in: SAI Computing Conference (SAI), 2016, IEEE, pp. 1093–1099; 2016.
- M. Guarascio, E. Ritacco, D. Biondo, R. Mammoliti, A. Toma. Integrating a Framework for Discovering Alternative App Stores in a Mobile App Monitoring Platform. In: NFMCP 2017. LNCS, vol 10785.
- R. Holland, S. Balaouras, K. Mak. Five Steps To Build An Effective Threat Intelligence Capability, Forrester research, inc.; 2013.
- NIST 2016. Guide to Cyber Threat Information Sharing. NIST Special Publication 800-150. http://dx.doi.org/10.6028/NIST.SP.800-150
- O. Or-Meir, N. Nissim, Y. Elovici, and L. Rokach. 2019. Dynamic Malware Analysis in the Modern Era—A State of the Art Survey. ACM Comput. Surv. 52

References

- S. Piper Definitive guide to next generation threat protection, CyberEdge Group, LLC, 2013.
- A. Ramsdale S. Shiaeles, N. Kolokotronis, A Comparative Analysis of Cyber-Threat Intelligence Sources, Formats and Languages. Electronics. 2020; 9(5):824.
- S. Samtani, W. Li, V. Benjamin, and H. Chen. 2021. Informing Cyber Threat Intelligence through Dark Web Situational Awareness: The AZSecure Hacker Assets Portal. Digit. Threat.: Res. Pract. 2, 4, 2021
- S. Samtani, K. Chinn, C. Larson and H. Chen, "AZSecure Hacker Assets Portal: Cyber threat intelligence and malware analysis," 2016 IEEE Conference on Intelligence and Security Informatics (ISI), 2016, pp. 19-24
- W, Tounsi, H Rais, A survey on technical threat intelligence in the age of sophisticated cyber attacks, Computers & security, 2018 - Elsevier
- W. Tounsi, What is Cyber Threat Intelligence and How is it Evolving? In: Cyber-Vigilance and Digital Trust: Cyber Security in the Era of Cloud Computing and IoT, Wiley, 2019
- C. Sauerwein, I. Pekaric, M. Felderer, R. Breu, An analysis and classification of public information security data sources used in research and practice, Computers & Security, 82, 2019, Pages 140-155,
- C. Sauerwein, C. Sillaber, A. Mussmann, R. Breu, 2017. Threat intelligence sharing platforms: An exploratory study of software vendors andresearch perspectives. Wirtschaftsinformatik und Angewandte Informatik
- C. Sauerwein, D. Fischer, M. Rubsamen, G. Rosenberger, D. Stelzer, and R. Breu. 2021. From Threat Data to Actionable Intelligence: An Exploratory Analysis of the Intelligence Cycle Implementation in Cyber Threat Intelligence Sharing Platforms. In The 16th International Conference on Availability, Reliability and Security (ARES 2021).
- M. Sahin and S. Bahtiyar. A Survey on Malware Detection with Deep Learning. In 13th International Conference on Security of Information and Networks (SIN 2020).
- F . Skopik, G . Settanni, R. Fiedler. A problem shared is a problem halved: a survey on the dimensions of collective cyber defense through security information sharing. Comput Secur 2016;60:154–76.
- B. Stojkovski, G. Lenzini, V. Koenig, and S. Rivas. What's in a Cyber Threat Intelligence sharing platform? A mixed-methods user experience investigation of MISP. In Annual Computer Security Applications Conference (ACSAC 2021).
- Wagner et al. 2016. MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform. In Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security (WISCS '16).
- A. Zibak and A. Simpson. 2019. Cyber Threat Information Sharing: Perceived Benefits and Barriers. In Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES '19).

References

- A curated list of pointers on threat intelligence: https://github.com/hslatman/awesome-threat-intelligence
- Collection of Cyber Threat Intelligence sources from the Deep and Dark Web https://github.com/fastfire/deepdarkCTI
- Github topic: threat intelligence https://github.com/topics/threat-intelligence
- CS4E deliverables:
  - Deliverable D3.3: Research Challenges and Requirements to Manage Digital Evidence
    - https://cybersec4europe.eu/wp-content/uploads/2020/02/D3.3-Research-challenges-and-requirements-to-manage-digital-evidence-Submitted.pdf
  - Deliverable D3.14: Cooperation With Threat Intelligence Services For Deploying Adaptive Honeypots
    - https://cybersec4europe.eu/wp-content/uploads/2021/10/D3.14-Cooperation-with-Threat-Intelligence-Services-for-deploying-adaptive-honeypots_2.05_submitted.pdf