HOW CRIMINALS USE AI & ML

Federica Cesti – Tri Nguyen - Quynh Tran - Abuelgasim Elfadul Gafar

Big Data - Spring 2019

Big Data - Spring 2019

Introduction
Overview of Criminal Uses
Survey of Methodologies
Specific framework
Conclusion

UNIVERSITY OF OULU

Big Data - Spring 2019

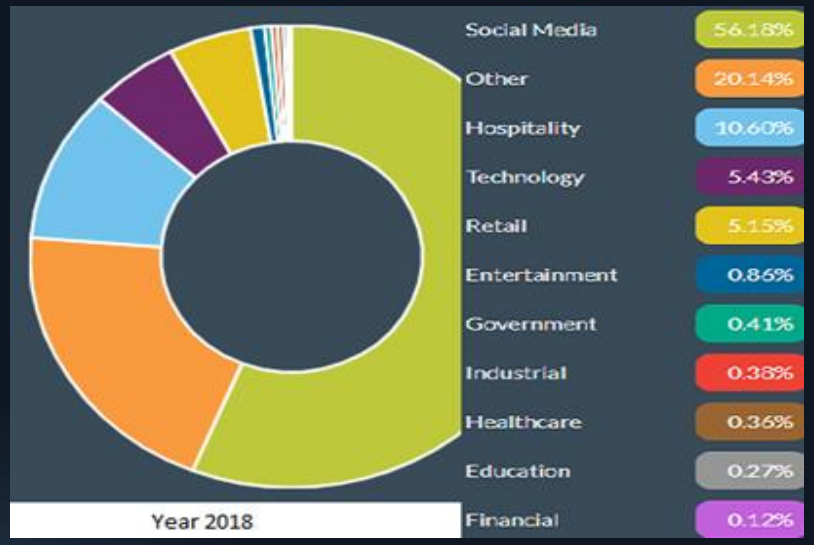
Introduction

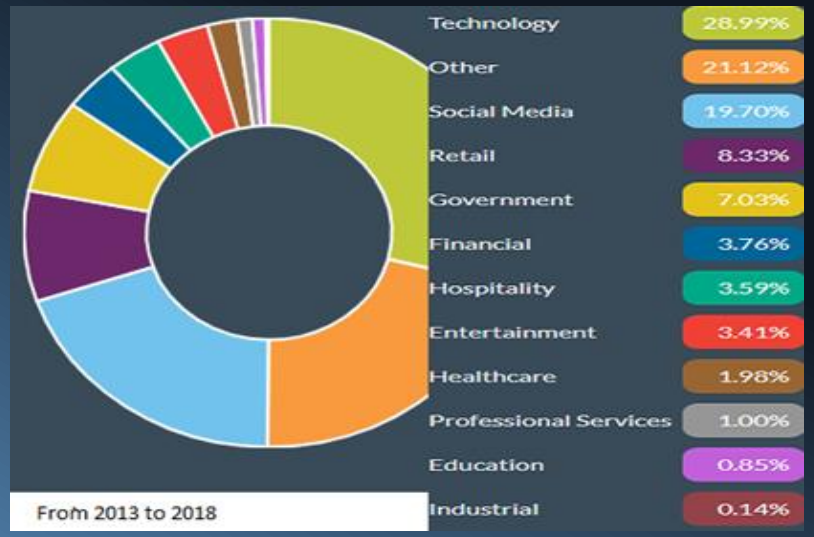
Overview of Criminal Uses

Survey of Methodologies

Specific framework

Conclusion



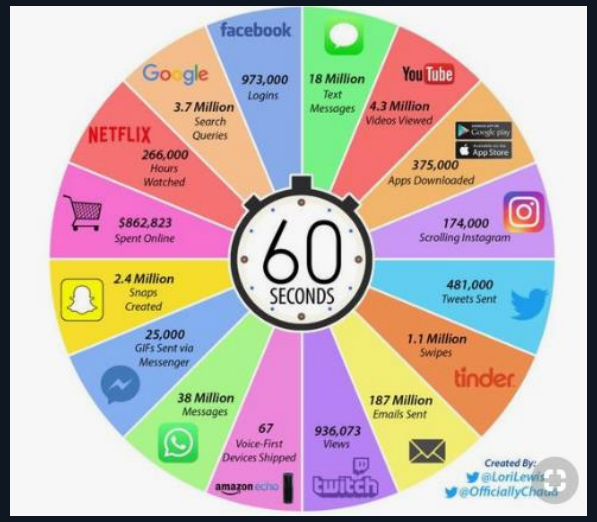


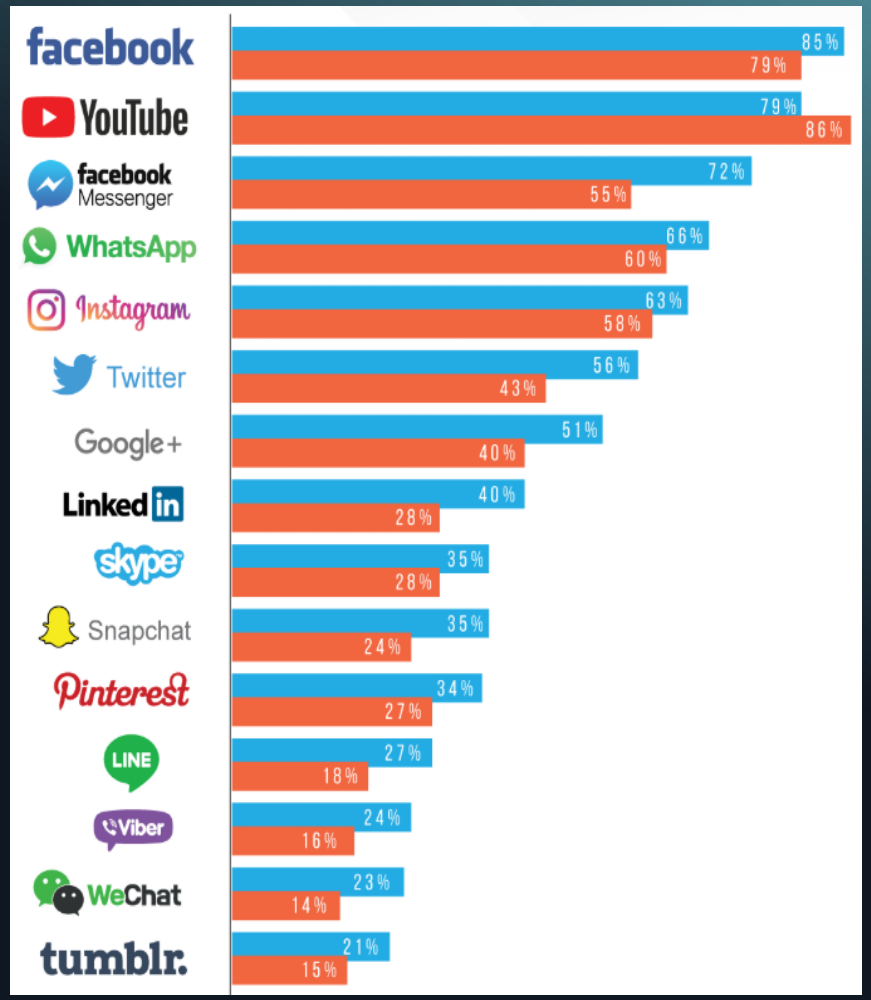
Introduction

- Digitalized life and big data, our information on cloud-high risk of losing data
- Technology good but also has dark bad side
- Possible threats from the criminal perspective
- Four most prominent methods
- Frameworks and tools that serve criminals
- The frequency of stolen data is about 75 records every second

Introduction

Top 15 social media and most bigdata generators





https://fi.pinterest.com/pin/314196511501816049/

UNIVERSITY OF OULU

Big Data - Spring 2019

Introduction

Overview of Criminal Uses

Survey of Methodologies

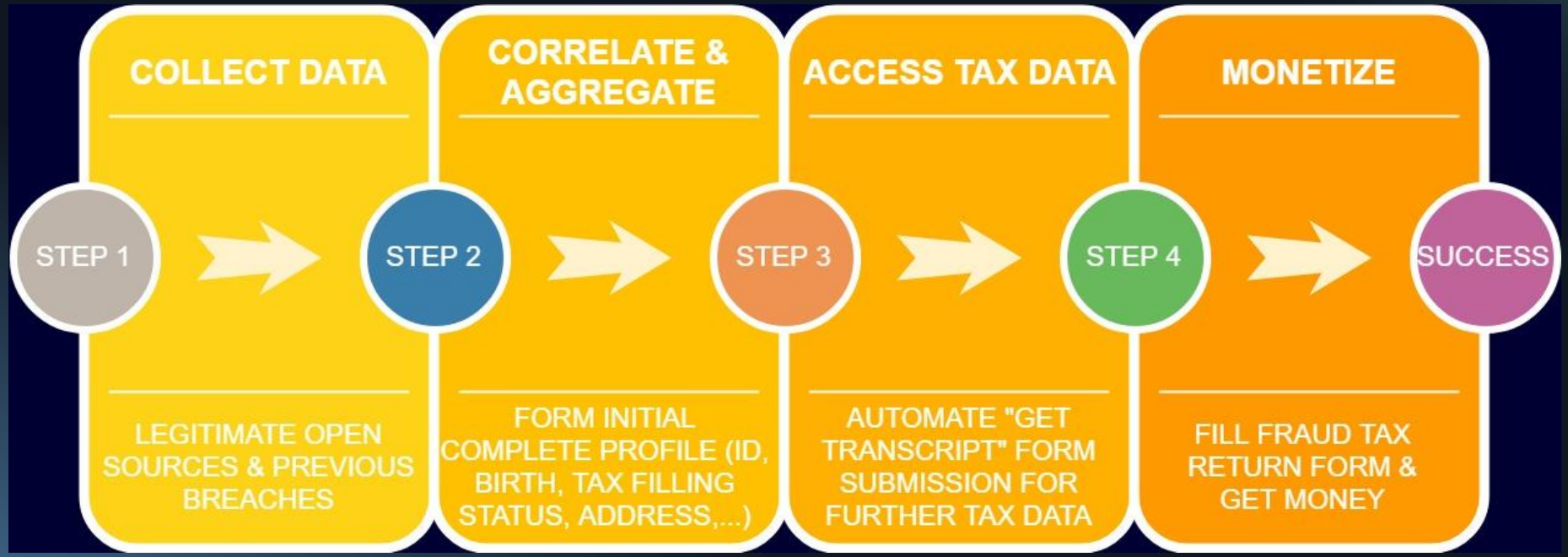
Specific framework

Conclusion

OVERVIEW OF CRIMINAL USES

CASE STUDY - US TAX FRAUD 2015

Tax data breach: 300.000 taxpayers accounts + 600.000 suspected fail attempts to access data



https://www.information-age.com/irs-reveals-how-cyber-criminals-use-big-data-too-tax-breach-200-worse-feared-123460016/

OVERVIEW OF CRIMINAL USES

CYBER-CRIMES

- Automated spearphishing
- Automated malware
- Exploits vulnerabilities from nature of systems: Data poisoning, accelerate flaws findings

PHYSICAL THREATS

- Ai weaponization: Swarm drones, robot killers
- Hijacking personal automated vehicles

INFORMATIVE THREATS

- Impersonation (Masquerade)

(2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Workshop Report.

CYBERCRIMES

❖ SPEARPHISHING

- Traditional labor-intensive cyberattacks become automated
- Low-skill groups and individuals are able to perform
- Highly personalized and accurate
- Target massive victims, even ones that are now seen as unworthy under cost-benefit perspective

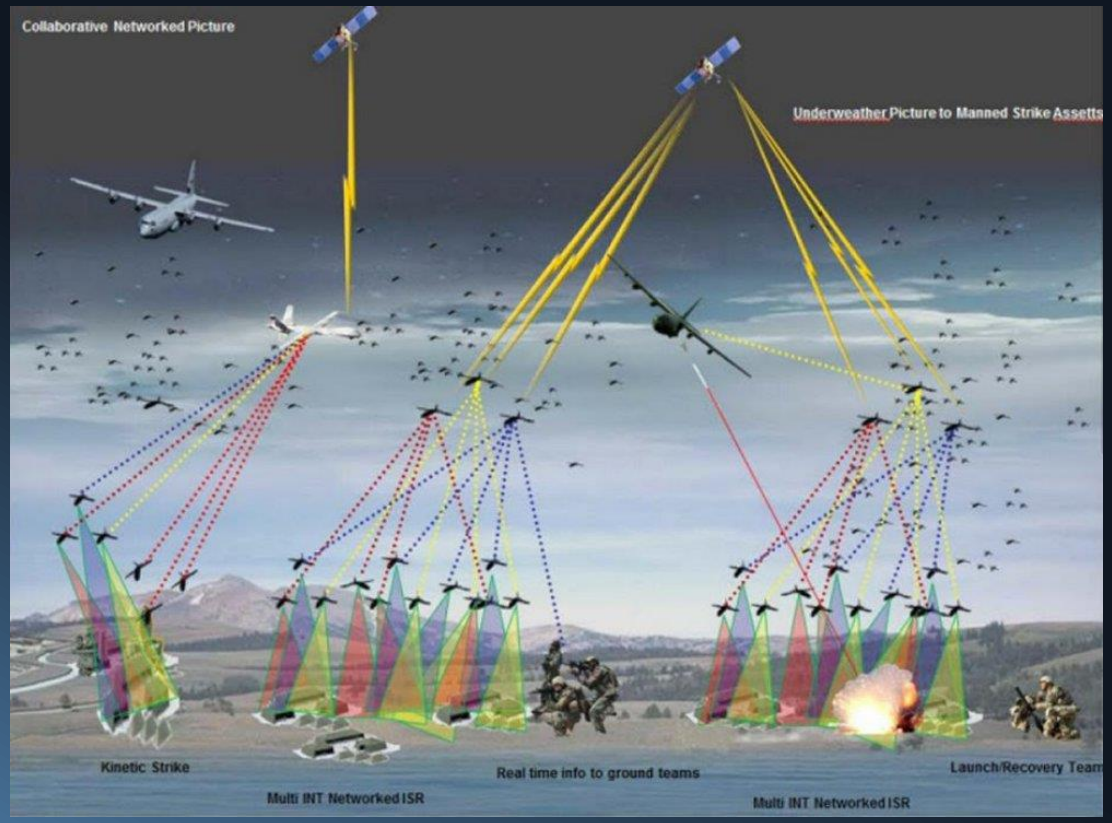
❖ MALWARE

- Autonomous
- Quickly propagate
- Learn context from target environment → select most suitable attacking techniques

❖ Exploits vulnerabilities from nature of systems:

- Data poisoning: Feed modified data to confuse machines
- Accelerate flaws findings: automate the discovery of vulnerabilities, using past code flaws to speed up the new flaws findings

PHYSICAL THREATS

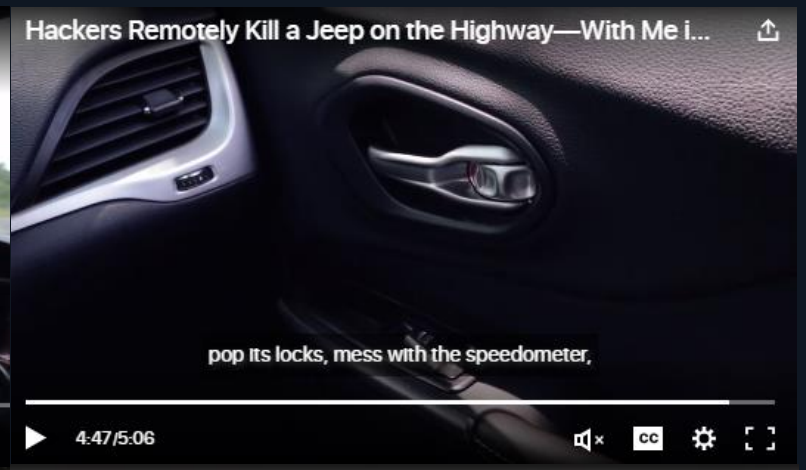


❖ AI WEAPONS

- Exploit face recognition and navigation system
- Swarm unmanned aerial vehicles (UAVs) attacks like drones
- Risk of engineering commercial machines like drones into weapons

Photo credit: https://www.uasvision.com/2018/06/28/suppressing-air-defenses-by-uav-swarm-attack/

PHYSICAL THREATS



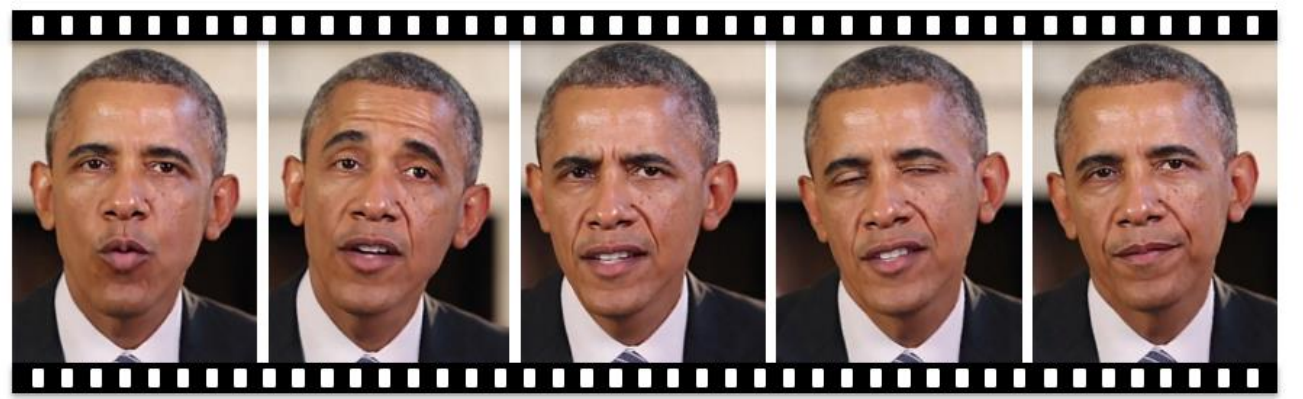
HIJACKING

- Take control of automomous vehicles and machines (cars, cleaning robots,..) → civilian attacks
- Might combine digital clickbaits to steal authorization first

[HACKERS REMOTELY KILL A JEEP ON THE HIGHWAY—WITH ME IN IT]

INFORMATIVE THREATS

IMPERSONATION



Output Obama Video

- It's the practice of pretexting as another person to obtain information, access to a person, company, or computer system and compromise the person image.
- Exploit speech and image synthesis to create fake videos and chatbots.
- Current prediction to target presidential election, can expand to normal citizens to steal data and money

Suwajanakorn, Supasorn & M. Seitz, Steven & Kemelmacher, Ira. (2017). Synthesizing Obama: learning lip sync from audio. ACM Transactions on Graphics. 36. 1-13

Introduction
Overview of Criminal Uses
Survey of Methodologies
Specific framework
Conclusion

STATE-OF-THE-ART

Where AI and ML can help criminals?

- INFORMATION GATHERING
- IMPERSONATION
- BYPASSING RESTRICTIONS
- AUTOMATED ATTACKS

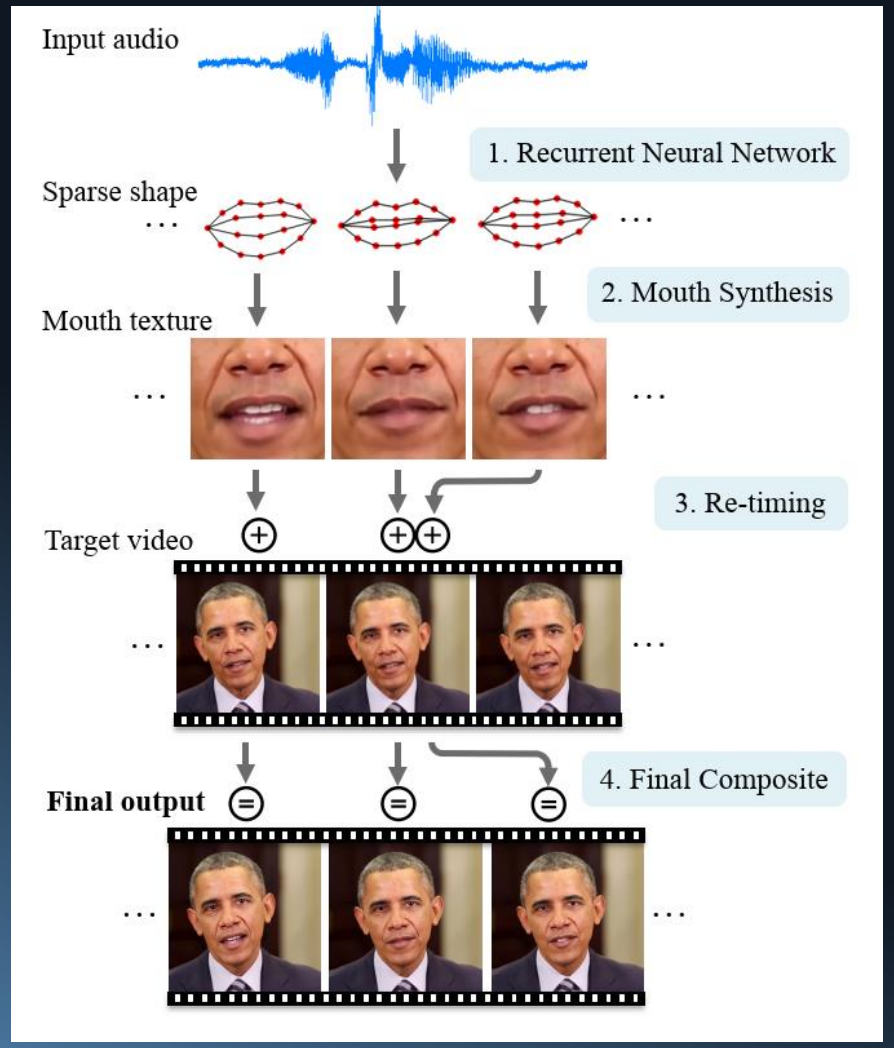
Information gathering

The aim is to steal confidential and personal data from users and companies.

Hackers use Machine Learning classifying algorithm to drive phishing attacks, targeted to specific individuals.

- Social Mapper: image recognition tool
- Know Your Enemy attack: to steal configuration information form a SDN network
- DirBuster: designed to brute force directories and files names on web/application servers.
- SNAP_R: tool to increase phishing campaign

Impersonation



[Synthetizing Obama Video]

Alternative tools:

- Google WaveNet: tool to create bots that speaks exactly like a human
- DeepFake: tool to generate fake videos

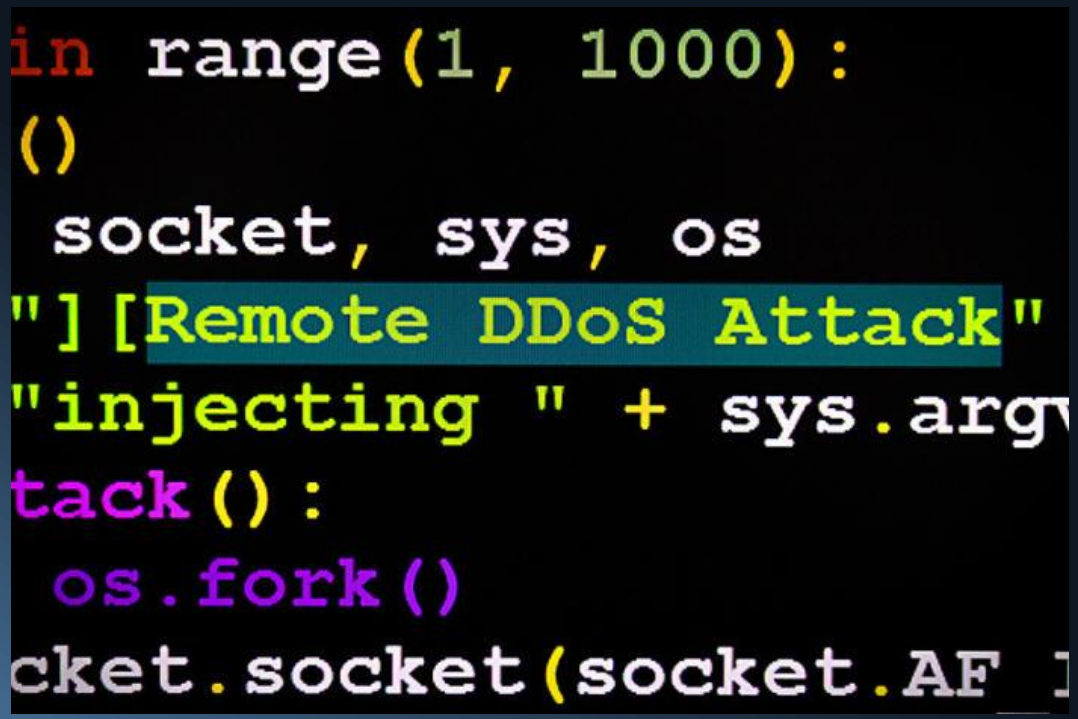
https://www.youtube.com/watch?v=9Yq67CjDqvw

Bypassing restrictions

The aim is to solve tests in order to get access to blocked resources or accounts.

- Solve CAPTCHA tests
- Solve "select all pictures containing a bus"
- PassGAN: tool to generate and guess passwords

Automated attacks



Several tools can automate malware spreading or applications' crashes.

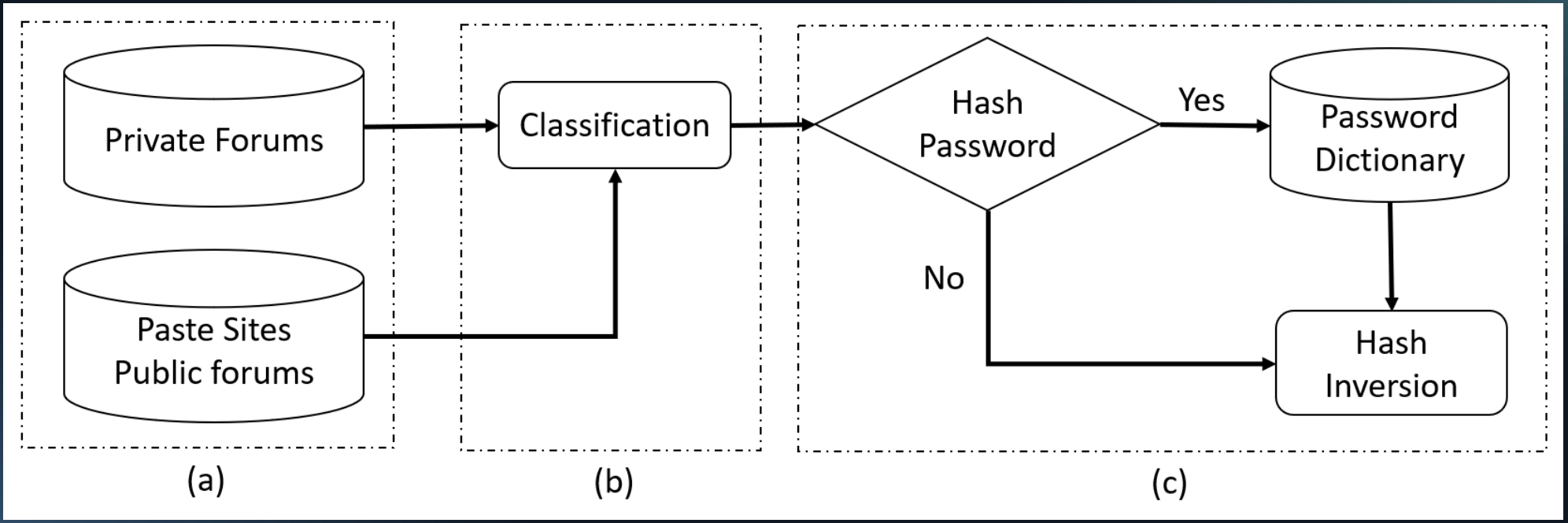
Also,

- DeepLocker: software that hides itself until the detection of a particular event.
- AI DDoS: their aim is to make an online service unavailable.

https://www.networkworld.com/article/3289108/the-rise-of-artificial-intelligence-ddos-attacks.html

Introduction
Overview of Criminal Uses
Survey of Methodologies
Specific framework
Conclusion

Credential Leaks

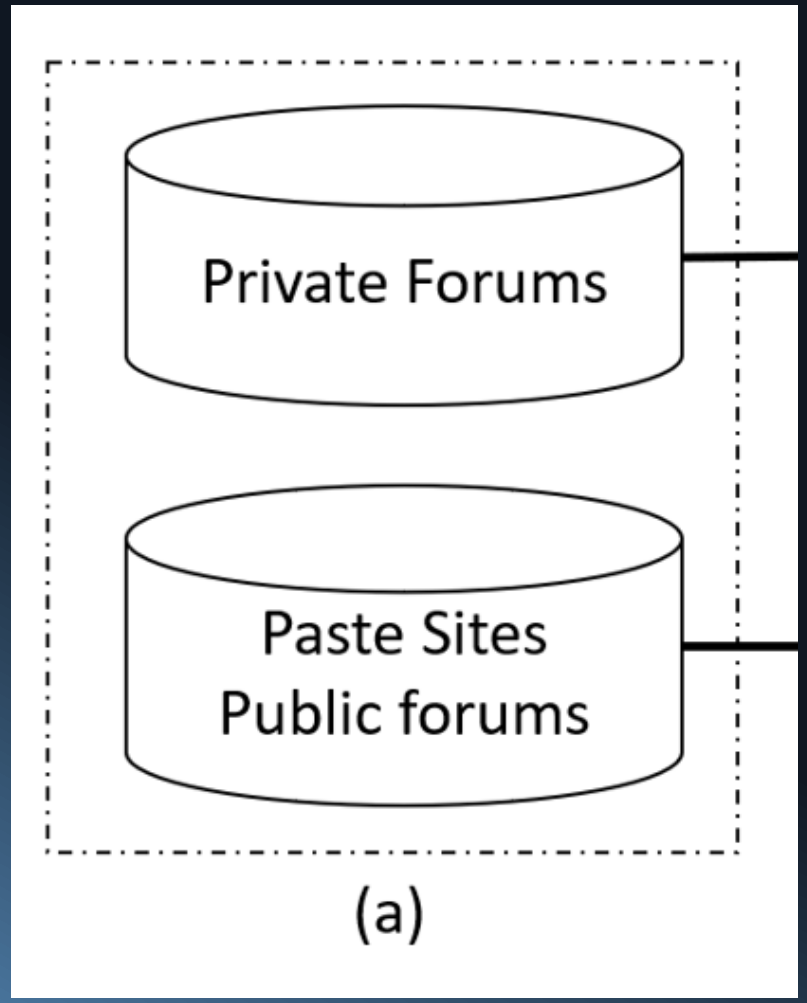


Framework for identifying credential leaks: (a) Crawling data, (b) Parsing and analyzing data, (c) Hash inversion

- Proving the data breach can be from public sites (not only black markets)
- It can be a model to collect and detect current credential leaks

K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki, *et al.*, "Data breaches, phishing, or malware?: Understanding the risks of stolen credentials," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1421–1434, ACM, 2017.

Credential Leaks (a) – Crawling



❖Crawling data

- Public forums
  - 115 paste sites
  - 5 blackhat forums
- Current Google's history
- Remove sites having less than 100 email addresses
- Result
  - 31446 candidate documents from public places and 258 credential leaks from 11 private forums

❖Private forums

- 258 confirmed leaks (from 11 private forums)

Credential Leaks (b) - Classification

❖Data is separated based on a delimiter detection

❖A parser is used to detect records having at least 2 columns

- Recognize email column (regular expression)
- Password column (2 main types)

- Hashing values (fixed-length values)
- Plain-text password (binary classification)

❖Binary classification

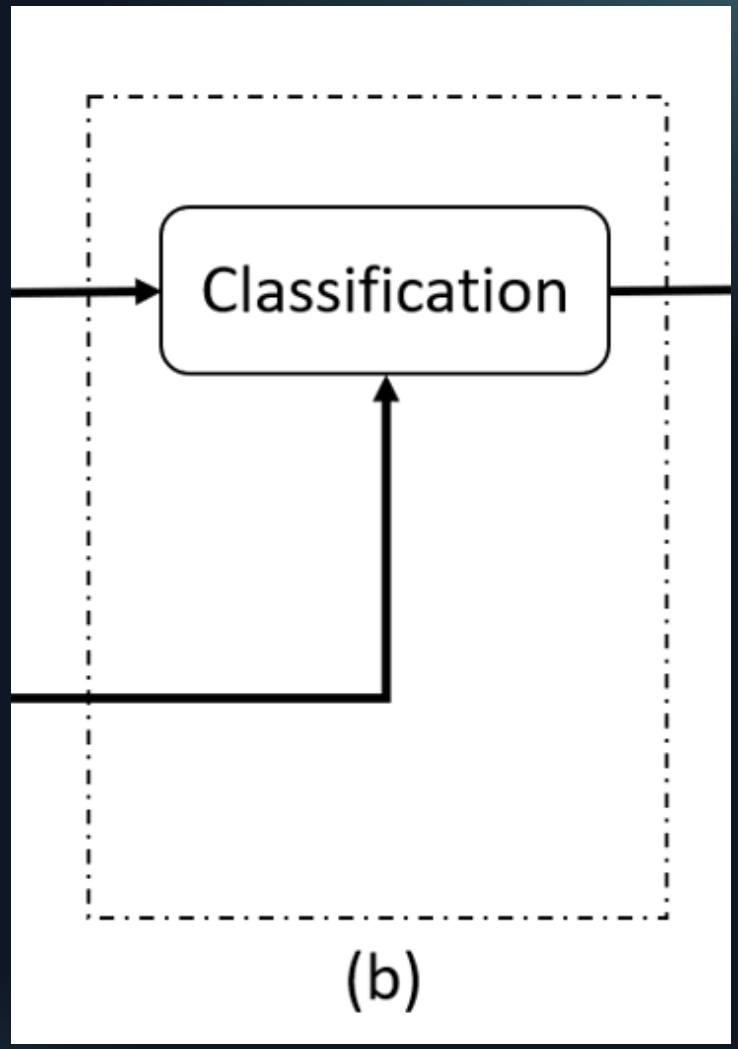
- Plain-text passwords are segmented as n-grams

- A grid search with 10-fold cross validation on data from private sites
- N-grams [1, 10] and binary vectors have top 1k to 100k most common n-grams

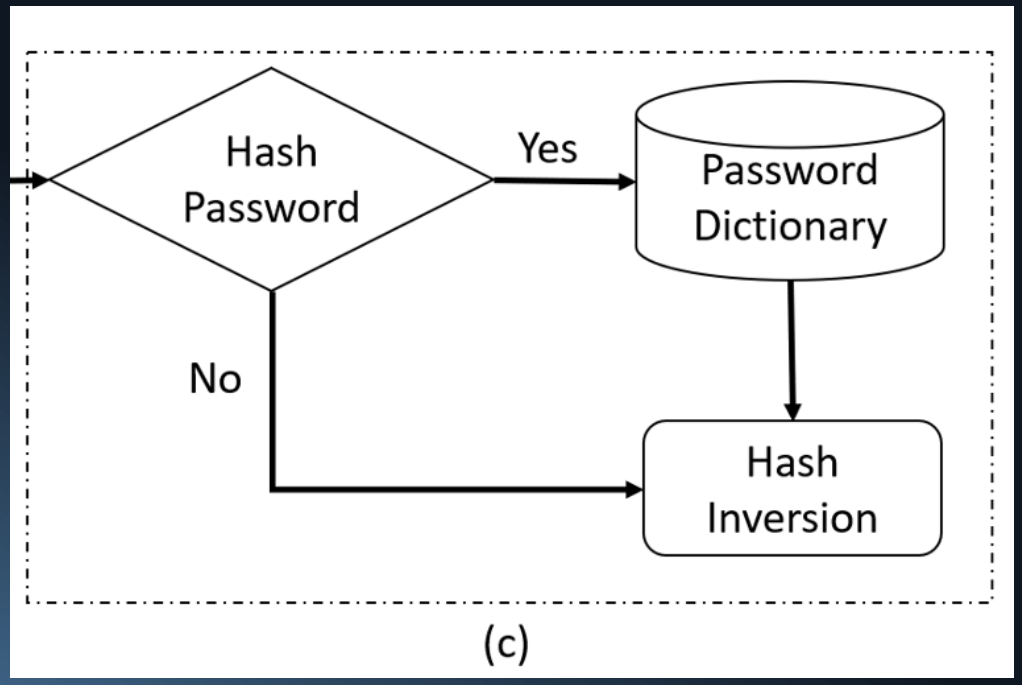
Result: N-grams is [2, 5] and 10k n-grams per class

❖Result: 94%

- 3527 candidate documents
- 123055697 emails and passwords



Credential Leaks (c) – Hash Inversion



- ❖Password dictionary
  - Plain-text passwords
  - Current dictionaries, a set of 3416701663 keywords
- ❖Inversion Hashing values
  - 35,8% of hashed passwords are inverted
- ❖Reasons
  - Salted passwords
  - Noise from black-market sites
  - Pay attention only on MD5 vs SHA-1

Table 3: Top 20 largest credential leaks in our dataset and the fraction of inverted (or existing plaintext) passwords.

Rank	Source	Number of credentials	Plaintext after inversion
1	Unknown [p]	558,862,722	100.0%
2	MySpace [p]	322,014,681	100.0%
3	Badoo	125,322,081	33.0%
4	Adobe[◇]	123,947,902	0.0%
5	LinkedIn	112,322,695	85.6%
6	VK [p]	76,865,954	99.6%
7	Tumblr[*]	73,355,694	0.0%
8	Dropbox[†]	68,669,208	0.0%
9	Zoosk	57,085,529	68.2%
10	IMesh[‡]	51,283,424	0.0%
11	LastFM	41,631,844	85.4%
12	Fling [p]	40,724,332	100.0%
13	Neopets [p]	35,822,980	100.0%
14	Mate1 [p]	27,383,966	100.0%
15	Unknown [p]	26,351,372	99.8%
16	000webhost [p]	15,249,241	100.0%
17	Taobao [p]	15,051,549	100.0%
18	NexusMods [p]	6,759,631	100.0%
19	Unknown [p]	5,728,163	99.7%
20	Unknown [p]	4,901,088	100.0%
–	Total	1,922,609,265	76.0%

Table 2: Breakdown of where we source credential leaks.

Source	Candidate documents	Confirmed leaks	Credentials extracted
Paste sites	3,317	1,666	4,855,780
Search index	26,208	1,304	10,856,227
Public forums	1,921	557	107,343,690
Private forums	–	258	1,799,553,568

Table 4: Top 10 passwords across all plaintext leaks.

Rank	Top Passwords	Number of Credentials	Percent of Credentials
1	123456	6,387,184	0.35%
2	password	2,759,747	0.15%
3	123456789	2,249,344	0.12%
4	abc123	985,709	0.10%
5	password1	888,836	0.05%
6[*]	homelesspa	855,477	0.05%
7	111111	855,257	0.05%
8	qwerty	829,835	0.05%
9	12345678	828,848	0.05%
10	1234567	740,464	0.04%

K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki, *et al.*, "Data breaches, phishing, or malware?: Understanding the risks of stolen credentials," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1421–1434, ACM, 2017.

Introduction
Overview of Criminal Uses
Survey of Methodologies
Specific framework
Conclusion

Conclusion

- The extremely impacts of criminal nowadays from digital data
- Several general types of attacks based on AI and ML
  - Information gathering
  - Impersonation
  - Bypassing restrictions
  - Automated attacks
- Many successful frameworks can be easily built support different types of attacks
- The sophisticated criminals are automated with state-of-the-art technologies

Thank you

Questions and Comments