

Robust Federated Learning with Integrated Access Control for Distributed Edge Cloud Big Data Systems

Vinod K. Barupati^{1*}, Shivshankar M. Nagrik²

¹ IT professional, JS university, Uttar Pradesh, India.

² Dept. of Pharmaceutics, Datta Meghe Institute of Higher Education and Research (Deemed to be University), Sawangi (M), Wardha, Maharashtra, India.

ABSTRACT

Federated learning is a distributed machine learning paradigm in which participating devices or edge nodes train a shared model without transferring their raw data to a central repository. This property makes federated learning attractive for edge cloud big data environments where data are geographically distributed, continuously generated, heterogeneous, and subject to privacy requirements. However, decentralization also introduces security and governance challenges. An attacker may impersonate a participant, submit poisoned model updates, exploit weak permissions, infer private information from model updates, or compromise an edge node. Robust federated learning therefore requires more than an aggregation algorithm; it requires coordinated access control, privacy protection, participant validation, secure communication, and resilience against malicious behavior. This review examines the integration of federated learning with access control for distributed edge cloud big data systems. It discusses the foundations of federated learning, edge intelligence, role based and attribute based access control, threat models, robust aggregation, secure aggregation, privacy preservation, communication efficiency, scalability, and application requirements. The review also compares major security mechanisms and identifies research directions for adaptive authorization, decentralized trust, privacy preserving identity management, resource aware robustness, and explainable security decisions. An integrated architecture is proposed in which identity, role, attributes, device state, data sensitivity, and training behaviour are jointly considered before model updates are accepted. Such an approach can provide a practical security layer for collaborative learning while retaining the principal benefit of federated learning: keeping sensitive data close to their source.

Keywords: Federated learning, Access control, Edge computing, Big data, Robust aggregation, Privacy preservation, Distributed systems

INTRODUCTION

The rapid growth of connected devices, cloud services, sensors, mobile applications, industrial systems, and intelligent infrastructure has created large volumes of distributed data. Much of this information is generated at the edge rather than inside a centralized data centre. Traditional centralized machine learning requires data to be collected and transferred to a common training location, but this approach can create privacy, bandwidth, latency, ownership, and governance problems [1]. Federated learning addresses part of this challenge by allowing local participants to train a shared model and send model updates instead of raw training records. McMahan and colleagues introduced Federated Averaging as a practical ap-

proach for learning from decentralized and non independent and identically distributed data while reducing communication requirements. The edge cloud environment extends this idea by placing computation and intelligence close to data sources while retaining access to cloud scale storage and processing[2,3]. Edge computing can reduce latency and network traffic, but it also creates a larger and more heterogeneous attack surface. Devices differ in hardware capability, operating system, network quality, trust level, and administrative ownership. Therefore, a federated system cannot assume that every participating node is equally trustworthy. A participant may be legitimate but compromised, or it may be authorized for one task but not another. Access control provides a mechanism for deciding who or what is

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



permitted to perform an operation on a resource. Role based access control associates permissions with roles, while attribute based access control can use multiple contextual properties such as device type, location, data sensitivity, time, or organizational affiliation. In a federated learning system, access control can be extended beyond ordinary data access. It can govern who may join a training round, which model a participant may receive, which data category may contribute to training, which aggregation service may be used, and whether a submitted update should be admitted [4,5]. The central challenge is therefore integration. Robust aggregation can reduce the influence of malicious updates, but it does not prove that the sender was authorized. Conversely, strong authorization does not prevent a legitimate but compromised participant from submitting poisoned updates. Secure aggregation can hide individual updates from the server, but this privacy property can make individual anomaly detection more difficult. A practical architecture must combine these mechanisms rather than treating them as isolated security controls [6,7].

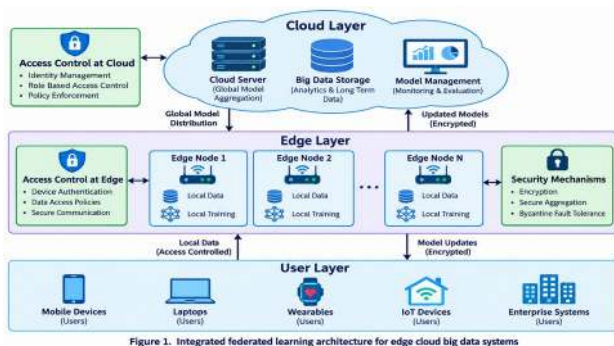


Figure 1: Integrated Federated Learning Architecture for Edge Cloud Big Data Systems

Figure 1: The conceptual workflow distributed edge nodes retain local big data and perform local training. Before participation, identity and authorization policies are evaluated. During aggregation, secure communication and privacy mechanisms protect model exchange, while robust aggregation and anomaly detection limit the effect of malicious updates. The cloud layer provides coordination and long term model management rather than requiring raw data centralization[9,10].

2. FUNDAMENTALS OF FEDERATED LEARNING IN DISTRIBUTED EDGE CLOUD SYSTEMS

Federated learning generally consists of a set of clients, an aggregation service, and a communication protocol. A client receives a current global model, trains it using local data, and returns an update. The server combines updates to construct a new global model. In the classical Federated Averaging method, client contributions are weighted according to local sample counts. The approach reduces direct movement of raw data, but model updates can still contain information about local data and can be manipulated by adversaries [11,12]. Edge environments make the federated setting more complex. Clients may have limited battery, memory, computation, and wireless bandwidth. Data distributions are frequently non independent and identically distributed because devices observe different populations or operating conditions. Li and colleagues highlighted the importance of heterogeneous networks and federated optimization, while Karimireddy and colleagues introduced SCAFFOLD to address client drift caused by statistical heterogeneity. Adaptive federated optimization methods further demonstrate that optimizer design can affect convergence under heterogeneous participation [13,14]. Federated learning can be synchronous, asynchronous, hierarchical, or personalized. In synchronous training, the server waits for a selected group of clients. In asynchronous designs, updates can arrive at different times. Hierarchical designs can use edge aggregators before cloud aggregation, which may reduce communication overhead and improve locality. Personalized federated learning allows related but non identical models for different participants [15]. These variations affect access control because authorization may need to operate at several levels rather than only at the final cloud server. The basic advantage of federated learning is not equivalent to complete privacy. Raw data may remain local while gradients or model parameters reveal information. Research on gradient leakage, membership inference, and model inversion demonstrates that model information can sometimes expose characteristics of training data. Consequently, a secure edge cloud federated system should treat model updates as potentially sensitive resource [16,17] .

3. ACCESS CONTROL REQUIREMENTS AND SECURITY MODELS

Access control in a federated environment should answer several questions: Is the participant authentic, Is the device currently trusted, Is the participant authorized for the requested model or task, Is the data category permitted for the training objective, Is the participant allowed to submit updates in the current round, What should happen if its behaviour changes, These questions require authorization policies that can operate dynamically [17,18]. Role based access control is useful when organizational responsibilities are stable. For example, an edge gateway, hospital analyst, industrial controller, and cloud administrator can receive different permissions. Attribute based

access control provides more flexibility because decisions can combine user, device, resource, action, and environmental attributes. A hybrid policy can therefore grant a training privilege only when the user role, device security state, data classification, model purpose, and session context satisfy defined conditions [19]. For federated learning, access control should cover the complete model lifecycle. Registration determines whether an entity may join. Model distribution determines which model version it can receive. Local training policies determine which data sources and computation tasks are allowed. Update submission determines whether an update is accepted. Aggregation determines whether a participant may influence the global model. Audit mechanisms record security relevant events so that suspicious behaviour can be investigated [20].

Table 1: Access Control Requirements for Federated Edge Cloud Big Data Workflows

Sr. No	Control element	Decision input	Purpose	Example
1	Identity authentication	User, device, certificate	Verify participant identity	Registered edge device [21]
2	Role authorization	Role and permissions	Limit available actions	Trainer, aggregator, auditor [22]
3	Attribute policy	Data, device, context attributes	Fine grained decisions	Sensitive data requires trusted device [23]
4	Model access	Model ID and task	Prevent unauthorized model use	Only approved disease model [24]
5	Update admission	Behaviour and update quality	Reduce malicious influence	Reject anomalous update [25]
6	Audit	Event and policy logs	Accountability and investigation	Record training round decisions [26]

4. THREAT MODEL AND SECURITY CHALLENGES

A useful threat model distinguishes honest, curious, compromised, and malicious participants. An honest participant follows the protocol. A curious server follows the protocol but attempts to infer private information. A compromised client may have valid credentials but behave maliciously. A malicious participant may attempt unauthorized access, model poisoning, backdoor insertion, Sybil participation, denial of service, or identity abuse [27]. Model poisoning is particularly important because the attacker changes the model update rather than directly modifying the central database. Backdoor attacks can cause a model to behave incorrectly for selected inputs while retaining apparently normal overall performance. Byzantine robust learning addresses arbitrary client

behaviour through robust aggregation rules such as coordinate wise median or trimmed mean, but subsequent work has shown that convergence alone does not guarantee resistance to all high dimensional attacks [28]. Access control itself has threats. Stolen credentials, compromised devices, excessive permissions, weak revocation, policy misconfiguration, and collusion can undermine a security architecture. A system that grants permanent access based only on initial registration may fail when a device becomes compromised. Continuous authorization is therefore relevant for edge environments in which device state and network conditions can change quickly [29,30]. Privacy attacks represent another category. Membership inference attempts to determine whether particular records were used for training. Gradient leakage demonstrates that information about local training examples can sometimes be reconstructed from

shared updates. Differential privacy can limit such leakage by adding calibrated noise, while secure aggregation prevents the server from seeing individual client updates in the clear. These mechanisms have different privacy and utility costs and should be selected according to application requirements [31,32].

5. ROBUST AGGREGATION AND BYZANTINE RESILIENCE

Fed Avg is effective under many practical conditions, but simple averaging is vulnerable when a minority of clients submit extreme or strategically manipulated updates. Robust aggregation replaces or augments averaging with methods designed to reduce the influence of outliers. Median and trimmed mean methods are among the classical robust approaches studied in distributed learning. Krum and related rules use distances between client updates to identify contributions that are more consistent with the majority [33]. The choice of robust aggregation depends on the threat model, dimensionality, number of participating clients, and expected data heterogeneity. A naive outlier detector can mistakenly classify legitimate non independent and identically distributed updates as malicious. For example, a device serving a rare population may naturally produce an update that differs from the majority. Therefore, robust security should combine statistical evidence with authorization context and historical behaviour rather than relying on a single distance threshold [34]. A practical architecture can maintain an update trust score based on historical participation, policy compliance, update norm, similarity to peer updates, validation performance, and device security state. Such a score should not permanently label a participant from one event. Instead, it can trigger additional verification, reduced influence, or temporary quarantine. This design connects access control with robust learning while allowing legitimate participants to recover after anomalous network or data conditions [35]. Robust aggregation also has a privacy trade off. Secure aggregation hides individual updates, which makes server side inspection of each update difficult. One solution is to perform privacy preserving pre validation at trusted edge aggregators or use secure protocols that reveal only aggregate statistics needed for anomaly detection. Designing such mechanisms without recreating a centralized data collection point remains an open problem. Ro-

bust aggregation can also be improved through temporal analysis. Instead of evaluating each update independently, the system can compare the current contribution with the client's previous contributions [36,37]. A sudden change in update direction, magnitude, or model behavior may indicate compromise or poisoning. Temporal analysis is particularly useful in edge environments where devices repeatedly participate in training. Nevertheless, sudden changes may also occur because of legitimate changes in local data, workload, environment, or operating conditions. Therefore, temporal behavior should be combined with contextual information rather than used as an independent decision rule [38].

6. PRIVACY PRESERVATION AND SECURE AGGREGATION

Federated learning reduces the need to centralize raw data but does not automatically provide differential privacy or cryptographic confidentiality. Secure aggregation is designed to allow a coordinator to compute an aggregate without learning each participant's individual contribution. Bonawitz and colleagues proposed a practical secure aggregation protocol designed for high dimensional vectors, client dropout, and communication constraints [39]. Differential privacy provides another protection layer by limiting the effect of any individual training record on the released result. Deep learning with differential privacy demonstrated how gradient clipping and noise can be incorporated into training. In federated learning, privacy budgets must be managed across multiple rounds. Excessive noise can reduce model accuracy, while weak noise may provide insufficient protection. The appropriate balance depends on data sensitivity and application risk [40]. Gradient leakage research shows why access control should treat updates as sensitive. A participant may be authorized to train but still have its update exposed to an unauthorized observer. Encryption, secure aggregation, trusted execution environments, and differential privacy can reduce this risk. However, each introduces computational, communication, or implementation overhead. Edge devices require resource aware privacy mechanisms because a method that is practical in a cloud server may be too expensive for low power devices [41]. Privacy protection should also include metadata. Participation timing, device identity, update frequency, and model version can reveal information

even when model parameters are protected. Therefore, a mature architecture should minimize unnecessary metadata, enforce least privilege, protect logs, and define retention periods [42].

7. INTEGRATED ACCESS CONTROL ARCHITECTURE

An integrated architecture can be organized into five functional layers: identity and trust, policy management, local learning, secure communication, and robust aggregation. The identity layer authenticates users and devices using certificates, cryptographic keys, or organizational identity services. The policy layer evaluates roles and attributes. The local learning layer controls which model and data sources may be used. The communication layer protects messages and model transfers. The aggregation layer validates and combines updates [43]. The policy decision can be represented conceptually as an authorization function that considers subject attributes, device attributes, resource attributes, action, and environment. For example, a participant may be allowed to contribute to a low sensitivity model from a managed device, while the same participant may be denied access to a highly sensitive model unless additional conditions are satisfied. Such context aware authorization is more flexible than a simple user identity check [44]. Dynamic revocation is essential. If a device reports compromise, repeated authentication failures, abnormal update behaviour, or policy violations, its authorization can be suspended without waiting for the end of a long training campaign. The architecture should also support delegated edge authorization so that every request does not have to travel to a distant cloud service. Auditability provides the final layer. Policy decisions, model versions, participant identifiers, security events, and aggregation outcomes can be recorded in tamper resistant logs. The audit trail should avoid storing unnecessary sensitive content. A useful audit record can contain a pseudonymous participant identifier, decision result, policy version, time, model version, and reason code [45].

8. COMMUNICATION, RESOURCE, AND SCALABILITY CONSIDERATIONS

Communication is one of the main constraints in federated learning. McMahan and colleagues showed that local training can substantially reduce the number

of communication rounds compared with synchronized stochastic gradient descent. Nevertheless, edge cloud systems may contain thousands or millions of devices, and each device may have limited bandwidth. Compression, quantization, client selection, local computation, and hierarchical aggregation can reduce communication burden [46]. Resource heterogeneity affects security as well as learning. Cryptographic protocols and differential privacy can increase computation and memory use. A resource aware architecture should therefore adapt security mechanisms to device capability without weakening the required security level. More capable edge gateways can perform heavier cryptographic or validation tasks, while constrained sensors may use lightweight authentication and delegated processing. Wireless federated learning introduces additional variability because channel quality, mobility, and interference influence participation. Resource allocation methods can jointly consider communication quality and learning contribution. Client selection should not systematically exclude devices with poor connectivity if doing so creates biased training data. Fair participation and secure participation therefore need to be considered together [47]. Scalability also requires policy management at large scale. Rather than defining individual permissions for every device, hierarchical roles, reusable attributes, policy templates, and automated provisioning can reduce administration. Federated identity and decentralized trust mechanisms may further reduce dependence on a single authorization service [48].

9. APPLICATIONS IN DISTRIBUTED BIG DATA SYSTEMS

Healthcare is a major application area because hospitals and clinics hold sensitive patient data under different administrative controls. Federated learning can support collaborative model development without requiring direct exchange of patient records. Access control can restrict models by clinical purpose, institution, user role, and device security state. Secure aggregation and privacy mechanisms can provide additional protection for model updates. Industrial Internet of Things environments generate continuous sensor streams from machines, production lines, vehicles, and energy systems. Edge based federated learning can identify equipment states or anomalies while keeping raw operational data close to the

source. Access policies can distinguish machine controllers, maintenance personnel, analytics services, and external researchers [49]. Smart transportation and connected vehicle systems can use distributed learning for traffic prediction, driver assistance, road condition analysis, and resource management. Vehicle participation is dynamic and devices may connect through heterogeneous networks. Identity, short lived authorization, and robust aggregation are particularly important because compromised participants can affect safety related models. Financial services can use federated learning for fraud detection and risk modelling when institutions cannot freely share raw customer data. Strong access policies are needed because financial information is highly sensitive. Model governance should also include auditability, policy versioning, and controlled model deployment. Smart cities combine data from cameras, environmental sensors, public infrastructure, and mobile devices. Federated learning can reduce central data collection, while edge processing can lower latency. However, the large number of heterogeneous participants increases the importance of scalable identity, attribute based authorization, privacy protection, and robust model validation [50].

10. EVALUATION METRICS AND RESEARCH METHODOLOGY

Evaluation of integrated federated security should measure both learning performance and security performance. Learning metrics include accuracy, precision, recall, F1 score, loss, convergence rounds, and generalization. System metrics include communication volume, latency, computation time, memory use, energy consumption, and client participation rate. Security evaluation should include attack success rate, malicious update detection rate, false positive rate, unauthorized access rate, privacy leakage, and recovery time after compromise. Robustness experiments can vary the percentage of malicious clients and compare different poisoning strategies. Access control experiments can test unauthorized model requests, privilege escalation, stolen credentials, and device state changes [51,52]. A strong experimental design should compare a baseline federated learning system with progressively integrated controls. For example, the study can compare plain aggregation, robust aggregation, robust aggregation with access

control, and a full architecture with secure aggregation and privacy protection. This allows researchers to quantify the security benefit and overhead of each component. Datasets should reflect the intended application and, where possible, preserve realistic non independent and identically distributed characteristics. Reproducibility requires reporting client partitioning, number of rounds, local epochs, learning rate, client selection method, attack assumptions, aggregation rule, privacy parameters, and hardware or simulation environment [53]. Current research often studies federated learning, access control, privacy, and robust aggregation separately. As a result, a method that performs well in one security dimension may create a weakness in another. For example, secure aggregation protects individual updates but can reduce the visibility needed for anomaly detection. Differential privacy can protect against inference but may reduce utility when strong noise is required. Another limitation is the difficulty of distinguishing malicious behaviour from legitimate statistical heterogeneity. A rare client may naturally produce a model update that differs from the majority. Therefore, fixed thresholds can create false positives and unfairly exclude minority data sources. Robust methods should account for data diversity and not assume that the majority is always correct. Many evaluations are performed in simulated environments with a limited number of clients. Real edge cloud systems have intermittent connectivity, device failures, changing trust states, software updates, and operational constraints. Security mechanisms must therefore be evaluated under realistic churn, mobility, and resource limitations. Access control policies can also become difficult to manage when models, devices, users, and data sources grow rapidly. Policy conflicts and configuration errors can become security risks. Automated policy verification, explainable authorization decisions, and continuous auditing are important areas for practical deployment [54,55].

11. CONCEPTUAL FRAMEWORK

The conceptual framework integrates federated learning, access control, edge cloud computing, privacy preservation, and robust aggregation into a single security oriented architecture. The framework is designed for distributed environments where large volumes of data are generated by multiple edge devices, organizations, or data sources. Instead of

transferring raw data to a central cloud server, each participating device performs local model training and shares only the required model parameters or updates. The first layer of the framework is access control, which verifies the identity and authorization status of participating clients. Only trusted and policy compliant devices are allowed to participate in the learning process. The second layer performs local federated training, where each client trains the model using its own locally stored data. The third layer provides secure communication and privacy protection to reduce the risk of unauthorized interception or disclosure of model information. The next layer is robust aggregation, where the server evaluates the received model updates and reduces the influence of abnormal, manipulated, or potentially malicious contributions. Techniques such as median based aggregation, trimmed mean, Krum, trust based weighting, and anomaly detection can be incorporated into this layer. Finally, the verified updates are combined to generate an improved global model, which is redistributed to participating clients for subsequent training rounds. The framework can therefore be represented as:

Edge Data Sources → Authentication and Access Control → Local Federated Training → Secure Update Transmission → Robust Aggregation → Global Model → Model Redistribution This framework provides a coordinated approach to improving data privacy, model integrity, access security, attack resilience, communication efficiency, and scalability in distributed edge cloud big data systems. Continuous monitoring and repeated validation across training rounds can further strengthen the reliability of the overall federated learning environment [56,57]. The framework also establishes a relationship between security and model performance. Excessive security controls may increase computational and communication overhead, whereas insufficient protection may expose the global model to attacks. Therefore, the proposed conceptual framework aims to maintain a balance between privacy, security, accuracy, communication cost, and computational efficiency. This balanced approach is particularly important for large scale edge cloud environments where numerous heterogeneous devices continuously generate and process big data. Overall, the conceptual framework provides a structured foundation for integrating access control with federated learning and robust aggregation.

12. RESEARCH GAPS AND FUTURE PERSPECTIVE

Future federated systems should move toward continuous and adaptive authorization. Instead of checking identity only during registration, authorization can be reevaluated using current device posture, behaviour, model sensitivity, and contextual attributes. This approach can support rapid revocation when a device becomes compromised [61,62]. A second gap is the integration of robust learning with privacy preserving aggregation. Robust aggregation normally benefits from access to individual updates, while secure aggregation deliberately hides them. Future protocols should support privacy preserving anomaly detection, secure reputation calculation, or trusted edge validation without exposing raw client updates. Another research direction is decentralized trust. Blockchain and distributed ledger mechanisms have been proposed for identity, reputation, audit, and coordination, but their overhead can be significant for constrained edge devices. Future designs should determine when decentralized trust is genuinely beneficial and when conventional lightweight authorization is more efficient. Personalized security policies are also needed. Different participants may have different risk profiles, data sensitivities, and resource capabilities. A one size fits all policy can either create unnecessary overhead or provide insufficient protection. Context aware policy engines could dynamically select authentication strength, privacy level, aggregation method, and model access rights [63,64]. Explainable security decisions should receive greater attention. When an update is rejected or a device is quarantined, the system should provide an auditable reason such as expired authorization, device posture failure, abnormal update behaviour, or policy mismatch. Explainability can help administrators distinguish attacks from legitimate changes in data distribution. Finally, future research should evaluate complete end to end systems rather than isolated algorithms. Benchmarking should include learning accuracy, attack resistance, privacy leakage, communication overhead, energy consumption, policy latency, fairness, and recovery behaviour. Standardized datasets and security benchmarks for edge cloud federated learning would make comparisons more meaningful [66,67].

CONCLUSION

Federated learning provides an important foundation for collaborative machine learning in distributed edge cloud big data systems because it allows model training to occur close to data sources while reducing the need to centralize raw information. However, decentralization changes the security problem rather than eliminating it. Participants can be compromised, credentials can be misused, model updates can be poisoned, and information can potentially be inferred from shared parameters. Robust aggregation, secure aggregation, differential privacy, and access control therefore need to be considered as complementary mechanisms. An integrated architecture can use identity, roles, attributes, device state, data sensitivity, and behavioural evidence to control participation while robust aggregation reduces the effect of malicious updates. Secure aggregation and privacy mechanisms can further limit exposure of individual contributions. The principal research challenge is to combine these controls without creating excessive communication, computation, or policy management overhead. Future systems should support continuous authorization, privacy preserving anomaly detection, decentralized trust where justified, resource aware security, and explainable policy decisions. End to end evaluation is also needed to measure security, privacy, learning utility, fairness, and system efficiency together. Overall, integrated access control should be treated as a core component of robust federated learning rather than as an external security feature. Such integration can provide a stronger foundation for trustworthy collaborative intelligence across healthcare, industrial systems, transportation, finance, and smart infrastructure while preserving the distributed nature of edge cloud data.

REFERENCES

1. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* 2017 Apr 10 (pp. 1273-1282). Pmlr.
2. Kairouz P, McMahan HB. Advances and open problems in federated learning. *Foundations and trends in machine learning*. 2021 Jun 23;14 (1-2):1-210.
3. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2019 Jan 28;10 (2):1-9.
4. Lim WY, Luong NC, Hoang DT, Jiao Y, Liang YC, Yang Q, Niyato D, Miao C. Federated learning in mobile edge networks: A comprehensive survey. *IEEE communications surveys & tutorials*. 2020 Apr 8;22 (3):2031-63.
5. Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y. A survey on federated learning. *Knowledge based systems*. 2021 Mar 15;216:106775.
6. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, Ramage D, Segal A, Seth K. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* 2017 Oct 30 (pp. 1175-1191).
7. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*. 2020 Mar 15;2:429-50.
8. Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning* 2020 Nov 21 (pp. 5132-5143). PMLR.
9. Wang J, Liu Q, Liang H, Joshi G, Poor HV. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*. 2020;33:7611-23.
10. Li Q, He B, Song D. Model-contrastive federated learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2021 Jun 20 (pp. 10708-10717). IEEE.
11. Rodio A, Neglia G. FedStale: leveraging Stale Updates in Federated Learning. In *ECAI 2024: 27th European Conference on Artificial Intelligence, 19to24 October 2024, Santiago de Compostela, Spain Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024)* 2024 Oct 19 (pp. 3071-3078). 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP: SAGE Publications Pvt. Ltd.

12. Li X, Huang K, Yang W, Wang S, Zhang Z. On the convergence of fedavg on non iid data. arXiv preprint arXiv:1907.02189. 2019 Jul 4.
13. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated learning with non iid data. arXiv preprint arXiv:1806.00582. 2018 Jun 2.
14. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492. 2016 Oct 18.
15. Wang S, Tuor T, Salonidis T, Leung KK, Makaya C, He T, Chan K. Adaptive federated learning in resource constrained edge computing systems. *IEEE journal on selected areas in communications*. 2019 Mar 11;37 (6):1205-21.
16. Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: Vision and challenges. *IEEE internet of things journal*. 2016 Jun 9;3 (5):637-46.
17. Satyanarayanan M. The emergence of edge computing. *computer*. 2017 Jan 5;50 (1):30-9.
18. Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*. 2019 Jun 12;107 (8):1738-62.
19. Mao Y, You C, Zhang J, Huang K, Letaief KB. A survey on mobile edge computing: The communication perspective. *IEEE communications surveys & tutorials*. 2017 Aug 25;19 (4):2322-58.
20. Mach P, Becvar Z. Mobile edge computing: A survey on architecture and computation offloading. *IEEE communications surveys & tutorials*. 2017 Mar 15;19 (3):1628-56.
21. Roman R, Lopez J, Mambo M. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*. 2018 Jan 1;78:680-98.
22. Yi S, Qin Z, Li Q. Security and privacy issues of fog computing: A survey. In *International conference on wireless algorithms, systems, and applications* 2015 Aug 1 (pp. 685to695). Cham: Springer International Publishing.
23. Stojmenovic I, Wen S. The fog computing paradigm: Scenarios and security issues. In *2014 federated conference on computer science and information systems* 2014 Sep 7 (pp. 1-8). IEEE.
24. Nguyen DC, Ding M, Pathirana PN, Seneviratne A, Li J, Poor HV. Federated learning for internet of things: A comprehensive survey. *IEEE communications surveys & tutorials*. 2021 Apr 26;23 (3):1622-58.
25. Nguyen DC, Ding M, Pham QV, Pathirana PN, Le LB, Seneviratne A, Li J, Niyato D, Poor HV. Federated learning meets blockchain in edge computing: Opportunities and challenges. *IEEE Internet of Things Journal*. 2021 Apr 13;8 (16):12806-25.
26. Lu Y, Huang X, Zhang K, Maharjan S, Zhang Y. Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles. *IEEE Transactions on Vehicular Technology*. 2020 Feb 13;69 (4):4298-311.
27. Qu Y, Pokhrel SR, Garg S, Gao L, Xiang Y. A block chained federated learning framework for cognitive computing in industry 4.0 networks. *IEEE Transactions on Industrial Informatics*. 2020 Jul 7;17 (4):2964-73.
28. Wan Y, Qu Y, Gao L, Xiang Y. Privacy-preserving blockchain enabled federated learning for B5G Driven edge computing. *Computer Networks*. 2022 Feb 26;204:108671.
29. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, Zhang L. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* 2016 Oct 24 (pp. 308-318).
30. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*. 2014 Aug 11;9 (3-4):211-487.
31. Zhu L, Liu Z, Han S. Deep leakage from gradients. *Advances in neural information processing systems*. 2019;32.
32. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)* 2017 May 22 (pp. 3-18). IEEE.
33. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer*

- and communications security 2015 Oct 12 (pp. 1322-1333).
34. Nasr M, Shokri R, Houmansadr A. Comprehensive privacy analysis of deep learning: Passive and active white box inference attacks against centralized and federated learning. In 2019 IEEE symposium on security and privacy (SP) 2019 May 19 (pp. 739-753). IEEE.
35. Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V. How to backdoor federated learning. In International conference on artificial intelligence and statistics 2020 Jun 3 (pp. 2938-2948). PMLR.
36. Blanchard P, El Mhamdi EM, Guerraoui R, Stainer J. Machine learning with adversaries: Byzantine tolerant gradient descent. *Advances in neural information processing systems*. 2017;30.
37. Yin D, Chen Y, Kannan R, Bartlett P. Byzantine-robust distributed learning: Towards optimal statistical rates. In International conference on machine learning 2018 Jul 3 (pp. 5650-5659). Pmlr.
38. El Mhamdi EM, Guerraoui R, Rouault S. The hidden vulnerability of distributed learning in byzantium. In International conference on machine learning 2018 Jul 3 (pp. 3521-3530). PMLR.
39. Fang M, Cao X, Jia J, Gong N. Local model poisoning attacks to {Byzantine Robust} federated learning. In 29th USENIX security symposium (USENIX Security 20) 2020 (pp. 1605-1622).
40. Fung C, Yoon CJ, Beschastnikh I. Mitigating sybils in federated learning poisoning. *arXiv preprint arXiv:1808.04866*. 2018 Aug 14.
41. Bhagoji AN, Chakraborty S, Mittal P, Calo S. Analyzing federated learning through an adversarial lens. In International conference on machine learning 2019 May 24 (pp. 634-643). PMLR.
42. Sun Z, Kairouz P, Suresh AT, McMahan HB. Can you really backdoor federated learning?. *arXiv preprint arXiv:1911.07963*. 2019 Nov 18.
43. Sandhu RS, Coyne EJ, Feinstein HL, Youman CE. Role-based access control models. *Computer*. 1996 Feb 29;29 (2):38-47.
44. Ferraiolo DF, Kuhn DR. Role-based access controls. *arXiv preprint arXiv:0903.2171*. 2009 Mar 12.
45. Sandhu R, Ferraiolo D, Kuhn R. The NIST model for role based access control: towards a unified standard. In *ACM workshop on Role based access control* 2000 Jul 26 (Vol. 10, No. 344287.344301).
46. Hu VC, Ferraiolo D, Kuhn R, Friedman AR, Lang AJ, Cogdell MM, Schnitzer A, Sandlin K, Miller R, Scarfone K. Guide to attribute based access control (abac) definition and considerations (draft). NIST special publication. 2013 Apr;800 (162):1-54.
47. Bethencourt J, Sahai A, Waters B. Ciphertext policy attribute based encryption. In 2007 IEEE symposium on security and privacy (SP'07) 2007 May 20 (pp. 321-334). IEEE.
48. Goyal V, Pandey O, Sahai A, Waters B. Attribute based encryption for fine grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security* 2006 Oct 30 (pp. 89-98).
49. Han J, Susilo W, Mu Y, Yan J. Privacy preserving decentralized key policy attribute based encryption. *IEEE transactions on parallel and distributed systems*. 2012 Jan 31;23 (11):2150-62.
50. Rao KR, Ray IG, Asif W, Nayak A, Rajarajan M. R PEKS: RBAC enabled PEKS for secure access of cloud data. *IEEE Access*. 2019 Sep 16;7:133274-89.
51. Vyhmeister E, G. Castane G. When Industry meets trustworthy AI: a systematic review of AI for Industry 5.0. *AI and Ethics*. 2026 Apr;6 (2):200.
52. Conti M, Kumar ES, Lal C, Ruj S. A survey on security and privacy issues of bitcoin. *IEEE communications surveys & tutorials*. 2018 May 31;20 (4):3416-52.
53. Zyskind G, Nathan O. Decentralizing privacy: Using blockchain to protect personal data. In 2015 IEEE security and privacy workshops 2015 May 21 (pp. 180-184). IEEE.
54. Caldas S, Duddu SM, Wu P, Li T, Konečný J, McMahan HB, Smith V, Talwalkar A. Leaf: A

- benchmark for federated settings. arXiv preprint arXiv:1812.01097. 2018 Dec 3.
55. Hard A, Rao K, Mathews R, Ramaswamy S, Beaufays F, Augenstein S, Eichner H, Kiddon C, Ramage D. Federated learning for mobile keyboard prediction. arXiv preprint arXiv:1811.03604. 2018 Nov 8.
 56. Yang T, Andrew G, Eichner H, Sun H, Li W, Kong N, Ramage D, Beaufays F. Applied federated learning: Improving google keyboard query suggestions. arXiv preprint arXiv:1812.02903. 2018 Dec 7.
 57. Smith V, Chiang CK, Sanjabi M, Talwalkar AS. Federated multi task learning. *Advances in neural information processing systems*. 2017;30.
 58. Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning with theoretical guarantees: A model agnostic meta-learning approach. *Advances in neural information processing systems*. 2020;33:3557-68.
 59. T Dinh C, Tran N, Nguyen J. Personalized federated learning with moreau envelopes. *Advances in neural information processing systems*. 2020;33:21394-405.
 60. Li J, Wang S, Zhang Y. Federated learning with matched averaging. *IEEE Transactions on Neural Networks and Learning Systems*. 2020;31(1):28-38.
 61. Dinh CT, Tran NH, Nguyen MN, Hong CS, Bao W, Zomaya AY, Gramoli V. Federated learning over wireless networks: Convergence analysis and resource allocation. *IEEE/ACM Transactions on Networking*. 2020 Nov 17;29(1):398-409.
 62. Amiri MM, Gündüz D. Federated learning over wireless fading channels. *IEEE transactions on wireless communications*. 2020 Feb 26;19(5):3546-57.
 63. Yang Z, Chen M, Wong KK, Poor HV, Cui S. Federated learning for 6G: Applications, challenges, and opportunities. *Engineering*. 2022 Jan 1;8:33-41.
 64. Theertha Suresh A, Yu FX, Kumar S, Brendan McMahan H. Distributed Mean Estimation with Limited Communication. arXiv e prints. 2016 Nov:arXiv-1611.
 65. Yin D, Chen Y, Kannan R, Bartlett P. Defending against saddle point attack in Byzantine robust distributed learning. In *International Conference on Machine Learning 2019 May 24* (pp. 7074-7084). PMLR.
 66. Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, Kiddon C, Konečný J, Mazzocchi S, McMahan B, Van Overveldt T. Towards federated learning at scale: System design. *Proceedings of machine learning and systems*. 2019 Apr 15;1:374-88.
 67. McMahan HB, Ramage D, Talwar K, Zhang L. Learning differentially private recurrent language models. arXiv preprint arXiv:1710.06963. 2017 Oct 18.

HOW TO CITE: Vinod K. Barupati, Shivshankar M. Nagrik, Robust Federated Learning with Integrated Access Control for Distributed Edge Cloud Big Data Systems, *Int. J. Sci. R. Tech.*, 2026, 3 (10), 1005-1015. <https://doi.org/10.5281/zenodo.23455670>