Domain Watch

Blocking phishing domains at registration

[1] Presented by Bryony Hill (Data Scientist)

NOMINET REGISTRY

Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans

Background

Data science at Nominet
- Classification and use of domains
- Modelling retention of domains

Other activities around DNS abuse:

- Other routes for suspension (incl. Law Enforcement Agencies, landing pages on suspended domains)
- Illegal Terms reporting
- Domain Health (ranking registrars by domain abuse)
- Benchmarking technical abuse on .UK (developing KPIs)

NOMINET REGISTRY

Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans

Aim of Domain Watch

Make .UK a safer registry by suspending domains registered for abuse at registration.

With a focus on maliciously registered phishing domains, e.g.

- verifypaypal-id7289.co.uk
- facdbook.co.uk
- account-support-center.org.uk
- wwwicloud.co.uk

NOMINET REGISTRY

Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans

Background

Phase 1 (2018): Netcraft Phishing API used to score and suspend domains

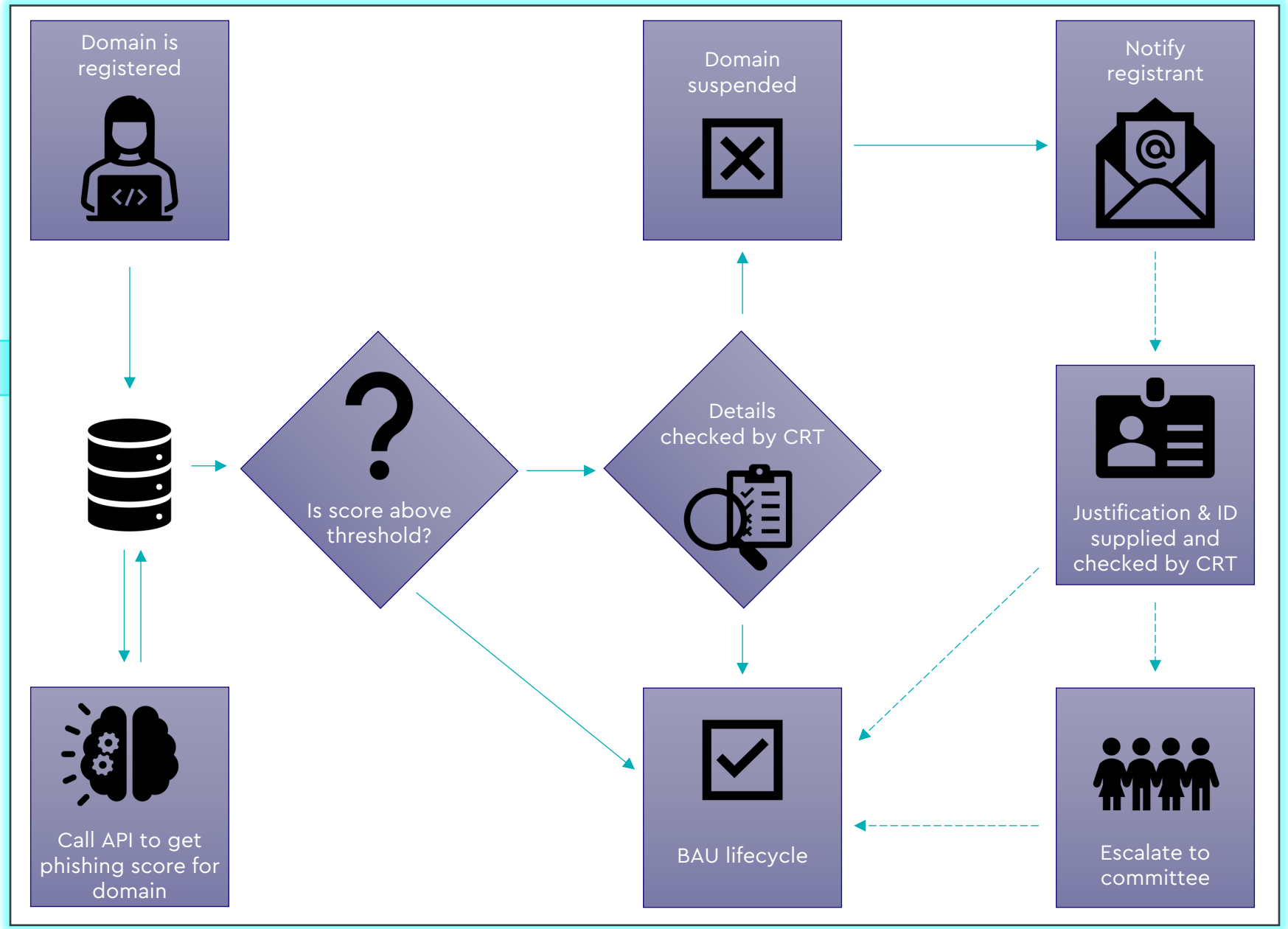
Internal model development (to reduce costs & improve accuracy)

Phase 2 (2019): TensorFlow (neural-network) and regular-expressions (string match) models implemented as new API

NOMINET REGISTRY

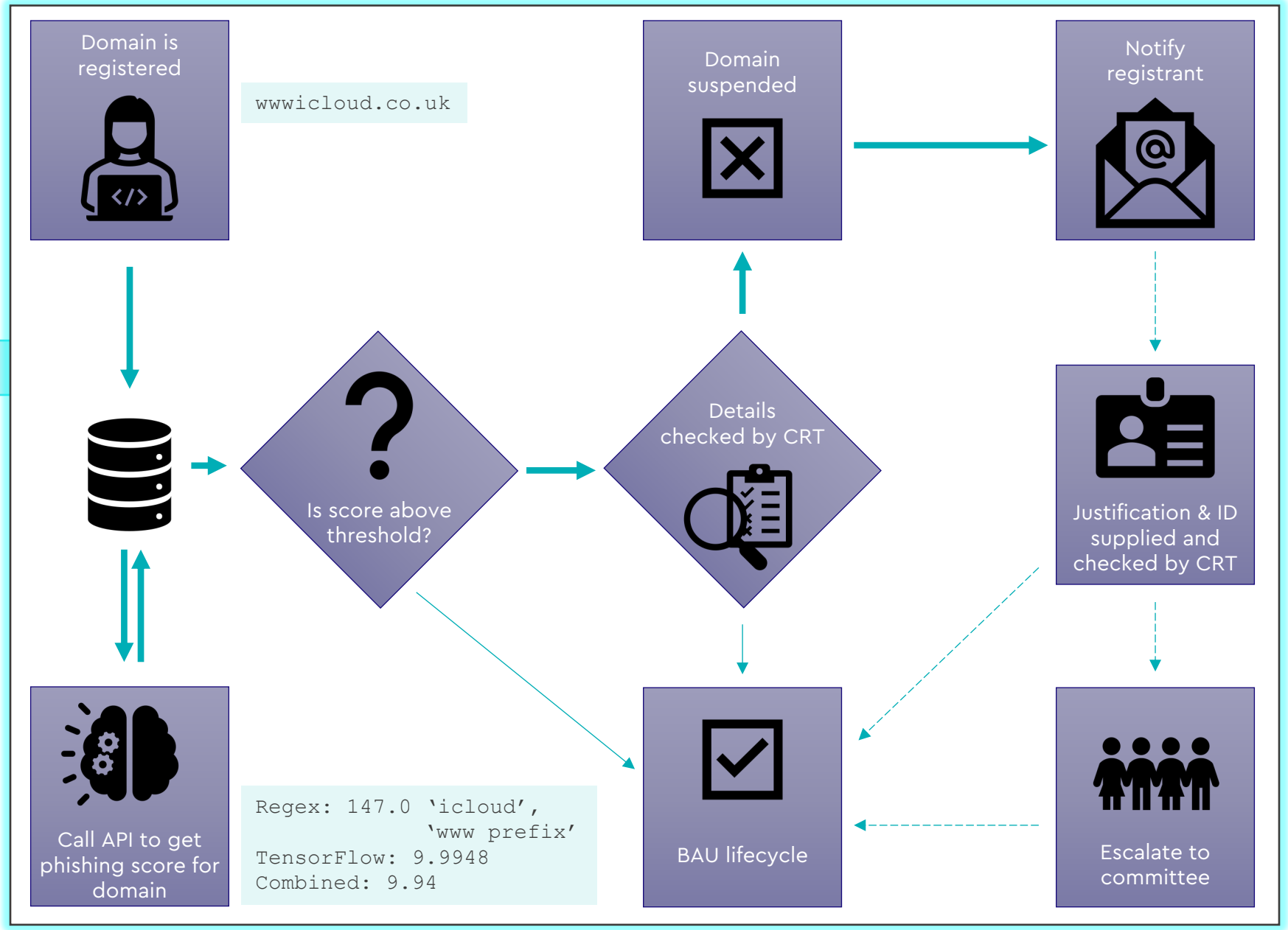
Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans



Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans



Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans

Regular expression model (substrings in domains)

- Manually built
- Checks for a fixed list of terms/brands in domain name
- Allows for some spelling mistakes

TensorFlow model (patterns learnt from data)

- Neural network model built on domain name
- Built from security feed data and registry data (with filtering to improve data quality)
- 98% accuracy on test data (balanced classes), ~60% precision on real data

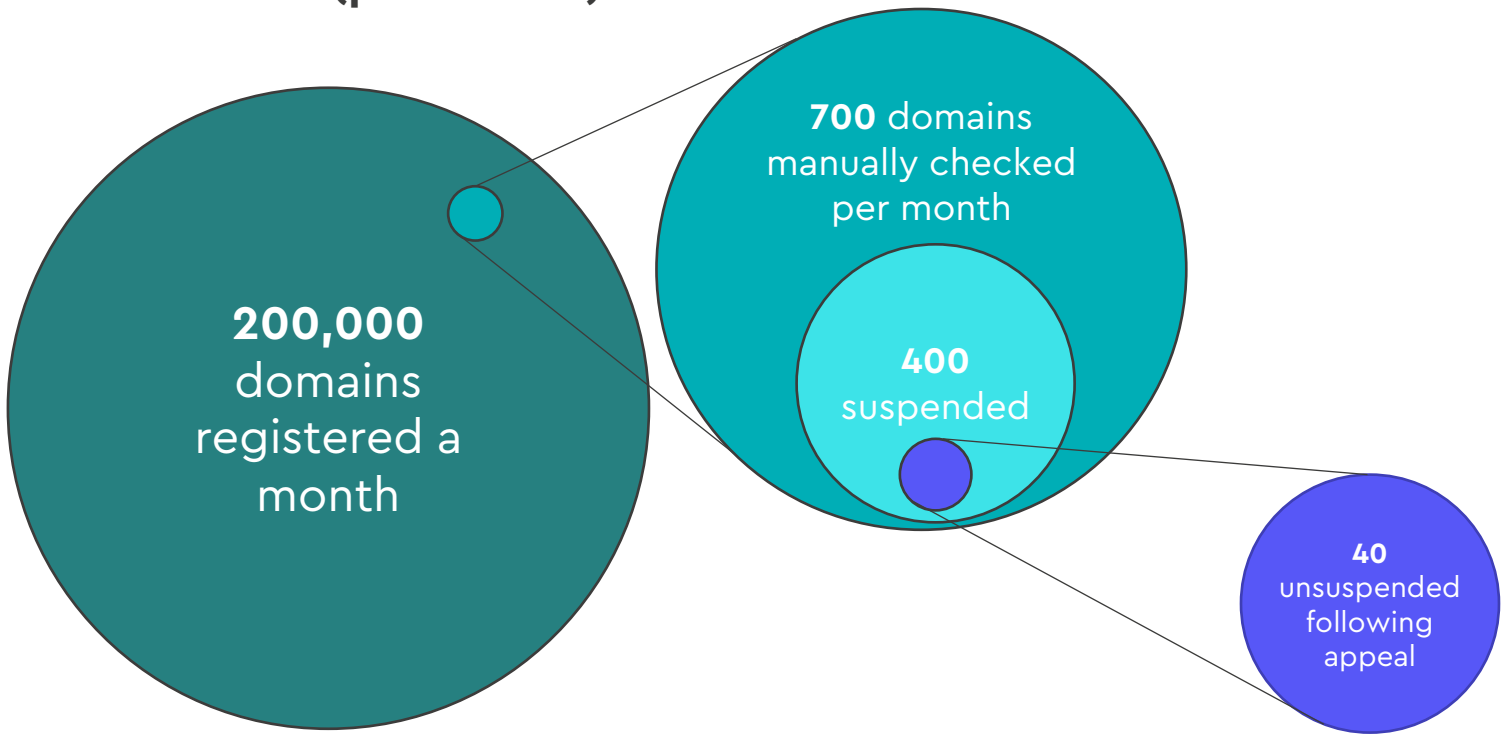
Additional data fields (future developments)

- Important variables include: e-mail address attributes, nameserver, phone number type, registrar, day of week/time of day of registration

Domain Watch

- Background
- Process
- Models
- Performance
- Future Plans

2022 Statistics (per month)



Collecting information:

- Legitimate use cases for suspicious domains (e.g. penetration testers)
- Repeat offenders
- Patterns in registrant data (e-mail providers, registrant country, registrars, terms in domain name)

Domain Watch

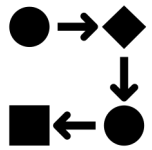
- Background
- Process
- Models
- Performance
- Future Plans

Common Terms
Monitoring frequent tokens/words in suspended registrations

Domain Watch

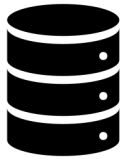
- Background
- Process
- Models
- Performance
- Future Plans

Future plans



Process

- Drop threshold to check more domains
- Trial of taking action against existing registrations



Data/Analysis

- Use web crawler to collect additional data for detecting abuse

Modelling

- Re-implement in new architecture
- Improve model accuracy
- Re-assess how models are combined
- Automate brand detection


Thank you for listening!