Comparing methods that identify malicious registrations

Thymen Wabeke | CENTR R&D

30 March 2022

Motivation

- 62% of abusive domains are registered with malicious intents
- For the majority, time between registration and misuse is short

- Verifying new registrations could prevent malicious registrations
  - But: +/- 2580 registrations per day
  - But: only 3 (0.11%) reported at Netcraft within 30 days

Goal

Identify registrations Support would like to review

- Support will assess whether a suspicious registration is malicious
- No resources wasted on verifying legit registrations

Assumptions:

- Manual review after delegation, no algorithmic decision making
- Use only data that’s available during registration

Research questions

- What approaches can we use?
- How would this impact operations?

Today's agenda:

- Discuss results
- Introduce 3 policy choices

Candidates studied

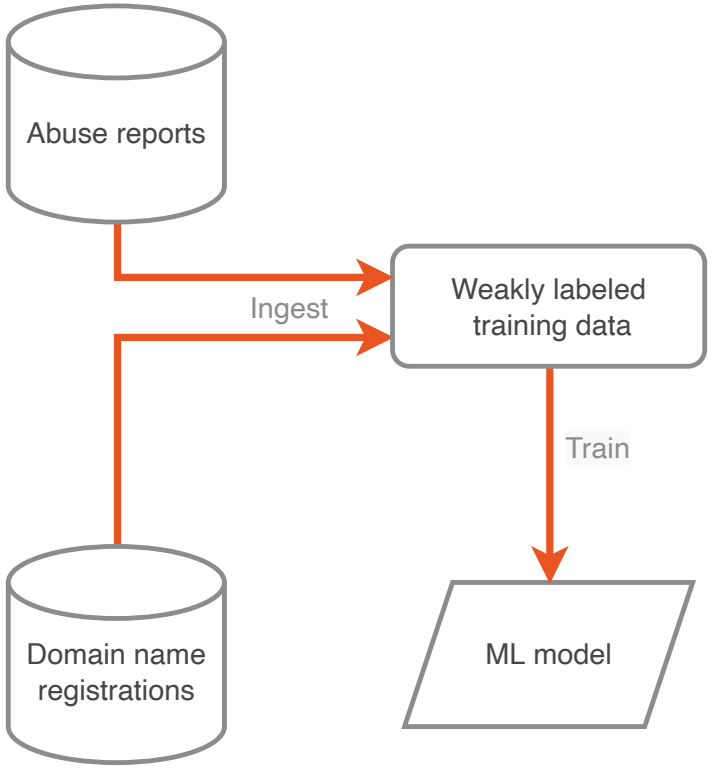
Knowledge-driven:

1. Score system: uses static rules to score suspiciousness

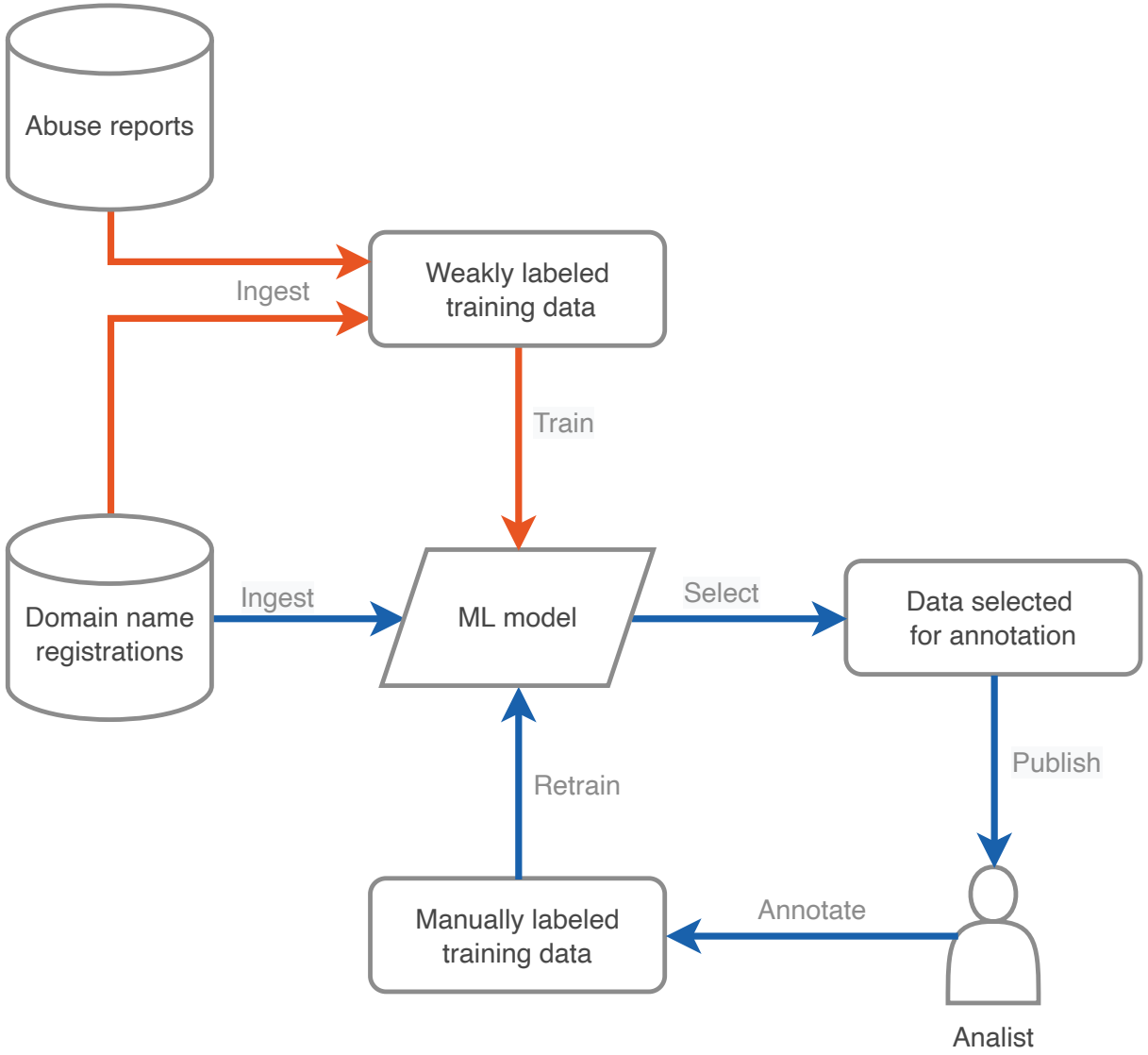
Data-driven:

2. Weak supervision: machine learning model trained using Netcraft data
3. Active learning: updated model using feedback loop

Candidate 2: Weak supervision

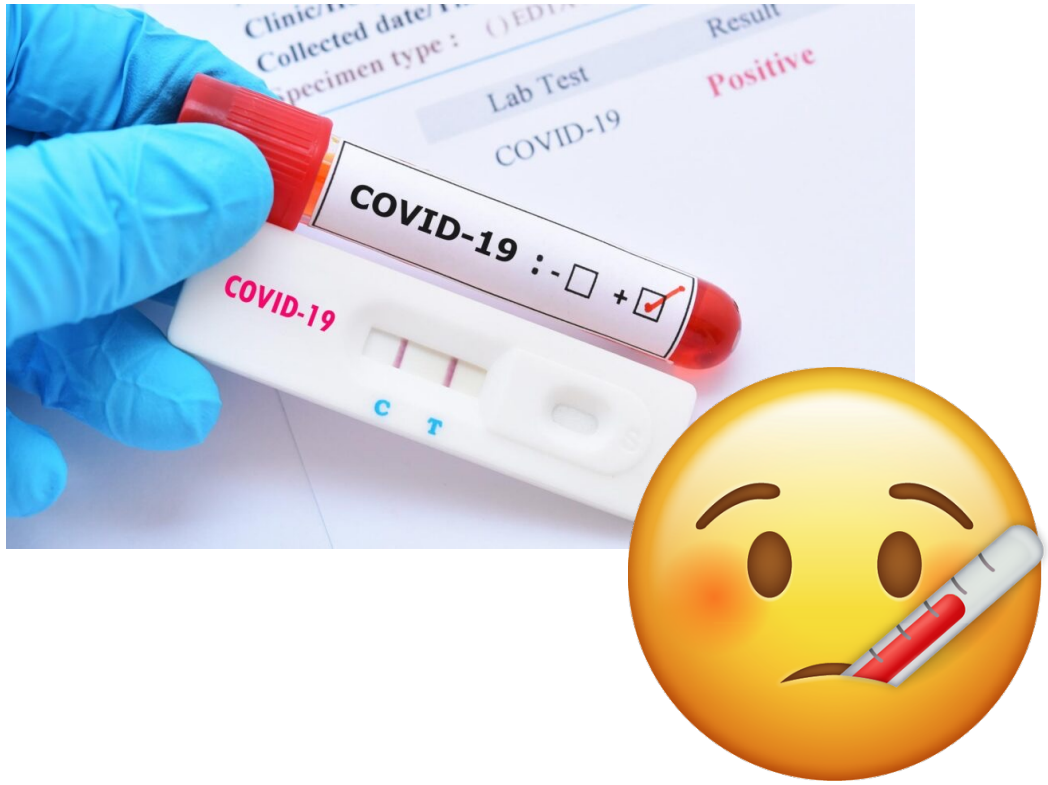


Candidate 3: Active learning



Evaluation metrics

Sensitivity
% positives identified correctly



Specificity
% negatives identified correctly

SIDN LABS

Evaluation datasets

Sensitivity:

- 150 Netcraft reports

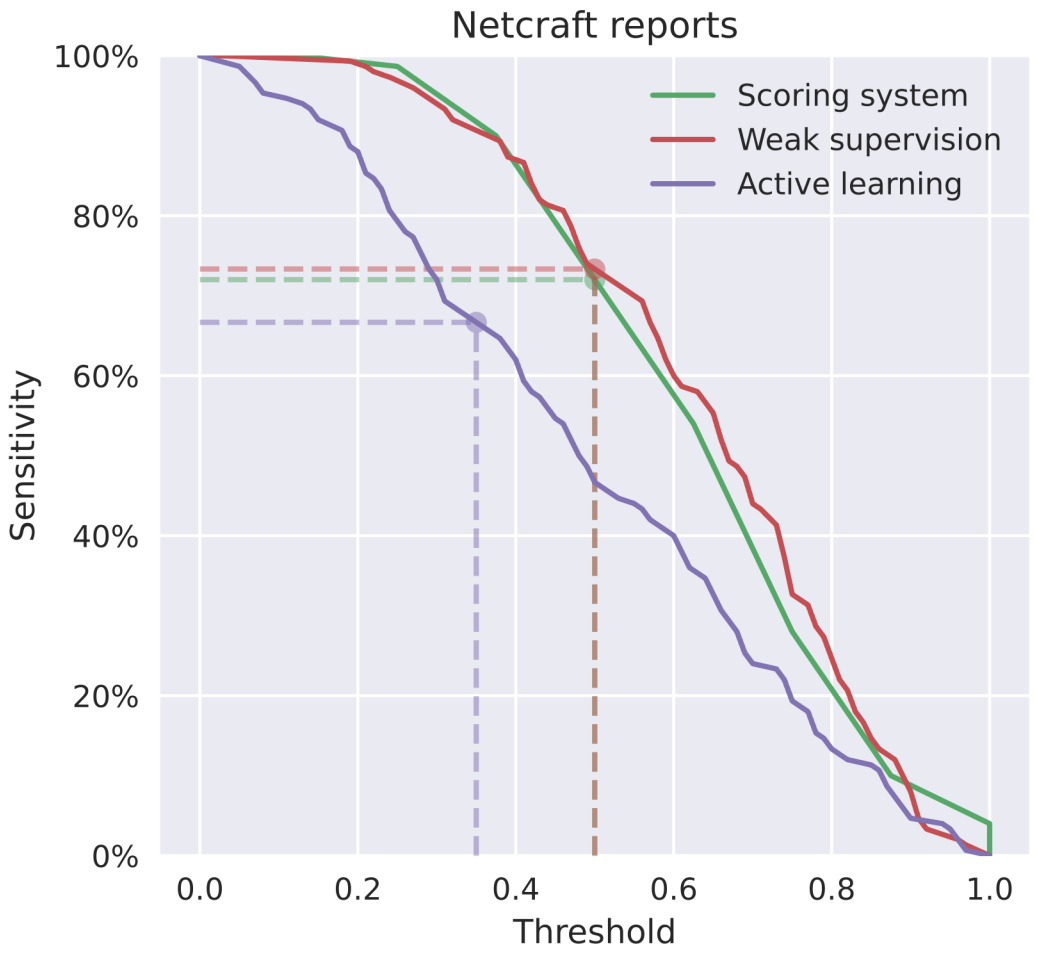
Specificity:

- 968 random registrations
- Manually labeled by Support team

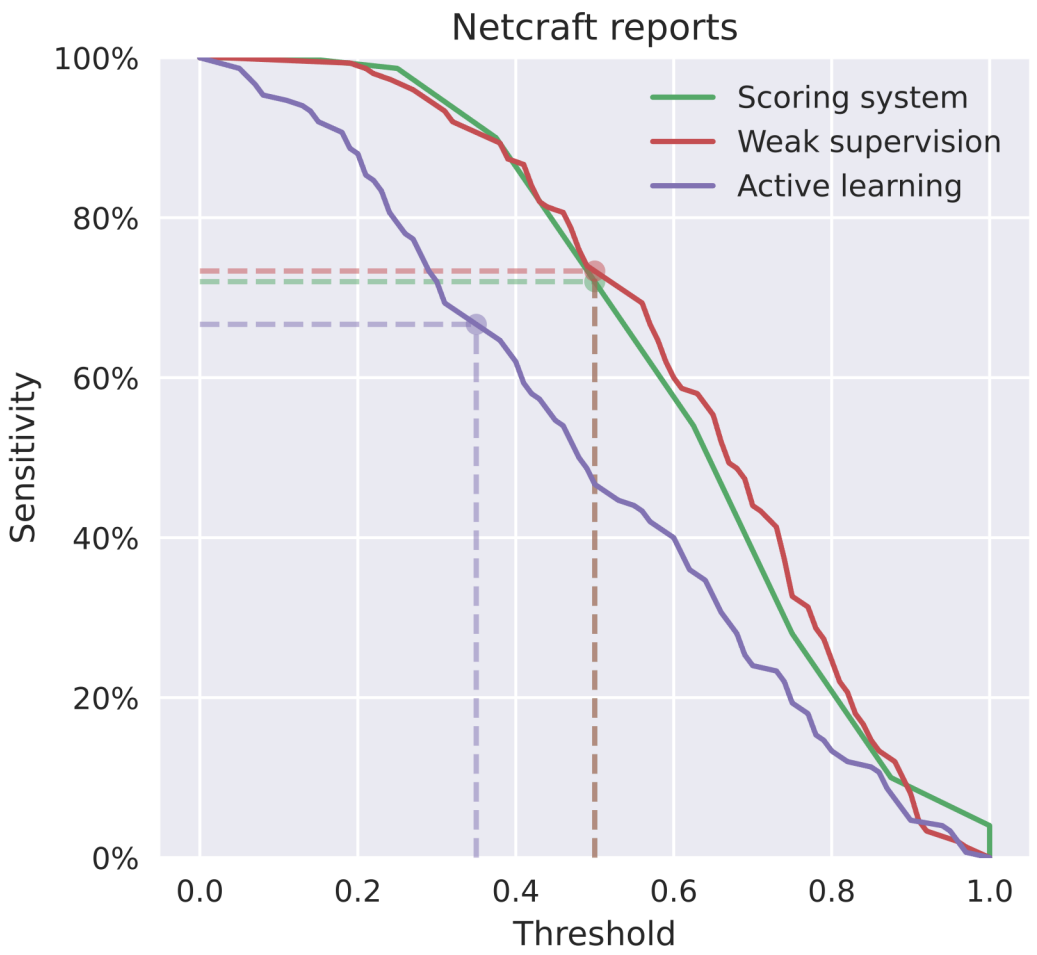
Source	Target	Count	Unique
Netcraft	Bona fide	0	0
	Malicious	150	118
Random	Bona fide	920	695
	Malicious	48	43

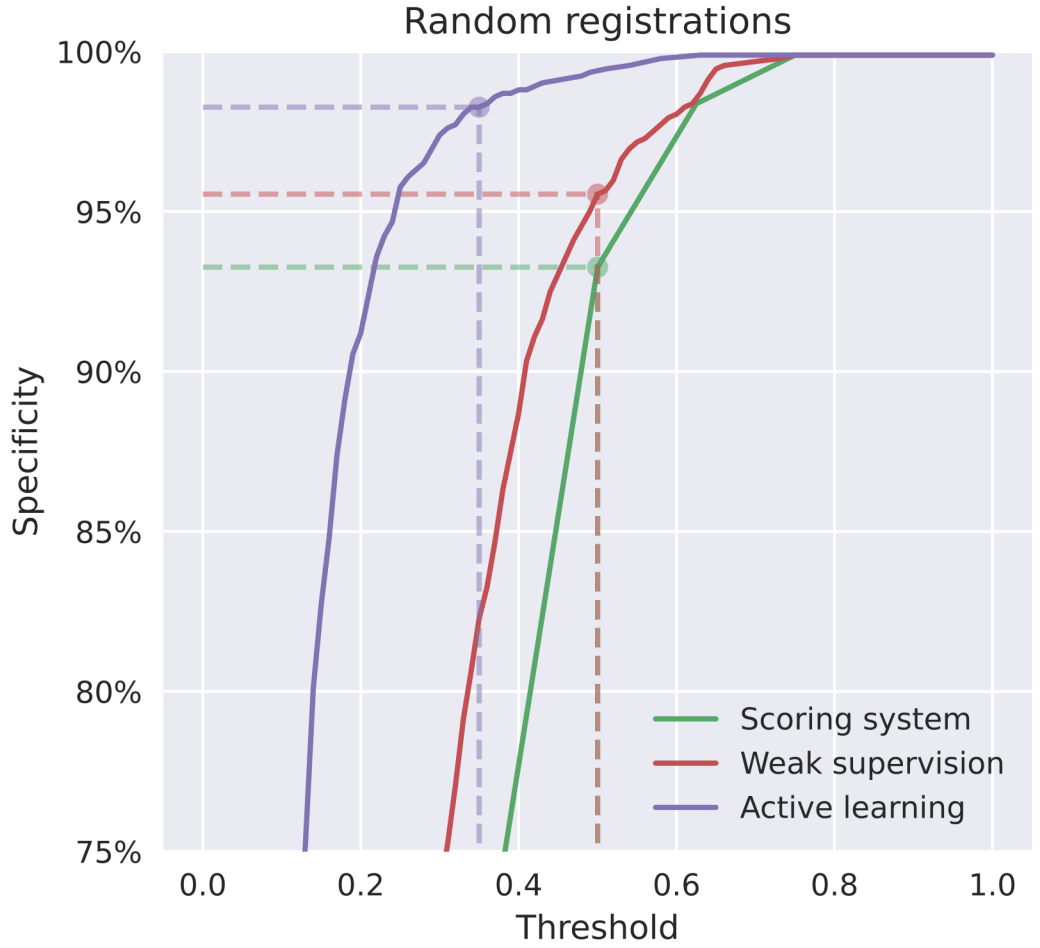
SIDN LABS

Sensitivity and specificity at different thresholds



Sensitivity and specificity at different thresholds





1st choice: detect everything or accept abuse?

Findings:

- All approaches are (more or less) sensitive and specific
- Sensitivity and specificity can be tuned using threshold

Choices:

- Detect all malicious registrations?
- Prevent reviewing false positives?

Expected number of daily reviews (1/2)

- Select a threshold per candidate
- Compute number of reviews we expect per day using two scenarios
  1. True abuse ratio = 0.11% (based on Netcraft reports)
  2. True abuse ratio = 5% (based on labels by Support)

Candidate	Threshold	Sensitivity	Specificity
Scoring system	0.5	72.0%	93.2%
Weak supervision	0.5	73.3%	95.6%
Active learning	0.35	66.7%	98.3%

SIDN LABS

Expected number of daily reviews (2/2)

Scenario 1: 0.11% malicious

	Review			No review		
	Σ	✓	✗	Σ	✓	✗
Score system	176	2	174	2404	2403	1
Weak supervision	117	2	115	2463	2462	1
Active learning	48	3	45	2532	2532	1

Scenario 2: 5% malicious

	Review			No review		
	Σ	✓	✗	Σ	✓	✗
Score system	258	93	165	2322	2286	36
Weak supervision	204	95	109	2376	2342	34
Active learning	129	86	43	2451	2408	43

Expected number of daily reviews (2/2)

Scenario 1: 0.11% malicious

	Review			No review		
	Σ	✓	✗	Σ	✓	✗
Score system	176	2	174	2404	2403	1
Weak supervision	117	2	115	2463	2462	1
Active learning	48	3	45	2532	2532	1

Scenario 2: 5% malicious

	Review			No review		
	Σ	✓	✗	Σ	✓	✗
Score system	258	93	165	2322	2286	36
Weak supervision	204	95	109	2376	2342	34
Active learning	129	86	43	2451	2408	43

2nd choice: how many analysts do we need?

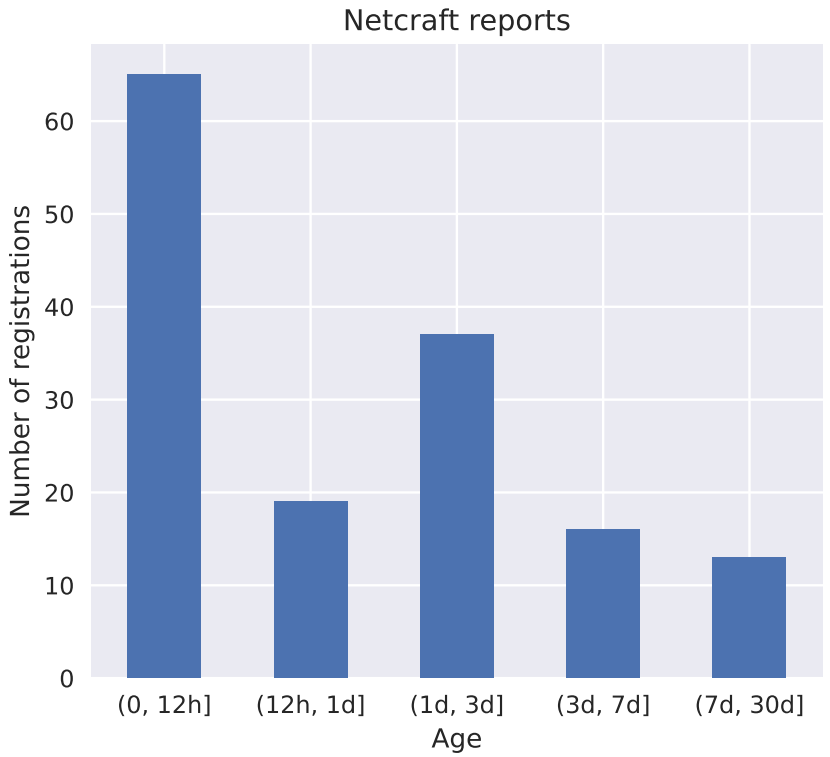
Finding:

- We can expect 50-250 registrations per day
- Review can take up to 15 minutes
- Many false positives, especially with a low abuse ratio

Choices:

- How much time do we want to invest?
- Can we speed up the review process?
- Does this influence our previous choice? Specificity more important?

Time between registration and abuse report



3rd choice: identify more or faster ?

Findings:

- Majority of Netcraft reports has age < 1 day

Choices:

- How fast can we review registrations? What about weekends?
- Identifying *unknown* malicious registrations or find them *faster*?
- Should we automatically defer registrations?

Future work

- Works towards “operational prototype”
  - Implement reputation features and lessons learned
  - Continue comparing 3 candidates

- Consider sharing our candidates and evaluation code

- Discuss policy choices

Are there any questions?

Follow us

SIDN.nl


@SIDN

SIDN

Thank you for your attention!

SIDN LABS