PSL-Based Adversarial Filtering at the CDN Edge: A Production Evaluation Across 310 Points of Presence

Hugo Slopwell* Madison Chudwell Brennan Slopkowski

Chudmaxx Labs, Chudflare Inc.

Agartha Research Park, Denver, CO 80202, USA

{slopwell, chudwell, slopkowski}@chudflare.com

Abstract

We present the design, deployment, and evaluation of an adversarial request-filtering system based on the Perceived Sigma Level (PSL) metric, a 1.0–10.0 continuous-valued classifier trained on 4.2×10^9 labeled HTTP request samples. PSL inference is performed in-line at the CDN edge across 310 points of presence (PoPs) with a median added latency of 0.91 ms (p99: 4.3 ms). The classifier achieves 98.7% true-positive rate at a 0.4% false-positive rate against a held-out adversarial corpus. We further introduce *Chad Fight Mode*, a deployment posture in which all requests scoring above 5.5 on the PSL axis are silently downgraded via HTTP 403, and demonstrate that the approach reduces "unwanted cheekbone exposure" by 93.4% at the edge. We discuss the limitations of PSL as a security primitive, the ethical considerations of cosmetic-coded filtering, and outline a pipeline for federated retraining of the classifier from user-supplied hunch telemetry. We release neither the model weights nor the labeled corpus, on the grounds that we have not yet found a buyer.

Keywords: content delivery, web security, adversarial classification, edge computing, looksmaxxing, parody.

1. Introduction

The proliferation of high-jawline traffic on the public Internet poses an emerging challenge for content delivery providers [1, 2]. While prior work has focused on rate limiting [3], IP reputation [4], and JA3 fingerprinting [5], none of these signals adequately capture the *cosmetic posture* of an incoming request, defined here as the joint distribution over visitor facial-harmony characteristics inferred from User-Agent string, request cadence, and TLS ClientHello entropy.

This paper makes the following contributions:

1. We define the **Perceived Sigma Level (PSL)**, a continuous metric in $[1.0, 10.0]$ summarizing the cosmetic posture of an HTTP request (§3).
2. We describe a production system, CHUD FIGHT MODE, that performs in-line PSL inference at the edge of a global CDN with median added latency of under 1 ms (§4).
3. We evaluate the classifier on 1.43×10^{10} requests captured in March 2026, demonstrating that it removes 93.4% of cheekbone-bearing traffic at a 0.4% false-positive rate (§5).
4. We discuss the limitations of the approach, including its documented inability to reliably classify visitors who recently showered (§6).

A note on terminology. "Chud" in this paper refers to the operational class of *desired* CDN visitor, characterized in our dataset by $\text{PSL} \in [1.0, 3.5]$, a measurable hunching angle, and at least one Monster Ultra Zero opened in the prior 24 h window. We acknowledge the term carries connotations elsewhere; in this work it is used in its purely operational sense.

2. Related Work

Edge filtering. The use of edge computing for security has been extensively studied [11, 12, 13]. Most prior systems filter on lexical or structural request features. Our work extends this line of inquiry by introducing a cosmetic axis to the classification boundary.

Facial-harmony metrics. The PSL scale traces its origins to anonymous fora active between 2014–2019 [6, 7]. Rigorous benchmarks remain scarce. Hunched et al. [8] attempt a normalization to the [0,1] interval but report a Cohen's kappa below 0.2, suggesting fundamental inter-rater disagreement. We do not address this; our classifier is trained against a single rater (§5.1).

Looksmaxxing detection. Recent work has examined the detection of looksmaxxing communities in short-form video [9, 10]. To our knowledge this is the first work to apply analogous detectors at the CDN layer.

3. The PSL Metric

We define the Perceived Sigma Level of a request r as a weighted combination of three submetrics:

$$\text{PSL}(r) = w_1 \cdot \phi_{\text{cb}}(r) + w_2 \cdot \phi_{\text{jl}}(r) + w_3 \cdot \phi_{\text{cad}}(r), \quad (1)$$

where $\phi_{\text{cb}}(r)$ is the estimated cheekbone visibility, $\phi_{\text{jl}}(r)$ is the estimated jawline angle from horizontal, and $\phi_{\text{cad}}(r)$ is request cadence relative to the population median. Weights $(w_1, w_2, w_3) = (0.42, 0.39, 0.19)$ were fit by five-fold cross-validation against a hand-labeled corpus of 50,000 requests collected by Chudwell over the spring 2025 quarter.

Crucially, none of ϕ_{cb}, ϕ_{jl}, or ϕ_{cad} has access to any biometric data. All three are estimated purely from HTTP-layer signals; we present this estimation as a feature rather than a limitation.

3.1. Calibration

We calibrate PSL such that the median chud scores approximately 2.1 and fewer than 1% of chuds exceed 3.5. The scale is open at the top: a score of 10.0 corresponds to the theoretical Pareto-optimal Gigachad, which we have never observed in our deployment[1].

4. System Architecture

The Chud Fight Mode classifier is deployed as a hot path in the request pipeline of all 310 Chudflare PoPs. Inference proceeds as follows:

Algorithm 1 In-line PSL inference at the edge.

Require: Incoming HTTP request r, threshold τ.
$h \leftarrow \texttt{hashClientHello}(r)$
$f \leftarrow \texttt{extractFeatures}(r, h)$
$p \leftarrow \texttt{PSLModel}(f)$ ▷ $O(\text{model depth})$
if $p > \tau$ **then**
  return `Response`(403, "you got mogged.")
else if $p > \tau - 1.0$ **then**
  return `ChudChallenge`(r) ▷ 5-second hunch CAPTCHA
else
  return `forward`(r) ▷ originate to backend
end if

The classifier is implemented as a 3-layer perceptron with hidden dimension 64, exported to ONNX and compiled to WebAssembly for execution inside our Chudders serverless runtime [14]. Inference cost is dominated by feature extraction; the model itself runs in under $200\,\mu$s at p99.

4.1. Deployment Postures

Customers select among four operational postures, summarized in Table 1.

Posture	Threshold τ	Action above τ
Chuddle	7.5	forward, log only
Chud Mode	5.5	forward, log only
Chad Fight Mode	5.5	403 "you got mogged."
Under Mew Mode	4.0	102 "processing in silence"

Table 1: Deployment postures available to Chudflare customers. Under Mew Mode is recommended only for users actively being mogged in production.

5. Evaluation

We evaluate the classifier on 1.43×10^{10} requests captured across all 310 PoPs during the four-week window of March 2–29, 2026. Of these, 9.2% were labeled (post hoc) as adversarial; the remainder served as our negative class.

[1] Slopwell reports one suspected sighting in February 2026, but the offending request was filtered before headers could be captured.

5.1. Methods

Labels were assigned by Chudwell in a single sitting on March 30, 2026, while seated, hunched, and consuming a Monster Ultra Zero. We acknowledge this is methodologically suboptimal; see §6 for a discussion. All experiments were conducted on production traffic; we did not maintain a clean test set.

5.2. Results

Table 2 summarizes our results against three baselines: IP reputation lookup, JA3 fingerprinting, and a generic rule-based WAF.

Method	TPR	FPR	Med. lat.
IP reputation only	41.2%	1.8%	0.1 ms
JA3 fingerprint match	68.4%	2.3%	0.3 ms
Rule-based WAF (94 rules)	79.1%	4.6%	1.2 ms
PSL (ours)	**98.7%**	**0.4%**	0.9 ms

Table 2: True-positive rate, false-positive rate, and median added latency across 1.43×10^{10} March 2026 requests. PSL achieves a 24.9 percentage-point absolute improvement in TPR over the strongest baseline (rule-based WAF), at lower FPR and comparable latency.

5.3. Ablation

To isolate the contribution of each submetric, we re-train PSL with each of ϕ_{cb}, ϕ_{jl}, ϕ_{cad} ablated. Removing ϕ_{jl} produces the largest drop in TPR (-7.2 points), suggesting that jawline estimation carries the bulk of the classifier's signal. Removing ϕ_{cad} has negligible effect, which we attribute to its high collinearity with ϕ_{cb}.

5.4. Discussion

The PSL classifier achieves strong performance against our held-out adversarial corpus, but several observations warrant care. First, the true-positive rate is unusually high relative to prior work; this may indicate label leakage during corpus construction. Second, the FPR of 0.4% corresponds to approximately 5.7×10^7 misclassified chuds per day at peak traffic, which our customer support team characterizes as "a lot of angry emails."

6. Limitations and Ethical Considerations

We identify the following limitations.

Single-rater labels. All adversarial labels were assigned by a single annotator over a four-hour interval. Inter-annotator agreement was not measured, in part because we did not have a second annotator.

Cosmetic axis. The premise of filtering by inferred cosmetic posture is, strictly speaking, indefensible. We deploy it because our customers ask for it. We do not endorse the practice and recommend its use only as a parody.

Adversarial robustness. A motivated attacker can lower their PSL by hunching, opening Doordash, or replacing their TLS

stack with that of an unmaintained 2014-era distribution. We have observed all three in the wild.

Ethical disclosure. The authors hold equity in Chudflare, which sells the Chad Fight Mode product evaluated in this paper. The labeled corpus was collected without IRB review, on the grounds that no humans were classified (only their requests). We acknowledge this distinction is thin.

7. Conclusion

We have presented PSL, an edge-deployed adversarial classifier for HTTP requests, and CHAD FIGHT MODE, the deployment posture in which it is most usefully embodied. The classifier achieves a 98.7% true-positive rate at 0.4% FPR on a production corpus of 1.43×10^{10} requests. Future work includes (i) extending the classifier to support federated retraining from customer-supplied hunch telemetry, (ii) introducing a continuous variant of Under Mew Mode in which all responses are stochastically lowercased, and (iii) finding a buyer for the labeled corpus.

Acknowledgments

The authors thank the Chudflare ChudVerse 2026 program committee for helpful early feedback, the Slop Operations team for keeping inference latency under p99 targets, and Brennan's couch for hosting the entire labeling session.

Reproducibility. We release neither the labeled corpus nor the trained weights. Source code for the inference path is available on request, by sending an email and never receiving a reply. A reference implementation of the classifier (in Chudscript) is available at `https://chudflare.com/playground`.

References

[1] M. Chudwell. "Cosmetic Coded Filtering at the CDN Edge: Position Paper." In *Proc. ChudVerse '24*, pp. 12–19, 2024.

[2] B. Slopkowski. "A Taxonomy of High-Jawline Traffic in Production HTTP Logs." *ACM Trans. Slop Sys.*, 14(3):1–28, 2025.

[3] M. Nottingham and R. Fielding. "Additional HTTP Status Codes." RFC 6585, IETF, 2012.

[4] The Spamhaus Project. "The Spamhaus Block List." `https://www.spamhaus.org/sbl/`, accessed 2026.

[5] J. Althouse, J. Atkinson, and J. Atkins. "JA3: A Method for Profiling SSL/TLS Clients." *Black Hat USA*, 2017.

[6] Anonymous. "A Note on Sigma Calibration." *[forum thread]*, 2018. Citation withheld for taste.

[7] Anonymous. "Re: A Note on Sigma Calibration (followup)." *[same forum]*, 2019. Citation withheld for taste.

[8] H. Hunched, P. Posture, and L. Levator. "A Benchmark for Inter-Rater Agreement on Cosmetic Scales." *Proc. Slop Conf. '22*, pp. 244–251, 2022.

[9] TikTok Trust & Safety Team. "Trend Detection in Looksmaxxing Communities: Internal Report." *Unpublished*, 2024.

[10] M. Chudwell. "Concept Drift in the TikTok Algorithm: A Longitudinal Study." *ChudVerse '25*, pp. 88–104, 2025.

[11] A. Real Company. "Cloudflare Workers: Serverless on the Edge." `https://workers.cloudflare.com`, 2017. *Not us. They are real. We are not.*

[12] Fastly Inc. "Compute@Edge: Architectural Overview." *Fastly White Paper*, 2019.

[13] Akamai Technologies. "Origin Shield for the Anti-Looksmaxxing Era." *Akamai Tech Doc*, 2020.

[14] H. Slopwell. "Chudders: Edge-Native JavaScript for the Hunched Engineer." *Chudflare Engineering Blog*, March 2026. `https://chudflare.com/blog/`.

Compliance Notice

This document is a parody. *Chudflare* is a parody of *Cloudflare, Inc.* It is not affiliated with, endorsed by, sponsored by, or in any way related to Cloudflare, Inc. Cloudflare ships excellent infrastructure that this paper has nothing to do with. All cited works marked "Anonymous" or attributed to *Chudmaxx Labs*, *Hugo Slopwell*, *Madison Chudwell*, or *Brennan Slopkowski* are fictional and should not be entered into any reference manager that you intend to use in a serious work. No humans were classified in the preparation of this manuscript. For real infrastructure research, visit `https://blog.cloudflare.com/research`.