IARIA Congress 2024

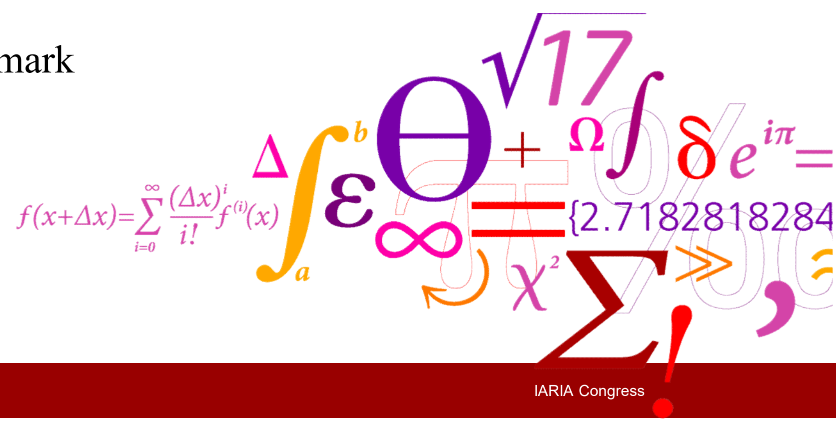
Machine Learning in Intrusion Detection: Introduction and Challenges

Weizhi Meng

Associate Professor

DTU Compute

Technical University of Denmark, Denmark



Research Directions

Weizhi Meng
weme@dtu.dk

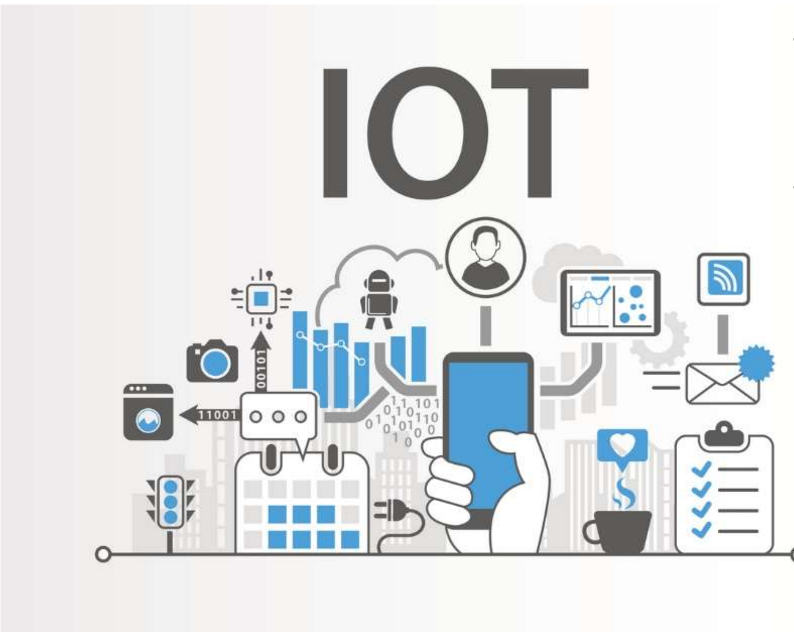
http://www.staff.dtu.dk/weme

- Intrusion Detection
- Biometric Authentication
- Trust Management
- HCI Security (Smartphone Security
- Blockchain

Outline

- Background on Intrusion Detection
- Machine Learning in IDS
- Open Challenges
- Discussion on Blockchain

Internet of Things



- Healthcare
- Vehicular
- Smart-home
- Smart-city …

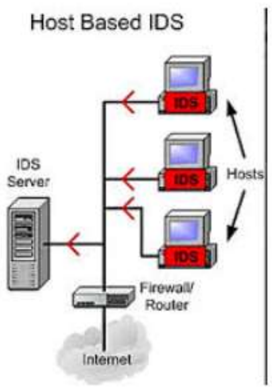
- Privacy and data sharing, where hackers could gain access to employees' personal devices and expose sensitive client data or even company trade secrets.
- Security threats, increased numbers of IoT devices susceptible to attack, risky if an IoT device is compromised (botnet, ransonware).
- How to handle the massive amounts of data produced by all of these IoT devices.

Intrusion Detection System (IDS)
definition

- *Intrusion detection is the process of identifying and responding to malicious activity targeted at computing and networking resources*
  - Process
    - Interaction between people and tools, it takes time
  - Identifying
    - Before, during or after the intrusion
  - Responding
    - Collect evidence, limit damage (honey pots), shut-out
  - Malicious activity
    - Intentional attempts to do harm
  - Computing and networking resources
    - Logical intrusions as opposed to physical intrusions

IDS Types

- **Host Based Systems**
  - Inspects local information
    - Application log-files
    - System log-files
    - etc.



- **Network Based Systems**
  - Inspects traffic on the network
  - Network Events

Signature Based Detection Systems

Relies on established patterns (“signatures”) in malicious network traffic

Similar to the signatures used in virus checkers

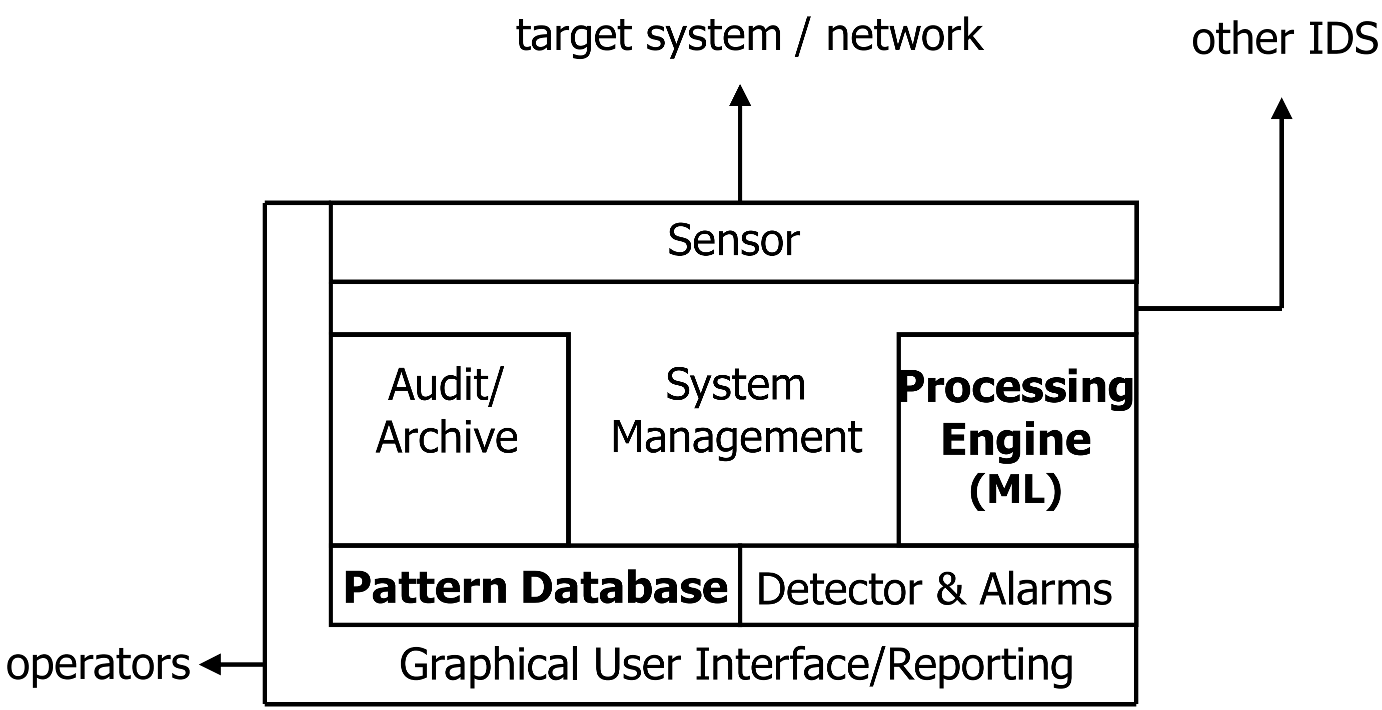
Anomaly-Based Detection Systems

Establishes base line for normal communication

Detects abnormal traffic pattern

Employs techniques from machine learning, pattern matching, big data

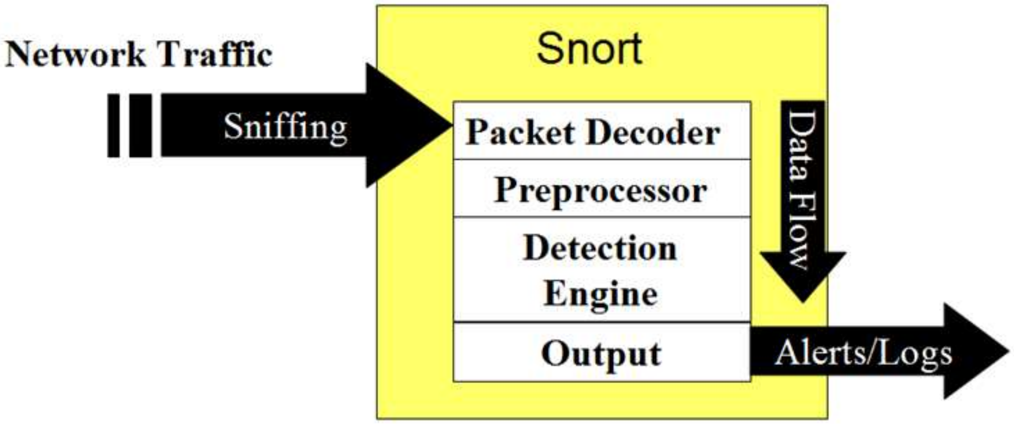
IDS Components



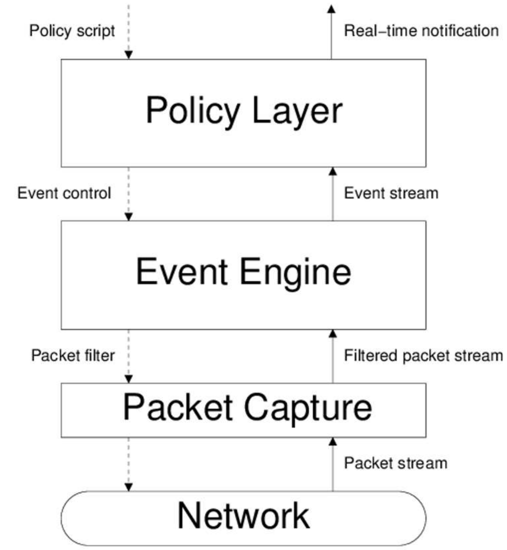
In Academia: Snort / Bro (Zeek)

Homepage: https://www.snort.org/

- ***Snort*** is an open-source signature-based network intrusion detection system.



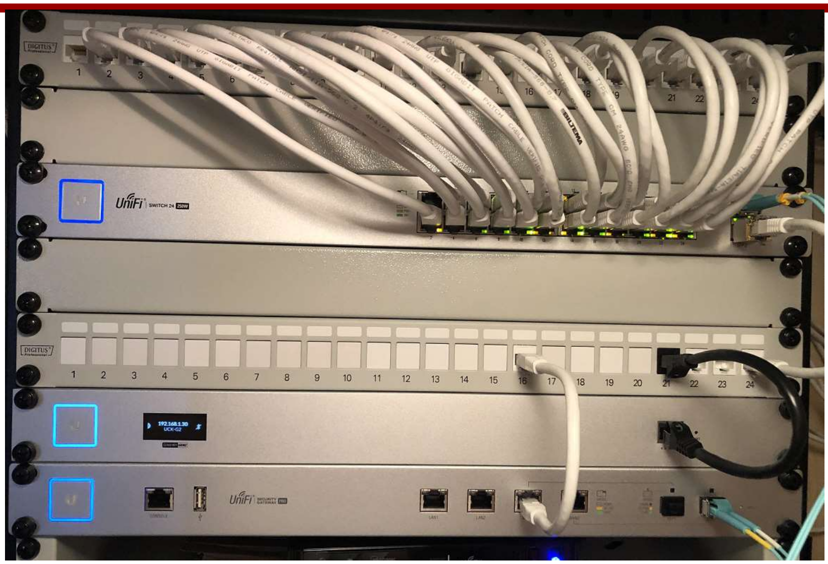
- ***Bro*** is an open-source anomaly-based network intrusion detection system.



DTU

In Industry: Intrusion Prevention System (IPS)

PaloAlto IPS 1.2 Gbps



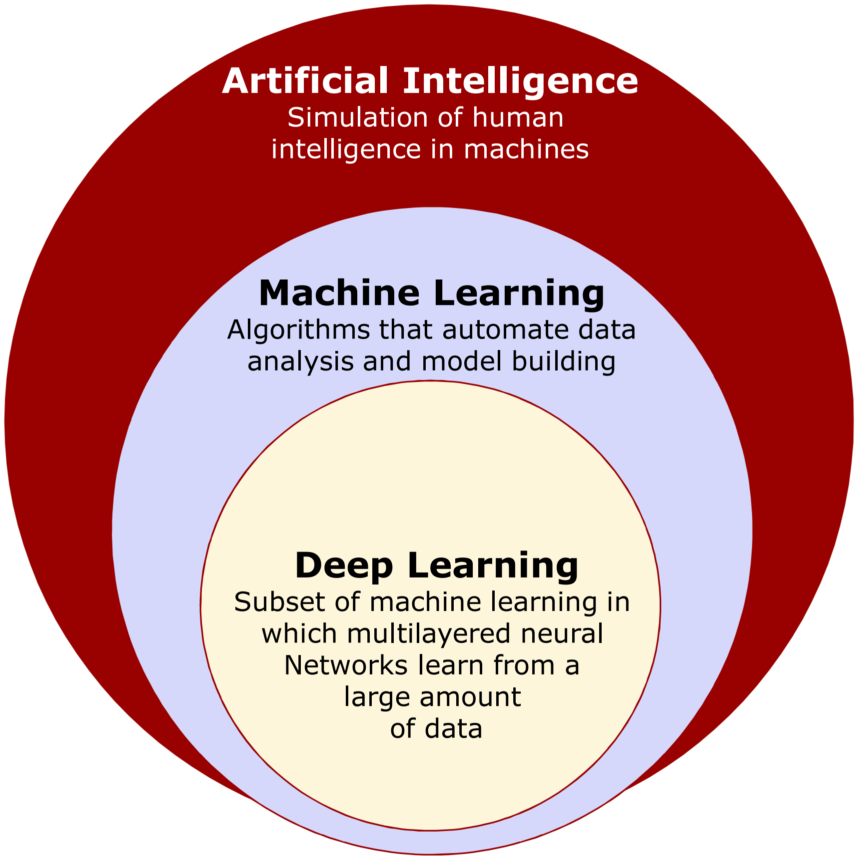
Unifi Security Gateway Pro

Trend Micro IPS

Outline

- Background on Intrusion Detection
- Machine Learning in IDS
- Open Challenges
- Discussion on Blockchain

Machine Learning Scope



Year of 1999

A Data Mining Framework for Building Intrusion Detection Models*

Wenke Lee
Salvatore J. Stolfo
Kui W. Mok
Computer Science Department, Columbia University
500 West 120th Street, New York, NY 10027
{wenke,sal,mok}@cs.columbia.edu

Abstract

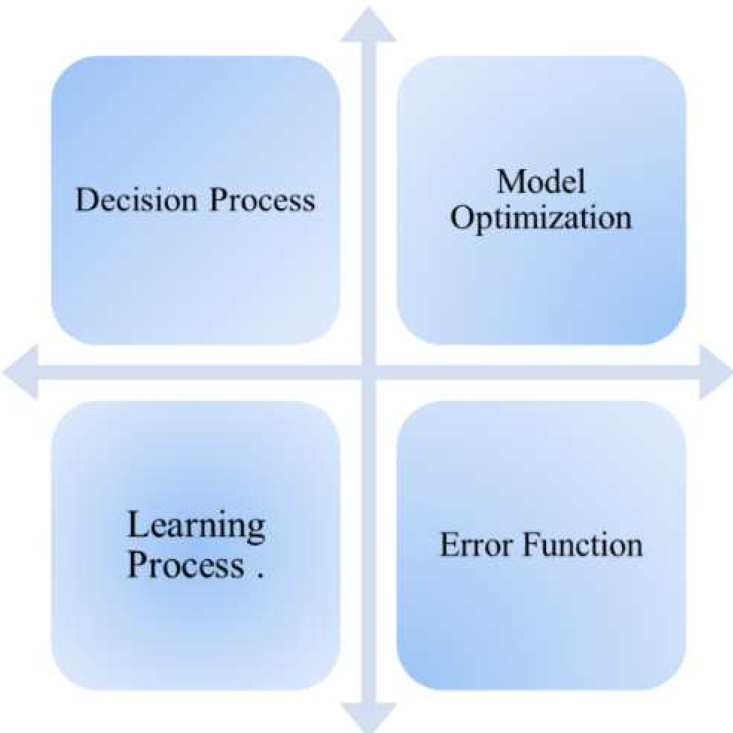
There is often the need to update an installed Intrusion Detection System (IDS) due to new attack methods or upgraded computing environments. Since many current IDSs are constructed by manual encoding of expert knowledge, changes to IDSs are expensive and slow. In this paper, we describe a data mining framework for adaptively building Intrusion Detection (ID) models. The central idea is to utilize auditing programs to extract an extensive set of features that describe each network connection or host session, and apply data mining programs to learn rules that accurately capture the behavior of intrusions and normal activities. These rules can then be used for misuse detection and anomaly detection. New detection models are incorporated into an existing IDS through a meta-learning (or co-operative learning) process, which produces a meta detection model that combines evidence from multiple models. We discuss the strengths of our data mining programs, namely, classification, meta-learning, association rules, and frequent episodes. We report our results of applying these programs to the extensively gathered network audit data for the 1998 DARPA Intrusion Detection Evaluation Program.

computer systems.

Intrusion detection techniques can be categorized into *anomaly detection* and *misuse detection*. Anomaly detection systems, for example, IDES [14], flag observed activities that deviate significantly from the established normal usage profiles as anomalies (i.e., possible intrusions). Misuse detection systems, for example, IDIOT [9] and STAT [5], use patterns of well-known attacks or weak spots of the system to match and identify known intrusion, patterns or signatures.

While accuracy is the essential requirement of an IDS, its extensibility and adaptability are also critical in today's network computing environment. There are multiple "penetration points" for intrusions to take place in a network system. For example, at the network level carefully crafted "malicious" IP packets can crash a victim host; at the host level, vulnerabilities in system software can be exploited to yield an illegal root shell. Since activities at different penetration points are normally recorded in different audit data sources, an IDS often needs to be extended to incorporate additional modules that specialize on certain components (e.g., hosts, subnets, etc.) of the network systems. The large traffic volume in security related mailing lists and Web sites suggest that new system security holes and intrusion methods are

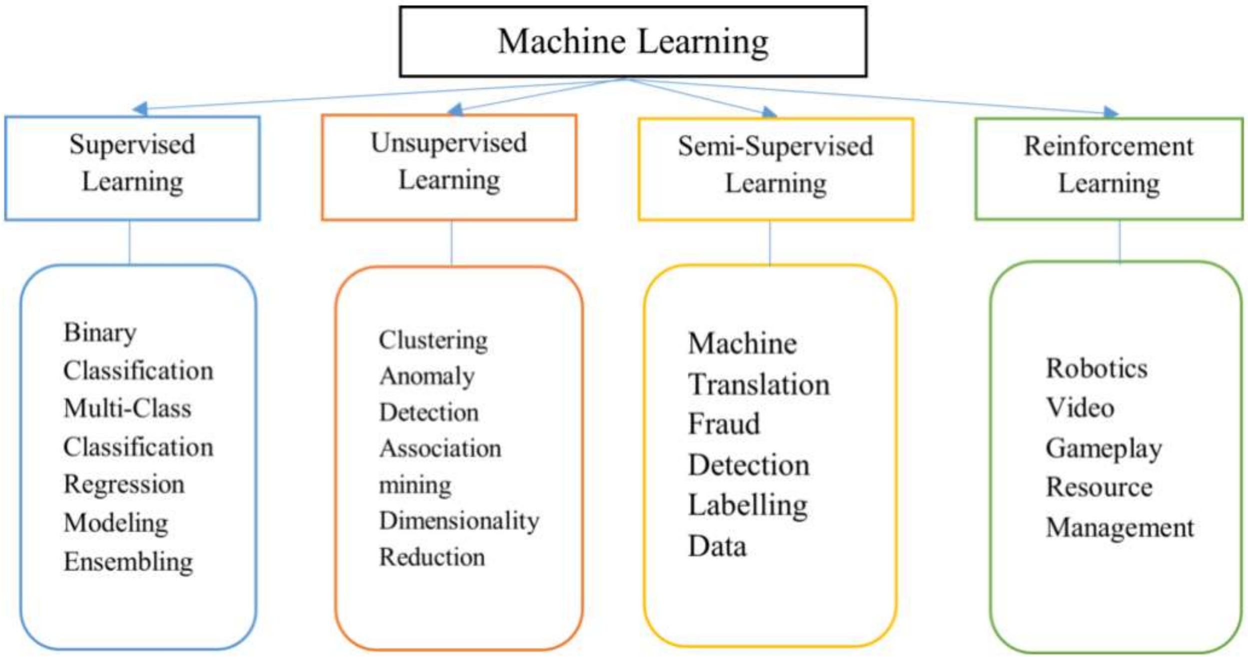
Learning Process



- **Decision making process:** when some raw dataset is given to the algorithm, it predicts a pattern.
- **An Error Function:** depicts the percentage of failing to achieve the desired output.
- **Model Optimization process,** weights are modified to minimize the difference between estimated output and actual output – update weights.

https://insights2techinfo.com/future-scope-of-machine-learning-and-ai-in-2022/

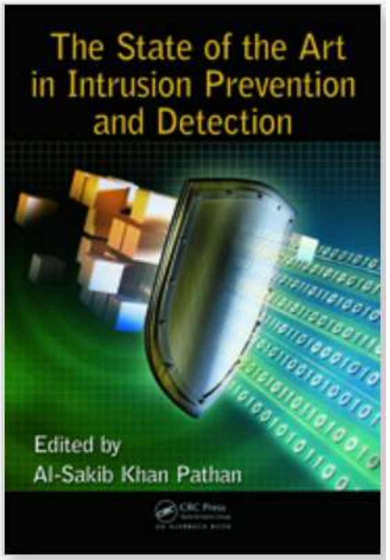
Machine Learning Classification



- **Reinforcement Learning**
  - learning the optimal behavior in an environment to obtain maximum reward.
- **Deep Learning**
  - machine learning methods based on artificial neural networks with three or more layers
- **Transfer Learning**
  - Train on one dataset but apply to another dataset / domain

https://insights2techinfo.com/future-scope-of-machine-learning-and-ai-in-2022/

Machine Learning in IDS (1)



Supervised Learning

K-Nearest Neighbor Classifier

Support Vector Machine

Decision Tree

Naïve Bayes Classifier

Neural Networks

Fuzzy Logic

Genetic Algorithms

Hybrid Classifiers

TABLE 13.1
Hybrid Classifiers Based on Genetic Algorithms

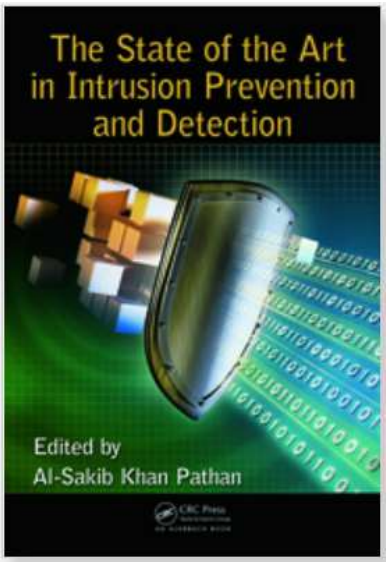
	Neural Network	SVM	Fuzzy	KNN	Decision Tree
Genetic Algorithm	[82,90]	[79–81]	[83,84,87,89]	[86,88]	[91]

100+ references

Applications of Machine Learning in Intrusion Detection (Chapter). The State of the Art in Intrusion Prevention and Detection (Book), Al-Sakib Khan Pathan (eds), CRC Press, Taylor & Francis, pp. 311-332, January 2014.

DTU

Machine Learning in IDS (2)



Supervised Learning

K-Nearest Neighbor Classifier

Support Vector Machine

Decision Tree

Naïve Bayes Classifier

Neural Networks

Fuzzy Logic

Genetic Algorithms

Hybrid Classifiers

100+ references

Applications of Machine Learning in Intrusion Detection (Chapter). The State of the Art in Intrusion Prevention and Detection (Book), Al-Sakib Khan Pathan (eds), CRC Press, Taylor & Francis, pp. 311-332, January 2014.

EAI Endorsed Transactions
on Security and Safety
Research Article EAI.EU

A Comprehensive Survey on Intrusion Detection based Machine Learning for IoT Networks

Hela Mliki[1,2,*], Abir Hadj Kaceam[2], Lamia Chaari[2]

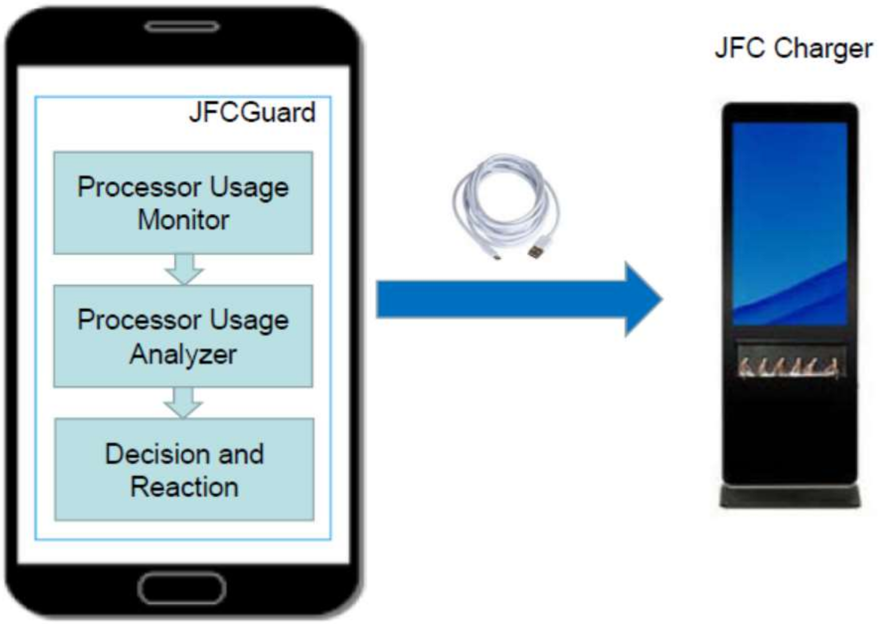
[1]Science Faculty of Gabes, University of Gabes, Tunisia
[2]Laboratory of Technology and Smart Systems (LT2S), Digital Research Center of Sfax (CRNS), University of Sfax, Tunisia

refine by year
2022 (31)
2021 (116)
2020 (66)
2019 (42)
2018 (39)
2017 (19)
2016 (11)
2015 (5)
2014 (7)
2013 (4)

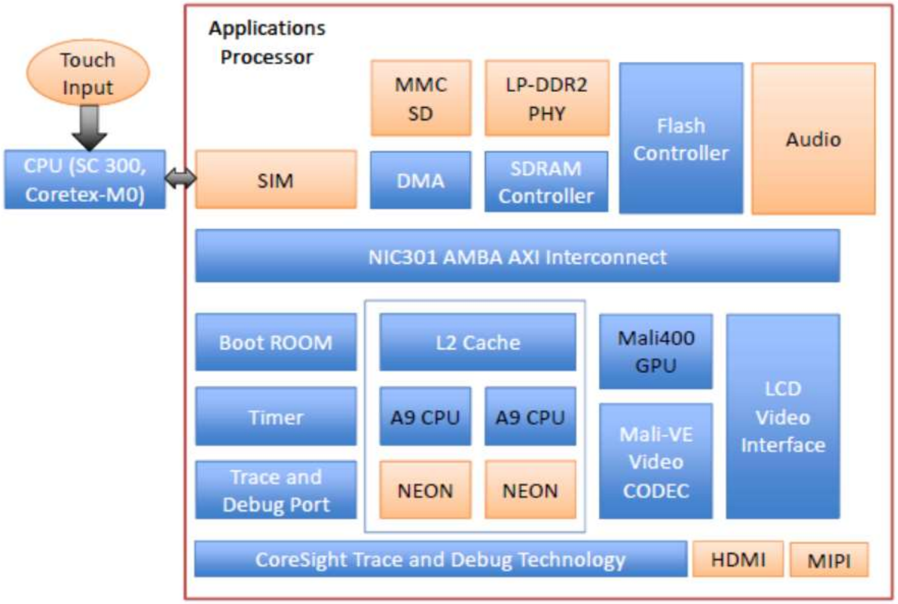
Keyword: machine learning + intrusion detection

[+]
[–] Search dblp
powered by CompleteSearch, courtesy of Hannah Bast, University of Freiburg

JFCGuard Framework

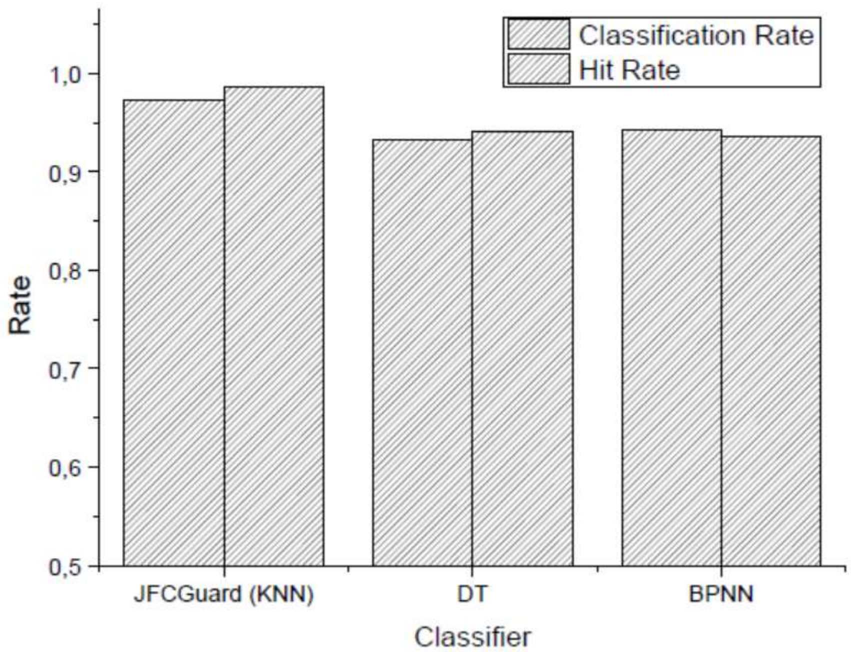


A Typical ARM-based Smartphone Hardware Architecture



W. Meng et al.: JFCGuard: Detecting juice filming charging attack via processor usage analysis on smartphones. Comput. Secur. 76: 252-264 (2018)

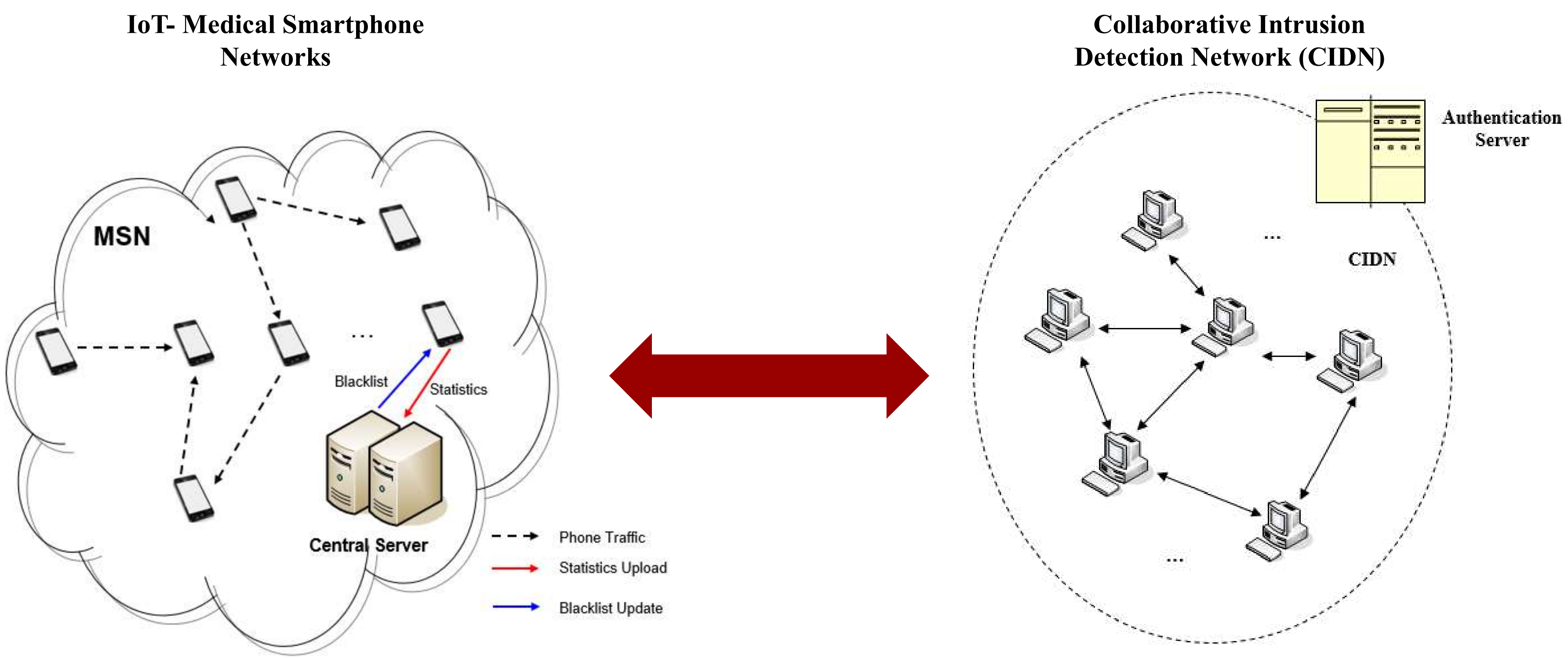
W. Meng et al.: Towards Detection of Juice Filming Charging Attacks via Supervised CPU Usage Analysis on Smartphones. Computers and Electrical Engineering, vol. 78, pp. 230-241 (2019)



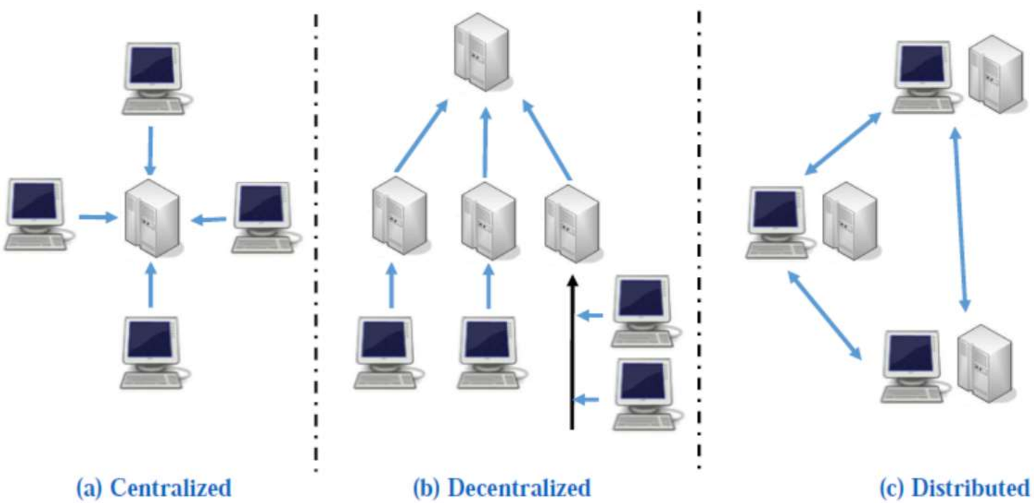
We ran the training and classification for ten times. The average classification accuracy and average hit rate are shown in the figure. **It is found that JFCGuard could achieve the best performance with a CA of around 0.97 and a HR of around 0.99, as compared with DT and BPNN.**

Moreover, we tested JFCGuard on smartphones and identified that the time consumption of JFCGuard (KNN), BPNN and DT is 1.4, 10.2, and 2.9 seconds. These results demonstrate that KNN is an appropriate classifier that can be adopted by JFCGuard, and that JFCGuard is promising and effective in detecting JFC attack.

Collaborative Intrusion Detection



Collaboration Topology

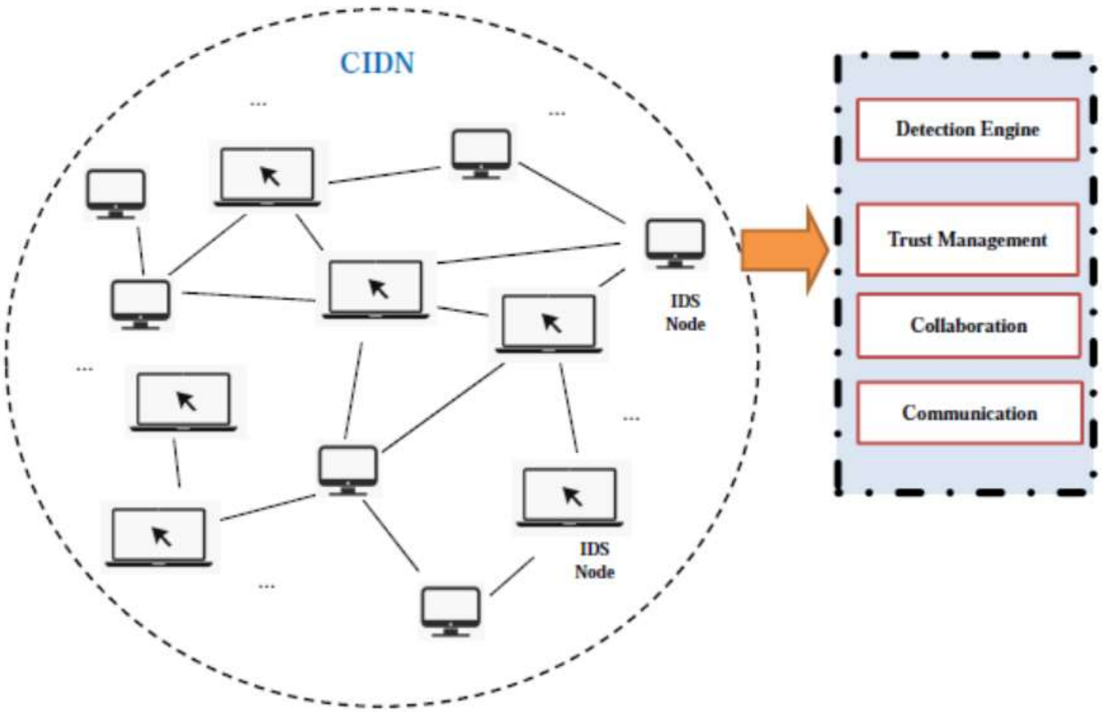


Centralized architecture employs a central server to collect and analyze the data collected from various monitor nodes.

Decentralized architecture adopts a hierarchical structure to organize the collaboration among monitor nodes and central server.

Distributed architecture enables a node having both monitoring and analyzing capability in the network without a central server.

CIDN framework with major components



Detection engine. Similar to a single IDS, detection engine here can employ both signature-based and anomaly-based detection to help examine incoming events.

Trust management. Help build the trust relationship among various nodes and identify malicious nodes.

Collaboration. This component is used to help coordinate the information exchange among different nodes.

Communication. This component is mainly responsible for building physical connections with different nodes.

W. Li and W. Meng. Collaborative intrusion detection in the era of IoT: Recent Advances and Challenges. Security and Privacy in the Internet of Things: Architectures, Techniques, and Applications (Book), Wiley-IEE, 2021.

Outline

- Background on Intrusion Detection
- Machine Learning in IDS
- Open Challenges
- Discussion on Blockchain

Discussion - 1

Algorithm Development

- When having a set of system or network features, a challenge is how to develop an appropriate algorithm to improve or optimize the detection performance.

Feature Selection

- How to select, decide and optimize an appropriate set of system or network features is an open problem

Impact of historical data

- With more historical data, the impact of such old data may decrease the detection sensitivity.

Discussion - 2

Overload traffic

- high-speed networks become common but may cause packet / event loss.
- Algorithm training / testing time

Trust Management - Incentive mechanism

- The purpose of this mechanism is to motivate and reward sensors for behaving in a benign manner or sharing the correct alarms. However, a non-suitable incentive mechanism may degrade the detection performance.

Detection privacy

- As CIDSs/CIDNs have to take over and monitor the whole network, data privacy issues have received much attention (e.g., GDPR).

Discussion - 3

CIDN deployment

- Deployed at small-scale
- Product?

Shekari, T., Bayens, C., Cohen, M., Graber, L., Beyah, R.: RFDIDS: radio frequency-based distributed intrusion detection system for the power grid. In: Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS) (2019)

Data Exchange

- Alarms used.
- Other messages / data?

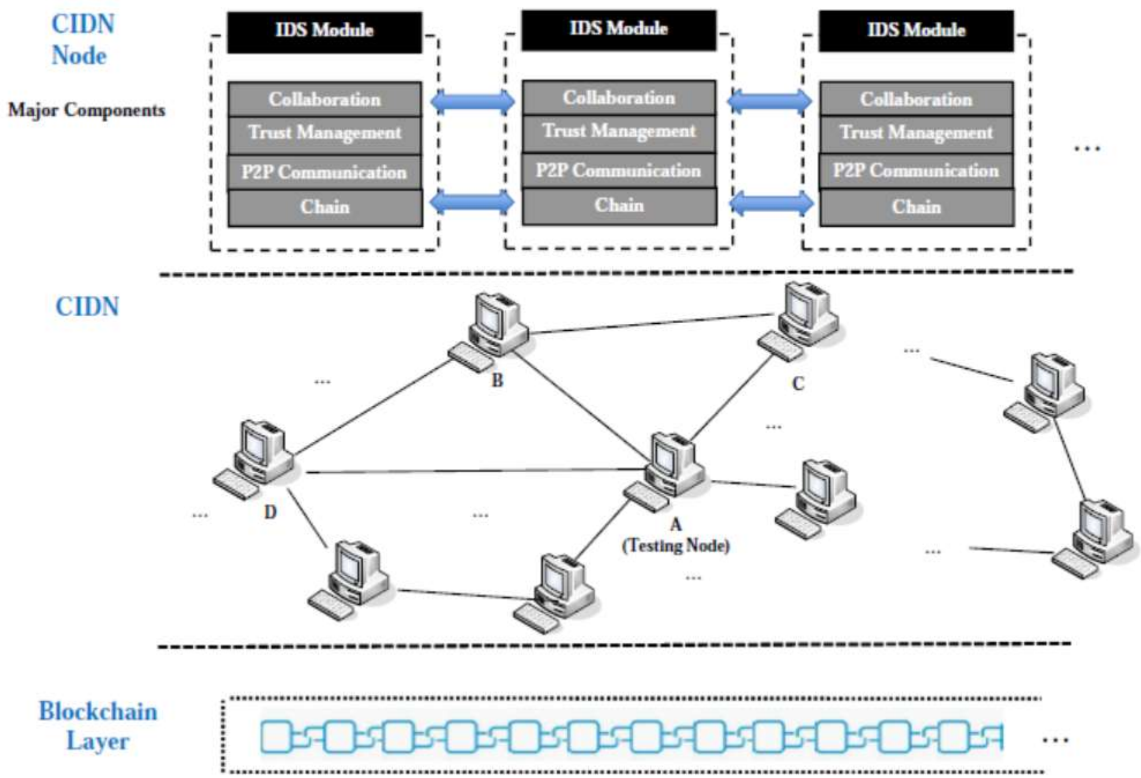
Existing security mechanisms

- Security information and event management (SIEM) system.
- Algorithm input / output

Outline

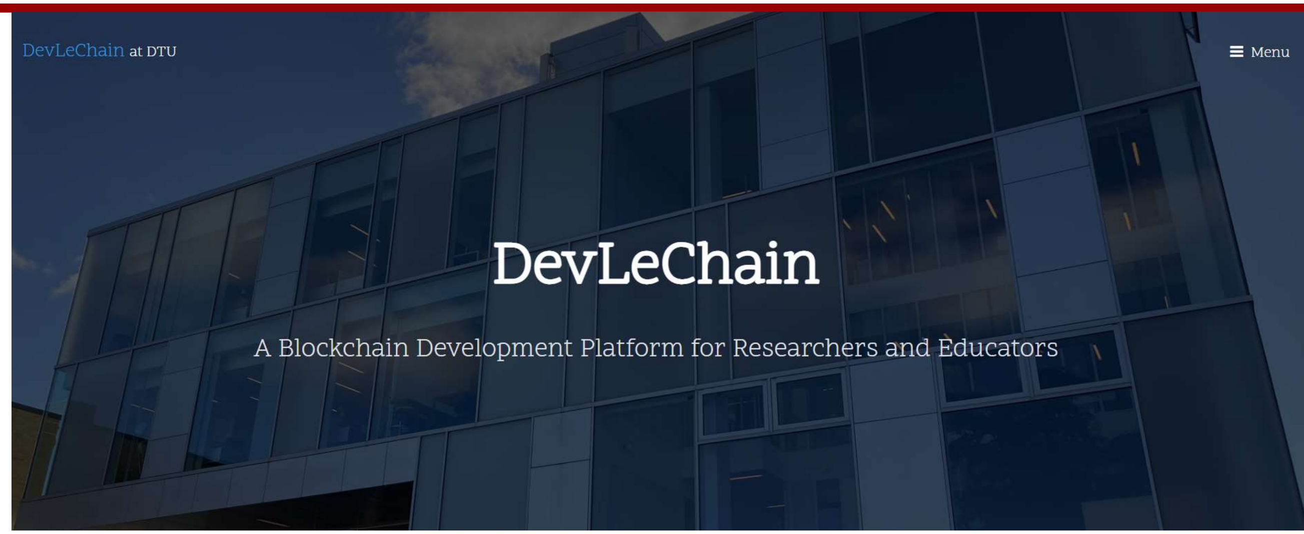
- Background on Intrusion Detection
- Machine Learning in IDS
- Open Challenges
- Discussion on Blockchain

Blockchain with ML-based IDS

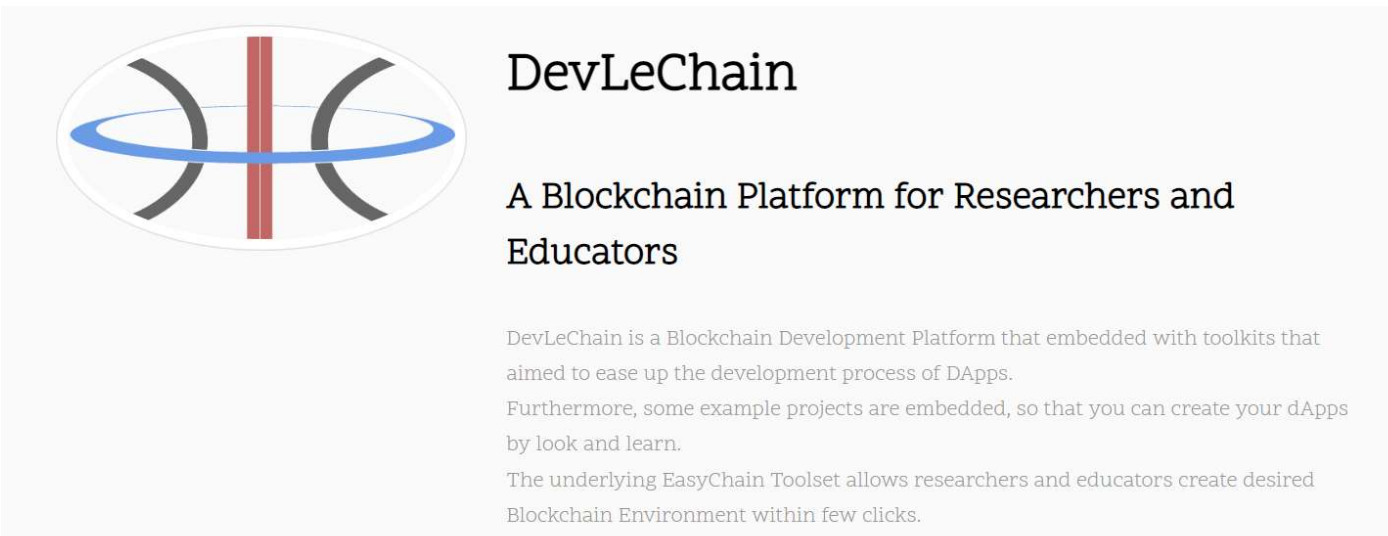


- This layer makes the framework different from traditional CIDN architectures, through allowing to establish a consortium blockchain.
- To verify ML model / events / messages via blockchain.

DevLeChain



https://sptagelab.github.io/DevLeChain/



DevLeChain

A Blockchain Platform for Researchers and Educators

DevLeChain is a Blockchain Development Platform that embedded with toolkits that aimed to ease up the development process of DApps.
Furthermore, some example projects are embedded, so that you can create your dApps by look and learn.
The underlying EasyChain Toolset allows researchers and educators create desired Blockchain Environment within few clicks.

Q&A

If you have any question, you can also contact via
weme@dtu.dk