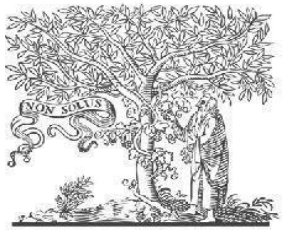
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2016 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating newcollective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper; all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 7th Jan 2016. Link

https://ijiemr.org/downloads.php?vol=Volume-05& issue=issue01

DOI: 10.48047/IJIEMR/V05/ISSUE 01/12

Title Privacy-Preserving Federated Threat Intelligence Learning for Heterogeneous Cloud Environments

Volume 05, ISSUE 01, Pages: 79 – 86

Paper Authors
Dr. Satya Venkata Surya Narayana Murthy



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Privacy-Preserving Federated Threat Intelligence Learning for Heterogeneous Cloud Environments

Dr. Satya Venkata Surya Narayana Murthy
Wipro Limited
Tower Center Blvd, East Brunswick, NJ 08816
nimusha73@gmail.com

Abstract
Attacks such as credential stuffing frequently move between organizations, yet security teams rarely share raw logs because of confidentiality obligations and because their event schemas and definitions of "suspicious" differ. This paper presents FedShield, a research design for cross-organization threat detection in which no raw security log leaves its owner. Each participating cloud operator trains a regularized logistic classifier on its own firewall, authentication and endpoint telemetry, and only governed aggregates, namely model parameters and indicator counts that pass disclosure thresholds and noise addition, are exchanged through standard threat-intelligence formats. The participating organization is the unit of analysis, which makes accountability and the decision boundary explicit. We formalize local training, parameter averaging and the disclosure mechanism; define evaluation with macro-averaged F1, worst-participant recall, bytes exchanged and indicator re-identification risk; and specify comparisons against independent local models, a pooled-data upper bound and rules-only indicator exchange. A worked scenario follows a credential-stuffing campaign moving between two cloud tenants. The design is restricted to technologies available in 2016. No experimental results are claimed; the contribution is an implementable, independently testable research design and an explicit account of its privacy and validity risks.

Keywords: threat intelligence sharing, privacy-preserving machine learning, federated learning, differential privacy, logistic regression, cloud security, STIX/TAXII.

Introduction
Cloud operators observe only a fraction of any campaign. A credential-stuffing attack that tests stolen passwords against one tenant is often visible, moments earlier or later, at another. Sharing evidence across

organizations could improve detection coverage, and threat-intelligence standards and guidance encourage such cooperation [1], [7], [16]. In practice, however, organizations are reluctant to transfer raw security logs, which contain user identities, internal addresses and business-sensitive activity, and which may be subject to regulatory restrictions [20].

Two frictions dominate. The first is institutional confidentiality: participants will share only what they can justify to their own governance bodies. The second is semantic inconsistency: each operator uses its own firewall, authentication and endpoint schemas, and the same label ("suspicious login") may mean different things in different places. Pooling heterogeneous records naively would hide both problems and produce a model whose apparent accuracy does not transfer [13], [17].

This paper proposes FedShield, which keeps learning local and exchanges only governed aggregates. Its contributions are:

1) A problem formulation in which the **participating organization** is the unit of analysis, so that coverage, disclosure and accountability are measured per participant.
2) An architecture combining locally trained logistic classifiers, a shared event taxonomy and a governed aggregation step that applies disclosure thresholds and calibrated noise before anything leaves a site.
3) A mathematical specification of local training, parameter averaging, the disclosure mechanism and the evaluation metrics, including worst-participant recall and re-identification risk.
4) A reproducible evaluation protocol and threat analysis, presented without unperformed experiments.

Section II reviews related work, Section III states the research questions, Section IV the data model, Section V the architecture and Section VI the formalization. Sections VII–IX cover the illustrative scenario, implementation and evaluation protocol; Section X discusses validity and ethics, and Section XI concludes.

2. Background and Related Work

A. Threat-Intelligence Sharing

The intrusion kill chain [2] framed defense as disrupting campaign stages, motivating the exchange of indicators across organizations. STIX [1] and TAXII [15] standardized how indicators are expressed and transported, NIST SP 800-150 gave guidance on sharing programs [7], and platforms such as MISP operationalized community sharing [16]. These approaches exchange indicators, not models, and typically rely on manual or rule-based curation.

B. Privacy-Preserving and Distributed Learning

Differential privacy offers a formal bound on what released statistics reveal about any individual record [3], [4], and has been applied to empirical risk minimization including logistic regression [10]. Distributed optimization methods such as ADMM [9] and, more recently, model averaging across clients [6] and collaborative deep learning with selective parameter sharing [5] allow learning without centralizing data. FedShield applies these ideas to the specific governance setting of security operations.

C. Disclosure Risk

Anonymized data can often be re-identified when records are sparse or distinctive [8], [11], and released models can leak information about their training data [19]. Rare security indicators are particularly risky: an unusual outbound destination

reported by only one participant can reveal that participant's incident. FedShield therefore treats re-identification risk as a first-class metric.

D. Machine Learning in Operational Security

Sommer and Paxson describe why anomaly and misuse detectors that work in the laboratory often fail in operation [13]. Class imbalance [12], concept drift [18] and the hidden maintenance costs of machine-learning pipelines [14] all apply strongly to cross-organization settings

Research gap: Indicator-sharing standards do not learn from local evidence, while privacy-preserving learning research rarely addresses heterogeneous security schemas, per-participant fairness of detection or the disclosure risk of rare indicators. FedShield is designed to connect these strands

3. Problem Formulation and Research Questions

The setting comprises three independent cloud operators whose firewall, authentication and endpoint-event schemas are incompatible. The objective is to improve cross-site detection coverage while bounding disclosure risk. Four research questions are posed:

1) **RQ1 (effectiveness):** Do locally trained classifiers with governed aggregate exchange improve detection coverage compared with independent local models and rules-only indicator exchange, and how close do they come to a pooled-data upper bound?
2) **RQ2 (generalization):** How does performance change when the distribution of evidence differs between participants or over time, and does a shared taxonomy with preserved provenance reduce that degradation?
3) **RQ3 (governance and efficiency):** What detection coverage is achievable under fixed disclosure budgets and communication limits?
4) **RQ4 (explainability):** What evidence lets each participant reconstruct why a local event was scored as it was, including the contribution of shared knowledge?

Each question is a testable comparative proposition. The scope excludes universal intrusion detection, automated cross-organization response and privacy guarantees beyond the stated mechanism.

4. Data Model and Provenance

Each participant maps its local telemetry to a shared event taxonomy (Table I). Records keep stable local identifiers, collection timestamps, source-system identifiers and a versioned mapping rule, so that the meaning of each shared feature is traceable.

TABLE I. Evidence types and the shared representation

Evidence type	Local sources (vary by participant)	Shared representation
Failed authentication bursts	Identity-provider and VPN logs	Count of failures per account and source within a fixed window
Unusual outbound destinations	Firewall and proxy logs	Rarity of destination relative to the participant's own baseline
Host quarantine events	Endpoint protection consoles	Binary label plus time since quarantine, used as weak supervision

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Leakage and provenance. Provenance fields are kept separate from features, and features
use only information available at decision time [14].

Normalization. Mappings from local categories to the shared taxonomy are documented, count features have explicit windows, and missing measurements are never treated as zeros.

Governance. Each participant ingests only authorized data, logs every transformation and documents what aggregation or anonymization removes.

Partitioning. Evaluation partitions are defined by time and by participant (leave-one-organization-out), because random splits of related records overstate performance [18].

5. The FedShield Architecture

Fig. 1 shows the information flow. Raw logs remain inside each organization; only governed aggregates cross organizational boundaries.



Fig. 1. FedShield architecture. Local training and governed aggregation (dark) are the stages where privacy controls apply.

- Local telemetry and schema mapping convert each participant's events into the shared taxonomy and validate timestamps, formats, authorization and duplicates.
- Local logistic training fits a regularized classifier on local data only.
- Governed aggregation applies minimum-count thresholds and calibrated noise, records every release in an audit log, and rejects releases that exceed the disclosure budget.
- Shared indicators and model are distributed through STIX/TAXII as an averaged parameter vector and thresholder indicator counts.
- Local scoring and review combine the local and shared models and route high-consequence results to a human analyst. Model confidence alone never authorizes a change to security controls.

Failure handling. If a participant is offline or its schema mapping fails validation, the round proceeds without it and the omission is recorded. Fallbacks are monitored to check whether any participant is persistently excluded.

6. Mathematical Formulation

A. Local Training

Organization **j** holds **n_j** labeled events **(x_{ij}, y_{ij}) with $y_{ij} \in \{-1, +1\}$.** It minimizes the regularized logistic loss

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

$$L_j(w_j) = \Sigma_i \log(1 + \exp(-y_{ij} \cdot w_j \cdot x_{ij})) + \lambda \|w_j\|^2 \quad (1)$$

B. Governed Aggregation

Shared parameters are combined by sample-weighted averaging [6]:

$$\bar{w} = \Sigma_j (n_j / n) \cdot w_j, \quad n = \Sigma_j n_j \quad (2)$$

Before release, each statistic q is perturbed with the Laplace mechanism [4], where Δq is its sensitivity and ε the privacy parameter:

$$\tilde{q} = q + Lap(\Delta q / \varepsilon) \quad (3)$$

Indicator counts below a minimum threshold κ are suppressed, in the spirit of k-anonymity [8].

C. Decision Rule

Let S(X) be the calibrated score of an event. With C_FN the cost of a missed attack and C_FP the cost of an unnecessary escalation, the threshold is chosen on validation data subject to analyst capacity K:

$$\delta(X) = 1[S(X) \geq \tau], \quad \tau^* = \arg\min_\tau C_{FN} \cdot FN(\tau) + C_{FP} \cdot FP(\tau) \quad s.t. \quad \Sigma_i \delta(X_i) \leq K \quad (4)$$

D. Evaluation Metrics

Because the goal is coverage for every participant, the protocol reports the macro-average and the minimum across participants:

$$F1_{macro} = (1/J) \Sigma_j F1_j, \quad R_{min} = \min_j Recall_j \quad (5)$$

Re-identification risk is the fraction of released indicators that could be attributed to a single participant:

$$R_{id} = |\{k \in K_{released} : c_k < \kappa\}| / |K_{released}| \quad (6)$$

These definitions do not by themselves establish privacy guarantees. Composition across repeated releases, membership inference from rare indicators [19] and incompatible event taxonomies can each invalidate an apparently convincing result.

7. Illustrative Scenario

Consider a credential-stuffing campaign that first targets tenant A and, two hours later, tenant B.

Step 1: **Define the observation boundary:** each operator can inspect only its own logs; neither can see the other's users.

Step 2: **Collect and map evidence:** tenant A observes a burst of failed logins from a small set of source networks, followed by an unusual outbound destination from one host that is later quarantined.

Step 3: **Train and aggregate:** tenant A's local model learns that this combination is predictive; its updated parameters, not its logs, contribute to the next averaged model.

Step 4: **Evaluate a simple baseline:** with rules-only exchange, tenant B receives the source addresses only if an analyst at A curated and shared them in time.

Step 5: **Score at tenant B:** the shared model raises the score of B's similar failed-login burst, and the event is escalated with the local evidence and the model version attached.

If the source-network indicator was seen by only one participant, the governed aggregation step suppresses it rather than revealing A's incident. This scenario illustrates the workflow; it is not a measured experiment, and no detection rate or runtime is inferred from it.

8. Implementation Considerations

8 Table II lists a period-appropriate technology stack. These are examples of technologies available in 2016, not a claim that an experiment was executed with them.

TABLE II. Candidate implementation stack (2016)

Component	Technology and role
Indicator exchange	STIX 1.x and TAXII 1.x for expressing and transporting shared indicators [1], [15]
Cloud telemetry	OpenStack telemetry (Ceilometer) and native firewall, identity and endpoint logs
Local processing	Apache Spark 1.x for windowed feature extraction [21]
Learning	Python 2.7/3.5 with scikit-learn logistic regression [22]
Governance	Audit log of every release; disclosure-budget accounting per participant

Engineering practice. A versioned data contract defines the shared taxonomy. Parsing, mapping, training, aggregation, scoring and audit are separate modules [14]. Configuration management records software versions, mappings, seeds, privacy parameters and thresholds, and access control restricts who may view raw events, change the model or approve releases.

Trade-offs. Larger aggregation intervals reduce bytes exchanged but delay shared knowledge;

stronger noise lowers disclosure risk but also lowers utility. These trade-offs require measurement.

9. Evaluation Protocol

FedShield is compared with three baselines (Table III). All candidates use the same permitted information and horizon, except the pooled-data oracle, which is deliberately given more data to serve as an upper bound.

TABLE III. Baselines and the question each comparison answers

Baseline	Question answered by the comparison
Independent local models	Does governed sharing add coverage beyond what each site achieves alone?
Pooled-data oracle	How much utility is lost by not centralizing raw logs?
Rules-only indicator exchange [1], [16]	Does learning add value beyond curated indicator sharing?

Metrics. Macro-averaged F1, worst-participant recall, bytes exchanged per round and re-identification risk are reported with bootstrap confidence intervals. Accuracy alone is not reported because of class imbalance [12].

Robustness and ablation. Tests vary missing fields, observation windows, label noise, the number of participants and the privacy parameter ε. Ablations remove the shared taxonomy, the noise mechanism and the threshold κ one at a time.

Reporting standard. No dataset is supplied with this paper, so the protocol is specified without results. A valid evaluation should provide reproducible code, disclosed partitions, documented exclusions and uncertainty estimates, and should report unfavorable outcomes.

10. Threats to Validity and Ethical Considerations

Internal validity. Inconsistent event taxonomies can make a shared feature mean

different things at different sites, and weak labels from quarantine events may be noisy.

Privacy. Repeated releases consume privacy budget, and rare indicators or model parameters may still leak information through membership or model-inversion attacks [19]. Budgets must be tracked across all rounds.

External validity. Three participants may not represent larger consortia; drift in attacker behavior and in each participant's infrastructure can erode shared models [18].

Feedback. Analysts may focus on events the model highlights, so later labels are not unbiased. Periodic review of a random sample of low-scored events mitigates this.
Ethics and governance**.** Participation requires written agreements on permitted use, retention and incident handling. Testing is limited to authorized data, and any automated response remains proportionate to the confidence and consequence of the recommendation.

11. Conclusion and Future Work

FedShield is a technically specified design for cross-organization threat detection that never transfers raw security logs. It links the twin problems of institutional confidentiality and inconsistent definitions of suspicious activity to local logistic learning, a shared taxonomy and governed aggregate exchange, with per-participant metrics for coverage and disclosure.
The credential-stuffing scenario shows how one participant's local learning can raise the alert at another without revealing either party's logs. The contribution is an architecture and analytical argument, not a claim of demonstrated superiority. Future work will run the protocol on authorized multi-organization data, quantify the privacy–utility trade-off and study larger and more heterogeneous consortia.

Acknowledgment

Technology and product names are used for identification and historical context only; no endorsement by their respective owners is implied.

References

1. S. Barnum, "Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX)," MITRE Corporation, McLean, VA, USA, White Paper, 2012.
2. E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," *Leading Issues in Information Warfare & Security Research*, vol. 1, no. 1, pp. 80–106, 2011.
3. C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloq. Automata, Languages and Programming (ICALP)*, Venice, Italy, 2006, pp. 1–12.
4. C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Theory of Cryptography Conf. (TCC)*, New York, NY, USA, 2006, pp. 265–284.
5. R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 22nd ACM SIGSAC Conf. Computer and Communications Security (CCS)*, Denver, CO, USA, 2015, pp. 1310–1321.
6. H. B. McMahan, E. Moore, D. Ramage, and B. Agüera y Arcas, "Federated learning of deep networks using model averaging," arXiv:1602.05629, Feb. 2016.
7. C. Johnson, L. Badger, D. Waltermire, J. Snyder, and C.

Skorupka, "Guide to cyber threat information sharing," NIST Special Publication 800-150, National Institute of Standards and Technology, Gaithersburg, MD, USA, Oct. 2016.
8. L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.
9. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
10. K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization," *J. Mach. Learn. Res.*, vol. 12, pp. 1069–1109, 2011.
11. A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. IEEE Symp. Security and Privacy*, Oakland, CA, USA, 2008, pp. 111–125.
12. H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.
13. R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symp. Security and Privacy*, Oakland, CA, USA, 2010, pp. 305–316.
14. D. Sculley *et al.*, "Hidden technical debt in machine learning systems," in *Advances in Neural Information Processing Systems 28 (NIPS)*, Montréal, Canada, 2015, pp. 2503–2511.
15. J. Connolly, M. Davidson, and C. Schmidt, "The Trusted Automated eXchange of Indicator Information (TAXII)," MITRE Corporation, McLean, VA, USA, White Paper, 2014.
16. C. Wagner, A. Dulaunoy, G. Wagener, and A. Iklody, "MISP: The design and implementation of a collaborative threat intelligence sharing platform," in *Proc. ACM Workshop on Information Sharing and Collaborative Security (WISCS)*, Vienna, Austria, 2016, pp. 49–56.
17. S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
18. J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, Art. 44, 2014.
19. M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Computer and Communications Security (CCS)*, Denver, CO, USA, 2015, pp. 1322–1333.
20. European Parliament and Council, "Regulation (EU) 2016/679 (General Data Protection Regulation)," *Official Journal of the European Union*, L 119, pp. 1–88, Apr. 2016.
21. M. Zaharia *et al.*, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *Proc. 9th USENIX Symp. Networked Systems Design and Implementation (NSDI)*, San Jose, CA, USA, 2012, pp. 15–28.
22. F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.