Predicting network intrusions using federated and interpretable anomaly learning

Maya MOUHAMAD

21/06/2024

Supervised by : Didier VERNA
Pierre PARREND
Nida MEDDOURI
Ali ALAELDINE

Abstract

- **Network intrusion detection system (NIDS)** using **Federated Learning** is proposed to enhance detection capabilities by continuously monitoring and analyzing network traffic by leveraging federated learning algorithms to improve the accuracy and efficiency of intrusion detection.

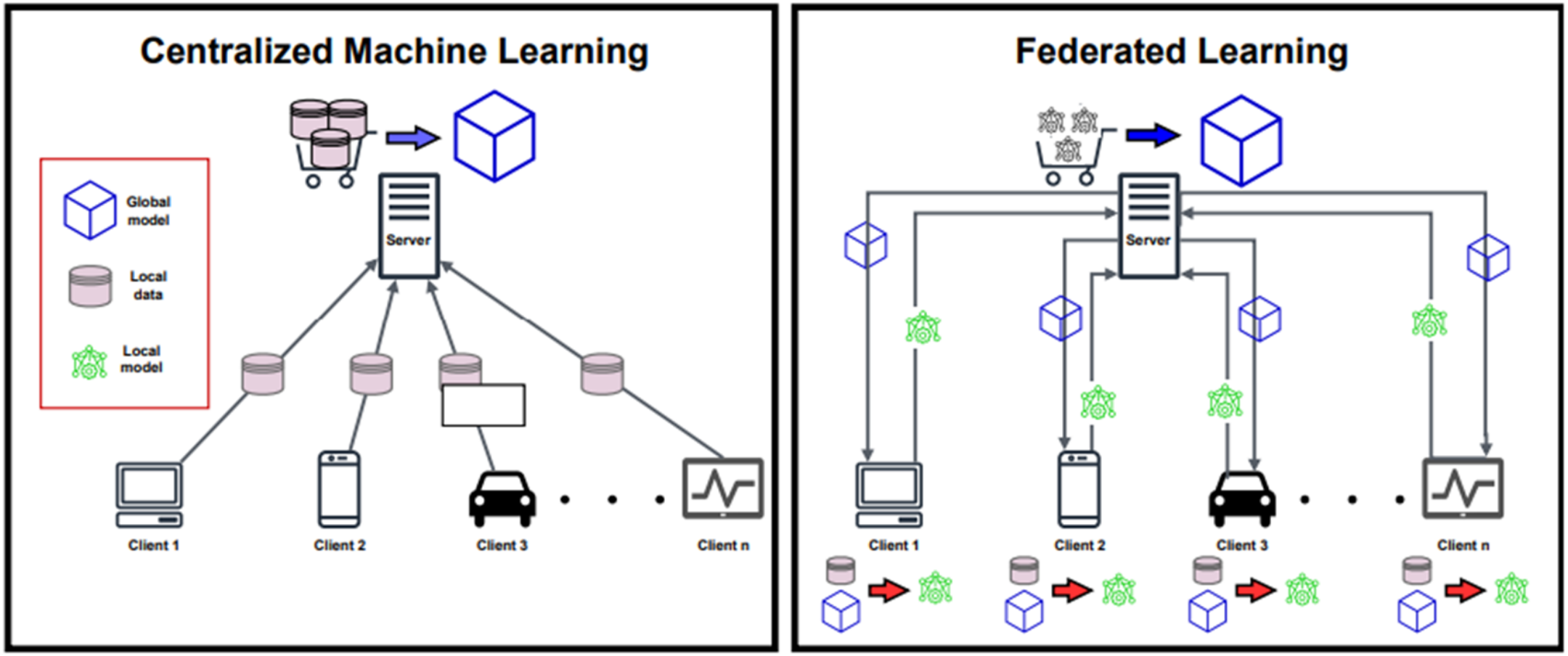
- This robust NIDS can detect and classify various types of network intrusions and can detect and alert network intrusions in real time, enabling prompt response and mitigation.

- The process is automated, reducing the need for manual monitoring and analysis.

Context

- Attacks seek to hide under the **layers of virtualization**.
- Interactions between third parties who have not established trust relationships
- Propose a multi-layer attack detection model
- Non-IID (non-independent and non-identically distributed) data.

Centralized Machine Learning vs. Federated Learning approach



Goal

A Federated Learning Approach for Detecting Cyber Attacks Across Multiple Cloud Systems

- Distributed detection
- Shared detectors and alerts

- **Challenges:**
  - Effectiveness
  - Efficiency
  - Privacy protection
  - Autonomy
  - Interpretability/Explainability of alerts.

- **Approach / Solution :**
  1. high-performing
  2. fast
  3. resource-efficient

Methodology

1. Detect anomalies related to attacks in Cloud infrastructures.
2. Enhance the interpretability of raised alerts.
3. Share alerts among different users of the Cloud infrastructure, without disclosing confidential information about the systems they manage.

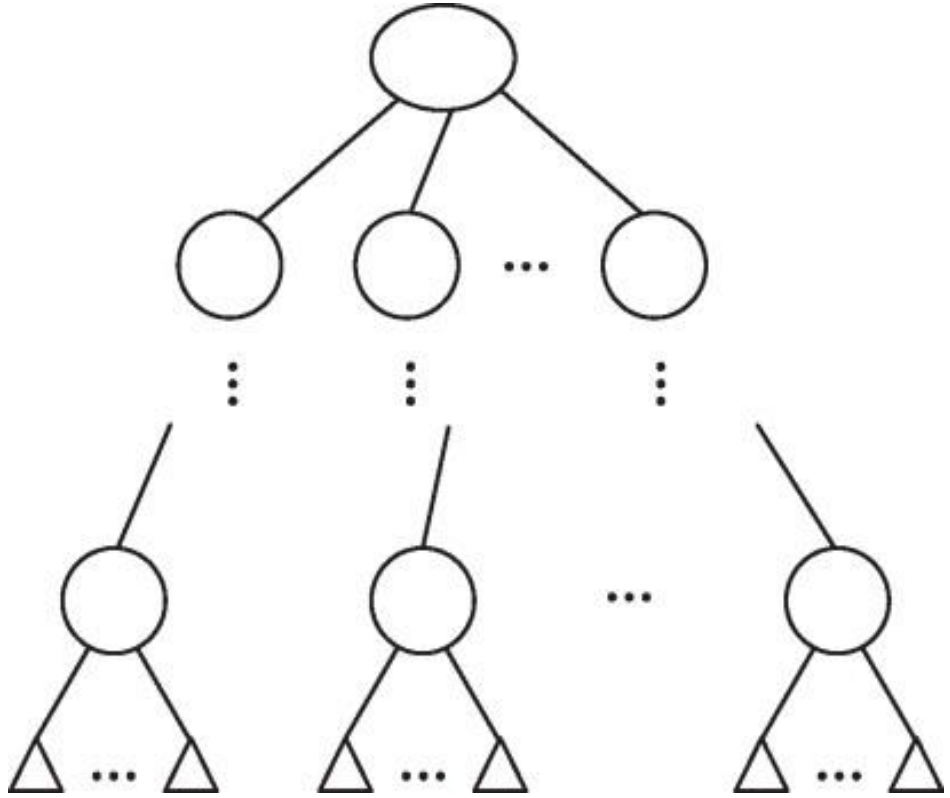
Methodology : prob/app1

Detection of Low-Level Attacks (at the level of physical cloud systems).

Methodology : prob/app 2

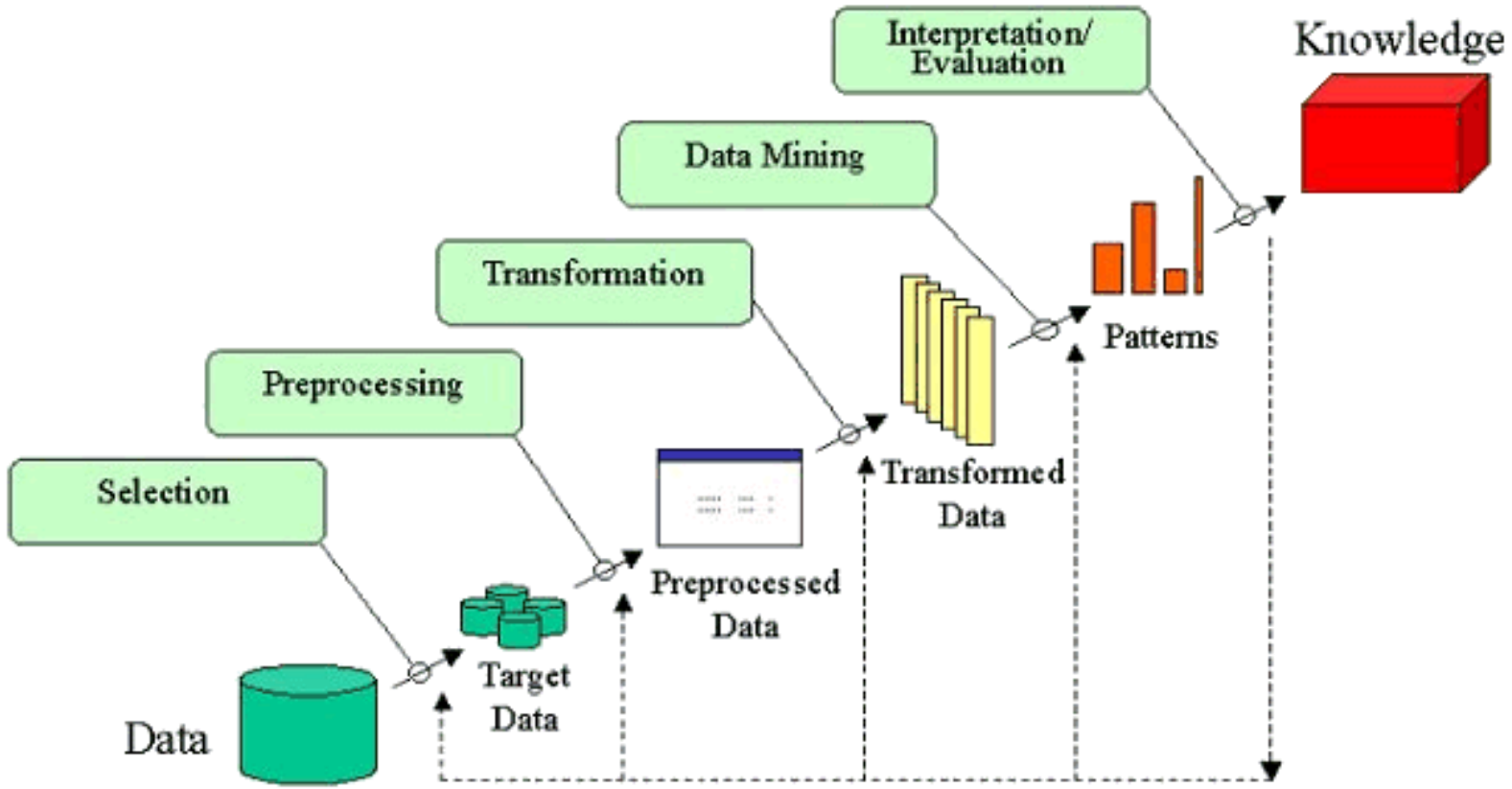
Efficient, effective, and privacy-respecting federated learning of multi-layer data.

- Decision tree approaches offer a competitive "detection capability" / "computation time cost" ratio.



Methodology : prob/app 3

Knowledge extraction for the sharing of alerts and detectors between parties without a trust relationship.



Expected Contributions

- Bibliography of federated learning
- Federated learning model by GBDT (Gradient Boosted Decision Trees)
- A model for extracting and sharing knowledge from GBDT trees

Actually

Nom de l'algorithme	Date
Smooth Private Forest	2021
CatBoost (Categorical Boosting)	2018
LightGBM (Light Gradient Boosting Machine)	2017
XGBoost (Extreme Gradient Boosting)	2016
Deep Forest	2016
Regularized Greedy Forest	2015
SysFor	2011
Bayesian Additive Regression Trees	2011
Spike and Slab Trees	2011
GUIDE	2011
SPAARC	2009
Oblique Random Forests	2009
Optimized Forest	2008
BFTree	2008
Rotation Forest	2006
Extra Trees (Extremely Randomized Trees)	2006
Conditional Inference Trees	2006
LMT	2005
JC Haid star	2005
LAD Tree	2004
JC Haid	2003
Forest PA	2001
Random Forest	2001
Random Subspace Method	2001
CDT	2000
Hoeffding Tree	2000
CRUISE	2000
ADTree	1999
PCT	1998
QUEST	1997
Bayesian Decision Trees	1997
NB Tree	1996
DBS Tree	1996
RIPPER	1995
J48 Consolidated	1993
J48	1993
MARS (Multivariate Adaptive Regression Splines)	1991
REP Tree	1987
SimpleCart	1986
ID3 (Iterative Dichotomiser 3)	1986
CART (Classification and Regression Trees)	1984
CHAID (Chi-squared Automatic Interaction Detector)	1980

Performances sur qlq Datasets

Dataset	Algorithme	Taux de correctly classified	Temps d'apprentissage (ms)	Temps de test (ms)
RT_IOT2022	CDT -SIDM	99.72(0.05)	12549.06(3437.20)	20.16(8.40)
	Decision stump	83.46(0.01)	7233.91(1611.22)	**18.91(7.79)**
	ForestPA	**99.81(0.04)**	611998.59(45312.69)	30.94(7.69)
	Hoeffding tree	95.70(0.35)	16778.13(732.42)	513.59(92.82)
	FT tree	99.68(0.05)	153432.50(11254.68)	3697.97(568.90)
	J48	99.79(0.04)	30713.59(1716.87)	30713.59(1716.87)
	J48 consolidated	99.73(0.05)	647191.41(25397.84)	24.37(8.97)
	JCHAID Star	99.66(0.06)	36834.53(1920.81)	25.00(8.31)
	RandomTree	99.74(0.06)	**2604.53(517.51)**	513.59(92.82)
	CART	99.74(0.06)	175162.97(4542.33)	19.84(8.56)
	SPAARC	99.79(0.05)	155066.72(6458.98)	22.03(14.06)
naticusdroid	AD TREE	3.59(6.61)	7650.63(346.51)	3.59(6.61)
	CDT -SIDM	96.24(0.34)	1860.16(114.33)	2.81(6.03)
	CSForest	2.81(6.03)	5.73(0.26)	**0.01(0.01)**
	Decision stump	85.68(0.68)	388.13(91.51)	2.81(6.03)
	FT tree	**96.64(0.35)**	9.88(0.58)	0.52(0.08)
	Hoeffding tree	94.98(0.56)	**0.34(0.03)**	**0.01(0.01)**
	J48	96.60(0.33)	3932.19(375.30	3932.19(375.30)
	ID3	96.52(0.33)	5407.81(344.19)	7.19(8.43)
	J48 Consolidated	96.60(0.33)	96.60(0.33)	5.00(7.33)
	JCHAID	95.86(0.36)	4611.09(313.93)	6.88(8.97)
	JCHAID Star	95.99(0.32)	6960.00(350.99)	4.06(6.89)

Thank you !