IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

Privacy Preserving Federated Intrusion Detection in Distributed Networks

Ram Kumar Chouhan[1], Dr. Kirti Jain[2], Mahesh Kumar Bagwani[3]
Sanjeev Agrawal Global Educational University, Bhopal, Madhya Pradesh, India*

Abstract: Collaborative network intrusion detection requires diverse training data while preserving organisational control over traffic records, therefore this study examines a federated detector that combines a convolutional recurrent classifier with update clipping and Gaussian perturbation alongside secure aggregation, using the original study tables to compare centralised training and ordinary federated averaging with the protected configuration on NSL-KDD and CICIDS2017, where reported protected model accuracies are 97.3% and 97.6% respectively compared with centralised accuracies of 98.6% and 99.1% and ordinary federated accuracies of 97.9% and 98.2%, while the reported communication totals increase from 142 MB to 148 MB corresponding to a calculated increase of approximately 4.23%, with the reported membership inference indicator decreasing from 34.1% to 9.7% although its attack definition and metric are unspecified, consequently these measurements provide descriptive evidence of a modest reported detection gap but do not establish statistical significance or a verified privacy guarantee, because run logs and the noise multiplier are unavailable and the corrected weighted aggregation formulation has not been rerun, while the accompanying analysis identifies the sensitivity and accounting conditions required to validate the target privacy budget before deployment claims can be justified.

Index Terms: Differential privacy; distributed networks; federated learning; intrusion detection; secure aggregation; traffic classification

I INTRODUCTION

Network operators collect flow statistics to identify scanning and service abuse as well as denial of service activity, yet an individual organisation observes only part of the wider attack landscape, which creates an incentive to learn jointly across administrative boundaries while retaining control over records that describe its users and infrastructure, therefore collaborative intrusion detection requires a training arrangement that accounts for both statistical diversity and information disclosure rather than assuming that pooling all observations is acceptable.

Federated learning coordinates local optimisation through a server that distributes a model and combines updates from participating sites [1], while the broader research agenda includes heterogeneous data and unreliable devices as well as privacy and governance [2], consequently its suitability for intrusion detection depends on how those issues interact in the intended deployment because a method designed for a large population of mobile devices may behave differently when only a few organisations participate in every training round.

Keeping traffic records at their source reduces direct exposure through central collection, although model gradients can encode information about training observations as demonstrated by gradient reconstruction research [3], while stronger reconstruction procedures show that averaging alone does not establish protection in the settings they examine [4], therefore the information released by an intrusion detector must be analysed at the update and aggregate levels as well as through the final classifier without assuming that findings from image reconstruction transfer quantitatively to network flows.

Membership inference provides a further route to disclosure because an adversary can use the behaviour of a classifier to distinguish records that contributed to training [5], while attack evaluation can produce misleading reassurance when it averages across unsuitable operating regions [6], consequently a privacy evaluation requires an explicit access model and a defined metric together with a credible attack baseline so that a lower numerical value represents an actual reduction in the specified risk.

We examine PP-IDS through the comparative measurements supplied in the original study tables and a technical analysis of its proposed privacy mechanism, where the contribution is a descriptive comparison of detection quality and reported communication cost together with a consistent account of weighted sensitivity and an explicit validation agenda, while the existing measurements are distinguished from the corrected formulation because no rerun or independent reproduction is available and the established classifier and protection primitives are not presented as new inventions.

The paper develops the detection setting and privacy model before presenting the data pipeline and candidate implementation details, after which the conditional analysis explains the requirements for a numerical guarantee and the Results section compares the original reported configurations, while the discussion separates directly reported values from calculated differences and identifies which conclusions remain dependent on unavailable experimental records.

II RELATED WORK AND RESEARCH POSITION

A Federated intrusion detection

Federated intrusion detection extends collaborative learning to traffic generated across gateways and organisational sites, with Rahman and colleagues comparing centralised and local approaches with federation in an Internet of Things setting [7], while DÏoT studies federated anomaly detection for connected devices [8] and DeepFed develops federated deep learning for industrial cyber physical systems [9], therefore these studies establish relevant design precedents although their models and threat assumptions must be compared individually rather than treated as a single privacy baseline.

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

The distinction between raw data locality and a formal privacy guarantee is central to this comparison because encrypted communication can protect messages during transport while the authorised recipient may still learn from the decoded information, whereas secure aggregation changes which intermediate values become visible and differential privacy limits disclosure under a defined change to the training data, consequently the related work should be assessed according to protected units and adversaries as well as accuracy rather than by the presence of the word privacy in a system description [10], [11], [12], [13], [14].

B Privacy mechanisms and accounting

Bagwani and colleagues examine real time signature based detection and prevention of distributed denial of service attacks in cloud environments [15], which provides relevant application context for traffic monitoring and operational response, while the present framework addresses a different question concerning the information disclosed during collaborative model training, therefore the cloud detection study motivates the deployment setting without serving as evidence for the differential privacy guarantee.

Khare and colleagues examine trustworthy intrusion detection under class imbalance and concept drift with attention to leakage and evaluation quality [16], thereby providing a directly relevant basis for reporting rare class performance and testing data separation in this study, while the present formulation adds the distinct requirements of private weighted aggregation and a declared honest participant cohort rather than treating benchmark detection quality as sufficient evidence of confidentiality.

Differential privacy gives a mathematical description of how the distribution of an output changes between neighbouring datasets [10], while private deep learning uses clipping and noise with composition accounting to control information released during optimisation [11] and client level federation extends the protected contribution from an individual example to a participant [12], therefore a system must state the neighbouring relation before choosing its sensitivity because example replacement and replacement of an organisation represent different privacy questions.

Secure aggregation enables an authorised aggregate without exposing every participant input under the protocol assumptions [13], while distributed discrete noise research explicitly considers the interaction between encoding and modular summation [14], consequently a floating point Gaussian equation followed by a modular mask equation is only an abstract description until the implementation specifies how values are encoded and how numerical effects enter the guarantee, with Rényi differential privacy providing one established route for composing the resulting releases [17].

C Statistical heterogeneity

Client datasets can differ in class frequency and feature distributions as well as collection conditions, while Dirichlet partitions provide a controlled way to vary label heterogeneity [18] and methods such as FedProx [19] and SCAFFOLD [20] address aspects of federated optimisation under heterogeneity, therefore an intrusion detection study should distinguish an improvement in optimisation from an improvement caused by a different partition or an additional class balancing procedure that is absent from its baseline.

Private learning on heterogeneous data has also been studied directly through DP-SCAFFOLD [21], while local batch normalisation provides a separate treatment of feature shift in FedBN [22], consequently the research gap cannot be stated as a general absence of work combining privacy with heterogeneous federation and is instead framed here as the need for a carefully specified assessment of those interactions for intrusion detection with a small number of organisational clients.

D Integrity and scope of contribution

Backdoor research demonstrates that collaborative training can be manipulated through submitted model updates [23], while defence work such as FLAME examines mechanisms for reducing this risk [24], therefore confidentiality mechanisms must be evaluated alongside a separate integrity model because hiding individual contributions can prevent the inspection required by some robust aggregation methods and a system that clips updates only inside compliant client software cannot assume that a malicious participant respects the same bound.

The present study combines a descriptive analysis of the original result tables with a corrected weighted release specification, while its evaluation guidance identifies the additional controls needed to validate those measurements, consequently the empirical contribution is limited to the reported configurations and the proposed mechanism corrections are not claimed to have generated the existing scores.

TABLE I COMPARISON OF RELATED RESEARCH DIRECTIONS

Direction	Representative work	Role in this study
Federated IDS	DÏoT and DeepFed	Application and architecture precedents
Formal privacy	Client DP and DP-SCAFFOLD	Protected unit and private optimisation
Secure summation	Bonawitz and distributed discrete Gaussian	Aggregate visibility and encoding
Integrity	Backdoor attacks and FLAME	Separate malicious update evaluation

III SYSTEM MODEL AND SECURITY OBJECTIVES

A Participants and information flow

We consider K organisations that each hold a labelled training dataset D_k and maintain a local preprocessing pipeline, while a coordination server maintains global parameters w_t at round t and a shared model schema determines the meaning of every input and parameter coordinate, therefore clients may retain different record counts and class distributions provided that their encoded features and output labels remain

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

compatible with the common training objective shown in Figure 1.

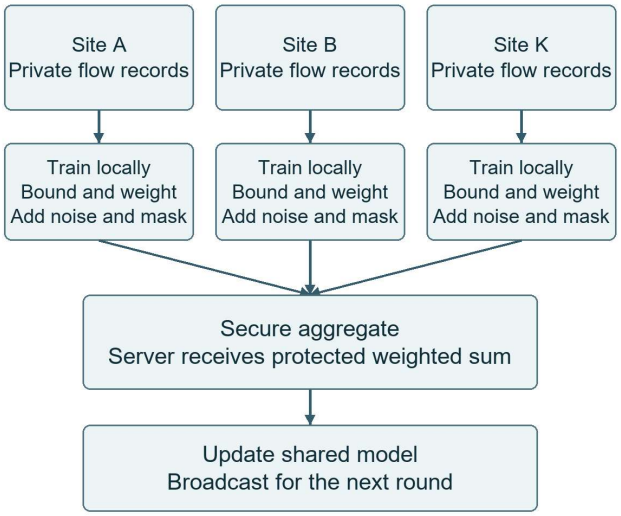


Figure 1 Federated architecture with private local training and protected weighted aggregation

The local site retains raw records and its fitted transformation parameters unless an explicitly authorised statistic is released, while the server receives authenticated protocol messages and the decoded aggregate together with any public participation information, therefore the observable transcript includes more than the final model and can contain membership lists and auxiliary counts as well as repeated aggregates that must be included in the privacy analysis whenever they depend on protected data.

B Adversary capabilities

The confidentiality analysis considers a server that follows the specified protocol but attempts to infer information from its observations, potentially with cooperating clients that disclose their own data and messages, while an external adversary may query the released classifier to conduct membership inference, consequently server observation and external model access require distinct empirical attack settings even when the same formal privacy bound covers the relevant released objects under its assumptions.

For the integrity analysis a malicious client may alter its training procedure and submitted contribution, including omission of clipping or noise, while the current formulation does not provide a proof that hidden inputs obey the prescribed bound, therefore confidentiality is conditional on sufficient honest noise contributors and poisoning resistance is not established, with fully malicious server behaviour and inference time evasion remaining outside the primary model rather than being implicitly covered by secure summation.

C Protected unit and adjacency

The main analytical case uses replacement adjacency in which one registered organisation replaces its complete private dataset while its public identity and fixed weight remain unchanged, so the public roster has the same K entries in both worlds and only one local training contribution differs, therefore the result protects information associated with that dataset replacement but does not claim to hide whether the organisation appears in the public participation roster [10], [12].

An add or remove relation would require a different treatment of participation and aggregation denominators, while data dependent dataset size weights can also change between neighbouring worlds, consequently the following expressions use fixed public weights and cannot be transferred unchanged to private dataset size weighting, with an implementation required either to justify public bounded counts or to use a mechanism that protects any released size information and accounts for its effect on sensitivity.

D Security objectives

The confidentiality objective is to limit distinguishability of the complete permitted training transcript under the chosen neighbouring relation, while the detection objective is to recognise attack classes without imposing an unacceptable benign false alarm rate and the operational objective is to quantify the total communication and computation required for that combination, therefore none of these objectives can be replaced by a single accuracy value or by a statement that raw traffic remains local.

TABLE II ADVERSARIES AND LIMITS OF PROTECTION

Adversary	Observation or action	Required condition
Curious server	Repeated aggregates and metadata	Account for complete transcript
Colluding clients	Reveal known inputs and noise	Minimum honest release cohort
External model user	Queries released classifier	Defined membership attack access
Malicious client	Can bypass local protection	Integrity is not established

IV DATA PREPARATION AND TEMPORAL REPRESENTATION

A Dataset scope

NSL-KDD provides a recognised historical intrusion detection benchmark with an established training and test separation [25], while CICIDS2017 provides flow records from a controlled capture campaign with documented attack periods [26], therefore both can support reproducible comparisons but neither alone demonstrates current operational coverage and their collection conditions must be retained in the interpretation because high benchmark accuracy does not establish performance on newly deployed networks.

The proposed evaluation preserves the NSL-KDD training and test distinction while taking validation data only from the training portion, whereas the CICIDS2017 evaluation uses a documented group or chronological separation suitable for the intended claim, consequently the exact capture files and attack labels must be recorded alongside cleaning decisions and final record counts because different subsets or mappings can

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

change both class imbalance and the difficulty of the learning task.

An additional dataset such as UNSW-NB15 can test whether conclusions depend strongly on the original two sources [27], although combining datasets without reconciling their feature semantics would create a separate transfer problem, therefore an extension should either train and evaluate within each dataset under matched protocols or define a justified common representation before making cross dataset claims rather than concatenating incompatible attributes solely to increase sample size.

B Separation before transformation

The data pipeline begins with source identification and deterministic cleaning followed by allocation of independent train and validation and test groups, while all learned transformations are fitted using training records only, therefore window generation occurs after separation and never allows the same underlying flow to appear on both sides of an evaluation boundary, as illustrated in Figure 2 where separate streams prevent the overlap that can arise from randomly splitting precomputed windows.

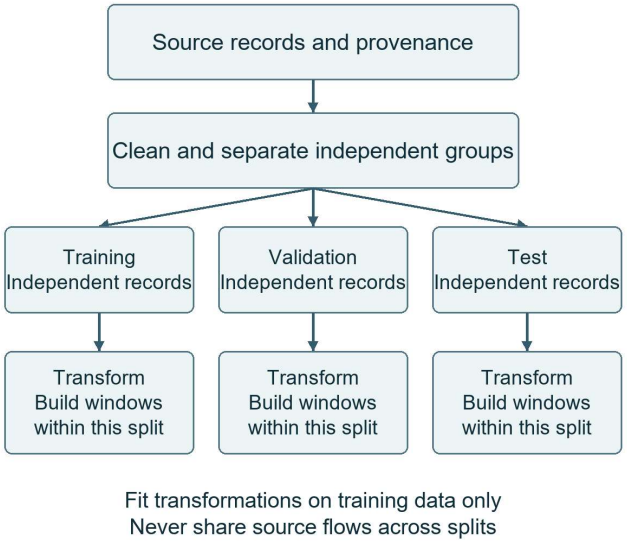


Figure 2 Separation of records before fitted transformations and window construction

Cleaning records the treatment of missing values and nonfinite measurements together with duplicate rows and excluded attributes, while potential capture identifiers are inspected for shortcuts that reveal the label or experiment schedule, consequently the reproducibility record contains both retained and removed feature names and their rationale so that a later implementation can distinguish legitimate traffic information from information that would not be available at the intended prediction time.

C Consistent feature semantics

Every client uses the same categorical vocabulary and numerical feature order because parameter averaging assumes a shared coordinate system, while a vocabulary can be established from public metadata or a separately protected discovery step, therefore a protocol must not independently assign one hot positions from each local dataset and then average models whose columns represent different categories even though the resulting tensors have the same apparent dimensions.

Local min max scaling avoids direct transmission of local extrema but can assign different physical meanings to the same transformed value across clients, while fixed public ranges offer a common interpretation when such ranges are defensible, consequently the evaluation should compare the selected scaling policy under an explicit deployment procedure that covers constant features and unseen values as well as new sites rather than treating local fitting as sufficient evidence of privacy or numerical compatibility.

D Window construction and targets

The temporal variant groups records by a declared traffic stream and orders them by an available valid timestamp before forming windows of length L, with a causal target defined as the class of the final flow in each window, therefore the classifier estimates that final label using its permitted context and does not use future flows, while the stride and minimum stream length determine the number of resulting training examples and must be recorded separately from the original record count.

For a dataset whose available representation does not establish meaningful chronology the individual flow classifier is the primary valid baseline, while an LSTM sequence built from arbitrary row order is treated as an artificial contextual experiment rather than evidence of behavioural evolution, consequently the NSL-KDD sequence variant requires an explicit ordering justification and any unsupported temporal interpretation is withheld even if the model produces a high numerical score.

When overlapping windows remain within the same training partition they are statistically dependent and can affect uncertainty estimates, while adjacent windows in the test stream also share context, therefore the reporting unit must be the final predicted flow or a declared event and uncertainty should respect appropriate grouping rather than treating every overlapping window as an independent observation, with class supports calculated after the target rule has been applied.

TABLE III DATASET AND REPRESENTATION CONTROLS

Component	Required specification
NSL-KDD	Preserve official split and justify any temporal order
CICIDS2017	Identify files and groups with capture periods
Features	Shared schema with recorded exclusions
Windows	Final flow label with declared stride and grouping
Inference	Apply training fitted transformation at the correct site

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

V CLASSIFIER AND FEDERATED OPTIMISATION

A Candidate convolutional recurrent model

The original measurements identify a CNN–LSTM classifier but do not supply a complete layer configuration or parameter count, while the architecture below is a proposed reproducible candidate for a future validation run, therefore its layer dimensions and normalisation choices must not be interpreted as recovered settings from the experiments that produced the reported tables.

The proposed temporal classifier receives an L by M matrix in which each row represents a flow and each column represents a consistently encoded feature, while a shared projection maps each row into a compact embedding before temporal convolution and recurrent processing, therefore convolution operates along an explicitly defined time axis rather than assuming that arbitrary adjacent feature columns have the spatial relationship found between neighbouring pixels in an image.

A concrete candidate uses a shared dense projection with thirty two outputs followed by a temporal convolution with thirty two channels and a kernel width of three, then a long short term memory layer with thirty two hidden units and a dense layer with sixteen units before class probabilities, while dropout with probability one fifth is applied before the final classifier, therefore Table IV specifies a reproducible starting architecture rather than asserting that this configuration has already produced validated benchmark results [28].

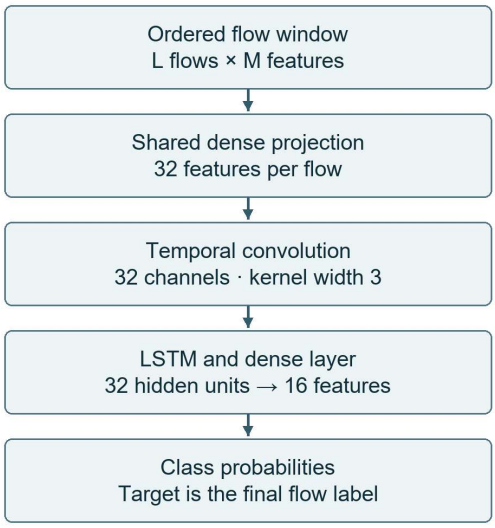


Figure 3 Candidate temporal classifier with a causal final flow target

Layer normalisation is used in this candidate to avoid silently aggregating incompatible batch running statistics, while a batch normalised comparison must specify whether statistics remain local or become part of the released state, consequently a FedBN style alternative can be evaluated separately [22] without being described as an equivalent implementation of the same model because local normalisation state changes both the deployed predictor and the aggregation procedure.

B Learning objective and class imbalance

The local objective is categorical cross entropy over the labels of the final flows in valid windows, while class weights may be defined from public development statistics or fitted privately within each client, therefore the main comparison keeps the weighting policy fixed across all relevant baselines and reports a separate reweighting ablation so that improvements on rare attacks cannot be incorrectly attributed to noise or secure aggregation.

The shared objective uses nonnegative public client weights a_k that sum to one, where equal weights give each organisation the same direct aggregation coefficient while public bounded size weights emphasise larger datasets, consequently the choice expresses an operational preference and a sensitivity constraint at the same time because a larger maximum coefficient permits a larger change in the aggregate when one participant dataset is replaced.

$$F(w) = \sum_k \quad a_k F_k(w) \quad (1)$$

C Local training and update state

The original study specifies ten clients and fifty communication rounds with five local epochs and batches of sixty four examples using stochastic gradient descent at a learning rate of one hundredth, while it also identifies a window length of ten and a clipping threshold of one, therefore these settings describe the reported study configuration although the unavailable seeds and noise calibration prevent reconstruction of its exact training trajectory.

At each round every selected client starts from the same broadcast model and computes a parameter difference after local optimisation, while the server aggregates model differences rather than independently initialised network weights, consequently the implementation must specify whether optimiser state is reset or retained locally and must ensure that all participants interpret the same parameter ordering before contribution clipping and encoding occur.

D Heterogeneity controls

The label partition study includes an IID reference and Dirichlet settings that vary concentration while recording the resulting client sizes and class supports [18], whereas feature heterogeneity is examined through groups or capture conditions where the dataset permits such a split, therefore the analysis distinguishes label skew from feature shift rather than assuming that one Dirichlet parameter represents every difference between organisational networks.

FedProx and SCAFFOLD provide useful optimisation comparisons [19], [20] while DP-SCAFFOLD provides a relevant privacy aware comparison [21], although any additional control variates or released state must be included in its privacy and communication accounting, therefore matching only the number of global rounds is insufficient when methods exchange different objects or perform substantially different amounts of local work.

TABLE IV CANDIDATE MODEL AND INITIAL TRAINING SETTINGS

Component	Proposed setting
Input	L = 10 valid ordered flows with M encoded features

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

Component	Proposed setting
Projection	Shared dense layer with 32 outputs and ReLU
Convolution	Temporal Conv1D with 32 channels and width 3 using same padding and ReLU
Recurrent layer	One LSTM with 32 hidden units and final state output
Normalisation	Layer normalisation on projection outputs
Classifier	Dense 16 with ReLU then dropout probability 0.2 and softmax over J classes
Local training	SGD with learning rate 0.01 and categorical cross entropy
Federation	K = 10 and T = 50 with E = 5 and batch size 64
Privacy	C = 1 as an initial threshold with noise to be calibrated

VI WEIGHTED PRIVATE AGGREGATION

A Contribution bounding and weighting

Let Δw_k denote the parameter difference submitted by a compliant client before protection and let C be a positive clipping threshold, while a zero vector remains unchanged to avoid division by zero, therefore the clipping operation bounds the unweighted update norm before multiplication by the public coefficient and produces a weighted contribution v_k whose norm is at most a_k C [29], as expressed in the following formulation.

$$u_k = \Delta w_k \min\{1, \frac{C}{\|\Delta w_k\|_2}\} \quad (2)$$

$$v_k = a_k u_k, \|v_k\|_2 \le a_k C \quad (3)$$

Under replacement adjacency with fixed weights the two possible clipped updates from the changed client can differ by at most twice the clipping radius, while every other contribution is unchanged conditional on the same previous transcript, therefore the sensitivity of the round aggregate is bounded by twice C times the largest public coefficient, with this conditional argument allowing adaptive training rounds to be composed through an appropriate privacy accountant [10], [17].

$$S = 2C \max_k a_k \quad (4)$$

The order of operations is important because the server needs the sum of already weighted values, whereas applying unequal weights after observing only an unweighted sum cannot recover the required quantity, therefore weighting precedes masking in this design and the sensitivity follows that same order rather than combining an unweighted clipping argument with a different weighted release at the server.

B Noise and the honest participant set

Each compliant noise contributor adds an independent Gaussian vector ξ_k with coordinate variance τ_k squared to its weighted contribution, while colluding clients may disclose their own noise or omit it entirely, therefore the effective noise for the server and colluder coalition is the sum contributed by the honest set H_t and its coordinate variance equals the sum of those honest variances rather than the variance claimed for all registered clients.

$$z_k = v_k + \xi_k, \xi_k \sim N(0, {\tau_k}^2 I) \quad (5)$$

$${v_t}^2 = \sum_{k \in H_t} {\tau_k}^2 \quad (6)$$

A protocol that assumes at least h honest contributors can calibrate independent equal noise shares so their residual variance reaches the required aggregate level when that assumption holds, although this aggregate guarantee does not imply the same bound for an isolated individual message, therefore the collusion threshold and minimum release cohort are privacy parameters that must be enforced operationally and disclosed with the final budget rather than appearing only as limitations after the experiments.

C Secure summation

Clients establish pairwise secret material through an authenticated key agreement procedure and expand it into masks with opposite signs at the two endpoints, while M_k denotes the signed sum of masks for client k and their encoded weighted noisy contributions enter the same modular domain, therefore summing all messages cancels every intact pairwise mask and yields the encoded aggregate as shown in Figure 4 without exposing individual unmasked inputs under the secure aggregation assumptions [13].

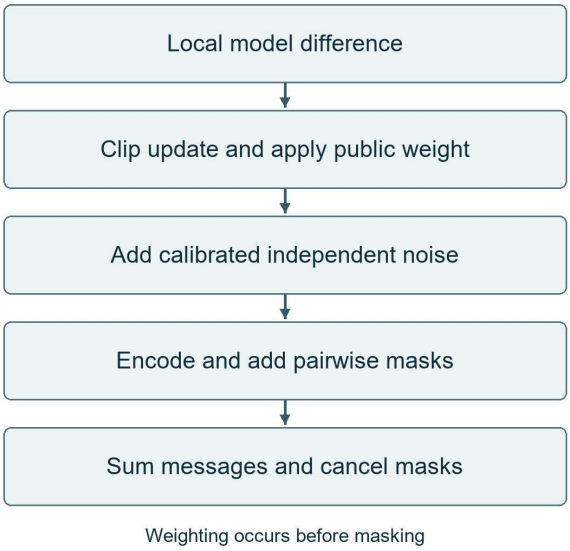


Figure 4 Correct order of weighting and protection before secure summation

$$y_k = \text{Encode}(z_k) + M_k (\text{mod} R) \quad (7)$$

$$\sum_k \quad y_k = \sum_k \quad \text{Encode}(z_k) (\text{mod} R) \quad (8)$$

Mask cancellation alone does not implement dropout recovery because a missing endpoint leaves a mask without its opposite contribution, while practical protocols include additional rounds and recovery information, therefore the system must adopt a complete secure aggregation protocol with declared thresholds and abort rules and must count those messages in its overhead instead of treating a pairwise mask equation as a complete fault tolerant implementation.

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

D Encoding and numerical behaviour

Real valued Gaussian perturbations require a representation compatible with integer modular arithmetic, while rounding and wraparound can change the observable distribution, therefore a practical implementation either supplies an analysis for its finite precision mechanism or adopts an analysed discrete mechanism such as the distributed discrete Gaussian construction [14] rather than assuming that the continuous equations remain exact after arbitrary quantisation.

The aggregate is decoded only when participation and numerical validity requirements are satisfied and the server then applies the declared update learning rate, while aborts and retries are recorded as part of the protocol behaviour, consequently any data dependent retry or diagnostic release is included in the privacy analysis and operational log because an apparently unsuccessful round can still disclose information if messages or partial statistics were made visible.

$$w_{t+1} = w_t + \eta_s \mathrm{Decode}(\sum_k \quad y_k \mathrm{mod} R) \quad (9)$$

VII CONDITIONAL PRIVACY AND SECURITY ANALYSIS

A Continuous Gaussian reference analysis

For an ideal continuous Gaussian round with sensitivity S and residual coordinate noise standard deviation ν_t the Rényi privacy cost at order λ greater than one is bounded by λ times S squared divided by twice ν_t squared [17], while this expression assumes the same covariance under the neighbouring worlds and the declared conditional release model, therefore it is a reference calculation for checking calibration rather than a substitute for analysing the final encoded cryptographic implementation.

$$\rho_t(\lambda) \leq \frac{\lambda S^2}{2\nu_t{}^2} \quad (10)$$

For adaptive rounds whose conditional mechanisms satisfy those bounds the costs add at each order, while conversion to an approximate differential privacy budget uses a chosen failure probability δ [17], therefore the total budget is obtained by minimising the resulting expression over the accountant orders and must cover all released training information rather than considering only the last model or treating every local epoch as an independent server release.

$$\varepsilon(\delta) \leq \min_{\lambda>1}\{\sum_t \quad \rho_t(\lambda) + \frac{\log(\frac{1}{\delta})}{\lambda-1}\} \quad (11)$$

If all rounds use the same aggregate noise multiplier m equal to ν divided by S then a convenient continuous order bound becomes T divided by twice m squared plus twice the square root of that value times log of one over δ, while this follows by minimising the preceding expression and is consistent with a concentrated privacy view [30], therefore it illustrates the cost of repeated full participation without claiming a computed budget for the unimplemented system.

$$b = \frac{T}{2m^2}, \varepsilon \leq b + 2\sqrt{b\log(\frac{1}{\delta})} \quad (12)$$

No amplification from client subsampling is assumed when all ten clients participate every round, while a sampled implementation requires its own selection model and accountant, consequently the intended budget of two with failure probability one in one hundred thousand remains an evaluation target until actual noise and participation values are supplied, with private tuning and any additional released statistics also requiring treatment in the end to end accounting.

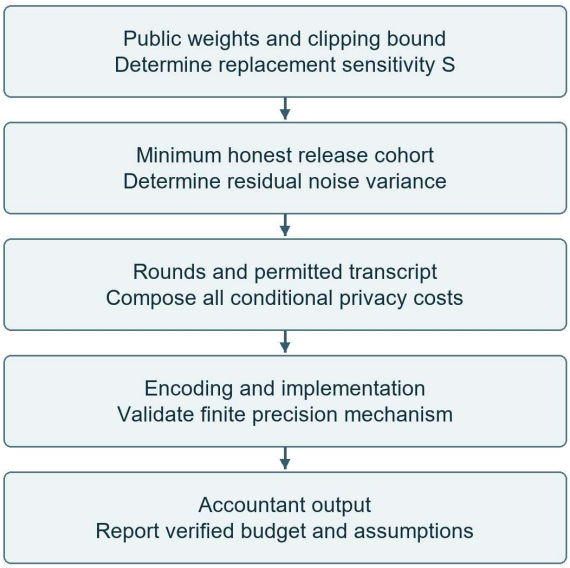


Figure 5 Inputs required to establish a numerical privacy guarantee

B Collusion and inference

Secure aggregation hides individual values only to the extent permitted by the aggregate and adversary assumptions, because a server that learns all but one contribution can subtract those known values and isolate the remaining noisy contribution, therefore a claim of resistance to collusion must specify how many honest participants remain and how their unknown noise preserves the claimed protection under the same adjacency relation rather than asserting that the server always learns nothing about any individual.

Record membership evaluation and client dataset privacy answer related but different questions, while a replacement relation covering arbitrary client datasets can include a change to a record within one client when the rest of the public setup remains fixed, therefore the paper states the exact formal relation and the empirical record attack separately rather than equating a client level budget with a numerical membership success percentage or with secrecy of the public client roster.

C Gradient reconstruction and malicious updates

Gradient reconstruction studies motivate testing the information contained in exposed updates [3], [4] although their quantitative results do not directly predict reconstruction of flow statistics, therefore an IDS evaluation must define valid feature constraints and reconstruction metrics as well as the attacker observations, with the present formulation supporting a restricted visibility argument under secure aggregation rather than an empirical claim that reconstruction is impossible.

Clipping bounds a compliant contribution but provides no evidence that a malicious client followed the clipping operation, while small norm backdoors can remain harmful even when a norm bound is enforced [23], therefore resistance to poisoning would require a compatible integrity mechanism and explicit attack experiments, with robust defences such as

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

FLAME [24] requiring a careful compatibility study before they can be integrated into an aggregation pipeline that conceals individual updates.

D Interpretation of formal protection

Differential privacy bounds how output distributions change under the specified neighbouring relation rather than guaranteeing that every sensitive fact remains unknown, while a coalition can still infer population patterns that hold across many participants, therefore privacy claims should identify the protected contribution and released objects and should avoid treating encrypted transport or local retention as evidence that all legal or organisational obligations have been satisfied [31].

VIII STUDY CONFIGURATION AND VALIDATION PROTOCOL

The available evidence consists of the detection and communication tables in the original conference manuscript together with its stated training settings, while this revision performs arithmetic comparisons of those entries without training a new model or recovering missing observations, therefore all measured values below are labelled as originally reported and all differences and ratios are identified as calculations from those values.

The reported study uses NSL-KDD and CICIDS2017 with ten simulated clients and describes a Dirichlet label allocation at concentration one half, while it compares a centralised CNN–LSTM and nonprivate FedAvg with the protected configuration at a target privacy budget of two and failure probability one in one hundred thousand, consequently the Results section covers those three configurations and the target budget remains unverified because the original account does not provide the noise multiplier or accountant output.

A Baselines and controlled comparisons

The reported experiment compares centralised training and ordinary FedAvg with the full protected configuration, whereas clipping only and noise only variants as well as secure aggregation without noise remain proposed validation ablations rather than completed experiments, therefore Figure 6 summarises the broader evaluation required to establish causal effects and its additional configurations are not attributed measured outcomes in the present Results section.

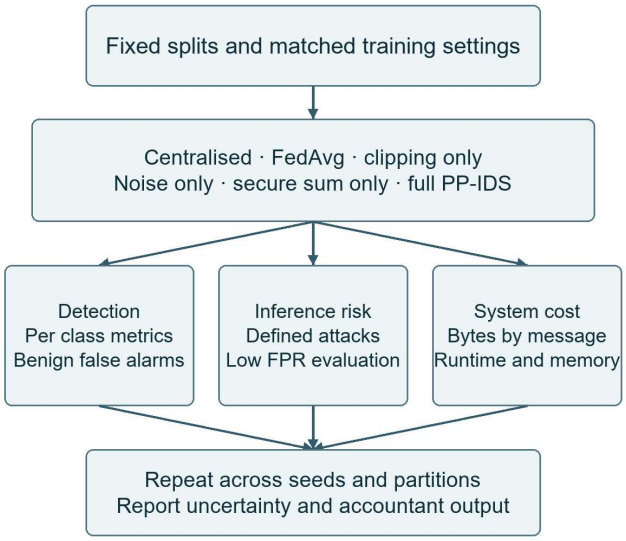


Figure 6 Controlled evaluation of detection utility and privacy attacks with complete system costs

The centralised model provides a practical reference for pooled training although it is not a mathematical accuracy upper bound, while an individual flow model tests whether temporal context adds useful information and privacy aware federated comparisons test whether the selected optimiser is appropriate [21], consequently a fair assessment reports both matched privacy budgets and computational effort because two systems can reach different utilities through different numbers of private releases or different amounts of local training.

B Detection measures and operating points

Detection reporting includes per class precision and recall with support counts plus macro F1 and the confusion matrix, while benign false positive rate provides an operational view of alert burden, therefore an aggregate recall value is not interpreted as the fraction of all attacks detected unless the averaging and target population justify that interpretation, especially when rare classes contribute little to a weighted metric despite being important to the deployment objective.

The decision threshold for binary attack alerts is selected on validation data and fixed before test evaluation, while multiclass reporting retains the full label space and explains any mapping to an attack versus benign decision, consequently the paper records both class prediction quality and alert behaviour in keeping with the evaluation concerns discussed by Khare and colleagues [16] rather than changing the target definition after observing which presentation gives the most favourable headline score.

C Membership inference procedure

The empirical privacy evaluation includes a loss threshold attack and a stronger reference model based attack where computationally feasible [5], [6], while member and nonmember examples follow matched eligibility rules with their selection disclosed, therefore an attacker must not gain trivial separation from different capture days or label distributions unless that is explicitly part of the intended threat scenario and the resulting confounding is analysed.

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

Attack performance is reported through receiver operating characteristic curves and true positive rates at fixed low false positive rates together with ROC AUC, while balanced attack accuracy includes a fifty percent chance baseline and attack advantage is labelled separately if used, consequently a number below chance is investigated for metric definition or reversed decisions rather than automatically presented as evidence of strong privacy.

Attack training and threshold fitting use separate development information and do not optimise on the final attack test set, while results are repeated over suitable target models and attack samples, therefore uncertainty reflects the training and attack procedures and the disclosure budget is reported beside those measurements so that empirical attack weakness is not substituted for a formal privacy guarantee.

D Communication and computation

Communication logging records actual transmitted bytes by direction and message type, including model downloads and protected uploads as well as authentication and key establishment and dropout recovery, while the unit and aggregation scope are stated explicitly, therefore an overhead percentage is computed from complete matched runs and cannot be inferred merely from the fact that clipping and noise preserve vector length.

Runtime reporting separates local preprocessing and optimisation from cryptographic preparation and server aggregation, while memory and inference latency are measured on the stated hardware under a declared workload, consequently the term lightweight is supported by model size and device measurements rather than by a qualitative architecture description, with energy or gateway suitability remaining outside the conclusion unless those measurements are actually performed.

E Repetition and sensitivity studies

Experiments repeat across multiple training seeds and independently generated client partitions using paired configurations wherever possible, while uncertainty intervals and the number of repetitions are reported with each main metric, therefore small differences in accuracy are interpreted relative to variability rather than presented as stable effects based on a single random initialisation and a single allocation of attack classes.

The planned sensitivity study varies the privacy budget and clipping threshold together with client count and heterogeneity, while keeping the interpretation of the protected unit fixed, consequently each setting requires a fresh accountant output and an explicit minimum honest cohort because a budget that holds for a large federation cannot be assigned unchanged to a smaller cohort after dropout or collusion.

TABLE V EXPERIMENTS AND REQUIRED REPORTING

Experiment	Controls and outputs
Detection	Per class metrics and support with benign false positive rate
Privacy ablations	Clipping and noise with secure sum isolated
Membership	Attack access and ROC with TPR at fixed FPR
Heterogeneity	IID and multiple label skew partitions
Privacy sweep	Recompute accountant for each noise and cohort setting
Communication	Actual uploads and downloads with cryptographic traffic
Repetition	Paired seeds and partitions with uncertainty

IX RESULTS AND DISCUSSION

A Reported detection performance

Table VI reproduces the detection values from the original study for the three reported configurations on both datasets, while Figure 7 shows their accuracy comparison without uncertainty bars because repeated run outputs are unavailable, therefore the values describe the submitted measurements and must not be read as independently reproduced outcomes or as performance of the newly specified candidate architecture.

TABLE VI ORIGINALLY REPORTED DETECTION METRICS IN PERCENT

Model and dataset	Acc	Prec	Rec	F1
Central NSL-KDD	98.6	98.1	98.4	98.2
FedAvg NSL-KDD	97.9	97.3	97.6	97.4
PP-IDS NSL-KDD	97.3	96.8	97.0	96.9
Central CICIDS2017	99.1	98.8	98.7	98.7
FedAvg CICIDS2017	98.2	97.9	97.7	97.8
PP-IDS CICIDS2017	97.6	97.1	97.0	97.0

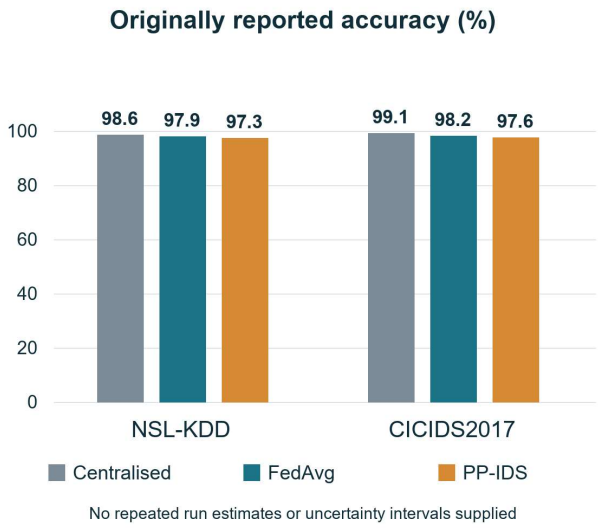


Figure 7 Accuracy values reproduced from the original study with no uncertainty estimates available

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

On NSL-KDD the reported protected accuracy is 97.3% compared with 97.9% for ordinary FedAvg and 98.6% for centralised training, yielding calculated gaps of 0.6 and 1.3 percentage points respectively, while on CICIDS2017 the corresponding values are 97.6% and 98.2% and 99.1% with calculated gaps of 0.6 and 1.5 percentage points, consequently both datasets show the same reported absolute difference between the protected and ordinary federated configurations although the consistency of that difference across random runs is unknown.

The reported F1 values for PP-IDS are 96.9% on NSL-KDD and 97.0% on CICIDS2017 compared with 97.4% and 97.8% for ordinary FedAvg, yielding calculated differences of 0.5 and 0.8 percentage points, while protected precision is 96.8% and 97.1% and the table gives recall of 97.0% for each dataset, therefore these aggregate values suggest a reported utility cost but cannot establish rare attack sensitivity because the original account does not define the averaging convention or provide per class supports.

The centralised to ordinary federated accuracy differences are 0.7 percentage points on NSL-KDD and 0.9 percentage points on CICIDS2017, while the additional protected configuration difference is 0.6 percentage points on each dataset, therefore the total reported gap can be described arithmetically in two stages although the available three way comparison cannot isolate the effects of clipping and noise from reweighting or other implementation differences.

B Communication measurements

Table VII reproduces the original communication totals and the original membership inference indicator, while Figure 8 visualises the communication comparison, therefore the two cost values are presented as study level totals because their attribution to a particular dataset and the inclusion of uploads and downloads and cryptographic setup are not documented in the original measurement description.

TABLE VII ORIGINALLY REPORTED COST AND INFERENCE INDICATOR

Configuration	Cost MB	MI indicator %
Centralised	Not reported	38.4
FedAvg	142	34.1
PP-IDS	148	9.7

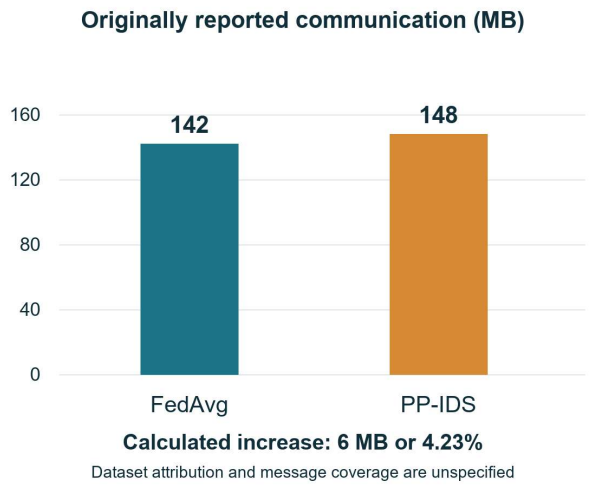


Figure 8 Reported communication totals with the relative increase calculated from those totals

The reported cost increases from 142 MB for ordinary FedAvg to 148 MB for PP-IDS, giving an absolute difference of 6 MB and a calculated relative increase of 6 divided by 142 multiplied by one hundred or approximately 4.23%, while the source attributes the difference to secure aggregation, therefore the result supports a descriptive comparison of the supplied totals but does not independently establish complete protocol overhead because message level logs and recovery traffic are unavailable.

No communication total is reported for pooled centralised training and no equivalent raw traffic transfer measurement is supplied, while cryptographic computation and device energy are also unmeasured, consequently the present comparison cannot establish that the protected federation is cheaper than central collection or suitable for a particular gateway solely from the difference between its two reported federated totals.

C Membership inference indicator

The original table gives a membership inference indicator of 38.4% for centralised training and 34.1% for ordinary FedAvg compared with 9.7% for PP-IDS, yielding a calculated reduction of 24.4 percentage points or approximately 71.55% relative to ordinary FedAvg, while the exact attack and success definition are absent, therefore this reduction is a numerical property of the supplied entries rather than a validated estimate of privacy protection.

If the indicator were balanced member versus nonmember classification accuracy then its chance baseline would be fifty percent and a value of 9.7% would require investigation of the decision direction or metric definition, whereas a true positive rate or attack advantage would require a different interpretation [5], [6], consequently the table retains the original indicator label for traceability but no conclusion about attack resistance is drawn until its definition and sampling and threshold procedure are recovered.

D Privacy budget and security interpretation

The target privacy budget of two with failure probability one in one hundred thousand is stated in the original experiment configuration but cannot be calculated from the available information, while the corrected weighted aggregation and

International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

finite precision requirements in Sections VI and VII introduce specifications that the original text does not establish as implemented, therefore the reported detection scores must not be presented as verified performance at that privacy budget or as evidence validating the corrected release mechanism.

Neither gradient reconstruction nor poisoning is empirically evaluated in the available results, while client side clipping can be bypassed by a malicious participant unless enforced, consequently the present result tables address reported classification and communication with an undefined inference indicator and do not establish resilience against inversion or backdoors or collusion beyond the explicitly conditional analysis.

E Implications and validation priorities

Taken at face value the supplied detection values place the protected configuration within one and a half percentage points of the centralised reference on both datasets, while the supplied cost totals differ by approximately four percent, therefore they motivate a reproducible investigation of the protection and utility trade off but remain insufficient for an operational recommendation because statistical variability and split integrity and complete cost accounting have not been established.

The immediate validation priorities are recovery of the architecture and run configuration together with the noise multiplier and accountant output, followed by leakage checks and paired repeated experiments with defined membership attacks, while the rare class and drift concerns discussed by Khare and colleagues remain relevant to the interpretation of aggregate detection quality [16], consequently a verified rerun could either confirm these preliminary comparisons or show that some reported differences arise from the evaluation procedure.

An organisation evaluating the framework also needs authenticated participation and a compatible feature schema together with an explicit model release policy, while compression and sparsification may change both the contribution transformation and its privacy sensitivity as considered in prior private federation research [32], therefore deployment extensions require separate measurements and accounting rather than being inferred from the present benchmark tables.

F Comparison with independently published results

Table VIII presents results attributed to two independent publications rather than additional PP-IDS experiments, while the cited values retain their original evaluation context and are not pooled with our reported measurements, therefore the comparison provides external context for detection and privacy evaluation without constituting a matched benchmark or establishing that one method outperforms another.

TABLE VIII PUBLISHED RESULTS FROM INDEPENDENT STUDIES

External study	Reported result	Evaluation context
Federated mimic learning [33]	NSL-KDD accuracy 98.11%	Student mimic model with ten users and a 90/10 study split
SecureTrust-FL [34]	CICIDS2017 accuracy 93.93% and weighted F1 94.60%	Binary classification with three datasets as three clients

Al-Marri and colleagues report 98.11% accuracy for federated student mimic learning on NSL-KDD [33], using a study split of ninety percent training and ten percent testing with ten users and twenty rounds alongside teacher and student data separation, while their FL classifier table records zero U2R recall despite high aggregate accuracy, therefore the study illustrates why a headline score must be assessed alongside minority class detection and split construction rather than directly ranked against our reported 97.3% NSL-KDD accuracy.

Alshammari and colleagues report 93.93% accuracy and 94.60% weighted F1 on the CICIDS2017 client in SecureTrust-FL [34], using binary classification with three datasets represented as three clients, while their separate noise experiment gives accuracies of 90.76% without noise and 91.00% at a noise scale of 0.001 compared with 88.52% at 0.01, therefore their results motivate a privacy utility sweep but these separate experiments do not establish a matched privacy budget or a directly comparable baseline for our reported 97.6% CICIDS2017 accuracy.

The externally reported scores describe different learning tasks and evaluation partitions as well as different privacy mechanisms, while a noise scale alone cannot be equated with a composed privacy budget and mimic learning does not automatically establish differential privacy, consequently the published figures remain contextual evidence and a defensible superiority claim would require implementation under a shared split and feature schema with matched labels and verified privacy assumptions.

The practical implication for the present study is to preserve per class reporting and evaluate several noise settings with uncertainty, while using separately documented teacher or student data whenever knowledge transfer is considered, therefore the literature comparison strengthens the validation agenda but does not fill missing PP-IDS measurements or convert published observations into results produced by our own system.

X LIMITATIONS AND FUTURE WORK

The available experimental evidence is restricted to aggregate values reported in the original manuscript without run logs or prediction files, while the architecture and membership attack remain incompletely specified and the corrected privacy mechanism has not been rerun, therefore the present study cannot quantify uncertainty or verify leakage free splitting or confirm that the claimed privacy target applies to the reported model outputs.

The security scope does not include a fully malicious coordinating server or enforced validation of every hidden contribution, while real networks may contain adaptive poisoning and compromised gateways as well as inference time evasion, consequently a future integrity extension must

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

preserve the required aggregation confidentiality while proving or testing the restrictions it places on inputs instead of assuming that compliant software behaviour extends to adversarial clients.

Further work should prioritise the finite precision privacy implementation and reproducible experiments before expanding to asynchronous or hierarchical federation, while additional datasets and real device trials can then test whether the method remains useful under greater variation, therefore personalised models and robust aggregation are evaluated as separate extensions with their own released state and privacy costs rather than being listed as automatic improvements to the current design.

XI CONCLUSION

The original study tables report PP-IDS accuracies of 97.3% on NSL-KDD and 97.6% on CICIDS2017 with an accuracy difference of 0.6 percentage points from ordinary FedAvg on each dataset, while the reported communication totals increase from 142 MB to 148 MB or approximately 4.23%, therefore the descriptive comparison indicates a modest reported utility and communication cost for the protected configuration under the supplied settings without establishing statistical significance or independently reproduced performance.

The reported membership inference decrease cannot establish attack resistance until its metric is defined and the numerical privacy target remains unverified without noise calibration and accountant output, while the corrected weighted aggregation formulation requires an actual implementation and rerun, consequently the results motivate further validation of private collaborative intrusion detection rather than a definitive claim that the reported utility was achieved under a proven privacy guarantee.

DECLARATIONS

Data availability: The result values reproduced here come from Tables II and III of the supplied conference manuscript titled Privacy Preserving Intrusion Detection in Distributed Networks Using Federated Learning with Differential Privacy and Secure Aggregation, while no underlying experiment logs or prediction files were available for this analysis and the benchmark datasets remain accessible through their providers [25], [26].

Funding and conflicts of interest: Statements remain to be supplied by the authors before submission because the available source material does not establish funding arrangements or competing interests.

Author contributions: Individual contributor roles require confirmation by the authors before submission so no specific role or corresponding author designation is assigned in this revision.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” AISTATS, PMLR 54, pp 1273–1282, 2017. Source

[2] P. Kairouz et al., “Advances and Open Problems in Federated Learning,” arXiv:1912.04977, 2019. Source

[3] L. Zhu, Z. Liu and S. Han, “Deep Leakage from Gradients,” NeurIPS, vol 32, 2019. Source

[4] J. Geiping, H. Bauermeister, H. Dröge and M. Moeller, “Inverting Gradients — How easy is it to break privacy in federated learning?,” NeurIPS, 2020. Source

[5] R. Shokri, M. Stronati, C. Song and V. Shmatikov, “Membership Inference Attacks against Machine Learning Models,” IEEE Symposium on Security and Privacy, 2017. Source

[6] N. Carlini, S. Chien, M. Nasr, S. Song, A. Terzis and F. Tramèr, “Membership Inference Attacks From First Principles,” IEEE Symposium on Security and Privacy, 2022. Source

[7] S. A. Rahman, H. Tout, C. Talhi and A. Mourad, “Internet of Things Intrusion Detection: Centralized, On-Device, or Federated Learning?,” IEEE Network, vol 34, no 6, 2020. Source

[8] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan and A. R. Sadeghi, “DÏoT: A Federated Self-learning Anomaly Detection System for IoT,” IEEE ICDCS, pp 756–767, 2019. Source

[9] B. Li, Y. Wu, J. Song, R. Lu, T. Li and L. Zhao, “DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems,” IEEE Transactions on Industrial Informatics, vol 17, no 8, pp 5615–5624, 2021. Source

[10] C. Dwork and A. Roth, “The Algorithmic Foundations of Differential Privacy,” Foundations and Trends in Theoretical Computer Science, vol 9, pp 211–407, 2014. Source

[11] M. Abadi et al., “Deep Learning with Differential Privacy,” ACM CCS, pp 308–318, 2016. Source

[12] R. C. Geyer, T. Klein and M. Nabi, “Differentially Private Federated Learning: A Client Level Perspective,” arXiv:1712.07557, 2017. Source

[13] K. Bonawitz et al., “Practical Secure Aggregation for Privacy-Preserving Machine Learning,” ACM CCS, pp 1175–1191, 2017. Source

[14] P. Kairouz, Z. Liu and T. Steinke, “The Distributed Discrete Gaussian Mechanism for Federated Learning with Secure Aggregation,” ICML, PMLR 139, 2021. Source

[15] M. K. Bagwani, A. Gangwar, K. Vishwakarma and V. K. Tiwari, “Real-time signature-based detection and prevention of DDOS attacks in cloud environments,” International Journal of Science and Research Archive, vol 12, no 2, pp 2929–2935, 2024, doi: 10.30574/ijsra.2024.12.2.1608. Source

[16] A. Khare, N. Choudhary and M. K. Bagwani, “Trustworthy and Adaptive Intrusion Detection for Imbalanced IoT Traffic,” International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies, vol 1, no 1, pp 80–86, 2026, doi: 10.68307/k9pxqe13. Source

[17] I. Mironov, “Rényi Differential Privacy,” IEEE Computer Security Foundations Symposium, 2017. Source

[18] T. M. H. Hsu, H. Qi and M. Brown, “Measuring the Effects of Non-Identical Data Distribution for Federated Visual Classification,” arXiv:1909.06335, 2019. Source

[19] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar and V. Smith, “Federated Optimization in Heterogeneous Networks,” Proceedings of MLSys, 2020. Source

[20] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich and A. T. Suresh, “SCAFFOLD: Stochastic Controlled Averaging for Federated Learning,” ICML, PMLR 119, pp 5132–5143, 2020. Source

[21] M. Noble, A. Bellet and A. Dieuleveut, “Differentially Private Federated Learning on Heterogeneous Data,” AISTATS, PMLR 151, pp 10110–10145, 2022. Source

ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Research Article | www.ijaicet.com | DOI: https://doi.org/10.68307/5ercgy27

[22] X. Li, M. Jiang, X. Zhang, M. Kamp and Q. Dou, "FedBN: Federated Learning on Non-IID Features via Local Batch Normalization," ICLR, 2021. Source

[23] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin and V. Shmatikov, "How To Backdoor Federated Learning," AISTATS, PMLR 108, pp 2938–2948, 2020. Source

[24] T. D. Nguyen et al., "FLAME: Taming Backdoors in Federated Learning," USENIX Security, pp 1415–1432, 2022. Source

[25] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," IEEE CISDA, pp 1–6, 2009; dataset documentation. Source

[26] I. Sharafaldin, A. H. Lashkari and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," ICISSP, pp 108–116, 2018; dataset documentation. Source

[27] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," MilCIS, pp 1–6, 2015; dataset documentation. Source

[28] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol 9, no 8, pp 1735–1780, 1997. Source

[29] X. Zhang, X. Chen, M. Hong, Z. S. Wu and J. Yi, "Understanding Clipping for Federated Learning: Convergence and Client-Level Differential Privacy," ICML, PMLR 162, 2022. Source

[30] M. Bun and T. Steinke, "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds," arXiv:1605.02065, 2016. Source

[31] N. Truong, K. Sun, S. Wang, F. Guitton and Y. Guo, "Privacy Preservation in Federated Learning: An Insightful Survey from the GDPR Perspective," Computers & Security, vol 110, article 102402, 2021. Source

[32] A. Cheng, P. Wang, X. S. Zhang and J. Cheng, "Differentially Private Federated Learning With Local Regularization and Sparsification," IEEE/CVF CVPR, pp 10122–10131, 2022. Source

[33] N. A. Al-Athba Al-Marri, B. S. Ciftler and M. M. Abdallah, "Federated Mimic Learning for Privacy Preserving Intrusion Detection," arXiv:2012.06974, 2020, Sections IV–V and conclusion. Source

[34] N. S. Alshammari, S. Mishra, M. Rathi, N. Goel and S. Tahzib, "SecureTrust-FL: trust-aware privacy-preserving federated learning for network intrusion detection," Scientific Reports, vol 16, article 28050, 2026, Tables 4–5, doi: 10.1038/s41598-026-58383-4. Source