RSAConference2019
San Francisco | March 4–8 | Moscone Center

SESSION ID: MLAI-T07

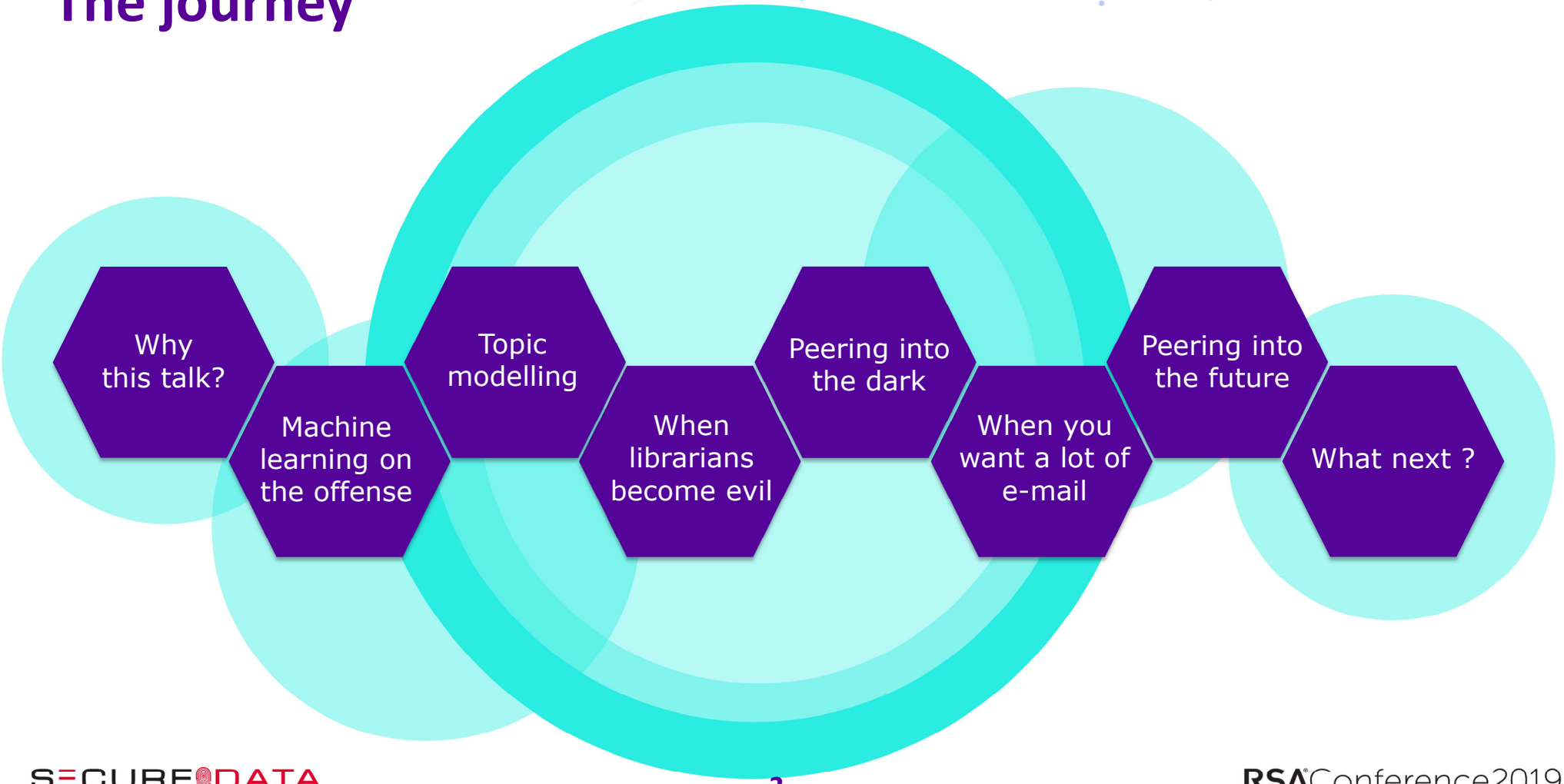
Rise of the machines
AI & ML based attacks demonstrated

Etienne Greeff and Wicus Ross

SecureData

@etienne_greeff
@wicusross

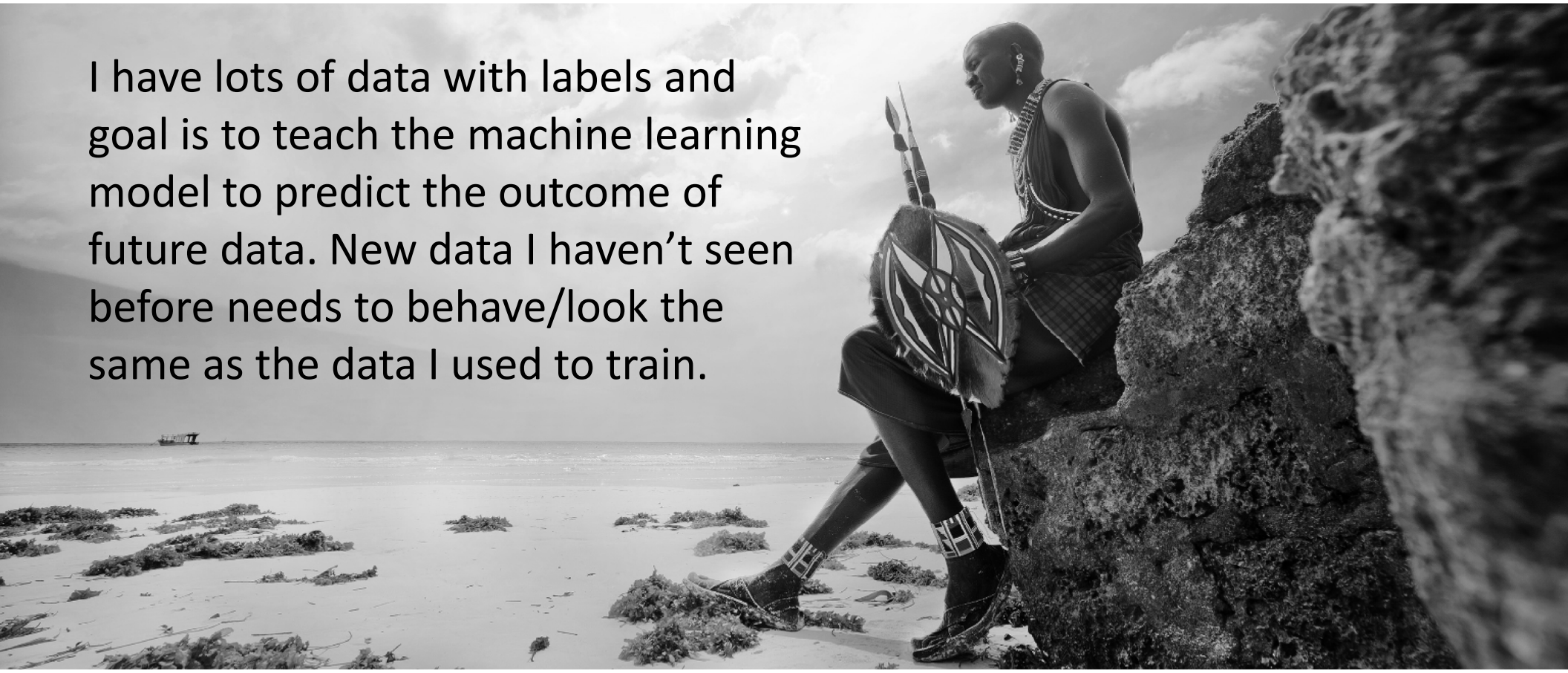
The journey


Why this talk?
Machine learning on the offense
Topic modelling
When librarians become evil
Peering into the dark
When you want a lot of e-mail
Peering into the future
What next ?

Why this talk ?

AI & Machine Learning (ML) is more suited to Offensive than Defensive applications.

Supervised learning

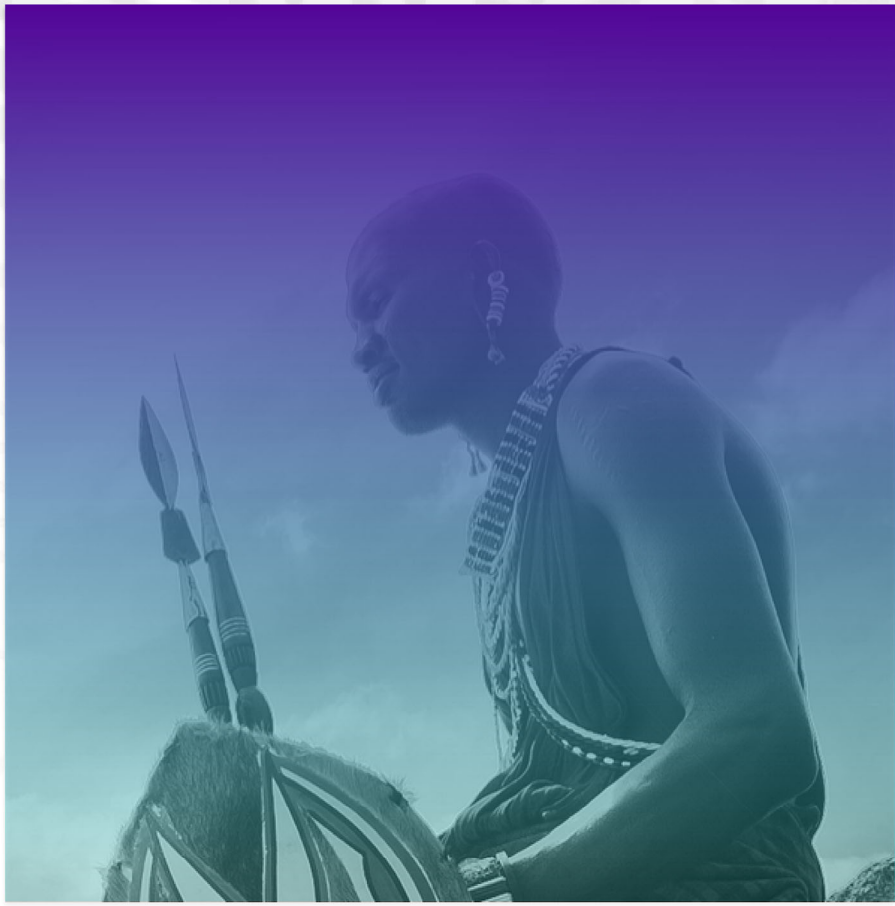


I have lots of data with labels and goal is to teach the machine learning model to predict the outcome of future data. New data I haven't seen before needs to behave/look the same as the data I used to train.

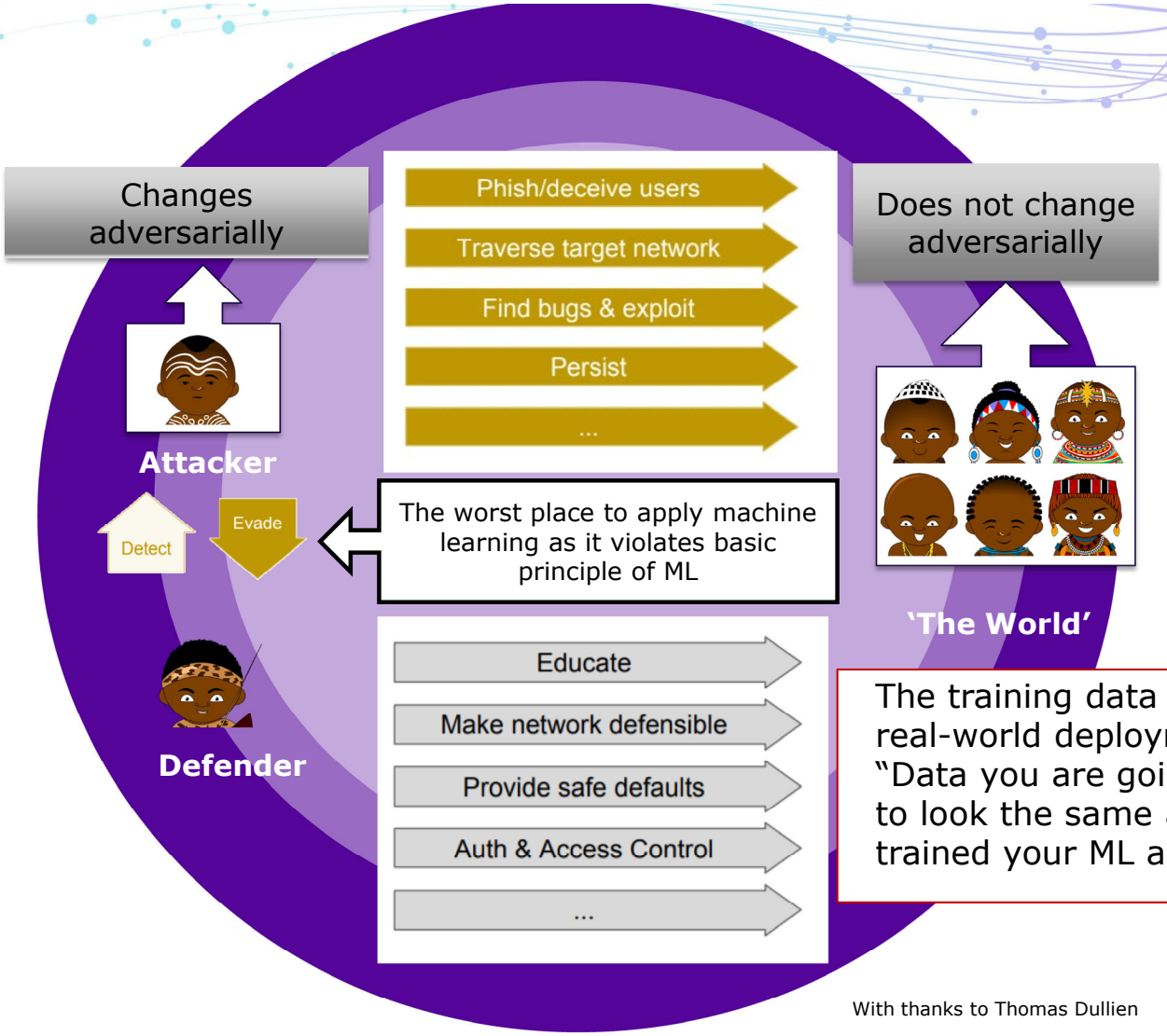
Unsupervised learning

I have lots of data but no labels …

For example: trying to find structure in text contained within files or e-mails i.e., topic modelling.




More suited to offensive than
defensive applications you say?



The training data will resemble their real-world deployment target i.e., "Data you are going to work on needs to look the same as the data you trained your ML algorithm on."

With thanks to Thomas Dullien

So where are these offensive applications?

A very short list of potential offensive uses

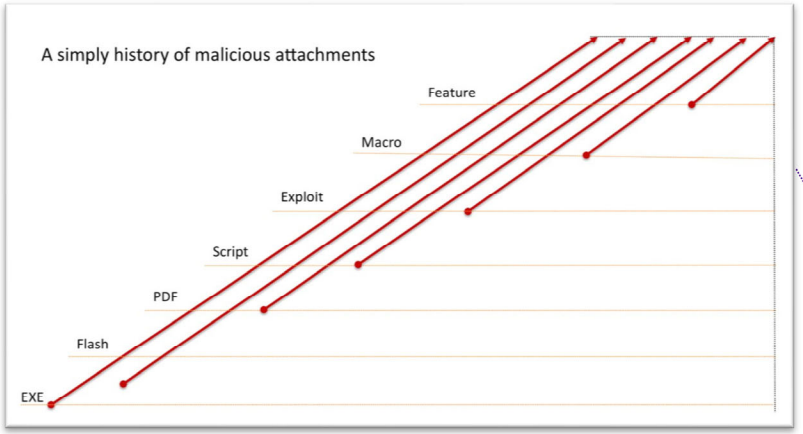
- DeepPhish - Using deep learning to bypass AI-based phishing detection
- Fooling deep learning-based image recognition
- Web application attacks using reinforcement learning
- Bug hunting in libraries using deep learning

Mostly in the realm of theory and proof of concept…

Current state of art seems to attempt to automate current attacks

We love features don't we?

Office



Feature

Macro

OLE

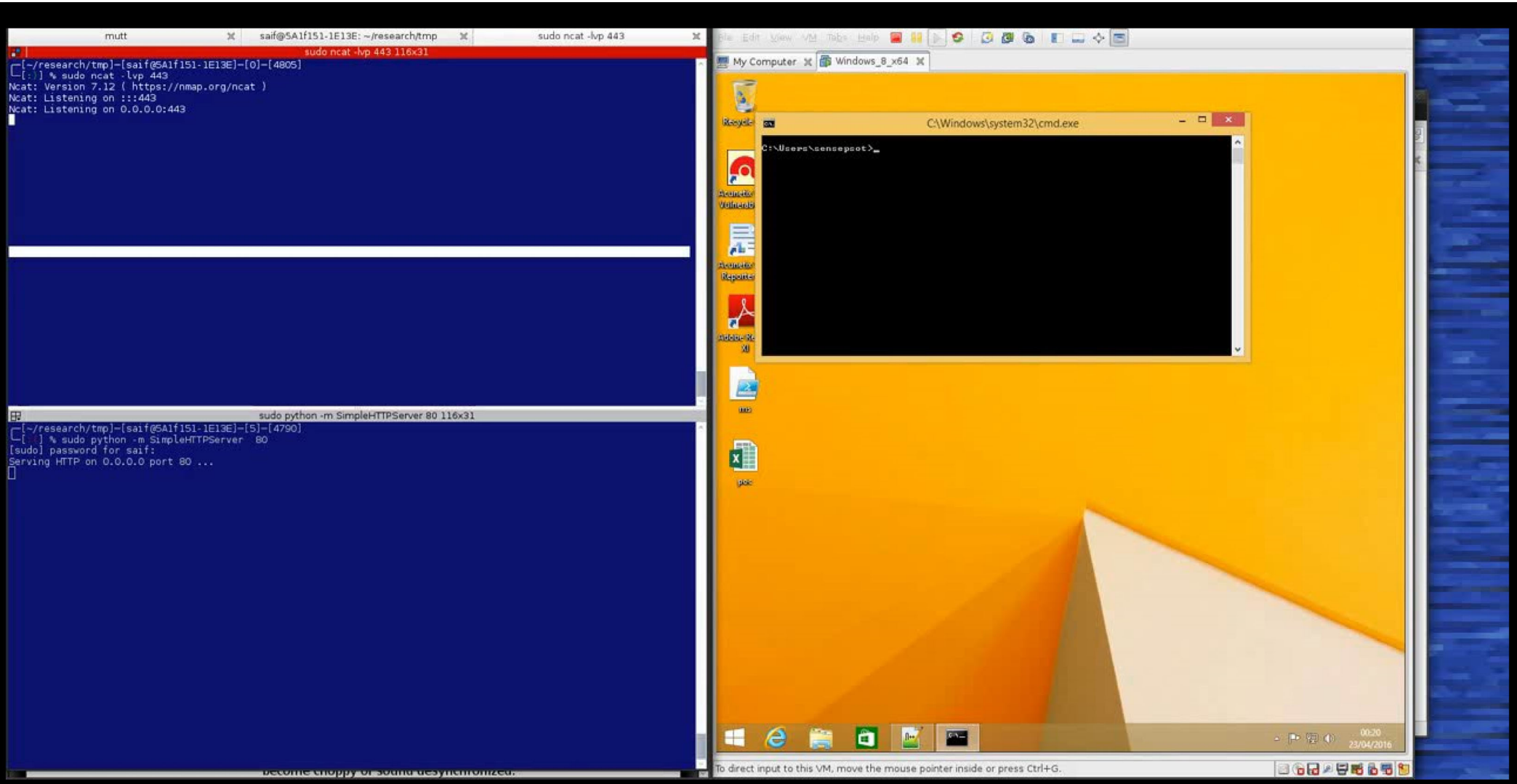
Malicious Rules and Forms

Responder

DDE

...


DDE : It’s a feature not a bug


mutt
saif@5A1f151-1E13E: ~/research/tmp
sudo ncat -lvp 443
sudo ncat -lvp 443 116x31
[~/research/tmp]-[saif@5A1f151-1E13E]-[0]-[4805]
[:] % sudo ncat -lvp 443
Ncat: Version 7.12 (https://nmap.org/ncat)
Ncat: Listening on :::443
Ncat: Listening on 0.0.0.0:443
sudo python -m SimpleHTTPServer 80 116x31
[~/research/tmp]-[saif@5A1f151-1E13E]-[5]-[4790]
[ ] % sudo python -m SimpleHTTPServer 80
[sudo] password for saif:
Serving HTTP on 0.0.0.0 port 80 ...
My Computer
Windows_8_x64
C:\Windows\system32\cmd.exe
C:\Users\sensepost>_
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.

Microsoft introduces Machine Learning as a feature …

Hmm what could possibly go wrong?

Introducing topic modelling

Models of text data

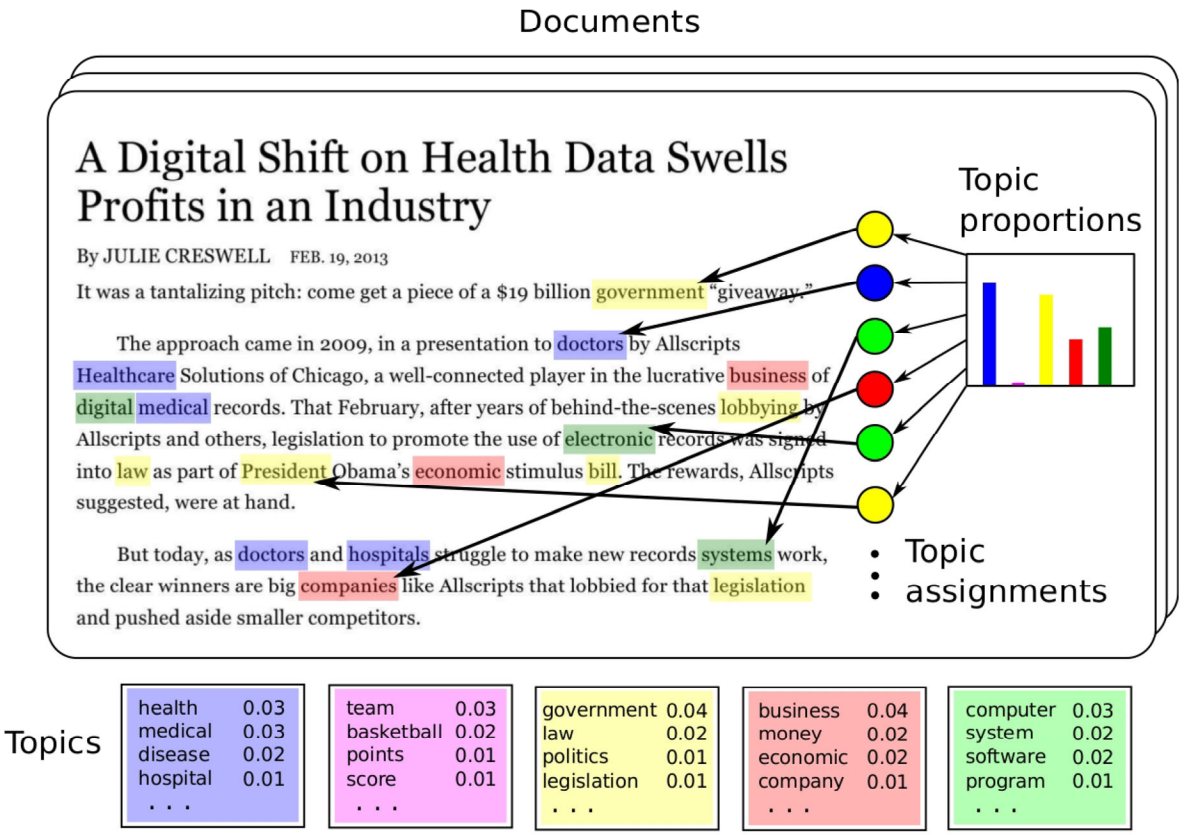
- Given text data we want to:
  - Organize
  - Visualize
  - Summarize
  - Search
  - Predict
  - Understand

- Topic models allow us to:
  - Discover themes in text
  - Annotate documents
  - Organize, summarize, etc.

Topic modelling

A probabilistic model

- Learns distributions on words called “topics” shared by documents
- Learns a distribution on topics for each document
- Assigns every word in a document to a topic



With thanks to Prof John Paisley Columbia University

Topic modelling

The New York Times

music band songs rock album jazz pop song singer night	book life novel story books man stories love children family	art museum show exhibition artist artists paintings painting century works	game knicks nets points team season play games night coach	show film television movie series says life man character know
theater play production show stage street broadway director musical directed	clinton bush campaign gore political republican dole presidential senator house	stock market percent fund investors funds companies stocks investment trading	restaurant sauce menu food dishes street dining dinner chicken served	budget tax governor county mayor billion taxes plan legislature fiscal

Topic modelling outputs two main things:

- A set of distributions on words (topics). Shown above are ten topics from NYT data. We list the ten words with the highest probability.
- A distribution on topics for each document (not shown). This indicates its thematic breakdown and provides a compact representation.

When librarians become evil

What if I could use topic modelling offensively?

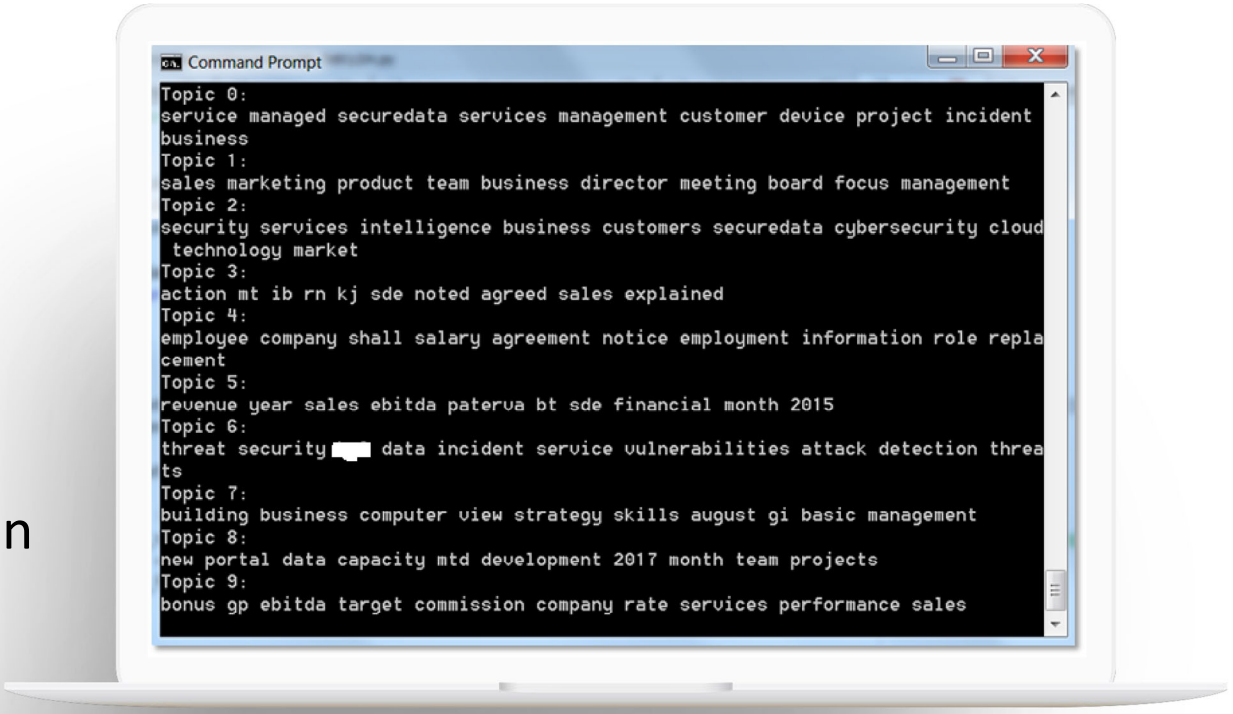
- Effectively map a network telling me where I should focus my malicious efforts
- Industrialise typo squatting

EXAMPLE 1:

Peering into the dark, creating an accurate data map of a network

Topic modelling for fun and profit on desktops

- My work machine Desktop
- 2 minutes, 800 files
- Scarily accurate
- Unparalleled insight
- Identify valuable information



Using Cobalt Strike to get to the juicy stuff in an automated scalable way

EXAMPLE 2:
When you WANT a lot of e-mail
aka Typo Squatting

Typosquatting

1 Select an Industry

Dubbed 'Friday Afternoon Fraud', the conveyancing scam has been known to take several forms, but generally occurs when the hackers intercept emails between home buyers or sellers, and their solicitors.

They generate lookalike emails which allow them to pose as the solicitor involved.
During the final stages of a property purchase or sale, they inform potential victims by email that certain bank account details have changed.

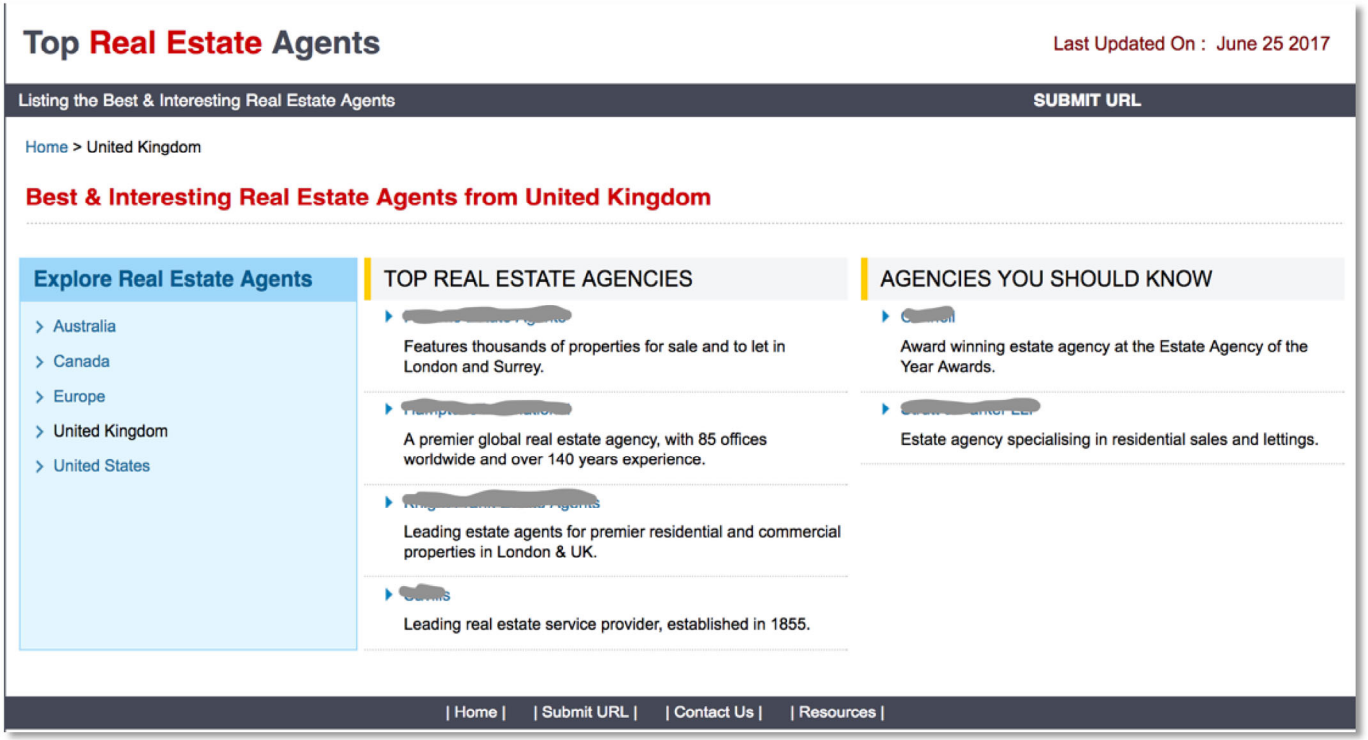
Home buyers stand to lose thousands in new cyberattack

Cybercriminals are hacking the email accounts of Irish solicitors in an attempt to steal tens of thousands of euro from unsuspecting home buyers, the Sunday Independent has learned. Stock photo: PA

Mark O'Regan
February 5 2017 2:30 AM

Typosquatting

2 Enumerate the players



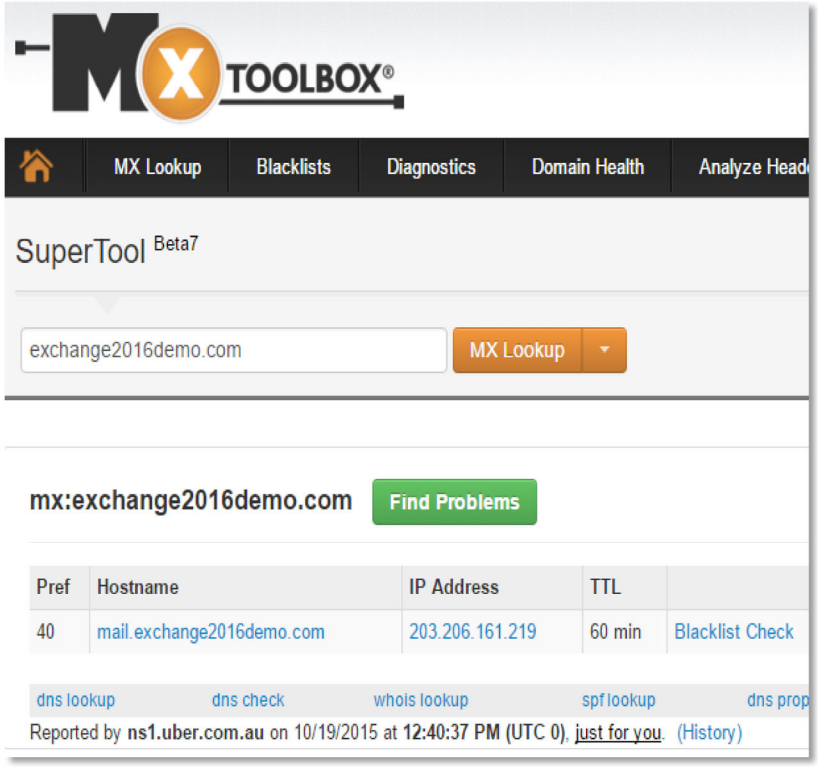
Typosquatting

3 Generate typos

- Skip letter
- Double letters
- Reverse letters
- Missed key
- Inserted key
- Phonemes and graphemes

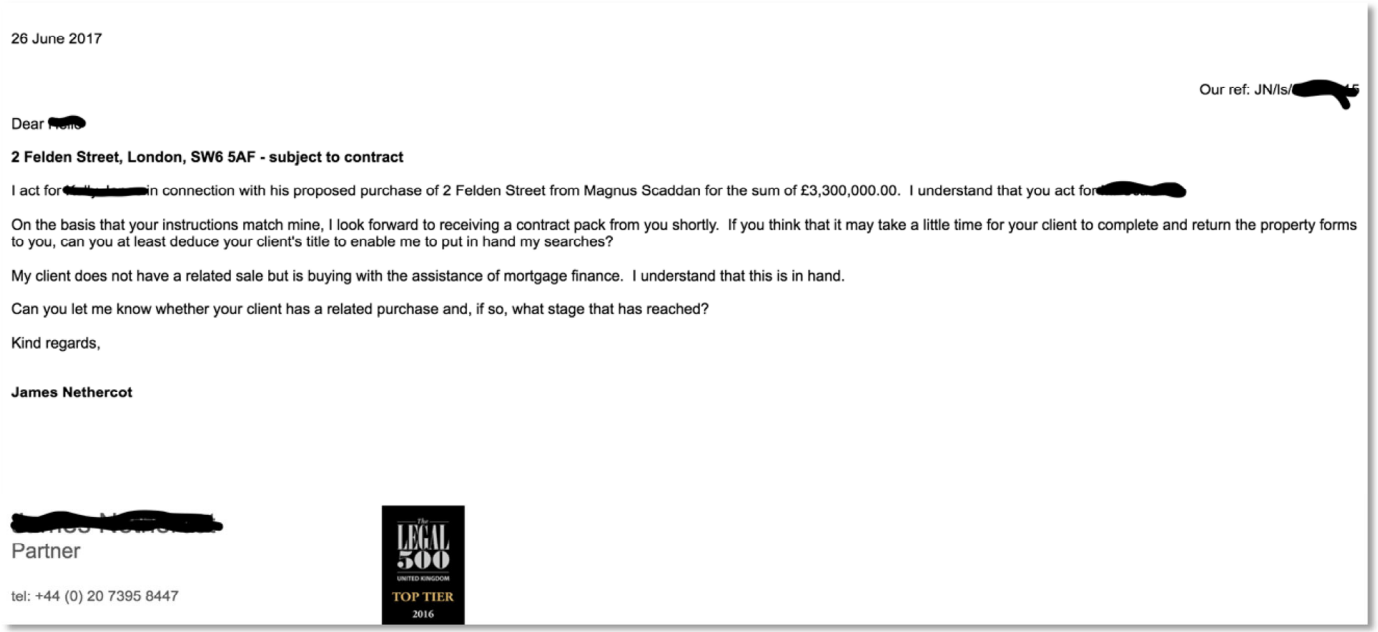
Typosquatting

4 Register the domains & setup mail server & wait just wait



Typosquatting

5 Great success!



26 June 2017

Our ref: JN/ls/[redacted]

Dear [redacted]

2 Felden Street, London, SW6 5AF - subject to contract

I act for [redacted] in connection with his proposed purchase of 2 Felden Street from Magnus Scaddan for the sum of £3,300,000.00. I understand that you act for [redacted]

On the basis that your instructions match mine, I look forward to receiving a contract pack from you shortly. If you think that it may take a little time for your client to complete and return the property forms to you, can you at least deduce your client's title to enable me to put in hand my searches?

My client does not have a related sale but is buying with the assistance of mortgage finance. I understand that this is in hand.

Can you let me know whether your client has a related purchase and, if so, what stage that has reached?

Kind regards,

James Nethercot

[redacted]
Partner

tel: +44 (0) 20 7395 8447

The LEGAL 500 UNITED KINGDOM TOP TIER 2016

But how do I scale this?

- Requires manual investigation of potentially 100s of mailboxes and 1000s of e-mails
- Use topic modelling to identify key mailboxes and key topics
- Examine all new email for topics of interest
- What if I could look for mailboxes that discuss
  - Financial information ?
  - Patent information ?
  - Confidential information ?

PEERING INTO THE FUTURE

Key takeaways

- Understand the new threat models that AI & ML may introduce
- Important to look at endpoint telemetry; i.e., try and spot feature based attacks
- Need to understand where data lives and how an attacker might see it
- Topic modelling and related techniques are powerful attack vectors creating new classes of attacks
- We believe that offensive applications of machine learning are very plausible and possible
- Have a response plan ready. No really !

Future work

- Automate typo squatting attacks with topic modelling using orchestration to automatically harvest interesting documents – small eek !
- Use topic modelling in defensive applications to map your network to have a real time view of where documents of interest reside
- Use topic modelling to make forum/github (no darkwebs here) information mining more productive i.e., what topics are discussed when my company is mentioned
- Start using built in libraries in Windows 10 for attacks without having to deploy a malicious payload


THANK YOU
Questions?
@etienne_greeff
@wicusross

RSA Conference2019