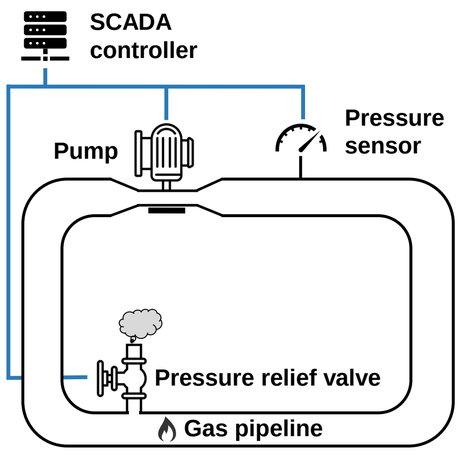
Ensemble Learning for Industrial Intrusion Detection

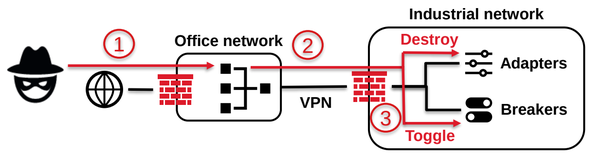
Motivation

Industrial Control Systems (ICSs) Need Security

- ICSs run **critical infrastructure**
  - Water treatment, power grid, production, …
- Today largely digitalized and increasingly Internet-connected
  - Enables automation, remote monitoring, …



- Cyberattacks can be devastating
  - Blackouts, physical damage, …
- Example: Ukrainian Power Grid attack
  - 250 000 people without electricity for hours

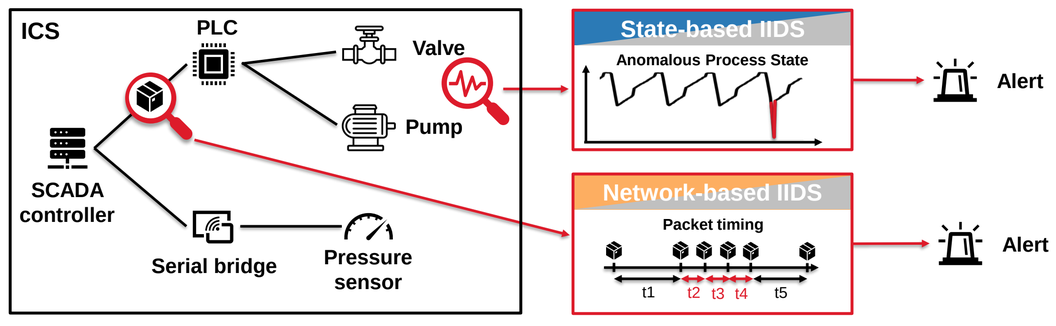


1. Infiltrate office network through spear phishing
2. Jump to industrial network through internal VPN
3. Toggle breakers in substations and destroy adapters to hamper recovery efforts

- ICSs demand reliable **security solutions**
  - Sophisticated attacks from resourceful adversaries
  - Missed attacks may result in devastating incidents
  - False positives cause downtime and significant costs

Industrial Intrusion Detection Systems (IIDSs)

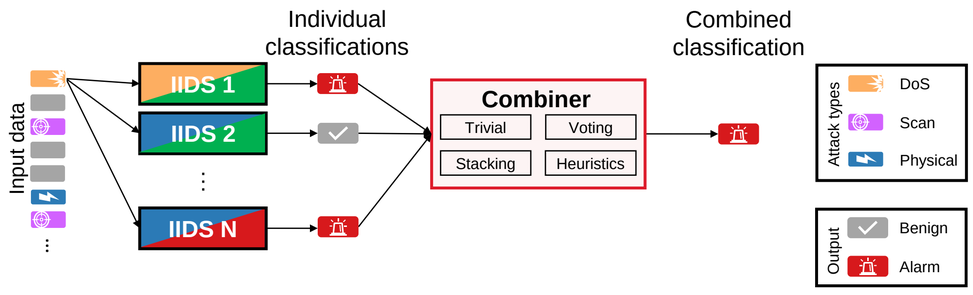
- IIDSs monitor an ICS' behavior to detect anomalies and cyberattacks
  - Raise an alert and inform operators in case of suspicious behavior
- Research proposes various types of IIDSs
  - Monitored ICS **properties**: Network Packets or Physical State
  - Detection **technology:** Supervised ML or Unsupervised ML



- Research mainly focuses on **monolithic** detectors
  - Each having different capabilities (e.g., high precision or ability to detect unknown attacks)
  - Approaches claim to be the single-best solution

Idea

- Idea: **Combine** multiple monolithic IIDSs
  - Build upon existing diverse approaches
  - Leverage their unique capabilities



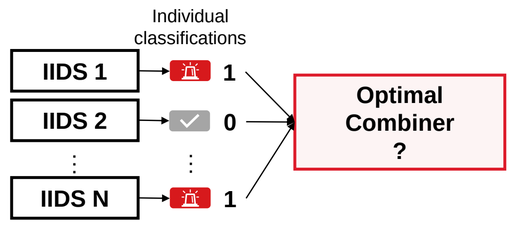
- Combining IIDSs can achieve different **goals**

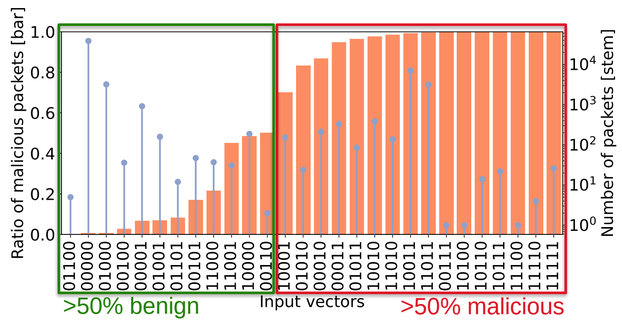
IIDS 1	+	IIDS 2	→ Improved accuracy/F1 score
IIDS 1	+	IIDS 2	→ Wider range of detectable attacks
IIDS 1	+	IIDS 2	→ Good accuracy with potential to detect novel attacks
IIDS 1	+	IIDS 2	→ Detect attacks on both network and physical level

- Unlocks additional **flexibility**
  - No IIDS lock-in: option to simply add new IIDS if its capabilities are useful
  - System can be adapted to different ICSs by choosing appropriate base IIDSs
- The **Combiner** computes a unified output
  - Different methods of combining are possible
  - Can be trainable (i.e., learning-based) or manually parameterized (e.g., weighted voting)

Upper Bound

- Calculate an upper bound for the optimal combiner
  - Theoretical combiner to maximize **accuracy**



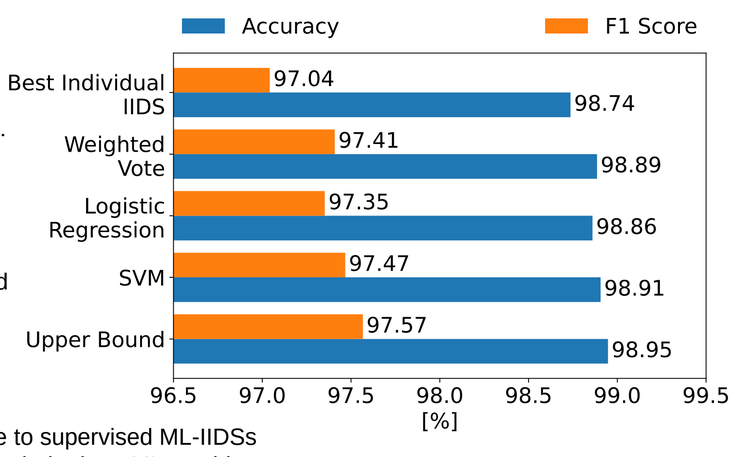


- Count occurrences for each input
  - Classify according to the majority to optimize accuracy

First Results

IIDS 1 + IIDS 2

- Combine 7 open-source supervised ML classifiers
  - fkie-cad/ipal_ids_framework
  - Random Forest, SVM, BLSTM, …
- Initial findings
  - Weighted votes outperform the best individual IIDS
  - SVM nearly matches upper bound
  - Upper bound indicates marginal headroom

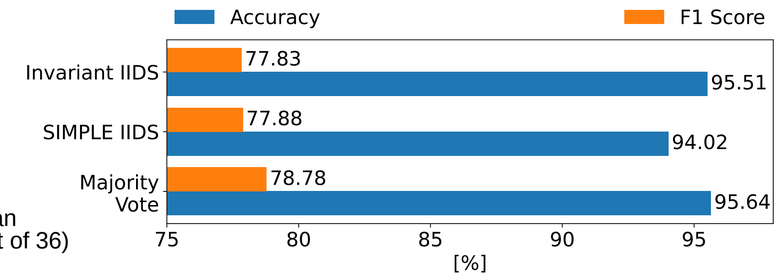


- **Lessons learned**
  - Ensemble methods are applicable to supervised ML-IIDSs
  - Simple weighted votes almost match the best ML combiner
  - Upper bound shows: more diverse set of classifiers needed

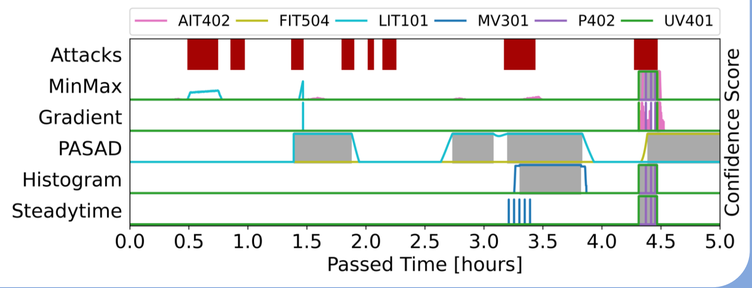
Ongoing Research

IIDS 1 + IIDS 2

- Challenge: Combine **unsupervised** IIDSs
  - Individual approaches differ a lot structurally compared to supervised IIDSs
  - Training of the combiner must use benign data only
  - Approaches like weighted voting are still applicable



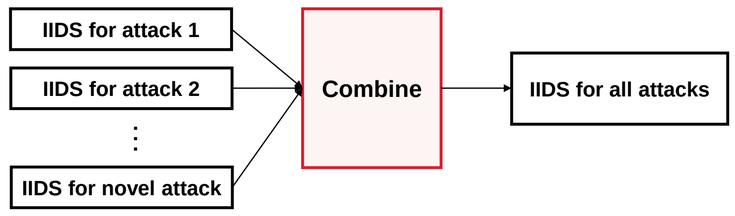
- Initial findings
  - A simple majority vote already improves the F1 score by ~1%
  - Only a combination of IIDSs can detect nearly all attacks (33 out of 36)
- **Time series-aware** IIDSs offer unique opportunities
  - A single alert during the attack suffices to notify ICS operators
  - Temporal effects can be leveraged by the combiner



Outlook

IIDS 1 + IIDS 2 + …

- Combine specialized IIDSs: **Mixture of Experts**
  - Get the best out of all worlds
  - E.g., anomaly detection for novel and signature-based IIDSs for known attacks



- Utilize **confidence** scores
  - Break up binary classification
  - Provides detailed data for the combiner

- Goal: combine IIDSs of arbitrary type
  - Additional challenges: required data, training methodology

Dominik Kus, Konrad Wolsing, Jan Pennekamp, Eric Wagner, Martin Henze, Klaus Wehrle

https://www.comsys.rwth-aachen.de

COMSYS | RWTH AACHEN UNIVERSITY