International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Negative Selection Algorithm in Static and Dynamic Data Environments: A Performance Analysis

Rejeesh E [1], Dr. Shijo M Joseph[2], Anupama M[3]

[1]Research Scholar, Department of IT, Kannur University
Email: *rejeesh_e[at]mgcollege.ac.in*

[2]Research Supervisor, Department of IT, Kannur University
Email: *shijomjose71[at]gmail.com*

[3]Research Scholar, Department of IT, Kannur University
Email: *anupama.cs.mgc[at]gmail.com*

Abstract: ***Network Intrusion Detection Systems (NIDS) increasingly rely on bio-inspired computational paradigms to cope with the volume, velocity, and variety of modern traffic. Among these, the Negative Selection Algorithm (NSA), a core mechanism of Artificial Immune Systems (AIS) modelled on self/non-self discrimination in the mammalian immune system, has been extensively investigated for anomaly-based intrusion detection because it requires only normal (self) samples for training and can, in principle, generalize to previously unseen (non-self) attacks. This paper reviews the evolution of NSA from its original binary, r-contiguous-bit formulation to real-valued, variable-sized-detector, density-guided, and hybrid deep-learning-assisted variants. The performance of these variants is examined separately for static benchmark datasets (e.g., NSL-KDD, KDD-Cup99, UNSW-NB15, CICIDS) and for dynamic, high-velocity, and concept-drifting environments typified by Internet of Things (IoT) and Industrial IoT networks. This paper is a comparative analysis of the performance of NSA variants and identified open challenges: detector-hole formation, curse of dimensionality, computational cost of detector regeneration, and the absence of native mechanisms for continual/incremental learning are discussed. This generated an insight to advancement of NSA to cover a decisive gap remains in ability to self-adapt to non-stationary, rapidly growing IoT data streams and lightweight NSA architectures suited to edge deployment.***

Keywords: Negative Selection Algorithm; Artificial Immune System; Intrusion Detection System; Anomaly Detection; Internet of Things; Concept Drift; Dynamic Data Streams; NSL-KDD

1. Introduction

The continuous expansion of networked infrastructure- from enterprise networks to large-scale IoT deployments- has amplified the attack surface available to adversaries and rendered signature-based Intrusion Detection Systems (IDS) increasingly insufficient against novel and zero-day threats [1]. Anomaly-based detection, which models normal behaviour and flags deviations, is therefore of enduring interest. Among anomaly-detection paradigms, Artificial Immune Systems (AIS)- inspired by the adaptive, distributed, and self-organizing properties of the biological immune system- have been applied to intrusion detection since the mid-1990s [1], [10]. The Negative Selection Algorithm (NSA), first proposed as a computational analogue of T-cell maturation in the thymus, generates a set of "detectors" that do not match any member of a self (normal) training set; any subsequent instance matched by a detector is classified as non-self (anomalous) [1].

NSA is particularly attractive for network intrusion detection because (a) it requires only normal traffic for training, sidestepping the scarcity and rapid obsolescence of labelled attack data, and (b) its detector population is, in theory, capable of covering an open-ended non-self space, offering some resilience to unseen attacks [2], [7]. However, three decades of research have shown that naive NSA implementations suffer from detector-hole formation, poor scalability in high-dimensional feature spaces, high false-positive rates, and a computationally expensive detector-generation phase [6], [27], [30]. These limitations have motivated a large and still-growing body of work refining detector representation, generation strategy, and matching rules, and- more recently- combining NSA with feature selection, clustering, and deep learning to meet the demands of large, heterogeneous, and fast-changing network data such as that produced by IoT ecosystems [8], [15].

This paper synthesizes recent advances in NSA-based intrusion detection with two objectives. First, it organizes the major algorithmic lineages of NSA (binary, real-valued, variable-sized-detector, density/clustering-guided, hybrid, and distributed variants) and summarizes their reported performance on standard static benchmark datasets. Second, it specifically examines how well these variants transfer to dynamic data environments — streaming, concept-drifting, and resource-constrained IoT settings- where the assumption of a fixed, representative self-set no longer holds. The paper closes with a comparative table and a set of research directions aimed at closing the static-to-dynamic performance gap [36].

2. Background: NSA Fundamentals

Classical NSA is a phased paradigm. In first phase, it generate candidate detectors in randomly, or via evolutionary or density-guided heuristics methods and matched it to a self-set. If any candidate among tgis matches a self-sample then it will be discarded. Finally the remaining "mature" detectors are retained as detector set. In the monitoring (detection) phase, incoming traffic instances are compared against the mature detector set; a match indicates a non-self (anomalous/intrusive) instance [1], [10]. In 1994, Forest et al. used a model using binary string representations with r-

contiguous-bit matching in early implementations [1]; from this idea, Hofmeyer and Forest's ARTIS developed NSA into a distributed architecture for network intrusion detection [2].In 1994, Forest et al. used a model using binary string representations with r-contiguous-bit matching in early implementations [1]; from this idea, Hofmeyer and Forest's ARTIS developed NSA into a distributed architecture for network intrusion detection [2]. A model that maps to a normalized hypercube and uses Minkowski-family distance matrices for matching was introduced by Gonzalez and Dasgupta. In this, real-valued NSA (RNSA) self/non-self events can be directly applied to continuous network-flow features [3], [4]. With the aim of significantly reducing the number of detectors required for a given target coverage, Ji and Dasgupta introduced the V-detector by incorporating variable-radius detectors with a statistical stopping criterion [5], [6].

3. Recent Advancements in NSA

Four new paths have been opened up with the emphasis on the RNSA/V-detector lineage. One is boundary- and density-awareness in detector generation. The second is DENSA, which uses flexible detector boundaries and dynamically determined detector counts to better fit irregular self-space geometries [13]. The third is an approach that adopts antigen-density-based approaches, using clustering of self-sets to bias detector placement to low self-density regions, reducing redundant coverage and narrowing detector holes near class boundaries [14], [15].

Hybridization of classical machine learning has achieved remarkable results. One example is the combination of NSA with k-nearest-neighbors or other classifiers, which can be more accurate than V-detector baselines, to resolve ambiguous boundary events that are misclassified in models using NSA alone [18]. There are also models that can reduce redundant coverage and detector-generation time by generating detectors close to the self-set using immuno-optimization variants [23].

Feature dimensionality is a factor that negatively affects the performance of NSA. Several models have been recently introduced, such as NIDS-NSA models, which seek to solve the problem of information-gain, PCA, or metaheuristic features before detector generation, with dimensionality reduction and feature selection as the main focus. These models have been shown to overcome the scalability and false-positive problems to some extent when working with NSL-KDD-style datasets [1], [32]. Feature selection and dimensionality reduction.

Distributed and hardware-aware generation. MapReduce-based and other partitioned schemes parallelise detector generation to scale NSA to larger offline corpora, while lightweight/hypercube-interface detector schemes aim to reduce the per-detector computational footprint for constrained devices [21], [25].

Domain-based extensions of NSA enable greater accuracy in general. Examples of this performance improvement include improved V-detector schemes for wireless sensor networks [16], dynamic immune algorithms developed for industrial cloud storage [17], and models incorporated into cyber-physical systems and fault/structural-health monitoring. [22], [23], [26], [27].

These advances have been shown to improve detection accuracy and reduce detector redundancy when compared to the original RNSA/V-Detector baselines on known static intrusion datasets [6], [27].

4. Performance in Static Data Environments

The overwhelming majority of NSA evaluations to date use static, pre-collected benchmark datasets- principally KDD-Cup99, NSL-KDD, UNSW-NB15, and, more recently, CICIDS and IoT-specific captures such as IoTID20 [1], [9], [14]. In this setup, the advanced density-aware, hybrid classifier-enhanced variants of NSA report robust results because a certain self-set is available offline and can thus be run thoroughly before the detector-generation phase is deployed. Variants such as nearly complete non-self coverage with a small number of detectors, low false-positive rates, , accuracy in competitiveness, and state-of-the-art supervised deep-learning classifiers on static splits have shown superior results compared to classical RNSA [13], [14], [13], [14], [18]. The main limitation of NSA, dimensionality sensitivity, is overcome by implementing feature-selection-augmented NSA pipelines by removing redundant or noisy attributes before detector generation[1], [32]. NSA has emerged as a reasonably competitive anomaly-detection technique on static datasets where only well-curated common-traffic labels exist.

5. Performance in Dynamic Data Environments

The picture changes markedly once the self-set is no longer static. Modern network environments- and IoT/Industrial-IoT deployments in particular- generate high-velocity, high-volume, heterogeneous traffic whose statistical properties drift over time as devices, firmware, usage patterns, and attack techniques evolve [8], [9], [31]. The classical NSA approach is to create a fixed detector population by collecting a large number of stationary self-sets based on a hypothetical representative. When such initially mature detectors deviate from the underlying "normal" distribution (concept drift), they are more likely to misclassify legitimate new-normal traffic as non-self, and the method may fail to capture newly emerging non-self regions as part of the detector, leading to an increase in false positives and miss detections [28], [31], [33]. Recent AIS surveys explicitly flag adaptability to real-time, evolving data as an unresolved challenge for negative-selection-based anomaly detection, noting that most published NSA variants- including RNSA, V-Detector, and Grid-based NSA- have been evaluated primarily on static corpora rather than genuine streaming settings [27], [30], [31]. Where NSA-related techniques have been extended toward dynamic operation- for example, dynamic immune algorithms for industrial cloud storage, or incremental adaptive-moment-estimation-based intrusion detection for cyber-physical grids- the emphasis has been on periodic retraining or incremental model updates rather than a detector-generation process that is intrinsically online [17], [26]. For IoT specifically, the dominant recent trend has instead been to pair lightweight feature selection or hybrid

deep/machine-learning classifiers with constrained-device constraints, largely bypassing rather than resolving NSA's own adaptation limitations [8], [9], [29], [32]. It can be seen that NSA works reliably when the data environment is stable, predictable, and constant. NSA cannot achieve this performance when the self-sets in IoT traffic are precisely that of dynamic and ever-growing.

6. Comparative Analysis

In this comparative analysis section, the major lineages of NSA discussed above are compared in the perspective of their reported static-benchmark performance with performance for dynamic, IoT-scale operation and notated in Table 1.

Table 1: Comparative summary of NSA variants across static and dynamic/IoT data environments

NSA Variant	Core Mechanism	Detector Rep.	Static-Data Performance	Dynamic/IoT Suitability
Binary NSA (Forrest et al., 1994)	r-contiguous bit matching	Binary strings	Accuracy - Moderate; Hole ratio - High	Poor — fixed self-set, no incremental update
ARTIS (Hofmeyr & Forrest, 2000)	Distributed binary detectors, r-chunk match	Binary	Good on small network self-profiles	Limited scalability to high-rate streams
RNSA (Gonzalez & Dasgupta, 2003)	Random real-valued detector generation	Hypersphere (fixed radius)	Good on NSL-KDD subsets	Weak — static radius unsuited to drifting self-region
V-Detector (Ji & Dasgupta, 2004/2009)	Variable-radius, statistical coverage estimate	Hypersphere (variable radius)	Accuracy - High Hole ratio - Low	single-pass, offline training
DENSA (Fouladvand et al., 2017)	Flexible boundary, adaptive detector count	Variable-shape	Very good on imbalanced self-space	Better adaptability; retraining still batch-based
Antigen-Density NSA (Yang et al., 2017/2020)	Density-clustering guided detector placement	Clustered real-valued	Strong on dense/sparse mixed data	Promising for evolving density but not streaming-native
Hybrid NSA + kNN / ML (2020-2024)	NSA for coarse detection + classifier at boundary	Hybrid	Best static accuracy (often >95%)	Moderate; added inference cost limits real-time IoT use
Distributed/MapReduce NSA (e.g., VOR-NSA)	Parallelised detector generation	Real-valued, partitioned	Scales to large offline datasets	Good throughput but not designed for concept drift
Deep-hybrid / GAN-assisted NSA (2023-2025)	NSA seeded/ augmented by deep generative or LLM features	Learned embeddings	Highest reported accuracy on benchmark sets	Emerging; still validated mostly on static IoT captures

7. Challenges and Open Issues

As feature dimensionality and self-space disorder, which are important factors in determining the nature of data sites, increase, the number of detector-holes that can form at self/non-self boundaries increases significantly [6], [21], [30]. The computational cost of detector regeneration is a major constraint that makes it impractical to provide high-throughput security through periodic full retraining for resource-constrained IoT nodes [17], [25].

Lack of native incremental/online learning:
All NSA variants considered in the survey follow the path of generating a detector set once from a fixed self-sample, which follows from classic NSA. In constantly growing data environments such as IoT, the lack of a mechanism to add more detectors and retire or redesign the less efficient ones as the traffic distribution drifts remains a drawback. [27], [30], [31].

Benchmark mismatch: Static datasets such as NSL-KDD and KDD-Cup99 are unable to reflect key features of IoT traffic, such as scale, diversity, or drift factors, which is important for limiting the environmental validity of reported accuracy estimates [8], [14].

Energy and memory constraints of edge/IoT devices, which restrict the size of the mature-detector population that can be held and matched against in real time [9], [29].

8. Conclusion and Future Directions

The Negative Selection Algorithm has evolved substantially since its original binary, r-contiguous-bit formulation-through real-valued and variable-radius detectors, density- and boundary-aware generation, hybridisation with classical and deep learning classifiers, and distributed detector generation- and these advances have delivered genuine, measurable gains in detection accuracy and detector efficiency on static, well-characterised benchmark datasets. However, this review shows that the same body of literature has not yet produced an NSA architecture that natively and efficiently adapts its detector population to continuously and rapidly growing, non-stationary data streams of the kind generated by modern IoT and Industrial-IoT networks. There is a clear gap between the real-world, always-on IoT security monitoring scenario and the efficiency that NSA achieves by performing expensive full-batch retraining, field-based reliance on static benchmark datasets, and detector-hole formation in well-designed laboratory conditions.

Therefore, this review clearly proposes the potential for research efforts to realize NSA architectures designed to accommodate the dynamic and constantly evolving characteristics of IoT data: models for creating detectors that update coverage without full retraining in an incremental fashion, and online possibilities in detector-generation schemes; for creating lightweight detectors that are optimal for constrained edge hardware, while also considering energy-consumption; for creating concept-drift-aware matching rules that are capable of distinguishing intrusions while incorporating legitimate distribution changes; and for authentically tracking IoT traffic evolution by incorporating continuously updated streaming benchmarks. Only by recognizing and overcoming the static-to-dynamic gap can NSA-based intrusion detection fulfill its founding promise. Creating a self-adaptive, immune-inspired defense capable of keeping up with the scale and velocity of growing networks is essential for a simple technology like NSA.

References

[1] Forrest, S., Perelson, A. S., Allen, L., & Cherukuri, R. (1994). Self-nonself discrimination in a computer. Proceedings of the 1994 IEEE Computer Society Symposium on Research in Security and Privacy, 202–212.
[2] Hofmeyr, S. A., & Forrest, S. (2000). Architecture for an artificial immune system. Evolutionary Computation, 8(4), 443–473.
[3] Gonzalez, F., & Dasgupta, D. (2003). Anomaly detection using real-valued negative selection. Genetic Programming and Evolvable Machines, 4, 383–403.
[4] Gonzalez, F., Dasgupta, D., & Nino, L. F. (2003). A randomized real-valued negative selection algorithm. Proceedings of the 2nd International Conference on Artificial Immune Systems (ICARIS), 261–272.
[5] Ji, Z., & Dasgupta, D. (2004). Real-valued negative selection algorithm with variable-sized detectors. In Genetic and Evolutionary Computation – GECCO 2004 (LNCS 3102), 287–298. Springer.
[6] Ji, Z., & Dasgupta, D. (2009). V-detector: An efficient negative selection algorithm with "probably adequate" detector coverage. Information Sciences, 179(10), 1390–1406.
[7] Dasgupta, D., & Gonzalez, F. (2002). An immunity-based technique to characterize intrusions in computer networks. IEEE Transactions on Evolutionary Computation, 6(3), 1081–1088.
[8] A Novel Feature-Selection Algorithm in IoT Networks for Intrusion Detection. Sensors, 23(19), 8153 (2023).
[9] An IoT intrusion detection framework based on feature selection and large language models fine-tuning. Scientific Reports (2025).
[10] Kim, J., & Bentley, P. (2001). Evaluating negative selection in an artificial immune system for network intrusion detection. Proceedings of GECCO 2001, 1330–1337.
[11] Aickelin, U., Bentley, P., Cayzer, S., Kim, J., & McLeod, J. (2003). Danger theory: The link between AIS and IDS? Proceedings of ICARIS 2003, 147–155.
[12] Stibor, T., Timmis, J., & Eckert, C. (2005). A comparative study of real-valued negative selection to statistical anomaly detection techniques. Proceedings of ICARIS 2005, 262–275.
[13] Fouladvand, S., Osareh, A., Shadgar, B., Pavone, M., & Sharafi, S. (2017). DENSA: An effective negative selection algorithm with flexible boundaries for self-space and dynamic number of detectors. Engineering Applications of Artificial Intelligence, 62, 359–372.
[14] Yang, T., Chen, W., Li, T., et al. (2017). An antigen space density based real-value negative selection algorithm. Applied Soft Computing, 860–874.
[15] Yang, C., et al. (2020). Negative selection algorithm based on antigen density clustering. IEEE Access, 8, 44967–44975.
[16] Sun, Z., et al. (2018). An intrusion detection model for wireless sensor networks with an improved V-Detector algorithm. IEEE Sensors Journal, 18(5), 1971–1984.
[17] Wang, W., et al. (2019). Intrusion detection and security calculation in industrial cloud storage based on an improved dynamic immune algorithm. Information Sciences, 501, 543–557.
[18] A hybrid real-valued negative selection algorithm with variable-sized detectors and the k-nearest neighbors algorithm. Knowledge-Based Systems (2021).
[19] Powers, S. T., & He, J. (2006). Evolving discrete-valued anomaly detectors for a network intrusion detection system using negative selection. Proceedings of the 6th Annual Workshop on Computational Intelligence (UKCI '06).
[20] Ayara, M., Timmis, J., de Lemos, L., de Castro, R., & Duncan, R. (2002). Negative selection: How to generate detectors. Proceedings of ICARIS 2002, 89–98.
[21] VOR-NSA: A MapReduce-based negative selection algorithm for distributed detector generation, as discussed in comparative NSA-kNN hybrid studies (2021).
[22] Abid, A., Khan, M. T., & de Silva, C. W. (2017). Layered and real-valued negative selection algorithm for fault detection. Proceedings of relevant AIS/fault-detection venue.
[23] Abid, A., Khan, M. T., Haq, I. U., Anwar, S., & Iqbal, J. (2020). An improved negative selection algorithm-based fault detection method. IETE Journal of Research, 61, 1–12.
[24] An immune optimization based real-valued negative selection algorithm. Applied Intelligence (2014).
[25] Gu, M., et al. (2024). A negative selection algorithm with hypercube interface detectors for anomaly detection. Applied Soft Computing, 154, 111339.
[26] Nie, Z., Basumallik, S., Banerjee, P., & Srivastava, A. K. (2024). Intrusion detection in cyber-physical grid using incremental ML with adaptive moment estimation. IEEE Transactions on Industrial Cyber-Physical Systems.
[27] Negative selection in anomaly detection—A survey. Computer Science Review (2023).
[28] Robust Anomaly Detection for Time-series Data (negative selection, URP, and ELM-AE). arXiv preprint (2022).
[29] Kaushik, S., et al. (2022). Efficient, lightweight cyber intrusion detection system for IoT ecosystems using mI2G algorithm. Computers, 11(10), 142.
[30] Singh, K., Kaur, L., & Maini, R. (2022). A survey of intrusion detection techniques based on negative selection algorithm. International Journal of System Assurance Engineering and Management, 13(Suppl 1), 175–185.
[31] Advancing Artificial Immune System-Based Anomaly Detection (comparative study of RNSA, V-Detector, and Grid-based NSA). Springer (2025).
[32] Continuous Features Discretization for Anomaly Intrusion Detectors Generation. arXiv preprint (2014).
[33] Adversarial Concept Drift Detection under Poisoning Attacks for Robust Data Stream Mining. arXiv preprint (2020).
[34] Saurabh, P., & Verma, B. (2016). An efficient proactive artificial immune system based anomaly detection and prevention system. Expert Systems with Applications, 60, 311–320.
[35] Generating detectors from anomaly samples via negative selection for network intrusion detection. Scientific Reports (2025).
[36] Vishnoi, V., Padhee, S., & Kaur, E. (2015). Controller Performance Evaluation for Concentration Control of Isothermal Continuous Stirred Tank Reactor. http://www.ijsrp.org/research_paper_jun2012/ijsrp-June-2012-26.pdf