GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

A Dynamic Intrusion Detection System Integrating Concept Drift Detection and Incremental Learning

Zhuoqi Liao [1,*]

[1] School of Computer Science and Technology, Zhejiang Normal University, Jinhua, Zhejiang, 321004, China
* Correspondence: Zhuoqi Liao, School of Computer Science and Technology, Zhejiang Normal University, Jinhua, Zhejiang, 321004, China

Abstract: The dynamic, evolving, and inherently non-stationary characteristics of modern network environments present a fundamental challenge to traditional intrusion detection systems that rely on static learning paradigms. As network traffic patterns, system usage behaviors, and threat manifestations continuously change over time, the statistical properties underlying detection data are prone to both explicit and implicit variations, commonly described as concept drift. Such drift leads to a gradual mismatch between previously learned models and current data distributions, resulting in performance degradation, delayed responses, and reduced practical effectiveness of fixed detection mechanisms. To address these limitations and support the construction of an intelligent defense system with long-term adaptability, this study conducts a systematic theoretical investigation into dynamic intrusion detection from the perspective of learning evolution. On this basis, a unified framework is proposed that tightly integrates concept drift detection mechanisms with incremental learning strategies, enabling models to identify distributional changes in a timely manner and update their knowledge without retraining from scratch. The framework emphasizes continuity, stability, and adaptability, aiming to balance detection accuracy with computational efficiency under continuously changing conditions. By clarifying the internal relationship between drift detection and incremental model updating, this work provides a structured theoretical foundation for the development of adaptive intrusion detection systems capable of maintaining robust performance in complex and evolving network scenarios.

Keywords: concept fusion; drift monitoring; incremental learning; dynamic intrusion detection system

Received: 03 November 2025
Revised: 26 November 2025
Accepted: 05 January 2026
Published: 11 January 2026

1. Research Background

Presently, information networks have become the core infrastructure underpinning societal operations, with their security and stability being of paramount importance. The security landscape within cyberspace is exhibiting increasingly severe complexity and dynamism. Attack techniques continually evolve, progressing from large-scale scanning and exploitation of known vulnerabilities to highly customized advanced persistent threats and covert attacks leveraging artificial intelligence. Concurrently, the network environment itself undergoes constant transformation due to the deployment of new services, equipment upgrades, and shifts in user behavior patterns. This dual dynamic evolution of attack methods and network environments gives rise to a fundamental issue: the concepts underlying network traffic data are not static but undergo drift over time [1].

Against this backdrop, traditional intrusion detection systems based on static learning paradigms face a fundamental theoretical dilemma. Data during training and deployment phases are assumed to follow identical probability distributions. However, the reality of concept drift utterly undermines this assumption. Over time, static models gradually become obsolete, their discriminative capabilities steadily deteriorating. This forces security administrators to periodically collect fresh data, retrain, and redeploy

models - a resource-intensive process with inherent latency that proves cumbersome and inefficient against rapidly evolving threats.

Concept drift detection and incremental learning, as two key technological directions, each offer partial solutions. Concept drift detection aims to identify moments of significant shift in data distribution by monitoring data streams or model performance, its core value lying in providing awareness of when change occurs. Incremental learning focuses on how models can continuously learn new knowledge from incoming data without forgetting existing knowledge, its core value being the provision of adaptive updating capabilities. Consequently, constructing a dynamic intrusion detection system framework that deeply integrates concept drift detection and incremental learning theories holds urgent theoretical significance and practical value. Thoroughly exploring the theoretical foundations, coupling mechanisms, and coordination strategies for their fusion is not only the inevitable path to overcoming the limitations of current static detection models but also the key to enabling intelligent security systems to possess sustainable evolutionary capabilities. This lays a solid theoretical foundation for building a genuine cyber defense system.

2. Theoretical Overview

2.1. Theoretical Framework for Concept Drift Detection

Within the context of network intrusion detection, the core objective of the theoretical framework for concept drift detection lies in constructing a mathematical and computational framework capable of automatically identifying fundamental shifts in data generation mechanisms [2]. This approach does not merely monitor fluctuations in network traffic metrics, but rather seeks to discern the underlying rules driving traffic generation.

Fundamentally, concept drift detection theory follows two parallel logical pathways. The first involves direct monitoring based on data distribution. This approach assumes that during conceptually stable periods, observed network traffic characteristics or their latent representations fluctuate around a stable statistical distribution. A significant deviation from this underlying distribution signals potential new attack patterns or alterations in normal network behavior. The second approach involves indirect inference based on model performance. This theory employs the detection model itself as a probe, continuously evaluating its classification performance on the latest data. A sustained decline in model performance, particularly abrupt deterioration in specific categories, is regarded as compelling evidence of concept drift. This indicates that the decision boundaries relied upon by the model are no longer applicable to the current environment.

Statistical process control theory treats data streams or model error rates as time-varying sequences. By employing tools such as control charts, it establishes confidence intervals or control limits. Should the cumulative deviation or instantaneous value of the sequence exceed predefined theoretical thresholds, a drift alert is triggered. Its advantages lie in its formal simplicity and computational efficiency, rendering it particularly suited for detecting sudden, abrupt shifts-such as the abrupt emergence of novel attacks within a network [3].

Hypothesis testing formalizes drift detection as a rigorous statistical hypothesis testing problem. Its core involves constructing two data windows and proposing a null hypothesis that both windows originate from the same distribution. By calculating statistics such as the Kolmogorov-Smirnov test, chi-squared test, or distance-based metrics, it assesses the confidence level for rejecting the null hypothesis, thereby conferring statistical interpretability upon detection outcomes.

Ensemble difference metric theory employs machine learning models themselves as sensors for distributional divergence. A typical approach involves using two or more learners to monitor performance differences between reference and new data, or directly training a classifier to distinguish between old and new data. The discriminative

capability of this classifier reflects the degree of distributional divergence between the two datasets, enabling capture of nonlinear changes in high-dimensional, complex feature spaces and demonstrating good sensitivity to gradual drift.

2.2. Theoretical Framework of Incremental Learning

From a formal definition perspective, incremental learning addresses an infinite or extremely lengthy data sequence. Upon receiving a new data batch at each time step, the model must complete its update immediately or within a finite timeframe, preparing to process the subsequent batch. Consequently, the core of incremental learning theory lies in investigating how to achieve efficient and stable knowledge accumulation under conditions where partial historical information is either missing or irreproducible [4].

The core theoretical challenge confronting this framework is the stability-plasticity dilemma. Stability denotes a model's capacity to retain acquired knowledge, while plasticity signifies its ability to adapt to new information. Excessive stability leads to rigidity, rendering the model incapable of learning emerging attack patterns; conversely, excessive plasticity induces forgetting, causing previously identified attacks to be misclassified as normal.

Consequently, incremental learning theory has developed several core solution paradigms. Firstly, regularization-based approaches impose constraints on parameter updates by introducing additional regularization terms into the loss function. Their theoretical core lies in identifying and preserving parameters crucial for old tasks. Secondly, dynamic architecture-based approaches permit the model's structure to expand or adjust as new tasks or knowledge emerge.

Thirdly, replay-based approaches. Systems maintain a finite-capacity memory buffer storing representative samples of past data or their features. When learning new data, samples from this buffer are mixed with fresh data for training, periodically replaying historical patterns.

3. Theoretical Modelling of Framework Components for Dynamic Intrusion Detection

3.1. Formalization of Intrusion Detection Problems in Dynamic Environments

The theoretical modelling of traditional intrusion detection systems rests upon an idealized assumption that the world is static. At a specific point in time, sufficient network traffic data is collected to represent all future possible scenarios, from which a fixed, optimal discrimination rule is learned. Once established, these rules are deployed as static filters, expected to remain effective indefinitely. Consequently, the primary task in formalizing intrusion detection for dynamic environments is to abandon the assumption of a static, closed world entirely, instead acknowledging and embracing the reality of a dynamic, open world. Within this new theoretical paradigm, intrusion detection ceases to be a mathematical optimization problem solvable in a single instance, becoming instead a continuous, never-ending process of adaptation.

Attack patterns exhibit non-stationarity. Novel attack vectors emerge from unforeseen directions, while established methods undergo camouflage and metamorphosis to evade detection. The set of anomalies or attack categories a system must recognize is not a static catalogue, but a dynamic collection where new members continually emerge and old ones may reappear in disguised forms. Natural drift in normal behavior baselines corresponds to deliberate alterations by attackers. The protected network environment itself undergoes legitimate, organic evolution. Deploying new applications, updating software versions, shifting user behavior patterns, and even adjusting network architecture all cause gradual yet persistent changes in the statistical characteristics of legitimate traffic.

3.2. Constructing Models for Network Flow Characteristics

The high-dimensionality, strong correlations, and extreme imbalance inherent in network flow data present the primary theoretical challenge for drift detection. The feature space dimensions of network connections or flow records are typically extremely high, with complex nonlinear dependencies between features. More critically, the volume of normal traffic differs by several orders of magnitude from that of attack traffic. This extreme class imbalance renders drift detection methods based on global statistics highly susceptible to failure. Subtle distributional shifts occurring only in a minority of attack samples are drowned out by the dominant signal of normal traffic, leading to detection delays or complete false negatives. Consequently, theoretical models must possess the capability to capture the evolution of microscopic anomaly patterns within the macrocosm of data flow, maintaining sensitivity unaffected by the scarcity of a class.

Moreover, the patterns of concept drift within network environments exhibit unique complexity and adversarial properties, demanding detection models that transcend mere identification of simple distribution shifts. The dynamism of cybersecurity stems not only from the natural evolution of technology but also from the ongoing strategic interplay between attackers and defenders. To evade detection, attackers deliberately execute evasive attacks, causing the traffic patterns they generate to drift slowly and purposefully from known anomaly zones towards normal regions.

Adversarial concept drift and benign concept drift arising from business updates may appear similar superficially, yet their nature is fundamentally distinct. The former constitutes deliberate, targeted evasion by attackers, aimed at progressively blurring classification boundaries; the latter represents spontaneous, purposeless evolution of the environment. Consequently, theoretical models must not only detect changes but also possess the preliminary capability to distinguish the 'intent' behind them. This differentiation is crucial for subsequent adaptive decision-making: Malicious drift necessitates decisive, reinforced learning to fortify defenses; benign drift may require more cautious, incremental adjustments to prevent the model from being disrupted by irrelevant noise.

To address these dual challenges, constructing a robust theoretical model must adhere to core design principles of hierarchical decoupling and collaborative monitoring. It should be deconstructed into distinct layers and groups, each equipped with sensitive detection submodules. Specifically, the model should undergo collaborative modelling across at least three logical layers:

First, at the macro-level, the model must monitor the statistical distribution of fundamental network traffic characteristics, such as the evolution of aggregate metrics including overall traffic volume, protocol-port distribution, and connection duration. This layer is sensitive to large-scale, global changes-such as traffic surges triggered by new service launches or alterations in connection request patterns caused by novel scanning tools. The theoretical basis lies in the fact that even if attack traffic constitutes a minor proportion, certain attacks (such as DDoS) or large-scale scans will still leave detectable anomalous perturbations in macro-level statistics.

Second, at the micro-level, the model must delve into the internal structure of traffic, focusing on the distributional stability of specific behavioral clusters or potential subspaces. This requires leveraging online clustering, community detection, or representation learning techniques to dynamically partition traffic into groups exhibiting similar behavioral patterns. Drift detection then operates within each independent cluster-for instance, monitoring the behavioral profile of a 'normal web server cluster' or tracking the evolutionary characteristics of an anomaly cluster like 'suspected C&C communications'. The core theoretical advantage of this approach lies in transforming a global, imbalanced detection problem into a series of locally balanced or at least more manageable subproblems. This significantly enhances the model's sensitivity to subtle variations in rare attack patterns.

Thirdly, at the correlation and sequence level, the model should transcend the assumption of independent and identically distributed events, focusing instead on the evolution of temporal dependencies and causal relationships between network incidents. The essence of an attack often manifests in anomalous patterns of interconnected events (such as a successful intrusion potentially encompassing correlated activities across multiple stages: scanning, exploitation, privilege escalation, and data exfiltration). Theoretical models require integration of monitoring for time series, graph structures, or state transition probabilities. For instance, monitoring changes in the probability of behavioral sequences from specific source IPs to target service ports may detect tactical adjustments by Advanced Persistent Threats (APTs) earlier than monitoring isolated connection characteristics.

Finally, a comprehensive theoretical model necessitates a meta-evaluation and decision fusion layer. This layer receives and synthesizes output signals from detection submodules across the aforementioned tiers (potentially including 'macro-metric anomalies,' 'shifts within specific anomaly clusters,' or 'ruptures in critical correlation patterns'). It must determine whether these signals indicate a single fundamental conceptual drift event based on predefined rules or learned experience, while assessing their overall confidence level and threat severity. Through this multi-perspective, multi-evidence fusion and validation, the system can more reliably trigger adaptive learning mechanisms while providing richer contextual information. For instance: 'Current changes primarily centre on the gradual expansion of normal operations within category X, accompanied by minor evasion signs of known attacks within category Y.' This guides the incremental learning module to adopt precisely matched, refined update strategies.

3.3. Theoretical Modelling of Adaptive Coordination Mechanisms

The theoretical modelling of adaptive coordination mechanisms aims to construct an intelligent dispatch centre endowed with decision-making and regulatory capabilities. By receiving raw perceptual signals regarding environmental changes from the drift detection module, it transforms these through analysis, fusion, and inference into a series of executable, measurable control commands. This precisely guides the incremental learning module to execute the model update strategy most suited to the current context. Generally, this process comprises three component steps.

The Context Parsing and Drift Semantic Understanding component processes the multidimensional signals output by the drift detection module. Its theoretical function involves signal fusion and semantic enhancement, providing an initial qualitative characterization of the drift's nature, which serves as the primary basis for subsequent decision-making.

The Strategy Mapping and Knowledge Base Management component serves as the coordinator's "decision-making brain". It stores mappings between different drift scenarios and their corresponding incremental learning strategy combinations. This mapping is not static but undergoes self-optimization through a meta-learning loop: the system evaluates the long-term effects after each strategy execution and adjusts the preference of the strategy mapping accordingly. This enables the coordination mechanism to learn from historical decisions, continuously refining the effectiveness of its scheduling strategies.

The execution control and feedback loop component is responsible for translating decisions into specific control parameters, issuing them to the incremental learning module for execution, and establishing a tightly coupled feedback loop. It not only initiates the learning process but also monitors resource consumption and immediate effects during learning, intervening dynamically when necessary. Simultaneously, it feeds back the model's preliminary performance on new data following learning completion to the coordinator. This data is utilized to update the system state and evaluate the

effectiveness of the current decision, thereby completing a full perception-decision-execution-evaluation adaptive cycle.

The dynamic decision-making process further necessitates the deep integration of resource awareness with multi-objective optimization theory. In real-world deployments, computational power, memory, and response time are finite resources. A theoretically optimal update strategy may prove impracticable under resource constraints. Consequently, the coordinator's decision model must incorporate a resource-constrained optimizer. For instance, when multiple drift signals are detected concurrently amidst resource scarcity, the coordinator must priorities resource allocation by ranking each drift based on its threat level, confidence score, and potential impact on the system's overall security posture. This necessitates that the coordinator not only comprehends security but also understands the system's operational state, making real-time Pareto-optimal trade-offs between objectives such as detection performance, response velocity, and resource efficiency.

The theoretical completeness of adaptive coordination mechanisms is further demonstrated through their design for interpretability and intervenability. As an autonomous decision-making hub, its logic must not constitute a black box. Theoretical models should output rational explanations for their decisions, such as: 'Strategy Y was selected for rapid model reconstruction due to detecting high-confidence mutations in feature cluster X, coupled with sufficient current system memory.' This explainability is crucial for establishing security analysts' trust in the system, while also providing a clear interface for analysts to intervene manually when necessary (e.g., vetoing a high-risk update or injecting domain-specific prior knowledge). Through this hybrid human-machine collaborative decision-making model, the system's autonomy is combined with human expertise and overall control capabilities, forming a more reliable and robust dynamic defense system.

4. Conclusions

This study systematically constructs a theoretical framework integrating concept drift detection with incremental learning to address core challenges in intrusion detection within dynamic network environments. The research first elucidates the intrinsic mechanism by which traditional static detection models fail due to concept drift, thereby demonstrating the necessity and superiority of deeply integrating environmental awareness (drift detection) with self-renewal (incremental learning). By establishing a closed-loop 'detection-adaptation' theoretical model, this study provides a comprehensive paradigm for enabling autonomous, continuous evolution in dynamic intrusion detection systems.

Incremental learning, whilst designed for efficient updates, still requires its computational and storage overhead to be strictly constrained in resource-constrained edge devices or high-throughput backbone network environments. Furthermore, after months or even years of continuous evolution, a model's internal knowledge structure may become exceptionally complex, potentially giving rise to 'cognitive debt'. Designing lightweight algorithms and periodic knowledge distillation and reconstruction mechanisms to maintain model simplicity and efficiency without sacrificing accumulated discriminative capabilities presents a theoretical and practical challenge for ensuring long-term robust system operation.

Traditional intrusion detection evaluations rely on static dataset partitioning and fixed performance metrics, which prove wholly inadequate for assessing dynamic adaptive systems. Future developments necessitate the establishment of dynamic benchmarking platforms capable of simulating concept drift, attack evolution, and shifts in normal behavior. These should employ metrics that better reflect long-term efficacy, such as cumulative detection gain, average adaptation recovery time, and forgetting rate,

to scientifically measure a system's intelligence and robustness throughout its entire lifecycle.

The theoretical significance of this research lies in shifting the focus of intrusion detection studies from pursuing static optimal solutions towards constructing viable adaptive learning systems. It not only provides systematic theoretical tools for addressing concept drift but, more importantly, outlines a feasible theoretical blueprint for the 'meta-learning' capability-the ability to 'learn how to learn'-that future intelligent security systems should possess. Future research may build upon this framework to further explore theoretical mechanisms for handling complex drift patterns such as periodic or context-dependent shifts, investigate theories for rapid adaptation to novel threats under sparse sample conditions, and strive to enhance the interpretability of the entire adaptive process. Ultimately, this will propel dynamic intrusion detection theory towards a more mature and comprehensive stage of development.

References

1. Z. Ouyang, Y. Gao, Z. Zhao, and T. Wang, "Study on the classification of data streams with concept drift," In *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, July, 2011, pp. 1673-1677. doi: 10.1109/fskd.2011.6019889
2. G. Gong-De, L. Nan, and C. Li-Fei, "Classification for concept-drifting data streams with limited amount of labeled data," In *International conference on automatic control and artificial intelligence (ACAI 2012)*, March, 2012, pp. 638-644.
3. N. Liu, and J. Zhao, "Streaming data classification based on hierarchical concept drift and online ensemble," *IEEE Access*, vol. 11, pp. 126040-126051, 2023. doi: 10.1109/access.2023.3327637
4. Z. Lin, and D. Hongle, "Research on SDN intrusion detection based on online ensemble learning algorithm," In *2020 International Conference on Networking and Network Applications (NaNA)*, December, 2020, pp. 114-118. doi: 10.1109/nana51271.2020.00027