Detecting Lookalike Domains

MSc Thesis

Introduction

11th August 2020

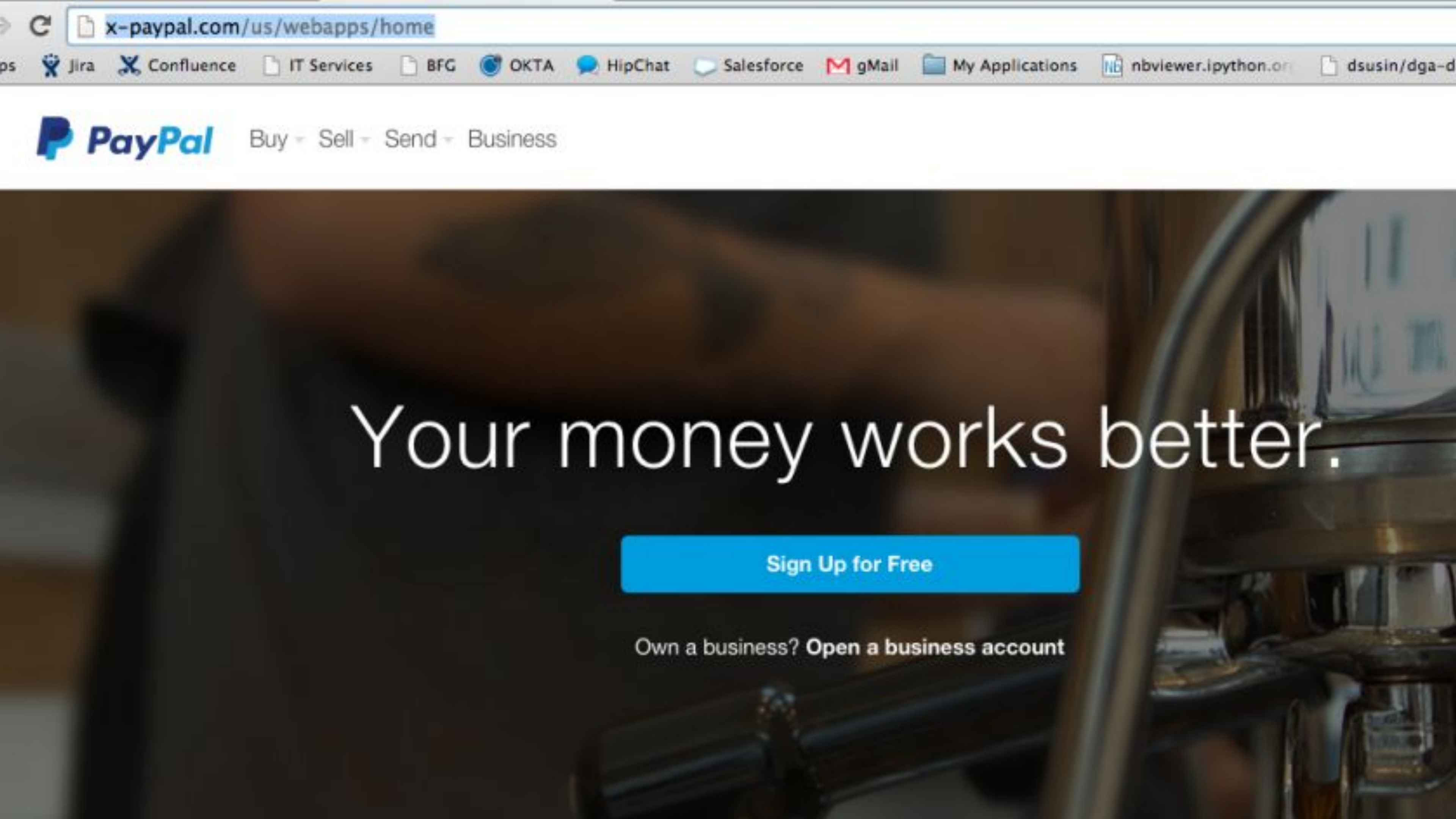
Patrick Frischknecht

Lookalike Domains

Domain names that look, read, or sound similar to other legitimate domains

facebook.com => faceboook.com

The Threats



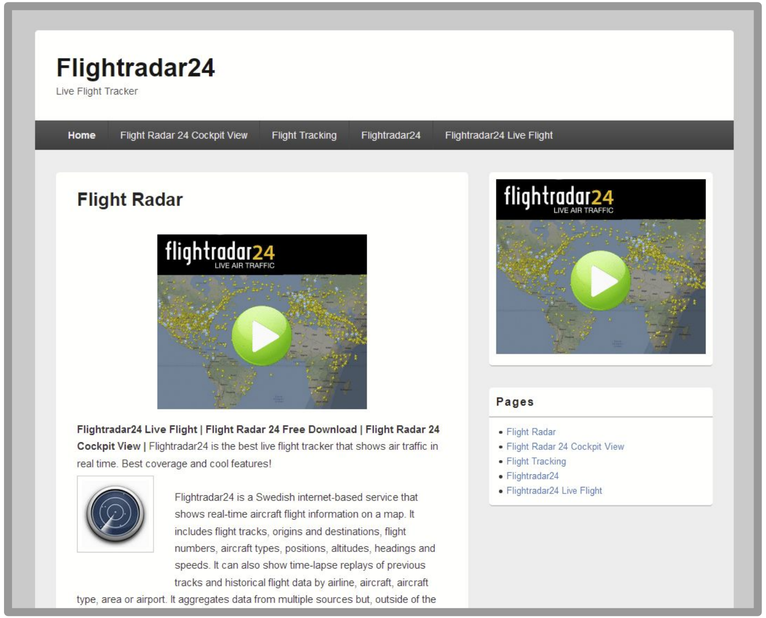
x-paypal.com/us/webapps/home

Your money works better.

Own a business? Open a business account

...lookalike domains are not just about phishing!

- Trademark abuse
- Social engineering
- Benign purpose: additional reach



Flightradar24

Live Flight Tracker

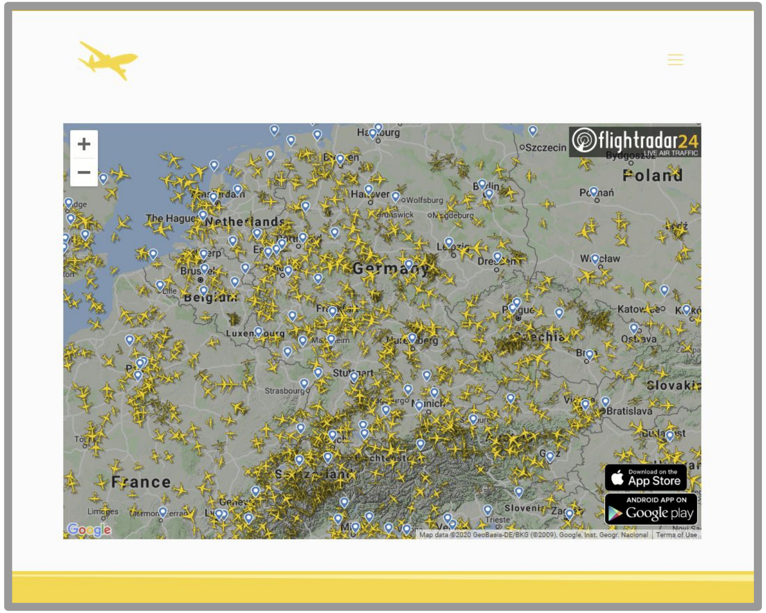
Flight Radar

Flightradar24 Live Flight | Flight Radar 24 Free Download | Flight Radar 24 Cockpit View | Flightradar24 is the best live flight tracker that shows air traffic in real time. Best coverage and cool features!

Flightradar24 is a Swedish internet-based service that shows real-time aircraft flight information on a map. It includes flight tracks, origins and destinations, flight numbers, aircraft types, positions, altitudes, headings and speeds. It can also show time-lapse replays of previous tracks and historical flight data by airline, aircraft, aircraft type, area or airport. It aggregates data from multiple sources but, outside of the

Pages

- Flight Radar
- Flight Radar 24 Cockpit View
- Flight Tracking
- Flightradar24
- Flightradar24 Live Flight



Detection Methods

Website features

URL-related

- Domain name [our focus]
- Domain history
- WHOIS information

Content-related

- HTML code
- Text
- Images [our focus: logos]

URL-related: Domain name squatting

- Attacker tries to impersonate another domain using a similar “lookalike” domain

  facebook.com => faceb00k.com

- Lookalikes reduce the chance a user recognizes the wrong page

Three squatting types are commonly used:
homographs, typosquatting, and combosquatting

Homographs

Using homographs, characters look like others:

facebook.com => faceb00k.com

apple.com => äpple.com

Detection:

Generative: generate all homograph domains and check

Threshold on string edit distance

Typosquatting

Assuming typos like missing or mistyped characters on the keyboard:

facebook.com => facebool.com

apple.com => aple.com

Detection:

Generative: generate all homograph domains and check

Threshold on “fat finger distance”

Combosquatting

Assuming word recombinations:

facebook.com => facebook-site.com

apple.com => applestore.com

Detection:

Partial matching (?)

Difficulty: Partial matching

attack.com => contains “att”, really a lookalike?

bikeandride.com => contains “ikea”

ORGANIC
NORFOLK FARMHOUSE
SCRUMPY
CALL IN FOR A
FREE SAMPLE
THE APPLE SHOP
KINGFISHER
FARM CIDER
100% PURE
NORFOLK
PEAR
WITH
APPLE
JUICE
70p A CUP
NORFOLK
CIDER CO.
PURE PRESSED
APPLE JUICE
WARM SPICED
CIDER
PROPER CIDER
IT'S
NEW,
IT'S
DELICIOUS,
IT'S
NORFOLK
PEAR CIDER
7.5% ALC!

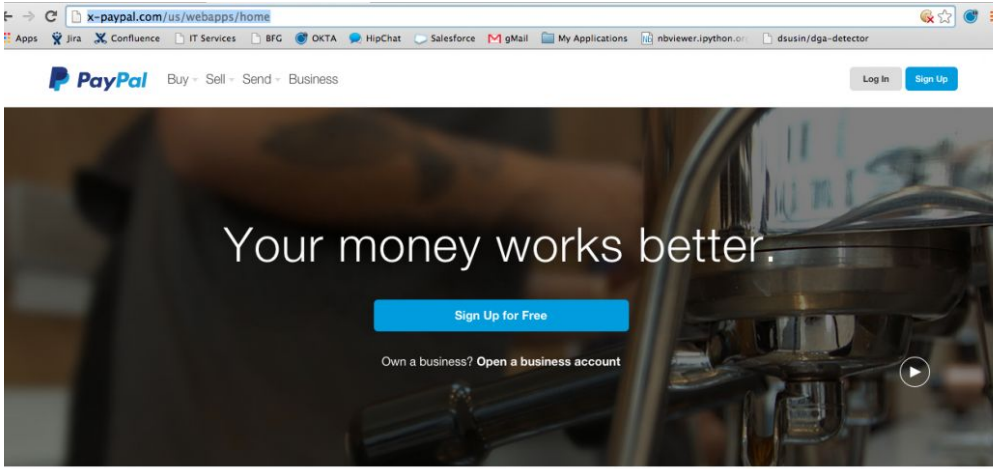
Difficulty: Ambiguity

How to handle false positives and ambiguity?

Check the website content!

Logo Detection

present in?



x-paypal.com/us/webapps/home

Your money works better.

Own a business? **Open a business account**

Is template present in image?

The Idea

Project overview

- User provides at least:
  logo image and **domain name**

- Our tool does:
  1) **generate** a list of **lookalike domains**
  2) **validate** the list and **capture a screenshot** of each website
  3) **search for similarities** in the screenshot
  4) **report** the results for manual review

The Current State

+ Difficulties

Project overview

Our tool currently does:

1) **generate** a list of **lookalike domains**

2) **validate** the list and **capture a screenshot** of each website

3) **search for similarities** in the screenshot

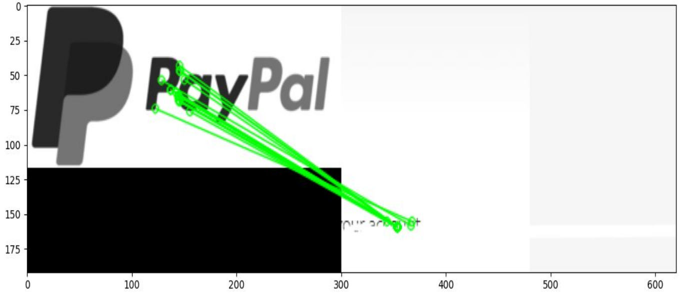
4) **report** the results for manual review

Difficulty: Logo detection

Experienced poor matching performance on websites

=> looking for better algorithms

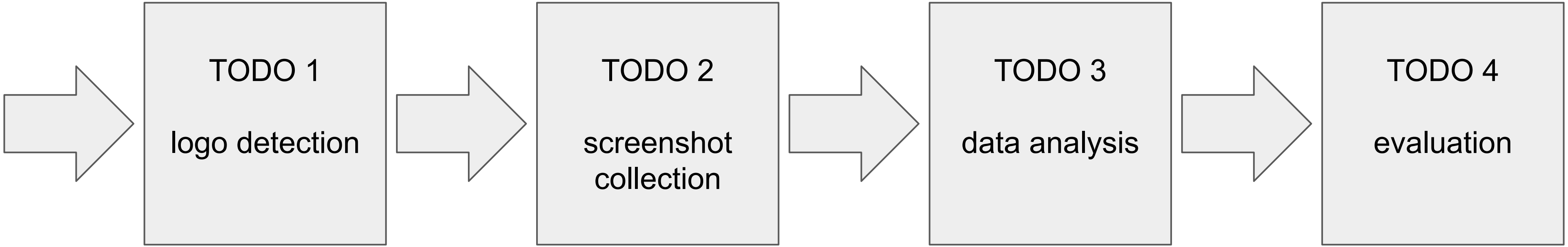
SIFT with sliding window



- detects and matches feature vectors in two images

(used in other papers)

The Next Steps



TODO 1

logo detection

TODO 2

screenshot collection

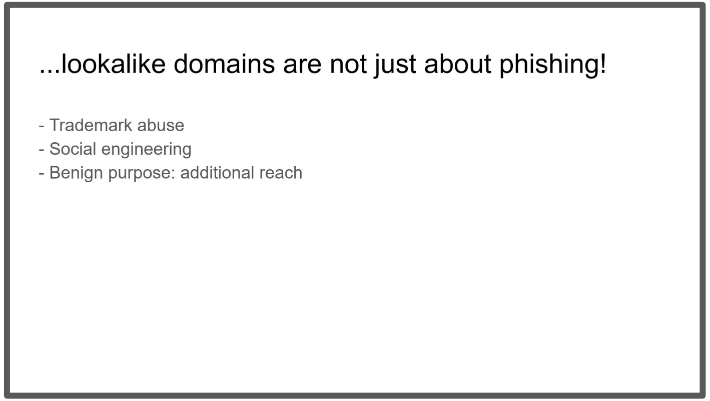
TODO 3

data analysis

TODO 4

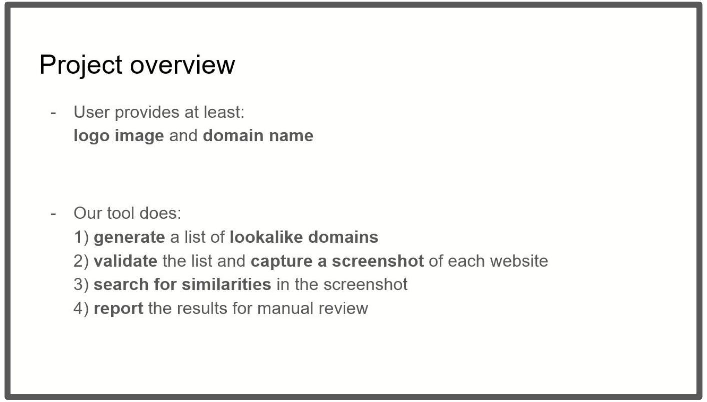
evaluation

Recap



...lookalike domains are not just about phishing!

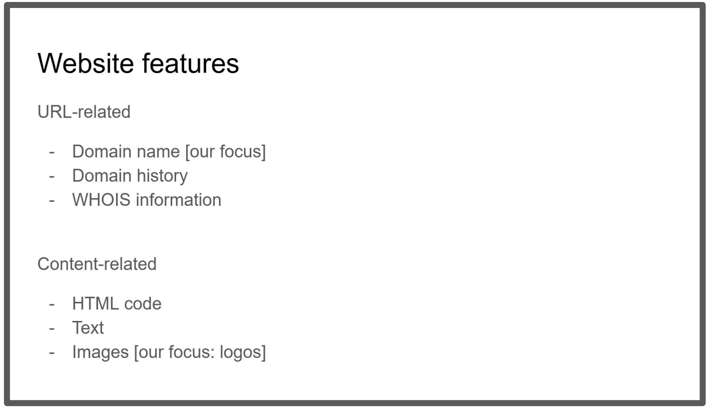
- Trademark abuse
- Social engineering
- Benign purpose: additional reach



Project overview

- User provides at least:
  logo image and **domain name**

- Our tool does:
  1) **generate** a list of **lookalike domains**
  2) **validate** the list and **capture a screenshot** of each website
  3) **search for similarities** in the screenshot
  4) **report** the results for manual review



Website features

URL-related

- Domain name [our focus]
- Domain history
- WHOIS information

Content-related

- HTML code
- Text
- Images [our focus: logos]

ORGANIC NORFOLK FARMHOUSE SCRUMPY
CALL IN FOR A FREE SAMPLE
THE APPLE SHOP
KINGFISHER FARM CIDER
IT'S NEW, IT'S DELICIOUS, IT'S NORFOLK PEAR CIDER 7.5% ALC!