A policy compliance architecture for secure data sharing

University of Amsterdam
Lu Zhang
16th June 2022

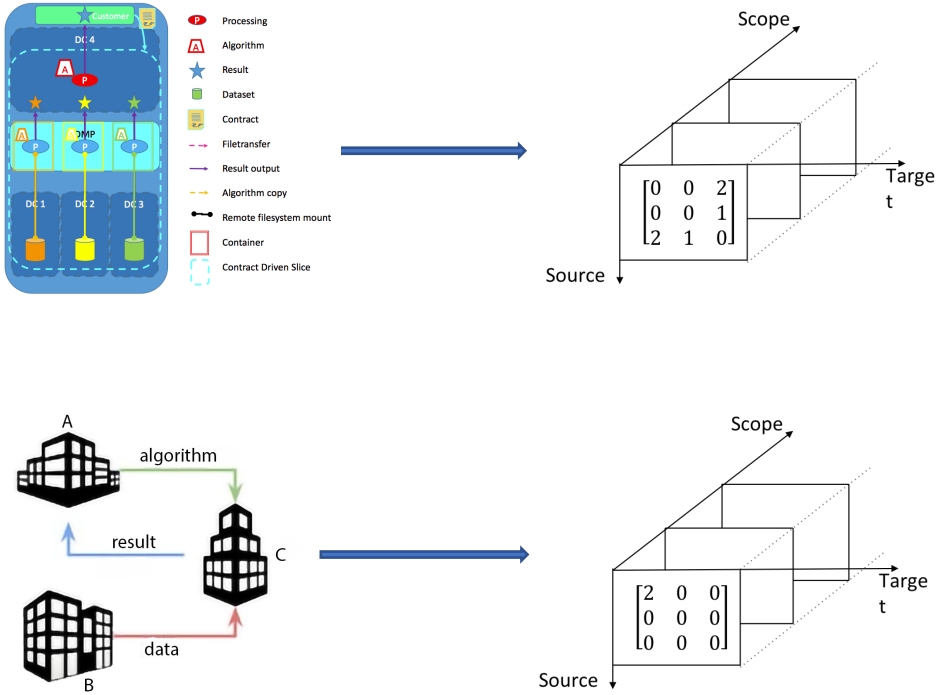
This project has received funding from the European Union's Horizon 2020 research and innovcation programme under grant agreement No 769288

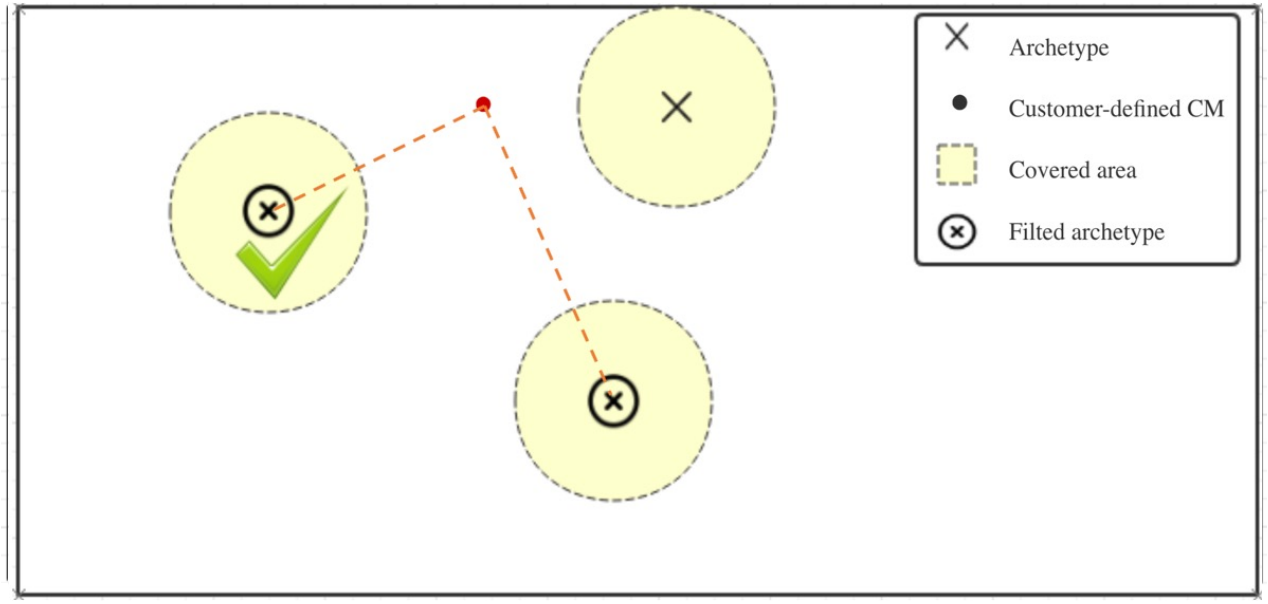
Main research question

How to select application-tailored infrastructure patterns and enhance policy compliance capabilities in a DDM infrastructure

- RQ1: How to map an application request to a best-fit digital infrastructure pattern based on collaboration models?
- RQ2: How to select an optimal digital infrastructure with minimum risk?
- RQ3: How to develop policy compliance detection components during execution?
- RQ4: How to defend against adversarial machine learning attacks for the monitoring components?

RQ1: Map application request to a best-fit infrastructure pattern

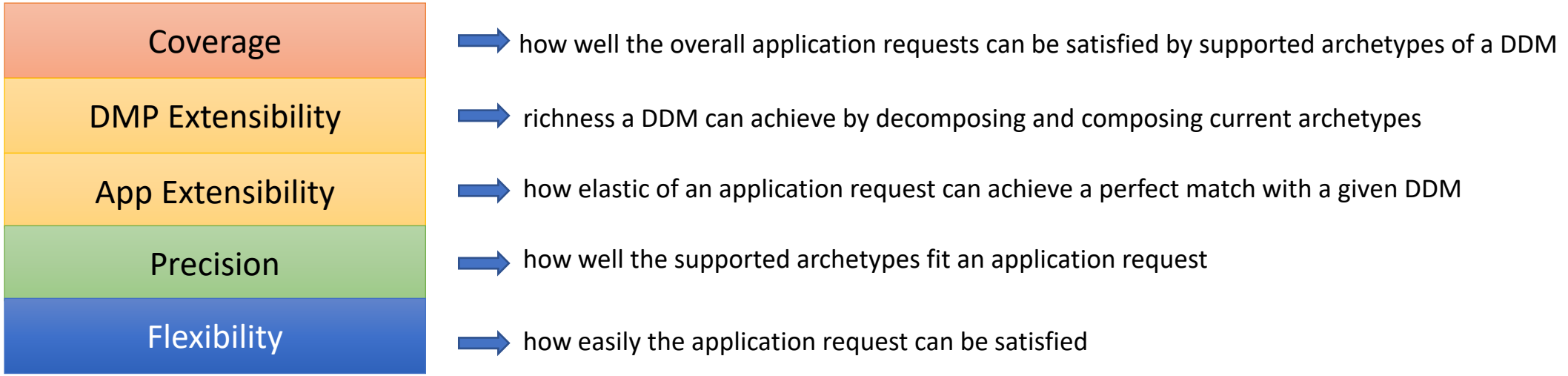




Lu Zhang "Management of collaborations in digital marketplaces" in proceedings of the 2019 International Conference on High Performance Computing and Simulation (HPCS 2019).
Lu Zhang, Reginald Cushing, Leon Gommans, Cees De Laat, and Paola Grosso, "Modeling of collaboration archetypes in digital marketplaces" in journal IEEE Access, DOI: 10.1109/ACCESS.2019.2931762

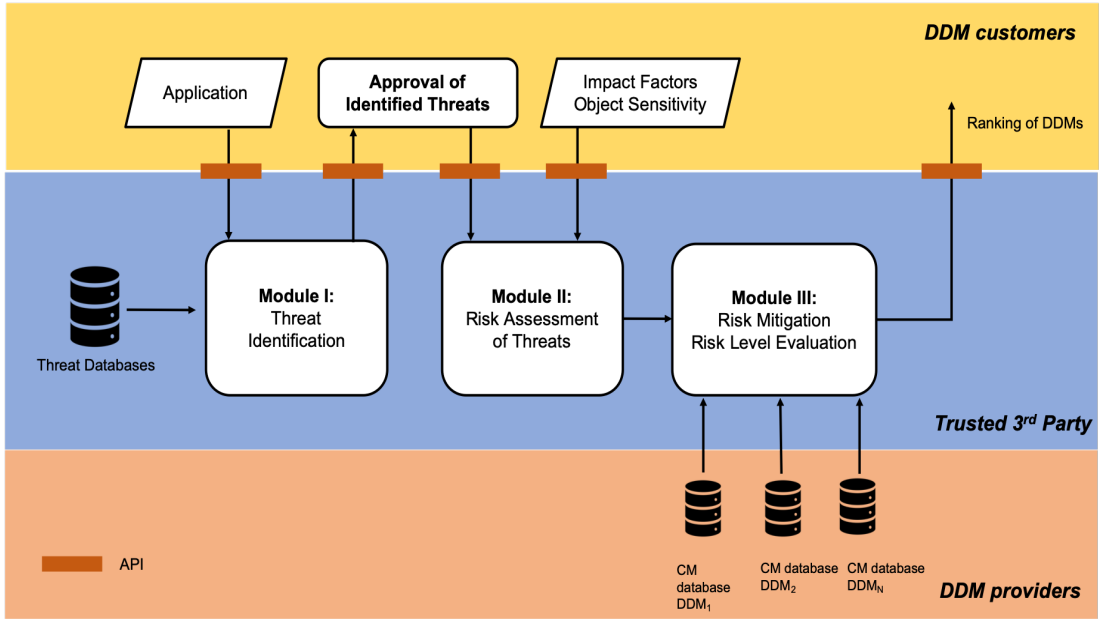
RQ1: Evaluation metrics of a DDM

- Provide a-priori information for DDM providers and potential customers
- Allow for comparison and intelligent selection of competing DDMs



RQ2: Select an optimal digital infrastructure with minimum risk

- Collaborative
- Application-based
- Robust
- Risk analysis-driven

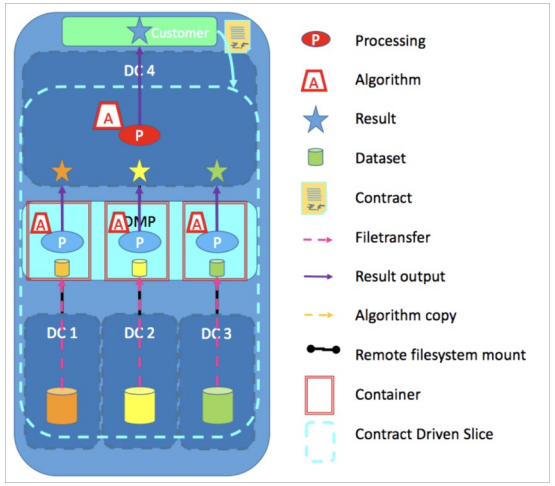


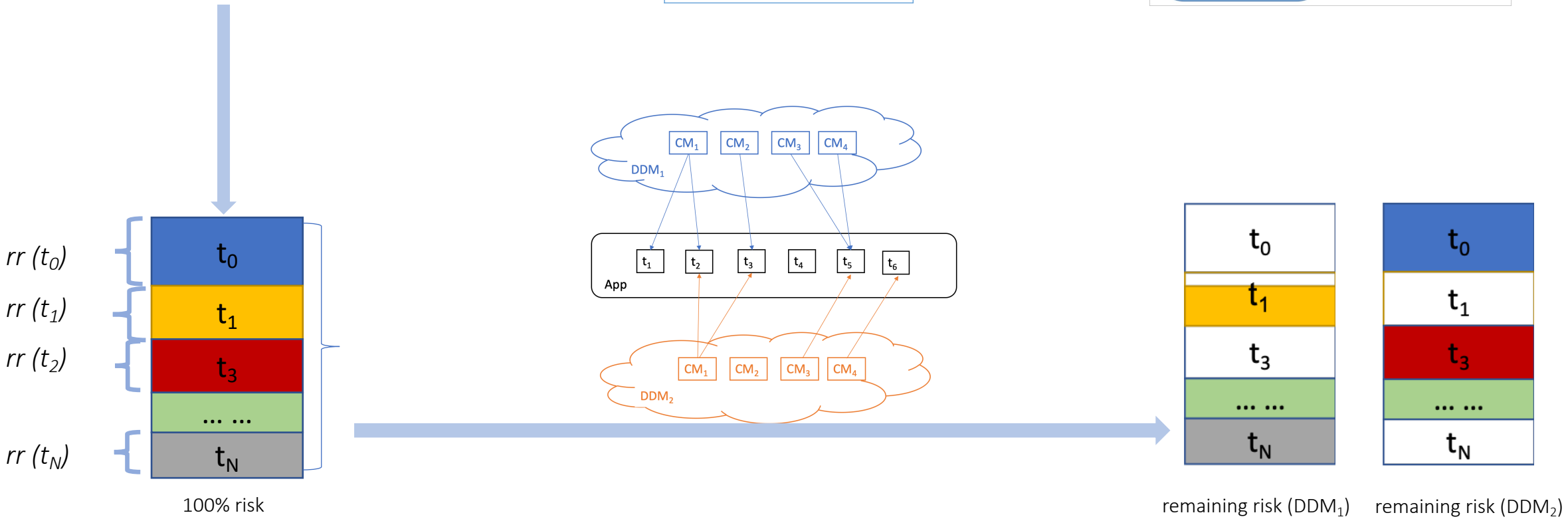
Lu Zhang, Arie Taal, Reginald Cushing, Cees de Laat, Paola Grosso, “A risk level assessment system based on the STRIDE/DREAD model for Digital Data Marketplaces” in journal International Journal of Information Security

Modified Microsoft STRIDE/DREAD model

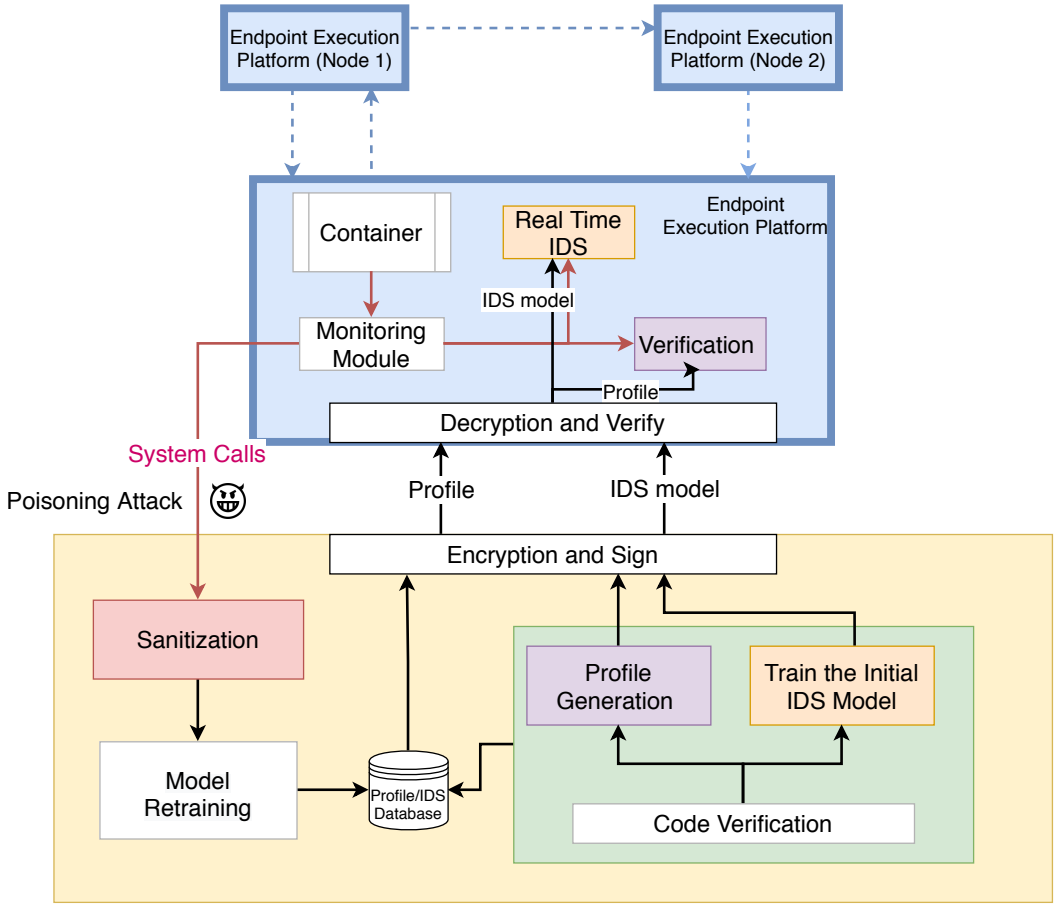
Risk Attributes	Low (0)	Medium (5)	High(10)
Damage Potential (DP)	Depending on sensitivity value of Data Object, Compute Object and Result Object (Low, Medium, High)		
Accessibility (AC)	Only by consortium party member	By involving party e.g. 3rd party	By outsiders
Skill Level (SL)	Advanced skills	Malware existing in Internet or using attack tools	Simple tools
Affected Users (AU)	One party member	Partial party members	All party members
Intrusion Detectability (ID)	Detectable without monitoring	Detectable by monitoring	Very hard to detect by monitoring

Threat List (Approved)
Not-trustable computing env
Eavesdropping
Malicious code: high result correlation
Man-in-the-middle
Container runtime escape
Data loss: Physical attack
Dos on other containers



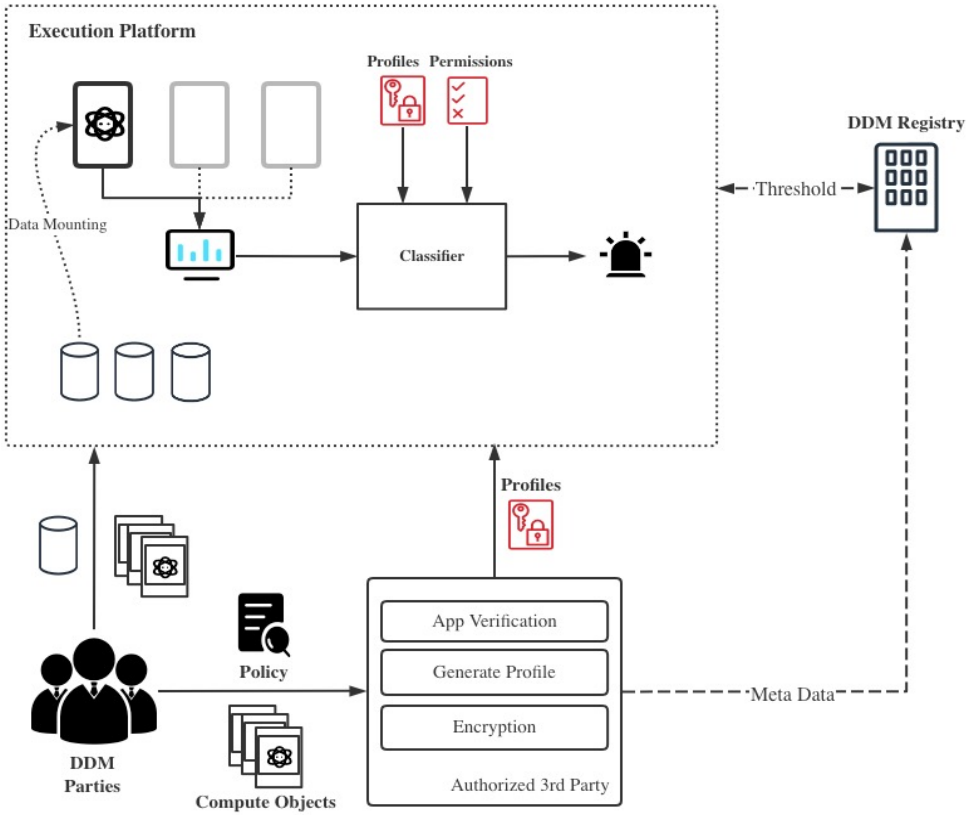


Policy compliance detection architecture



- Monitor the run-time behaviors of containerized applications with **system call tracing**
  - Profile Generation and Verification
  - Intrusion Detection System (IDS)
  - Sanitization

RQ3: Profile generation and verification



- Allow data owners or a DDM infrastructure provider to accurately **identify** which algorithms are running inside the container
  - Profiling of a container image with occurrence distribution of n-grams
  - Compute dissimilarity with Laplace smoothing and mutual cross-entropy

Lu Zhang, Reginald Cushing, Ralph Koning, Cees de Laat, Paola Grosso, “Profiling and discriminating of containerized ML applications in Digital Data Marketplaces (DDM)” in Proceedings of the 7th International Conference on Information Systems Security and Privacy (ICISSP 2021).

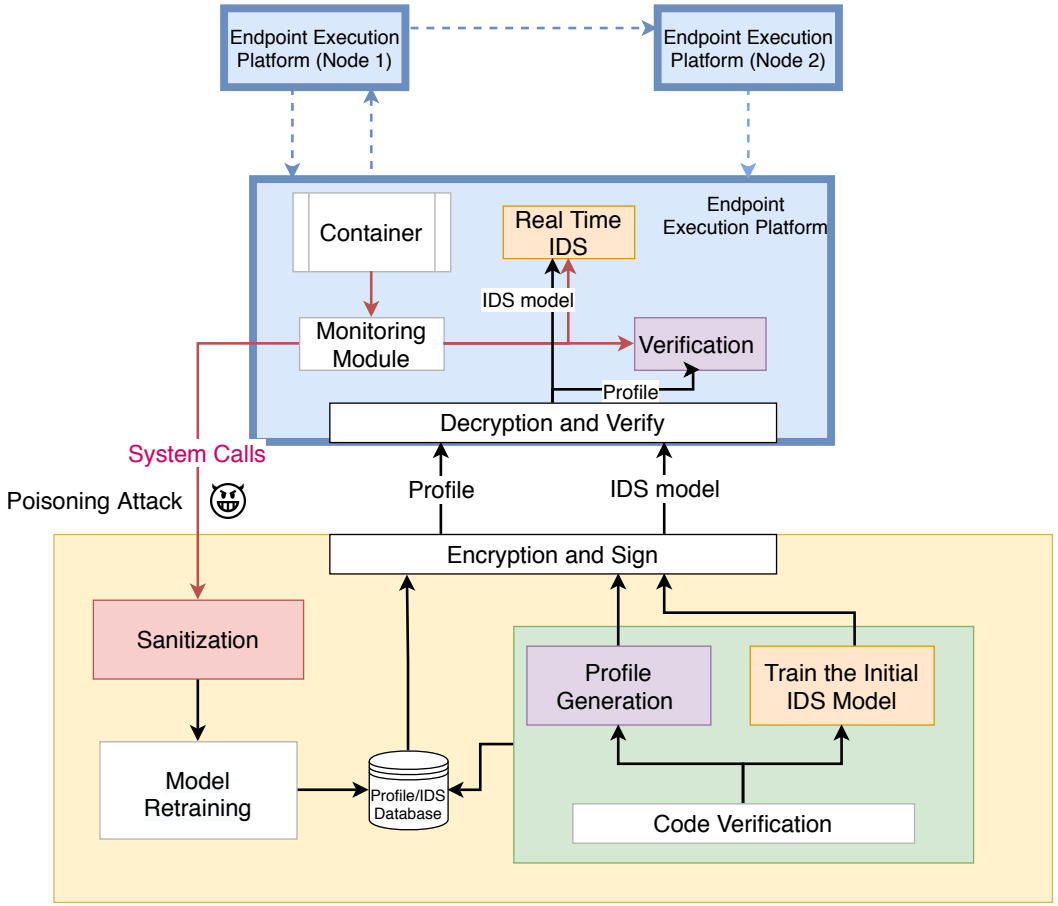
RQ3: Experimental Results

The confusion matrix of the classifier for 6 applications running with various platform OSs and training data sets

	APP 1	APP 2	APP 3	APP 5	APP 6	APP 7	mean (%) ± std
APP 1	**1529**	0	209	0	22	0	86.7 ± 0.15
APP 2	0	**1760**	0	0	0	0	100 ± 0
APP 3	0	0	**1623**	137	0	0	92.2 ± 0.15
APP 5	0	0	61	**1483**	216	0	84.2 ± 0.21
APP 6	0	0	0	0	**1760**	0	100 ± 0
APP 7	0	0	0	0	0	**1760**	100 ± 0
							93.85

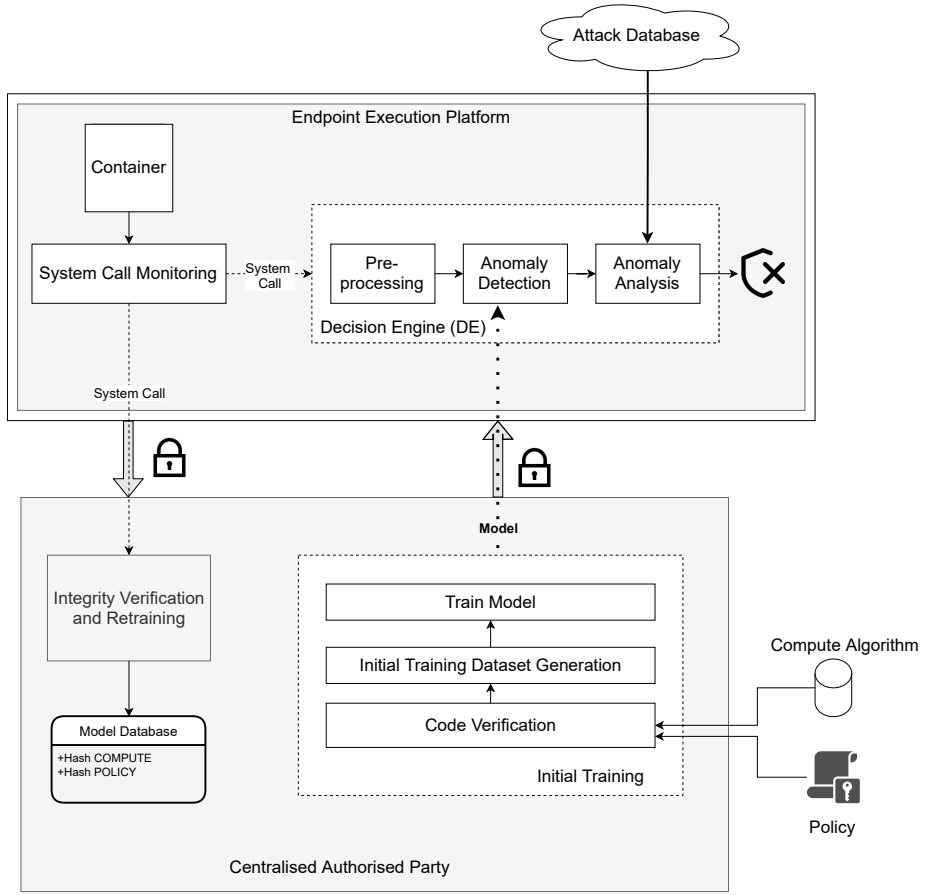
- The accuracy varies with applications
- Overall accuracy for all applications is as high as 93.85%.

Policy compliance detection architecture



- Monitor the run-time behaviors of containerized applications with **system call tracing**
  - Profile Generation and Verification
  - Intrusion Detection System (IDS)
  - Sanitization

RQ3: Hybrid real time Intrusion Detection System



- One Class Support Vector Machine (OC-SVM) for anomaly detection
- Model retraining to adapt to dynamic characteristics of the application behavior

Lu Zhang, Reginald Cushing, Cees de Laat, Paola Grosso, “A real-time intrusion detection system based on OC-SVM for containerized applications” in proceedings of the 24th IEEE International Conference on Computational Science and Engineering (CSE 2021).

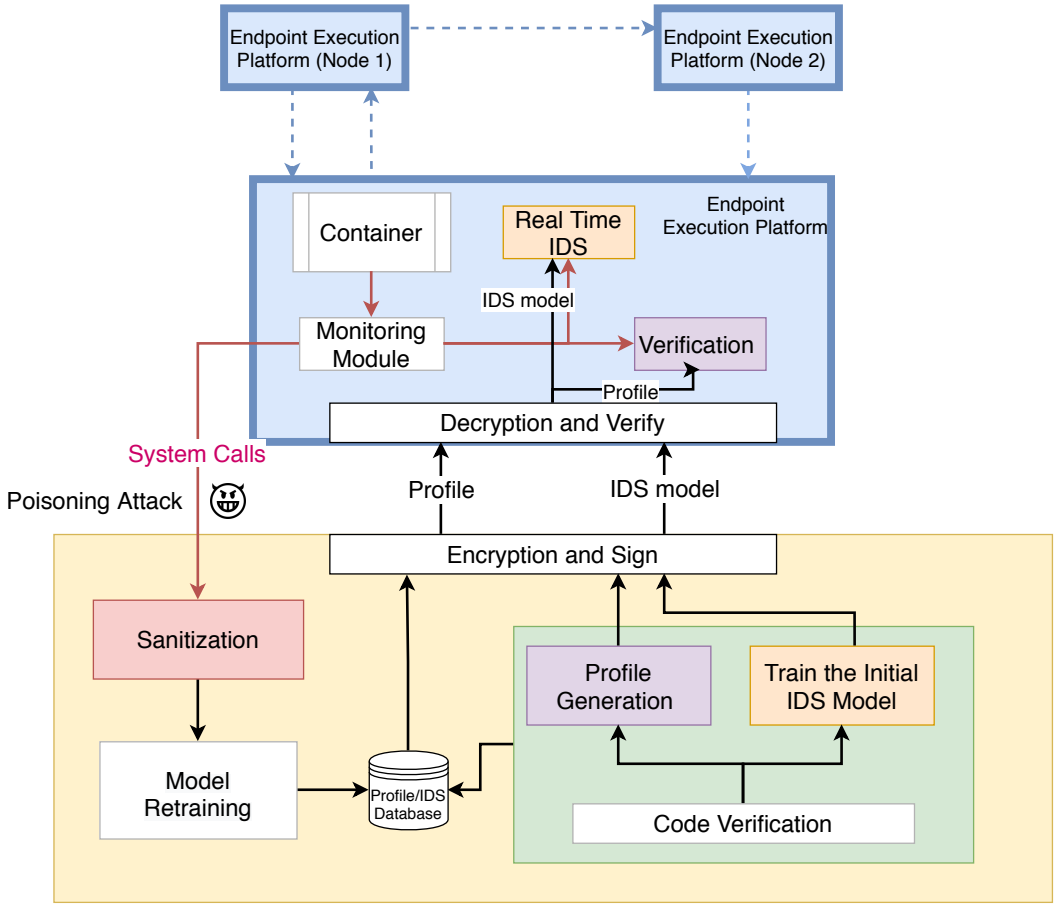
RQ3: Experimental Results

TABLE III: AUC, TPR, FPR values of different applications and attacks.

Application	Attack	AUC	TPR	FPR
CouchDB	Execute Arbitrary Code	0.995	1	0.067
Mongodb	Brute Force	0.959	1	0.020
Image Classification	PGD	0.917	1	0.12
	BIM	0.949	0.972	0.12
	CW	0.929	0.988	0.12
	FAB	0.951	0.961	0.12
	MIFGSM	0.851	1	0.12
	PGDDLR	0.857	1	0.12
	Square	0.858	1	0.12
	TPGD	0.799	0.55	0.12

- Optimal Configuration:
  - Feature extraction with *tf*
  - *Segmentation Length = 30000*
  - *Gaussian Kernel*
- The attacks arbitrary code execution and brute force performed on dynamic applications are easier to detect.
- More difficult to detect adversarial ML attacks who generates the adversarial samples in the runtime

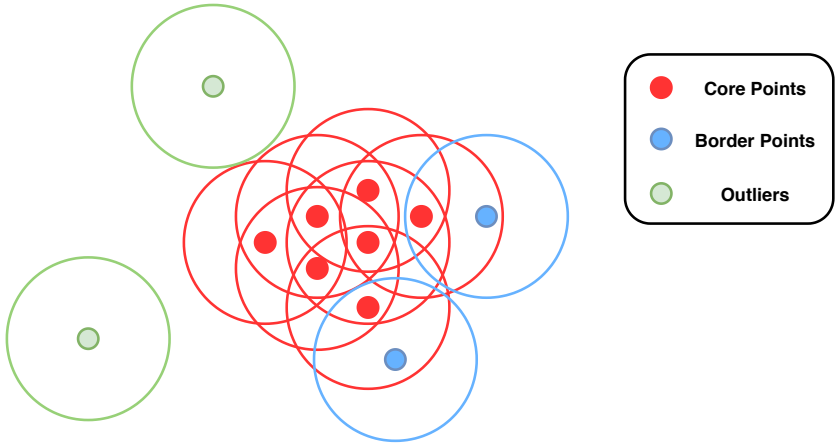
Policy compliance detection architecture

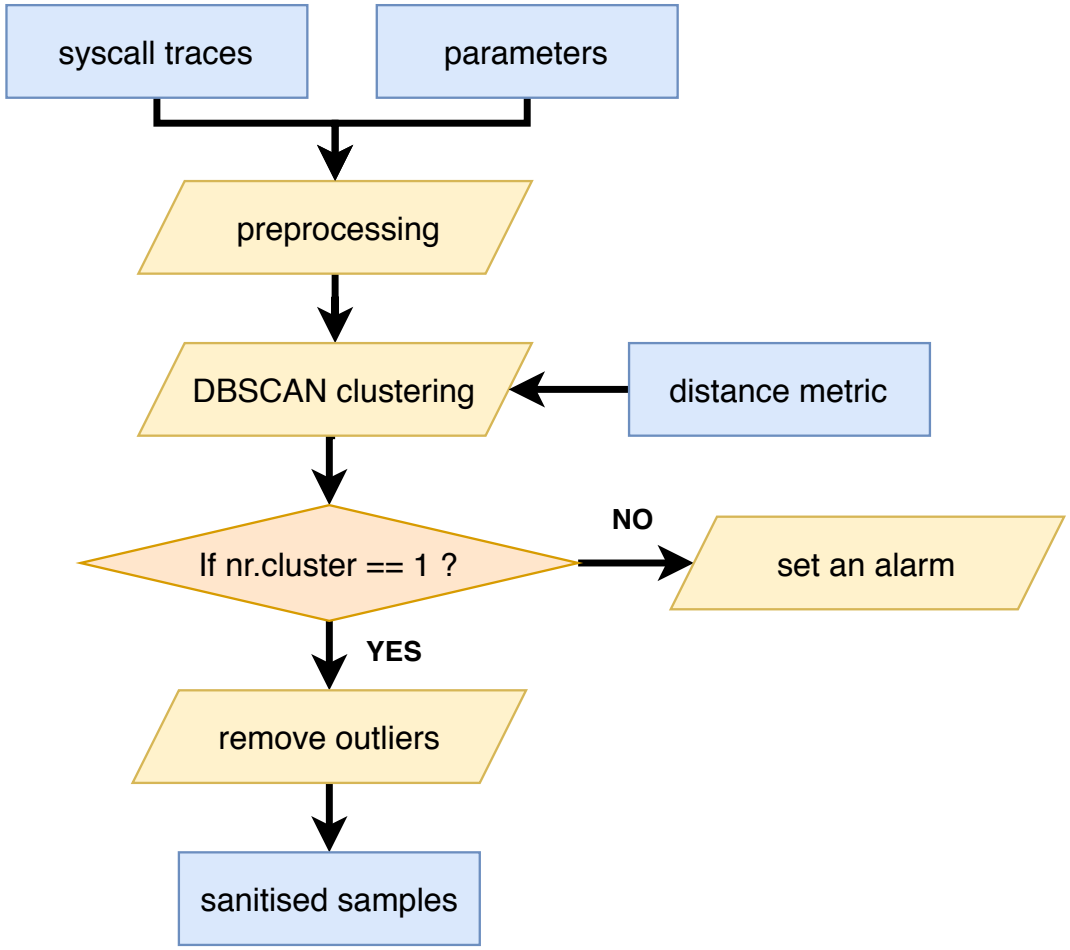


- Monitor the run-time behaviors of containerized applications with Linux **system call tracing**
  - Profile Generation and Verification
  - Intrusion Detection System (IDS)
  - Sanitization

RQ4: Sanitization with DBSCAN algorithm

- Clustering is done based on density, not related to shapes
- Deal with non-linear issues
- Do not require initial normal points
- Independent of machine learning algorithm and attack types



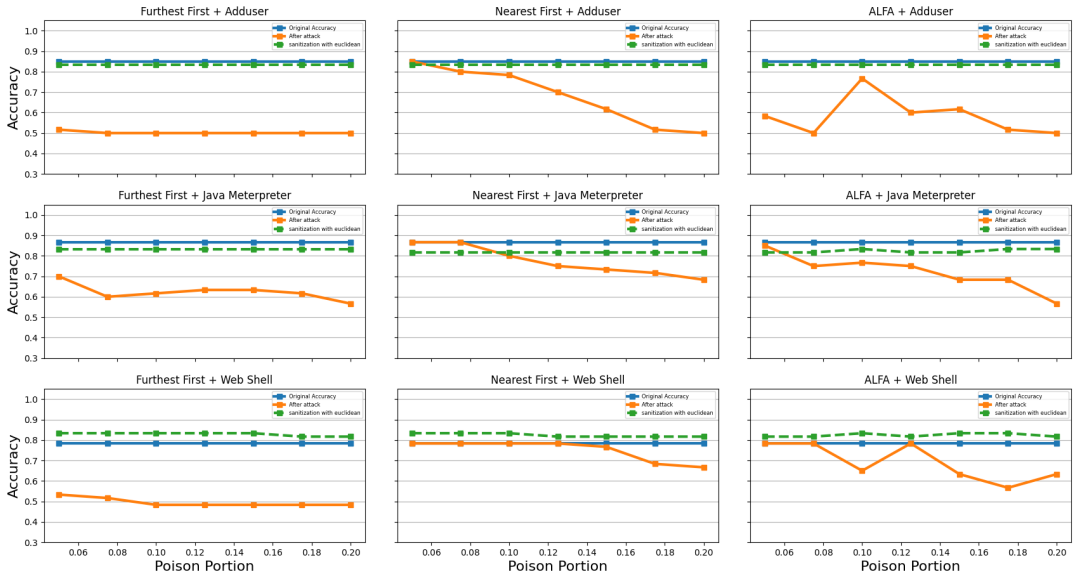


Lu Zhang, Reginald Cushing, Paola Grosso, “Defending OC-SVM based IDS from poisoning attacks” in Proceedings of the 2022 5th IEEE Conference on Dependable and Secure Computing (IEEE DSC 2022)

RQ4: Experimental Results

- We apply 3 classic **label flipping attacks** to a public dataset and measure the accuracy
  - Before attack
  - After attack
  - After implementing defending mechanisms
- Poisoning attacks can degrade the performance of the OC-SVM classifier to a large degree, defending mechanisms are necessary
- Accuracy after the sanitization process is pretty close to the original accuracy

Public Dataset



Conclusions

- An approach to model and measure mutual similarities of multi-lateral collaboration relationships
- A framework to quantitatively assess and compare risk exposure of data exchange infrastructures
- A hybrid intrusion detection system
- A methodology to profile and discriminate running behaviors of a containerized algorithm
- A defence mechanism for poisoning attacks targeted machine learning based IDS

THANK YOU AND ANY QUESTIONS?

www.dl4ld.nl
www.dl4ld.net