Data Security

Mustakimur R. Khandaker

AI in the Presence of Attacker

Important to consider the presence of attacker.

- History has shown attacker always follows footsteps of new technology development (or sometimes even leads it).
- The stake is even higher with AI.
  - As AI controls more and more systems, attacker will have higher & higher incentives.
  - As AI becomes more and more capable, the consequence of misuse by attacker will become more and more severe.

Continue ...

Attack AI:

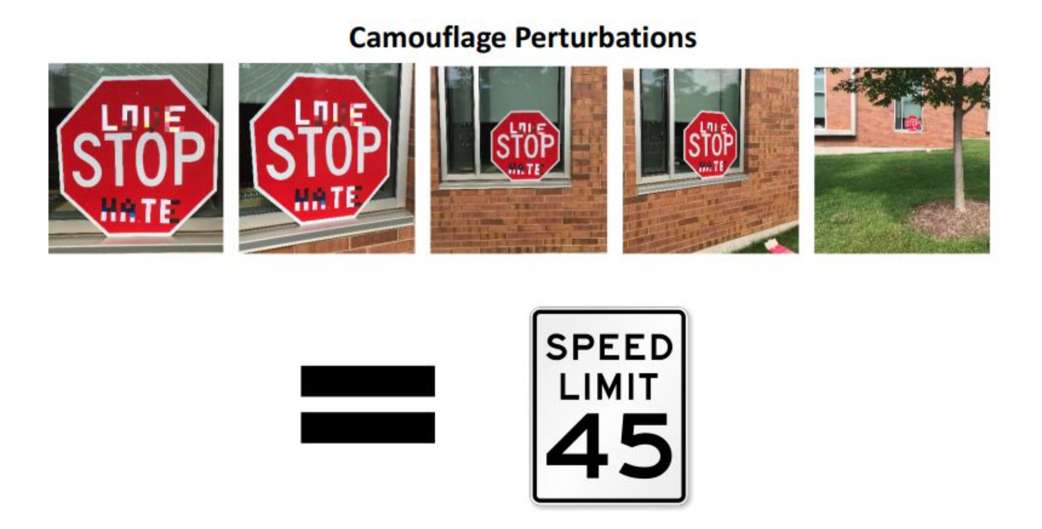
- Cause learning system to not produce intended/correct results.
- Cause learning system to produce targeted outcome designed by attacker.
- Learn sensitive information about individuals.
- Need security in learning systems.

Misuse AI:

- Misuse AI to attack other systems.
    - Find vulnerabilities in other systems.
        - Target attacks.
        - Devise attacks.
- Need security in other systems.

Adversarial Examples in Physical World

Can we generate adversarial examples in the physical world that remain effective under different viewing conditions and viewpoints, including viewing distances and angles?



Camouflage Perturbations

Adversarial perturbations are possible in physical world under different viewing conditions and viewpoints, including viewing distances and angles.

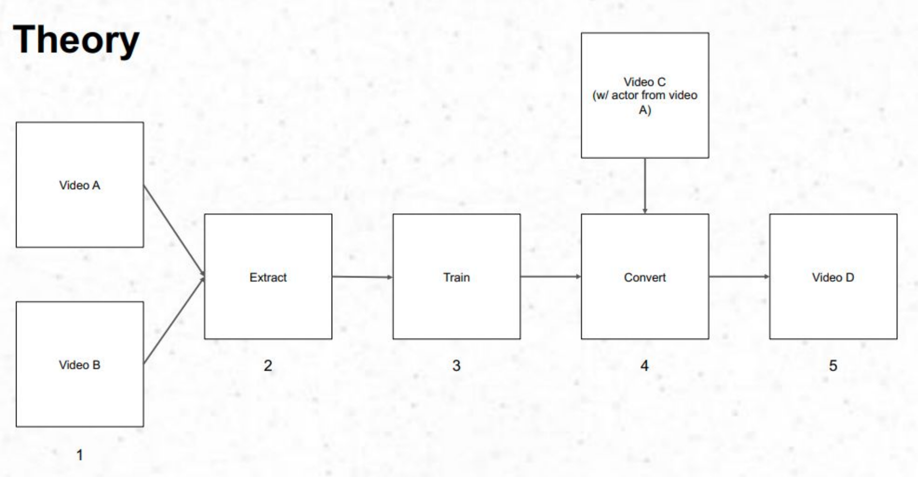
Deepfake

A deepfake is generally understood to be a video in which the face of one person has been swapped with the face of another person.

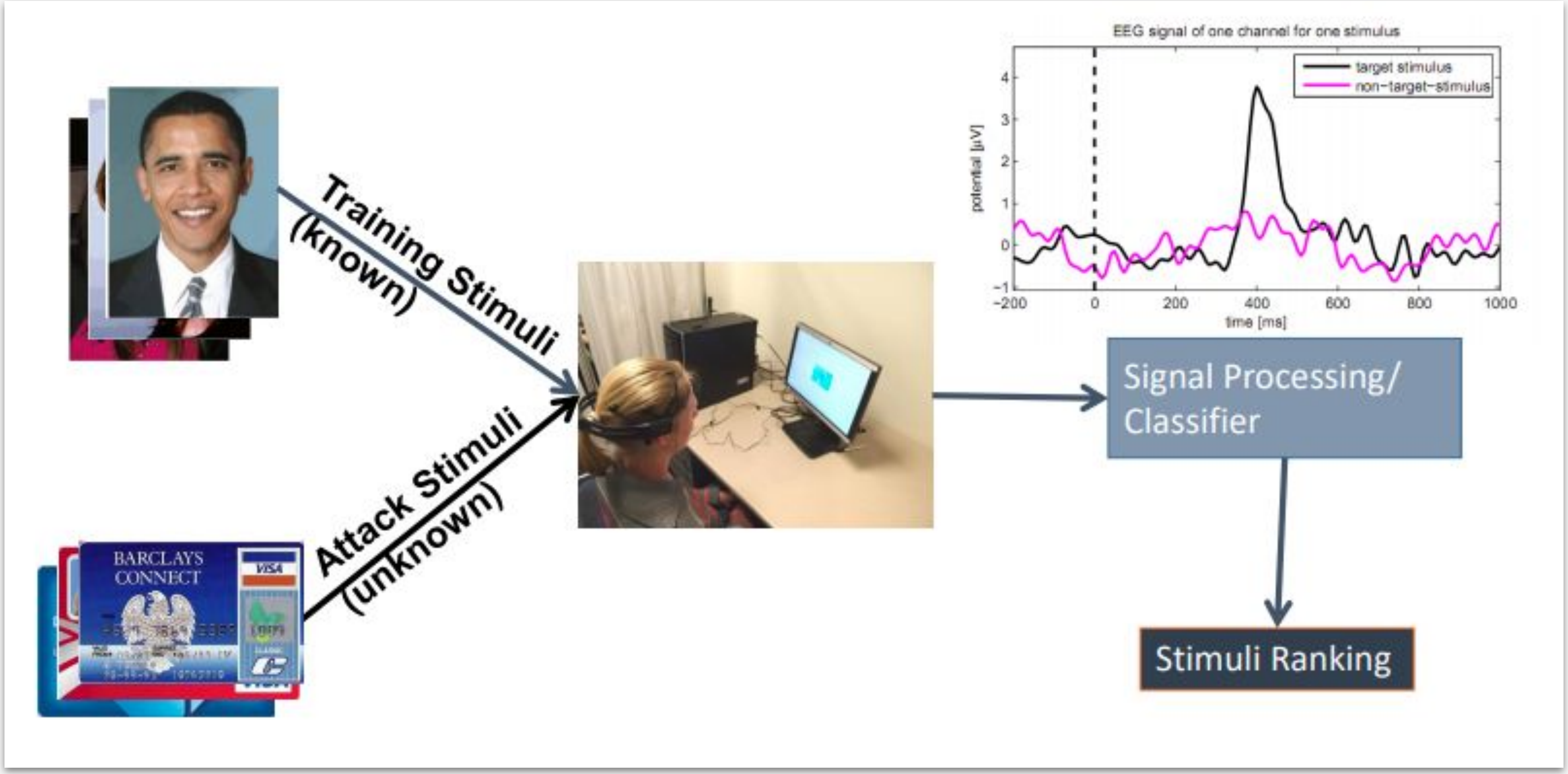
- There are variations on this theme (face swap, puppet-master, lip-sync).
- The scope of this presentation is limited to video.

Problems:

- Online content is demonstrably being used for criminal/reputation/influence operations.
- It has not been possible to easily fake people in videos.
- People therefore trust it more than, say, images, which can be easily faked by anyone with Photoshop.



BCI as Side-Channel to the Brain



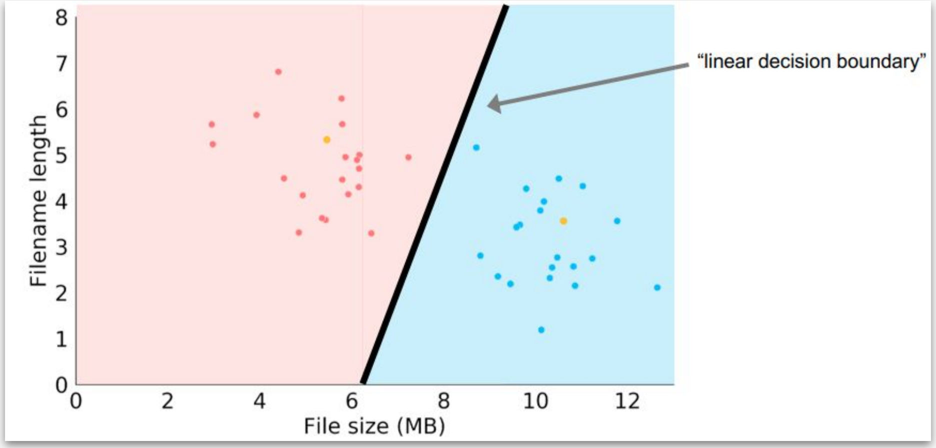
Training Stimuli (known)

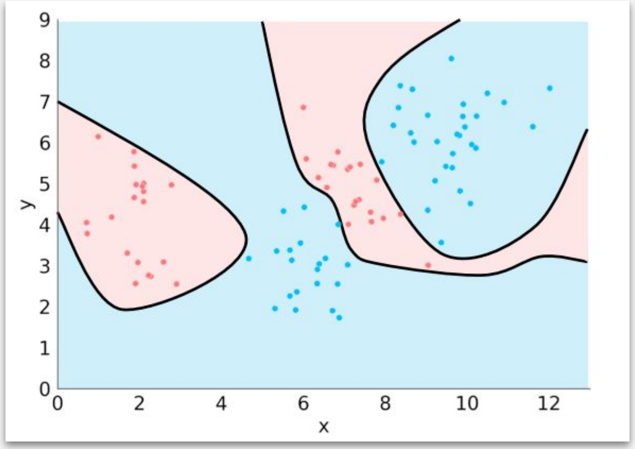
Attack Stimuli (unknown)

Signal Processing/ Classifier

Stimuli Ranking

Malware Detection





NON-LINEAR DECISION BOUNDARIES

Static features -extract “static” information about the file

- File size?
- String in the file?
- File header metadata?
- What DLLs does it use? What functions from each DLL does it call?
- Are parts of it encrypted? Compressed?

Dynamic features - watch the program as it executes

- What API functions does the sample call?
- Is it talking to the network?
- Is it reading/writing/removing files or registry entries?
- Is it operating crypto libraries?

Privacy

Duality of Privacy and Security

Security is about keeping unwanted traffic from entering one’s network.
Privacy is about keeping wanted information from leaving one’s network.

- Is this always true?
- Is it too simplistic?
- Does it yield any valuable insights for the privacy conundrum?
- What have been the success stories of security?
- How applicable are lessons from security to privacy?

Privacy is Essential for Survival

Required for strategy, to compete over limited resources.

Example: play a game while revealing to other players his hand or strategy!

Privacy reflects the autonomy and free will of the individual.

Example: in Christianity, individual choice between good and evil.

Privacy provides a mechanism for “forgetting” or not knowing of some forms of indiscretions.

Example: “go west young man”.

Anonymity

A key component of privacy is the notion of **anonymity**.

- What is anonymity? Closely tied to identity.
- Anonymity can thus be defined as ”without attribution to an individual”.
  - “On the Internet nobody knows you are a Sam”.

“one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.”

Anonymization techniques:

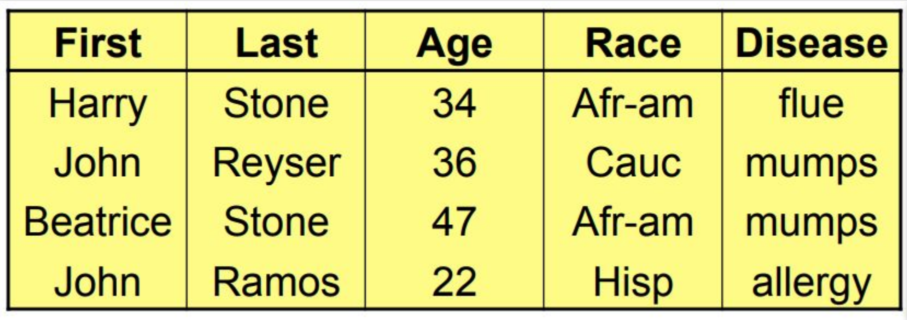
- **Context-dependent:** may require significantly different techniques depending on nature of data being anonymized.
- Metrics involve identifying a **balance between making data safe to be shared yet have high utility**.

K-Anonymity

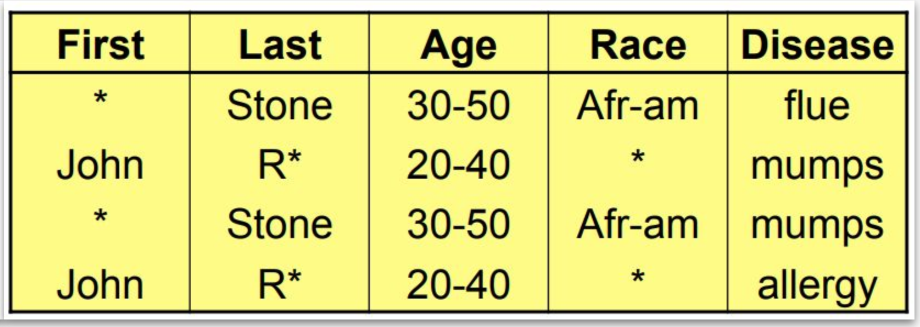
Definition: each tuple is equal to at least k-1 others.

Anonymizing: through suppression and generalization.

- Useful, but not really private.
- Hard: NP-complete for suppression only.
    - Approximations exists.



First	Last	Age	Race	Disease
Harry	Stone	34	Afr-am	flue
John	Reyser	36	Cauc	mumps
Beatrice	Stone	47	Afr-am	mumps
John	Ramos	22	Hisp	allergy



First	Last	Age	Race	Disease
*	Stone	30-50	Afr-am	flue
John	R*	20-40	*	mumps
*	Stone	30-50	Afr-am	mumps
John	R*	20-40	*	allergy

Forensic Analysis

Example

“Prove that email NOT was received”

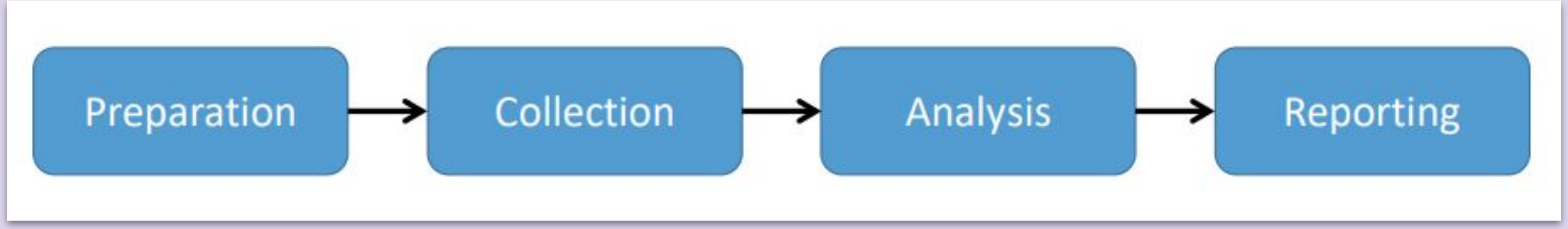
- Easy to prove that an email was received:
    - Look at email servers for logs.
- Does not finding traces mean that it was not received?
    - Maybe the investigation was not complete.
    - Maybe it was (cleverly) deleted.
- Reason for third-party logs.

What is Computer Forensics

Interpretation of computer media for digital evidence.

- **Who uses it:** Industry, Government, Individual.

Simplified Process Model:



Goal:
Provide an understanding of the sources of law and policy and how these concepts drive permissible actions based on various roles.

Questions to Cover (Preparation)

Has something happened at all?

- Hacked vs Bugs vs Random effect.

When did it happen and for how long?

What happened with what effects?

- Read vs Delete vs Modification.

Who did it and why?

- IP vs Username.

How did he do it?

Actions (Collection)

Conduct a search for evidence:

- e.g., Data, Logs, Passwords, Waste paper basket.

Package evidence:

- When possible, duplicate the systems.
- e.g., make a copy of the hard drive.

Limit access to the collected data.

- Maybe the who is an insider.

Practical actions for systems:

- Show active processes.
- List network configuration.
- Dump memory (including swap area).
- Stop system.
- Duplicate HDD.

Methods of Hiding Data (Analysis)

OS-level:

- Use dot-files, mark as hidden/system.

Change extension/naming:

- “password.txt” -> “fun.png”.

Unallocated disk sectors.

Steganography (usable for small data, e.g., secrets).

Encryption.

HDD Analysis:

- If possible (e.g., magnetic HDD) look at the physical level: platter, interfaces, etc.
- Second, look at low-level partitions, volumes, etc.
- Look at the boot process (ROM->BIOS->BOOTLOADER->KERNEL).
- Inspect virtualization configuration (if implemented).
- Inspect OS configuration.

Report

- Info of investigator/Confidence levels.
- Identification of case: Date, ID, etc.
- Subject of examination: system, network, etc.
- Summary of findings.
- Procedural history: steps taken, detailed findings.
- Results and conclusion: server was hacked by X, from Y to Z timeframe, A&B data was accessed; etc.
- Annex.

Place to Practice/Get Involve

CTF:

- Pwnable.kr
- pwnable.tw
- picoCTF
- Google CTF

RSS:

- Security Now
- Threatpost
- Google Project Zero

Conference:

- Blackhat, Defcon, RSA, SASCon.
- USENIX Security, CCS, S&P, NDSS.

Twitter:

- Security Community e.g. LiveOverflow.

Bug Bounty Market:

- Bug Bounty Program
- Intel's Bug Bounty Program
- Microsoft Bounty Programs
- Program Rules – Application Security
- Apple Security Bounty Program

Jobs:

- US Cyber Command (Home USCYBERCOM)
- NSA (National Security Agency)
- Federal Cyber Career (CyberCareers.gov)
- Google, Microsoft, Intel, Qualcomm, Tesla, Uber, etc.

< Data Security />