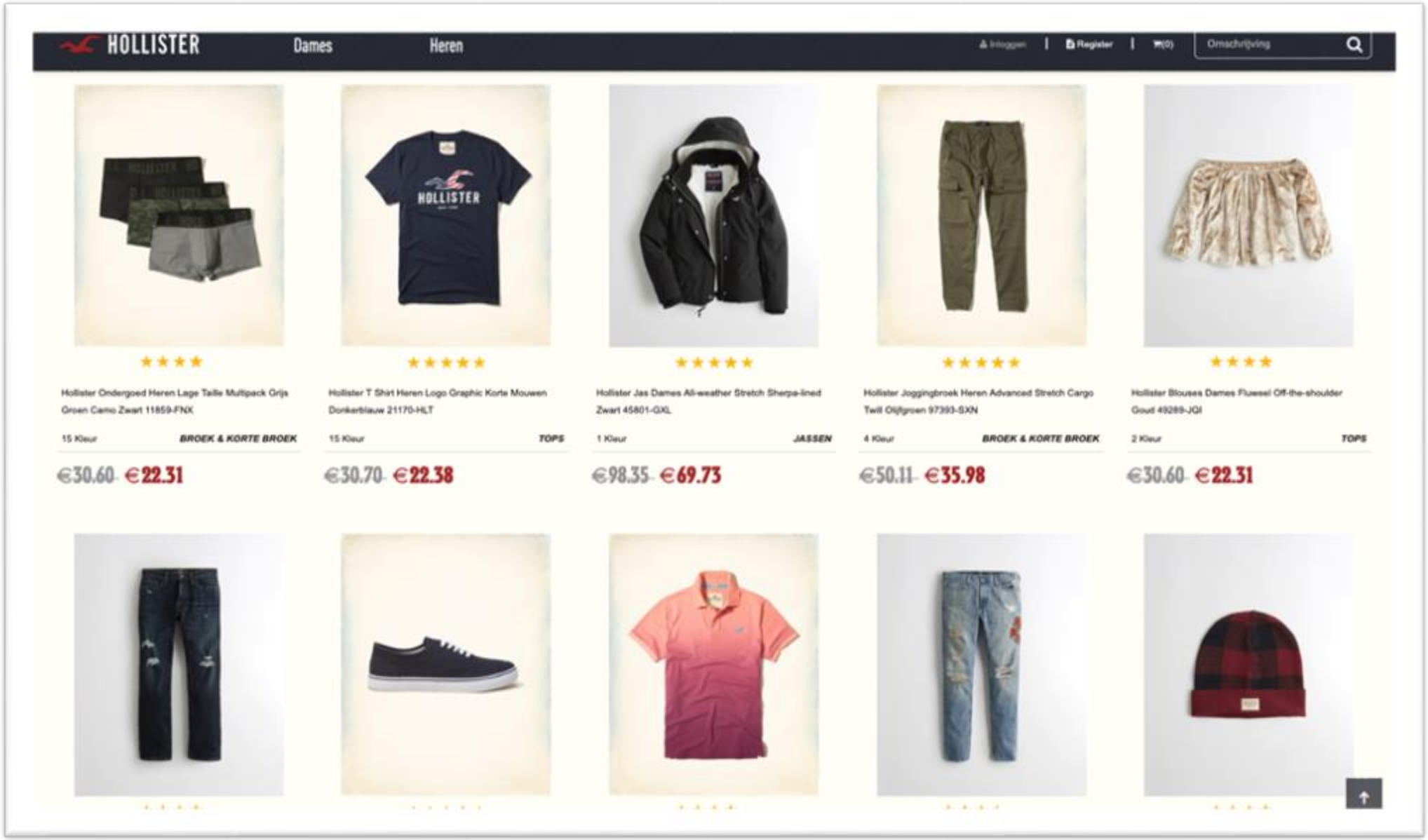
Counterfighting Counterfeit
Detecting and taking down fraudulent webshops at a ccTLD

Thymen Wabeke, Giovane C. M. Moura, Nanneke Franken, and Cristian Hesselman
{firstname}.{lastname}@sidn.nl

Published at PAM2020: https://doi.org/10.1007/978-3-030-44081-7_10

SIDN LABS

nederlandwebshop.nl



SIDN's interest

- Consumer losses [1-4]
- Trust in Internet may decrease

Perfect vantage point:

- List of *all* .nl-domains;
- Registration data and measurements.

Results so far

- Detected thousands since 2016
- Protected users form being scammed

- 2 detection systems, 2 case studies
  - BrandCounter (2018 Q1-2)
  - FaDe (2019 Q1)

Research questions of paper

Q1: How many counterfeit webshops?
Q2: How to take counterfeit shops offline?

Q3: How do counterfeiters operate?

BrandCounter

Observation:

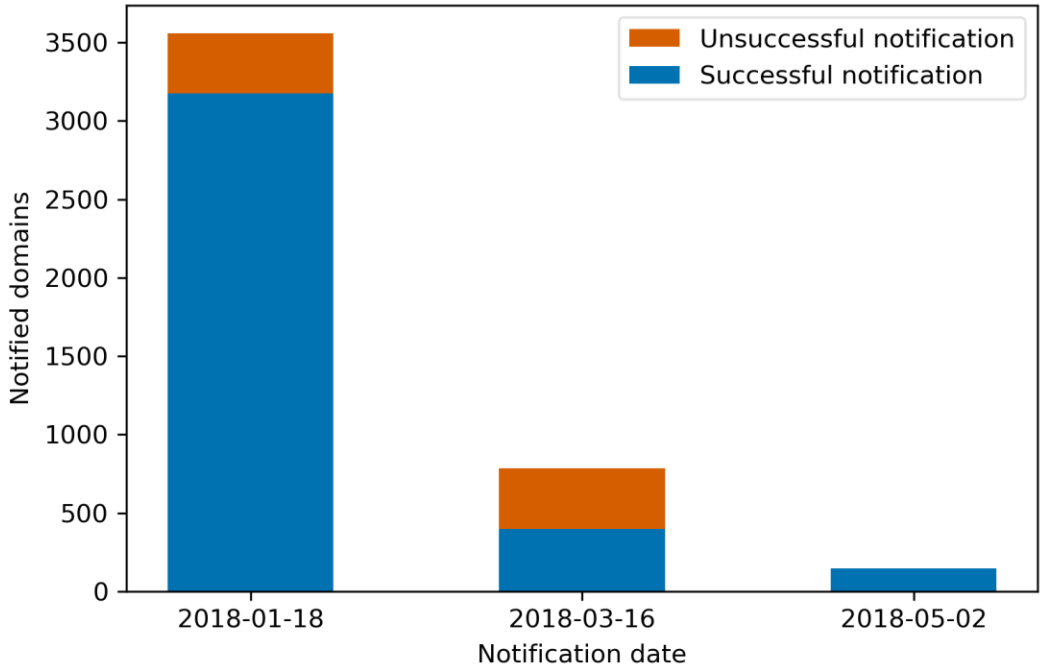
- Long html <title> tags listing brands (Nike, Reebok, Gucci, etc.)
- Helps rank high on search engines (SEO) [5]

Method:

- Create a list with 1100 brands and discount words
- Count suspicious words in the html <title> of .nl-websites
- >5 words (arbitrary), mark as suspicious

BrandCounter notification

- 18,952 suspicious domain names found by counting brands
  - 42.3% registered with *Registrar A*
- We (SIDN) have limited possibilities to take down domains directly
- Sent 4107 notifications to *Registrar A*
  - 3708 took down (90.31%)



Fake Detector (FaDe)

- Have counterfeiters given up?
- Learned to avoid BrandCounter?

Towards FaDe:

- Collaboration with ICS, a credit card issuer
- ICS provided 231 shops involved in scams
- Classification model based on supervised machine learning

Photo by Sigmund on Unsplash

SIDN LABS

- Support Vector Machine
- Optimized using grid search

Dataset

- 231 counterfeit
- 229 legitimate

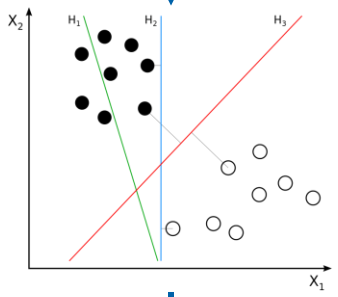
Features

- 6 registration
- 3 infrastructure

Training samples

Testing samples

Train model



Apply model

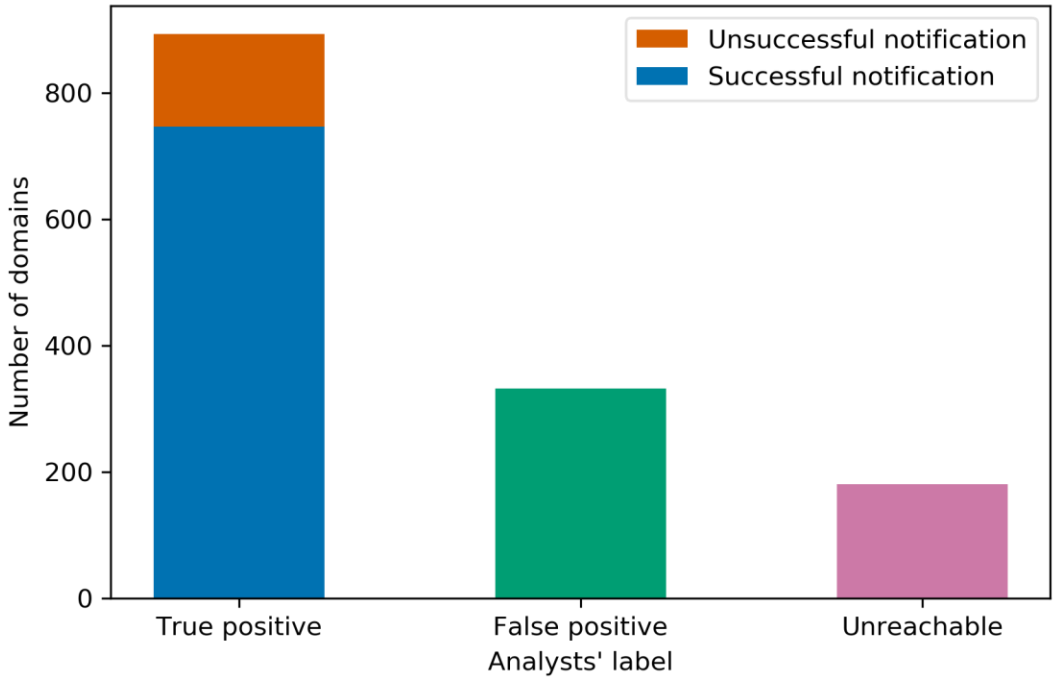
Samples	Precision	Recall
Train (cross-validation)	0.98	0.97
Test	1.0	1.0

SIDN LABS

FaDe notification

- Applied model to 30k .nl-domains
  - 1407 classified as suspicous
  - 894 true positives (73%)

- Sent 894 notifications to registrars
  - 747 took down (84%)

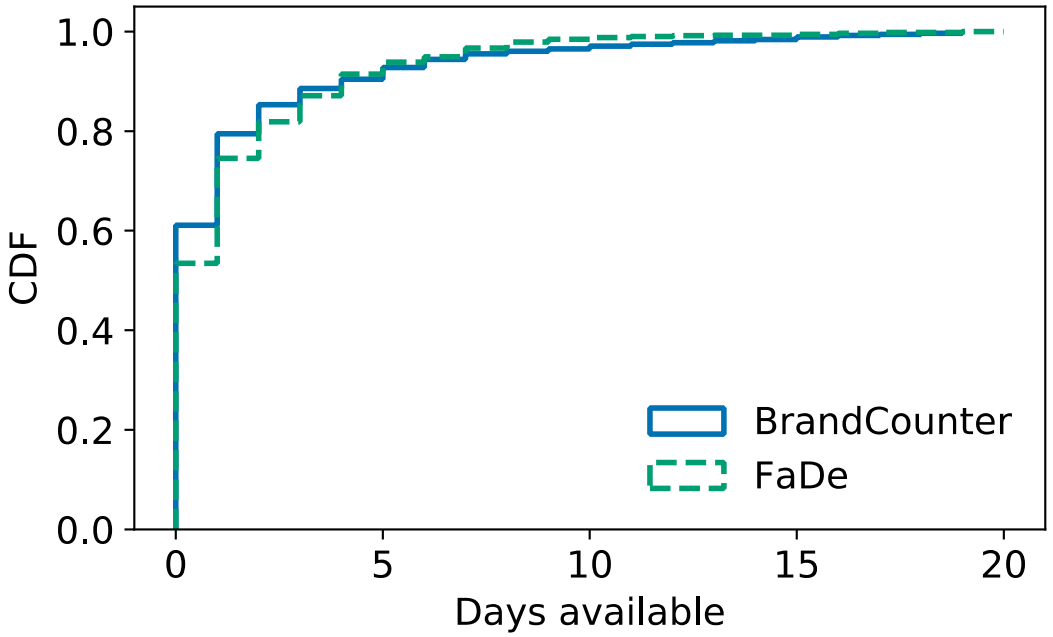


How do counterfeiters operate?

Photo by JESHOOTS.COM on Unsplash

Production farm of shops: automation

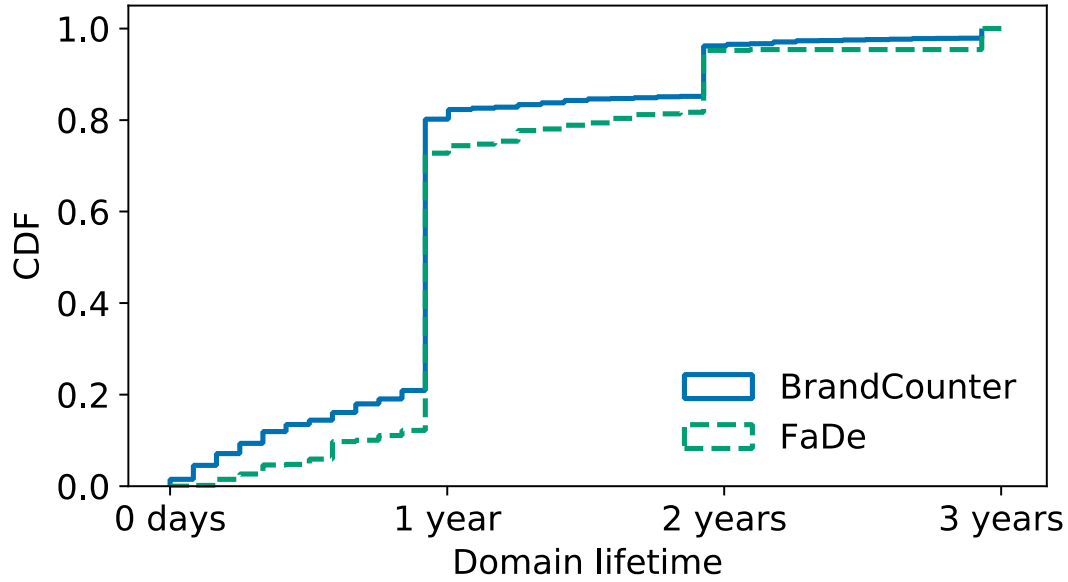
- Mostly cheap registrars that offer APIs
- Similar yet different website templates
- 80% is a re-registered domain
  - Majority re-registered immediately
  - Benefit from “residual reputation” [6]



Days in between domain expiration and re- registration.

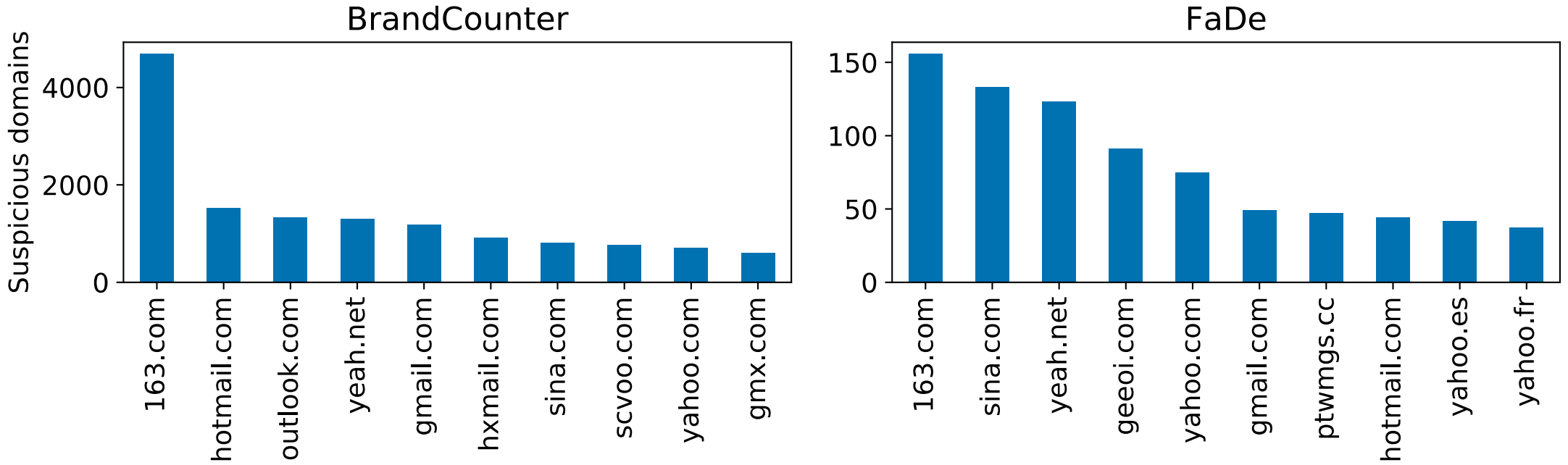
Spreading chances: domains cheap and disposable

- Domain names do not match content
- Spelling mistakes, translation errors
- Domains have short lifetimes

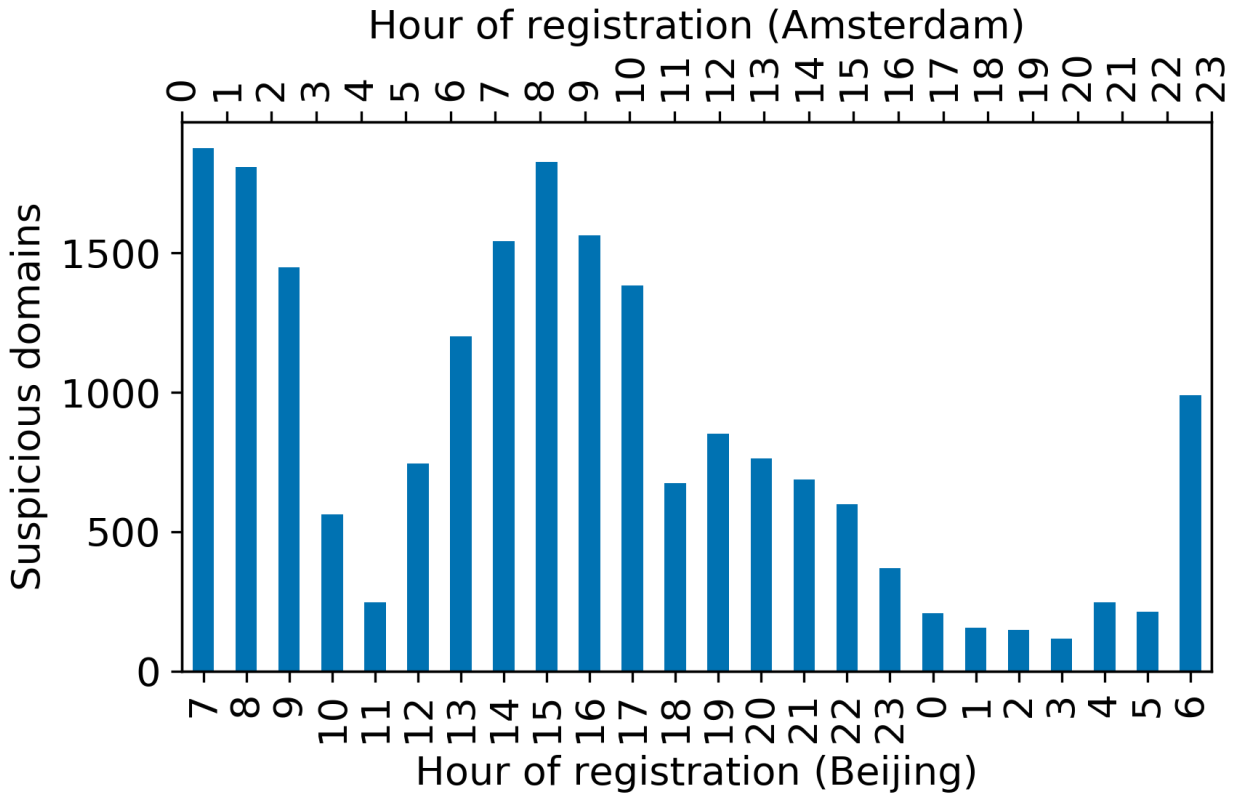


Most domains not renewed after 1 year— the registration period.

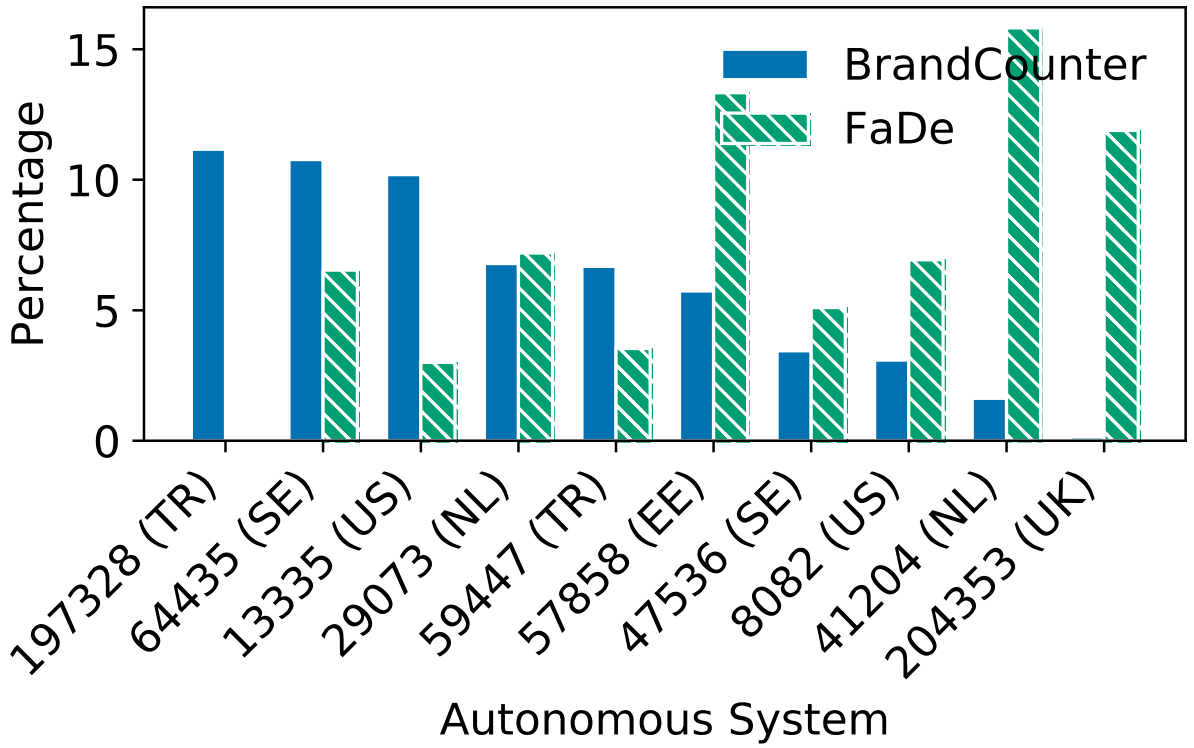
Registrations from China



Registrations from China



Not hosted in China



We helped to take down 4455 counterfeit webshops

Less counterfeit webshops (not peer-reviewed)

- Learned to avoid us again?
- Changed tactics due to proactive detection and quick notifications?

Year	Taken down
2018	~12,000
2019	4,340
2020	481

Number of counterfeit webshops taken down

Less domains ≠ less scams:

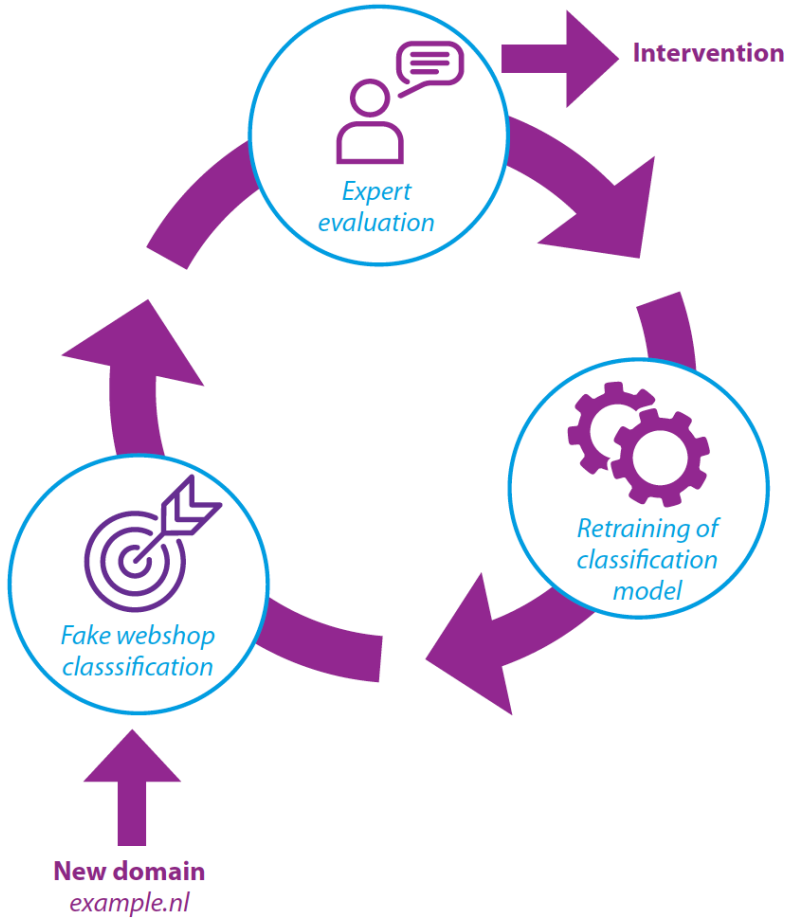
- Share signals with other organizations
- Research on wider disruption

We shared COVID-19 webshops that promoted iDEAL with Currence for broader analysis and wider disruption

https://www.sidnlabs.nl/en/news-and-blogs/machine-learning-method-identifies-brand-logos-on-fake-webshops

Lessons learned

- Registrars and ICS collaboration was key
- Detectors are simple yet effective
  - Registries have perfect vantage point
  - Suggests little pressure
- It's an ever going wack-a-mole game
  - Monitor features and evaluate model regularly
  - We collaborate with others on this topic



References

1. RTL Nieuws: Dit jaar al 307 nep-webwinkels oine gehaald door politie (in Dutch) (Dec 12 2018), https://www.rtlnieuws.nl/geld-en-werk/artikel/4520646/dit-jaar-al-307-nep-webwinkels-offline-gehaald-door-politie
2. NOS: Consumenten voor 5 miljoen euro opgelicht via nepwinkels op sociale media (in Dutch) (Dec 12 2018), https://nos.nl/artikel/2258095-consumenten-voor-5-miljoen-euro-opgelicht-via-nepwinkels-op-sociale-media.html
3. NOS: Waar komen al die nep-webshops toch vandaan? (in Dutch) (May 5 2018), https://nos.nl/artikel/2230087-waar-komen-al-die-nep-webshops-toch-vandaan.html
4. Peter Hornung: Gef□alschte Sneaker von der FDP? (In German). https://www.tagesschau.de/wirtschaft/fakeshops-plagiate-sneaker-china-101.html (2019)
5. Wang, D.Y., Der, M., Karami, M., Saul, L., McCoy, D., Savage, S., Voelker, G.M.: Search + seizure: The effectiveness of interventions on seo campaigns. In: Proceedings of the 2014 Conference on Internet Measurement Conference. pp. 359--372. IMC '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2663716.2663738
6. Lever, C.,Walls, R., Nadji, Y., Dagon, D., McDaniel, P., Antonakakis, M.: Domainz: 28 registrations later measuring the exploitation of residual trust in domains. In: 2016 IEEE Symposium on Security and Privacy (SP). pp. 691{706 (May 2016). https://doi.org/10.1109/SP.2016.47