PATENT SPECIFICATION

United States Patent No. US11387654

AI-Based Training Data Processing System for Industrial Control Network Security

Filing Date: March 15, 2021

Issue Date: September 22, 2022

Assignee: DeepShield Systems, Inc., Delaware

Inventors: Rodriguez, Elena; Chen, Marcus; Morrison, James

ABSTRACT

A system and method for processing training data for artificial intelligence-based security monitoring of industrial control systems (ICS) and operational technology (OT) networks. The invention provides novel techniques for preprocessing, anonymizing, and augmenting network traffic data while preserving critical security-relevant features. The system employs a multi-stage pipeline architecture for converting raw industrial protocol data into normalized training sets optimized for anomaly detection models.

BACKGROUND

[0001] Industrial control systems face increasing cybersecurity threats requiring advanced detection capabilities. Traditional signature-based approaches prove insufficient for zero-day attacks and sophisticated threats targeting critical infrastructure. Machine learning models require extensive high-quality training data, but collecting and processing such data from operational industrial environments presents significant challenges.

[0002] Prior approaches fail to adequately address the unique characteristics of industrial protocols and operational patterns while maintaining data fidelity for AI model training.

SUMMARY OF THE INVENTION

[0003] The present invention provides systems and methods for processing industrial network traffic data to generate optimized training datasets for AI-based security monitoring systems. Key innovations include:

a) Protocol-aware feature extraction preserving operational context

b) Automated anonymization of sensitive industrial parameters

c) Synthetic data augmentation maintaining statistical properties

d) Adaptive sampling based on operational patterns

e) Multi-stage validation ensuring training data quality

DETAILED DESCRIPTION

[0004] System Architecture

The system comprises:

- Data ingestion module supporting major industrial protocols
- Protocol parsing and feature extraction pipeline
- Anonymization engine with configurable rules
- Synthetic data generation module
- Quality validation and verification system
- Export interface for training pipeline integration

[0005] Data Processing Pipeline

The processing pipeline implements:

Protocol-specific parsing and normalization

Temporal feature extraction

Operational context preservation

Parameter anonymization

Statistical augmentation

Quality validation

[0006] Feature Extraction Method

The system extracts features including:

- Command sequences and timing patterns
- Process variable relationships
- Control loop characteristics
- Network topology attributes
- Protocol-specific metadata

[0007] Anonymization Process

Sensitive data protection through:

- Parameter value randomization
- Topology obfuscation
- Identity masking
- Relationship preservation
- Configurable sensitivity rules

CLAIMS

What is claimed is:

A method for processing industrial network traffic data comprising:

a) Receiving raw network capture files

b) Extracting protocol-specific features

c) Applying anonymization rules

d) Generating synthetic augmentation data

e) Validating training set quality

f) Outputting processed datasets

The method of claim 1 wherein feature extraction preserves:

a) Temporal relationships

b) Process variable correlations

c) Control system behaviors

d) Network topology attributes

A system implementing the method of claim 1 comprising:

a) Data ingestion interface

b) Processing pipeline modules

c) Anonymization engine

d) Synthetic data generator

e) Quality validation system

f) Export interface

DRAWINGS

[0008] FIG. 1 illustrates the system architecture

[0009] FIG. 2 shows the processing pipeline flow

[0010] FIG. 3 depicts the feature extraction process

[0011] FIG. 4 demonstrates anonymization examples

TECHNICAL FIELD

[0012] The invention relates to cybersecurity, artificial intelligence, industrial control systems, operational technology, and data processing methods for machine learning applications in critical infrastructure protection.

INDUSTRIAL APPLICABILITY

[0013] This invention has direct application in:

- Industrial cybersecurity systems
- Critical infrastructure protection
- Manufacturing operations security
- Maritime and offshore facility monitoring
- Smart grid and utility network security

LEGAL NOTICES

The information and designs disclosed herein are proprietary to DeepShield Systems, Inc. and protected under applicable intellectual property laws. All rights reserved. Unauthorized use, reproduction, or distribution is prohibited.

Patent Attorney of Record:

Sarah Williams, Reg. No. 65432

Williams & Associates, LLP

101 Technology Drive

Boston, MA 02110