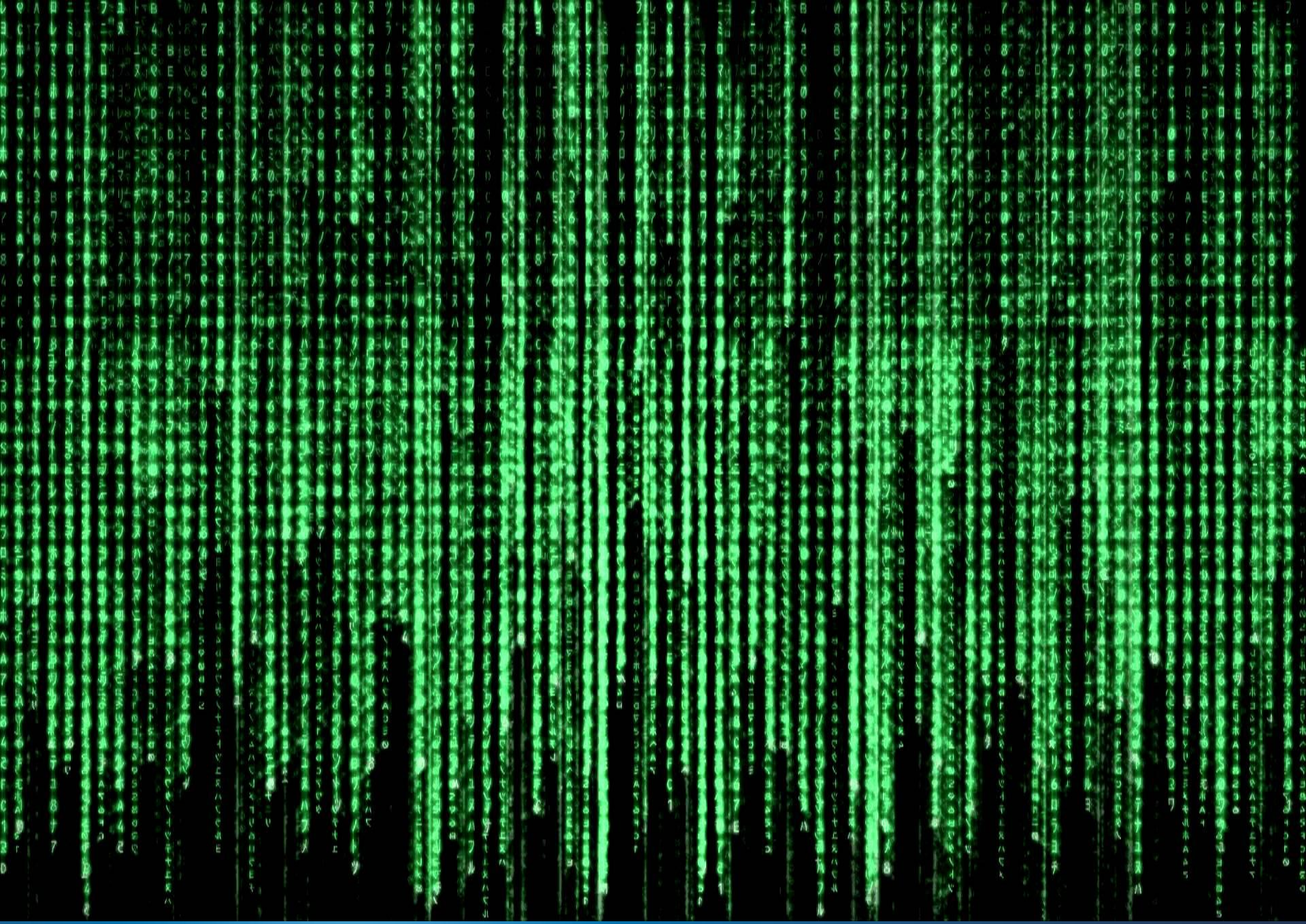
Actionable information for security incident response

Cosmin Ciobanu 2015



Agenda

- Background
- Definition & properties
- Processing pipeline
- Case studies and inventory
- Conclusions

Background

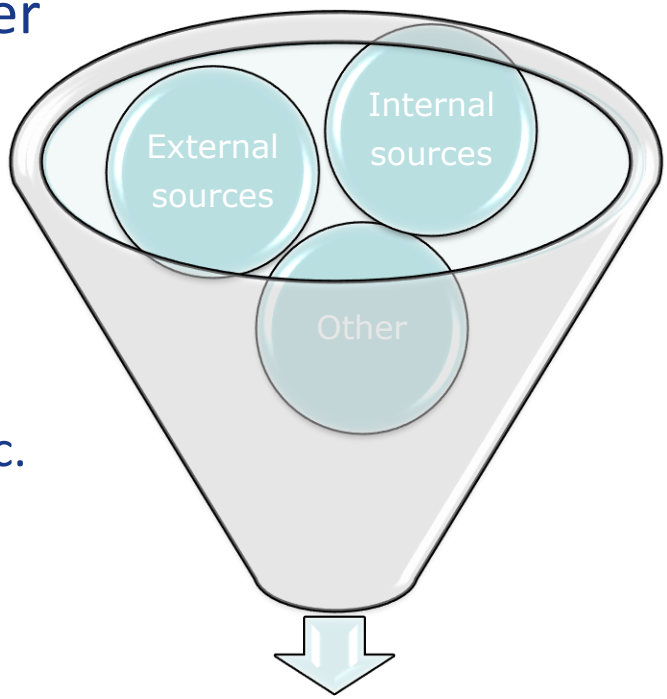
- In 2014 ENISA together with CERT Polska created the study called “Actionable information for security incident response”. Valuable input from expert group was received
- The goal of the study is to provide CERTs with guidance how to obtain actionable information.
- The study is complemented by an inventory of standards, tools , formats used in collection, exchange & processing of actionable information.
- Exercise scenario “Using indicators to enhance defence capabilities”

Description

Incident response process relies on information.

- Information can be retrieved either

from internal or externals sources.



Actionable output

Internal sources:

Firewall, router, waf, proxy, dns,

domain controller honeypots, webserver etc.

External sources:

Security feeds, osint, emails, advisories etc

Definition

- Business lexicon:

actionable information = data that can be used to make specific business decisions. It must be relevant, timely , accurate and complete relative to the business goals.

- Intelligence community lexicon:

actionable information = timely, accurate information resulting from the planning, direction, collection, exploitation, processing, analysis, production, dissemination, evaluation, and feedback of available information concerning threats to the country.

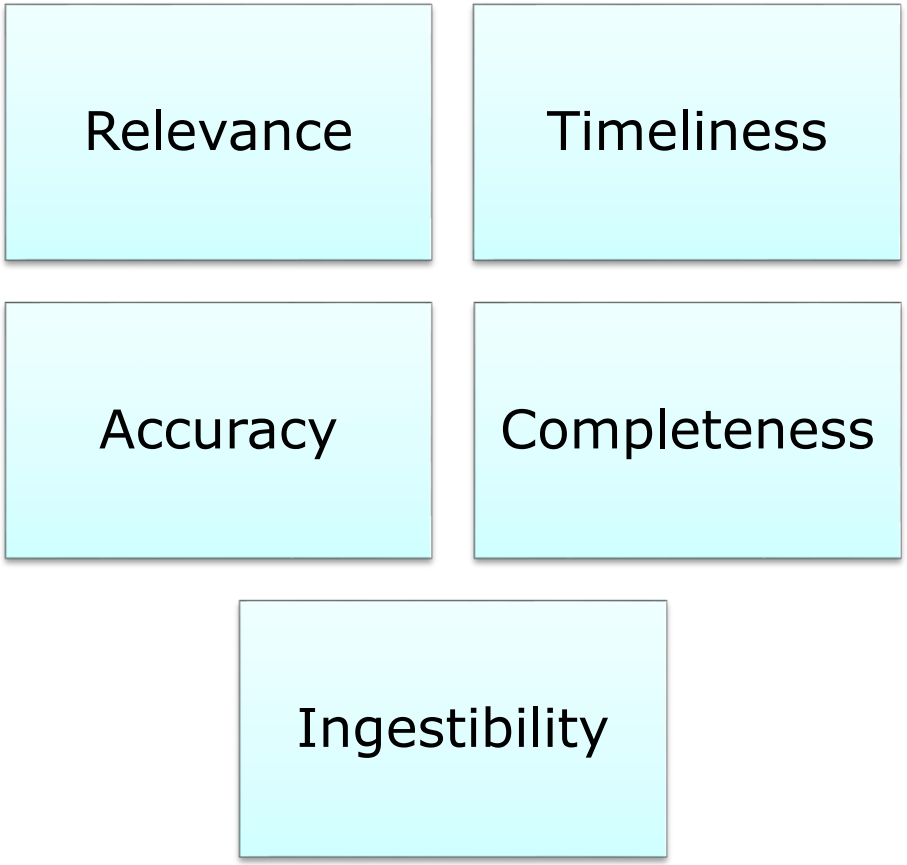
Similar concept:

- IT Security lexicon:

actionable information = relevant, timely, accurate data, that can be used to take action against attacks, threats or prevent future attacks.

Properties of actionable information

- Based on the feedback from the expert group, 5 criteria have been defined :



1. Properties: Relevance

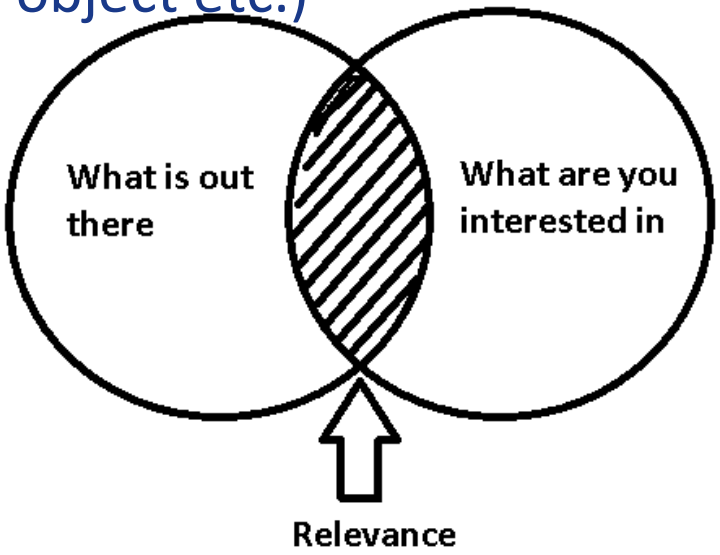
- In order for information to be **relevant**, it must be applicable to the recipients constituency, including networks, software and versions, hardware platforms.

Ex. List of IPs of compromised hosts is relevant

if the hosts belong to the constituency

hence the importance of properly describing

the constituency of CERTs. (RIPE IRT object etc.)



2. Properties: Timeliness

- Information must be **timely**

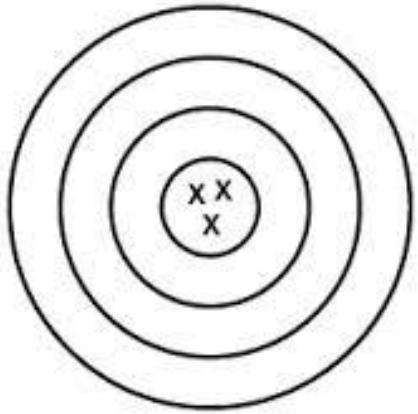
Due to the dynamics of attacks sometimes, information can become irrelevant hence non-actionable rather quickly.

Time synchronization is a critical factor which can have huge impact on the other properties of information, if not done proper.

Ex. Nowadays most CERTs share large amounts of data among them (e.g. passive DNS, malware data, blacklists) . Out of phase data with mismatched timestamps could render the date un-usable or irrelevant.

3. Properties: Accuracy

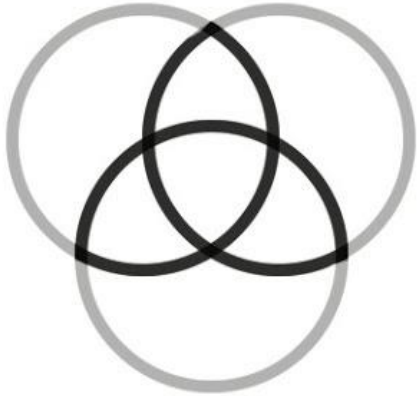
- Information needs to be **accurate.**



The consumer of the information should be able to immediately process the data under the assumption that the data is tidy , error free, with few false positives or false negatives. The accuracy is a result of a combination of mutual trust between source and receiver and local context at the receiver. Other important factors that impact source alleged confidence are the transparency of sources and means of collection.

4. Properties: Completeness

- Information needs to be **complete**.



Due to a certain number of reasons the producer of the data might omit pieces of information.

Ex. When a scan is detected , complete information would include not only the source ip, but also the destination ip , source port, destination port and other traffic characteristics.

Some producers limit their information due to different kinds of constrains (legal, not wanting to divulge their technologies etc)

5. Properties: Ingestibility

- Information must be **ingestible**

This implies that standardized data formats should be used by the producer in order to reduce complexity and make easier the extraction of observables & indicators on the consumer side.

Automation is an important component that should be taken into consideration when choosing the standard.

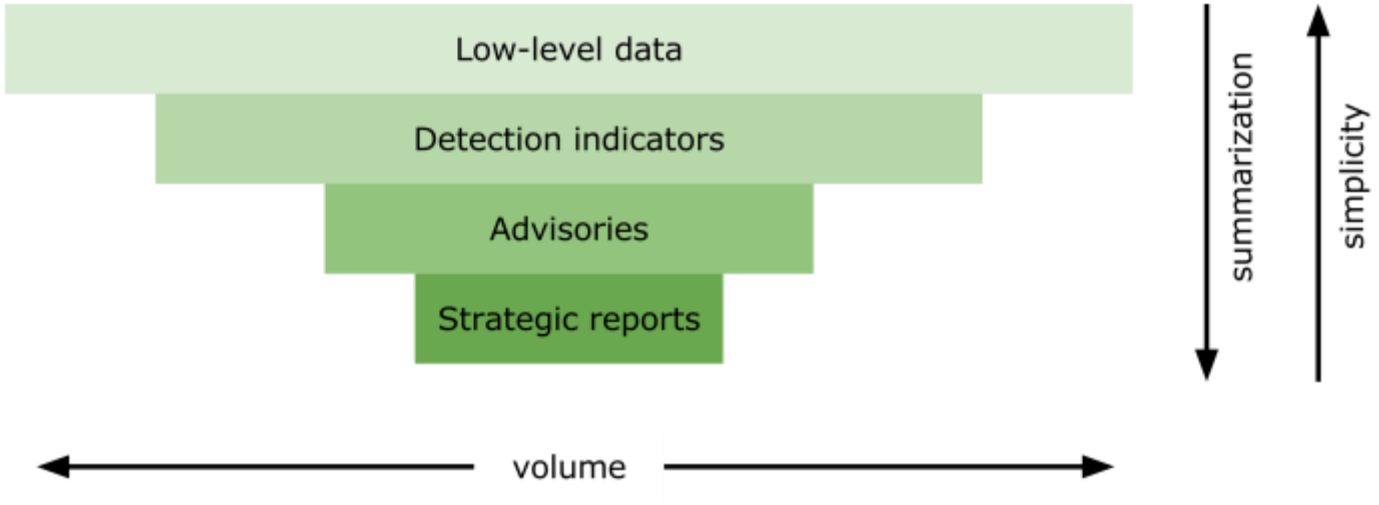
The choice for a particular standard can be influenced by the number of consumers, volume & frequency of the data etc.

The spectrum of actionable information I

Level of information	Types of information
Low-level	• network flow records and full packet captures • application logs, including typical IDS alerts • samples of executable files, documents, and email messages
Detection indicators	• IP addresses, DNS names, and URLs • specific values of format-specific fields, for example email headers • artifacts (e.g., hashes, registry, keys) related to malware • sequences of low-level events (e.g., syscalls, packets) linked to malicious behavior
Advisories	• vulnerabilities, exploit code, patches and patch status • high-level patterns of activity on a host, service, network or internet level
Strategic reports	• highly summarized threat analyses, written in prose

The spectrum of actionable information II

Figure 1. Levels of information



Levels of information

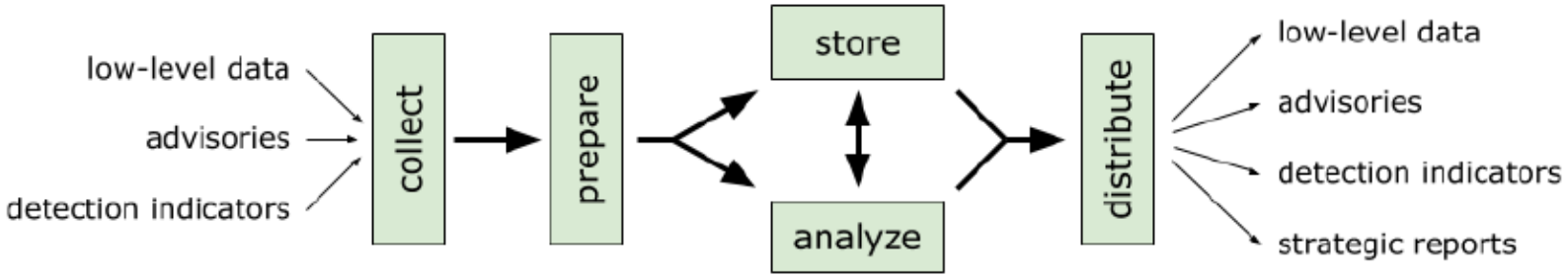
- **Low level data** = machine generated data, in large volumes , requires context in order to become actionable. (DNS logs, NetFlow, Proxy logs etc)
- **Detection indicator** = a pattern that can be matched against low-level data in order to detect threats. (IP, MD5/SHA1 hash, URL etc.). Usually it comes complemented by additional information in order to give more context. This type of information is the result of manual analysis (malware RE, IDS, honeypot etc.) or automated malware analysis using sandboxes. This category of information is in most cases actionable.
- **Advisories** = Information at this level it is considered high-level, becomes actionable if it's put in context with other information.

Info at this level is usually valuable from a defensive point of view.

- **Strategic reports** = highly summarized data that can provide overview of particular situations.

Generalized information pipeline

Figure 2. Generalized information processing pipeline.



- The 5 steps of the pipeline are: **collection** , **preparation**, **storage**, **analysis** and **distribution**.
- The above conceptual model is meant to provide a structure to the study.
- Its main goal is to describe multiple aspects of systematic information handling.

Collection

- The most important step of the pipeline, any issue at this step can have serious impact in next steps.

CERTs rely on external / internal data for incident handling:

Internal sources are more reliable and have better time accuracy, collection can be difficult when done at a scale. Can be correlated with external sources for threat detection.

There are certain number of issues when dealing with external data like (security feeds, reputation services, etc.):

- Doubt related to accuracy of the data
- Lack of transparency on the collection methods
- Some sources omit the time zone information (making time correlation difficult)
- Is the data generated automatically ?

Properties of data collection methods

- **Recurrence**: - is the source publishing distinct events in case by case manner or it behaves like a regular feed of info.
- **Consumption model**: - push / pull models

Pull model use RESTful APIs and could introduce latency to the distribution and also could have scalability issues.

Push model has limitation in controlling recipients over the time range, format and volume of info sent.(Ex. email)

- **Granularity**: - every element of an event is sent separately (or data is sent in batches (ex. daily digests).

Preparation - steps

- **Parsing**

Information comes in multiple formats and need to be parsed , so that important fields are extracted.

- **Normalization**

Everything that was collected needs to be unified in one common format. Due to heterogeneity of the data & lack of common ontology , this task is challenging.

- **Aggregation**

Similar elements from the data can be grouped (aggregated) into a single one.

- **Enrichment**

Adding more context to existing information.

- **Automation**

Whenever possible make use of automation of the processes and check the consistency of the results

Storage

The storage technology & implementation needs to be chosen carefully without impacting any properties of the data

- **Retention time** – might be influenced by Legal / Technical issues.
- **Scale** – chose scalable solutions that can cope with high-volumes of data.

The volume of data is revere proportional to the level of data.

High volume – Low-level data

Medium volume – Indicators

Low volume – Advisories & strategic reports.

Dataset management

When dealing with multiple sources and datasets, the management process needs to ne augmented by proper procedures. Integrity/confidentiality of the data must be ensured.

Technologies

RDBMS – SQL* or NoSQL dbs (scalability, security, performance, mgmt. options, ease of quering)

Analysis & Distribution

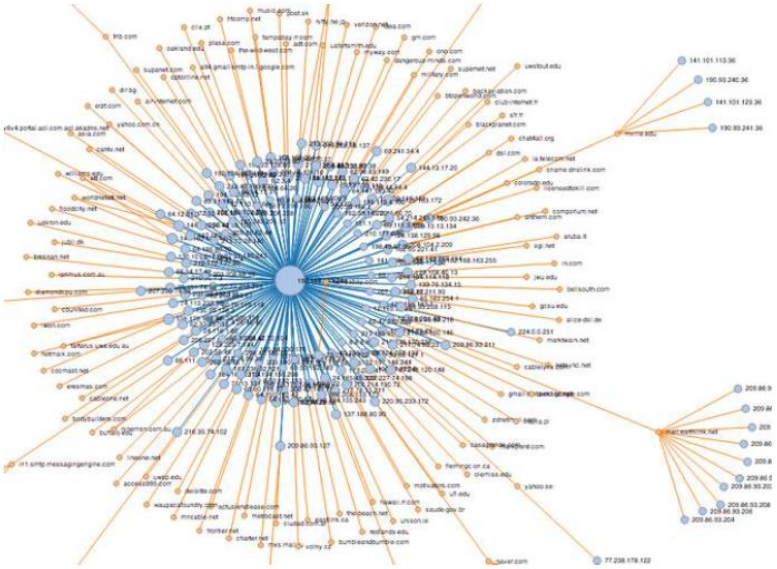
- **Analysis** yields more contextual results from original data, after triage & discarding the information which is not relevant and by going through enrichment process (passive DNS, reputation data, IP2AS etc)
- **Distribution** external / internal

Internal distribution can be fluent , e.g. sending the data to NOC or IT, but depending on the impacted infrastructure can be also difficult.

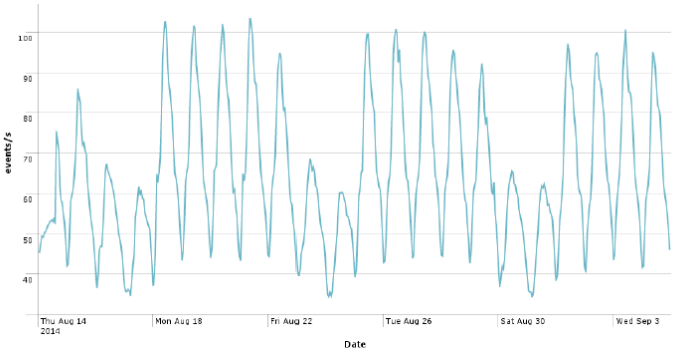
External distribution can be impacted by technical capabilities of the receivers, legal implications (NDA, laws).

Case studies

```
<?xml version="1.0" encoding="UTF-8"?>
<ioc xmlns="http://schemas.mandiant.com/2010/ioc" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" id="3b92a703-2f56-4b9d-862f-35c1fa2847d6"
last-modified="2014-10-15T09:16:04">
  <short_description>'New Unsaved Indicator'</short_description>
  <authored_date>2014-10-15T06:30:05</authored_date>
  <links/>
  <definition>
    <Indicator operator="OR" id="2655453b-799d-480d-a170-99d821f4ad1d">
      <IndicatorItem id="6a6cdce1-372e-4d25-9aed-76164c6a1983" condition="is">
        <Context document="FileItem" search="FileItem/StreamList/Stream/Name"    type="mir"/>
        <Content type="string">encrypted</Content>
      </IndicatorItem>
      <IndicatorItem id="77269c72-f649-4e7f-9dcb-4b1adadfe2cb" condition="contains">
        <Context document="Network" search="Network/DNS" type="mir"/>
        <Content type="string">home.windows-security.su</Content>
      </IndicatorItem>
      <IndicatorItem id="70005dc6-e884-4b15-b986-6060fe8586ec" condition="is">
        <Context document="DnsEntryItem" search="DnsEntryItem/Host" type="mir"/>
        <Content type="string">home.windows-security.su</Content>
      </IndicatorItem>
      <IndicatorItem id="3e3716f7-a8d3-4eb1-b990-dce2b441d11d" condition="is">
        <Context document="DnsEntryItem" search="DnsEntryItem/RecordData/IPv4Address" type="mir"/>
        <Content type="IP">184.128.98.111</Content>
      </IndicatorItem>
      <IndicatorItem id="f5156778-bc61-4b1c-ab60-02e2a7514141" condition="is">
        <Context document="DnsEntryItem" search="DnsEntryItem/RecordData/IPv4Address" type="mir"/>
        <Content type="IP">184.128.152.18</Content>
      </IndicatorItem>
      <IndicatorItem id="7bc87e9f-4da2-41c6-a143-660cf060b6a0" condition="is">
        <Context document="DnsEntryItem" search="DnsEntryItem/RecordData/IPv4Address" type="mir"/>
        <Content type="IP">184.128.2.34</Content>
      </IndicatorItem>
      <Indicator operator="AND" id="9dda7744-001d-49bf-a2ad-579daa4f5b2c">
        <IndicatorItem id="391b98de-0383-440e-a0cf-9cfe6663764d" condition="is">
          <Context document="RegistryItem" search="RegistryItem/KeyPath" type="mir"/>
          <Content type="string">HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Run</Content>
        </IndicatorItem>
        <IndicatorItem id="31afd6cd-7c98-44b8-b188-d2e92d374e65" condition="is">
          <Context document="RegistryItem" search="RegistryItem/ValueName" type="mir"/>
          <Content type="string">driver32</Content>
        </IndicatorItem>
      </Indicator>
    </Indicator>
  </definition>
```

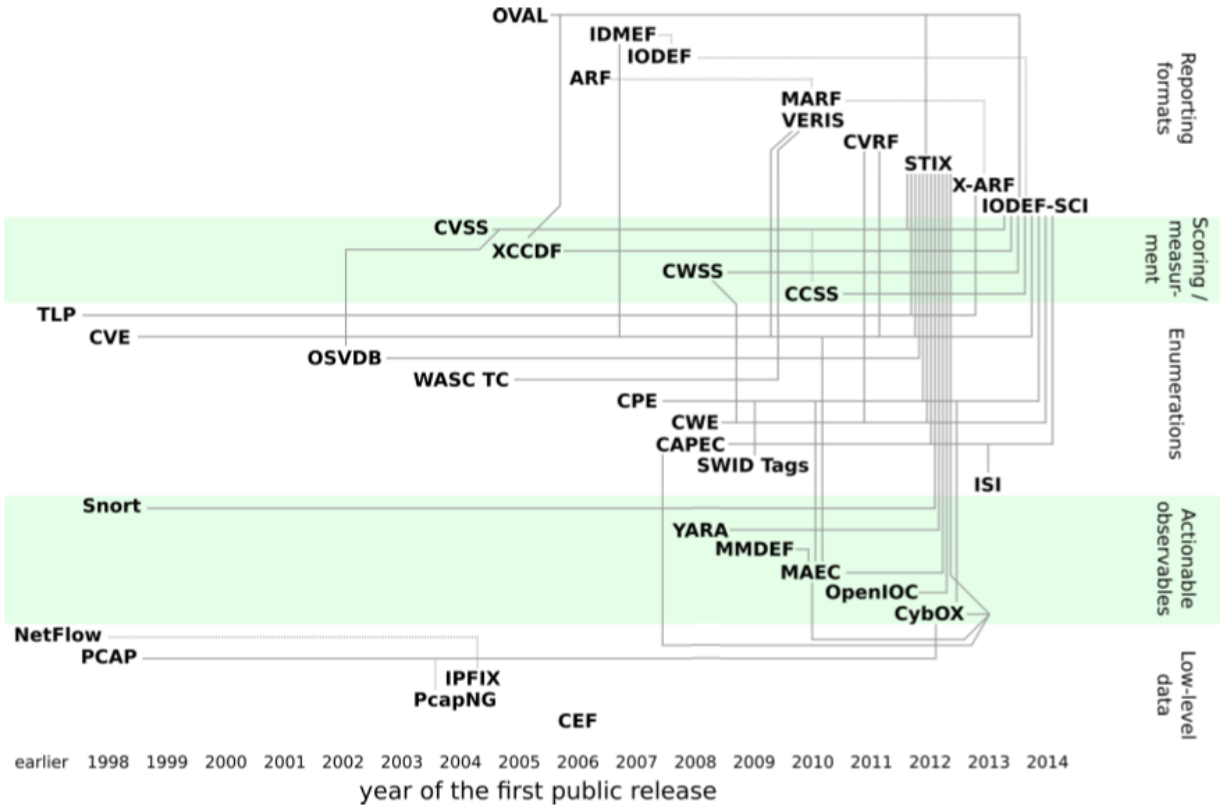


Phase	Detect	Deny	Distrupt	Degrade	Deceive	Destroy
Reconnaissance						
Weaponization	NIDS					
Delivery		Firewall ACL				
Exploitation		Patch				
Installation	HIDS					
C&C	NetFlow, NIDS	Firewall ACL			DNS redirect	
Action on Objectives	Audit log				Honeypot	



The inventory

- Standards, Transport and serialization, Information management tools



Conclusions

- Define clear sharing rules and labels on the exchange data.
- Don’t start from scratch , build upon existing tools.
- Assess the possibility of applying additional processes that can provide more context.
- Use security visualization in analysis.
- Use standardized formats & transport mechanisms (see inventory)
- Keep it simple, unless a data format is required.
- Scale based on the requirements of each data type and volume. Take into consideration the overhead that data formats coming in larger volumes

Details

The study , the inventory and training scenario can be found at the following address:

https://www.enisa.europa.eu/activities/cert/support/actionable-information

Short URLs:

The study

http://goo.gl/AAfc9a

The inventory

http://goo.gl/F68V4z

The training scenario

http://goo.gl/Ko3Eey

The VM & materials for the training scenario:

https://www.enisa.europa.eu/ftp/crits-headless.ova

https://www.enisa.europa.eu/ftp/crits-headless-files.zip

Thank you

Follow ENISA: