A Review of Research Work on Network-Based SCADA Intrusion Detection Systems

Slavica V. Boštjančič Rakas[1]
Mirjana D. Stojanović[2]
Jasna D. Marković-Petrović[3]

[1]Mihailo Pupin Institute, Serbia
[2]Faculty of Transport and Traffic Engineering, Serbia
[3] CE Đerdap Hydroelectric Power Plants Ltd., HPP Đerdap 2, Serbia

Introduction (1)

- SCADA systems:
  - control and monitor geographically dispersed process equipment on multiple sites, often spread over large distances
  - centralized data acquisition and control are essential to system operation
  - most widespread types of industrial control systems
- Failures and malfunctions:
  - serious consequences due to their strategic importance for national critical infrastructures
- Fourth-generation SCADA systems:
  - adopt IIoT and the Future Internet (FIN) technologies (cloud/fog computing, big data analytics, mobile computing, etc.)

Introduction (2)

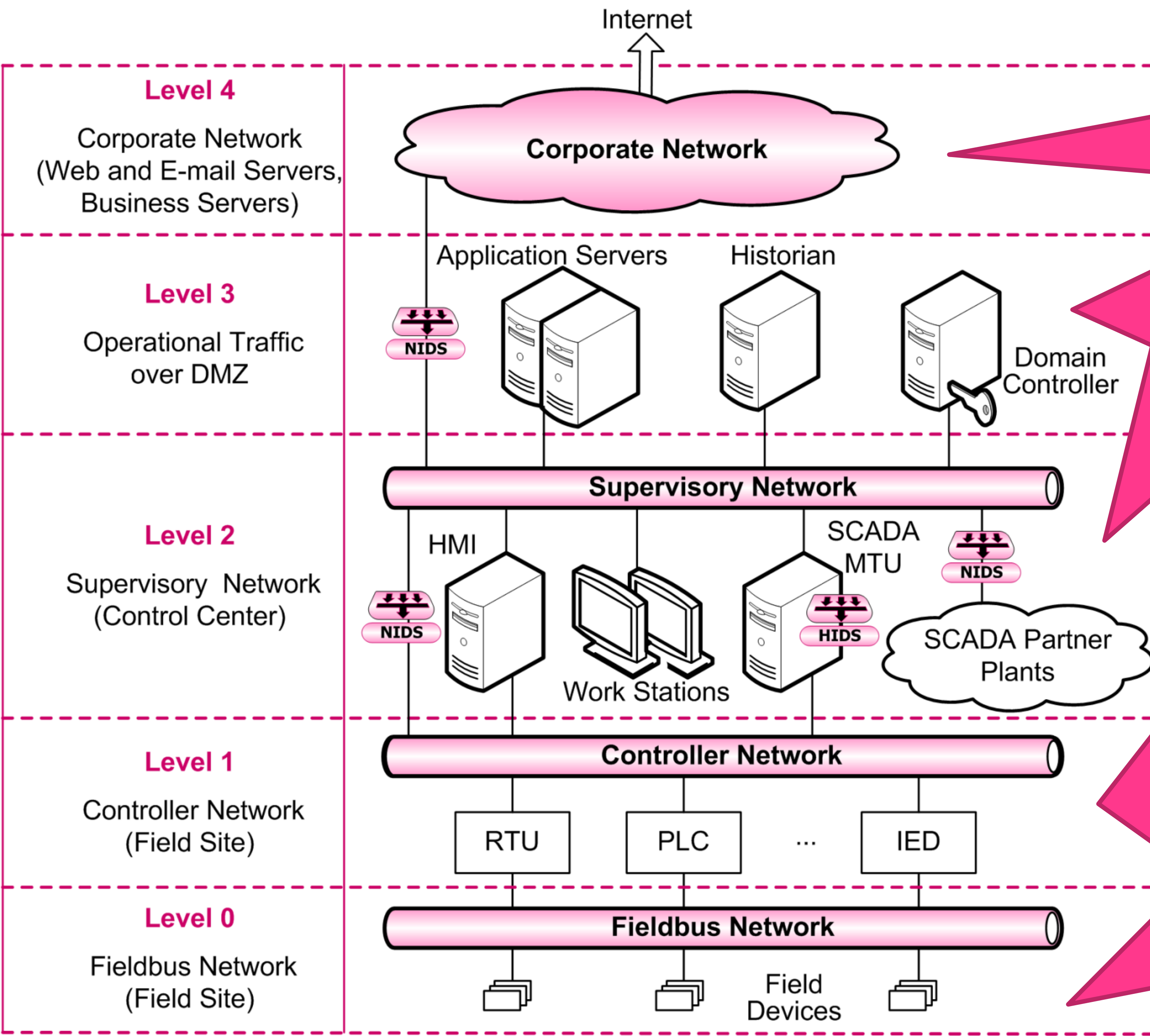
- Relevant standards and recommendations:
  - general IT security standards, common standards and directions for protecting SCADA and industrial control systems, and specific directions concerning particular industrial sectors
- Intrusion detection:
  - process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents
- Intrusion detection technologies:
  - NIDS (network-based IDS) and HIDS (host-based IDS).
- Basic methodologies for incidents detection:
  - signature-based detection,
  - anomaly-based detection
  - specification-based detection

Introduction (3)

- The main objectives of this work:
  - To propose a systematic and comprehensive evaluation methodology for SCADA-specific IDSs;
  - To perform a critical evaluation of recent IDS solutions
  - To assess their strengths, weaknesses, implementation maturity, as well as suitability to FIN environment;
  - To identify gaps in current research and to propose relevant research priorities for future work in the area

FACTORS THAT AFFECT THE DESIGN OF SCADA-SPECIFIC IDS

1. Hierarchical SCADA architecture



Corresponds to the corporate IT network, which is connected to the Internet.

- Control center collects and analyzes information from field sites, presents them on the HMI consoles, and

- Controllers process signals from field devices and generate appropriate commands for these devices.
- Processing results are

Represents physical devices that interact directly with industrial hardware, interconnected via fieldbus.

2. Traffic properties in SCADA networks

- SCADA networks are characterized by regular traffic patterns and a limited set of telecommunication protocols

QoS requirements

…eir configuration

- Updates should be performed on a regular basis, because the data is only valid in its assigned time period
- The order of updating is important for sensor data concerning monitoring of the same process or correlated processes
- The order of data arrival to the control center – an important role in presentation of process dynamics and influences decision making, by either a control algorithm (software) or a human operator who monitors the industrial process

3. Cyber vulnerabilities and attacks

- Vulnerabilities:
  - Policy and procedure, architecture and design, configuration and maintenance, physical, software development, and communication/network
  - Factors that affect SCADA vulnerabilities: human errors, resource limitations of physical devices, unsecure legacy systems and proprietary protocols, equipment failures and other accidents caused by negligence, and natural disasters
- Attacks - launched by external sources or internal sources
- Different taxonomies of attacks:
  - 4 classes: reconnaissance, response and measurement injection, command injection and DoS
  - 4 categories: key-based attacks, data-based attacks, impersonation-based attacks and physical-based attacks
  - Attacks on hardware, attacks on software, and attacks on network connections

REVIEW METHODOLOGY

Selection of Papers

- The initial set of research papers from well known databases (IEEE Xplore, SCOPUS, Web of Science)
- Keywords - “SCADA” and “intrusion detection”
- The search considered the 5-year period from 2015 to 2019
- Obtained 310 papers in total: 71 papers (IEEE Xplore), 131 papers (SCOPUS) and 105 papers (WoS)
- Resulting set - 68 papers (after exclusion of the replicated papers)
- Further, we focused on papers containing original proposals for SCADA-specific NIDS solutions - remaining set 86 papers
- The final selection – 26 comparable papers (solutions) – we eliminated similar papers by the same authors or papers describing the results of the same projects

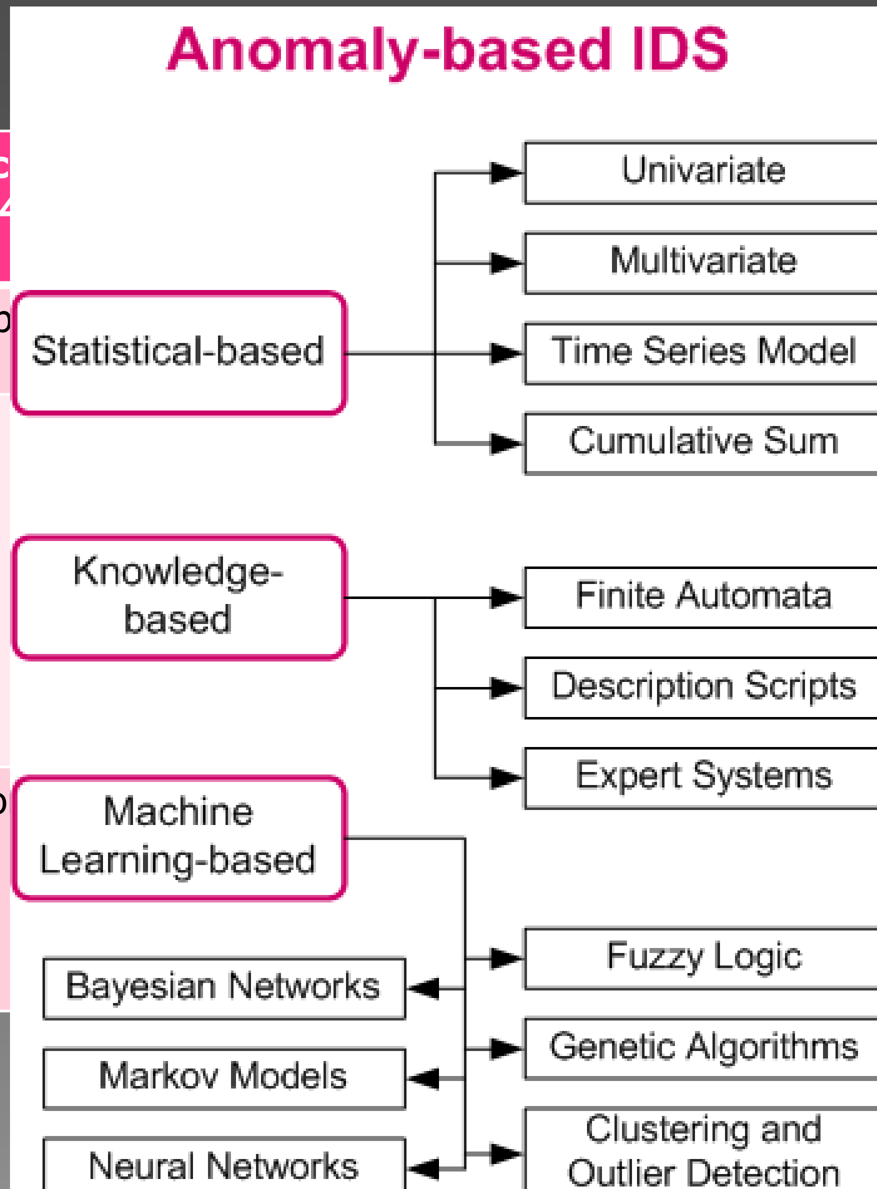
IDS Evaluation Methodology

1. Detection methodology
2. Protected protocols
3. Implementation tools
4. Test environment
5. Performance evaluation

- Overall assessment is performed based on the previous five evaluation properties.

1. Detection methodology

General classification [1]	Adopted c... [2], [3], [4...	...on	Alternative terms used in the literature
Blacklist	Signature-b...	...ased	Rule-based Misuse detection
Whitelist	Anomaly-based	...ed	Behavior-based
		...ed	Rule-based (expert systems)
		...ed	Behavior-based
	Specificatio... based	...based	Stateful protocol analysis Deep packet inspection Model-based



2. Protected protocols

- Most widespread SCADA-specific protocols: Modbus, IEC 60870-5 series, DNP3, IEC 61850 series, and EtherNet/IP
- Majority of protocols are created or extended to operate over TCP/IP networks
- Most of the current fieldbus protocols are Ethernet-based

3. Implementation tools

Snort: most widely deployed IDS worldwide: relies on a relatively simple language for specification of misuses and attack signatures

Suricata: newer network threat detection engine capable of real-time intrusion detection, inline intrusion prevention, network security monitoring and offline processing of captured packets

Bro (Zeek): a passive, open-source network traffic analyzer, which is organized into two major components: event engine and policy script interpreter

General-purpose programming languages: C, C++, C#, Perl, Python, Java, are also used to develop SCADA-specific IDS applications

General-purpose open-source tools (WEKA, TensorFlow, LIBSVM, Anaconda) are used to build SCADA-specific solutions

4. Test environment

1. Pen-testing activities (typical for non-industrial environments) – unacceptable for SCADA and other industrial control systems
2. New security solutions need reliable test environments that meet the requirements regarding fidelity, repeatability, measurement accuracy and safe execution
3. Test environment encompasses testbed, datasets and simulated attacks
4. Testbed is a platform for conducting exhaustive, transparent, and replicable testing of algorithms, methods, prototypes, etc.
5. SCADA security testbed can be implemented in one of the following ways:
   - **Cyber physical system (CPS) testbed**: uses real hardware and software to pursue lines of experimentation and exploration
   - **Emulation-based testbed**: may use different combinations of physical devices and software to simulate the control network and the physical process
   - **Software simulation testbed**: can be simple simulation-based (assumes a single software simulation package for testing purposes) or federated simulation-based (may have several interacting simulations such as plant, network, etc.)
   - **Virtualization-based testbed**: uses virtualization technology to build a low-cost, high-fidelity, reusable, and easy-to-maintain testbed

5. Performance evaluation

1. No dedicated performance evaluation techniques for SCADA-specific IDSs
2. General techniques developed for IDS evaluation in public and enterprise IT networks are used
3. We focus on the following criteria: detection accuracy, timeliness, response to incidents and efficiency

5. Performance evaluation – Detection accuracy

- **Known as classification accuracy or effectiveness:**
  - represents the ability of the system to distinguish between intrusive and non-intrusive activities
  - it is represented by a set of measures that determine how correctly

Confusion matrix	Derived evaluation metrics
(see confusion matrix below)	(see metrics below)

Confusion matrix

Actual	Predicted	
	Attack	Normal
Attack	*TP*	*FN*
Normal	*FP*	*TN*

Derived evaluation metrics

1. $FPR = FP/(FP + TN)$
2. $FNR = FN/(TP + FN))$
3. $DR = TPR = Recall = TP/(TP + FN) = 1 - FNR$
4. $TNR = TN/(FP + TN) = 1 - FPR$
5. $Accuracy = (TP + TN)/(TP + TN + FP + FN)$
6. $Precision = TP/(TP + FP)$
7. $F\text{-}measure = 2/(1/Precision + 1/Recall) = 2TP/(2TP + FP + FN)$

5. Performance evaluation - Timeliness

1. Refers to the system's ability to perform its analysis as quickly as possible
2. Objective – enable prompt response to incident to minimize the damage within a specific time period
3. Timeliness is usually estimated concerning the time needed to process the unit of analysis (packet, group of packets, traffic flow, communication session or dataset instance)
4. Detection latency – time between the attack detection and the actual moment of the attack
5. Total delay – time between the response of the system and the actual moment of the attack

5. Performance evaluation - Response to incidents

- Passive response - assumes alert generation after detection of an incident
- Active response – encompasses prevention capabilities and/or integration with the other security mechanisms
- Intrusion prevention system (IPS) – a tool that generates response to detected threats by attempt to preclude their realization
- Both IDS and IPS are integral parts of the overall security management system
- Efficient solution typically assumes combination of different technologies

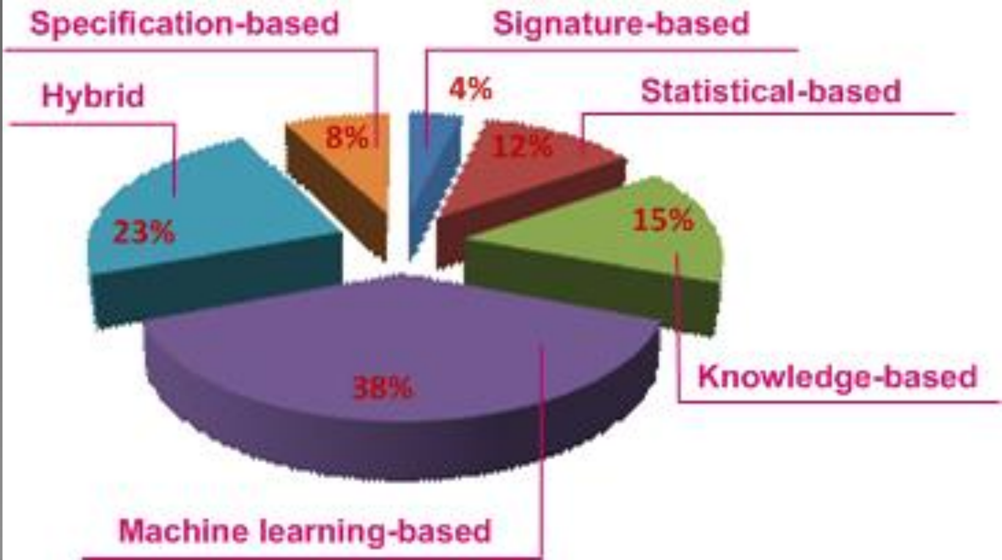
5. Performance evaluation - Efficiency

- Refers to the resources needed to be allocated to the system including CPU and memory usage
- IDS can collect and analyze data continually as the data is acquired or in blocks, after an event has occurred
- Continuous mode, also known as real-time processing, provides the opportunity for administrator to take action while the intrusion is in progress
- Performance of any Network IDS depends on its configuration, monitored network properties, and the system's placement in that network

EVALUATION AND COMPARISON OF SOLUTIONS

DETECTION METHODOLOGY

- Anomaly-based methods prevail
- Reason for their expansion:
  - inherent suitability for SCADA systems in terms of identifying traffic patterns
  - capability to support FIN technologies, high level of automation and continuous detection improvement
- Signature-based techniques are practically being abandoned

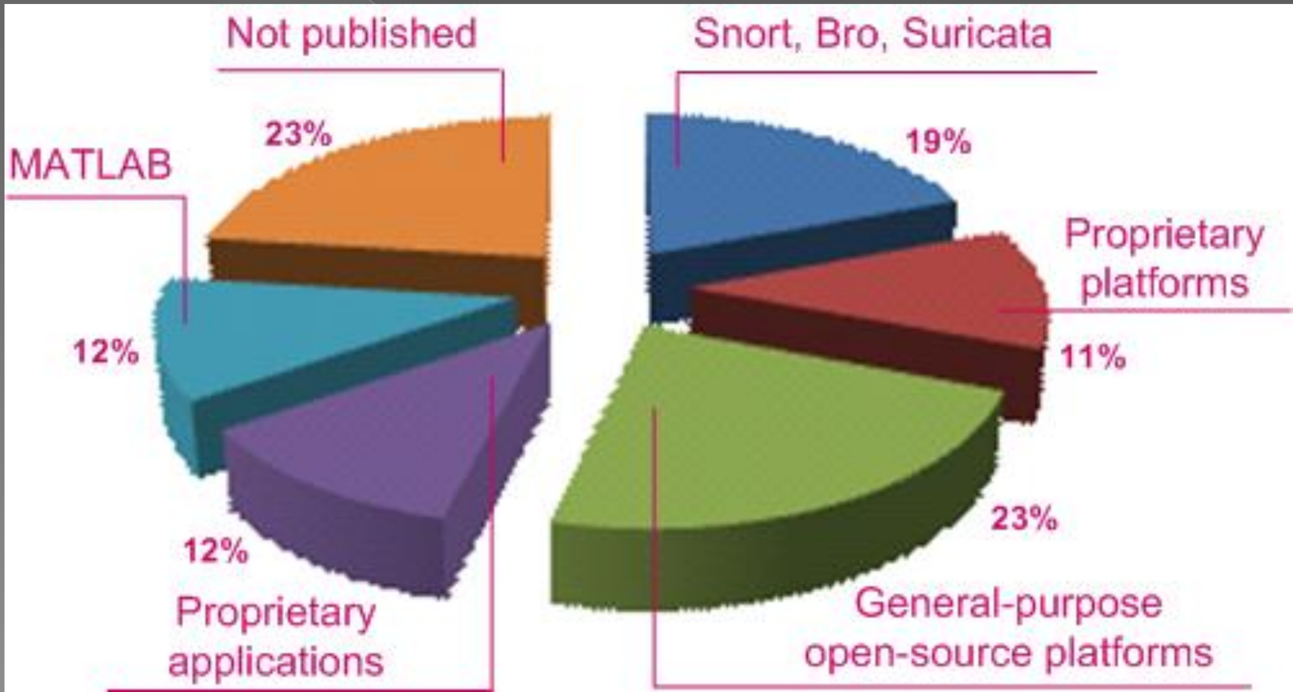


PROTECTED PROTOCOLS

- Most widespread SCADA protocols are comprised in surveyed studies, including Modbus, IEC 60870-5 series, DNP3, IEC 61850 and EtherNet/IP
- About 69% of surveyed papers consider only one protocol
- 12% of surveyed papers deal with multiprotocol environments
- The information about SCADA protocol is not available in 19% of surveyed papers

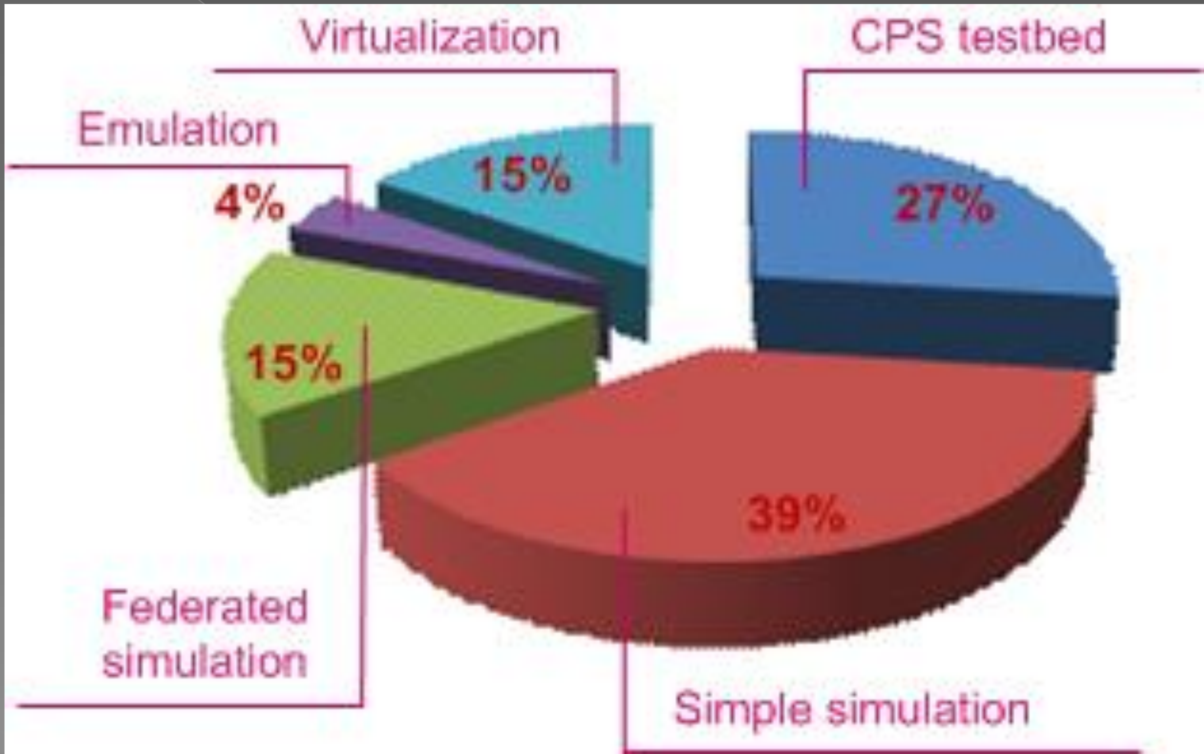
IMPLEMENTATION TOOLS

- Open-source NIDS (Snort, Suricata and Bro) are predominantly used for signature-based and hybrid techniques
- MATLAB is used for implementation of algorithms
- General-purpose programming languages (C/C++, Java, Python) for developing proprietary applications/platforms –> used to build SCADA-specific solutions



TEST ENVIRONMENT - TESTBEDS

- Seven solutions have been verified in powerful CPS testbeds
- Software simulation testbeds prevail, with majority of simple simulation based testbeds
- Four testbeds are virtualization-based
- One testbed is emulation-based



TEST ENVIRONMENT - DATASETS

- Only five papers include tests with real SCADA network data
- The other 21 papers include one or more experimental and/or synthetic datasets:
  - 15 datasets are publicly available
  - only two of public datasets are not SCADA-specific – KDD99 and UNSW-NB15

TEST ENVIRONMENT – SIMULATED ATTACKS

- The most diverse situation
- In some cases, system's behavior under attacks was not analyzed
  - tests were performed on a real system and limited to suspicious messages and events or
  - the tests were focused only on system's efficiency
- The most common simulated attacks comprise the following attacks:
  - Attacks on general Internet protocols – 11
  - Command/response injection or modification – 10
  - DoS – 10
  - Attacks on SCADA protocols – 7
- Other simulated attacks were:
  - Reconnaissance – 5
  - MITM – 3
  - Unauthorized access – 1
  - Probing – 1
- Six studies with thorough specification and simulation of a number of realistic attacks intended to jeopardize the particular control process
- Only two studies included independent validation performed by invited hackers and six independent teams

PERFORMANCE EVALUATION – DETECTION ACCURACY

- 6 papers – analysis is not presented or the results are given in a descriptive way
- 2 papers – descriptive results rather than well-defined evaluation metrics
- 8 papers – results presented through smaller number of evaluation metrics (typically Accuracy and FPR)
- 11 papers – comprehensive detection accuracy analysis
- Statistical-based techniques provide high accuracy, with low FPR and FNR rates
- Knowledge-based techniques provide good overall accuracy
- Among machine learning-based techniques, deep learning based on CNN outperforms techniques based on clustering and outlier detection
- Detection accuracy of hybrid methods depends on combined techniques

PERFORMANCE EVALUATION – TIMELINESS

- Timeliness analysis is available in 11 studies
- Among the results concerning packet as a unit of analysis, deep packet inspection outperforms other techniques at least for an order of magnitude
- If dataset instance is observed as a unit of analysis, deep learning method performs much worse than clustering and outlier detection and hybrid method
  - This is not surprising since deep learning inherently requires large datasets to obtain high accuracy
- The results concerning detection latency are hardly comparable probably due to different experimentation platforms
  - thus, detection latency seems lower in test scenarios with simple simulation

PERFORMANCE EVALUATION – RESPONSE TO INCIDENTS

- Only four systems provide active responses to detected attacks

Reference	Type of active response
[1]	Linked with the distributed multilevel correlation structure
[2]	(1) Alarm generation; (2) Automatic response – redirecting of anomalous flow to Honeypot, dropping malicious packets
[3]	Exploits reclose logic in relays to prevent physical damage caused by an attempt to disconnect multiple transmission lines
[4]	Data encryption using the Hybrid Elliptical Curve Cryptography

PERFORMANCE EVALUATION – EFFICIENCY

- Only five papers provide the results of efficiency evaluation
- Results concerning memory usage are comparable for 3 solutions, while the other two use less memory (of an order or two of magnitude)
- Results concerning CPU usage are hardly comparable due to different processor platforms – results presented in two papers confirm that CPU usage increases for higher traffic load
- Packet loss and/or alert loss under high traffic load are addressed in one paper

SUMMARY OF REVIEW FINDINGS (1)

1. Signature-based techniques are insufficient to secure SCADA systems –> due to their inherent drawbacks regarding inability to cope with new or unknown threats and the need to continuously update signatures
2. Machine learning-based techniques have gained a strong momentum in the past few years (stand-alone or in combination with other techniques)
3. Knowledge-based techniques perform better in terms of detection accuracy, but on the count of deteriorated timeliness, especially for large-scale systems
4. Statistical-based techniques are most useful in hybrid techniques because of their high detection accuracy

SUMMARY OF REVIEW FINDINGS (2)

5. Specification-based techniques gain in importance for SCADA application layer protocols –> perform well in terms of both detection accuracy and per-unit processing time
6. Integration of two or more detection methods may contribute to improvement of the IDS scope and detection accuracy
7. The most widespread SCADA protocols are addressed in the recent works:
   - Modbus TCP prevails
   - Additional research efforts are needed towards environments such as digital substations and smartgrids

SUMMARY OF REVIEW FINDINGS (3)

8. Open-source NIDS implementation tools (Snort, Bro and Suricata) are superseded by open-source and proprietary IDS platforms
9. Realistic and comprehensive cyber physical system testbeds are needed to allow for experimentation with different solutions
   - If they are unavailable, sophisticated simulation/emulation testbeds should be developed
   - virtualization may help to provide inexpensive, credible and reusable testbeds
   - Simple simulation-based testbeds should be avoided due to their low fidelity and poor reusability

SUMMARY OF REVIEW FINDINGS (4)

10. A strong need to use datasets from real SCADA networks:
    - National strategies for critical infrastructure protection should find a way to make them available to research community
    - Good strategy is to reuse datasets, either publicly available or obtained from CPS testbeds
11. A lack of proper attack models and scenarios in which the attackers try to exploit vulnerabilities in SCADA systems:
    - Reports on IDS performance evaluation might be insufficiently reliable and hardly comparable
    - Efforts are needed to improve frameworks for modeling cyber attacks and procedures to apply them in the appropriate testbeds

SUMMARY OF REVIEW FINDINGS (5)

12. Performance evaluation remains the most critical issue
13. The work is needed on identification and specification of requirements for IDSs in SCADA networks, and establishing a common set of performance metrics:
    - At least detection accuracy, timeliness, response to incidents and efficiency
    - Procedures for IDS performance testing should be established in accordance with the predefined set of requirements
    - Timeliness analysis should be presented in each new proposal, since it is crucial parameter to assess system's ability to respond to incident in real time
    - Efficiency analysis is important (under heavy traffic load, if possible) –> represents indirect measures that take into account the time and space complexities of intrusion detection algorithm

SUMMARY OF REVIEW FINDINGS (6)

14. Only four of surveyed papers discussed active responses to detected attacks
15. A strong need to perceive the overall SCADA security system architecture and to define procedures for real-time interaction of the IDS and other components of the security system like correlators, SIEM software, etc.
16. Particularly, work on IPS capabilities should be strongly encouraged
17. Evolution towards fourth-generation SCADA brings new research challenges related to security in industrial IoT environment that assumes the use of FIN technologies

CONCLUSION (1)

- Growth of solutions for SCADA IDS gains in importance with proliferation of advanced networking technologies and the ongoing evolution towards fourth-generation SCADA systems
- Evaluation methodology was proposed encompassing identification of general IDS features and analysis of system's characteristics regarding detection technique, protected protocols, implementation tools, test environment and performance evaluation
- Final assessment is performed based on the previous analysis, including strengths, weaknesses, maturity stage, as well as portability to FIN environment

CONCLUSION (2)

- Results of our study –> significant progress in developing new intrusion detection methods –> using open-source implementation tools and creating sophisticated security testbeds
- The most important future research directions:
  - development of proper attack models
  - establishment of procedures for IDS performance evaluation
  - integration of IDS with other components of ICS security system (bearing in mind the migration towards Future Internet environment)

S. V. B. Rakas, M. D. Stojanović and J. D. Marković-Petrović, "A Review of Research Work on Network-Based SCADA Intrusion Detection Systems," in *IEEE Access*, vol. 8, pp. 93083-93108, 2020, doi: 10.1109/ACCESS.2020.2994961.

https://ieeexplore.ieee.org/document/9094250