EVALUATE OR DIE TRYING

A METHODOLOGY FOR QUALITATIVE EVALUATION OF CYBER THREAT INTELLIGENCE FEEDS

Agenda

Problem Statement

Previous Work

Our Approach

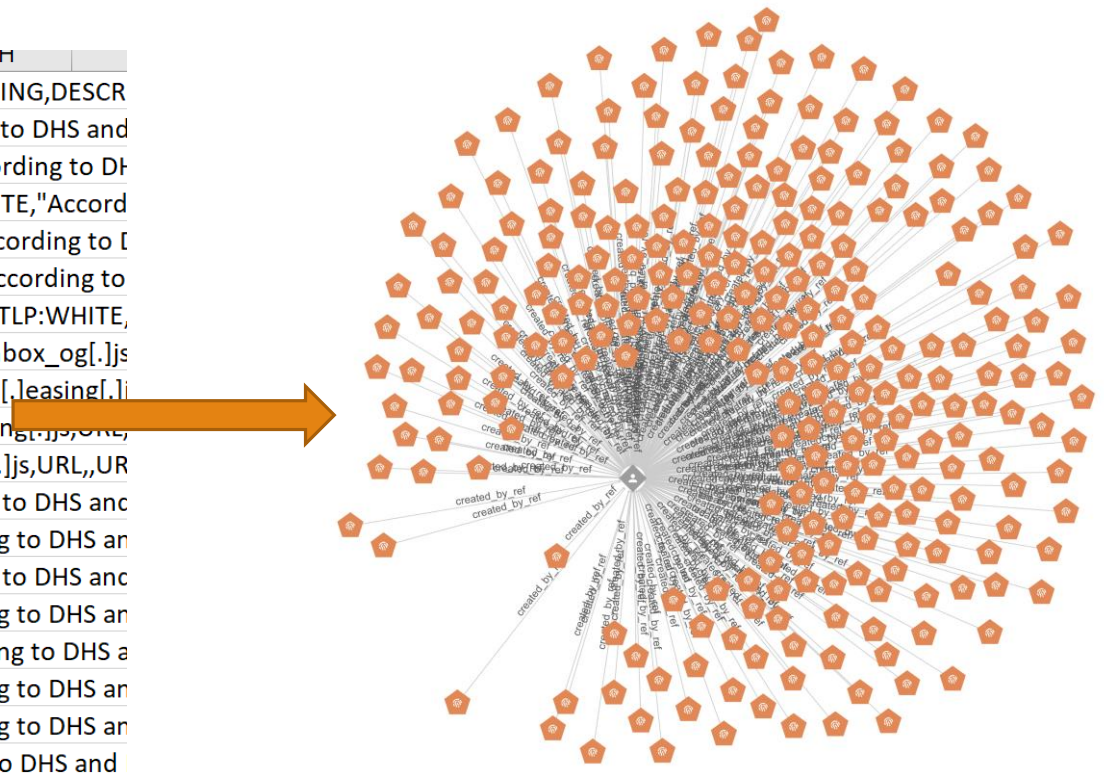
Metrics

So What?

Problem Statement

Evolution?

A	B	C	D	E	F	G	H
INDICATOR_VALUE,TYPE,COMMENT,ROLE,ATTACK_PHASE,OBSERVED_DATE,HANDLING,DESCR							
bit[.]ly/2m0x8IH,URL,,URL WATCHLIST,DELIVERY,2017-0						HITE,"According to DHS and	
tinyurl[.]com/h3sdqck,URL,,URL WATCHLIST,DELIVERY,2017-05-02,TLP:WHITE,"According to DH							
www[.]imageliners[.]com/nitel,URL,,URL WATCHLIST,DELIVERY,2017-05-02,TLP:WHITE,"Accord							
file://184[.]154[.]150[.]66/ame_icon[.]png,URL,,URL WATCHLIST,C2,,TLP:WHITE,"According to D							
hxxps://167[.]114[.]44[.]147/A56WY,URL,,URL WATCHLIST,DELIVERY,,TLP:WHITE,"According to							
hxxp://187[.]130[.]251[.]249/img/bson021[.]dat?0,URL,,URL WATCHLIST,DELIVERY,,TLP:WHITE,							
hxxp://www[.]oilandgaseng[.]com/fileadmin/templates/Redesign_2013_V2/js/loginbox_og[.]js							
hxxp://www[.]plantengineering[.]com/typo3conf/ext/t3s_jslidernews/res/js/jquery[.]easing[.]j							
hxxp://www[.]controleng[.]com/typo3conf/ext/t3s_jslidernews/res/js/jquery[.]easing[.]js,URL,							
hxxp://www[.]csemag[.]com/typo3conf/ext/t3s_jslidernews/res/js/jquery[.]easing[.]js,URL,,UR							
130[.]25[.]10[.]158,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS and							
167[.]114[.]44[.]147,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS an							
176[.]53[.]11[.]130,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS and							
184[.]154[.]150[.]66,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS an							
187[.]130[.]251[.]249,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS a							
193[.]213[.]49[.]115,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS an							
195[.]87[.]199[.]197,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS an							
2[.]229[.]10[.]193,IPV4ADDR,,IP_WATCHLIST,C2,2017-03-02,TLP:WHITE,"According to DHS and							



CTI in Security Operations

More organizations are consuming CTI, especially in the form of finalized intelligence reports, and integrating them into their defensive mechanisms. **Operationalizing narrative-based intelligence reports—reports that describe in detail a series of events related to an intrusion or incident—is time-consuming for CTI analysts. A lack of automation for these reports makes them especially time-consuming. CTI teams need to ensure that they are properly staffed and allocating enough time to make the best use of this type of reporting.**

The Evolution of Cyber Threat Intelligence (CTI): 2019 SANS CTI Survey
https://www.sans.org/reading-room/whitepapers/threats/paper/38790

Threat Intelligence Fatigue

- Organizations tend to obtain as much information as possible
- Sources not meeting **intelligence** and **production** requirements
- Customer cannot judge the **quality** of an intel feed
- Unknown business value.
- How to justify expenditures for intelligence sources?

Previous Work

Previous Work

Measuring the IQ of your Threat Intelligence
Alexandre Pinto, Kyle Maxwell, DEFCON 22, August 2014

Data-Driven Threat Intelligence:
Useful Methods and Measurements for Handling Indicators
Alexandre Pinto, Alexandre Sieira, FIRST Conference 2015, June 2015

Evaluating Threat Intelligence Feeds
Paweł Pawlinski, Andrew Kompanek, FIRST Technical Colloquium for Threat Intelligence Munich, 2016

This ↑ is still a must. Our work is NOT a replacement, but should co-exist with earlier work.

Our Approach

Our Approach

- We use STIX 2.0 as common format for comparison

- Ingest native STIX 2.0 feeds
- Convert existing STIX 1.2 feed into STIX 2.0
- Convert source specific JSON into STIX 2.0

- Store STIX 2.0 data in PostgreSQL DB
- Use Jupyter notebook for analysis

Consideration

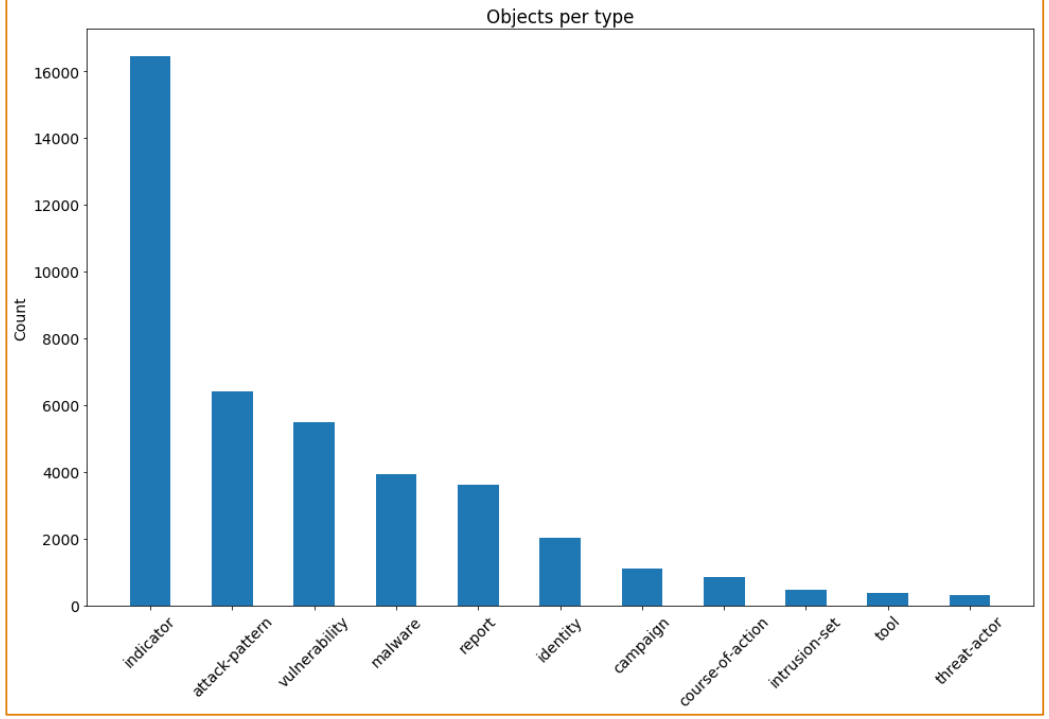
- We are looking at the feed of decent size (difficult to eye-ball)
- The feeds are updated daily, append-only.
- Mix of open and commercial sources
- We focus on STIX 2.0 objects (one feed contained STIX 2.1 entities)
- Convert existing STIX 1.2 / JSON feeds into STIX 2.0 with best effort

Metrics

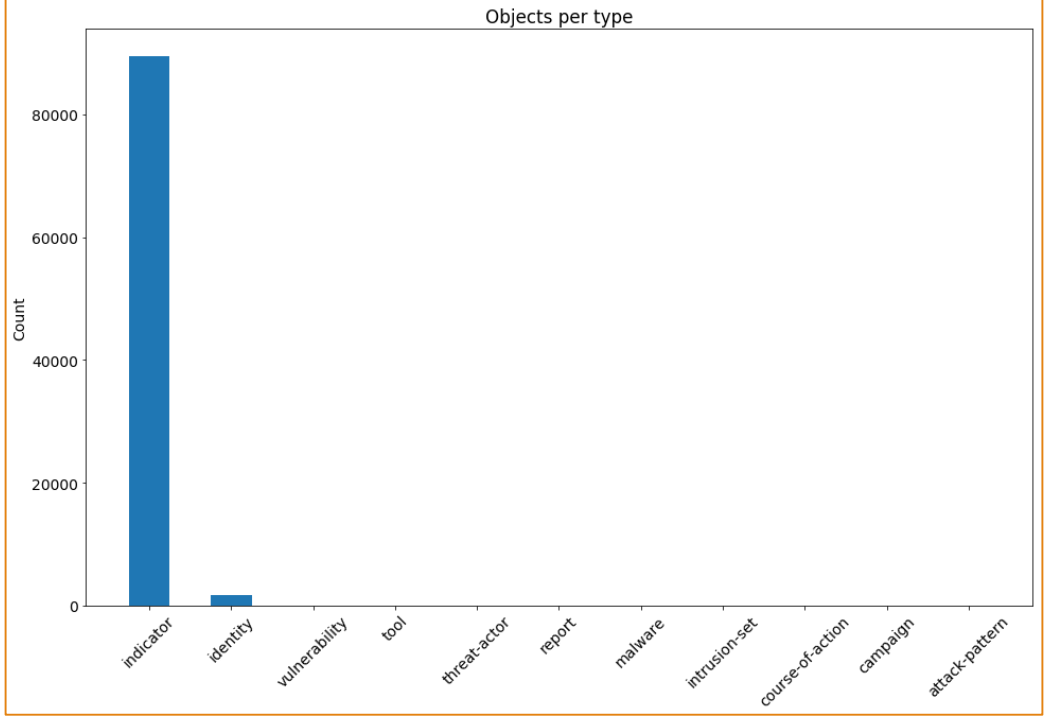
Objects & observables

Object Type Variability

Source A



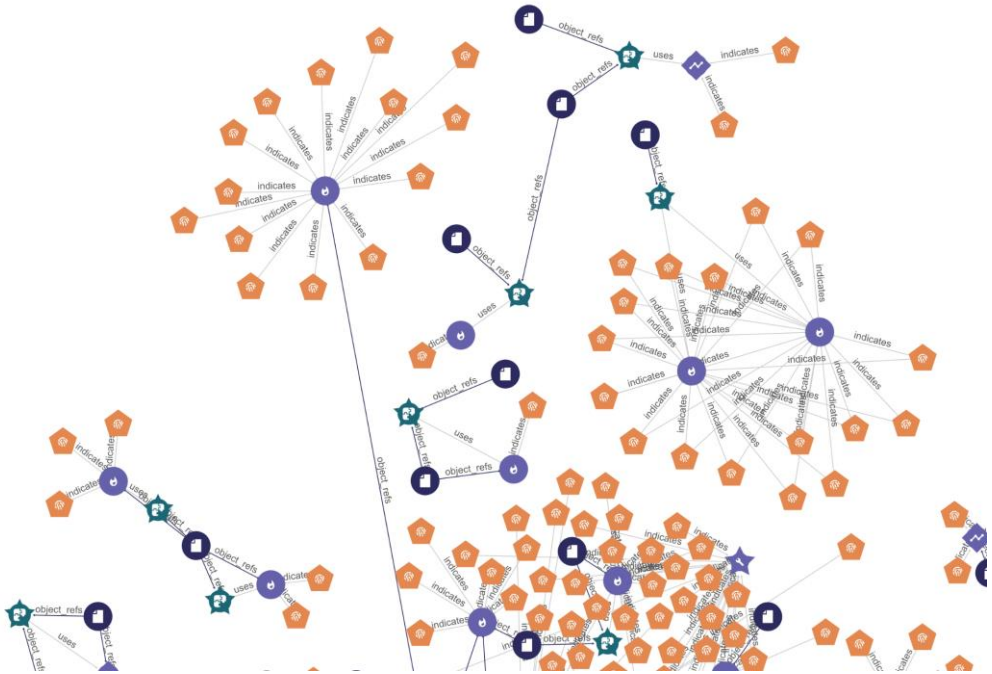
Source B



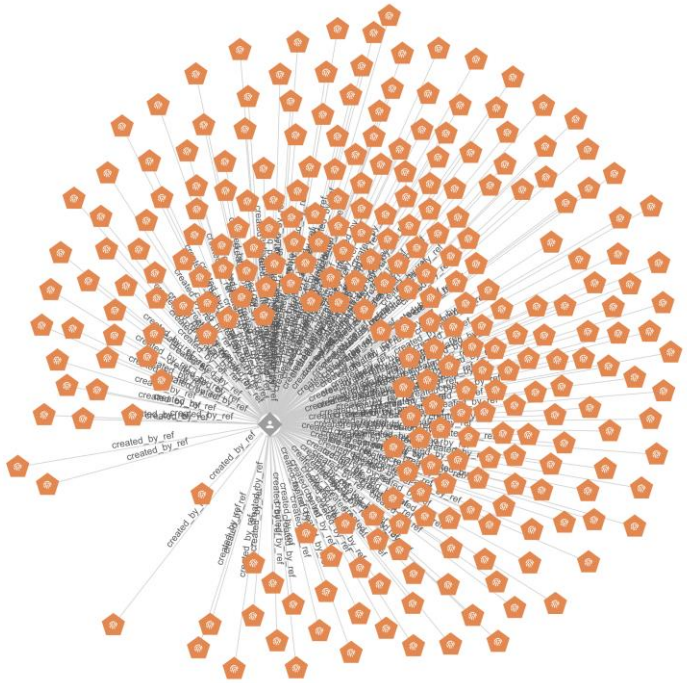
- Do these object types align with my needs?
- Is the feed balanced or is it heavily skewed to one particular object type?
- Are there custom STIX2 objects that might cause ingestion issues?

Object Type Variability

Source A



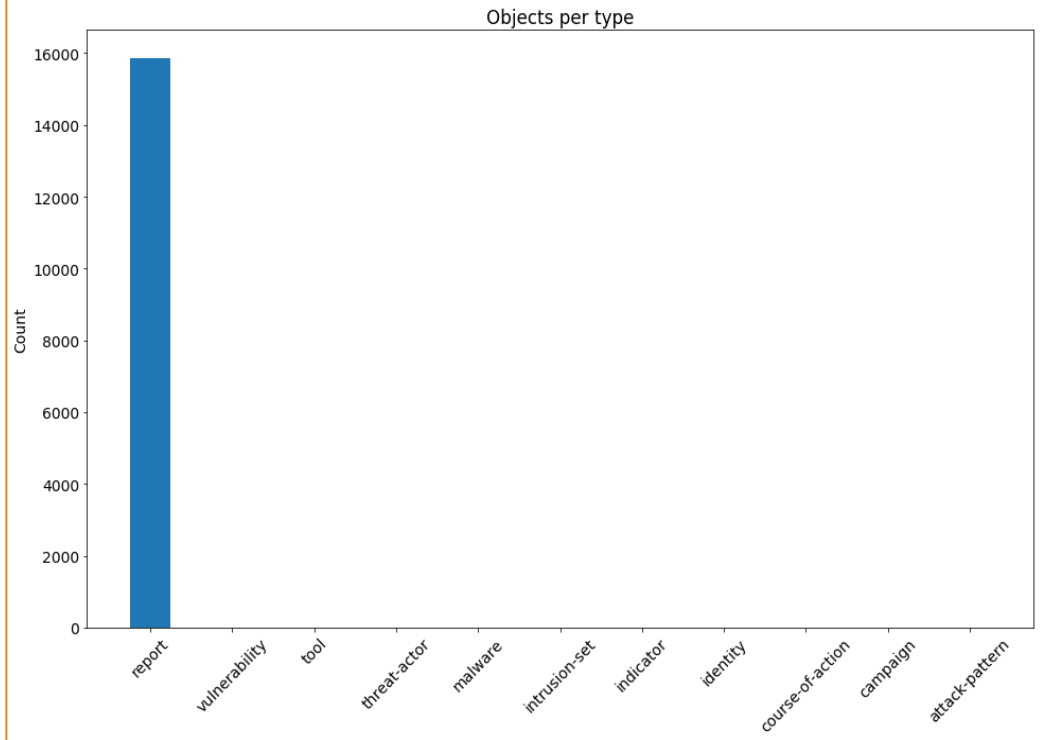
Source B



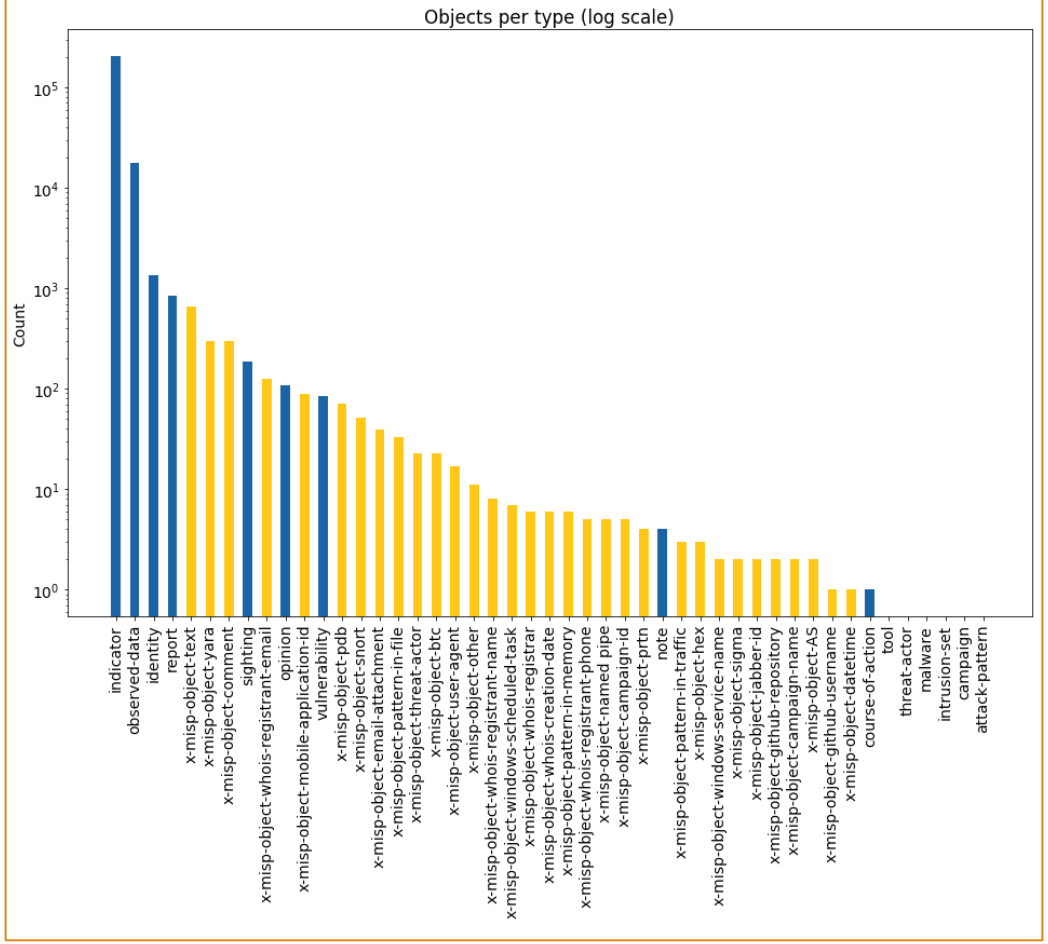
- Do these object types align with my needs?
- Is the feed balanced or is it heavily skewed to one particular object type?
- Are there custom STIX2 objects that might cause ingestion issues?

Object Type Variability

Source C

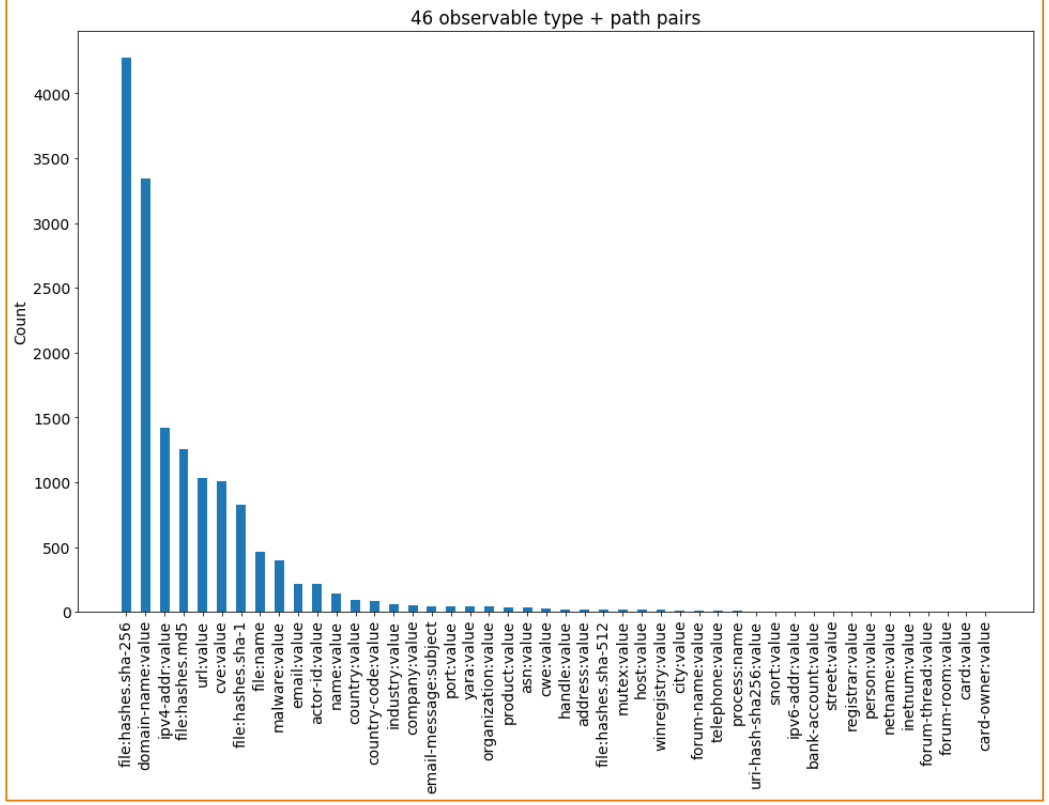


Source F

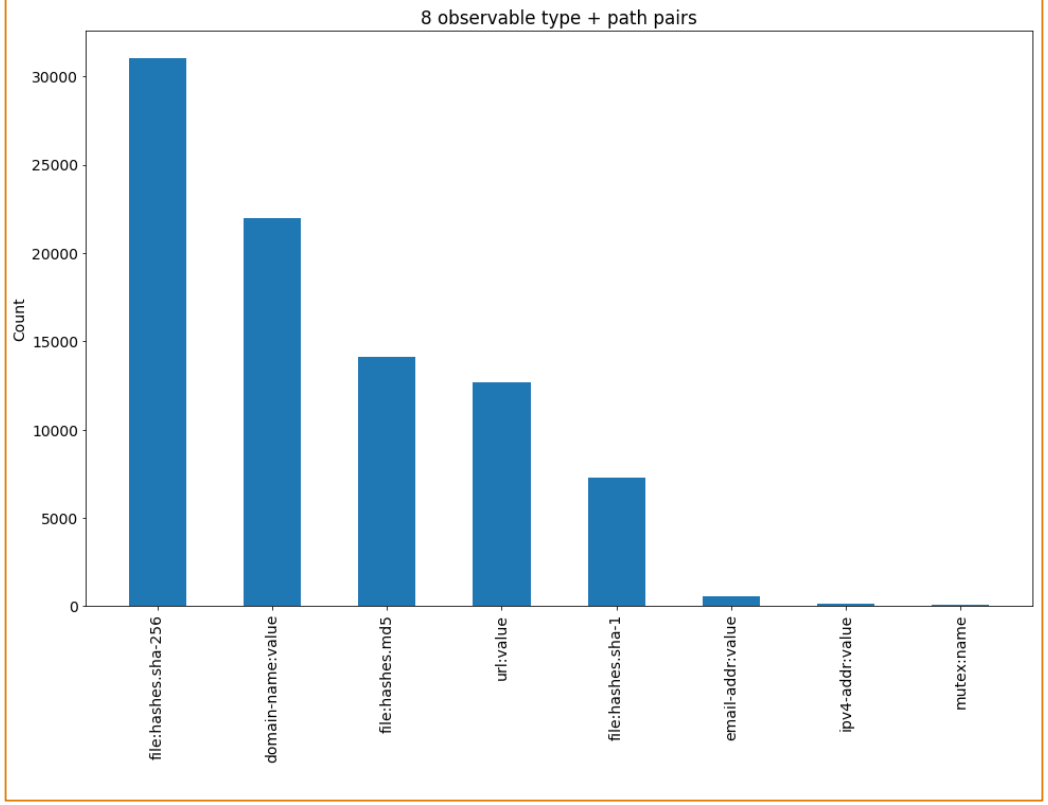


Observables Variability

Source A



Source B



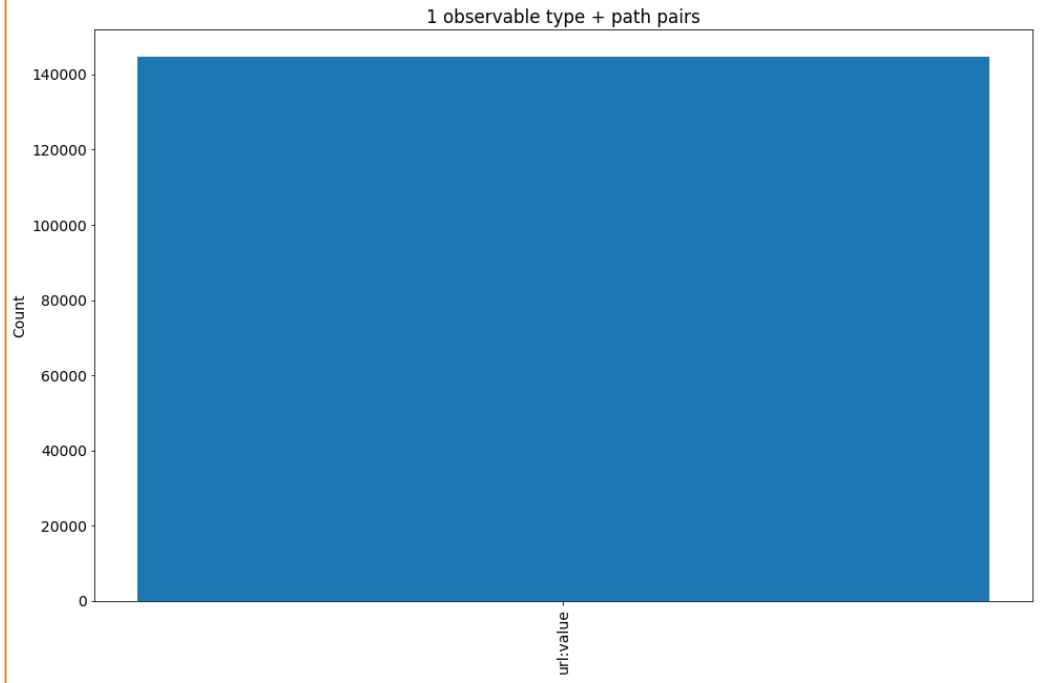
- Do these observable types align with my needs?
- Is feed balanced or is it heavily skewed to one particular observable type?

Observables Variability

Source C

No Observables

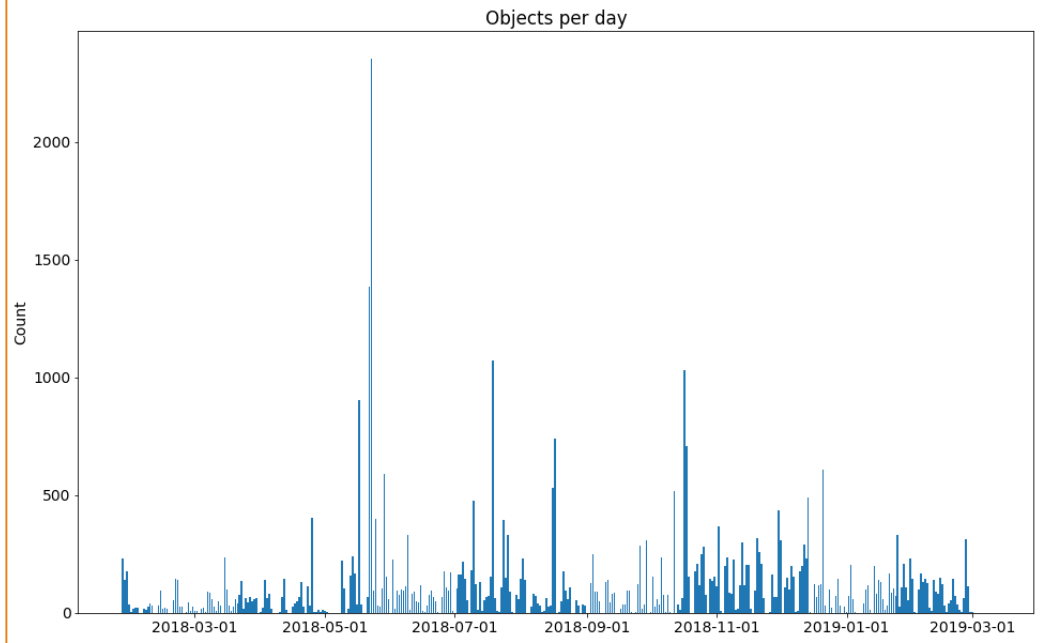
Source D



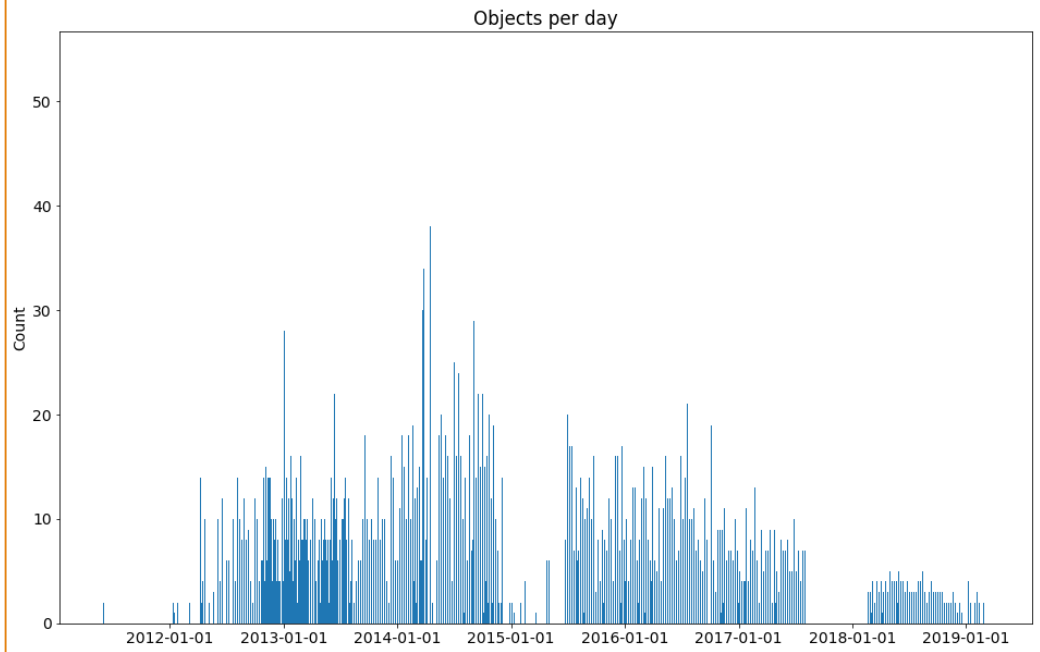
- Do these observable types align with my needs?
- Is feed balanced or is it heavily skewed to one particular observable type?

Timeframe & Gaps

Source A



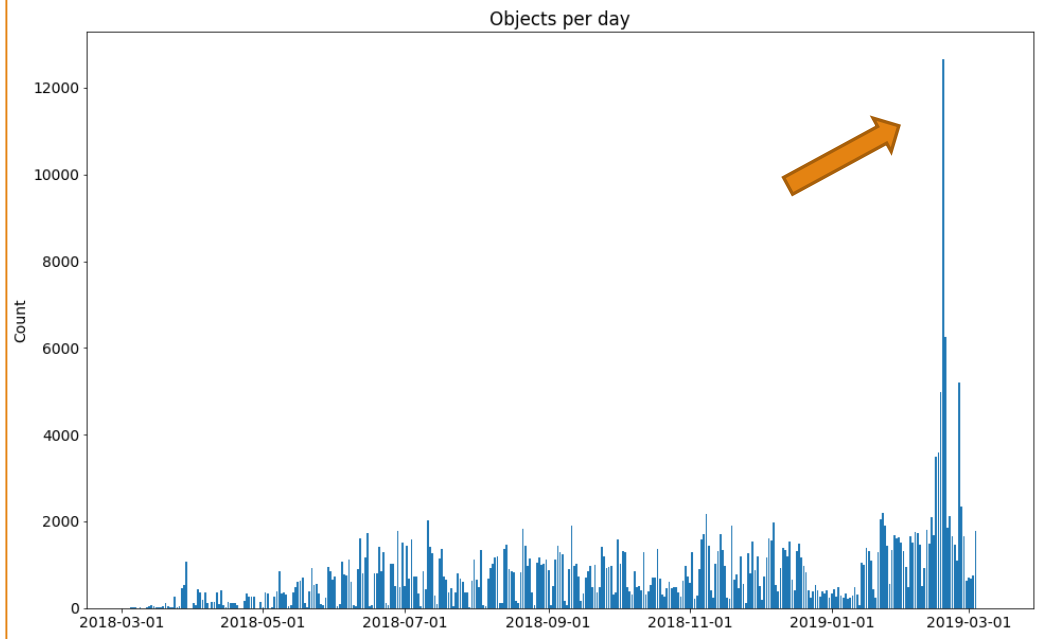
Source C



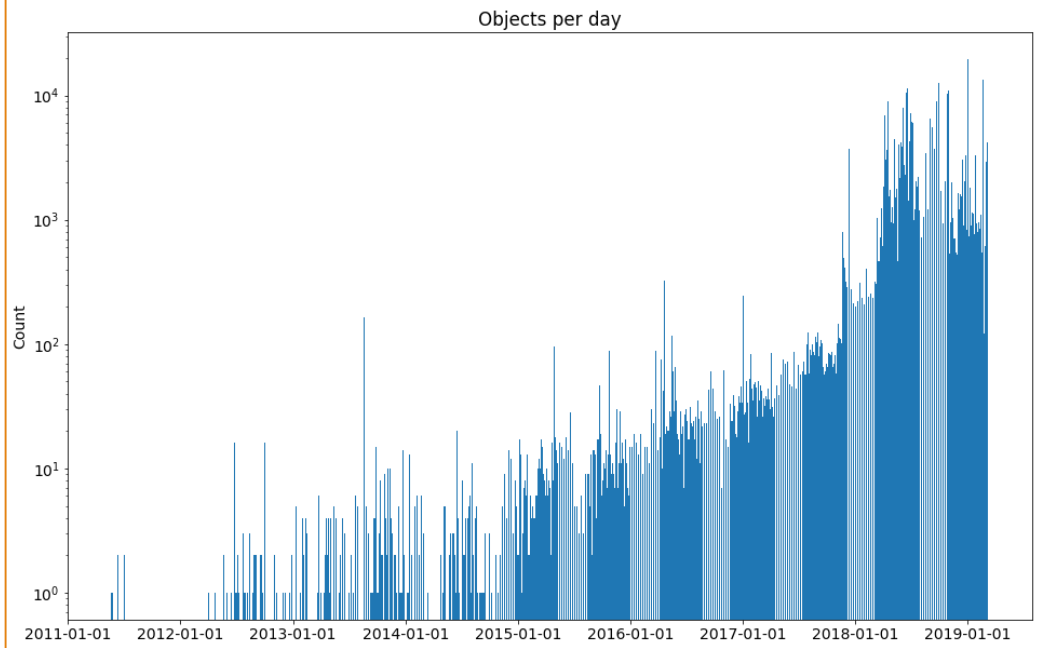
- Does the source contain enough historical data?
- Are there significant gaps in the dataset?
- Is daily data influx consistent over long period of time?

Timeframe & Gaps

Source D



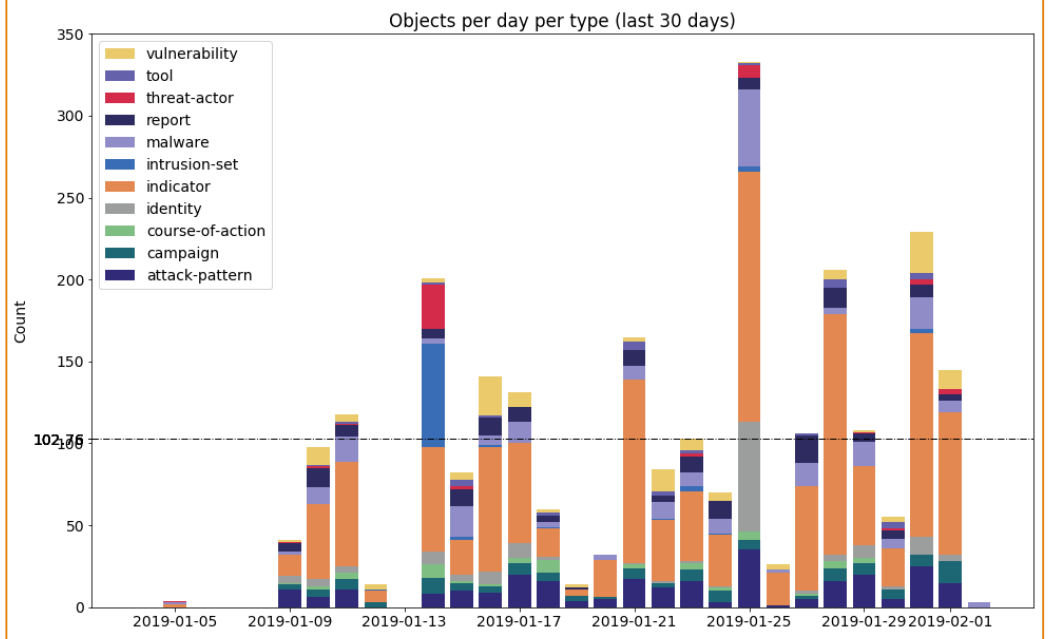
Source E



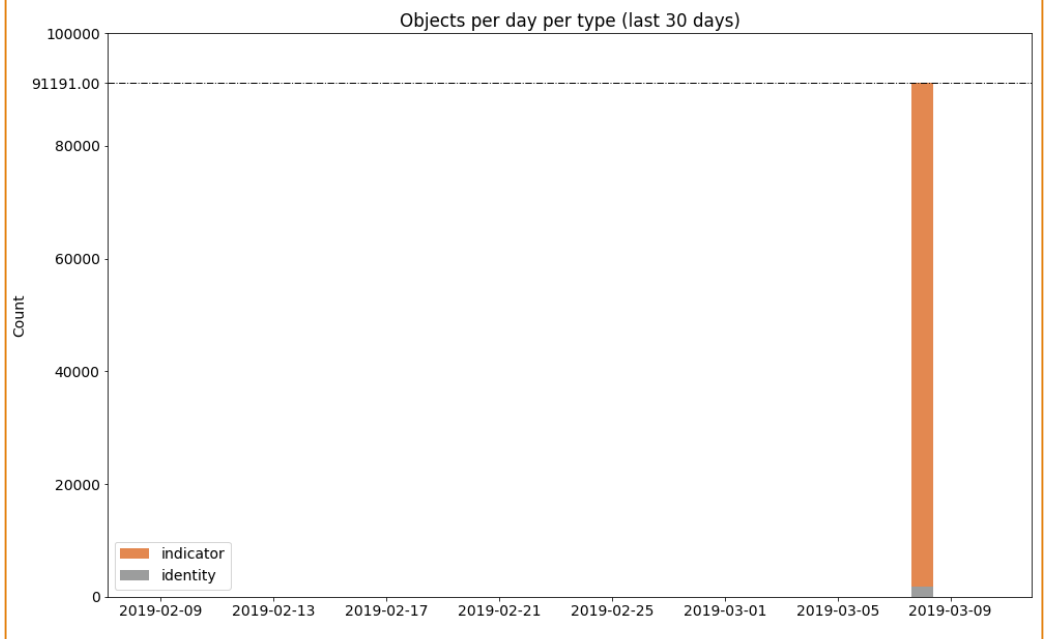
- Does the source contain enough historical data?
- Are there significant gaps in the dataset?
- Is daily data influx consistent over long period of time?

Influx

Source A



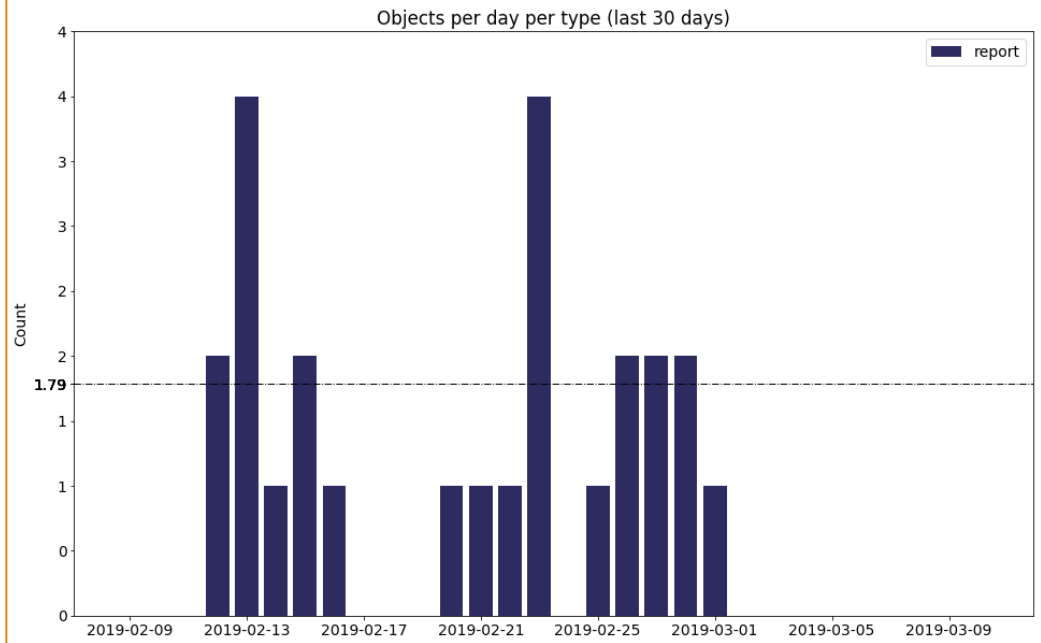
Source B



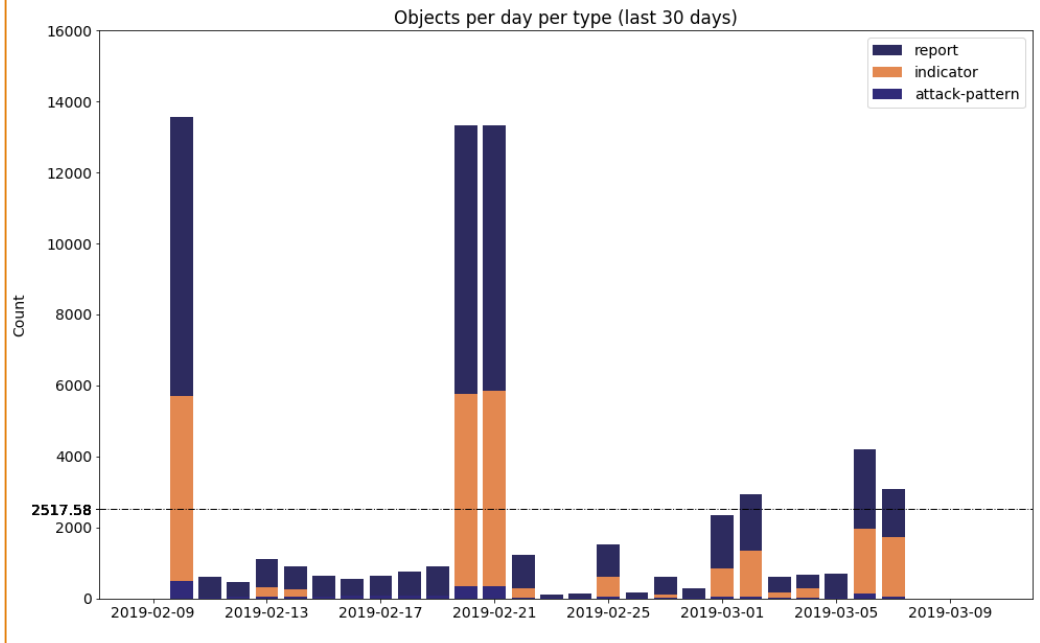
- What is the daily average for the last 30 days?
- Does the feed contain spikes that can cause performance issues during ingestion?
- Is the feed balanced across object types or is it skewed to one particular object type?

Influx

Source C



Source E



- What is the daily average for the last 30 days?
- Does the feed contain spikes that can cause performance issues during ingestion?
- Is the feed balanced across object types or is it skewed to one particular object type?

Fullness

2.4.1 Properties

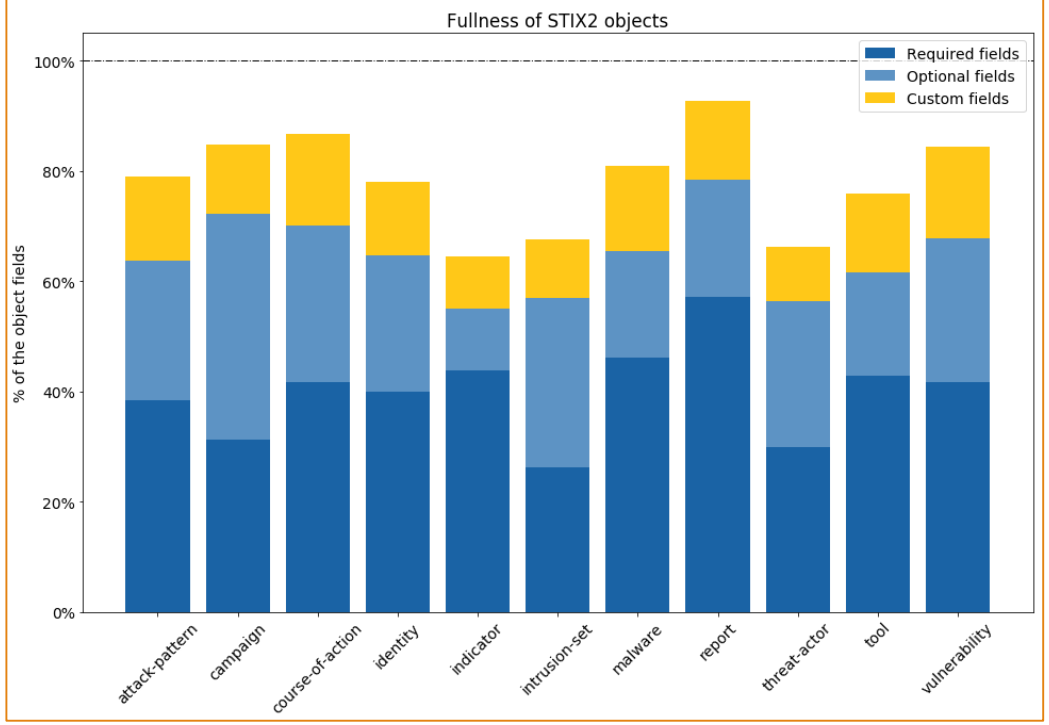
Common Properties
`type`, `id`, `created_by_ref`, `created`, `modified`, `revoked`, `labels`, `external_references`, `object_marking_refs`, `granular_markings`
Identity Specific Properties
`name`, `description`, `identity_class`, `sectors`, `contact_information`

Property Name	Type	Description
`type` (required)	`string`	The value of this property **MUST** be `identity`.
`labels` (optional)	`list` of type `string`	The list of roles that this Identity performs (e.g., CEO, Domain Administrators, Doctors, Hospital, or Retailer). No open vocabulary is yet defined for this property.
`name` (required)	`string`	The name of this Identity. When referring to a specific entity (e.g., an individual or organization), this property **SHOULD** contain the canonical name of the specific entity.
`description` (optional)	`string`	A description that provides more details and context about the Identity, potentially including its purpose and its key characteristics.
`identity_class` (required)	`open-vocab`	The type of entity that this Identity describes, e.g., an individual or organization. This is an open vocabulary and the values **SHOULD** come from the `identity-class-ov` vocabulary.
`sectors` (optional)	`list` of type `open-vocab`	The list of industry sectors that this Identity belongs to.

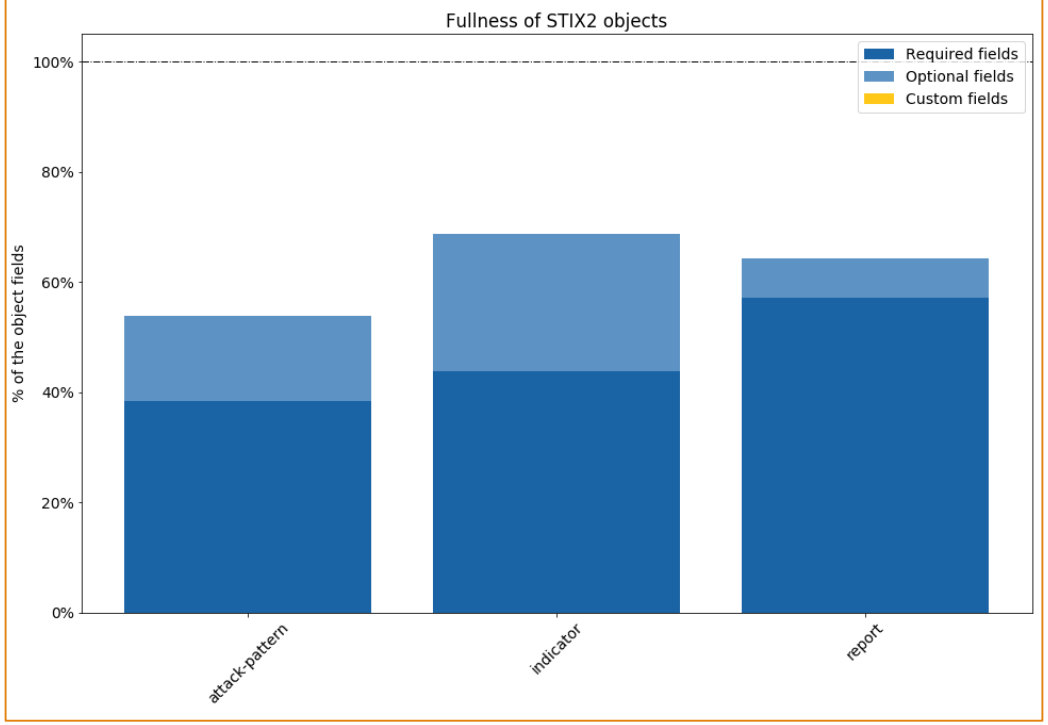
STIX™ Version 2.0. Part 2: STIX Objects
https://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part2-stix-objects.html

Fullness

Source A



Source F



- Does the source leverage optional fields or does it provide minimum context only?
- Does the source implement custom fields?

Relationships

Relationship by Type

Relationships are not restricted to those listed below. Relationships can be created between any objects using the `related-to` relationship type or, as with open vocabularies, user-defined names.

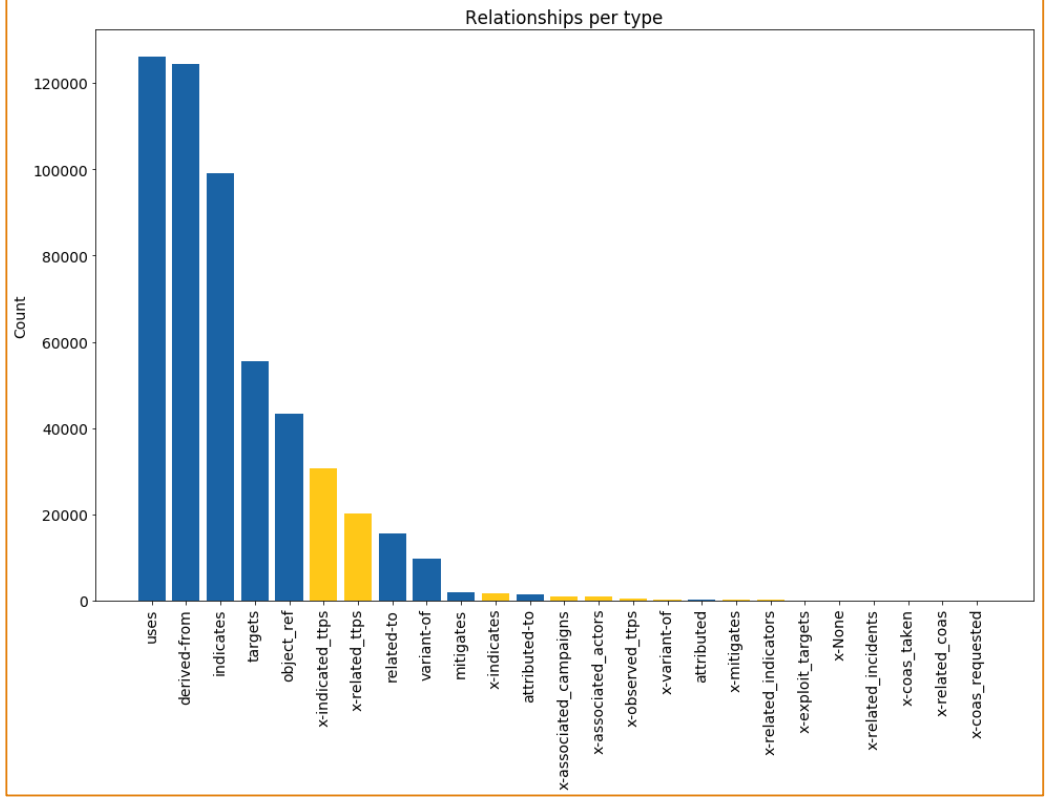
Embedded Relationships	
`created_by_ref`	`identifier` (of type `identity`)
`object_marking_refs`	`identifier` (of type `marking-definition`)
Common Relationships	
`duplicate-of`, `derived-from`, `related-to`	

Source	Relationship Type	Target	Description
—	—	—	—
Reverse Relationships			
`attack-pattern`, `campaign`, `intrusion-set`, `malware`, `threat-actor`, `tool`	`targets`	`vulnerability`	See forward relationship for definition.

STIX™ Version 2.0. Part 2: STIX Objects
https://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part2-stix-objects.html

Relationship by Type

Source A



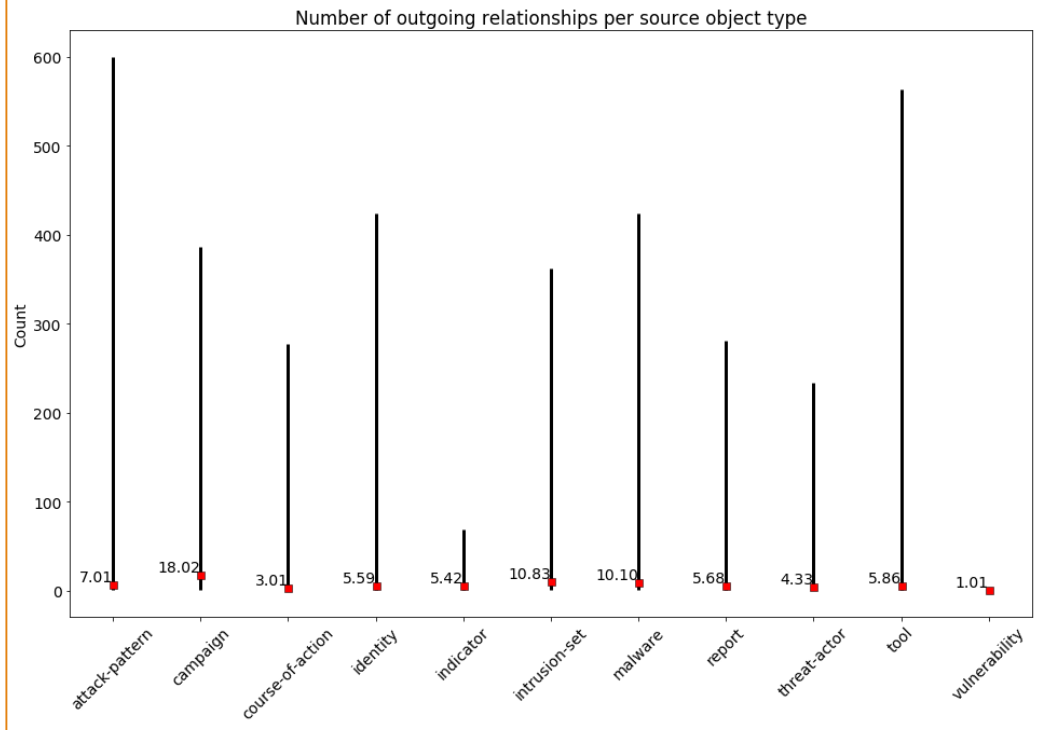
Source B

No Relationships

- Does the source use custom relations? Hint at unconventional data model.
- Custom relation types might also cause integration during ingestion.

No. of Outgoing Relationship

Source A (Log)



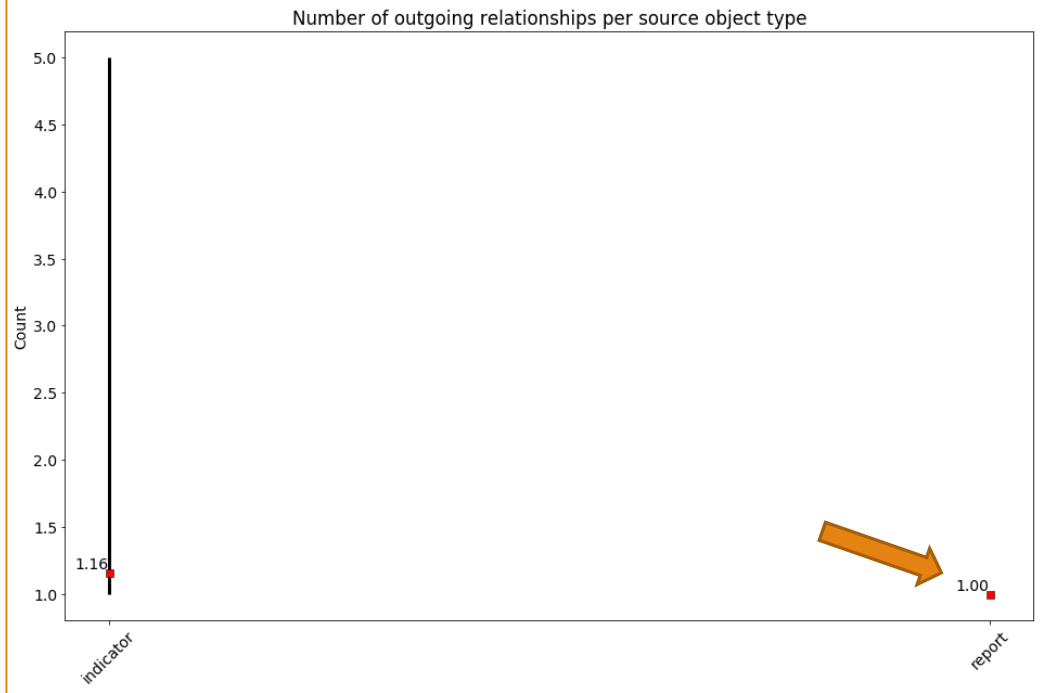
Source B

No Relationships

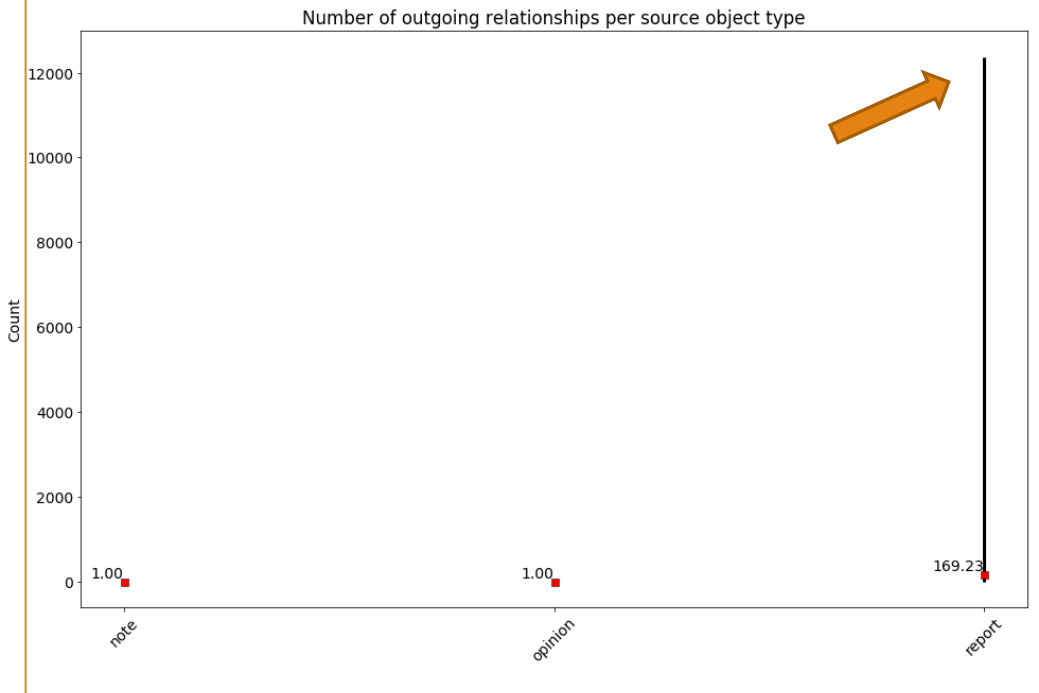
- Does the dataset have objects with unreasonable number of outgoing relations?
- This might be a symptom of a poor data model and might cause issues during ingestion.

No. of Outgoing Relationship

Source E



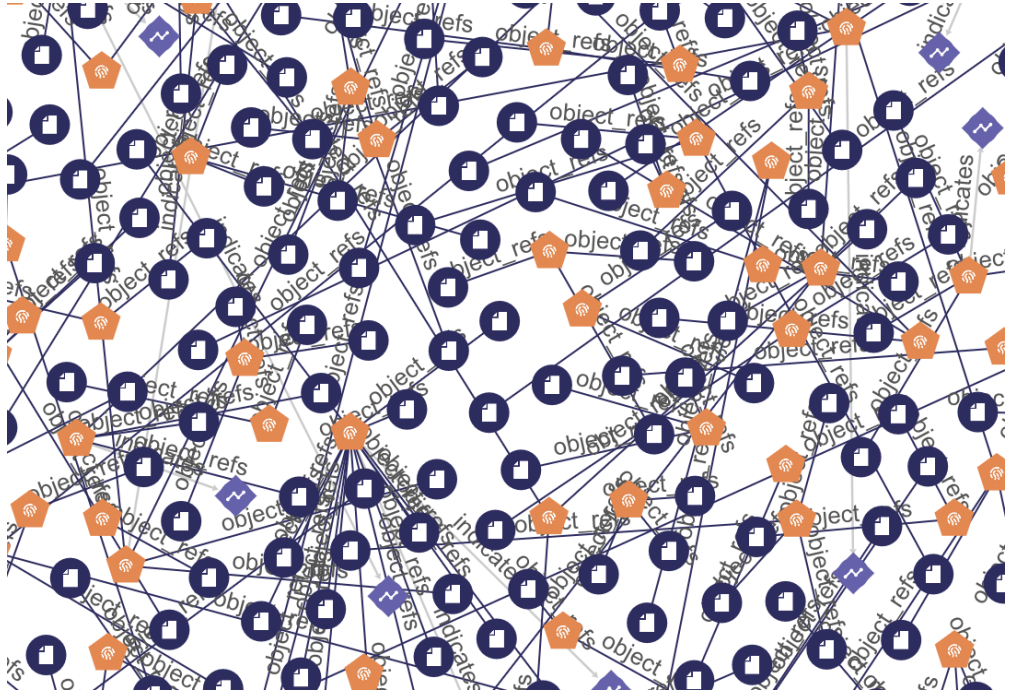
Source F



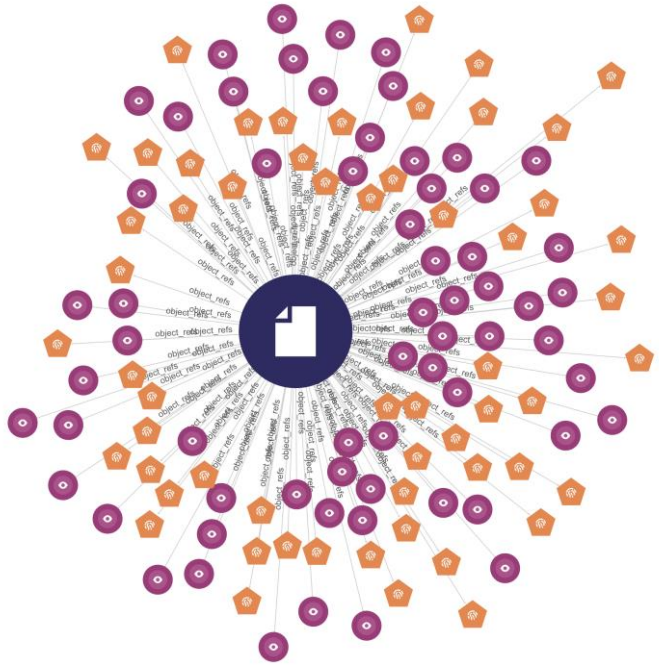
- Does the dataset have objects with unreasonable number of outgoing relations?
- This might be a symptom of a poor data model and might cause issues during ingestion.

No. of Outgoing Relationship

Source E



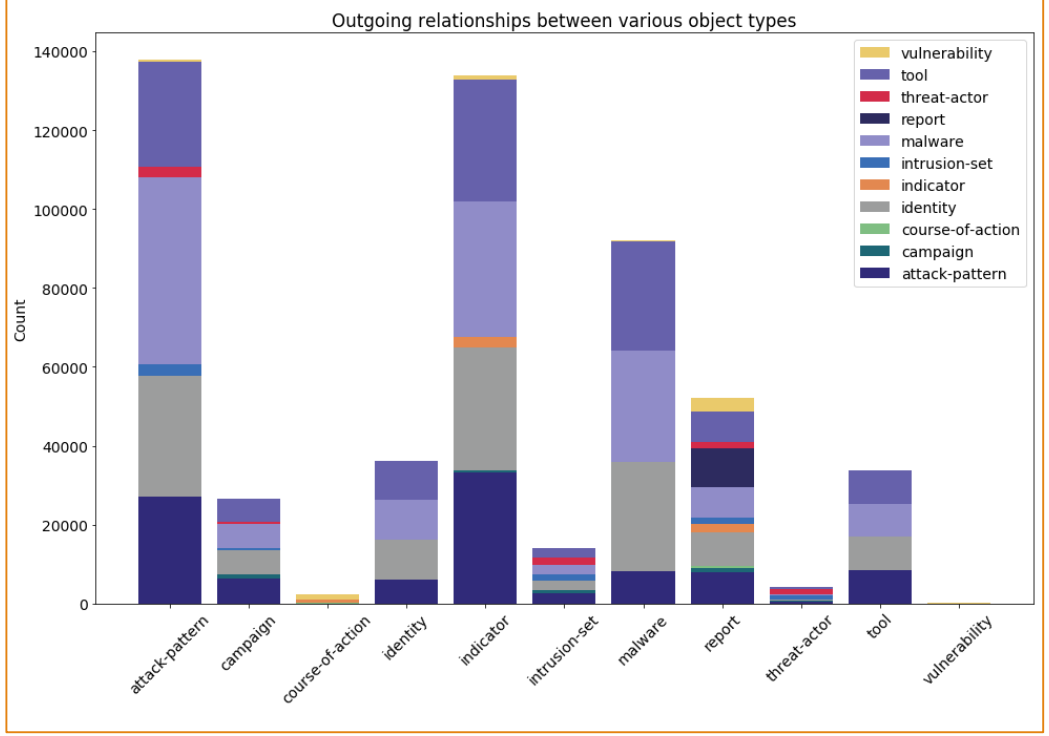
Source F



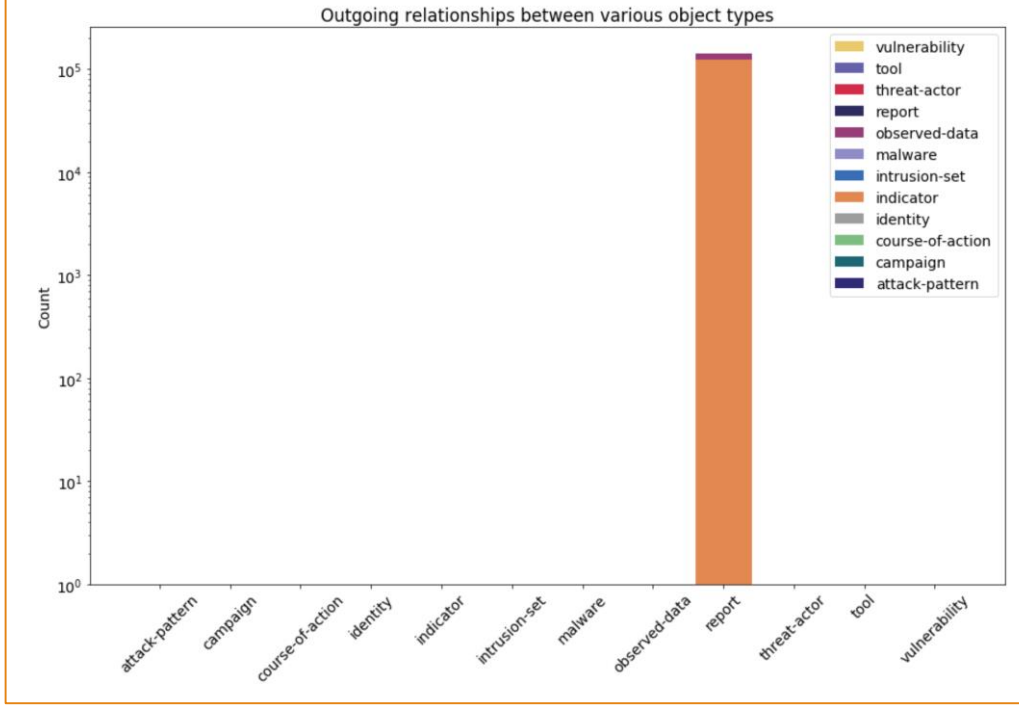
- Does the dataset have objects with unreasonable number of outgoing relations?
- This might be a symptom of a poor data model and might cause issues during ingestion.

Outgoing Relationship Between Objects

Source A



Source F

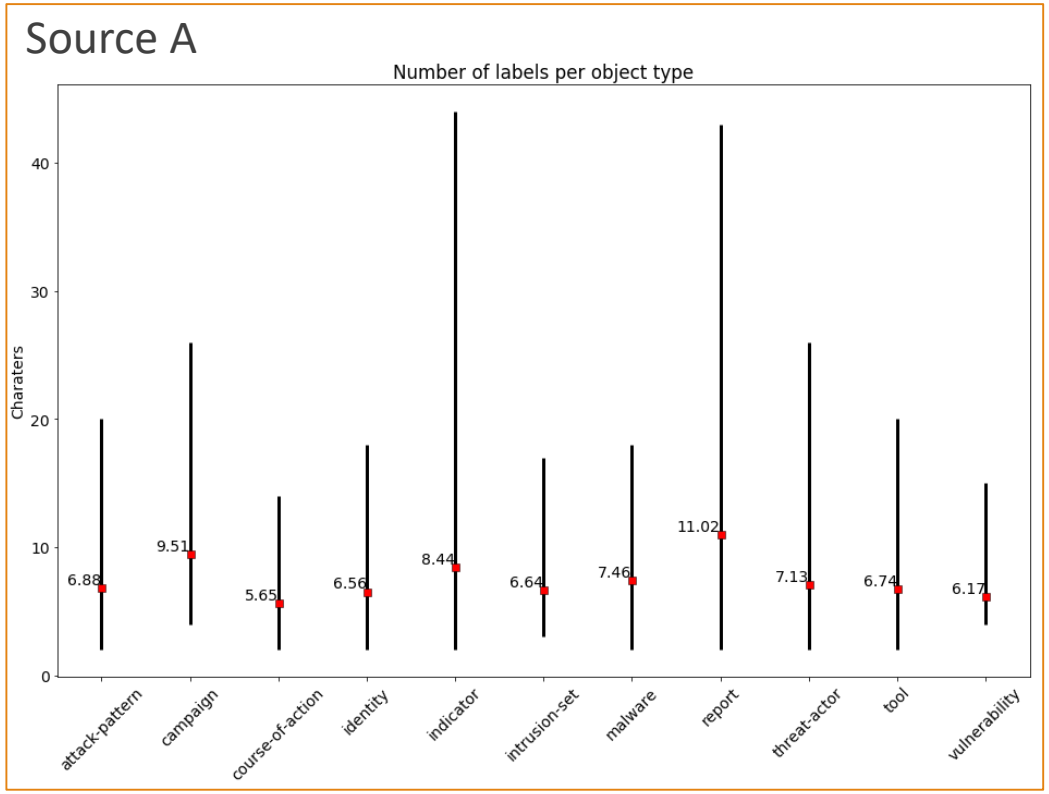


- How connected is the dataset?
- What data model the dataset has?

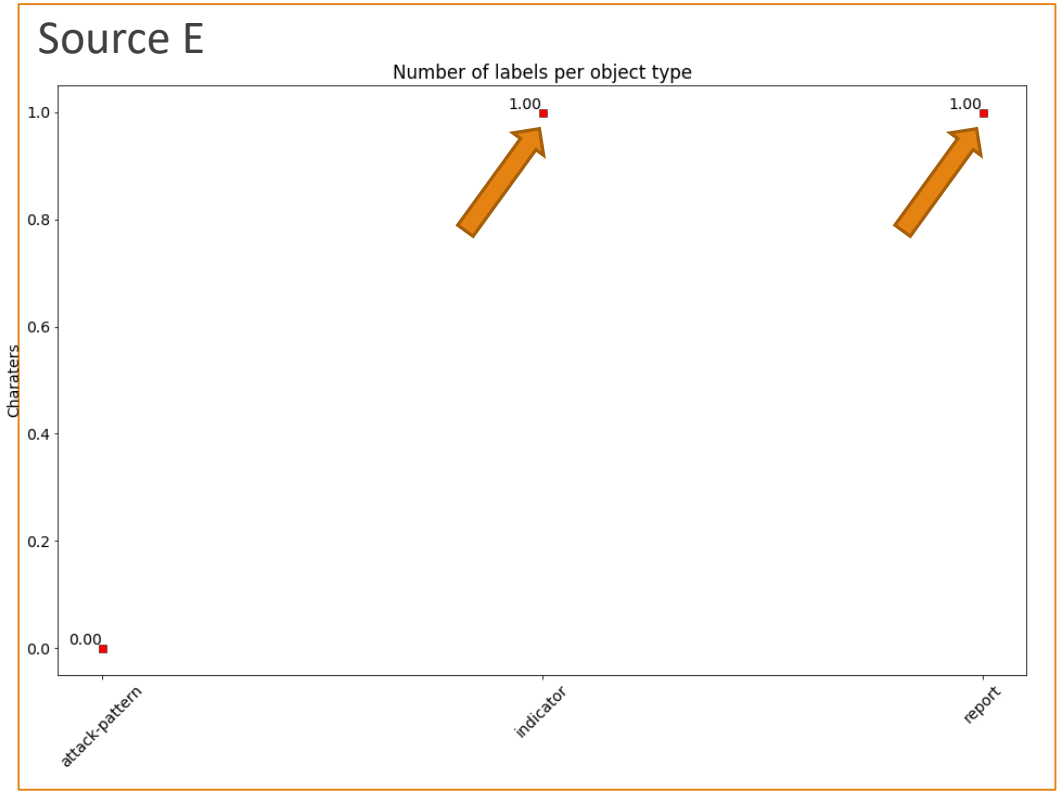
Content

No. of Labels per Object Type

Source A



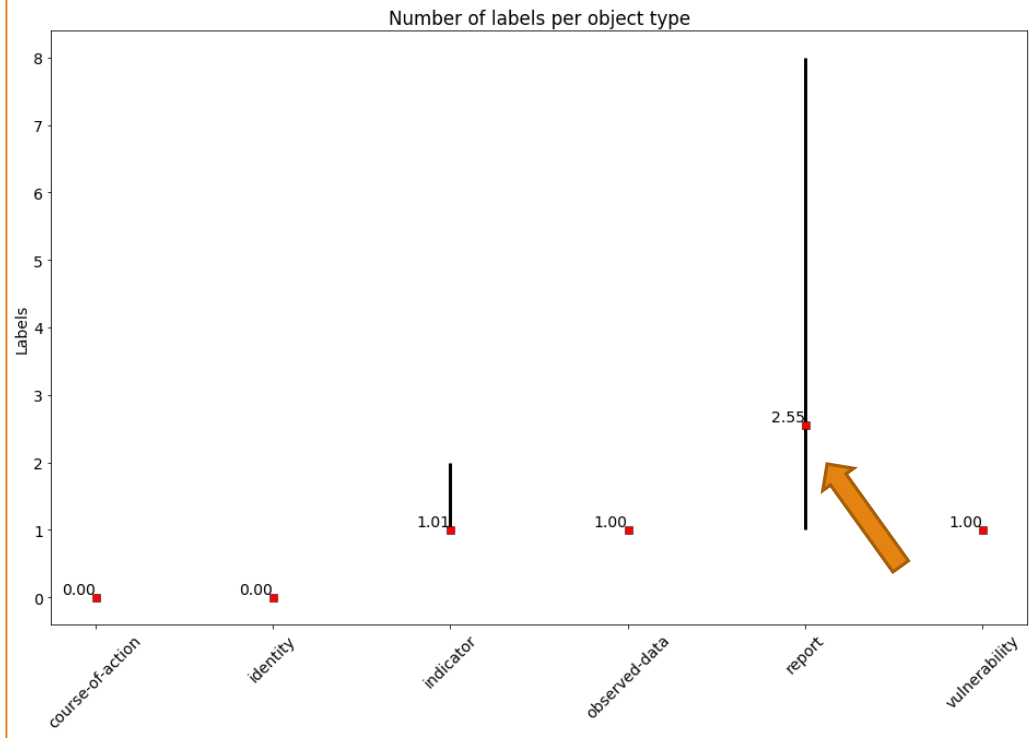
Source E



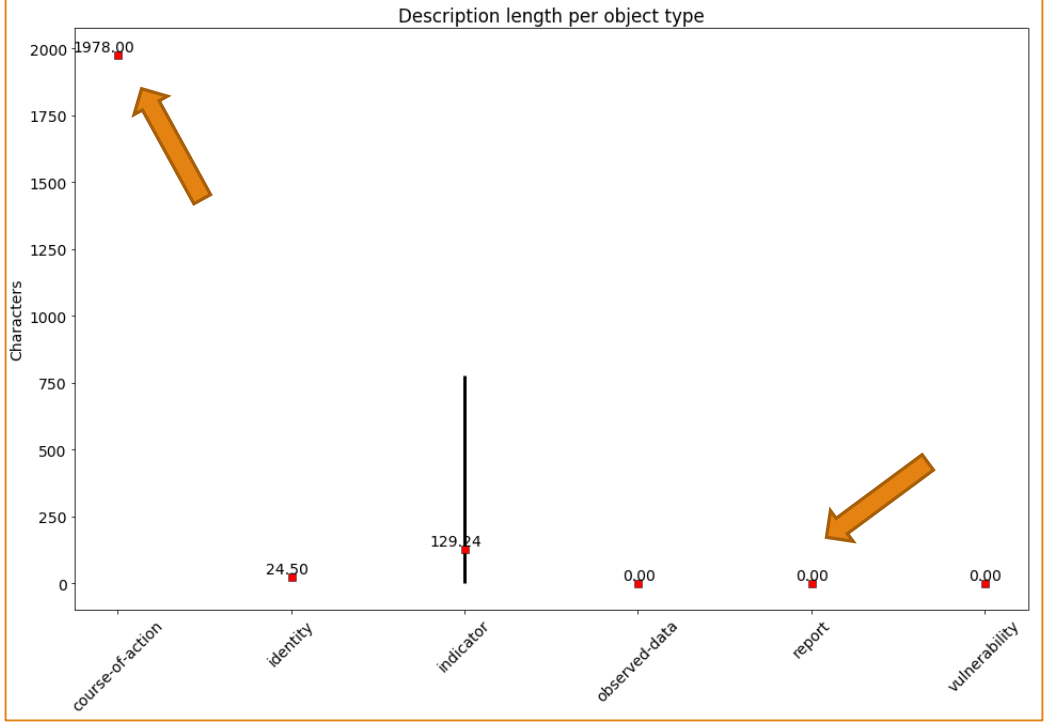
- How well is data labeled?

Quality of the content: labels vs description

Source F - No. of Label per Object Type



Source F – Description Length per Object Type



Source F

- Reports: not having any description length, but labeled on average with 2.55 labels
- CoA: an average description length of 1978 chars (with no variance), but no label?

Metrics - Full List

- Object types
- Observable types
- Time frame & gaps
- Objects per day per type
- Fullness
- Relationships by type
- No. of outgoing relationships
- Outgoing relationships between various object types
- Number of incoming relationships per target object type
- Incoming relationships between various object types
- No. of hanging or detached relationships
- Description length per object type
- Number of labels per object type
- Objects per TLP
- Unique and re-used observables
- Observables overlap between feeds
- Relevancy / Proximity

Metrics - Example

Metric	Weighting	Source A		Source B		Source C		Source D		Source E		Source F	
		Points	Score	Points	Score	Points	Score	Points	Score	Points	Score	Points	Score
Entity Variability	1,50	5	7,5	1	1,5	1	1,5	2	3	3	4,5	5	7,5
Observables Variability	1,20	4	4,8	5	6	0	0	1	2	1	1,2	4	4,8
Time Frame & Gaps	1,10	2	2,2	2	2,2	5	5,5	2	2	4	4,4	3	3,3
Influx per Day	1,00	5	5	1	1	1	1	2	2	3	3	2	2
Entity Thickness / Completeness	1,20	3	3,6	1	1,2	1	1,2	2	6	2	2,4	2	2,4
Relationship by Type	1,10	4	4,4	0	0	0	0	1	2	1	1,1	2	2,2
No. of Outgoing Relationship	1,00	2	2	1	1	1	1	1	1	1	1	1	1
Proximity	1,50	1	1,5	0	0	0	0	0	0	1	1,5	1	1,5
	Totals		**31**		**12,9**		**10,2**		**18**		**19,1**		**24,7**

Observations & Lessons Learned

- “Results produced by the stix2-elevator are not for production purposes”
  - python libraries used for STIX2 transformation require a lot of hand holding
- Some feeds can not be easily converted to STIX2.0 because of feed / spec limitations:
  - UUID4-only IDs
  - Reports must have `object_refs` field set
  - Indicators must have a pattern
- There are few STIX2.0 sources available (for now), feed providers are taking their time.
- Feed evaluation is a multi-step process of analyzing feed characteristics from intelligence requirements perspective.

So What?

Situation Today

Consumer Profile

Super Bank Inc.

Russia

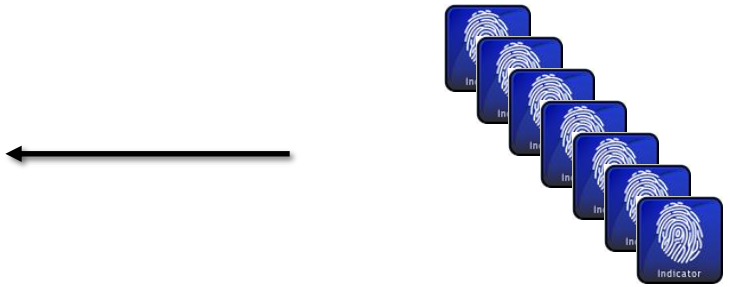
Sector: Finance

FlawedAmmyy

FastCash

RDP

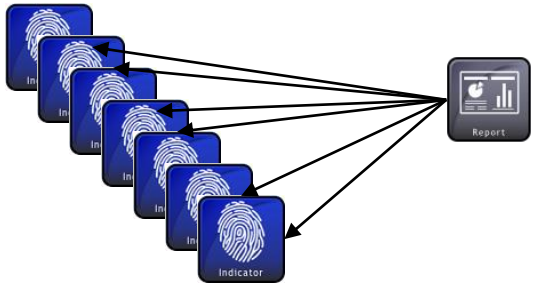
VNC



Source A
no structure

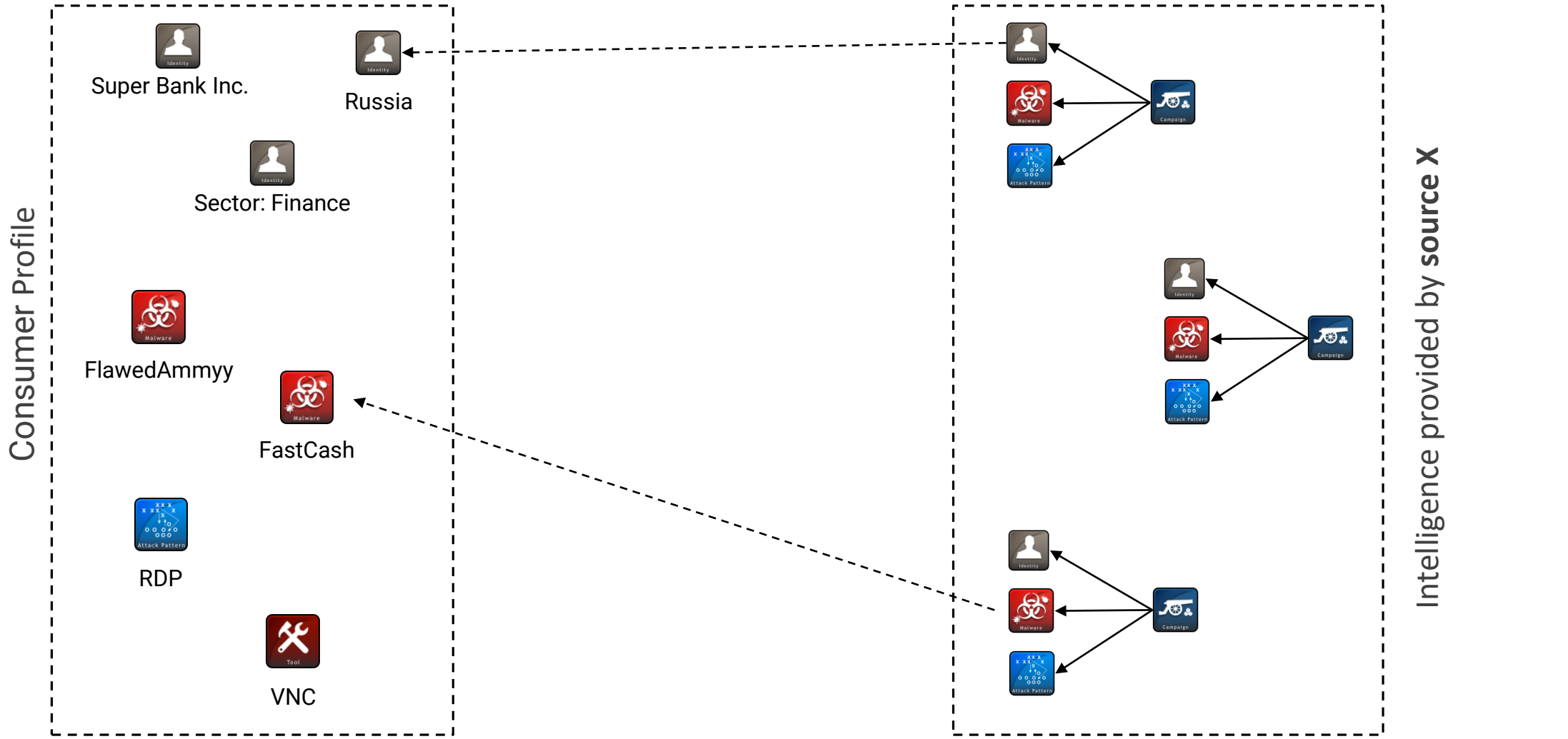
C2

Vendor B
poor structure

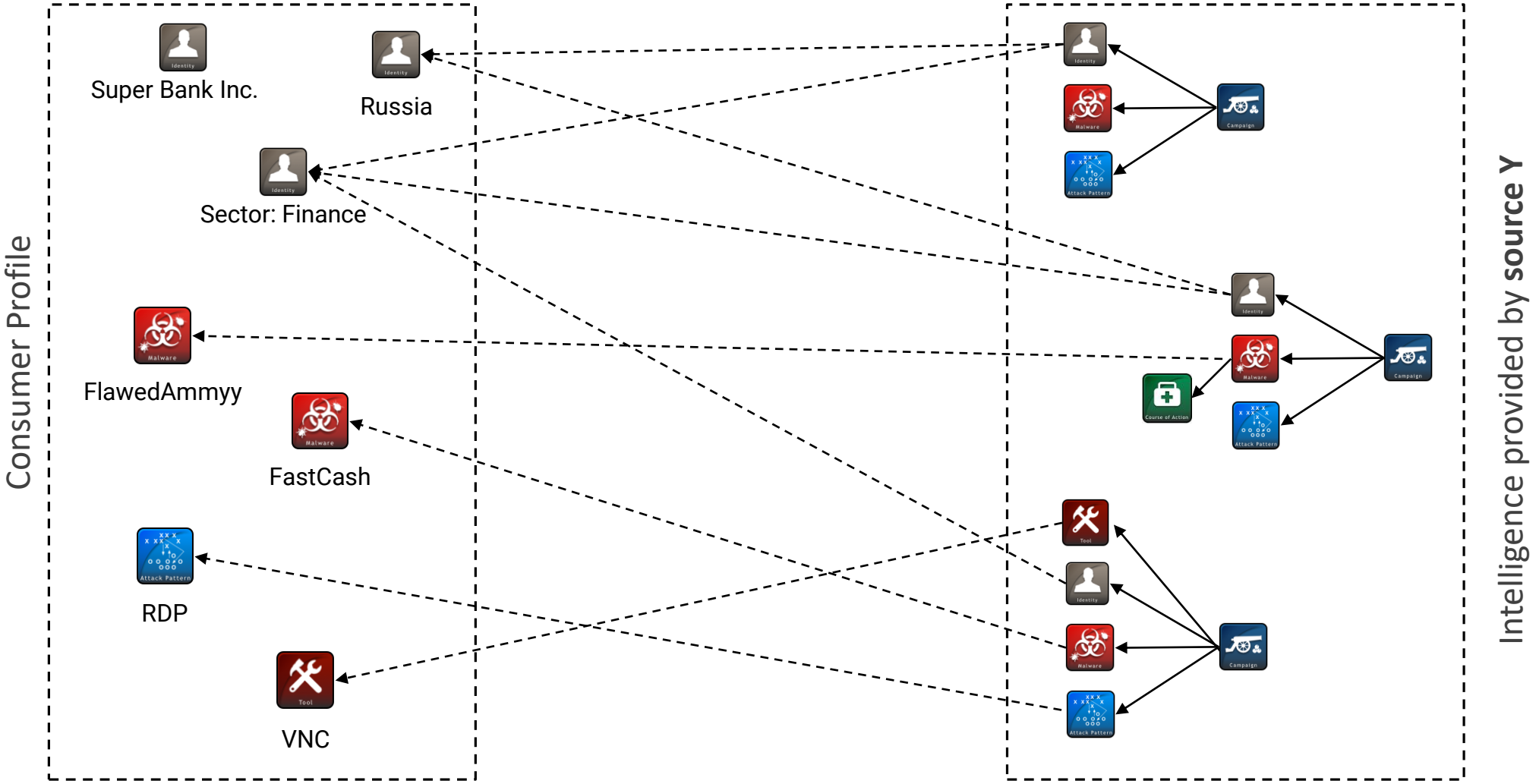


Vendor C
Intel "hidden" as unstructured data in reports

Proximity - Source X



Proximity - Source Y



Relevancy - Source A

Relevant malware

Malware names: njrat, shamoon, loki, lokibot, gandcrab

Object type	Objects matched	Objects of this type	% of objects of this type
attack-pattern	48	6395	0.751%
campaign	22	1119	1.966%
course-of-action	5	855	0.585%
identity	6	2028	0.296%
indicator	59	16442	0.359%
intrusion-set	3	461	0.651%
malware	114	3927	2.903%
report	90	3613	2.491%
threat-actor	1	304	0.329%
vulnerability	2	5478	0.037%

Relevant industry sectors

Industry sector: energy

Object type	Objects matched	Objects of this type	% of objects of this type
identity	84	2028	4.142%

Industry sector: government

No matches

Industry sector: financial-services

Object type	Objects matched	Objects of this type	% of objects of this type
identity	320	2028	15.779%

Relevant CVEs

CVEs: CVE-2017-11882, CVE-2017-0199, CVE-2018-15982

Object type	Objects matched	Objects of this type	% of objects of this type
attack-pattern	76	6395	1.188%
campaign	15	1119	1.340%
indicator	47	16442	0.286%
malware	18	3927	0.458%
report	62	3613	1.716%
tool	1	369	0.271%
vulnerability	11	5478	0.201%

Takeaways

- Consumers must understand and document intelligence & production requirements
- Measure and differentiate between good / bad STIX
- Calculate Proximity
- Leverage the power of intelligence consumers to influence feed providers
- Intelligence provider to improve their feed quality

Learn More / Challenge Us

Sergey Polzunov

sergey@polzunov.com

https://www.linkedin.com/in/polzunov

https://github.com/traut

Jörg Abraham

https://www.linkedin.com/in/joergabraham

@mod_tastic