SP

SpacSec 2026

A Reproducible Evaluation Framework for Immune-Augmented Intrusion Detection

Anton Shilov · Andrei Petrovski · Igor Kotenko

Presenter: Anton Shilov

International Conference on Cyber Warfare, Security & Space Computing
Saint Petersburg, 26–28 October 2026

1. Introduction — Research Goal

SP SpacSec 2026

Topic: anomaly detection in cybersecurity events and SOC alerting

Goal

Develop and rigorously evaluate a modular RF–AIS framework for anomaly detection and SOC alerting.

Three Objectives

- Build independent RF and AIS scoring branches with configurable fusion.
- Benchmark detection quality against strong RF and control baselines.
- Assess reproducibility, rare attack categories and alert load.

2. Background Research

SP SpacSec 2026

Strong RF baseline

Supervised learning works well on labelled, structured security events.

Changing attacks

Rare behaviours and distribution shifts may differ from training examples.

Missed incidents

False negatives leave attacks undetected; higher recall is valuable.

Alert overload

False positives consume analyst capacity; alert rate matters.

Research question: can AIS provide a complementary signal without excessive false alarms?

3.1 Experimental Setting and Data

SP SpacSec 2026

Primary benchmark: UNSW-NB15

40,000	10,000	30,000
training	validation	test

10 seeds • binary detection • multiclass attack categories • open-set stress tests

No test-set tuning: preprocess/train on training data; select fusion and thresholds on validation.

3.2 Baseline Components: RF and AIS

Random Forest

Learns discriminative patterns from labelled event features.
Output: supervised score s_RF(x).

Artificial Immune System

Measures deviation from normal/self patterns.
Output: anomaly score s_AIS(x).

The question is not “RF versus AIS”, but whether AIS adds information beyond RF.

3.3 Hybridization and Decision Rule

AIS SIGNAL GENERATORS

v1 Negative selection

Original detector branch; nearly inactive.

v2 kNN self-distance

Deviation from nearby normal examples.

v3 Detector pruning

Validation-based removal of unstable detectors.

v4 Immune memory

Targets RF false-negative patterns.

Fusion: max-rule • weighted • gated • stacking • AIS-as-features

3.4 Evaluation Protocol

SP SpacSec 2026

Compared configurations

RF-only and AIS-only
RF + validation threshold
Max-rule and weighted fusion
Random and shuffled auxiliary controls

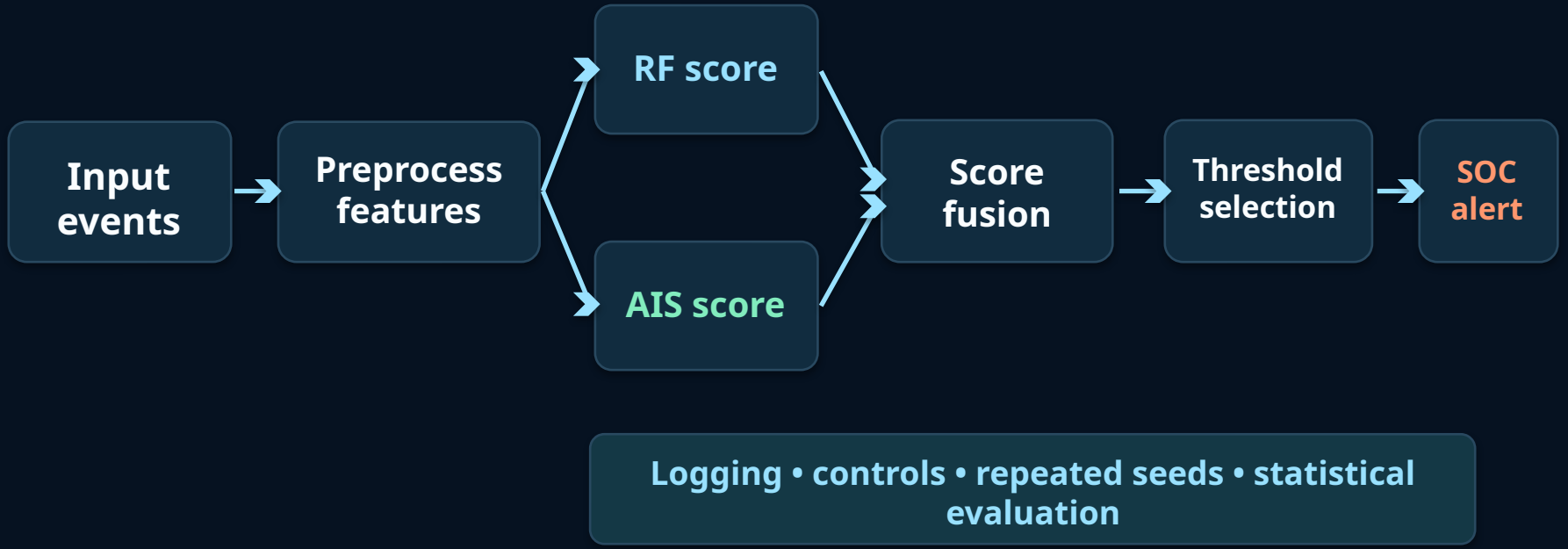
Reported evidence

Precision • Recall • F1-score
False positives / false negatives
Alert rate and ROC-AUC
Confidence intervals and paired tests

Criterion: a hybrid gain must be stable across seeds and not reproduced by control signals.

3.5 Software Framework Architecture

SP SpacSec 2026



RF-only path remains the strong benchmark; AIS and evaluation controls are interchangeable.

Paper §3.5 · Components corresponding to Fig. 1

4.2 External Validation on UNSW-NB15

SP SpacSec 2026

TWO OPERATING POINTS • ONE RF–AIS FRAMEWORK

MAX-RULE • HIGHER RECALL

−32.4%

MISSED ATTACK RECORDS

FN 241 → 163 | Recall 98.54% → 99.01%

Trade-off: FP 3,914 → 4,345

WEIGHTED • LOWER ALERT LOAD

−14.6%

FALSE-POSITIVE ALERTS

FP 3,914 → 3,342 | F1 0.8868 → 0.8972

Trade-off: FN 241 → 359

MAIN UNSW-NB15 TEST SEED · Configuration-specific gains; not a ten-seed guarantee

Paper §4.1–4.2 · External validation; measured FP/FN trade-offs

4.3 Multi-Seed Statistical Qualification

SP SpacSec 2026

−0.0004	0.9219	10
mean ΔF1 (max-rule)	Wilcoxon p-value	random seeds

95% CI for ΔF1: [−0.0064, +0.0060]

Takeaway: rigorous multi-seed testing rules out naive gains and guides selective AIS fusion.

Paper §4.3 · Ten-seed verification

4.4 AIS Diagnostics and Revised Variants

Metric	RF-only	Hybrid candidate	Difference
Mean F1-score	0.8886	0.8911	+0.0025
Mean precision	0.8076	0.8139	+0.0063
Mean false positives	3,886.6	3,726.0	−160.6
Mean alert rate	0.6734	0.6664	−0.0070
Mean recall	0.9876	0.9847	−0.0029

Across 10 seeds: 161 fewer FP on average and +0.0025 F1; effect not conclusive (p = 0.375).

Paper §4.4 · Validation-selected AIS v2/v3/v4

4.5 Multiclass Attack-Category Analysis

SP SpacSec 2026

+0.0064
mean F1 on rare classes

6
attack categories improved

0.4646
hybrid Macro-F1

Improved: Worms • DoS • Reconnaissance • Backdoor • Analysis • Generic

Overall Macro-F1: RF 0.4649 | RF–AIS 0.4646

Paper §4.5 · Rare-class and category-level gains

5. Discussion — Why the Framework Matters

SP SpacSec 2026

FROM A MODEL SCORE TO AN EVALUATION SYSTEM

01

Operational trade-offs

Measure FN, FP and alert rate — beyond F1-score.

02

Strict control baselines

Test whether AIS adds value beyond threshold and random-score effects.

03

Attack-category analysis

Six categories had local F1 gains; open-set remains exploratory.

Key insight: selective integration must be backed by reproducible evidence.

6. Conclusions and Future Work

THE RESEARCH GOAL WAS ACHIEVED

A modular RF–AIS evaluation framework was built, tested and characterized.

01 BUILT A REUSABLE FRAMEWORK
RF/AIS modules · selectable fusion · calibrated thresholds

02 MEASURED OPERATIONAL EFFECTS
Main seed: −32.4% FN (max-rule) | −14.6% FP (weighted)

03 MAPPED AIS APPLICABILITY
10-seed validation · 6 classes with local gains · rare-class F1 +0.0064

Caveat: single-seed gains are configuration-specific; overall F1 superiority is not established.