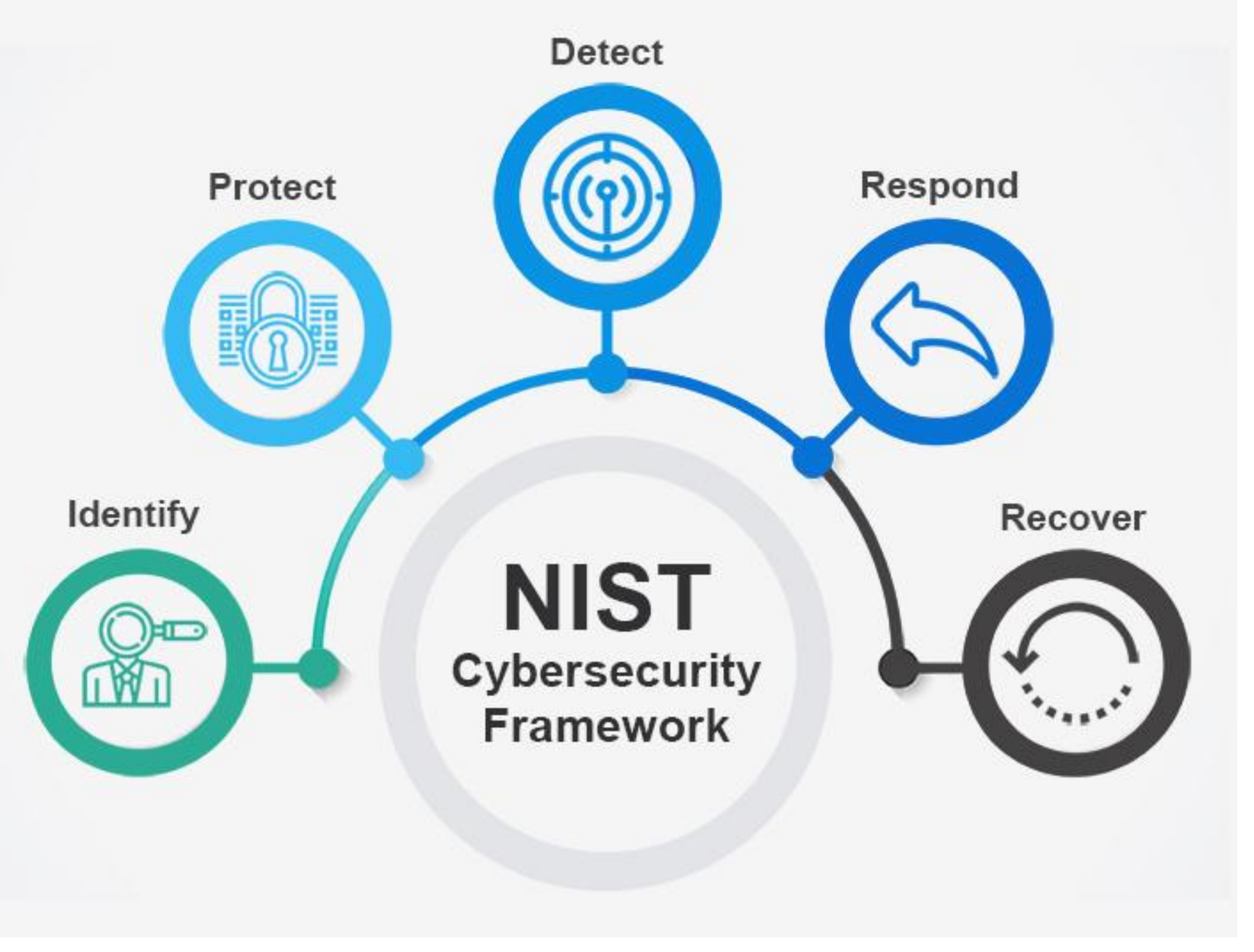
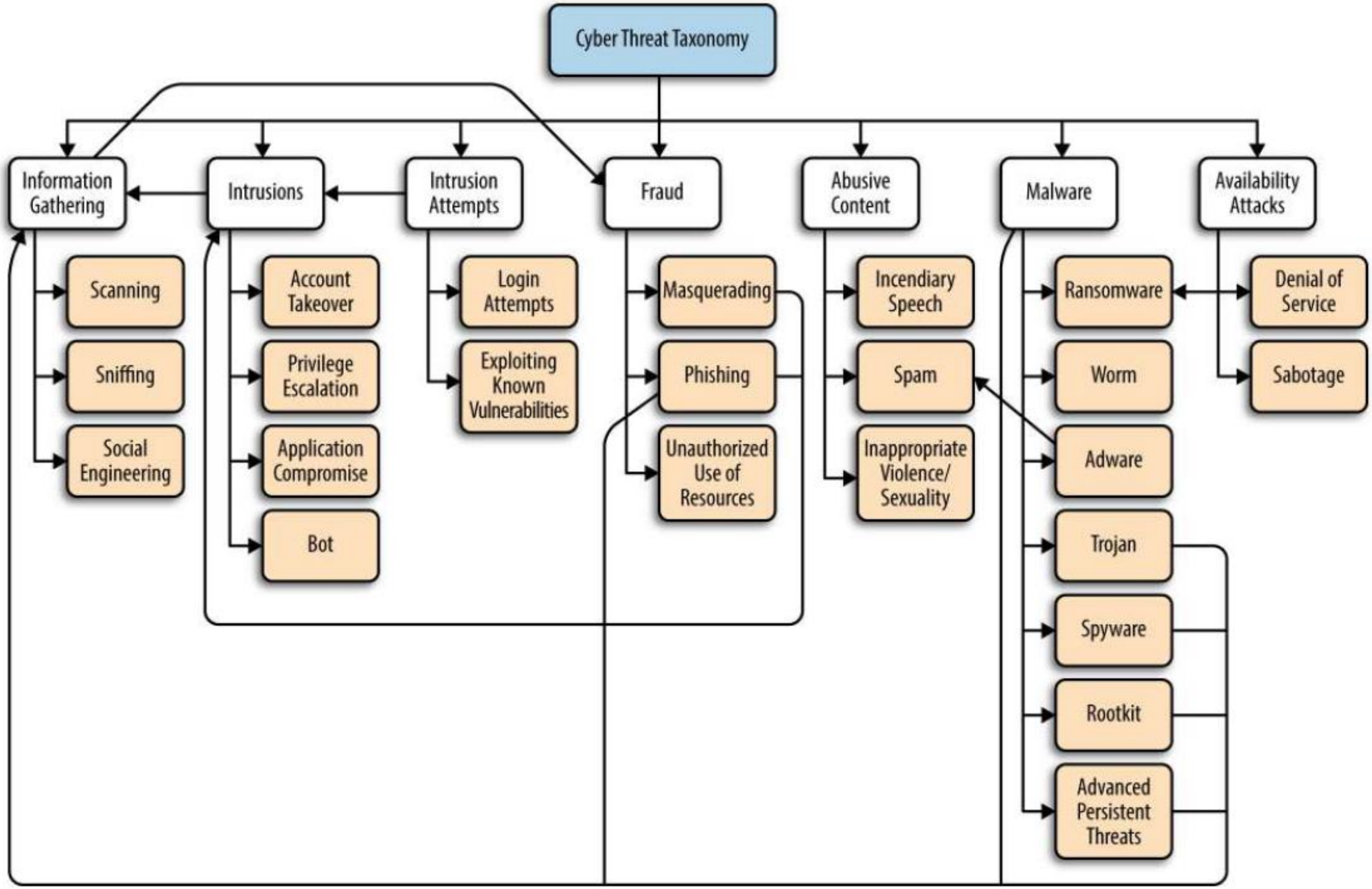
Machine learning and Applications in Cyber Security

Cyber Security

Cyber Security




Cyber Threat Taxonomy
Information Gathering
Intrusions
Intrusion Attempts
Fraud
Abusive Content
Malware
Availability Attacks
Scanning
Sniffing
Social Engineering
Account Takeover
Privilege Escalation
Application Compromise
Bot
Login Attempts
Exploiting Known Vulnerabilities
Masquerading
Phishing
Unauthorized Use of Resources
Incendiary Speech
Spam
Inappropriate Violence/ Sexuality
Ransomware
Worm
Adware
Trojan
Spyware
Rootkit
Advanced Persistent Threats
Denial of Service
Sabotage

AI in Cyber Security

Threat landscape:

- Changing very fast with billions of connected devices around the world
- Massive, largely automated botnets infecting consumer devices
- User data privacy: under threat of social engineering or phishing attacks

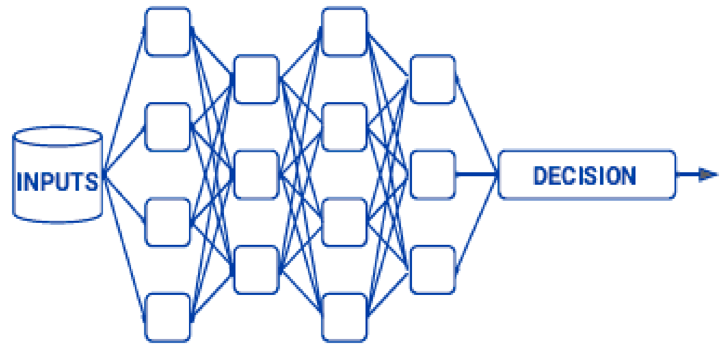
AI for cyber security todays not only the rule-based systems, but also ML- based ones

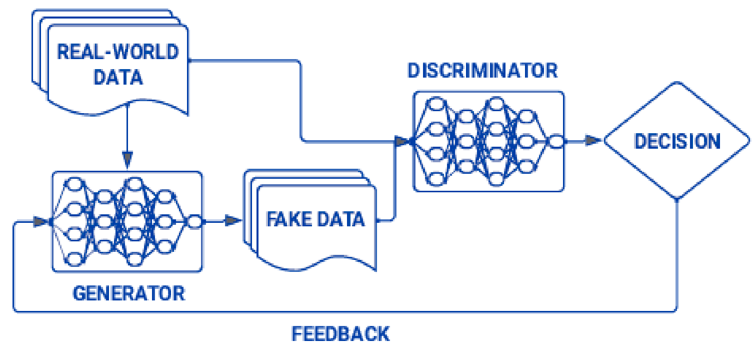
Main categories: Pattern recognition and Classification

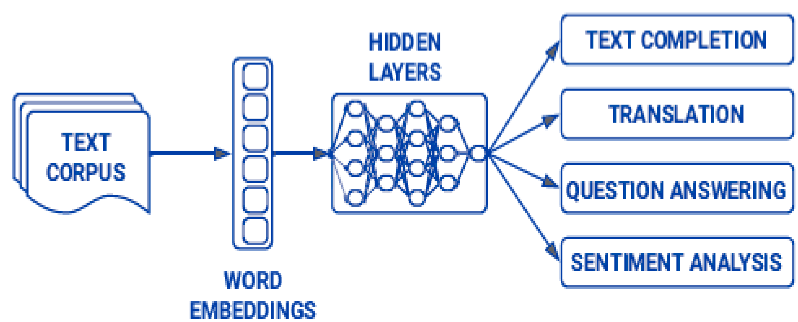
AI in Cyber Security

- 1. Spam mail and phishing page detection
- 2. Anomaly detection
- 3. DoS and DDoS attack detection
- 4. Malware detection and identification
- 5. Detection of advanced persistent threats
- 6. Detection of information leakage
- 7. Detection of hidden channels
- 8. Detection of software vulnerabilities.

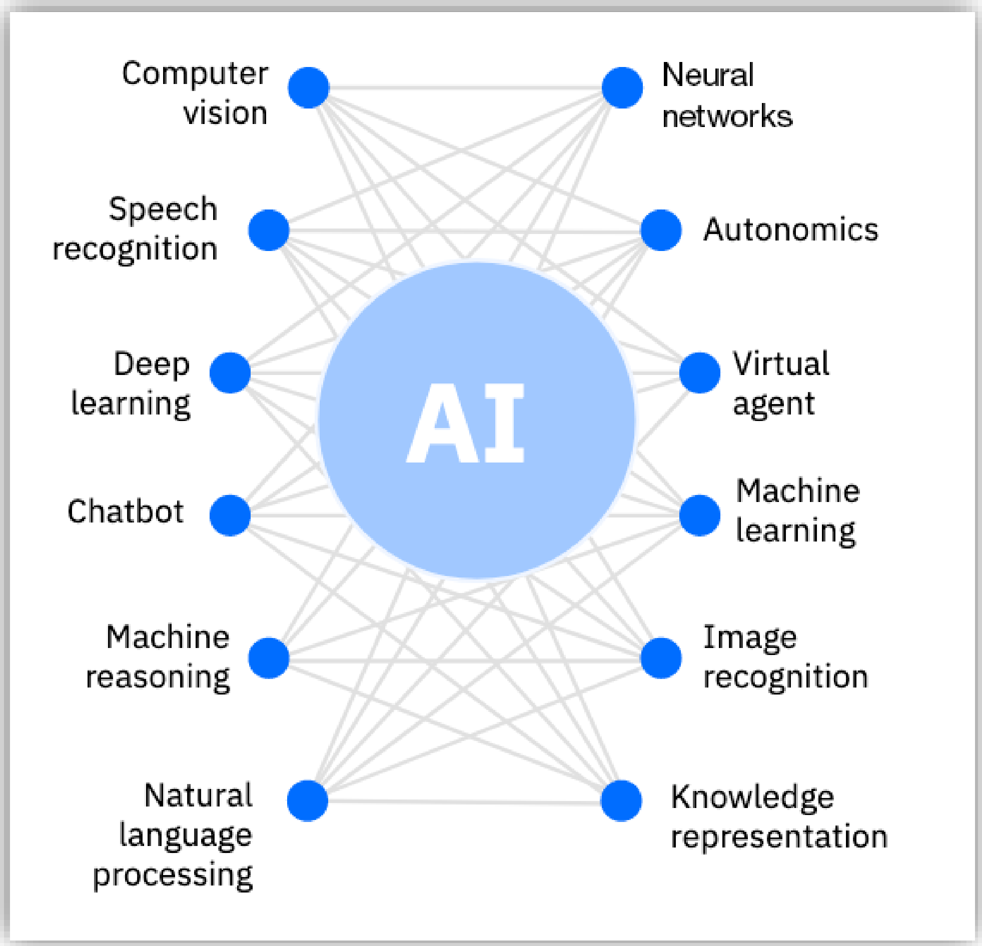
- 9. Biometric recognition
- 10. User identification and authentication
- 11. Detection of identity theft
- 12. Social media analytics

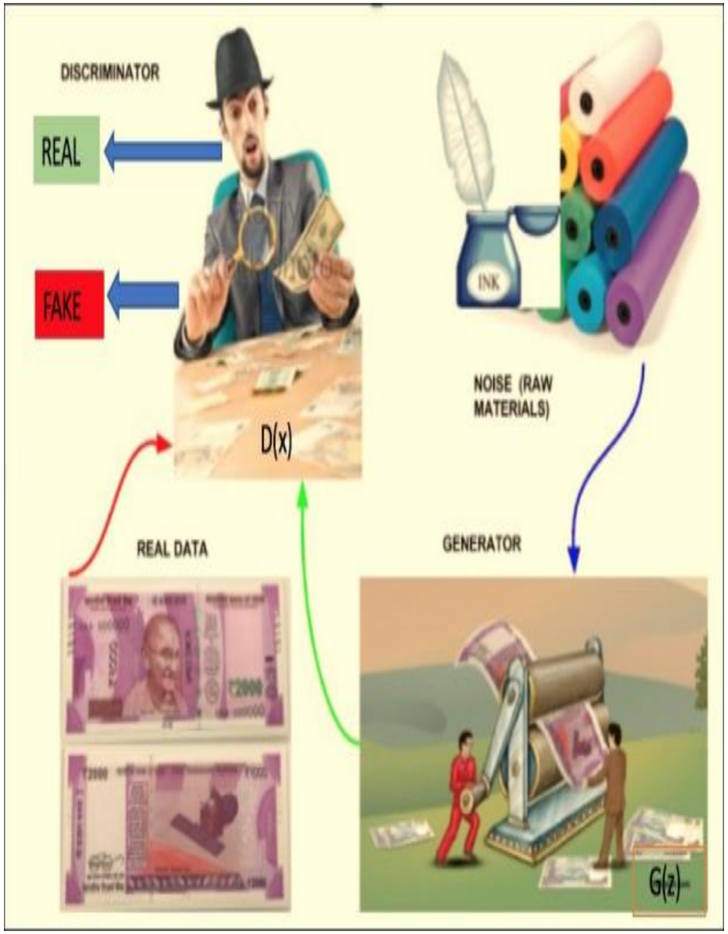






AI: Cutting Edges

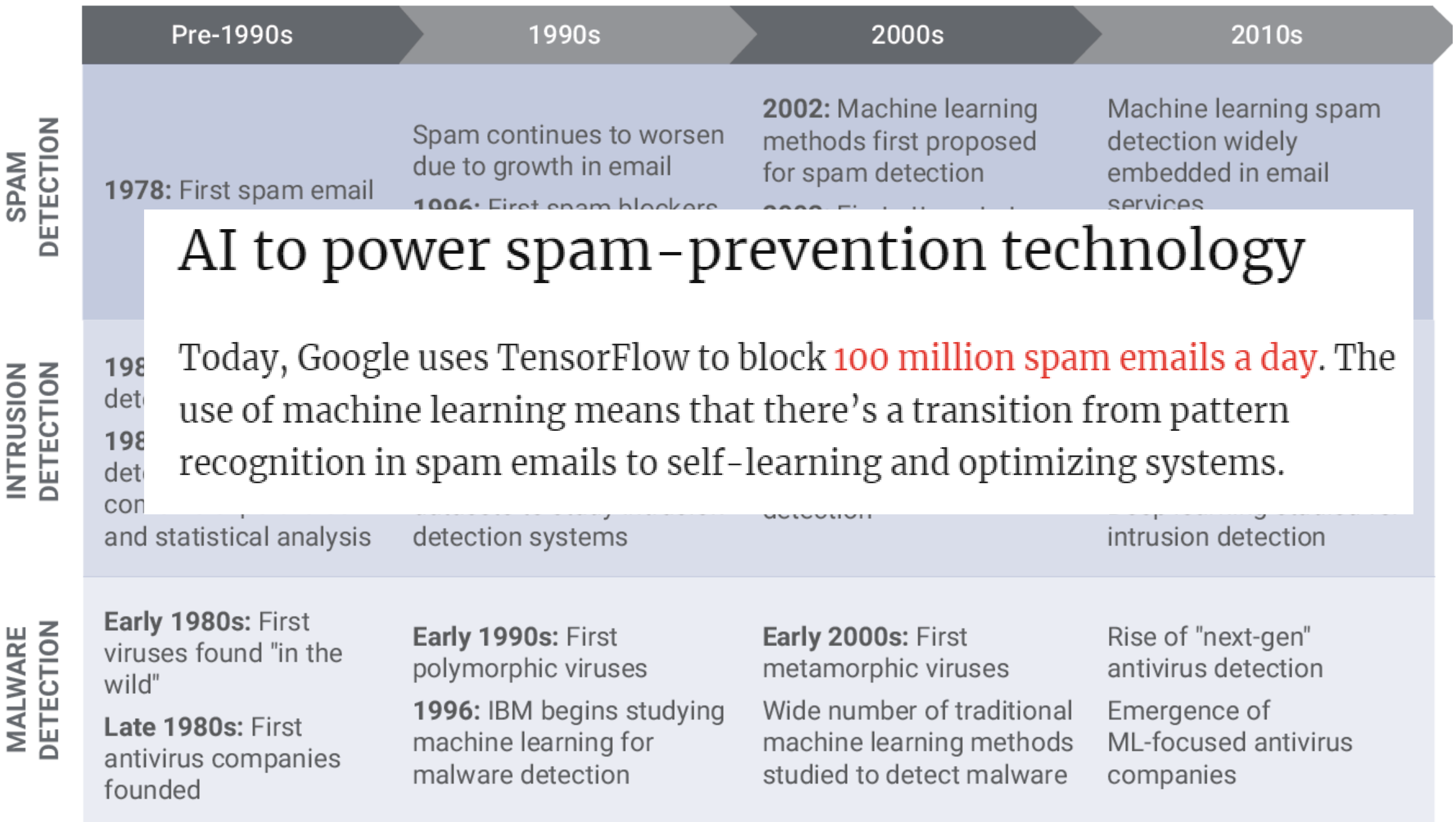




AI in Cyber Security

	Pre-1990s	1990s	2000s	2010s
SPAM DETECTION	**1978:** First spam email	Spam continues to worsen due to growth in email **1996:** First spam blockers	**2002:** Machine learning methods first proposed for spam detection **2003:** First attempts to regulate spam in the United States	Machine learning spam detection widely embedded in email services Emergence of deep learning-based classifiers
INTRUSION DETECTION	**1980:** First intrusion detection systems **1986:** Anomaly detection systems combine expert rules and statistical analysis	**Early 1990s:** Neural networks for anomaly detection first proposed **1999:** DARPA creates datasets to study intrusion detection systems	Machine learning further studied as a possible tool for misuse-based and anomaly-based intrusion detection	**Late 2010s:** Emergence of large-scale, cloud-based intrusion detection systems Deep learning studied for intrusion detection
MALWARE DETECTION	**Early 1980s:** First viruses found "in the wild" **Late 1980s:** First antivirus companies founded	**Early 1990s:** First polymorphic viruses **1996:** IBM begins studying machine learning for malware detection	**Early 2000s:** First metamorphic viruses Wide number of traditional machine learning methods studied to detect malware	Rise of "next-gen" antivirus detection Emergence of ML-focused antivirus companies

AI in Cyber Security



AI to power spam-prevention technology

Today, Google uses TensorFlow to block 100 million spam emails a day. The use of machine learning means that there's a transition from pattern recognition in spam emails to self-learning and optimizing systems.

AI in Cyber Security

	Pre-1990s	1990s	2000s	2010s
SPAM DETECTION	**1978:** First spam email	Spam continues to worsen due to growth in email **1996:** First spam blockers	**2002:** Machine learning methods first proposed for spam detection **2003:** First attempts to regulate spam in the United States	Machine learning spam detection widely embedded in email services Emergence of deep learning-based classifiers
INTRUSION DETECTION	**1980:** First intrusion detection systems **1986:** Anomaly detection systems combine expert rules and statistical analysis	**Early 1990s:** Neural networks for anomaly detection first proposed **1999:** DARPA creates datasets to study intrusion detection systems	Machine learning further studied as a possible tool for misuse-based and anomaly-based intrusion detection	**Late 2010s:** Emergence of large-scale, cloud-based intrusion detection systems Deep learning studied for intrusion detection
MALWARE DETECTION	**Early 1980s:** First viruses found "in the wild" **Late 1980s:** First antivirus companies founded	**Early 1990s:** First polymorphic viruses **1996:** IBM begins studying machine learning for malware detection	**Early 2000s:** First metamorphic viruses Wide number of traditional machine learning methods studied to detect malware	Rise of "next-gen" antivirus detection Emergence of ML-focused antivirus companies

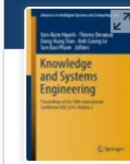
AI in Cyber Security

	Pre-1990s	1990s	2000s	2010s
SPAM DETECTION	**1978:** First spam email	Spam continues to worsen due to growth in email **1996:** First spam blockers	**2002:** Machine learning methods first proposed for spam detection **2003:** First attempts to regulate spam in the United States	Machine learning spam detection widely embedded in email services Emergence of deep learning-based classifiers
INTRUSION DETECTION	**1980:** First intrusion detection systems **1986:** Anomaly detection systems combine expert rules and statistical analysis	**Early 1990s:** Neural networks for anomaly detection first proposed **1999:** DARPA creates datasets to study intrusion detection systems	Machine learning further studied as a possible tool for misuse-based and anomaly-based intrusion detection	**Late 2010s:** Emergence of large-scale, cloud-based intrusion detection systems Deep learning studied for intrusion detection
MALWARE DETECTION	**Early 1980s:** First viruses found "in the wild" **Late 1980s:** First antivirus companies founded	**Early 1990s:** First polymorphic viruses **1996:** IBM begins studying machine learning for malware detection	**Early 2000s:** First metamorphic viruses Wide number of traditional machine learning methods studied to detect malware	Rise of "next-gen" antivirus detection Emergence of ML-focused antivirus companies

Example: Spam detection

the Spam Detection Rate (SDR) and the False Alarm Rate (FAR) seems to be most obvious criteria to measure the effectiveness of a spam detection resolution.

The final purpose of any Anti-Spam approach is to maximize the SDR and to minimize the FAR as much as possible.



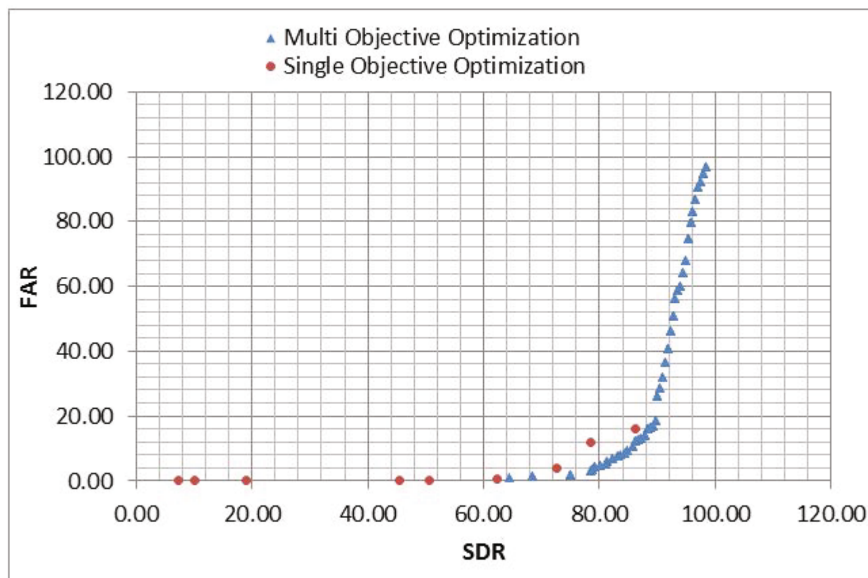
Knowledge and Systems Engineering pp 211–221 | Cite as

A Multi-objective Approach for Vietnamese Spam Detection

Minh Tuan Vu, Quang Anh Tran, Quang Minh Ha & Lam Thu Bui

Conference paper

Part of the Advances in Intelligent Systems and Computing book series (AISC,volume 245)



Example: Detection of software vulnerabilities

Software vulnerability detection is the process of confirming whether a software system contains flaws

Code injection attacks are one of the most powerful and common attacks against software applications

ML can be used to model the syntax and semantics of the code, infer code patterns to analyze large code bases, and assist in code auditing and code understanding

Example: Detection of software vulnerabilities

IEEE *Xplore*® Browse My Settings Help Institutional Sign In

All

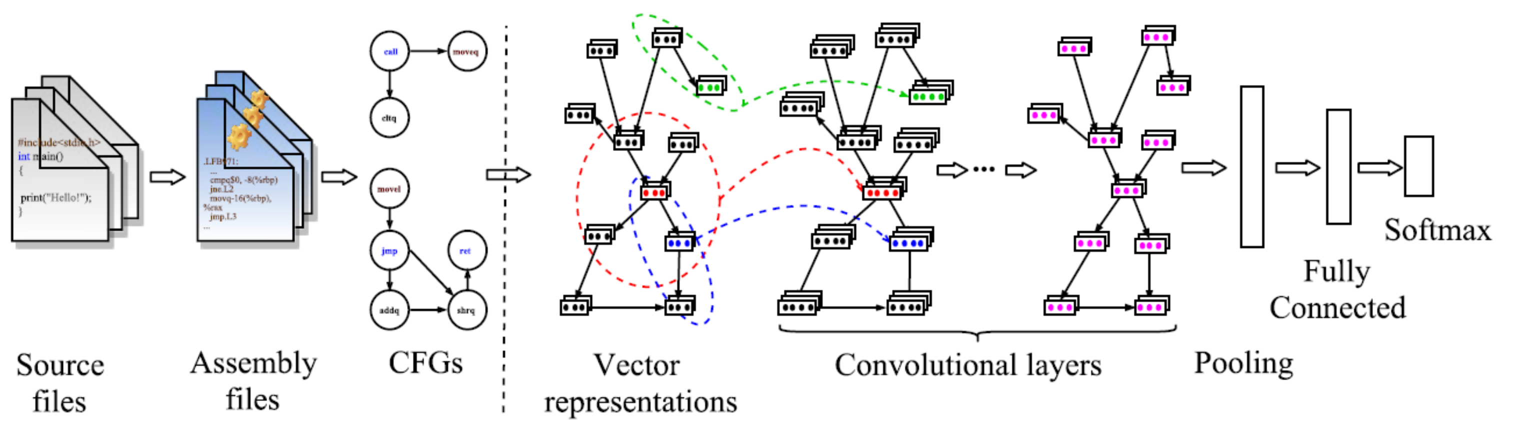
ADVANCED SEARCH

Conferences > 2017 IEEE 29th International ...

Convolutional Neural Networks over Control Flow Graphs for Software Defect Prediction

Publisher: IEEE Cite This PDF

Anh Viet Phan ; Minh Le Nguyen ; Lam Thu Bui **All Authors**



(a) Building Control Flow Graphs **(b) The convolutional neural network on directed graphs**

Fig. 3: The overview of our approaches for software defect prediction using convolutional neural networks on Control Flow Graphs of assembly code.

Example: Biometric recognition

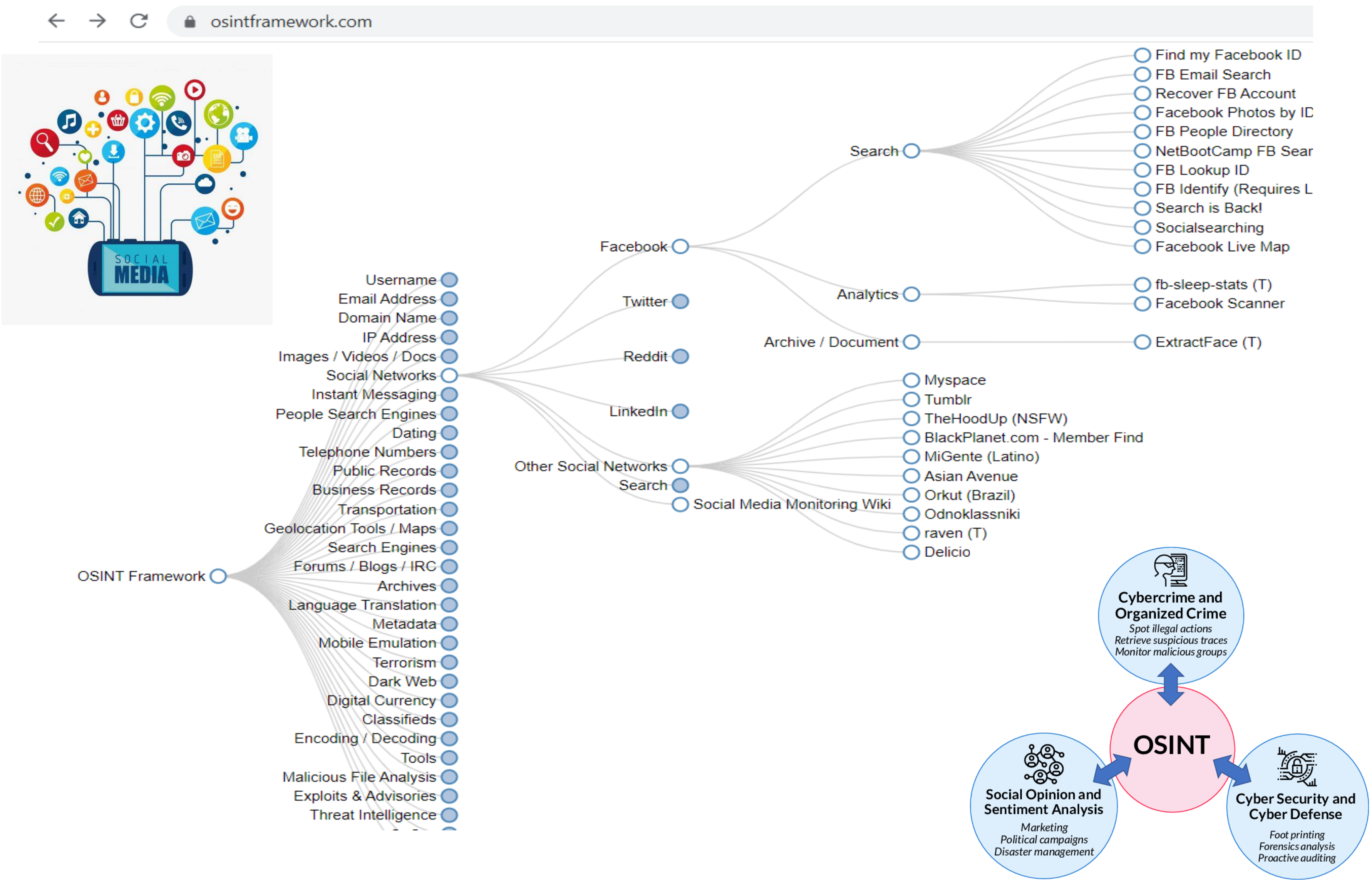
Modality	Features	ML techniques
Face	Distance between eyes, DCT, Fourier transform, Ratio of distance between eyes and nose, Principal components,	PCA, LDA, Kernel PCA, Kernel LDA, SVM, Deep neural network
Iris	DCT, Fourier transform, Wavelet transform, Principal components, Texture features,	PCA, LDA,
Fingerprint	Delta, Core points, Ridge ending, Island, Bifurcation, Minutiae, FFT	Artificial neural networks, Support vector machine, Genetic algorithms, Bayesian training, Probabilistic models
Finger vein	LBP, Minutiae, Bifurcation and end points, Pixel information	SVM, Deep learning
Palm print	Shape, Texture, Palm lines, PCA, LDA coefficients, DCT	Naive Bayes, k-nearest neighbor, HMM
Palm vein	LBP, Minutiae, Bifurcation and end points, Pixel information	SVM, Deep learning
Voice	Linear prediction coefficient (LPC), Cepstral coefficient (CC), MFCC features	Gaussian mixture models, HMM, ANN, SVM deep learning

Example: Biometric recognition

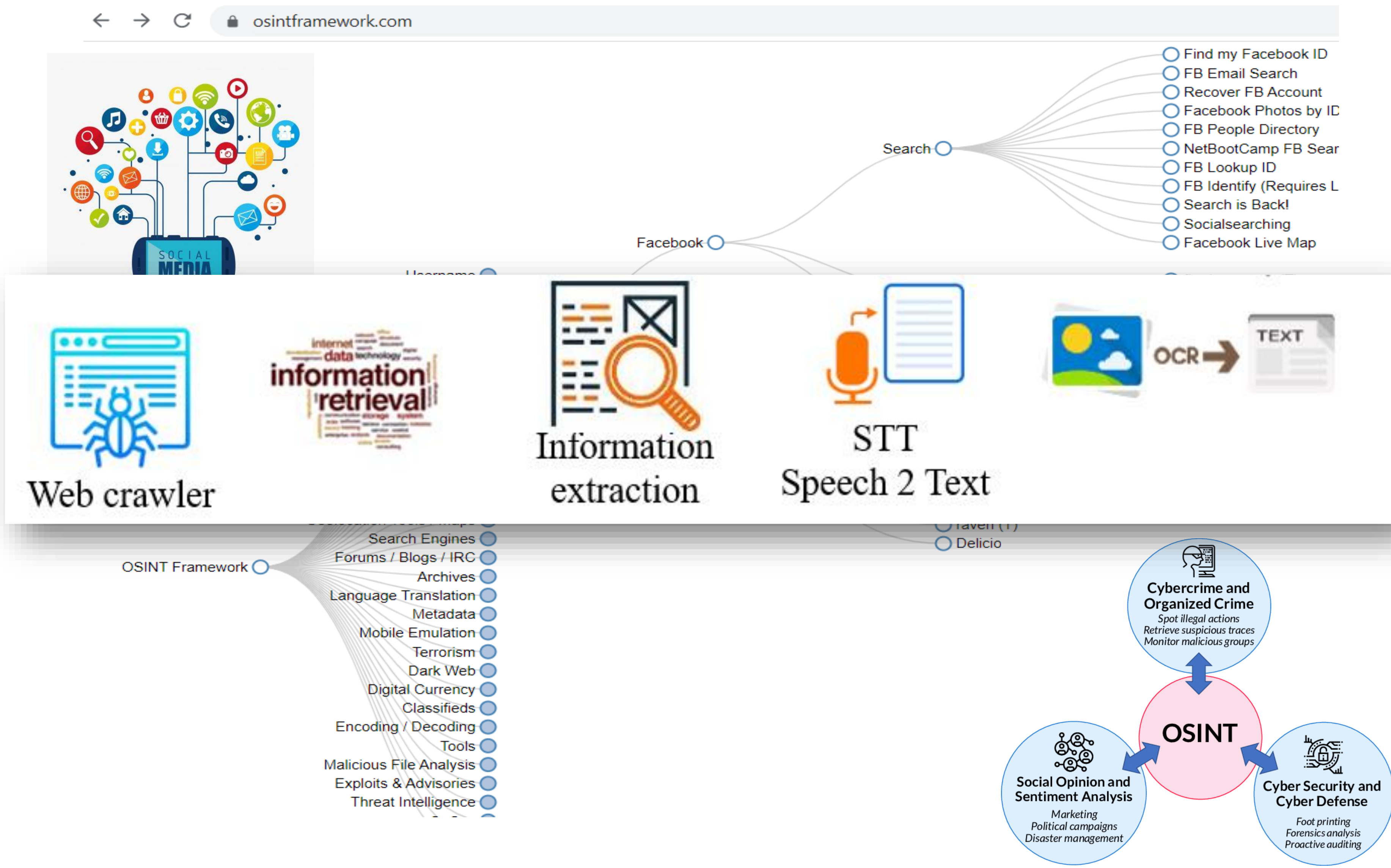
Related problems

1. **Face reconstruction:** using Generative Adversary Networks GAN
2. **Facial detection** YOLO và Haar-like
3. **Facial recognition** FaceNet
4. **Frame selection:** entropy based selection

Social media analytics and OSINT

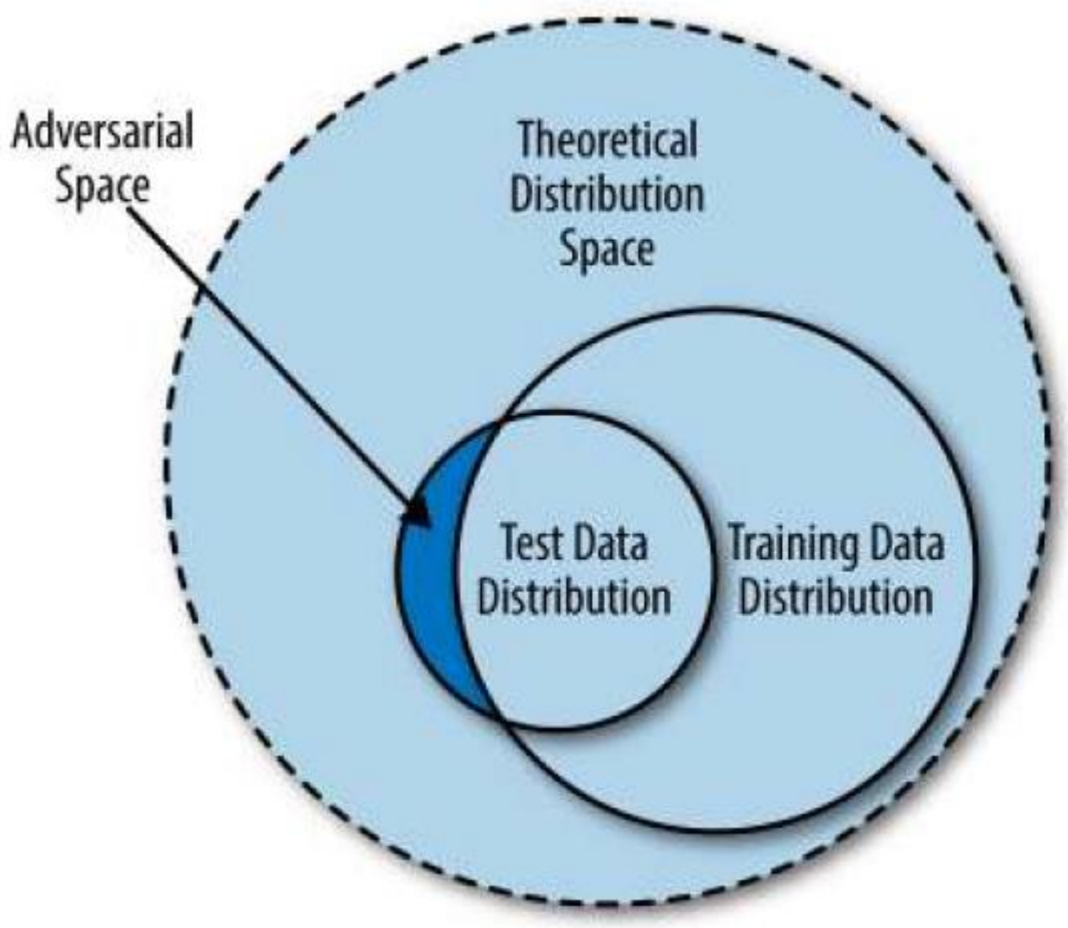


OSINT



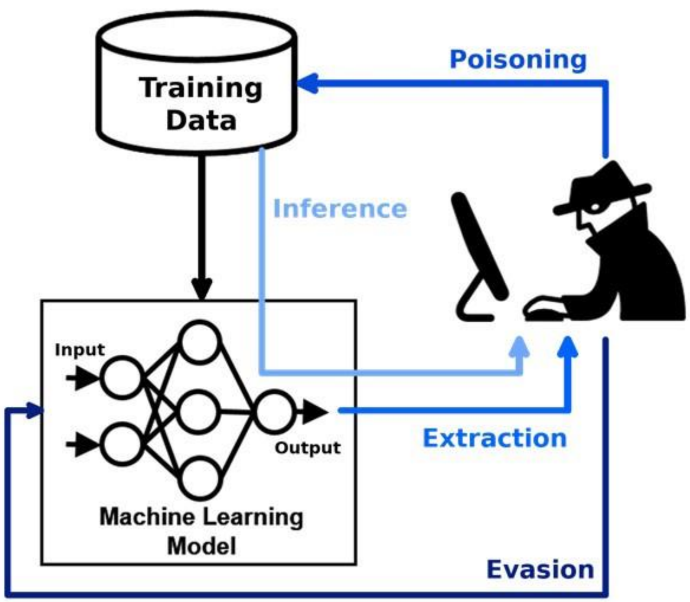
osintframework.com
Find my Facebook ID
FB Email Search
Recover FB Account
FB People Directory
FB Lookup ID
Search is Back!
Socialsearching
Facebook Live Map
Search
Facebook
SOCIAL
Web crawler
information
retrieval
Information extraction
STT
Speech 2 Text
OCR
TEXT
Delicio
Search Engines
Forums / Blogs / IRC
OSINT Framework
Archives
Language Translation
Metadata
Mobile Emulation
Terrorism
Dark Web
Digital Currency
Classifieds
Encoding / Decoding
Tools
Malicious File Analysis
Exploits & Advisories
Threat Intelligence
Cybercrime and Organized Crime
Spot illegal actions
Retrieve suspicious traces
Monitor malicious groups
OSINT
Social Opinion and Sentiment Analysis
Marketing
Political campaigns
Disaster management
Cyber Security and Cyber Defense
Foot printing
Forensics analysis
Proactive auditing

Security of machine learning models



Security of machine learning models

- Poisoning Attacks



Security of machine learning models

- Evasion Attacks

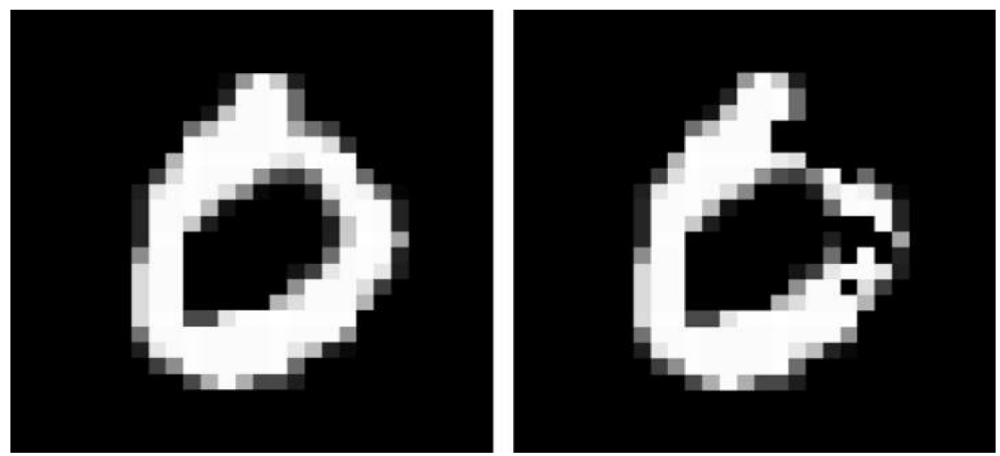
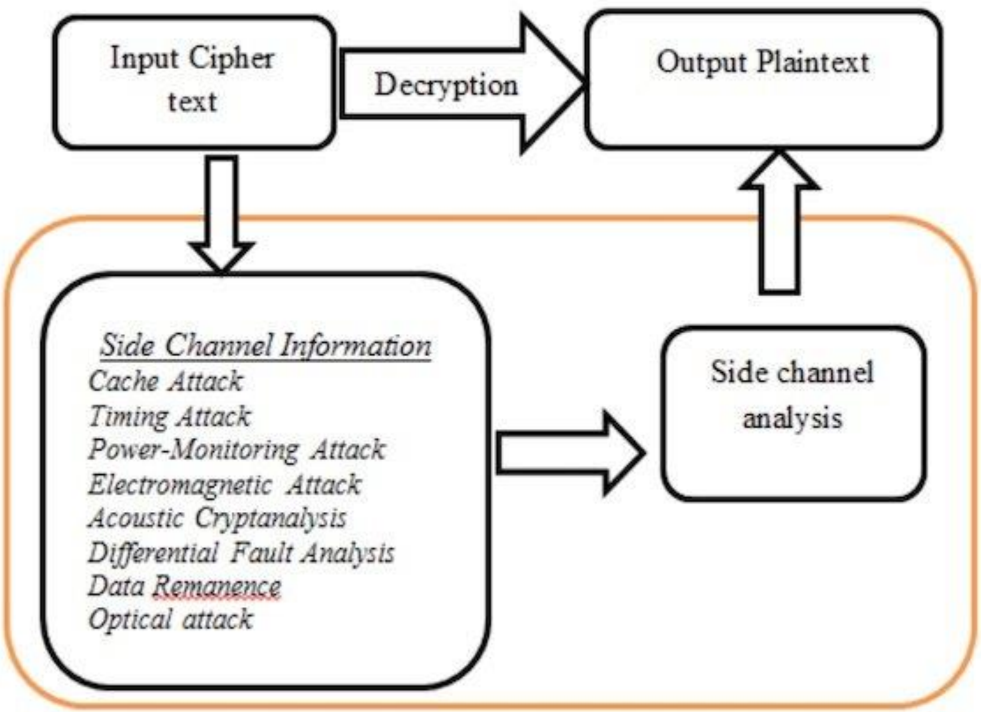
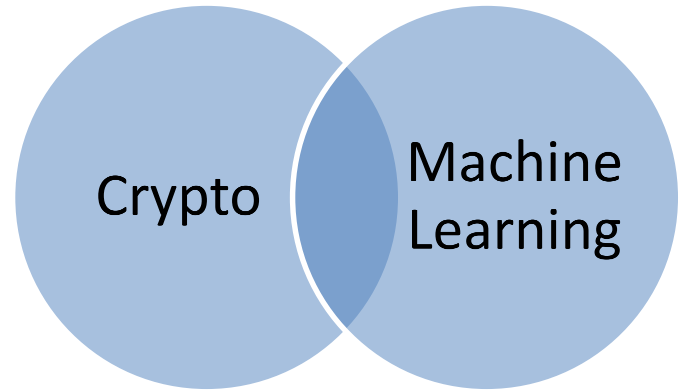


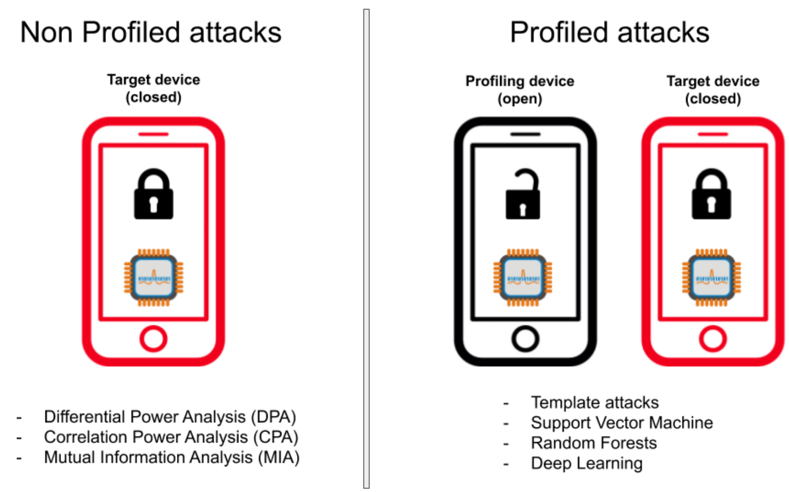
Figure 8-9. Comparison of the unaltered MNIST handwritten digit of a 0 (left) with the adversarially perturbed version (right)

Cryptography and Machine Learning

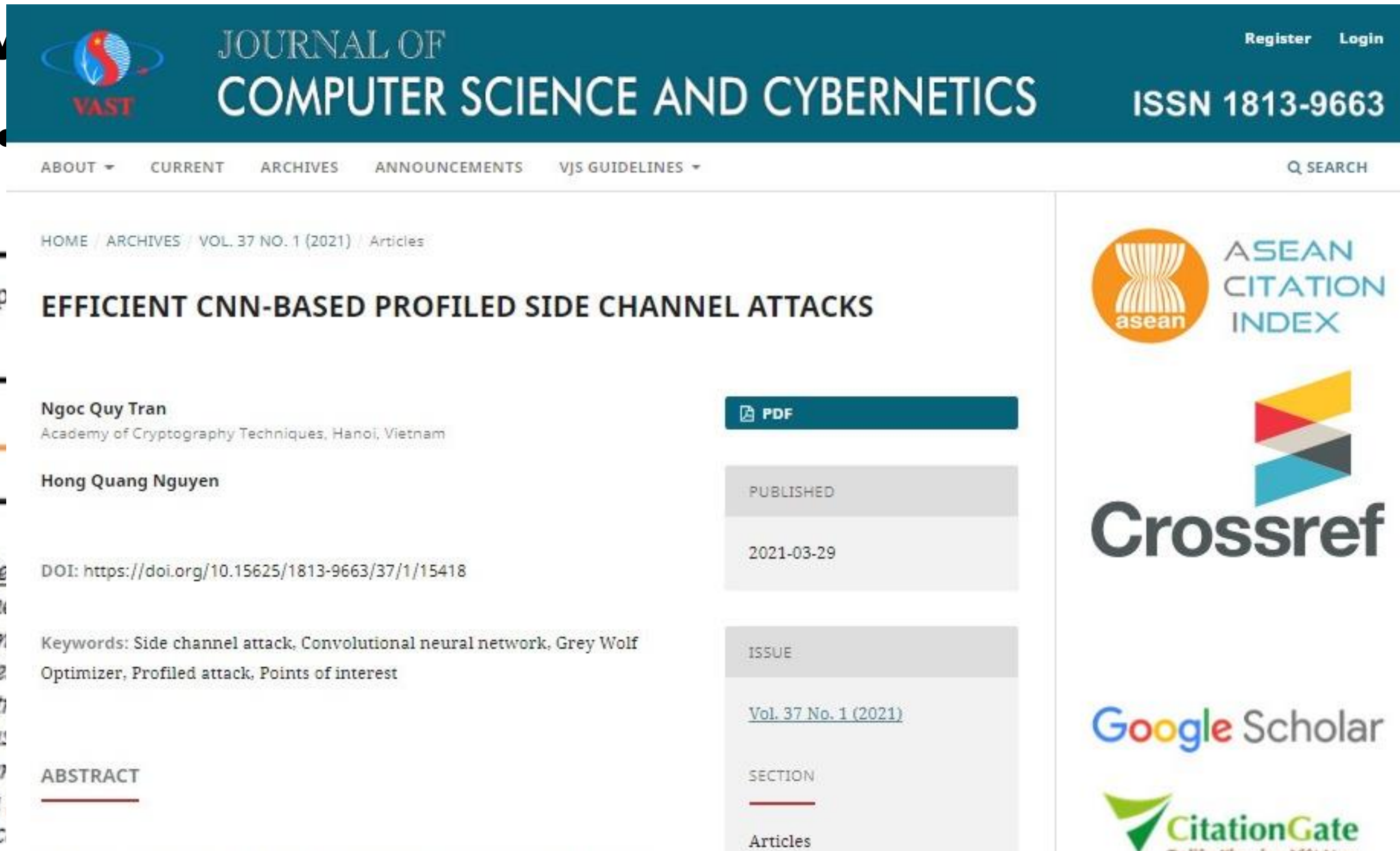
- **ML for Cryptography**
  - Side channel attacks







Cryptography and Machine Learning


JOURNAL OF COMPUTER SCIENCE AND CYBERNETICS
Register Login
ISSN 1813-9663
ABOUT CURRENT ARCHIVES ANNOUNCEMENTS VJS GUIDELINES SEARCH
HOME / ARCHIVES / VOL. 37 NO. 1 (2021) / Articles

EFFICIENT CNN-BASED PROFILED SIDE CHANNEL ATTACKS

Ngoc Quy Tran
Academy of Cryptography Techniques, Hanoi, Vietnam
Hong Quang Nguyen

PDF

PUBLISHED
2021-03-29
DOI: https://doi.org/10.15625/1813-9663/37/1/15418
Keywords: Side channel attack, Convolutional neural network, Grey Wolf Optimizer, Profiled attack, Points of interest
ISSUE
Vol. 37 No. 1 (2021)
SECTION
Articles

ABSTRACT

ASEAN CITATION INDEX
Crossref
Google Scholar
CitationGate
Tư liệu Khoa học Việt Nam

- Correlation Power Analysis (CPA)
- Mutual Information Analysis (MIA)

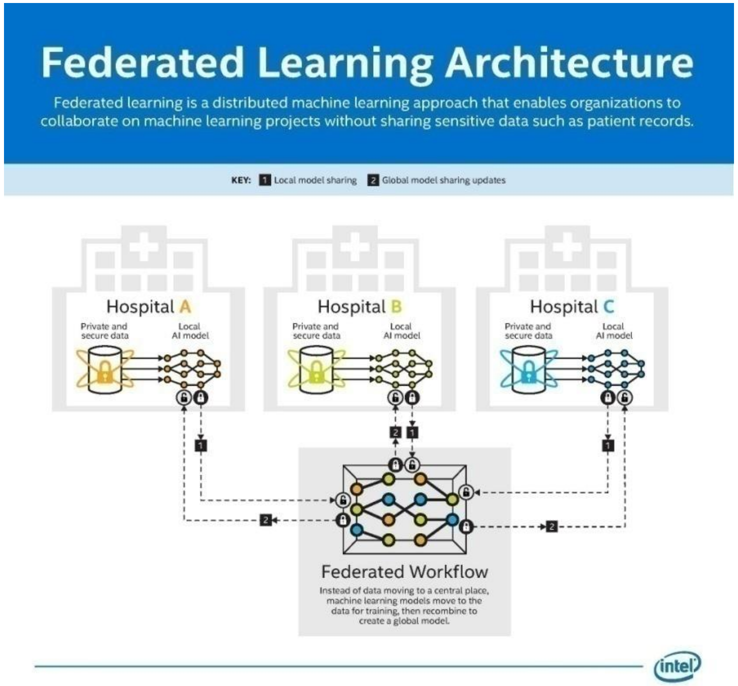
attacks
Target device (closed)
attacks
ector Machine
- Random Forests
- Deep Learning

Cryptography and Machine Learning

- Crypto for ML
  - Security for training data: Data encrypted

Security of Machine Learning models

- Security for training data: distributed learning
- Machine Learning as a Service (MLaaS)



Conclusion

- Machine Learning and Cryptography have a natural similarity
- The application of AI in cryptography and information security is inevitable.
- Challenges: Data for model training.