ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MAX PLANCK INSTITUTE FOR SOFTWARE SYSTEMS

Northeastern University

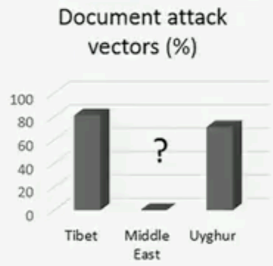
# A Broad View of the Ecosystem of Socially Engineered Exploit Documents

Steve [illegible] Upadhyay, Manuel Gómez Rodríguez and **David Choffnes**

work that we did with our colleagues at the Max Planck Institute

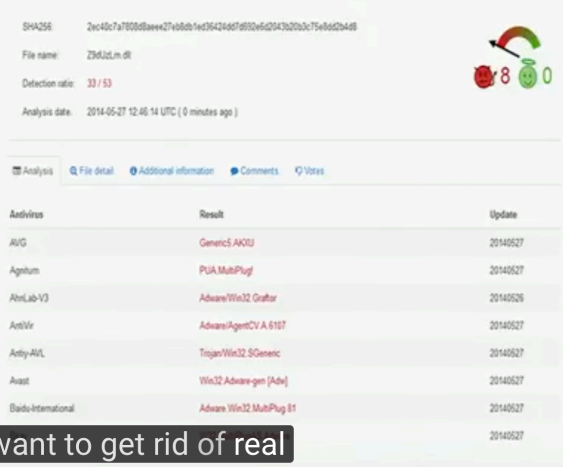
## Challenges with measuring targeted attacks

- Low-volume, socially engineered messages that convince specific victims to install malware
- Three studies published at Usenix Security'14
  - Tibet (Hardy et al.), Middle East (Marczak et al.), and Uyghur (Le Blond et al.)


Measuring targeted attacks is a long and difficult process

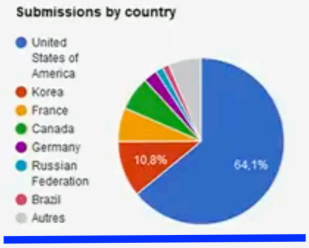
# Can Anti-Virus Aggregators (VirusTotal) help?



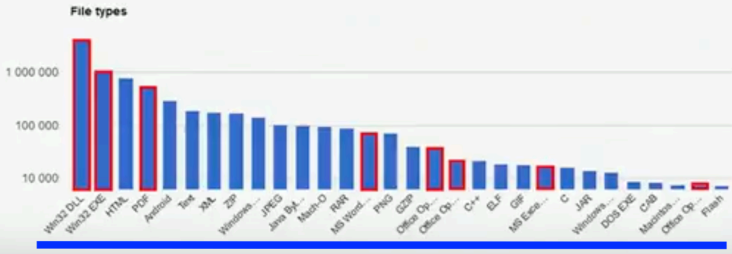

be something you want to get rid of real quick

3

# VirusTotal Statistics (one week)








4

# VirusTotal as a vantage point to measure targeted attacks

virustotal

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MPI but here I'm showing epfl and then we extracted the

5

# Research questions

- Do targeted groups upload exploit documents to VirusTotal?
- Can we scale our analysis to hundreds of thousands of samples?
- How do attacks faced by different groups compare with each other?
- Is VirusTotal used by other actors such as attackers and researchers?

6

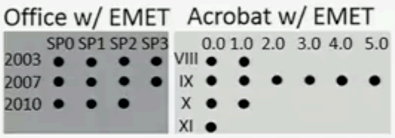
# Exploit document infection process


8



# Can we scale our analysis to hundreds ( thousands of samples? **Detection**





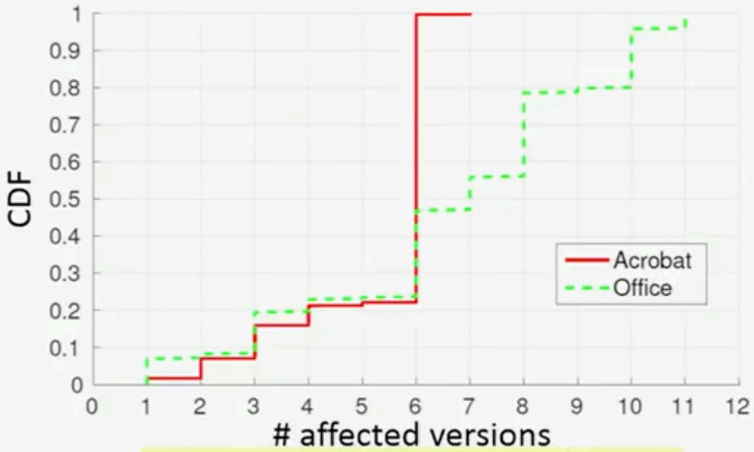

12

## How many versions of readers do we have to test?


before a piece of malware is identified
or another way to look at this is in

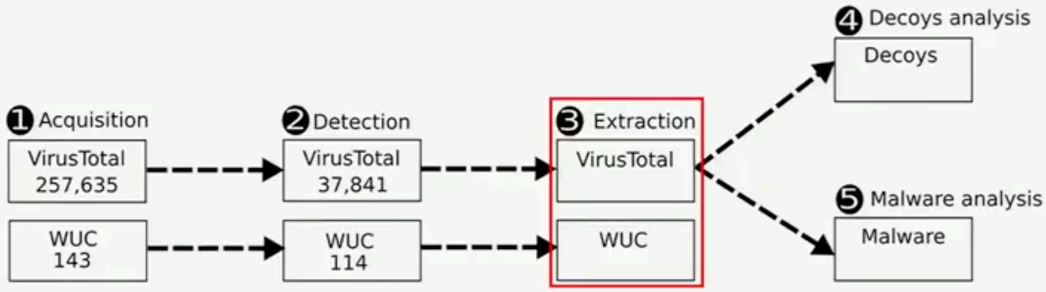


## How many versions of readers do we have to test?

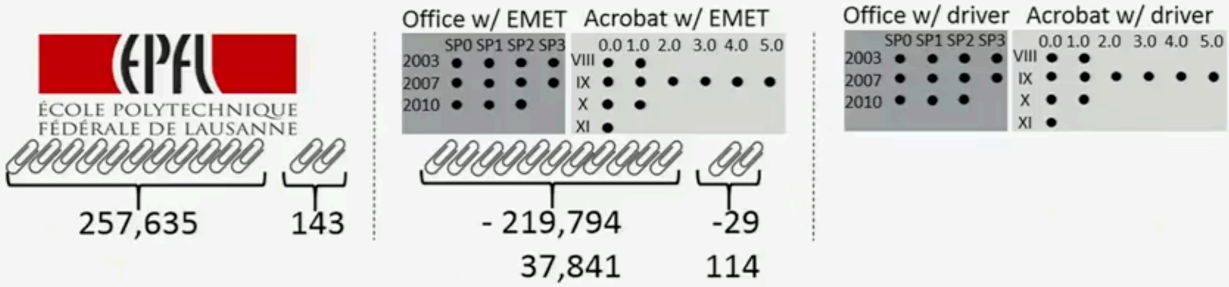


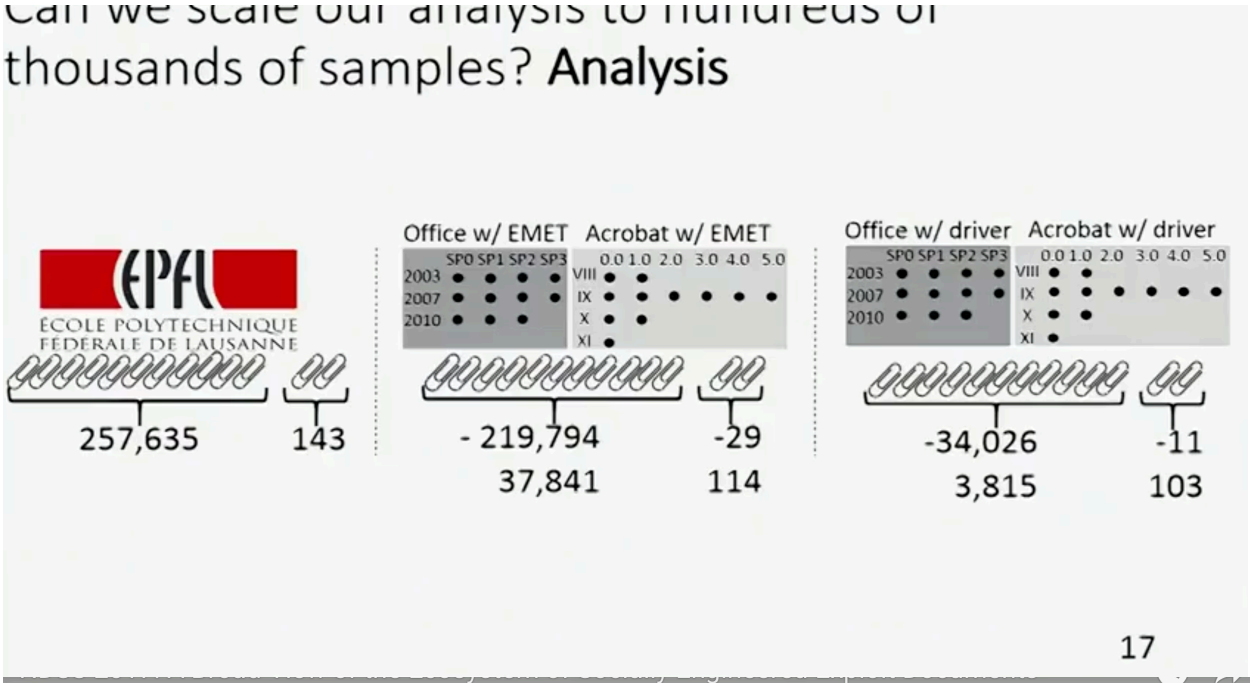

Few exploits are portable across all reader versions

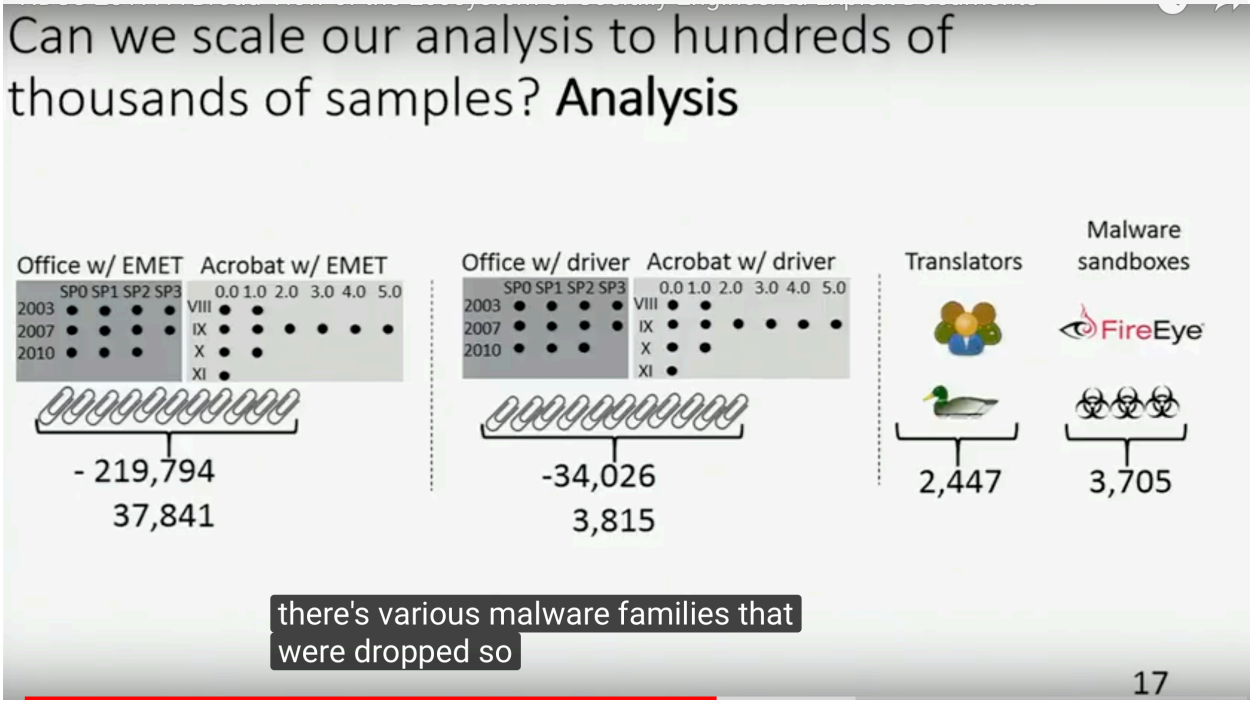
# Data acquisition and processing workflow




# Can we scale our analysis to hundreds of thousands of samples? **Extraction**




thousands of samples? Analysis
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
257,635
143
Office w/ EMET
Acrobat w/ EMET
SP0 SP1 SP2 SP3
2003
2007
2010
0.0 1.0 2.0 3.0 4.0 5.0
VIII
IX
X
XI
- 219,794
37,841
-29
114
Office w/ driver
Acrobat w/ driver
-34,026
3,815
-11
103
17


Can we scale our analysis to hundreds of thousands of samples? Analysis
Office w/ EMET
Acrobat w/ EMET
SP0 SP1 SP2 SP3
2003
2007
2010
0.0 1.0 2.0 3.0 4.0 5.0
VIII
IX
X
XI
- 219,794
37,841
Office w/ driver
Acrobat w/ driver
-34,026
3,815
Translators
Malware sandboxes
FireEye
2,447
3,705
there's various malware families that were dropped so
17

## Do targeted groups upload exploit documents on VirusTotal? **Likely targets (inferred from decoys)**

| *Group* | *Number* | *Fraction* |
|---|---|---|
| *Uyghur* | 237 | .16 |
| *Vietnam* | 145 | .10 |
| *USA* | 118 | .08 |
| *Tibet* | 115 | .08 |
| *Taiwan* | 100 | .06 |
| *India* | 72 | .05 |
| *Russia* | 51 | .03 |
| *Japan* | 50 | .03 |
| *Philippines* | 38 | .02 |
| *South Korea* | 19 | .01 |
| *Myanmar* | 17 | .01 |
| *Mongolia* | 14 | <.01 |
| *Thailand* | 9 | <.01 |
| *Indonesia* | 7 | <.01 |
| *Others* | 438 | .30 |
| *Total* | 1,430 | 1.00 |

found over 1,400 cases that were clearly targeted attacks you'll



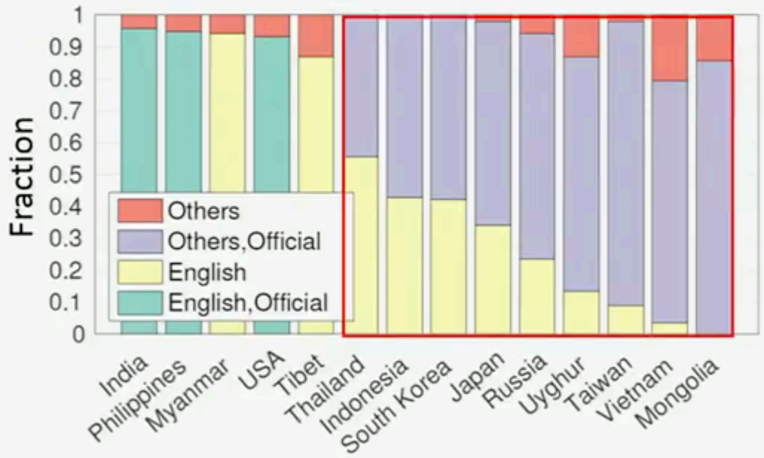
## Do targeted groups upload exploit documents on VirusTotal? **Likely targets (inferred from decoys)**

| *Group* | *Number* | *Fraction* |
|---|---|---|
| *Uyghur* | 237 | .16 |
| *Vietnam* | 145 | .10 |
| *USA* | 118 | .08 |
| *Tibet* | 115 | .08 |
| *Taiwan* | 100 | .06 |
| *India* | 72 | .05 |
| *Russia* | 51 | .03 |
| *Japan* | 50 | .03 |
| *Philippines* | 38 | .02 |
| *South Korea* | 19 | .01 |
| *Myanmar* | 17 | .01 |
| *Mongolia* | 14 | <.01 |
| *Thailand* | 9 | <.01 |
| *Indonesia* | 7 | <.01 |
| *Others* | 438 | .30 |
| *Total* | 1,430 | 1.00 |

VirusTotal gives visibility into attacks targeting numerous groups

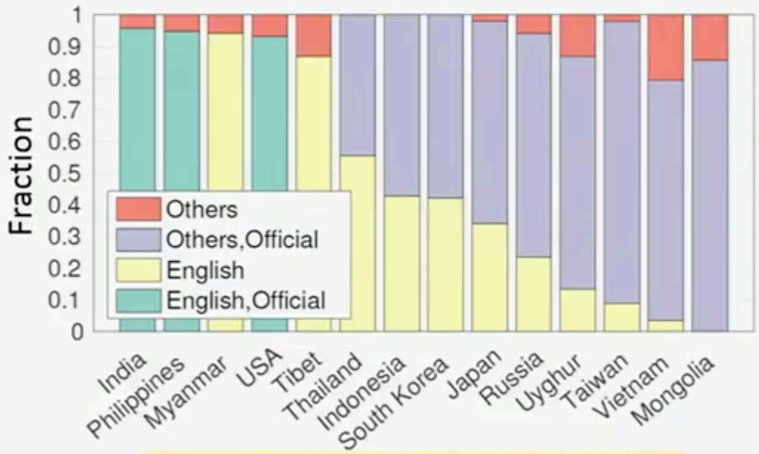
19

How attacks faced by different groups compare with each other? **Languages of decoys**



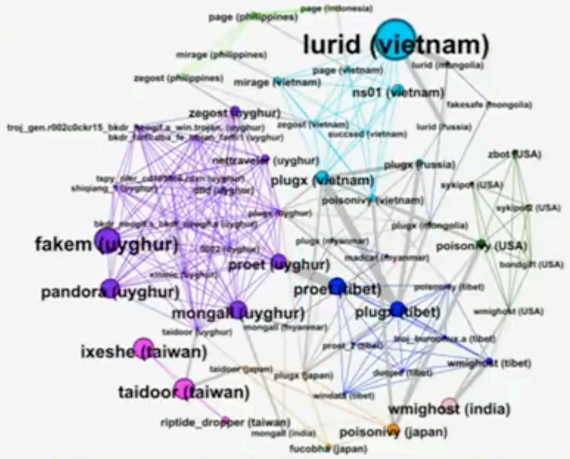



How attacks faced by different groups compare with each other? **Languages of decoys**




Decoys tend to use the official language of the groups they target

# How attacks faced by different groups compare with each other? **Malware targeting**



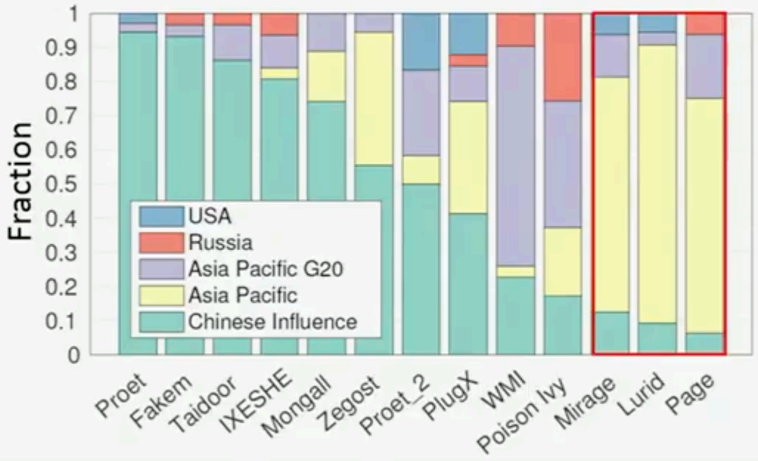

From our dataset, malware families tend to target one or two countries



# Targeted regions

- Chinese influence: Tibet, Uyghur, Taiwan
- Asia Pacific: Myanmar, the Philippines, Thailand, and Vietnam
- Asia Pacific, G20: India, Indonesia, Japan, and South Korea
- Russia and USA

## How do attacks faced by different groups compare with each other? **Malware targeting (cont.)**






## Future work

- Monitoring operator behavior of targeted malware
- Analysis of evasions techniques, attackers operations, and other attack vectors
- Deploy on-premises and cloud-based services for analysis of email attachments

# Take home messages

- Complementary methodology to measure targeted attacks at scale
- At-risk groups upload exploit documents to VirusTotal
- Groups tend to be targeted with tailored decoys and malware families
- Preliminary impact
  - Service deployed at email provider with 100,000+ users
  - Dataset and academic service available at https://slingshot.dedis.ch