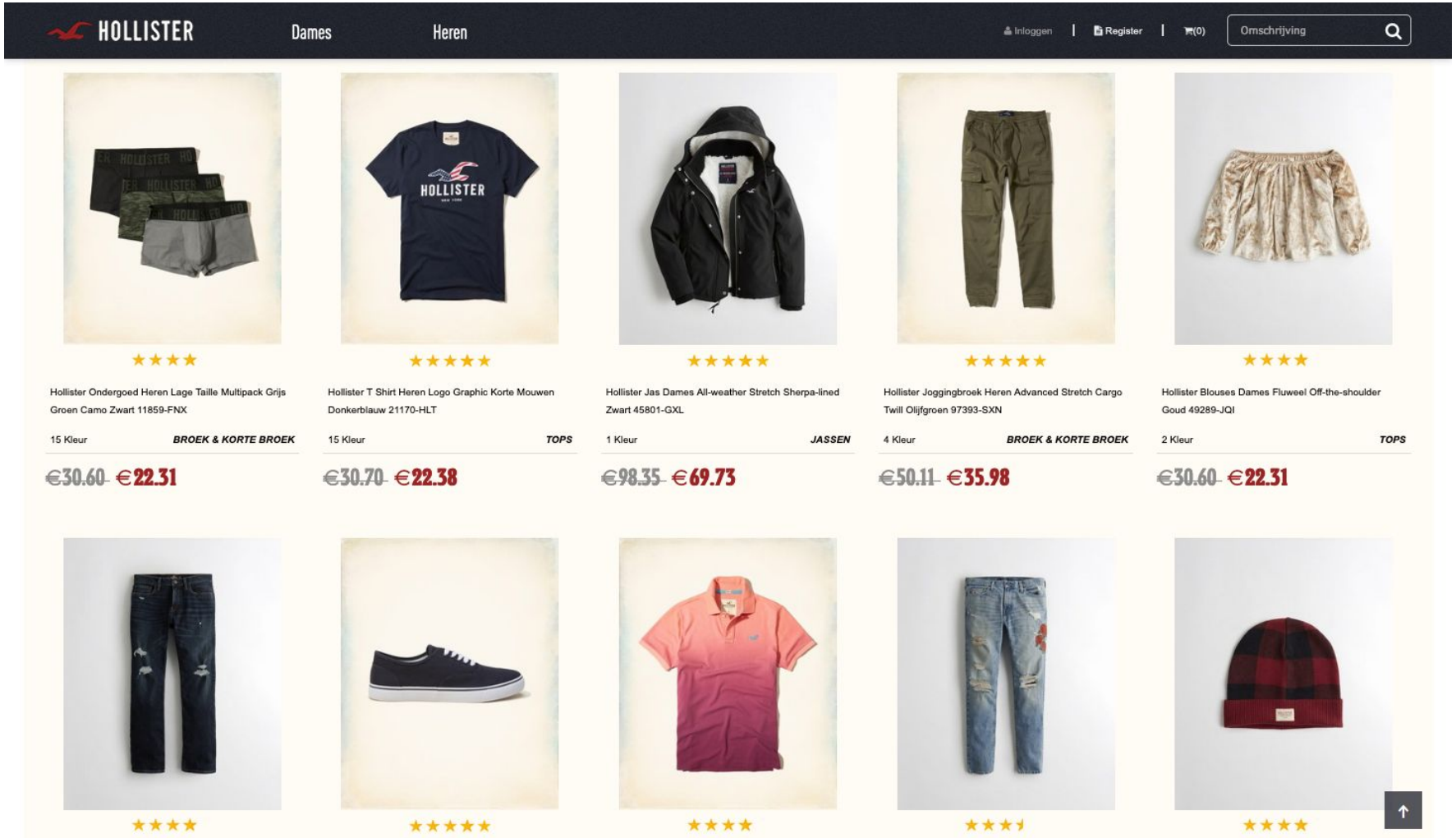
Fake webshop-bestrijding

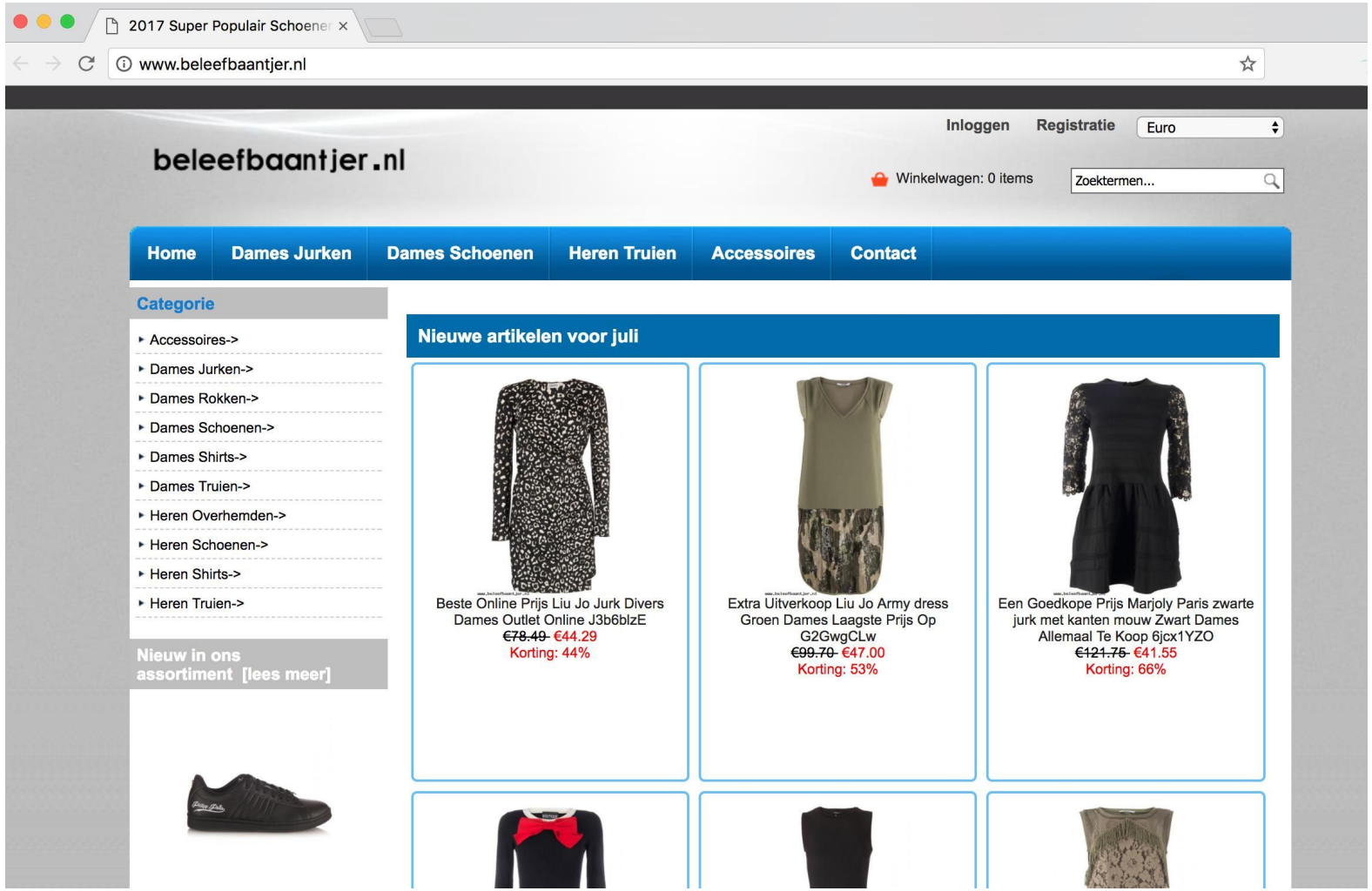
Detectie en bestrijding van frauduleuze webshops

Gepubliceerd bij PAM2020: https://doi.org/10.1007/978-3-030-44081-7_10

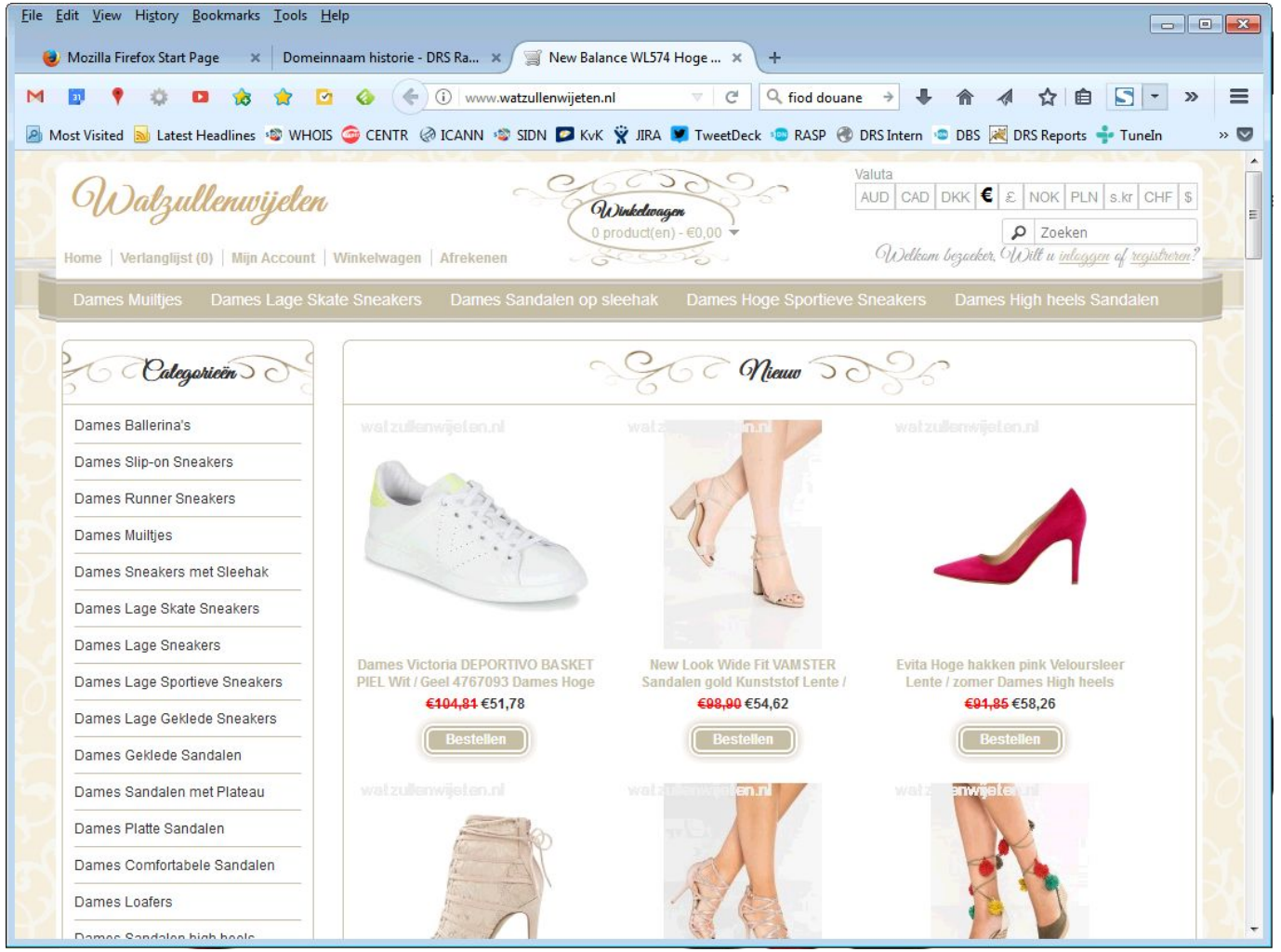
nederlandwebshop.nl



Beleefbaantjer.nl

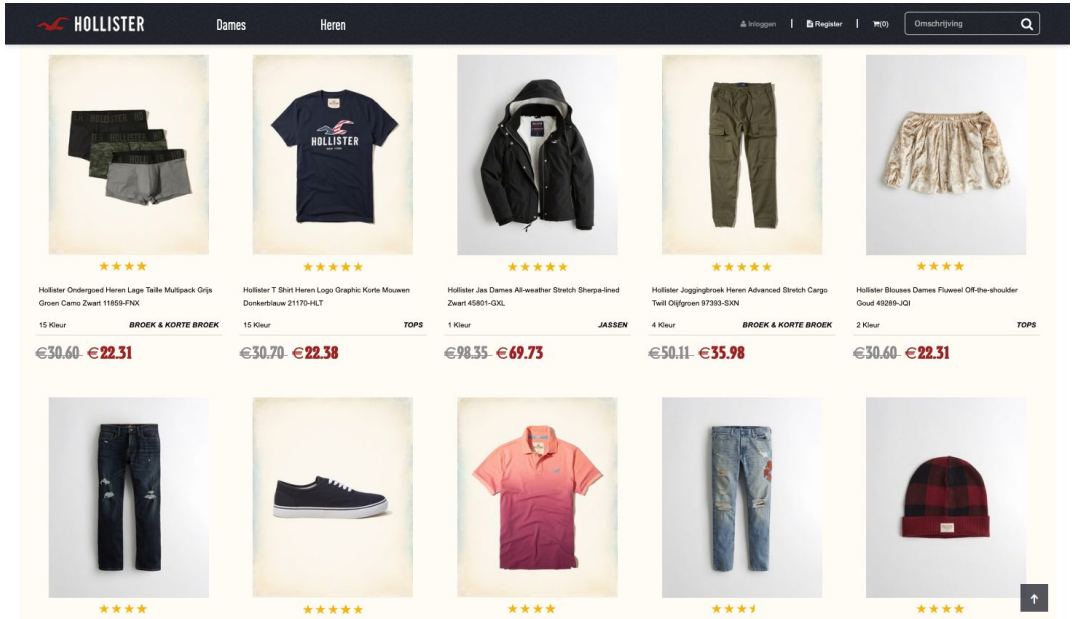


watzullenwijeten.nl



Nepwebshops werken omdat gebruikers het niet weten

of



Waarom houdt SIDN er zich mee bezig

- Schade bij consumenten
- Minder vertrouwen in Internet

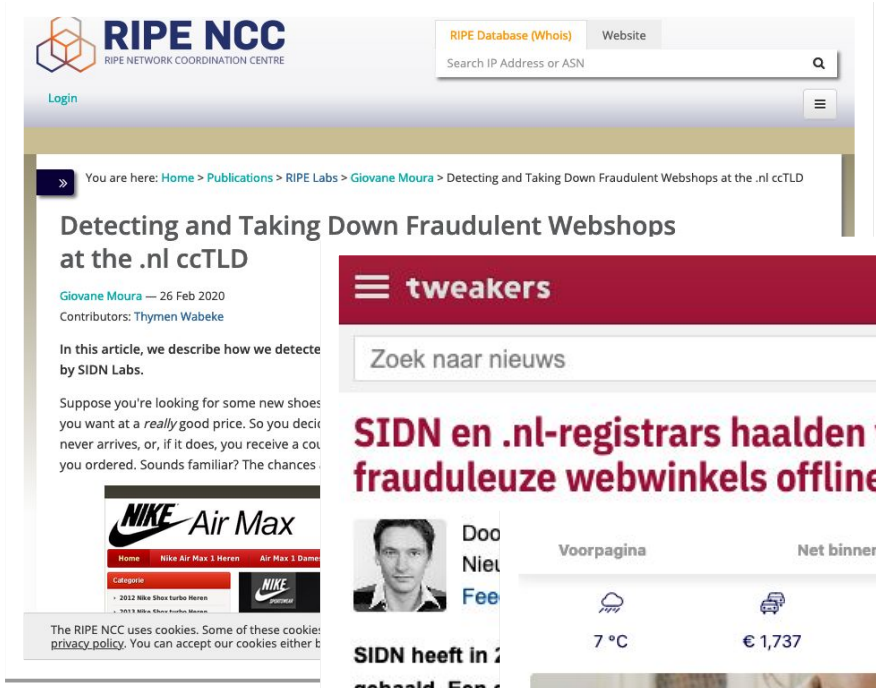
Goed uitgangspunt:

- Lijst van alle .nl-domeinnamen
- Registratiedata en bevragingsmetingen

Kleding, make-up en technische gadgets: mensen kopen het steeds vaker via sociale media. En dat brengt risico's met zich mee, zegt de Autoriteit Consument en Markt (ACM). De schade voor consumenten door oplichting bedroeg vorig jaar bijna 5 miljoen euro. Daarom start de waakhond vandaag een campagne waarin ze consumenten waarschuwt voor nepwinkels op sociale media.

Resultaten tot nu toe

- Sinds 2016 duizenden fake webshops aangepakt
- Gebruikers daarmee beschermd
- 2 detectie-systemen, 2 case studies
  - BrandCounter (2018)
  - FaDe (2019)



3 Vragen die we behandelen:

1) Hoeveel fake webshops?
2) Hoe haal je ze weg?
3)Hoe werken ze?

BrandCounter

Observaties:

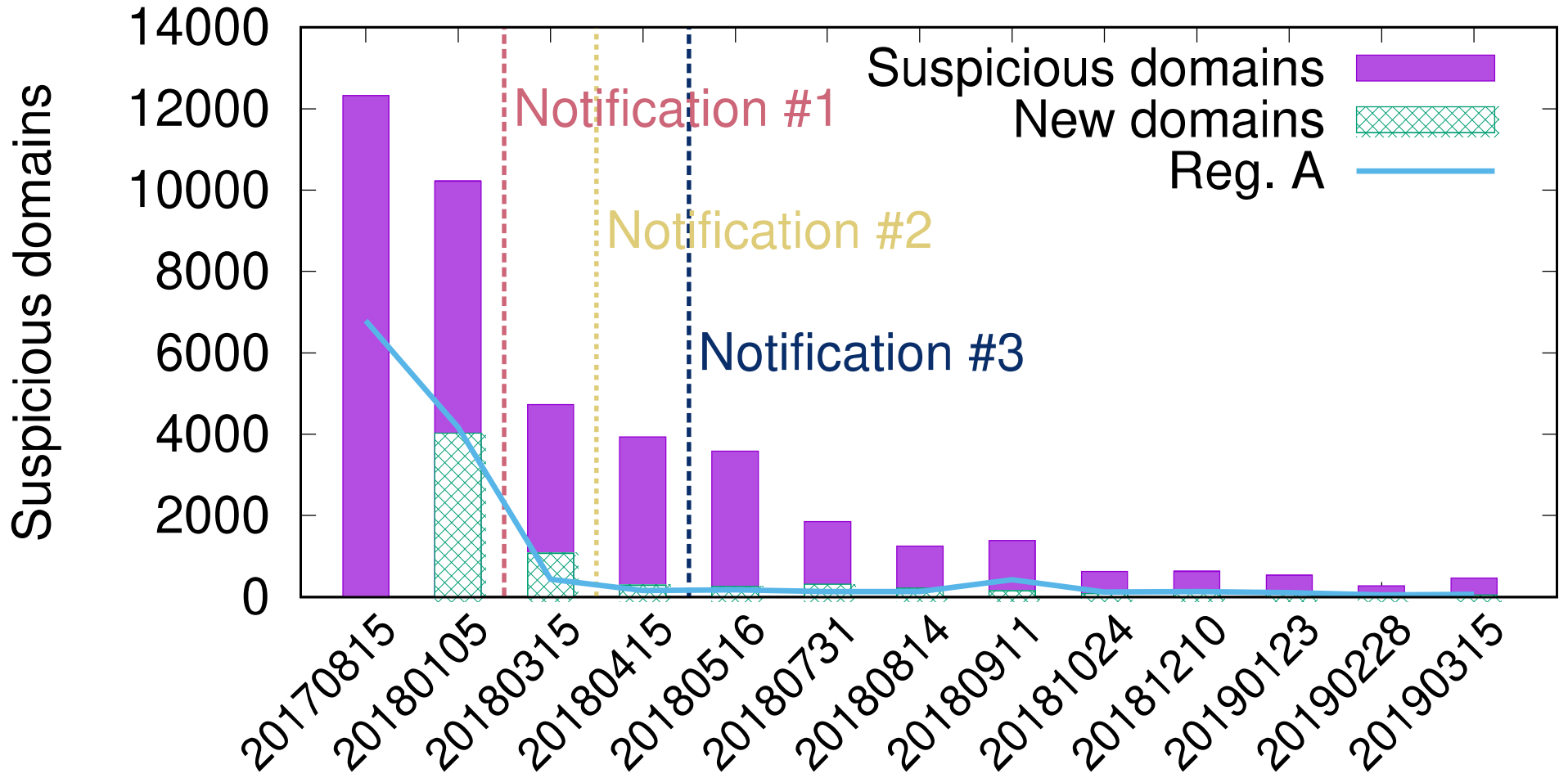
- Lange html <title> tags met merknamen (Nike, Reebok, Gucci, etc.)
- Daarmee hoge SEO-score op zoekmachines [5]

Methode:

- Lijst gemaakt van 1100 merknamen en woorden als 'korting'
- Aantal verdachte woorden geteld in <title> tags van alle .nl-sites
- Bij meer dan 5, markeer als verdacht

SIDN LABS

BrandCounter metingen



Een domeinnaam 'weghalen'

- Een domeinnaam **zelf** is nooit illegaal
- Wat ermee **gedaan wordt** kan dat wel zijn
- SIDN is ook geen 'content-politie'

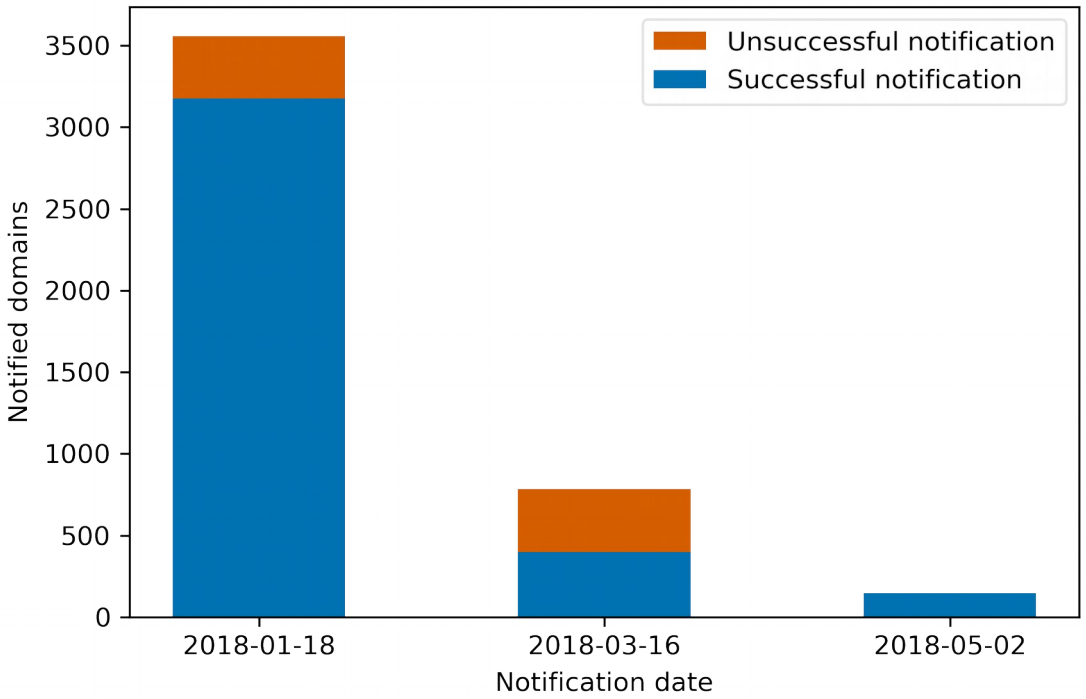
Dus kan (en mag) SIDN niet zomaar een domeinnaam weghalen, behalve in opdracht van een rechter.

Dat moet dus altijd in samenwerking met de registrar en/of hoster.

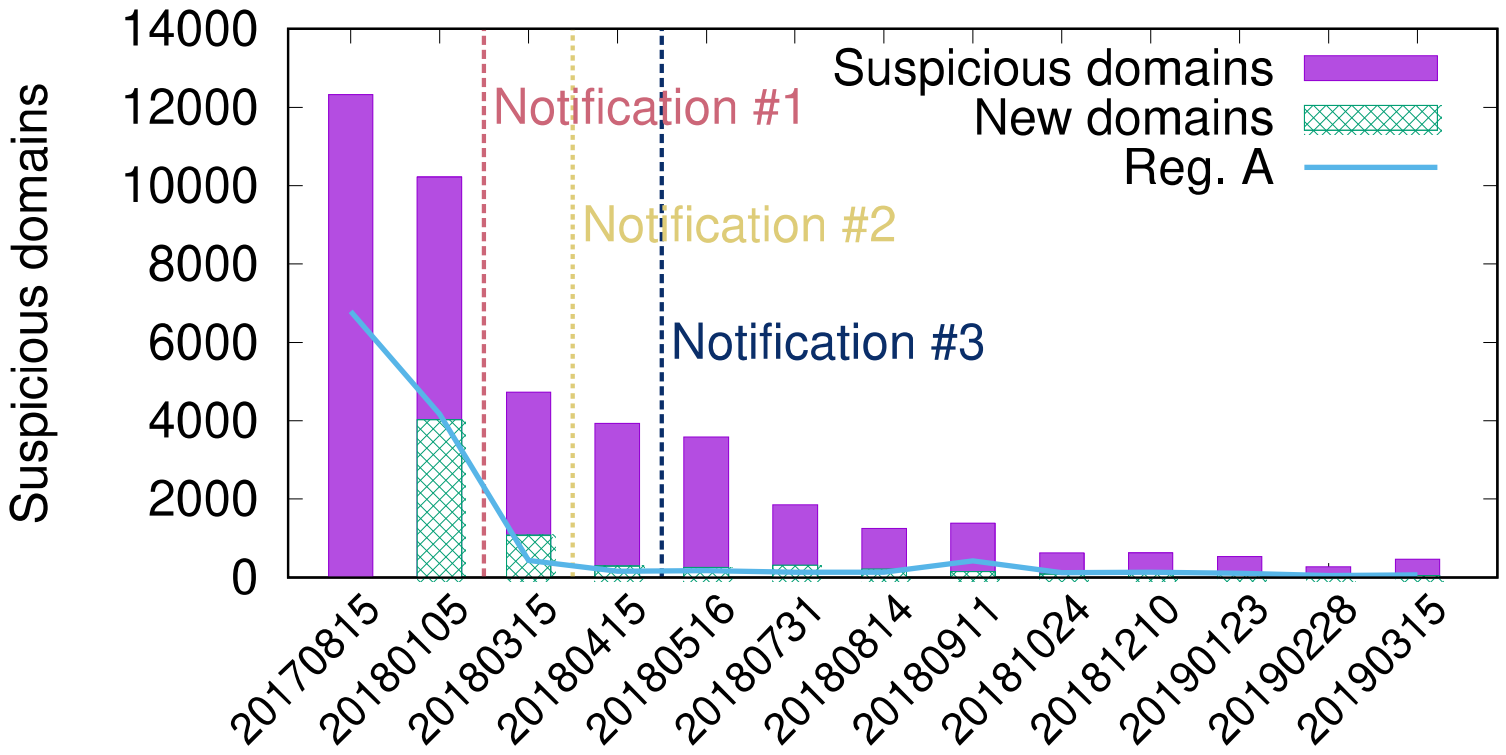
Melding bij “Registrar A”

- Wij (SIDN) hebben maar beperkte mogelijkheden om domeinnamen rechtstreeks te verwijderen
  - 42,3% van de verdachte domeinnamen bij *Registrar A*

- 4107 meldingen verstuurd naar *Registrar A*
  - 3708 door ze verwijderd (90,31%)



BrandCounter metingen



Hebben ze het opgegeven?
Of geleerd hoe BrandCounter te omzeilen?

Fake Detector (FaDe)

- Niet afhankelijk van de <title> tag
- Niet bevooroordeeld door SIDN's zichtpunt (.nl)

Aanpak:

- Samenwerking met ICS, een creditcard-maatschappij
- ICS leverde 231 webshops die betrokken waren bij fraude
- 'Supervised Machine Learning' om een classificatiemodel te maken

FaDe

- Support Vector Machine
- Optimized using grid search

Dataset

- 231 nep
- 229 echt

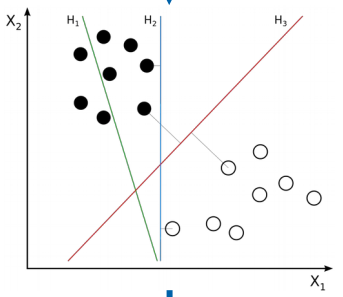
Features

- 6 registratiegegevens
- 3 infrastructuur

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 meldingen

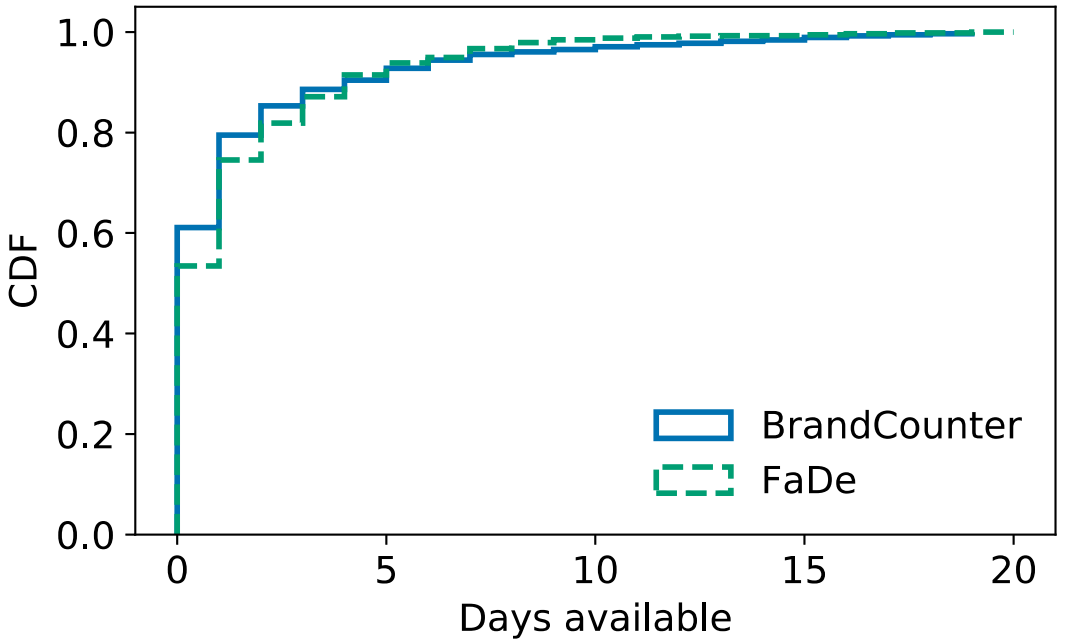
- Model toegepast op 30000 .nl-domeinen
  - 1407 als 'verdacht' aangemerkt
  - 894 (73%) daarvan terecht (true positives)

- 894 meldingen verstuurd naar registrars
  - 747 (84%) aangepakt

Hoe gaan nepwebshops te werk?

Fabriek van webshops: automatisering

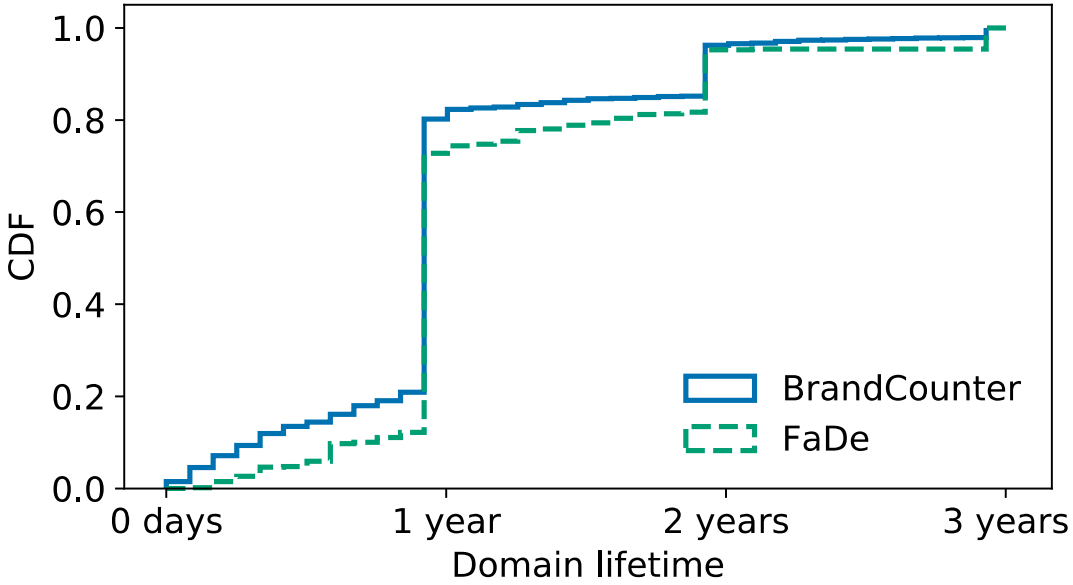
- Voornamelijk via goedkope registrars die een API aanbieden
- Website templates die dezelfde structuur hebben maar er anders uit zien
- 80% van de gebruikte domeinnamen heeft eerder bestaan
  - Merendeel direct geregistreerd na opheffing
  - Voordeel: “residual reputation” (nog bekend bij zoekmachines, etc.) [6]



Days in between domain expiration and re- registration.

Fabriek van webshops: automatisering

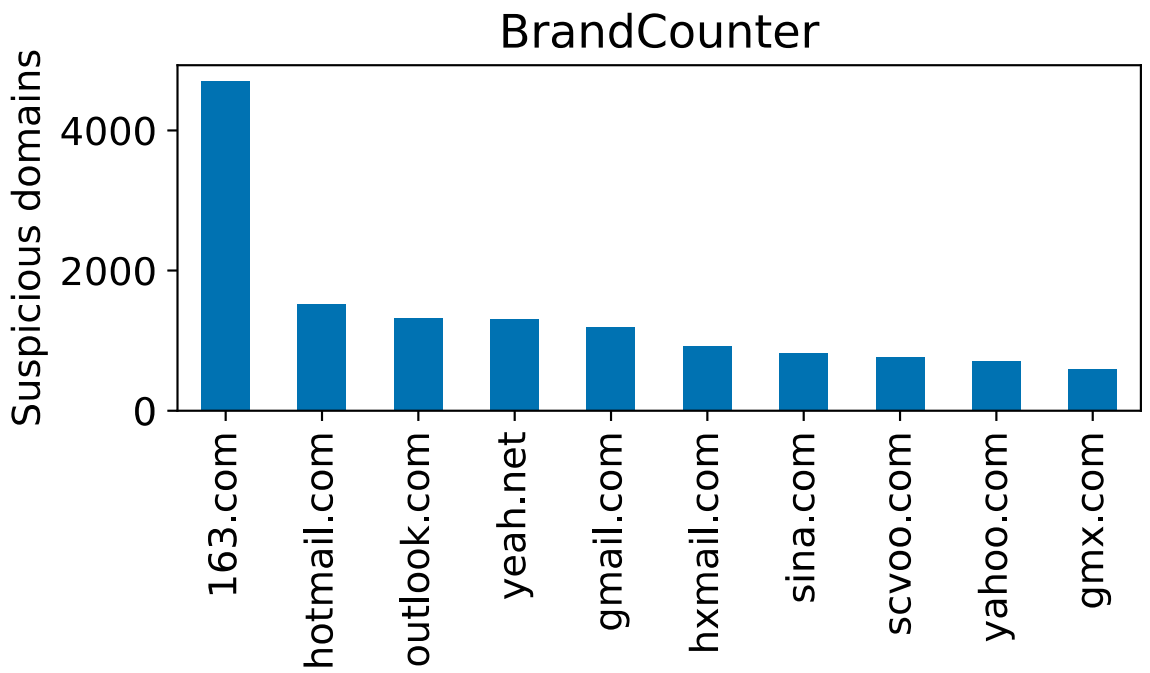
- Domeinnamen komen niet overeen met inhoud
- Spelfouten, vertalingsfouten
- Domeinnamen bestaan niet lang

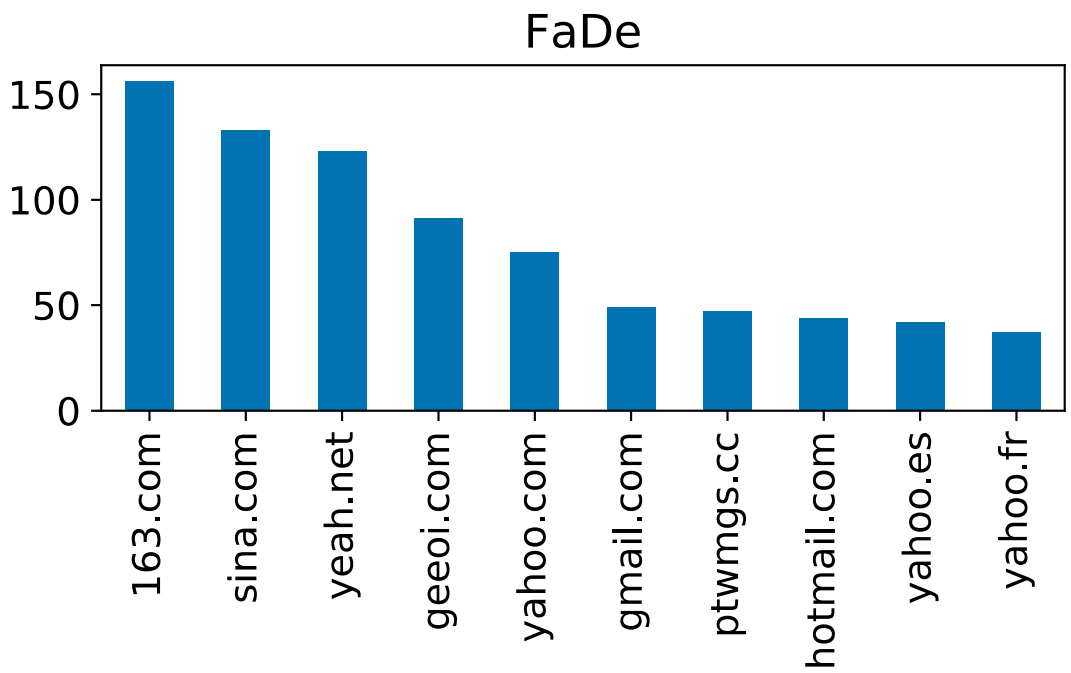


Most domains not renewed after 1 year— the registration period.

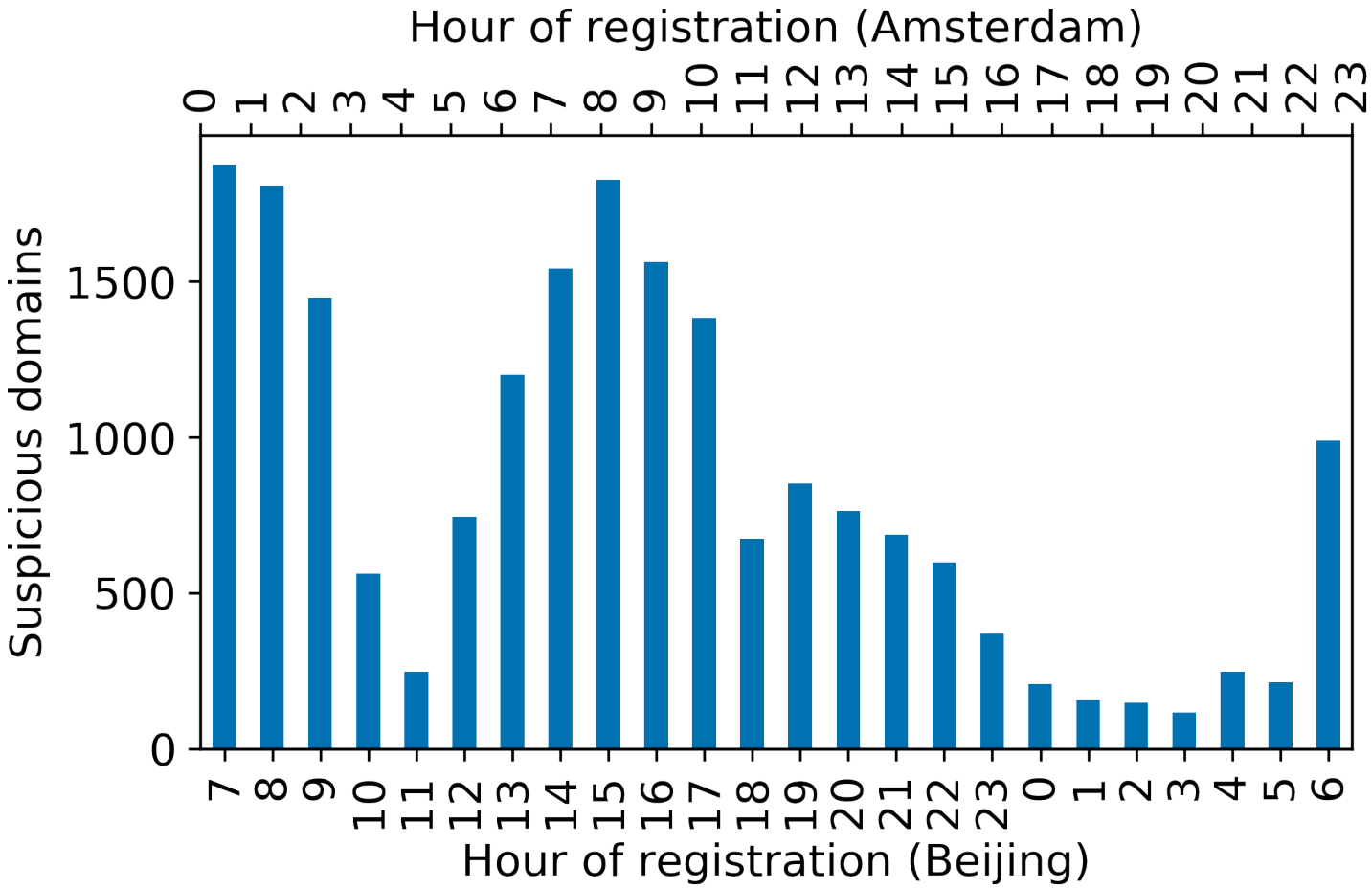
Registraties vanuit China

E-mailadres van de 'eigenaar'

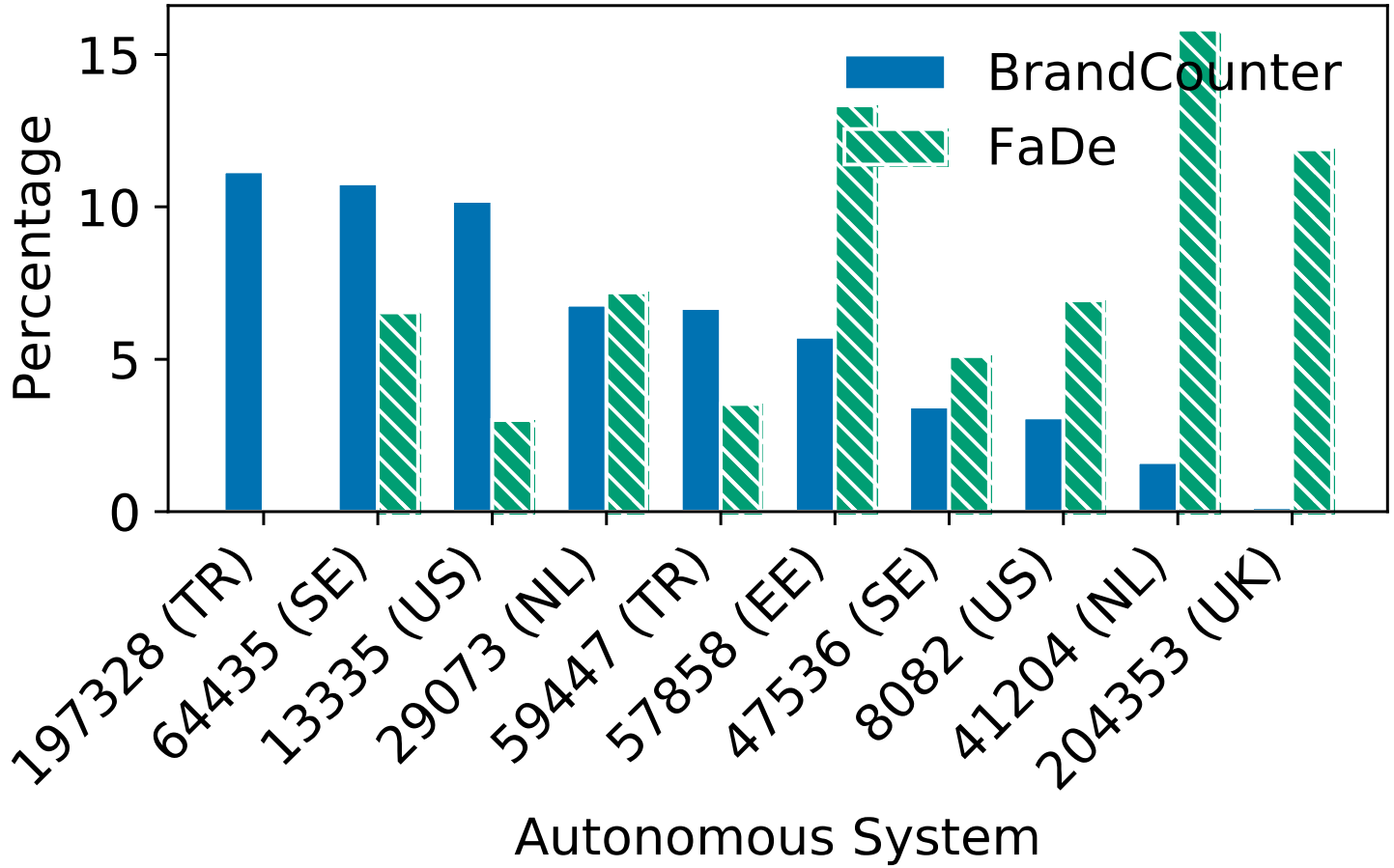




Registraties vanuit China



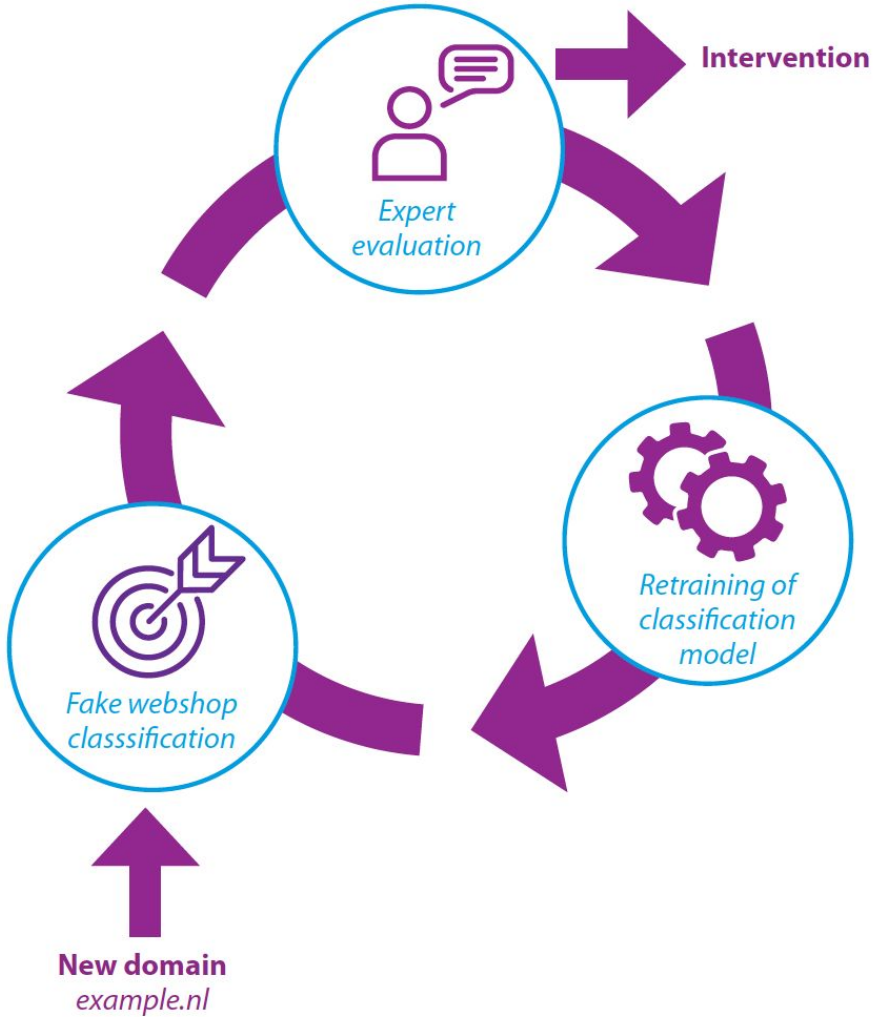
Maar niet gehost in China



Zo hebben wij 4455 fake webshops helpen neerhalen

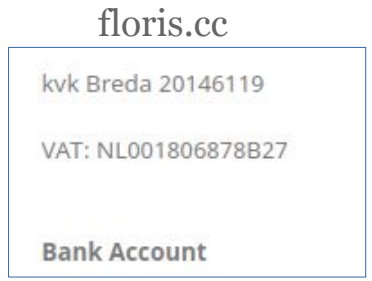
Lessen

- Samenwerking met registrars en ICS was essentieel
- Detectors zijn simpel maar effectief
  - Registries hebben goed zichtsveld
  - Moeten wel wat druk uitoefenen
- Het blijft 'mollenmeppen'
  - Webshops passen zich aan



Wat kun je zelf doen?

- Aanbiedingen te goed om waar te zijn? Dan is dat vaak zo
- Check keurmerken:
  - Thuiswinkel-waarborg etc. zijn links naar het officiele certificaat
- Slecht taalgebruik is een red flag
- Enigszins verdacht?
  - Check het kvk-nummer op kvk.nl, en vergelijk adresgegevens
  - Geen kvk-nummer op website? Wegblijven.



Referenties

- 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
- 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
- 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
- Peter Hornung: Gefalschte Sneaker von der FDP? (In German). https://www.tagesschau.de/wirtschaft/fakeshops-plagiate-sneaker-china-101.html (2019)
- 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
- 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

SIDN LABS