Breaking Big Data

Evading Analysis of the Metadata of Your Life

Dave Venable
Masergy Communications
@davevenable


black hat
EUROPE 2016


hooli

Big Data

Data sets that are so large or complex that traditional data processing applications are inadequate to deal with them.

Source: Wikipedia

What’s being collected?


COLLECT
ALL THE THINGS
memegenerator.net


...and figure out what to do with it later

Will Big Data Cure Cancer?

Big data technologies will play a key role in diversifying Saudi Arabia's economy away from a huge dependence on oil revenues.

How the Chicago Cubs Used Big Data to Win the World Series

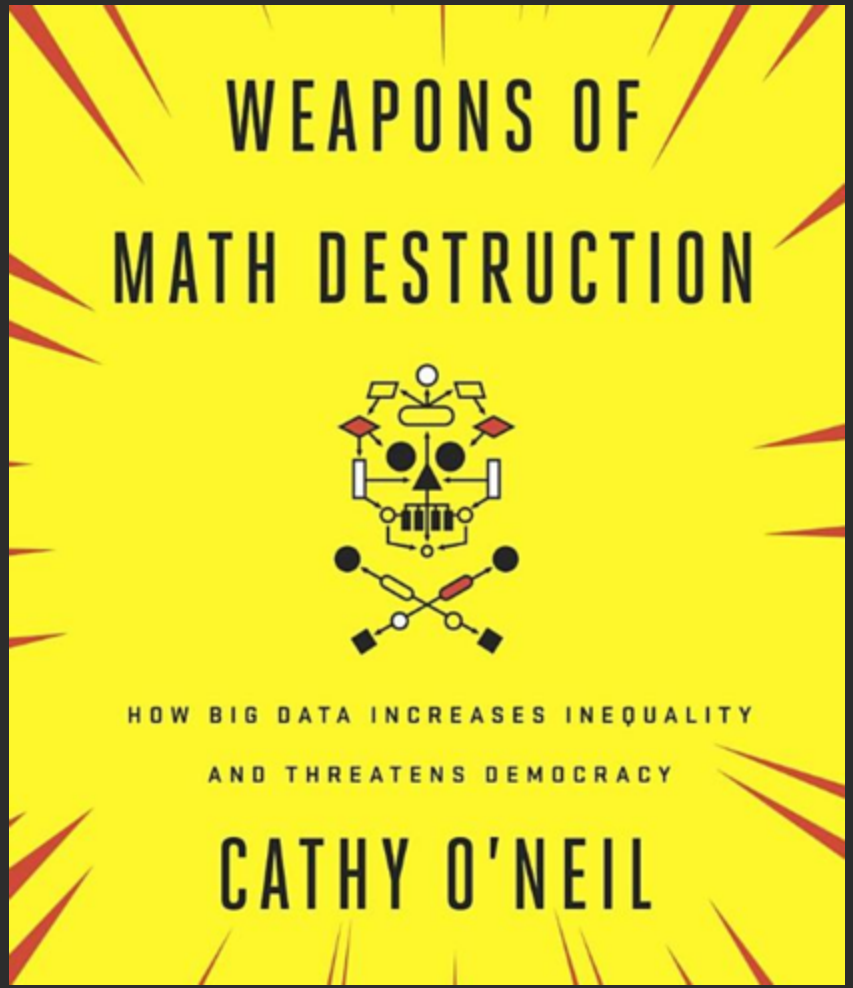
Data + Algorithms = **Big Data**

$3.1T

Source: IBM

Bad

Data + Algorithms = Big Data

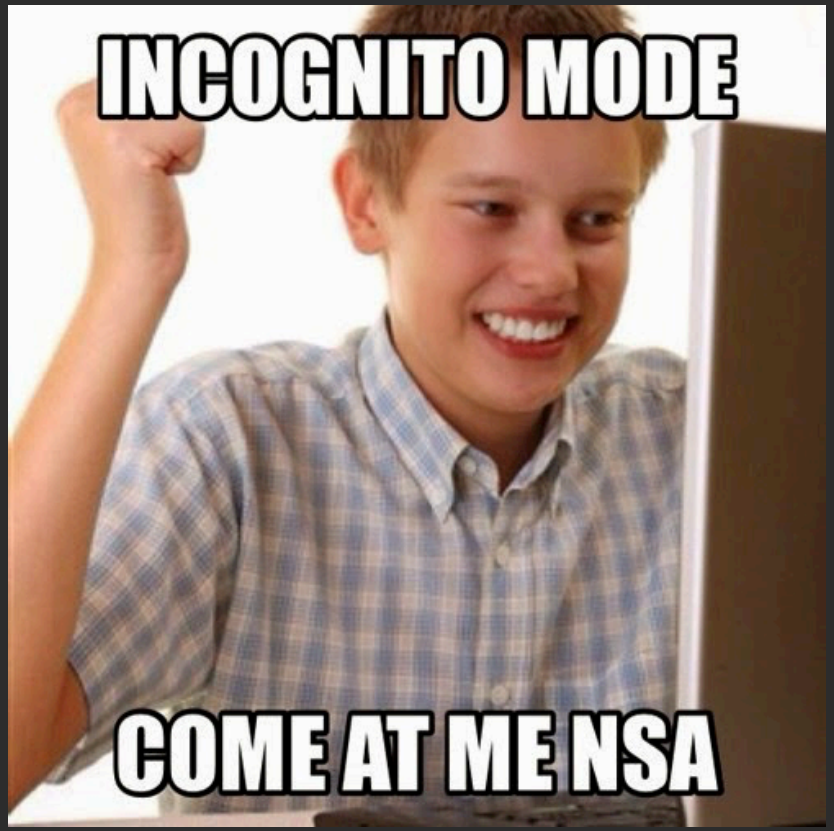


Bad Biased

Data + Algorithms = Big Data

Bad Biased Delusions

Data + Algorithms = Big ~~Data~~


INCOGNITO MODE
COME AT ME NSA

Encryption Works

“Properly implemented strong crypto systems are one of the few things that you can rely on.”

- Edward Snowden

But so what?

Email Content

Messaging

Phone Calls

Packet Payloads

HTTPS Content

Sender

Recipient

Participants

Length

Source

Destination

Encrypted

Alice and Bob


British Airways → Alice


Alice is in Seoul


“Seoul is so fun!”


Bob → Cindy


Cindy → Bob


Bob → Hotel


Cindy → Hotel

Bob - £500.00 Hotel


Hotel → Restaurant

Bob - £230.00 Restaurant


Restaurant → Hotel


Hotel → Bob’s


Hotel → Cindy’s


Bob → Cindy


Cindy → Bob

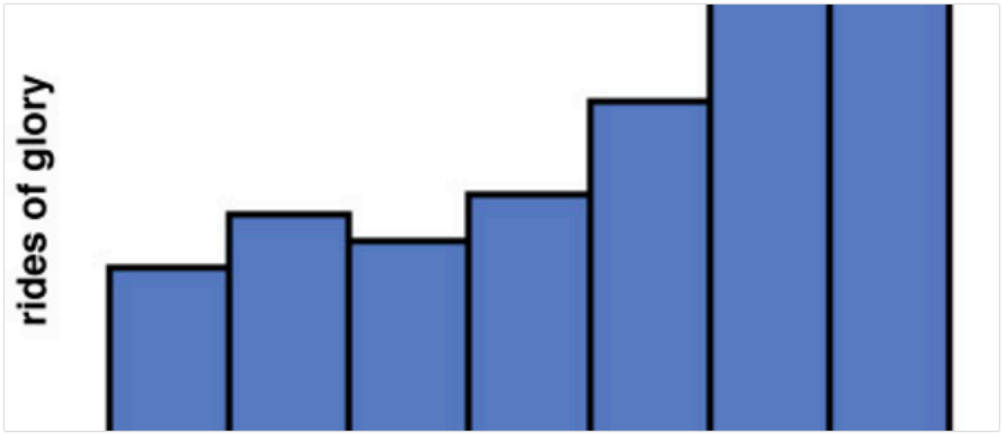

Alice is in London

#UBERDATA

RIDES OF GLORY

MARCH 26, 2012

POSTED BY VOYTEK



Original: https://goo.gl/t08BiP

Encryption Works

“Properly implemented strong crypto systems are one of the few things that you can rely on.”

- Edward Snowden

Encryption Works

“Properly implemented strong crypto systems are one of the few things that you can rely on. Unfortunately, endpoint security is so terrifically weak that NSA can frequently find ways around it.”

- Edward Snowden

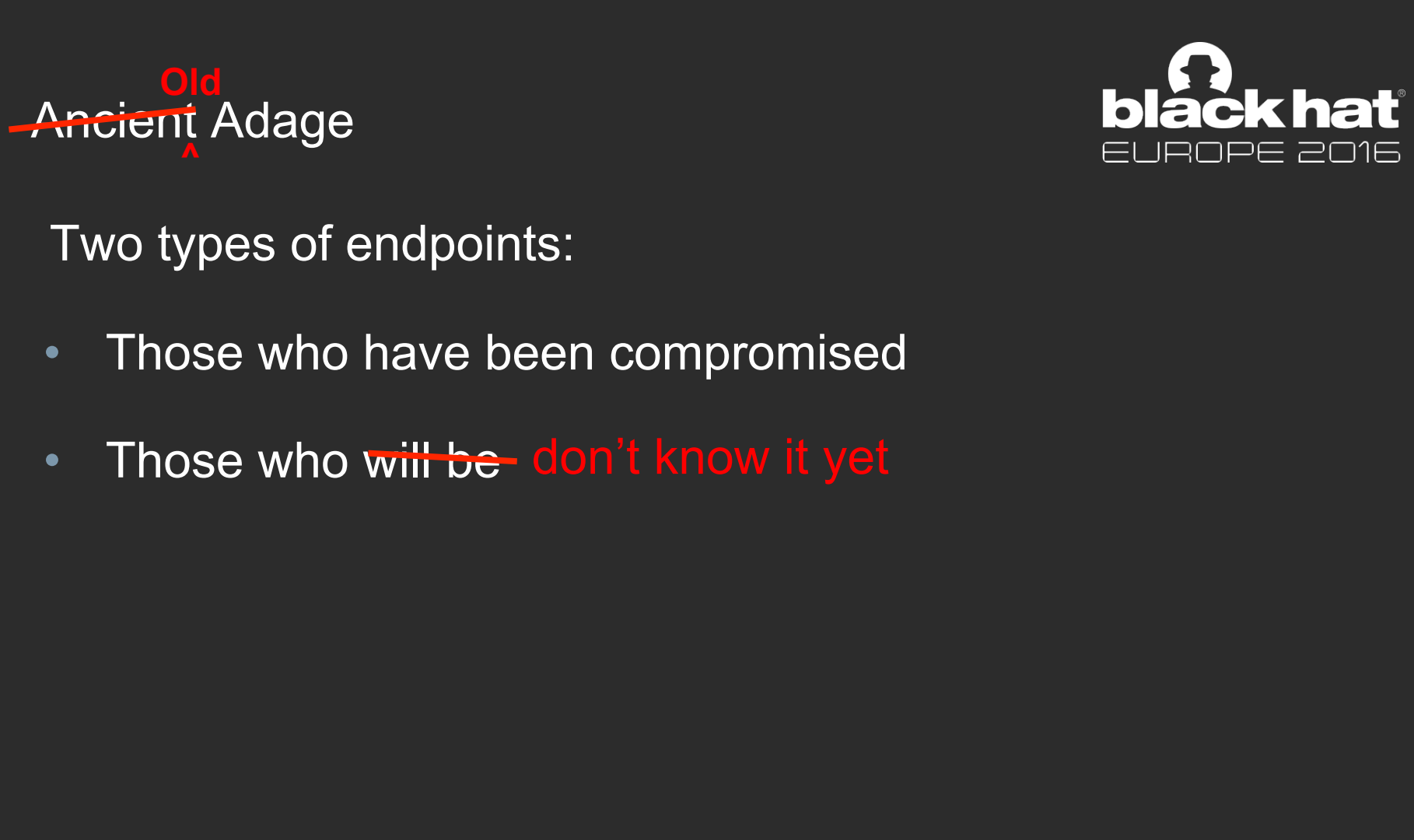
Endpoints Suck

Ancient Adage

Two types of endpoints:

- Those who have been compromised
- Those who will be

~~Ancient~~ Old Adage



Two types of endpoints:

- Those who have been compromised
- Those who ~~will be~~ don’t know it yet

New ~~Old~~

Ancient Adage

One

Two types of endpoints:

- Those who have been compromised
- Those who will be - don't know it yet

New ~~Old~~ Ancient Adage

One ~~Two~~ types of endpoints:

- ~~Those who have been compromised~~
- ~~Those who will be~~ ~~don't know it yet~~

Everything is compromised

...and what isn’t has permission.

Bob

9:00 pm / 21:00

Home → Pub

Bob is driving to the pub (too fast)

Check in at pub

6:30 am / 06:30

Driving to work (too carelessly)

2:00 am / 02:00

Bob - £100.00 @ pub

Pub → Home

Bob is driving home (too slow)

Insurance App

Your life has patterns

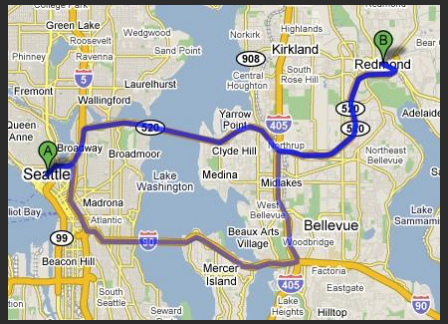
...that produce metadata

BBC

Your habits

#irc

Your locations



black hat
EUROPE 2016

Google photography

All News Images Maps Books More Search tools

About 1,450,000,000 results (0.65 seconds)

Your interests

Kali Linux Forums
Welcome to the Kali Linux Forums.

Kali Linux Forums
Kali Linux Official Forums

[root@netsec /]# hot new rising

Bob ↔ Cindy

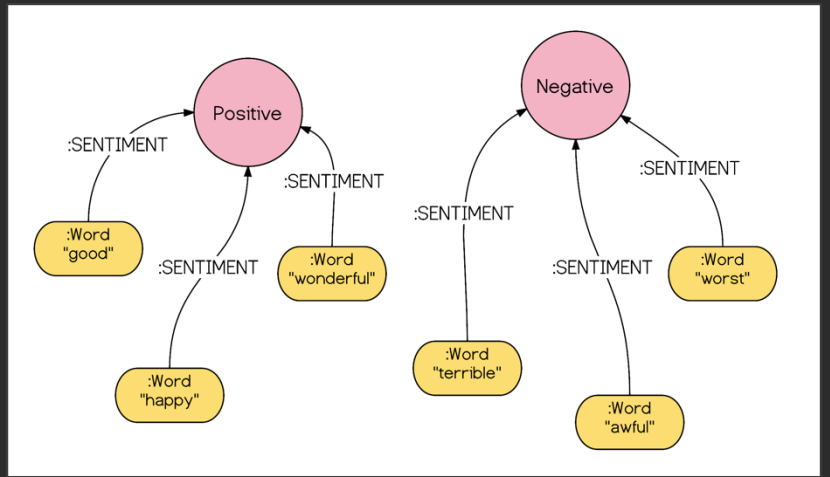
Same place, same time

Your friends

[Candidate 1] sucks

Loving this coffee shop

Your feelings



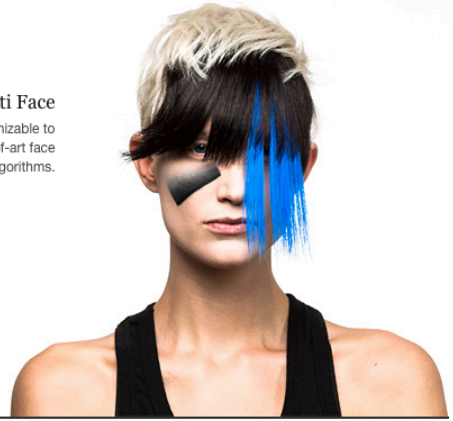
Face

Once computer vision programs detect a face, they can extract data about your emotions, age, and identity.

See how a face is detected

Anti Face

This face is unrecognizable to several state-of-art face detection algorithms.



Camouflage from face detection.

CV Dazzle explores how fashion can be used as camouflage from face-detection technology, the first step in automated face recognition.

From all appearances, deception has always been critical to daily survival—for human and non-human creatures alike—and,

GAO U.S. Government Accountability Office

Reports & Testimonies | Bid Protests & Appropriations Law | Key Issues | About GAO | Caree

FACE RECOGNITION TECHNOLOGY:

FBI Should Better Ensure Privacy and Accuracy [Reissued on August 3, 2016]

GAO-16-267: Published: May 16, 2016. Publicly Released: Jun 15, 2016.

HIGHLIG

What

The Dep
Interstat
of over
enforce
from a s
return a
2015, th

ope

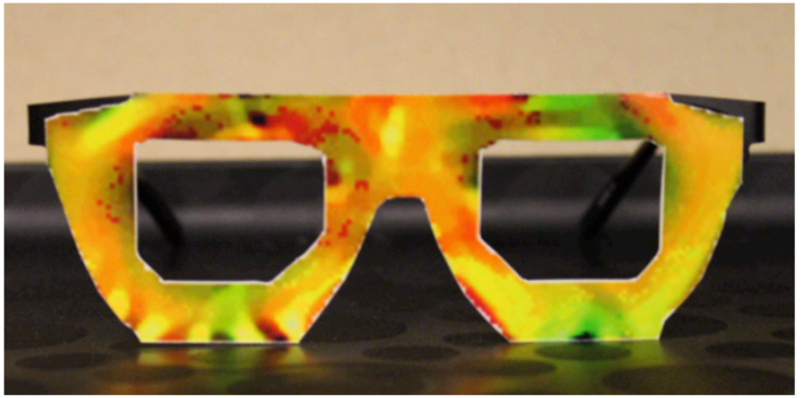
ch conside

ACY & SECURITY

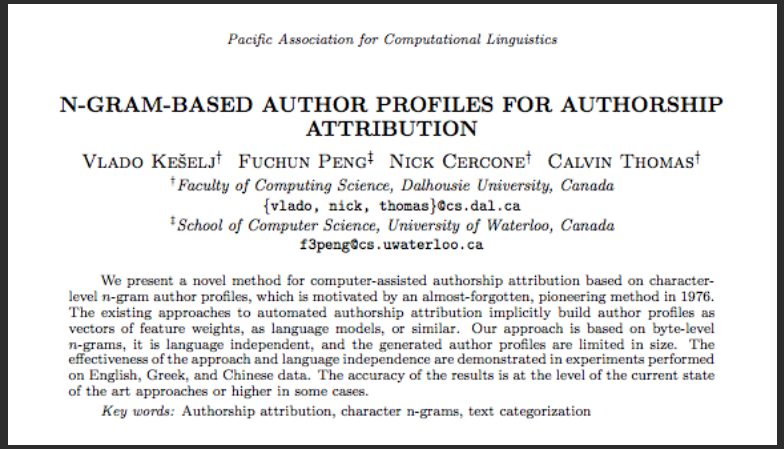
Facebook'
Is Differen

May 18, 2016 · 9:30 AM ET

NAOMI LACHANCE



black hat EUROPE 2016



Pacific Association for Computational Linguistics

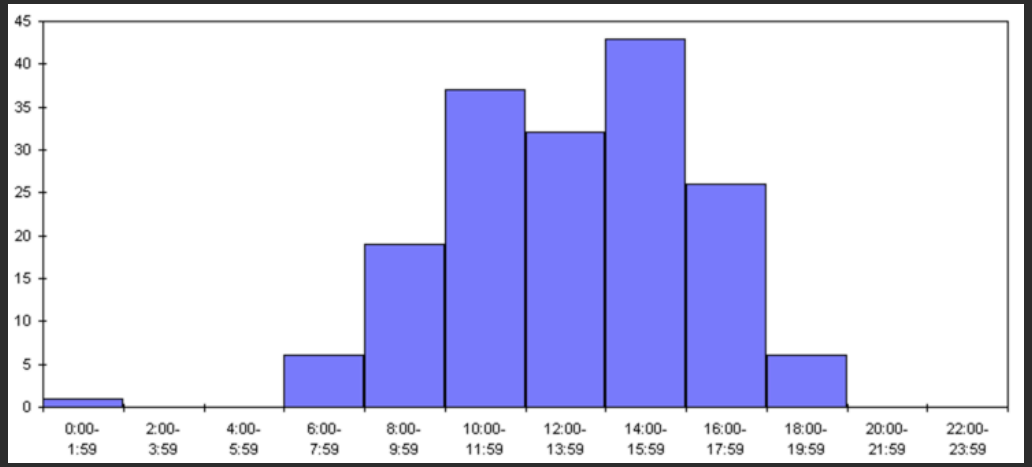
N-GRAM-BASED AUTHOR PROFILES FOR AUTHORSHIP ATTRIBUTION

VLADO KEŠELJ[†] FUCHUN PENG[‡] NICK CERCONE[†] CALVIN THOMAS[†]
[†]*Faculty of Computing Science, Dalhousie University, Canada*
{vlado, nick, thomas}@cs.dal.ca
[‡]*School of Computer Science, University of Waterloo, Canada*
f3peng@cs.uwaterloo.ca

We present a novel method for computer-assisted authorship attribution based on character-level *n*-gram author profiles, which is motivated by an almost-forgotten, pioneering method in 1976. The existing approaches to automated authorship attribution implicitly build author profiles as vectors of feature weights, as language models, or similar. Our approach is based on byte-level *n*-grams, it is language independent, and the generated author profiles are limited in size. The effectiveness of the approach and language independence are demonstrated in experiments performed on English, Greek, and Chinese data. The accuracy of the results is at the level of the current state of the art approaches or higher in some cases.

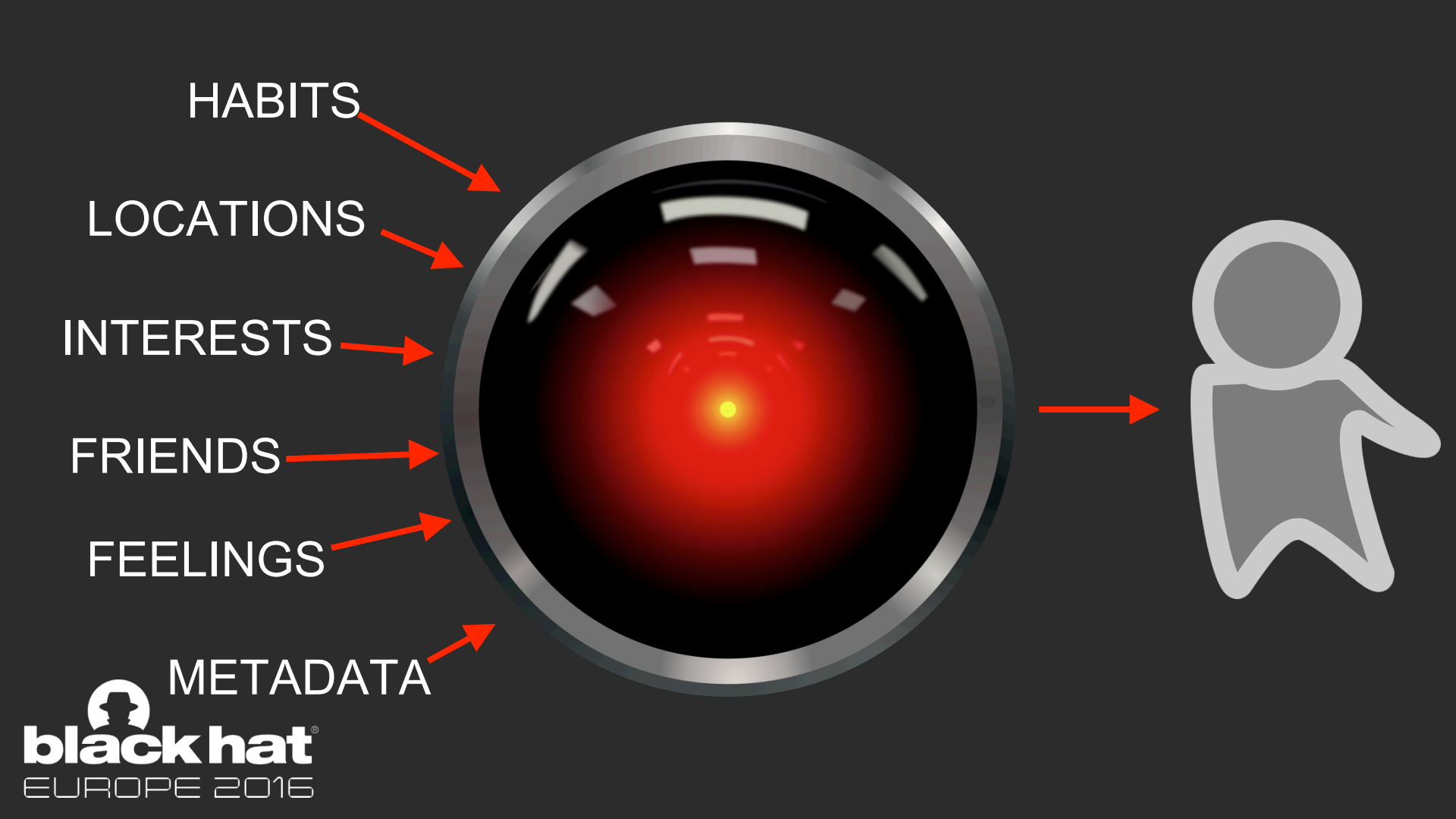
Key words: Authorship attribution, character *n*-grams, text categorization

Your writing


45
40
35
30
25
20
15
10
5
0
0:00-
1:59
2:00-
3:59
4:00-
5:59
6:00-
7:59
8:00-
9:59
10:00-
11:59
12:00-
13:59
14:00-
15:59
16:00-
17:59
18:00-
19:59
20:00-
21:59
22:00-
23:59

Your timing


Pattern of Life Analysis


HABITS
LOCATIONS
INTERESTS
FRIENDS
FEELINGS
METADATA

Personas

Everything you are, do, or use is linked to a persona.

If an activity wasn't intentionally done under an alt persona, it's linked to your true persona.

Anything can link two personas. Anything.

mmkay?

Links are bad.

The Grugq's OPSEC: Because Jail is for wuftpd

black hat EUROPE 2016

OPSEC Refresher

“Paranoia doesn’t work retroactively” - The Grugq

For any persona:

One account - One persona - One activity

One phone no. - One persona - One activity

One connection - One persona - One activity

One home - One persona - One activity

One device - One persona - One activity

One location - One persona - One activity

One *anything else*?

One. Use.

“But Dave, what about a _____?”

One.

Use.

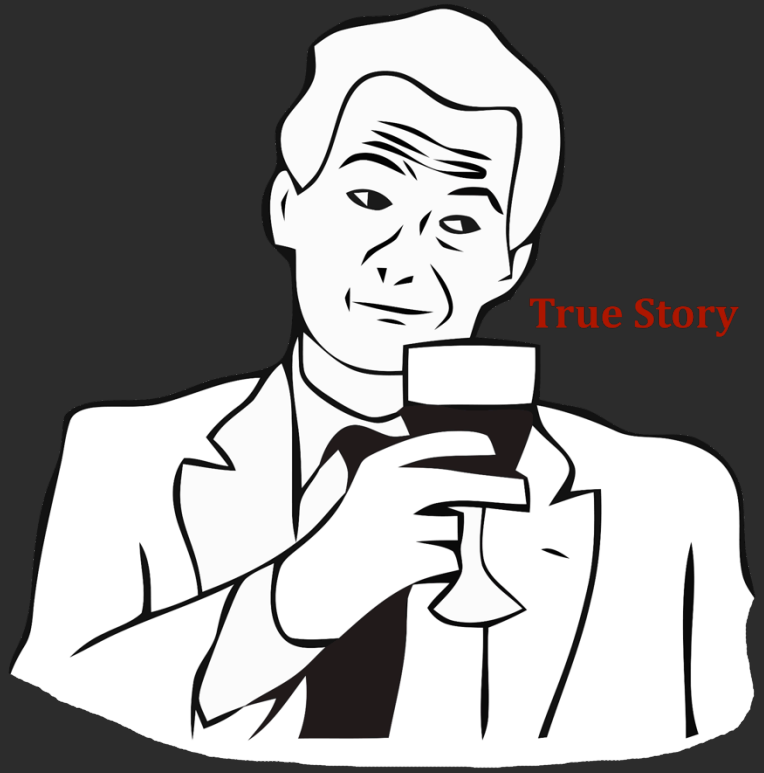
Other Options?

When a new American diplomat arrives for duty at the U.S. embassy in Moscow or Beijing, CIA official say, Russian and Chinese intelligence operatives run data analytics programs that check the "digital dust" associated with his or her name. If the newcomer's footprint in that dust – social media posts, cell phone calls, debit card payments – is too small, the "diplomat" is flagged as an undercover CIA officer.

Source: Reuters

OPSEC Fails

Bombing the Test



Bombing the Test

1. Go to Harvard
2. Take difficult class
3. Don’t study for final exam
4. How do you get out of taking the exam?

Email a Bomb Threat!

Bombing the Test

1. Started up Tor
2. Emailed bomb threat (via Tor) to:
    a. Two Harvard officials
    b. Harvard Police Department
    c. The Crimson (Harvard's newspaper)
3. Went to take the test
4. Evacuated with other students

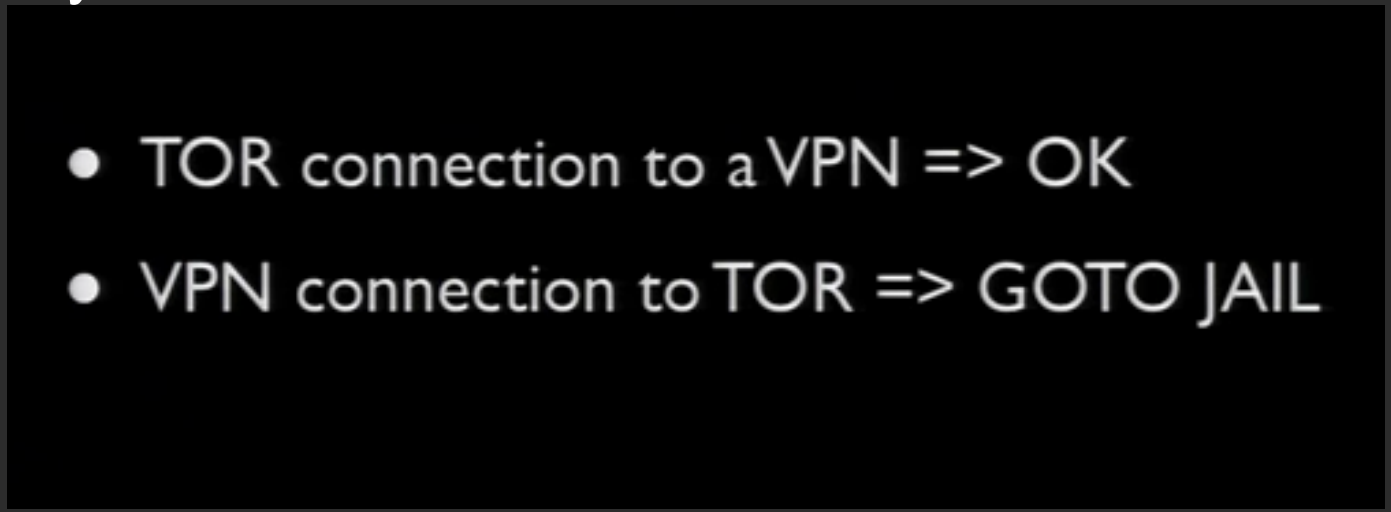
Bombing the Test

1. Police determined the email sent via Tor
2. Looked at Harvard's network logs
3. ID'd Tor users at the time email was sent
4. Knocked on doors

Conventional Advice

VPN = Privacy

Tor = Anonymity



Source: The Grugq

There’s no one size fits all for OPSEC

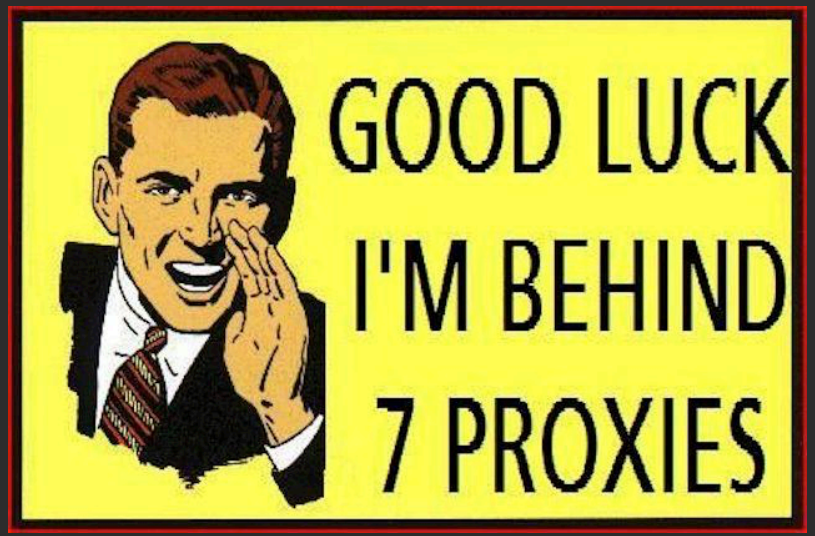
What about Bob?

What about Bob? #1

1. Lives upstairs from a cafe
2. Goes downstairs to use their wifi
3. VM on personal laptop
4. Does [activity]

What about Bob? #2

1. Drives across town to cafe
2. Brings phone but turns it off
3. Uses their wifi - with a proxy
4. VM on personal laptop
5. Does [activity]



What about Bob? #3

1. Public transport to park with wifi
2. Phone is on at home
3. Uses Tor
4. Separate hardware
5. Does [activity]

Information Leaks

black hat
EUROPE 2016

Summary

Assume everything is compromised

Be aware of links

Look at your own pattern of life

Adjust your personas' PoLs to be different

Watch out for leaks where you least expect them

Breaking Big Data

Evading Analysis of the Metadata of Your Life

Dave Venable
Masergy Communications
@davevenable