Not just indicators: automated data processing with n6

Paweł Pawliński
pawel.pawlinski@cert.pl

CERT.PL

FIRST Annual Conference

Kuala Lumpur, 2018-06-25

Today's workshop: open source n6 platform

https://github.com/CERT-Polska/n6

or get the demo VM image

keyword:

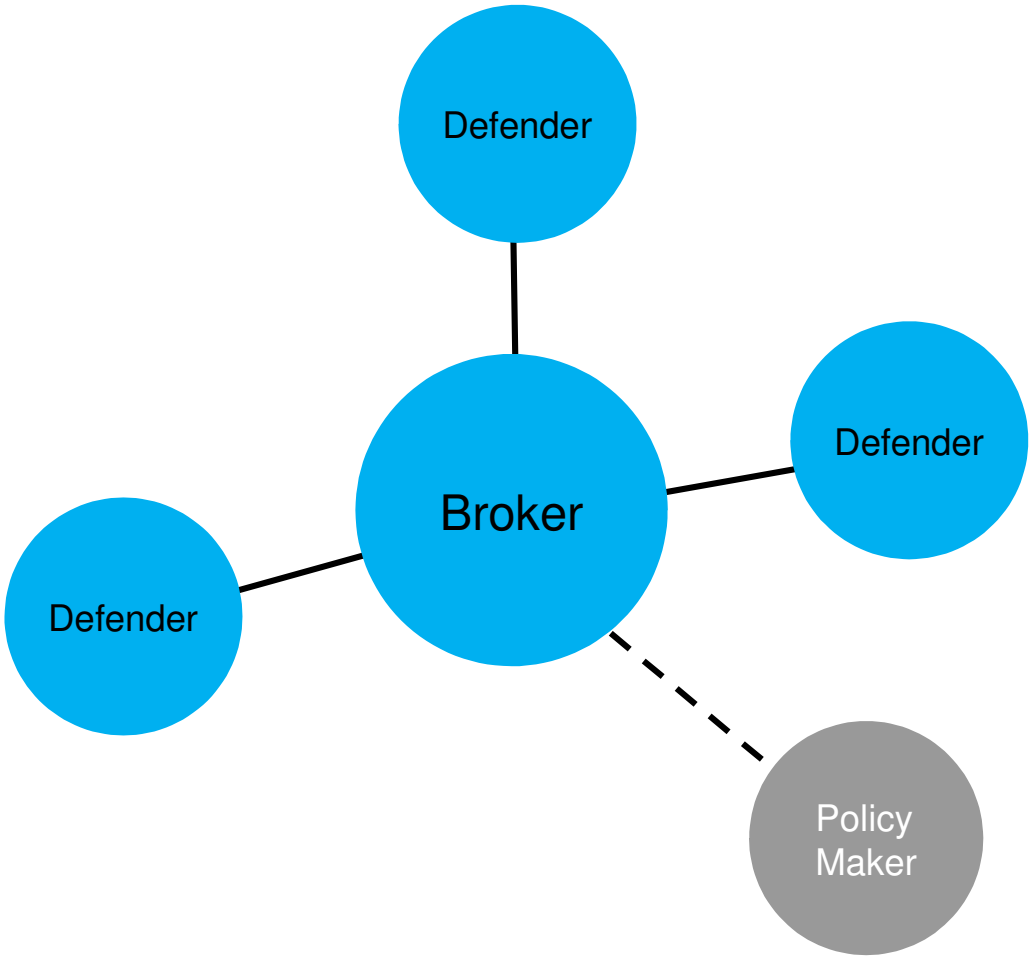
automation

Agenda

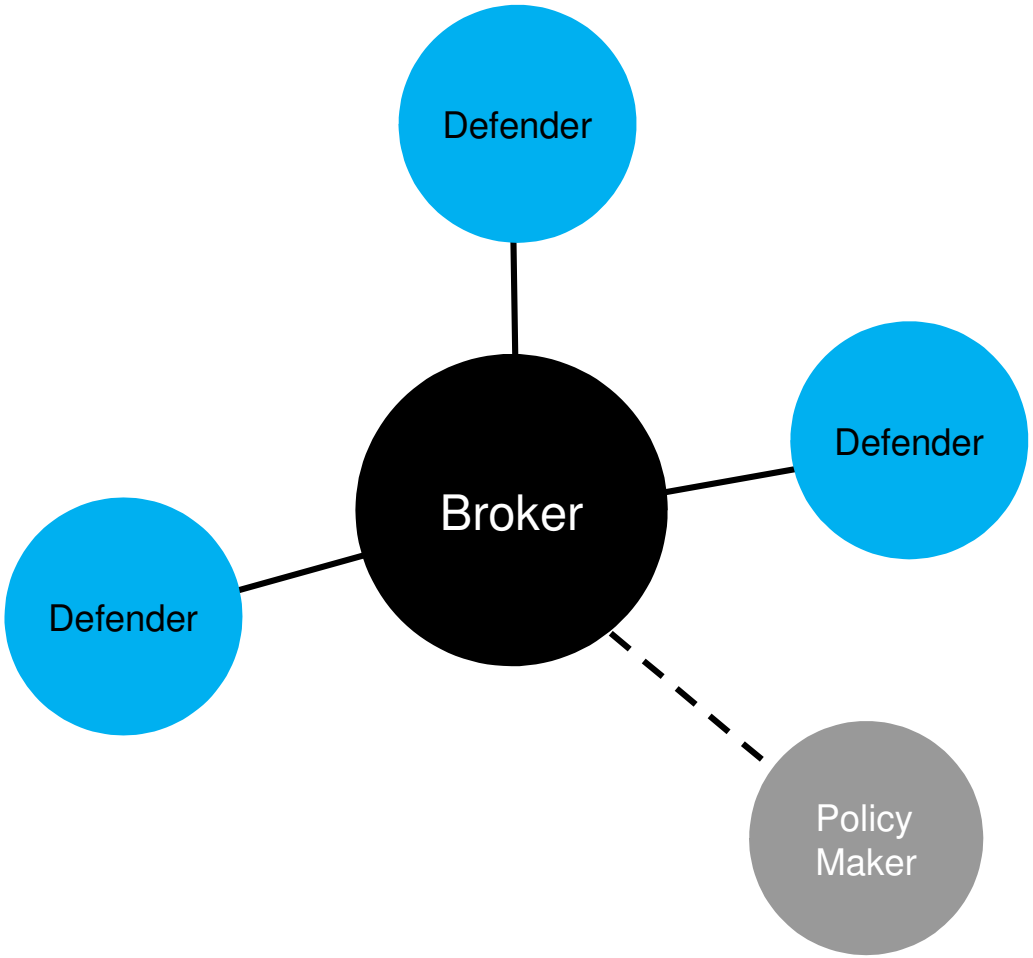
CERT.PL: quick introduction

- Established in 1996
- Constituency:
  - national CSIRT
  - except government, military, critical infrastructure
- Part of NASK:
  - research institute
  - **.pl** registry
  - software development
  - ISP
  - ...
- Trying to share information & tools

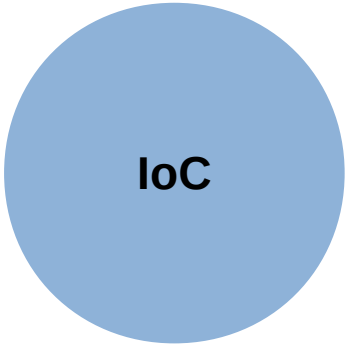
Our place in the information flow



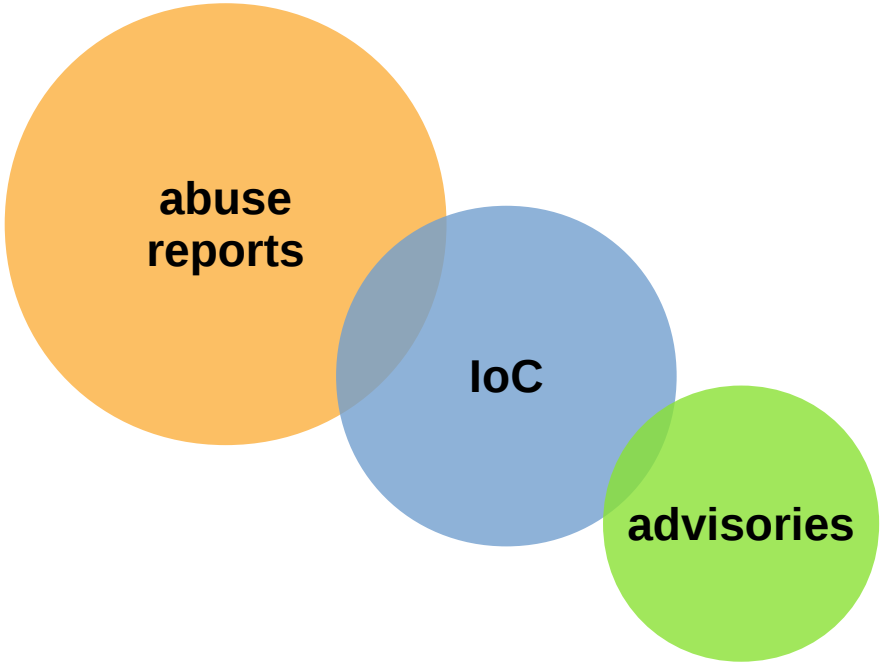
Our place in the information flow



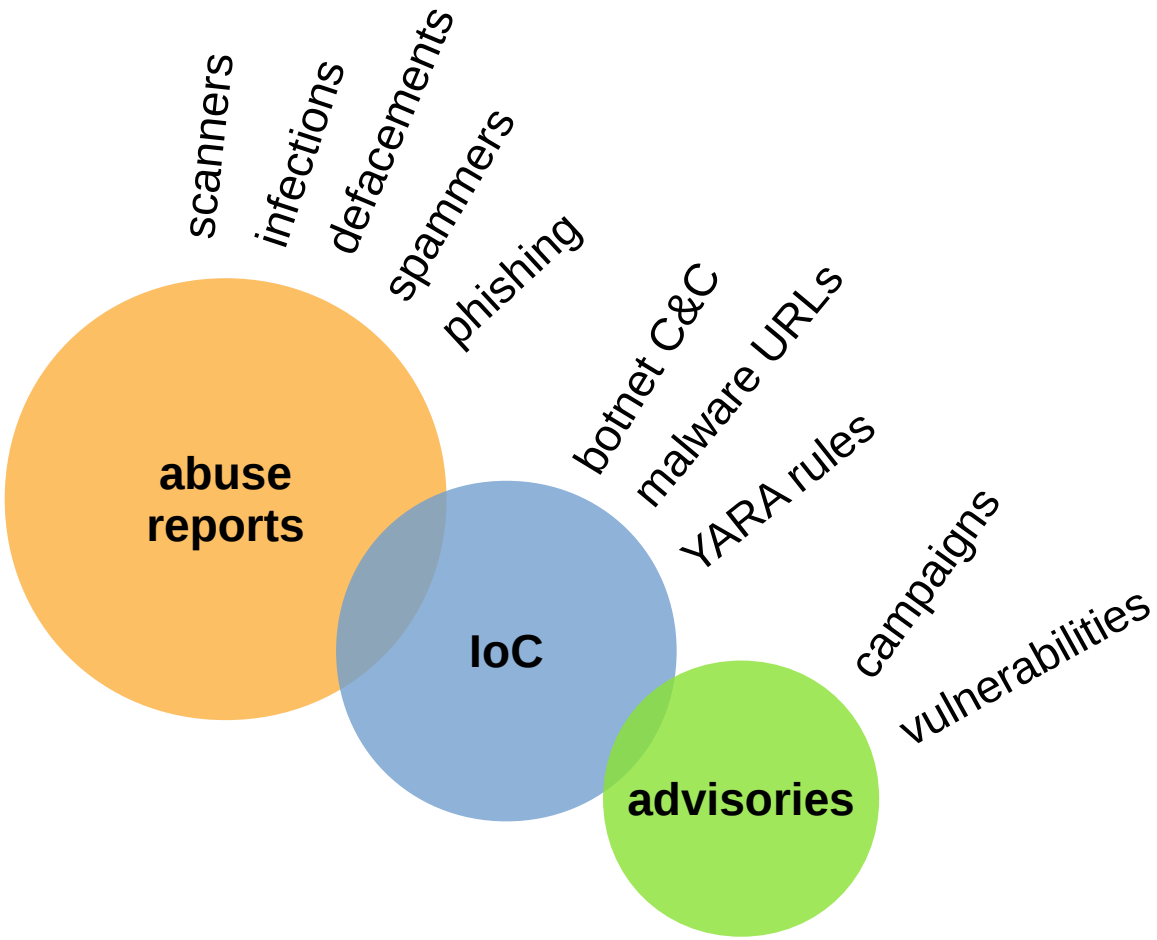
What kind of data you process?



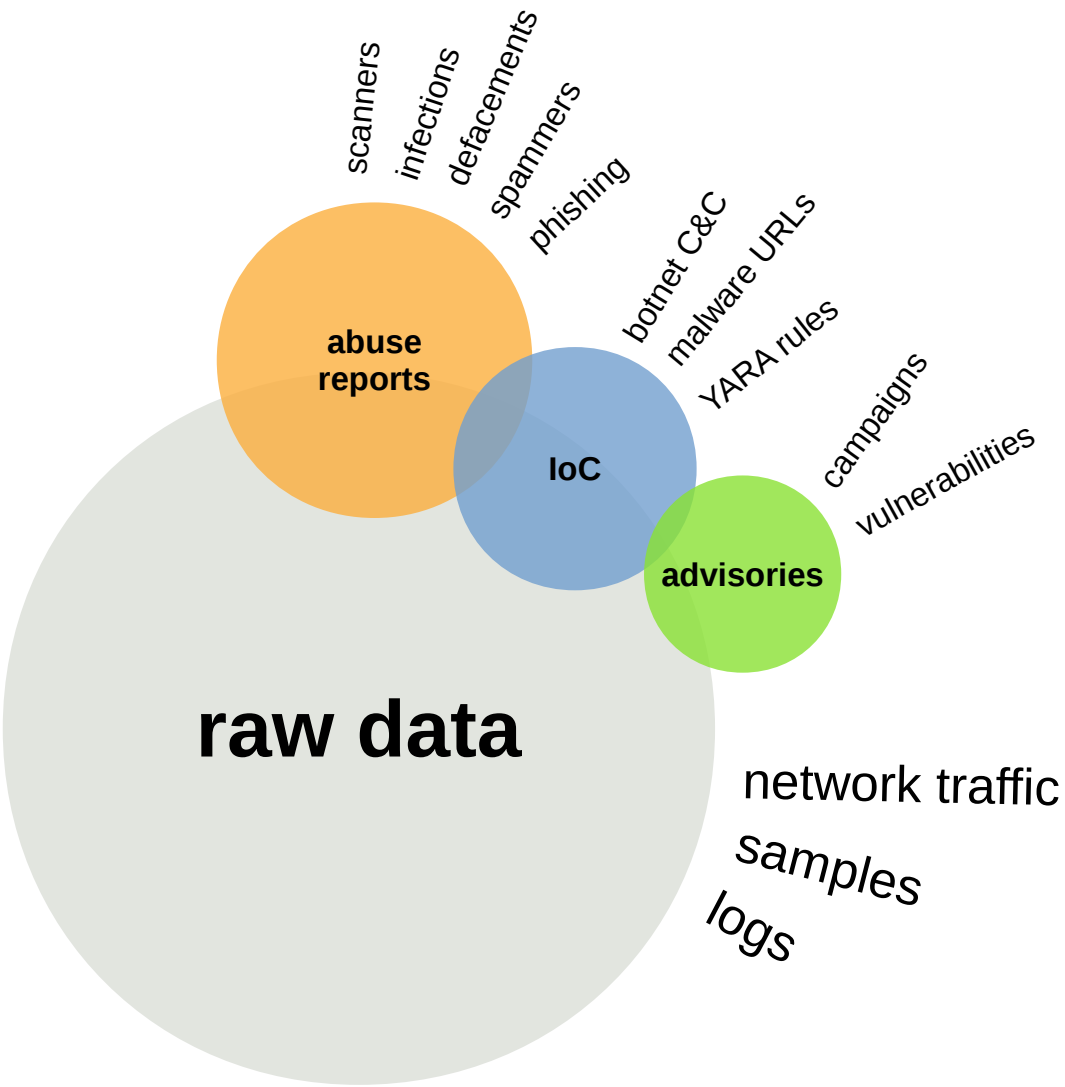
What kind of data you process?



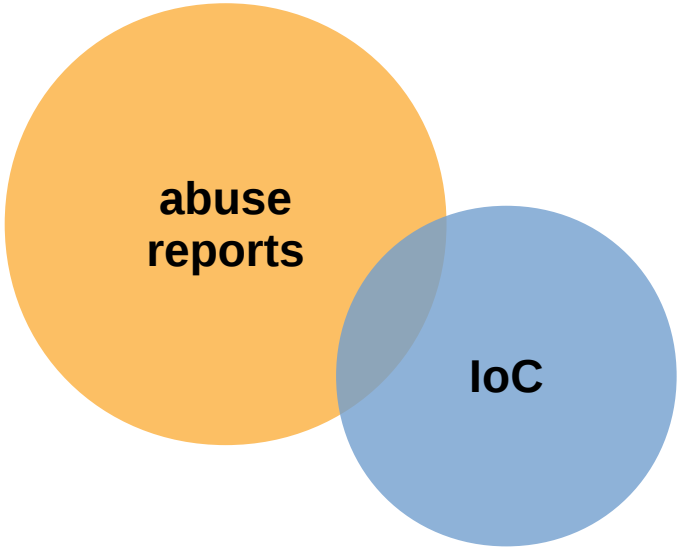
What kind of data you process?



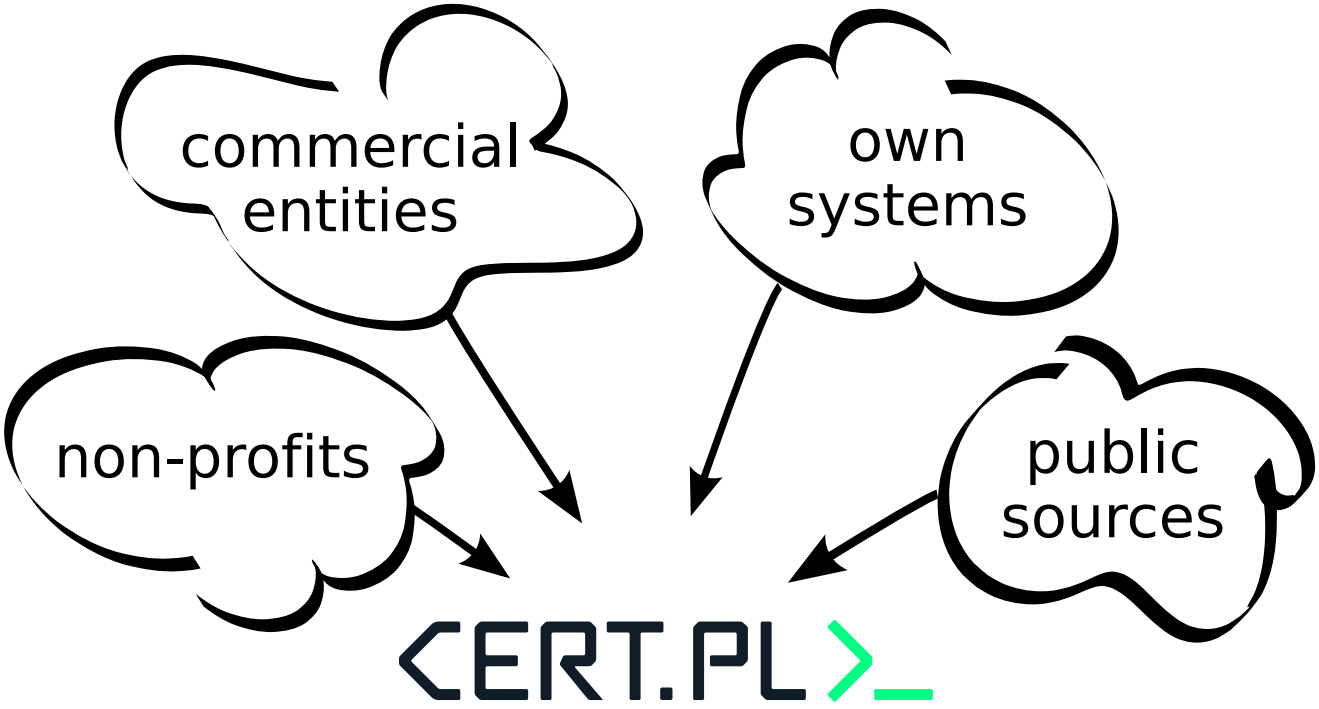
What kind of data you process?



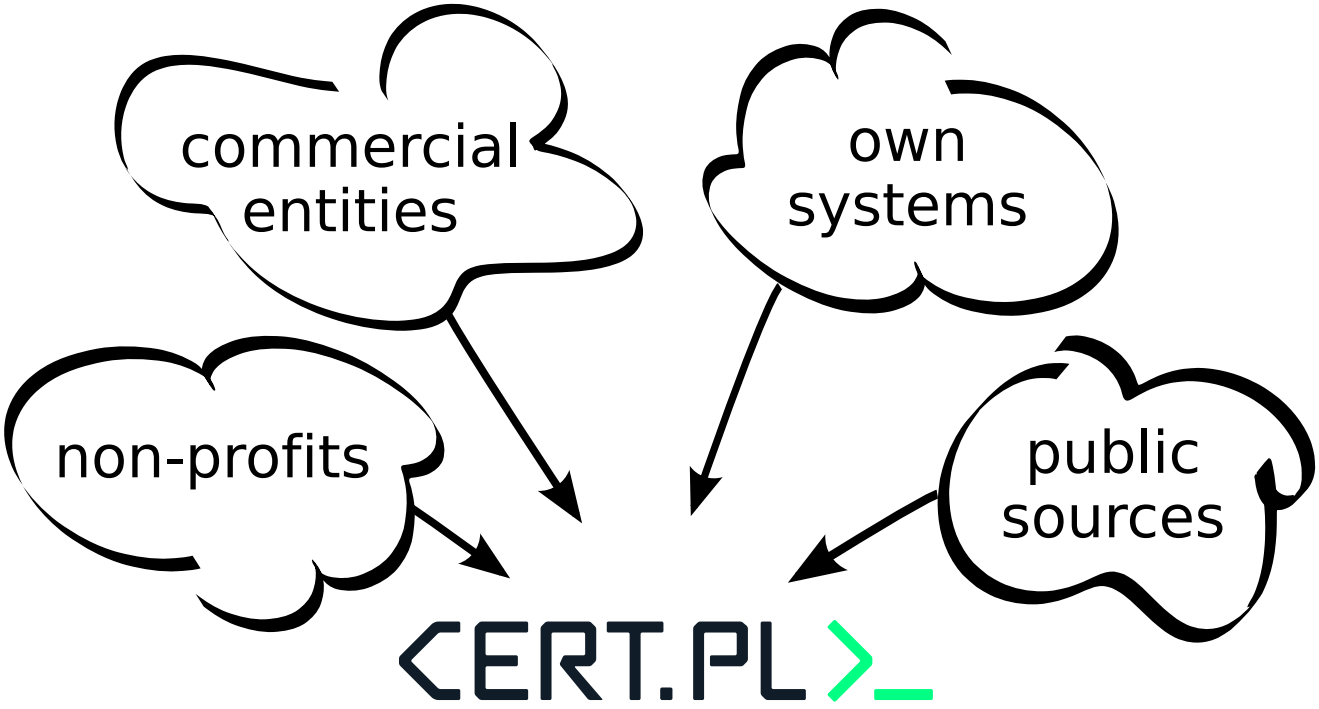
What kind of data you process?



Sources of information



Sources of information

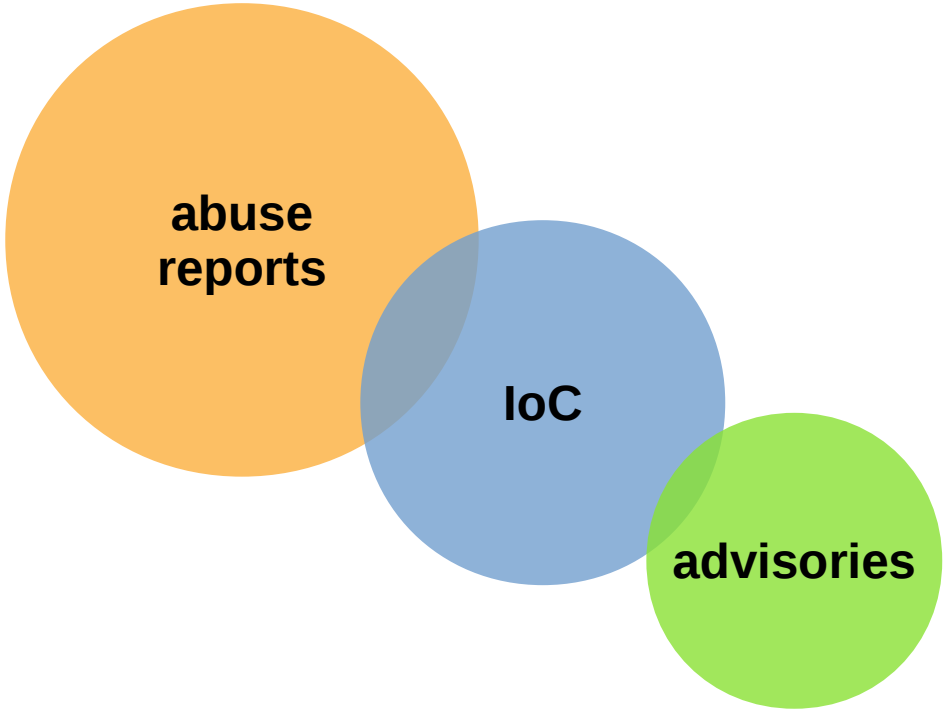


40+ data providers
80+ active incoming data feeds
1M+ events per day

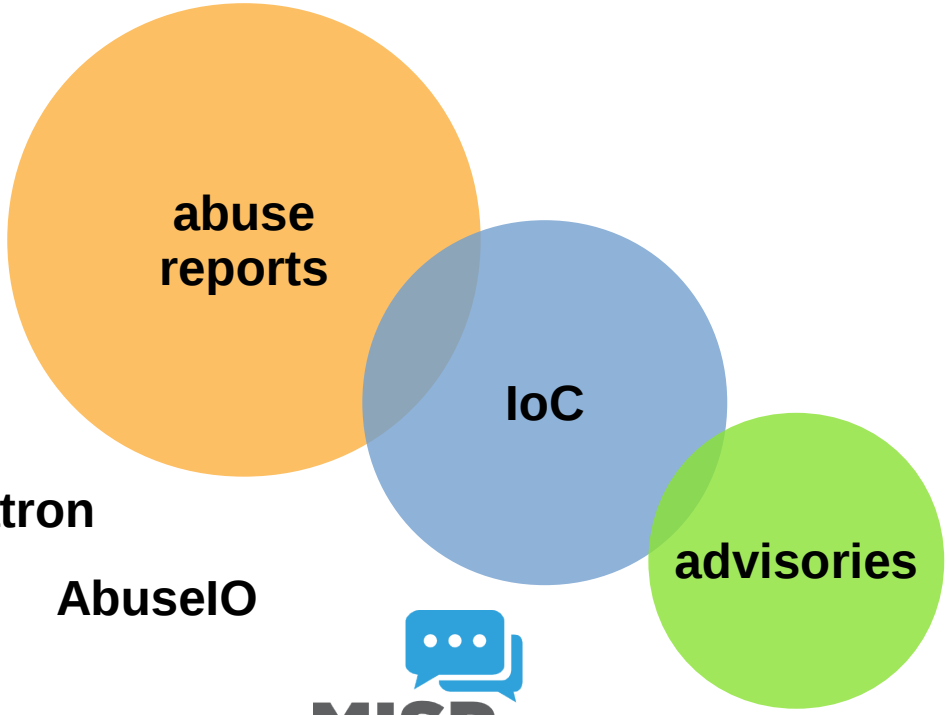
Own systems

- Sinkhole
  - infections
- Malware tracking
  - C&C infrastructure
  - configuration, injects
- Honeypots & darknet (network telescope)
  - attacks
  - scans
  - denial-of-service
  - see SISSDEN project
    `https://sissden.eu`
    presentation by Shadowserver on Tuesday

Tooling



Tooling



Megatron

AbuseIO

INTELMQ

MISP Threat Sharing

TARANIS Vulnerability Management

Commercial Threat Intelligence Platforms

Tooling

network security incident exchange

Tooling

network security incident exchange

Agenda

n6: first generation

- Deployed late 2011
- Minimalistic by design
- Filtering: client gets relevant data only
- Keeping original format
- Enrichment
- Flat files served directly by Apache
- Authentication using X.509 certificates
- Last commit 2015, shut down 2017

Original code

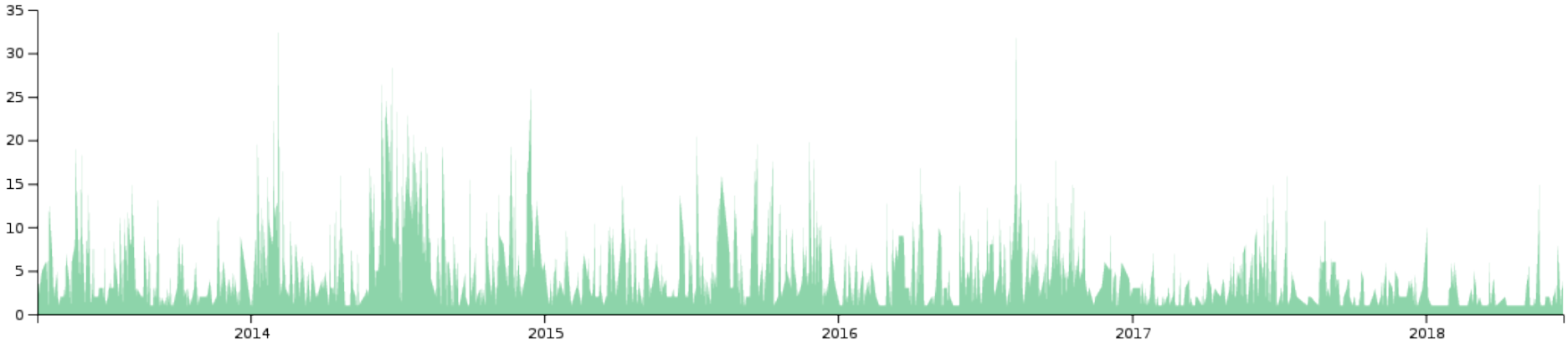
```
SLOC     Directory        SLOC-by-Language

2488     transfer         python=1982,sh=391,perl=115
1346     sources          perl=1198,python=82,sh=66
1280     pyn6toolkit      python=1280
886      manage           sh=886
517      engine           perl=517
```

Generated using David A. Wheeler's 'SLOCCount'.

Main objectives of n6 (2013)

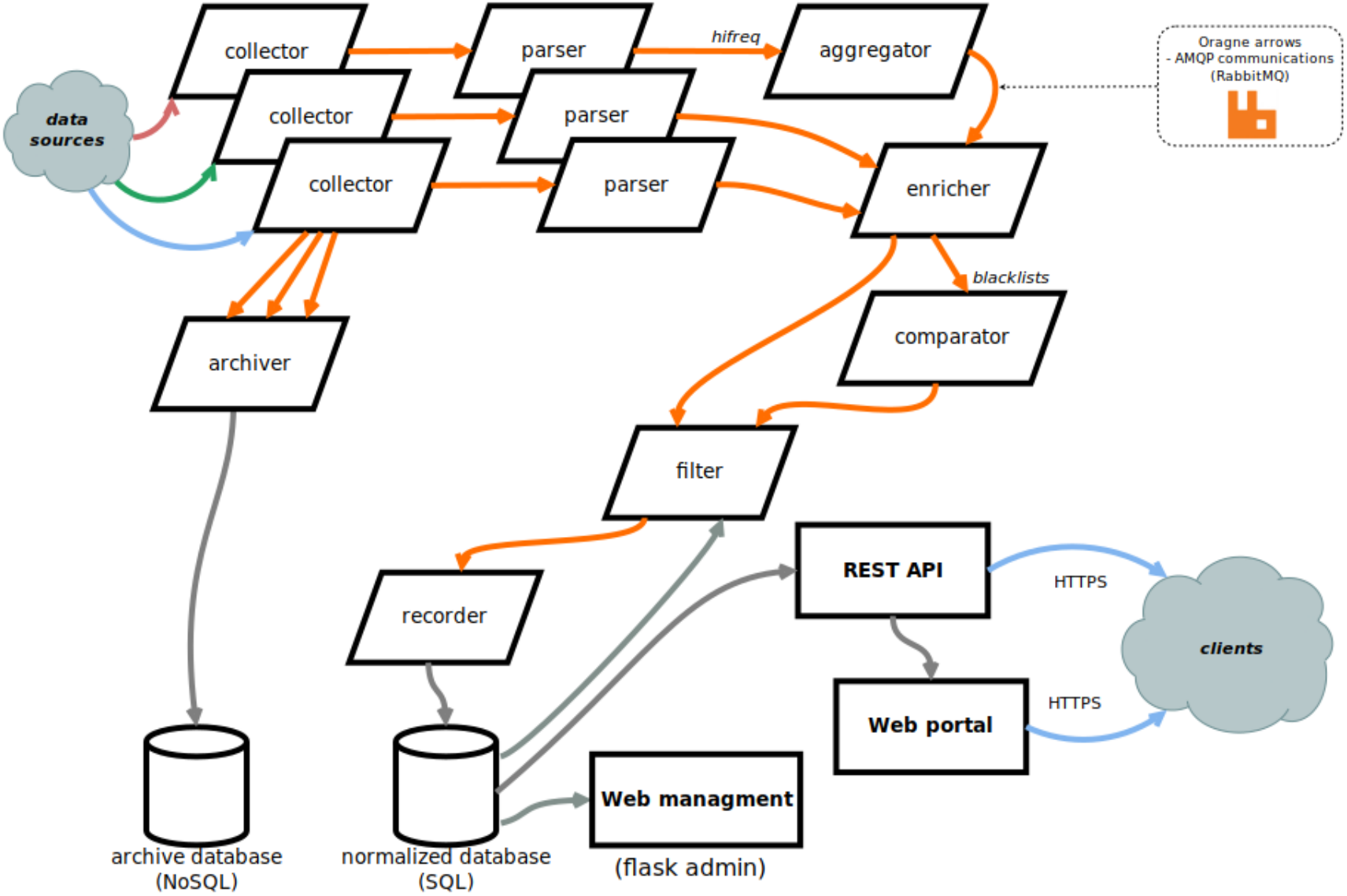
1. Provide information to our constituency
2. Get actionable conclusions value from data available
3. Obtain data from as many sources as possible
4. High throughput: gigabytes, 10M+ events daily
5. Easy to query, simple data model
6. Secure access, fine grained permissions
7. Maintain and improve quality of incoming data
8. Minimize manual & maintenance effort
9. Reliable (including HA)

n6 new generation: 2013+



commits

Architecture



Handling of incoming data

- Stream/message oriented architecture: RabbitMQ
  - AMQP: standard protocol
  - configurable flow of messages
  - integration with other services
  - web management
  - HA
- Collectors: specialized code to fetch data from sources
- Parsers: convert to event streams and normalize
- Aggregator: on the fly to deduplication
- Enricher: DNS, ASN, country code
- Comparator: blacklist state tracking
- Filter: organization (client) mapping
- Recorder & Archiver: persistence
- Web interfaces
  - clients: sign-up, browse data, manage access
  - admins: full management
- (upcoming) Notifier: send statistics on new data

Storage

- Original data
  - Document store: MongoDB
  - Collection per source
  - Files (GridFS) & arbitrary BSONs
  - Compressed size: 1.4 TiB
- Normalized events
  - SQL: MariaDB + TokuDB engine
  - Optimized schema
  - Indexes, partitioning
  - Transparent compression
  - 3 B records in total
  - 2 TiB disk space
  - Up to 500 inserts/s per recorder
  - Designed for batch reads (up to millions of events)
- Critical for overall performance

Sharing interface

- Simple to use REST API
- Multiple output formats: **JSON**, CSV, IODEF, (upcoming: STIX 2)
- Real-time stream API (STOMP)
- Flexible permission model, attribute-level granularity
- Authentication via client X.509 certificates
- Test endpoints with autogenerated data

n6 vs IntelMQ

- Similarities
  - inspired by AbuseHepler
  - Python
  - queues, modular / microservices
- IntelMQ:
  - focus on notifications (email)
  - more generic, build your own parts
  - active developer community
  - management tools build from scratch
- n6:
  - focus on feeds (API)
  - events aggregate on the fly
  - leveraging existing tools (RabbitMQ, supervisor, Flask admin)
  - included: complete database, ACLs, flexible queries
  - user web interface

IntelMQ integration

- New in n6: elastic pipeline
- Running IntelMQ bots in n6 pipelines
- Adapters for message conversions: n6 $\rightarrow$ IntelMQ, IntelMQ $\rightarrow$ n6
- Mapping: attributes, taxonomy

n6 vs MISP

- MISP:
  - focus on sharing IoC
  - broad set of use cases
  - very sophisticated data model (taxonomies, galaxies, etc)
  - multiple sharing arrangements (peer-to-peer & other)
- n6:
  - narrow focus: provide feeds, primary abuse data
  - filtering data: only relevant events
  - mostly network IoCs
  - simple data model
- Integration: MISP collector
  - support for incremental updates
- n6 can complement MISP for distribution of abuse data

Agenda

RabbitMQ

RabbitMQ

Overview Connections Channels Exchanges Queues Admin

Queues

All queues (8)

Pagination

Page 1 of 1 - Filter: Regex (?)(?)

Overview			Messages			Message rates			+/-
Name	Features	State	Ready	Unacked	Total	incoming	deliver / get	ack	
aggregator	D DLX	idle	0	0	0				
comparator	D DLX	idle	0	0	0				
dba	D DLX	idle	0	0	0				
dead_queue	D	idle	0	0	0				
enrichement	D DLX	idle	0	0	0				
filter	D DLX	idle	0	0	0				
spam404-com.scam-list	D DLX	idle	0	0	0				
zbd	D DLX	idle	0	0	0				

Add a new queue

HTTP API | Command Line

API & data format

```
curl --key key.pem     \
     --cert cert.pem   \
     --insecure        \
https://localhost:4443/\
search/events.json
```

API & data format

```
curl --key key.pem      \
     --cert cert.pem    \
     --insecure         \
https://localhost:4443/\
search/events.json
```


```
{
    "status": "replaced",
    "restriction": "public",
    "confidence": "low",
    "replaces": "3e193d38ab180cdc16f0ce9553c43498",
    "url": "http://api.ctp-line.ru/distrib",
    "expires": "2017-11-29T00:05:03Z",
    "modified": "2017-11-28T01:32:54Z",
    "fqdn": "api.ctp-line.ru",
    "source": "malwarepatrol.malurl",
    "client": [
        "cert.ee"
    ],
    "time": "2017-11-26T00:05:03Z",
    "rid": "24c03a5f4373d85f934e037f6aca3651",
    "category": "malurl",
    "id": "b2a85dd6cdf5816da8440ac6c8457d83",
    "address": [
        {
            "cc": "EE",
            "ip": "185.4.75.25",
            "asn": 198068
        },
        {
            "cc": "RU",
            "ip": "78.85.20.223",
            "asn": 12389
        },
        {
            "cc": "RU",
            "ip": "91.146.50.179",
            "asn": 3226
        }
    ]
},
```

Web interface

n6 Portal

Other threats | Search events | Threats inside my network | Export table | Admin panel | Logout

Select search criteria:

Country

Add new

Start date

2018-06-15T23:09:48

Country

PL

Remove

Max results: 100

Displayed fields

Search

Found 22 entries.

Time	Category	Name	IP	ASN	Country	FQDN	Source	Confidence	Origin	URL	Protocol
2018-06-22T16:46:43Z	scam		62.129.193.238	12824	PL	gamegensire.com	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		91.228.199.104	198414	PL	adultgameshacked.com	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		91.228.199.230	198414	PL	fastfilez.net	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		91.228.199.230	198414	PL	celebrityphone.net	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		62.129.193.238	12824	PL	robucs.com	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		178.32.205.96	16276	PL	lekturymp3.pl	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		212.91.26.153	57367	PL	giftcode.pl	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		85.128.131.76	15967	PL	pushcloud.org	spam404-com.scam-list	low			
2018-06-22T16:46:43Z	scam		46.242.165.31	12824	PL	flamehacks.com	spam404-	low			

https://localhost/report/inside

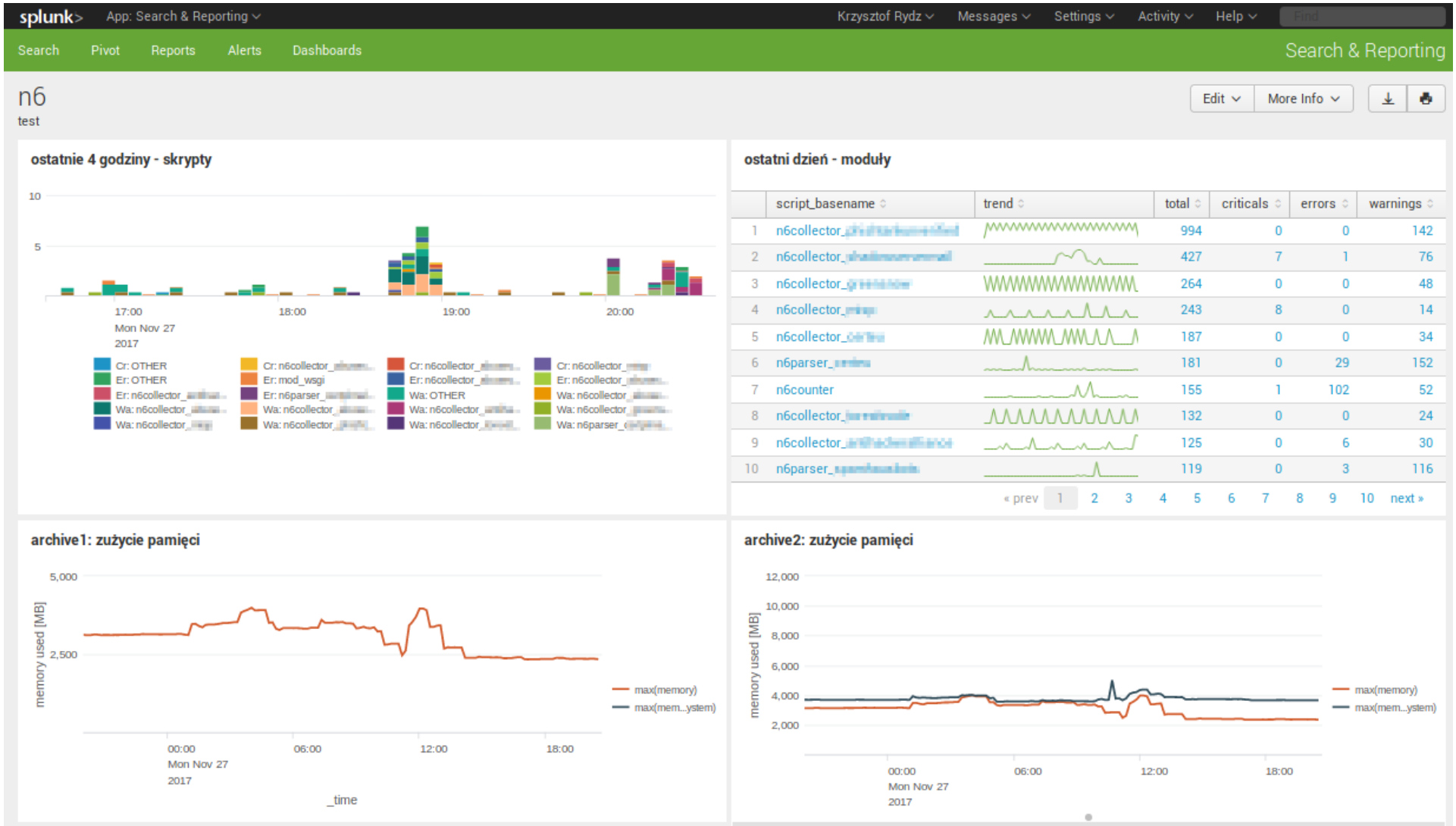
Configuration management

admin_panel | Home | Org | Org Group | User | Criteria Container | Source | Subsource | Subsource Group | System Group

List (2) | Create | With selected | Search

	Org Id	Full Access	Access To Inside	Access To Threats	Access To Search
	example.com	✓	✓	✓	✓
	testorg.com	⊖	✓	✓	⊖

Monitoring: logging to Splunk



Monitoring: performance monitoring

Agenda

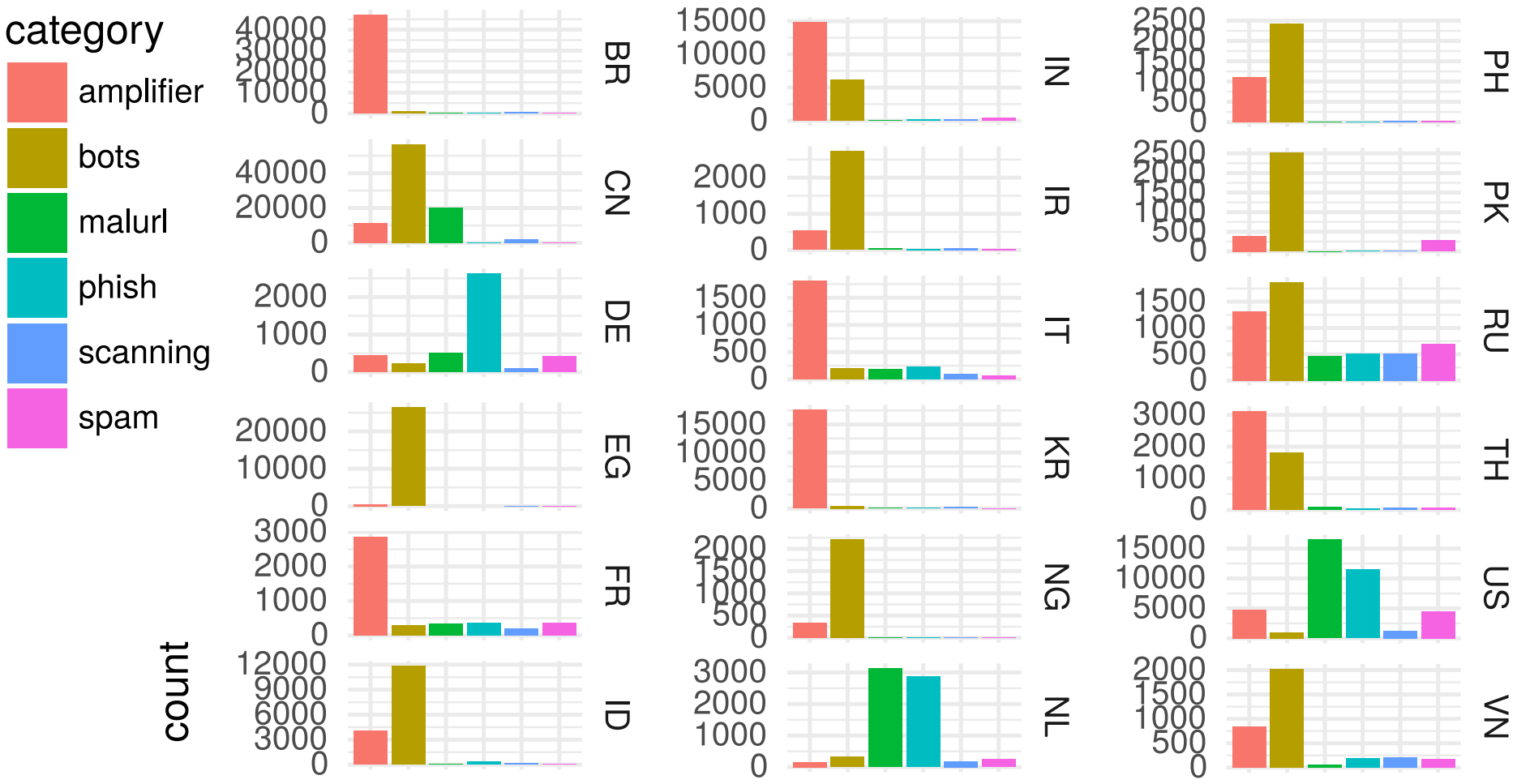
Sharing with organizations in Poland

- Primary use case
- Free service for network owners
- 250+ registered organizations
- 100+ active users

Sharing with organizations in Poland: conclusions

- Challenges:
  - low uptake by ISPs (data on customers is unused)
  - recipients might not know what to do with the data
  - troubles automating processing on the client side
  - rare feedback
- Many recipients require human interface
- Motivation for better delivery methods

Data on other countries: avg events daily in 2018



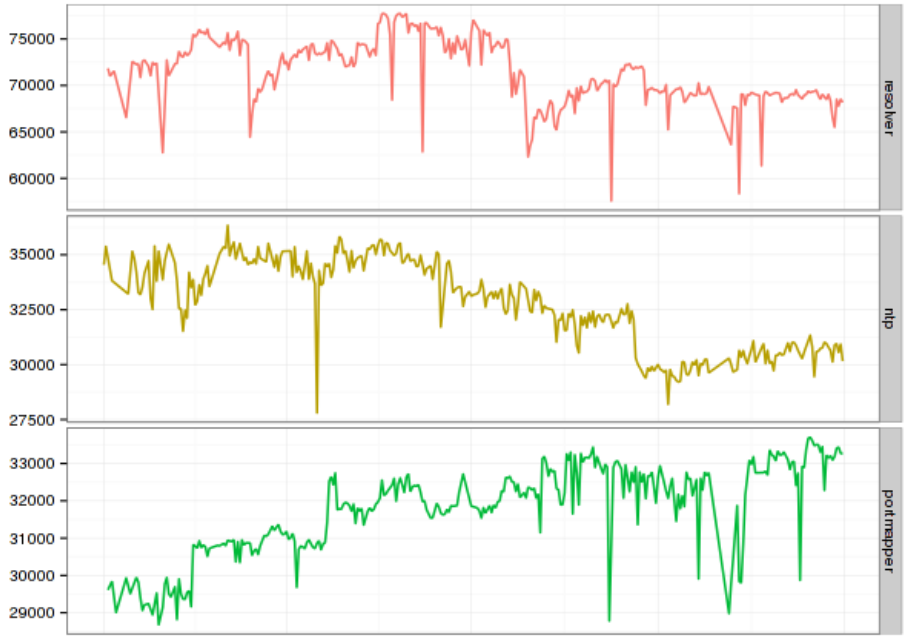
Data on other countries: conclusions

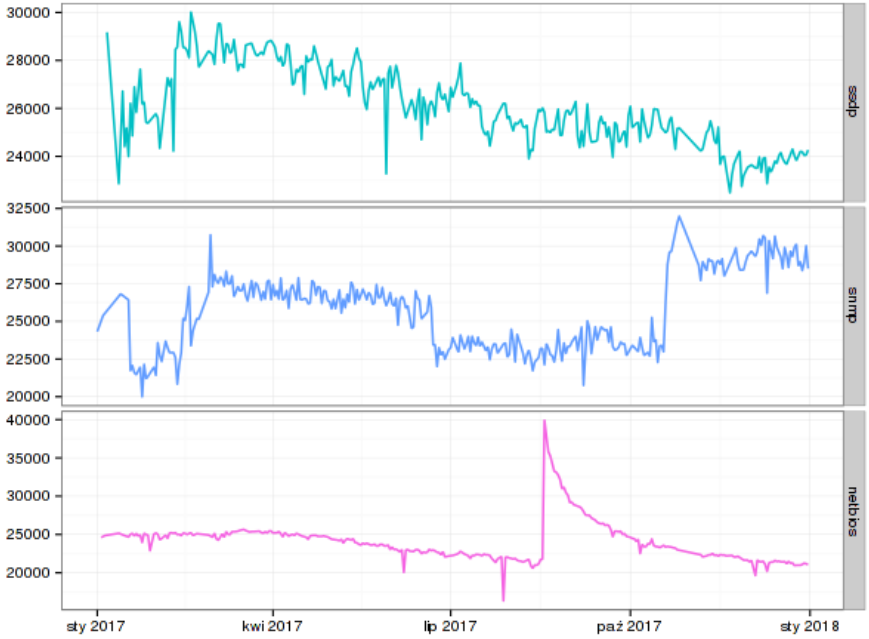
- Mostly nat/gov CSIRTs
- Rare feedback
- Limited uptake
- Some recipients have maintenance issues
- Feasible to have CSIRT-to-CSIRT exchange network?

Quantitative analysis: annual report

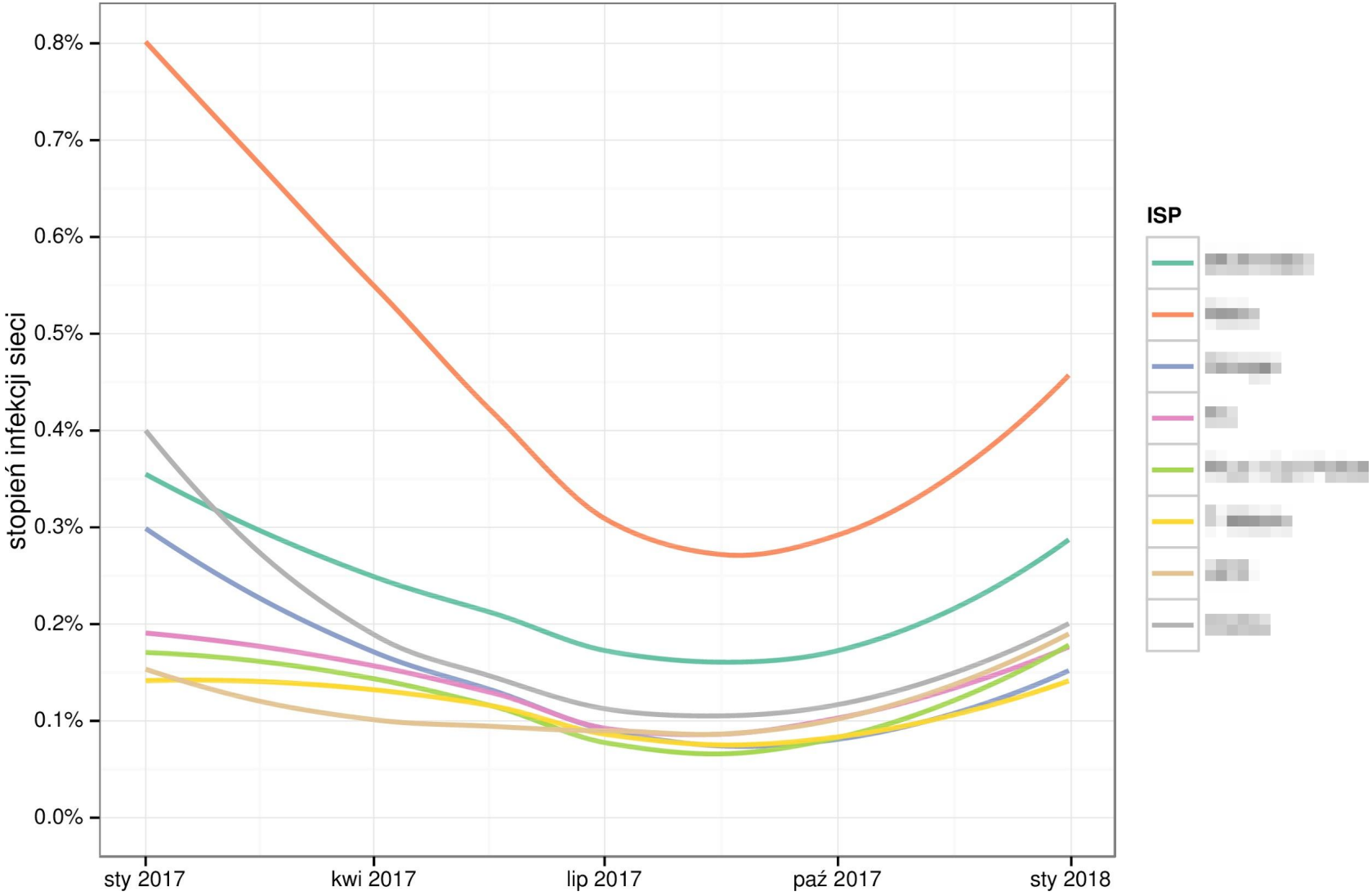
2017 soon ...

Annual report: amplifiers





Annual report: infection rates by ISP size



Quantitative analysis: conclusions

- In-depth analysis of collected data: possible to spot trends and anomalies
- Often not obvious how to use this knowledge
- Challenge: evaluation of data sources
  - quality of information
  - evaluation needs to be part of standard processes?
- Challenge: cross-comparable metrics
  - country- or ISP-level

Agenda

Future plans

- Prettier web interface (soon)
- Even more performant database schema
- Release of functionality missing from open source version
  - IntelMQ integration
  - notifier
  - stream API
  - additional collectors and parsers
- More enrichments
- Complete management functionality in web interface
- Provide metrics to clients (network health)
- Continuous quality evaluation

Discussion

- Do you have similar use-cases?
- What tools do you use?
- Can n6 automate some of your processes?
- What features should we add?
- Do you have/know good data sources to add?
- Other comments?
- Questions?

Opportunity to share ideas: IHAP

- Incident Handling Automation Project: informal dev/user group
- Mailing list & semi-annual meetings
- BoF session on Thursday (28.06), 18:00 – 19:00 @ Johor 2+5
- (see Additional programming section of the program)

Reading material: data processing and quality

- *Actionable Information for Security Incident Response*, 2014
  www.cert.pl/news/9684
- *Threat Intelligence: Collecting, Analysing, Evaluating*, 2015
  www.mwrinfosecurity.com/our-thinking/intelligent-threat-intelligence
- *Everything You Wanted to Know About Blacklists But Were Afraid to Ask*, Leigh Metcalf, Jonathan M. Spring, CERT / SEI, 2013
  resources.sei.cmu.edu/library/asset-view.cfm?assetid=83438
- *Paint it Black: Evaluating the Effectiveness of Malware Blacklists* Marc Kührer, Christian Rossow, Thorsten Holz, Ruhr-Universität Bochum, 2014
- NECOMA project, Deliverable 2.2: *Threat Analysis Platform,* Dataset rating, 2015
  www.necoma-project.eu

https://github.com/CERT-Polska/n6

Co-financed by the Connecting Europe Facility of the European Union