CoSALT: A Resource-Rational Model of Context-Adaptive Thresholds for Detecting Sparse Super-Spreaders

Zeyu Yang[1,2], **Shu Li**[1,2,*], **Meijie Du**[1,2], **Zhou Zhou**[1,2] **& Yong Wang**[3]

[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]National Computer Network Emergency Response Technical Team/Coordination Center of China

[*]Corresponding author: lishu@iie.ac.cn

Abstract

Detecting rare, meaningful signals in non-stationary, heavy-tailed environments challenges biological agents and artificial monitors alike. In network security, analysts must identify sparse super-spreaders within vast benign traffic; as baselines drift, fixed thresholds either overload attention with false alarms or miss emerging threats. We present CoSALT, a resource-rational anomaly detector that frames criterion setting as a Minimum Description Length (MDL) problem. CoSALT adaptively partitions observations into a compressible bulk and a salient tail, while an explicit cognitive selection cost penalizes attending to too many candidates. It further uses spline-based quantile regression to learn context-dependent sparsity expectations. On real backbone-traffic benchmarks, CoSALT improves F1 relative to strong baselines. In a behavioral study, its selection boundaries closely match human visual judgments, especially under explicit budget constraints. These findings are consistent with a resource-rational account of adaptive detection in dynamic, resource-constrained environments, while leaving room for alternative cognitive mechanisms.

Keywords: anomaly detection; resource rationality; minimum description length; adaptive thresholds

Introduction

Many real-world decisions require setting a detection criterion under shifting contexts: what counts as "large," "rare," or "suspicious" depends entirely on the current distribution of observations. Cognitive science offers a rich theoretical framework for this problem, explaining criterion shifts as adaptive trade-offs between accuracy and limited resources (Bhui et al., 2021; Lieder & Griffiths, 2020). Efficient coding theories (Młynarski & Hermundstad, 2021) and signal detection models posit that organisms adapt their sensitivity to the prevailing statistics of their environment to maximize information transmission under metabolic or attentional constraints. In this paper, we use a high-stakes network security task as an ecological testbed to operationalize and test these principles.

The domain of network telemetry presents a challenge strikingly similar to visual search in natural scenes (Figure 1). Analysts must identify sparse super-spreaders—sources contacting many destinations with low frequency, characteristic of stealthy attacks—within massive streams of data. Two properties make this environment cognitively demanding. First, the data is heavy-tailed: a power-law distribution means "extreme" values are common, blurring the line between salient outliers and benign heavy hitters. Second, the environment is non-stationary: traffic patterns drift due to diurnal cycles and workload shifts. In such dynamic settings, fixed thresholds fail: they either flood the analyst with false alarms or miss emerging threats when the background noise floor rises.

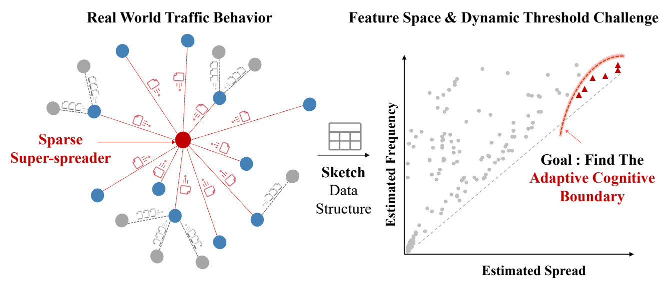

Figure 1: Detecting "Sparse Super-spreaders" as a cognitive search task. The transformation from traffic topology (Left) to feature space (Right) necessitates an Adaptive Cognitive Boundary (red dashed line). This boundary must dynamically separate stealthy anomalies (red triangles) from the heavy-tailed background noise (grey dots).

We approach this problem as a cognitive challenge: *How should a bounded agent (relying on compact sketch data structures) carve a moving, heavy-tailed distribution into "normal" and "noteworthy" components while paying a cost for every item flagged for attention?* This framing synthesizes three cognitive themes: (i) Simplicity: The preference for representations that compress the "bulk" of experience; (ii) Context Adaptation: The rapid renormalization of judgment criteria based on recent stimulus history; (iii) Resource Rationality: The optimization of behavioral policies subject to computational and attentional limits.

Contributions. We introduce CoSALT (Co-adaptive Simplicity-and-Attention Limited Thresholding), a model that instantiates these principles for streaming anomaly detection. CoSALT makes three contributions:

1. A resource-rational criterion: We derive a per-window threshold by minimizing a Minimum Description Length (MDL) objective. This objective explicitly balances the fit of the bulk distribution against a *selection cost*—a penalty term representing the cognitive load of attending to candidate items.
2. Context-sensitive structure learning: Instead of assuming a fixed relationship between spread and frequency, CoSALT learns a local "normative boundary" via spline-based

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

quantile regression, flagging only those sources that are anomalous relative to their structural context.

3. Human-model alignment: In a behavioral experiment ($N = 60$), we show that CoSALT's detection boundaries closely track human visual intuition in detecting anomalies within scatter plots. Crucially, human alignment improves when participants are placed under explicit "budget constraints," supporting the validity of our resource-rational selection cost.

Together, CoSALT serves a dual purpose: it is a robust, sketch-compatible tool for network defense, and a computational proof-of-concept for how rational agents can maintain sensitivity in drifting, heavy-tailed environments.

Related Work

Adaptive criteria and resource-rationality. Work on adaptive decision criteria argues that thresholds shift with recent stimulus statistics and resource limits, as captured by resource-rational analyses and efficient/adaptive coding accounts (Bhui et al., 2021; Griffiths et al., 2015; Lieder & Griffiths, 2020; Młynarski & Hermundstad, 2021). Classic signal detection and adaptation-level perspectives similarly predict criterion shifts under changing base rates and costs (Green & Swets, 1966; Helson, 1964). CoSALT instantiates these ideas in a streaming monitoring setting by treating alert volume as a first-class resource constraint rather than a downstream tuning knob.

MDL and simplicity-based accounts of cognition. Simplicity principles formalize the trade-off between fit and complexity via Minimum Description Length (MDL) (Chater & Vitányi, 2003; Rissanen, 1978), and recent work has used MDL to explain cognitive phenomena such as dual-process behavior (Moskovitz et al., 2024). CoSALT adopts an MDL objective to select a per-window bulk–tail boundary, but extends standard MDL model selection by adding an explicit *selection/attention cost* for forwarding n_T candidates. This connects to practical workload and queue-aware thresholding schemes that explicitly optimize for prioritized review under constrained investigator capacity (Butvinik et al., 2026).

Streaming anomaly detection under drift and tail risk. A large body of work addresses anomaly detection in non-stationary streams using lightweight, tail-sensitive methods such as Extreme Value Theory (EVT) thresholding (Li et al., 2021; Siffer et al., 2017) and tree-based streaming detectors like Robust Random Cut Forests (Guha et al., 2016). Practical monitoring systems often require adaptive quantile/threshold tracking in the presence of drift (Arandjelović et al., 2015; Isaac & Sharma, 2024; X. Yang et al., 2025). Recent evaluation and benchmarking studies emphasize that performance depends strongly on properties such as locality, memory span, and the intensity/frequency of concept drift (Cao et al., 2025; Iglesias Vázquez et al., 2023). In security settings, drift is often framed as a normality shift; systems such as CADE and OWAD explicitly detect and adapt to such shifts for deployed anomaly detectors (Han et al., 2023; L. Yang et al., 2021). CoSALT targets the same operational regime but focuses on closed-form, per-window criterion setting from sketch summaries, with an explicit cognitive workload term.

Boundary learning and human visual anomaly judgments. CoSALT's Stage 2 casts sparsity as deviation below a learned conditional boundary rather than deviation from the mean; quantile regression provides a principled way to estimate such boundaries (Koenker & Bassett, 1978). This boundary-centric view also aligns with findings from graphical perception and scatterplot judgment, where observers segment dense clouds from tails and treat outliers relative to local context (Ciccione et al., 2023; Cleveland & McGill, 1984).

The CoSALT Model

Problem Formulation

We consider a streaming monitoring setting where, for each time window t, a sketch summary provides estimates for N active source keys. Let $\hat{S}_i(t)$ denote the estimated distinct destination count (spread) and $\hat{F}_i(t)$ the total packet count for source i. Our goal is to identify sparse super-spreaders (Z. Yang et al., 2025): sources with high spread $\hat{S}_i$ but anomalously low packet frequency relative to that spread.

CoSALT approaches this not merely as a statistical filtering task, but as a *resource-rational attention allocation* problem (Griffiths et al., 2015). The system must decide which stimuli (sources) warrant the cost of downstream processing. This requires two cognitive adaptations: (1) learning the distinction between "background noise" and "salient signals" (Stage 1), and (2) learning the context-dependent relationship between features to detect structural anomalies (Stage 2).

To stabilize variance and operate on a perceptually relevant scale, we define the log-transformed spread and frequency:

$$x_i = \log(\hat{S}_i + 1) \tag{1}$$

$$f_i = \log(\hat{F}_i + 1) \tag{2}$$

The monotone transform in Eq. 1 compresses the heavy tail of the distribution, akin to the logarithmic coding of stimulus intensity in biological sensory systems.

Stage 1: Adaptive Thresholding via MDL

A core challenge in anomaly detection is distinguishing true anomalies from benign "heavy-tailed" events. Simple models often assume exponential decay for the tail. However, network traffic, like many natural signals, follows power-law distributions. A cognitive system that assumes a light tail when the environment is heavy-tailed will suffer from excessive false alarms—a failure of adaptation. To address this, we employ a flexible model selection framework.

The MDL Objective. We cast threshold selection as minimizing the description length of the data code. We seek a boundary τ_x that separates a predictable "bulk" from a surprising "tail." The objective minimizes:

$$\begin{aligned} \mathrm{DL}(\tau_x, M_{\mathrm{tail}}) = {} & \mathrm{NLL}_{\mathrm{bulk}} + \mathrm{NLL}_{\mathrm{tail}}(M_{\mathrm{tail}}) \\ & + \mathrm{Pen}_{\mathrm{params}} + \mathrm{Pen}_{\mathrm{select}}. \end{aligned} \tag{3}$$

For comparability across windows with varying N, we minimize the average description length DL/N; all terms are measured in the same log unit.

Bulk and Tail Models. The bulk ($x \leq \tau_x$) is modeled as Gaussian, representing the noise floor:

$$x \sim \mathcal{N}(\mu, \sigma^2). \tag{4}$$

For the tail exceedances $u = x - \tau_x > 0$, we introduce a competition between two hypotheses to balance flexibility and parsimony:

- **Exponential (Exp):** This assumes a standard "thin" tail:

$$p(u) = \lambda e^{-\lambda u}. \tag{5}$$

- **Generalized Pareto (GPD):** A more robust model derived from Extreme Value Theory, capable of capturing polynomial decay:

$$p(u) = \frac{1}{\sigma_u}\left(1 + \frac{\xi u}{\sigma_u}\right)^{-(1+1/\xi)}. \tag{6}$$

By allowing the system to select $M_{\text{tail}} \in \{\text{Exp}, \text{GPD}\}$ based on the data, CoSALT dynamically adapts its internal model of the environment, avoiding the cognitive bias of underestimating extreme events. To prevent fragile tail fits, we only consider tail modeling when the number of exceedances n_T is sufficient (we use $n_T \geq 20$); otherwise we default to the Exponential tail.

Complexity and Attention Costs. The penalty terms enforce Occam's razor. $\text{Pen}_{\text{params}}$ penalizes the extra parameter in GPD (ξ), ensuring it is chosen only when the data strongly deviates from exponential. Crucially, we include a *selection cost* corresponding to the code-length of specifying which n_T items (out of N) are sent to downstream attention:

$$\begin{aligned} \text{Pen}_{\text{select}} &= \log \binom{N}{n_T} \\ &\approx N\,H\!\left(\tfrac{n_T}{N}\right) \\ &= n_T \log \frac{N}{n_T} + (N - n_T) \log \frac{N}{N - n_T}. \end{aligned} \tag{7}$$

This term represents the cognitive cost of attending to n_T items. As the environment becomes noisier, this cost naturally shifts the optimal threshold upward, regulating the rate of false positives.

Stage 2: Context-Sensitive Sparsity via Quantile Regression

Sources exceeding $\theta_S(t)$ form a candidate set C. To detect sparsity, we must compare a source's frequency f_i against an expected baseline given its spread x_i.

Standard approaches (e.g., linear regression) estimate the average relationship. However, in both neural coding and traffic analysis, anomalies are defined by deviations from a boundary, not the mean. Furthermore, the data exhibits heteroscedasticity: variance is not constant but scales with magnitude. A linear model fails to capture the complex, non-linear saturation effects (e.g., CDN behaviors) at high spread. To robustly identify the lower bound of normal behavior, we upgrade from mean regression to **Quantile Regression**.

We model the q-th quantile (e.g., $q = 0.1$, representing the lower normative boundary) of frequency f_i given spread x_i. We use B-splines to accommodate non-linearities without assuming a specific functional form:

$$\hat{g}_q(x) = \sum_{j=1}^{K} \beta_j B_j(x). \tag{8}$$

In practice, we use $K = 6$ basis functions with internal knots placed at empirical quantiles of $\{x_i\}_{i \in C}$, and we add a mild roughness penalty to stabilize fits:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i \in C} \rho_q\big(f_i - \hat{g}_q(x_i)\big) + \lambda \|\mathbf{D}\boldsymbol{\beta}\|_2^2, \tag{9}$$

where $\rho_q(e) = e(q - \mathbb{I}(e < 0))$, $\mathbf{D}$ is a second-difference operator, and we set $\lambda = 10^{-2}$. When $|C|$ is small ($|C| < 50$), we fall back to a linear quantile regression (i.e., $K = 2$) to avoid overfitting.

Sparsity Scoring. The sparsity score z_i is defined as the deviation below this learned adaptive boundary:

$$z_i = \hat{g}_q(x_i) - f_i. \tag{10}$$

A large positive z_i implies the source is significantly sparser than the learned population norm for that specific spread level. This mirrors *predictive coding*, where signals are evaluated based on their deviation from top-down predictions. Finally, we apply the same MDL cut-selection (Eq. 3) to these scores $\{z_i\}$ to set a rigorous detection threshold τ_z.

Adaptation, Surprisal Ranking, and Complexity

To reduce jitter under gradual drift, we maintain an exponential moving average of the learned thresholds:

$$\tau^{(t)} \leftarrow (1 - \beta)\tau^{(t-1)} + \beta\tau_t^*, \tag{11}$$

where τ_t^* is the current window's optimum.

Flagged sources are primarily ranked by the surprisal of their sparsity score z_i under the tail model. For an exceedance $u_z = z_i - \tau_z > 0$, the surprisal scales with u_z: $-\log p_{\text{tail}}(u_z) \propto u_z$. Ranking by the sparsity margin u_z thus prioritizes sources that deviate most strongly from the expected spread–frequency baseline, aligning analyst attention with the most structurally anomalous events.

CoSALT is designed for high-throughput stream processing. The Stage 1 spread thresholding involves sorting and scanning, with complexity $O(N \log N)$ or $O(N)$ if using approximate quantiles. Stage 2 operates only on the small candidate set C (where $|C| \ll N$). Spline regression on $|C|$ points is $O(|C|K^2)$, which is negligible. Thus, the total per-window definition cost scales linearly with N and is compatible with real-time sketch-based monitoring.

Evaluation

To evaluate CoSALT as both an engineering method and a cognitively motivated model of adaptive thresholding, we conducted two complementary studies. **Experiment I** evaluates detection performance against strong baselines using real-world network traces with realistic attack traffic. **Experiment II** is a behavioral experiment comparing CoSALT's selection boundaries to human judgments in a visual anomaly detection task, testing whether the MDL-based criterion aligns with human intuitions about "noteworthy" observations.

Experiment I: Detection Performance on Real Network Traffic

Dataset and Setup. We merged CAIDA Witty Worm scanning traffic (CAIDA, 2004) into a 20-minute MAWI background trace (MAWI Working Group, 2016). The MAWI data introduces challenges like heavy tails and non-stationarity. By varying worm prevalence from 0.1% to 1.0% across 1-minute windows, we evaluate detection under controlled mixing, specifically targeting the distributional shift during peak volume (windows 8–12).

Baselines. We compare CoSALT against five strategies: (1) **Fixed-99**: a static rule flagging sources above the 99th spread percentile with sparsity ratio < 3; (2) **Top-K** ($K = 100$): which reports the top sparse spreaders under a fixed budget; (3) **SPOT** (Siffer et al., 2017): a streaming EVT-based detector with exponential forgetting to handle non-stationarity; (4) **RRCF** (Guha et al., 2016): Robust Random Cut Forest, a tree-based streaming anomaly detector; and (5) **OIF**(Leveni et al., 2024): Online Isolation Forest, an online adaptation of Isolation Forest designed for streaming anomaly detection under concept drift.

Metrics. We report Precision, Recall, F1-score, FPR, along with **Threshold Stability** (σ_τ): the standard deviation of the adaptive threshold across windows.

Table 1: Detection Performance on MAWI Background with CAIDA Witty Worm Traffic.

Method	Precision	Recall	F1	FPR	σ_τ
Fixed-99	0.52 ± 0.19	**0.97 ± 0.02**	0.68 ± 0.05	0.042	—
Top-K	**0.94 ± 0.04**	0.52 ± 0.14	0.68 ± 0.04	**0.006**	—
SPOT	0.79 ± 0.11	0.88 ± 0.07	0.83 ± 0.06	0.024	0.31
RRCF	0.82 ± 0.09	0.86 ± 0.06	0.84 ± 0.03	0.021	0.18
OIF	0.85 ± 0.07	0.89 ± 0.05	0.87 ± 0.02	0.018	0.15
CoSALT	0.92 ± 0.04	0.94 ± 0.03	**0.93 ± 0.01**	0.009	**0.11**

Overall Performance As shown in Table 1, CoSALT achieves the highest F1-score of 0.93, improving by 6 percentage points over the best-performing baseline among evaluated methods (Online IForest, F1=0.87). Importantly, the improvement is not achieved by "buying" recall with excessive false alarms: CoSALT reduces FPR from 0.042 (Fixed-99) to 0.009, a ∼4.7× reduction in benign sources forwarded for review. This gain stems from CoSALT's balanced precision–recall trade-off: Fixed-99 attains near-perfect recall of 0.97 but low precision of 0.58, leading to 4.2% of benign sources being incorrectly flagged—an analyst workload that scales poorly in production deployments. Top-K illustrates the opposite extreme: it enforces a hard budget and thus achieves very low FPR (0.006) and high precision (0.94), but misses nearly half of attacks (recall 0.52), making it brittle when prevalence rises. Among adaptive methods, Online IForest outperforms RRCF (F1=0.87 vs. 0.84) by leveraging incremental tree updates, but still falls short of CoSALT's MDL-based boundary optimization.

Adaptation vs. Static Rules. The performance gap between adaptive methods (SPOT, RRCF, Online IForest, CoSALT) and static baselines (Fixed-99, Top-K) widens during periods of distributional shift. Figure 2 reveals that Fixed-99's F1 drops below 0.62 during peak traffic windows (8–12), where static percentile cutoffs become miscalibrated as the bulk distribution shifts upward (raising "normal" spread). In contrast, CoSALT's per-window MDL optimization maintains F1 $>$ 0.90 throughout, effectively re-anchoring the decision boundary to the current bulk–tail transition rather than an absolute percentile. Online IForest shows improved stability over RRCF due to its incremental update mechanism, but still exhibits performance dips during peak windows. Importantly, CoSALT also yields the lowest threshold variability ($\sigma_\tau = 0.11$) among adaptive methods, suggesting smoother criterion updates and fewer abrupt swings in alert volume—consistent with adaptation-level theory predictions that optimal criteria should track "natural" distributional boundaries.

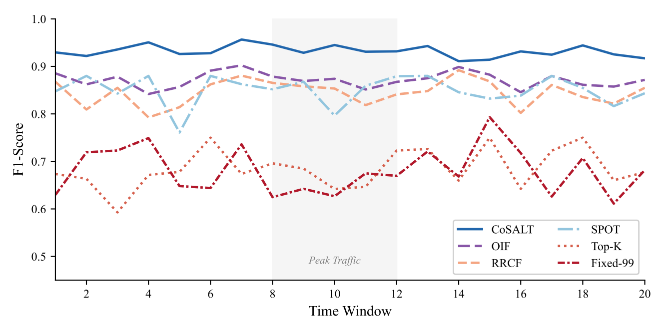

Figure 2: F1-score trajectories across 20 time windows.

Ablation Study To validate the contribution of each cognitive mechanism integrated into CoSALT—specifically the resource-rational selection cost, the adaptive environmental model, and the structural boundary learning—we conducted an ablation study. We compared the full CoSALT model against three degraded variants:

1. **w/o Selection Cost:** Removes the $\mathrm{Pen}_{\mathrm{select}}$ term from Eq. 3. This represents an "unbounded" agent that optimizes only for representational fidelity (NLL) without regard for attentional load.
2. **Fixed Exp Tail:** Disables the model selection for the tail distribution, forcing the agent to assume a thin-tailed Exponential distribution (M_{tail} = Exp). This represents an agent with a rigid internal model of the environment.

3. **Linear Mean Reg:** Replaces the Stage 2 Quantile Regression with standard linear regression, defining sparsity based on deviation from the mean rather than the normative boundary.

The Cost of Unbounded Attention. Removing the selection cost caused the most significant performance degradation (Table 2), with the F1 score dropping to 0.76. Without $\text{Pen}_{\text{select}}$, the MDL objective favored complex models that "explained" noise as signal, resulting in a near-perfect Recall of 0.98 but a catastrophic drop in Precision to 0.62. This quadrupled the False Positive Rate to a value of 0.038 compared to the full model. This finding supports the practical importance of selection regularization in noisy streams, but does not by itself uniquely identify the cognitive form of the underlying cost function.

Table 2: Ablation Study Results.

Variant	Precision	Recall	F1	FPR
CoSALT (Full)	**0.91**	**0.94**	**0.92**	**0.009**
w/o Selection Cost	0.62	0.98	0.76	0.038
Fixed Exp Tail	0.74	0.95	0.83	0.027
Linear Mean Reg.	0.88	0.79	0.83	0.012

Rigid Priors Fail in Heavy-Tailed Contexts. Forcing an Exponential tail assumption reduced Precision to 0.74. Because network traffic follows a power law, an agent assuming exponential decay systematically underestimates the probability of benign heavy hitters, misclassifying them as anomalies. The full CoSALT model, which dynamically selected the GPD model in 85% of the windows, successfully distinguished between "rare" (anomaly) and merely "large" (benign tail), validating the need for adaptive environmental modeling.

Boundary vs. Average Learning. Replacing Quantile Regression with Linear Mean Regression hurt Recall, which dropped to 0.79. The linear model captures the average relationship between spread and frequency but fails to identify the lower bound of the distribution where sparse super-spreaders reside. By modeling the boundary, CoSALT effectively separates structural anomalies from natural variance, confirming that anomaly detection is better framed as boundary learning than mean estimation.

Experiment II: Human–Model Alignment

Does CoSALT's MDL-based criterion capture something about human judgment? We addressed this question by comparing algorithmic and human selections in a visual anomaly detection task, with particular interest in whether explicit resource constraints shift human thresholds toward MDL-optimal values.

Participants & Stimuli. Sixty adults (age $M = 26.4$, $SD = 5.2$; 30 female) participated. Stimuli comprised 24 scatterplots generated from MAWI traffic windows, each displaying approximately 500 sources on log-transformed axes (Spread × Frequency). Anomaly density varied from 0% to 8% across stimuli.

Procedure. Participants used a lasso tool to "select sources that stand out from the normal pattern." In a between-subjects design, half performed the task without constraints (**Unbounded**), while the other half received a **Budget Constraint**: they were told that "the team can verify only a limited number of targets; select only the most salient ones."

Metrics and Analysis. We computed three alignment measures for each participant–algorithm pair: (1) **Jaccard Index (JI)** for point set overlap; (2) **Boundary Agreement (ρ_b)** as the Spearman correlation between the participant's minimum selected x-value and τ_x; and (3) **Sparsity Sensitivity (ρ_s)** relating selection frequency to the algorithmic sparsity score z_i.

Results. Table 3 summarizes alignment across methods. CoSALT achieved the highest Jaccard similarity (0.76), exceeding OIF by 12 points and Fixed-99 by 34 points.

Table 3: Human–Model Alignment.

Method	Jaccard (JI)	Boundary (ρ_b)	Sparsity (ρ_s)
Fixed-99	0.42 ± 0.11	0.31	0.28
Top-K	0.51 ± 0.09	0.44	0.35
SPOT	0.55 ± 0.09	0.52	0.45
RRCF	0.59 ± 0.08	0.61	0.52
OIF	0.64 ± 0.07	0.67	0.58
CoSALT	**0.76 ± 0.06**	**0.83**	**0.79**

Perceptual "Knee" Detection. Participants consistently placed their selection boundaries near the visual "knee" of the spread distribution—the inflection where the dense cluster gives way to a sparse tail. CoSALT locates this transition by finding where bulk Gaussian coding cost begins to exceed tail-model benefit. The boundary correlation ($\rho_b = 0.83$) indicates that MDL-derived thresholds align with perceptual segmentation, paralleling efficient-coding accounts of categorical perception (Młynarski & Hermundstad, 2021).

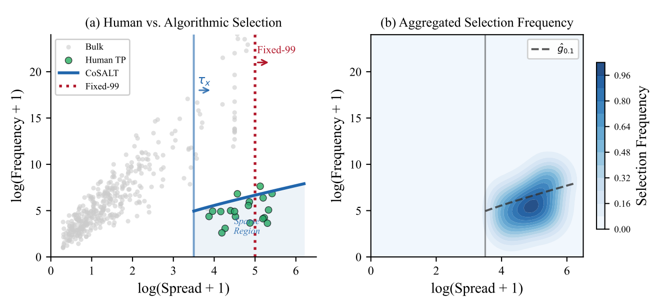

Figure 3: Human vs. algorithmic anomaly detection. (a) A representative stimulus with human selections overlaid. (b) Aggregated selection frequency across all participants and stimuli, peaking in the high-spread, low-frequency region that CoSALT's sparsity score targets.

Implicit Quantile Regression. Human selections were not driven by spread alone. As shown in Figure 3b, selection frequency peaked for sources with low frequency given their spread—the same region targeted by CoSALT's Stage 2 spline boundary. The sparsity correlation ($\rho_s = 0.79$) suggests that, even without explicit instruction, participants performed an implicit form of conditional boundary detection: judging each point relative to its expected baseline.

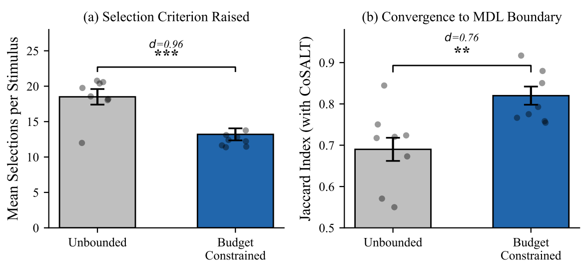

Figure 4: Resource constraints shift human detection criteria toward MDL-optimal boundaries. (a) Budget-constrained participants raised their selection threshold, flagging 29% fewer sources per stimulus. (b) This stricter criterion produced tighter alignment with CoSALT's MDL-derived boundaries, as measured by Jaccard Index.

Resource Constraints Shift Criteria. Resource-rational theories predict that explicit capacity limits raise decision criteria, making agents more selective. Figure 4(a) shows that budget-constrained participants selected significantly fewer points per stimulus than the unbounded group ($M = 13.2$ vs. 18.5; $t(58) = 3.72$, $p < .001$, $d = 0.96$), indicating a higher internal "noteworthiness" threshold under scarcity. Figure 4(b) shows the key consequence: this stricter criterion produced *better* alignment with CoSALT's MDL-optimal boundaries ($JI = 0.82$ vs. 0.69; $t(58) = 2.94$, $p = .005$, $d = 0.76$). Under scarcity, human criteria converged toward the boundary that CoSALT derives from the selection-cost term $\mathrm{Pen}_{\mathrm{select}}$. This convergence is consistent with the idea that an explicit selection penalty captures an attention-related cost, although it should not be taken as definitive proof of a unique cognitive mechanism.

Discussion

CoSALT connects operational detection to cognitive theories of adaptive thresholding in drifting, heavy-tailed environments. By balancing data compression against an explicit cognitive cost for allocating attention, CoSALT yields adaptive thresholds without parameter tuning. The model's two-stage architecture clarifies the adaptation process: Stage 1 identifies salient tails via model selection to handle extremes, while Stage 2 assesses contextual sparsity using quantile regression—measuring deviations from a normative boundary rather than the mean. This approach is supported empirically: CoSALT outperforms the evaluated baselines on network traffic and closely mirrors human selection boundaries, especially under resource constraints.

Beyond engineering performance, CoSALT suggests concrete cognitive predictions. First, increasing the selection cost (or imposing stricter reporting budgets) should shift criteria upward and reduce the number of "alarms," consistent with criterion shifts in signal detection theory (Green & Swets, 1966) and with resource–rational accounts of attention allocation. Second, the optional moving-average update predicts recency-weighted adjustments to criteria, consistent with adaptation-level ideas (Helson, 1964). Third, ranking by tail surprisal yields a psychologically interpretable notion of "unexpectedness" under local statistics, linking operational triage to efficient coding perspectives on salience (Młynarski & Hermundstad, 2021).

Limitations and future work. Our evaluation combines MAWI background with injected real worm traffic from CAIDA. While this provides realistic attack behavior and controlled prevalence, additional validation on fully labeled operational traces would strengthen external validity, particularly in the presence of benign high-spread behaviors and measurement noise introduced by distributed sketch merging. On the modeling side, CoSALT intentionally uses simple bulk+tail families to remain sketch-compatible; richer generative alternatives may capture multi-modality at the cost of complexity and computation.

For the behavioral study, our stimuli abstract away many cues available in real SOC work (e.g., temporal context, metadata, and feedback). A direct test of the cognitive interpretation would manipulate the reporting budget and payoff structure within participants, and measure how criterion placement and confidence evolve across windows with controlled drift. A promising paradigm is to present per-window scatterplots of $(\log S, \log F)$ (or $(\log S, \log F/S)$) while varying an explicit budget and base-rate, and to compare inferred human boundaries to the MDL-optimal thresholds. This design would allow sharper tests of whether the selection-cost term captures human sensitivity to workload and base-rate changes.

Conclusion

We introduced CoSALT, a resource-rational MDL model that selects per-window, context-sensitive thresholds for detecting sparse network super-spreaders from sketch summaries. By combining (i) adaptive bulk-tail coding with tail model selection and (ii) contextual sparsity scoring via quantile regression, CoSALT maintains high detection performance under drift while regulating analyst workload through an explicit selection cost. In addition to engineering gains on real traffic traces, CoSALT offers a computationally explicit hypothesis for how bounded agents may set criteria in heavy-tailed environments, and it yields testable predictions about how attention budgets shape human anomaly judgments.

Acknowledgments

This work is supported by the National Cyber Security-National Science and Technology Major Project 2025ZD1500204.

References

Arandjelović, O., Pham, D.-S., & Venkatesh, S. (2015). The adaptable buffer algorithm for high quantile estimation in non-stationary data streams. *2015 International Joint Conference on Neural Networks*, 1–7.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.

Butvinik, D., Boateng, N., & Hackmon, A. (2026). Adaptive kde for real-time thresholding: Prioritized queues for financial crime investigation. https://doi.org/10.48550/arXiv.2601.14473

CAIDA. (2004). *The caida ucsd dataset on the witty worm — march 19–24, 2004*. https://www.caida.org/catalog/datasets/witty_worm_dataset

Cao, Y., Ma, Y., Zhu, Y., & Ting, K. M. (2025). Revisiting streaming anomaly detection: Benchmark and evaluation. *Artificial Intelligence Review*, *58*(1), 8.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

Ciccione, L., Dehaene, G., & Dehaene, S. (2023). Outlier detection and rejection in scatterplots: Do outliers influence intuitive statistical judgments? *Journal of Experimental Psychology: Human Perception and Performance*, *49*(1), 129–144.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Guha, S., Mishra, N., Roy, G., & Schrijvers, O. (2016). Robust random cut forest based anomaly detection on streams. *International Conference on Machine Learning*, 2712–2721.

Han, D., Wang, Z., Chen, W., Wang, K., Yu, R., Wang, S., Zhang, H., Wang, Z., Jin, M., Yang, J., Shi, X., & Yin, X. (2023). Anomaly detection in the open world: Normality shift detection, explanation, and adaptation. *Proceedings of the Network and Distributed System Security Symposium*.

Helson, H. (1964). *Adaptation-level theory*. Harper & Row.

Iglesias Vázquez, F., Hartl, A., Zseby, T., & Zimek, A. (2023). Anomaly detection in streaming data: A comparison and evaluation study. *Expert Systems with Applications*, *233*, 120994.

Isaac, E. R., & Sharma, A. (2024). Adaptive thresholding heuristic for kpi anomaly detection. *2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 737–741.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Leveni, F., Weigert Cassales, G., Pfahringer, B., Bifet, A., & Boracchi, G. (2024). Online isolation forest. *Proceedings of the 41st International Conference on Machine Learning*, *235*, 27288–27298.

Li, J., Di, S., Shen, Y., & Chen, L. (2021). Fluxev: A fast and effective unsupervised framework for time-series anomaly detection. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*, 824–832.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1.

MAWI Working Group. (2016). *Mawi traffic archive: Samplepoint-f (2016)*. https://mawi.wide.ad.jp/mawi/samplepoint-F/2016/

Młynarski, W. F., & Hermundstad, A. M. (2021). Efficient and adaptive sensory codes. *Nature Neuroscience*, *24*, 998–1009.

Moskovitz, T., Miller, K. J., Sahani, M., & Botvinick, M. M. (2024). Understanding dual process cognition via the minimum description length principle. *PLOS Computational Biology*, *20*(10), e1012383.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*(5), 465–471.

Siffer, A., Fouque, P.-A., Termier, A., & Largouet, C. (2017). Anomaly detection in streams with extreme value theory. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1067–1075.

Yang, L., Guo, W., Hao, Q., Ciptadi, A., Ahmadzadeh, A., Xing, X., & Wang, G. (2021). Cade: Detecting and explaining concept drift samples for security applications. *30th USENIX Security Symposium*, 2327–2344.

Yang, X., Howley, E., & Schukat, M. (2025). Agent-based dynamic thresholding for adaptive anomaly detection using reinforcement learning. *Neural Computing and Applications*, *37*, 18775–18791.

Yang, Z., Li, S., Zhang, H., Yang, R., Liu, Q., & Wang, Y. (2025). Flask-sketch: Identifying sparse superspreaders in high speed network. *2025 IEEE Symposium on Computers and Communications (ISCC)*, 1–6.