An improved approach: adaptive learning for high-speed data encryption of low-earth orbit (LEO) satellites

Makhabbat Bakyt[1,†], Luigi La Spada[2,†], Sabyrzhan Atanov[1,†], Khuralay Moldamurat[1,†] and Mansur Moldakhanov[3,†]

[1] *L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan*

[2] *School of Computing, Engineering and the Built Environment, Edinburgh Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, United Kingdom*

[3] *LLP Samira, Pavlodar, 140000, Kazakhstan*

Abstract

This paper examines the critical problem of ensuring data security in low-Earth orbit (LEO) satellite communication systems, where resources are limited, and the risk of cyberattacks is heightened. An innovative approach to continuous user authentication is proposed, based on adaptive machine learning using logistic regression and support vector machines, combined with robust cryptographic protocols. The study's primary goal is to develop a lightweight and efficient encryption method that guarantees a high level of data security without significantly increasing the computational load on LEO satellite onboard systems. The article analyzes existing continuous authentication methods and identifies their vulnerabilities, such as dependence on specific features, scalability issues, and susceptibility to targeted attack scenarios. The proposed approach is based on the adaptive sliding window method, which allows dynamic adaptation to changes in user behavior and effectively detects anomalies indicating potential unauthorized access attempts. To safeguard data, the study proposes using functional encryption and decentralized key generation, enhancing the system's resilience to various types of attacks. Preliminary simulation results using mouse movement data demonstrate that the proposed approach achieves high anomaly detection accuracy (over 80%) with minimal computation time (less than 9 ms). This method can be applied to protect data in various communication systems with LEO satellites, including flight control, telemetry, and data transmission systems.

Keywords

authentication, machine learning, adaptive learning, artificial intelligence, data security, LEO satellites

1. Introduction

The rapid advancement of satellite communication technologies and the increasing number of LEO satellites offer unprecedented opportunities for collecting and processing vast amounts of data, such as high-resolution satellite images. However, this progress also brings significant challenges in ensuring the security of data transmitted between LEO satellites and ground stations. Limited onboard resources, high risks of data interception, and the increasing sophistication of cyberattacks necessitate the development of innovative data encryption approaches that provide robust security without overburdening the computational capabilities of these satellites.

Traditional encryption methods, including symmetric algorithms like AES and asymmetric algorithms like RSA, present several limitations when applied to LEO satellite communication. Symmetric algorithms require the pre-distribution of keys among all communication participants, a

DTESI 2024: 9th International Conference on Digital Technologies in Education, Science and Industry, October 16–17, 2024, Almaty, Kazakhstan

* Corresponding author.

† These authors contributed equally.

bakyt.makhabbat@gmail.com (M. Bakyt); L.LaSpada@napier.ac.uk (L. Spada)

0000-0002-1246-9696 (M. Bakyt); 0000-0003-4923-1320 (L. Spada)

complex task in a dynamic LEO satellite network with changing topology. Asymmetric algorithms, while suitable for secure key exchange, are computationally intensive and may be unsuitable for resource-constrained LEO satellites [1]. Moreover, both traditional encryption methods are potentially vulnerable to attacks based on quantum computing [2].

Federated Learning (FL) methods have emerged as a promising solution for training machine learning models on distributed data without requiring centralized data collection [1]. This approach holds potential for satellite communication systems, but current FL methods often overlook the unique characteristics of such systems and lack sufficient data security measures. For example, while some research [2] addresses data security in satellite communications using FL, it doesn't offer concrete solutions to protect against internal and external threats. Other studies [3, 4, 5, 6, 7, 8, 9, 10] explore various aspects of FL in satellite systems but lack a comprehensive approach that addresses all security and efficiency concerns. This paper proposes a novel federated learning approach for LEO satellites based on adaptive machine learning using logistic regression and support vector machines, integrated with cryptographic protocols. This approach aims to address the following challenges:

- Ensuring data privacy: Protecting sensitive data from unauthorized access and preventing information leakage.
- Improving efficiency: Minimizing computational and communication overhead while maintaining robust security.
- Ensuring high classification accuracy: Developing machine learning models capable of effectively classifying data collected by LEO satellites.

To achieve these goals, we propose using functional encryption, decentralized key generation, and on-orbit model aggregation methods. This paper also explores the potential of quantum key distribution (QKD) for secure communication between LEO satellites and ground stations [11-19, 20, 21]. QKD allows the generation of secret keys used to encrypt data transmitted between LEO satellites and ground stations, with security based on the fundamental principles of quantum mechanics, making it resistant to attacks from quantum computers.

2. Methods

Our approach utilizes functional encryption (FE) to ensure data privacy during the aggregation of machine learning models trained on different LEO satellites [12]. Each satellite generates its own encryption keys, eliminating the need for a central key generation authority and bolstering system security. This scheme employs the anonymous veto network protocol (AV-net) [15] for secure key exchange between satellites without relying on a trusted center. AV-net utilizes asymmetric encryption to ensure the privacy and authentication of transmitted data, protecting against unauthorized access and data manipulation.

To expedite model convergence, we propose on-orbit model aggregation. This minimizes delays associated with transmitting data to the ground station and optimizes communication bandwidth. It also enhances system fault tolerance, as the loss of communication with one satellite doesn't halt the training process. On-orbit model aggregation involves these steps:

1. Each satellite in orbit trains its model based on the global model received from the ground station.
2. The first visible satellite sends its model to its neighbor, which performs partial aggregation of its model with the received model.

3. This process continues until the final partially aggregated model reaches the initial satellite.
4. The initial satellite transmits this final model back to all satellites in the orbit (Fig. 1).

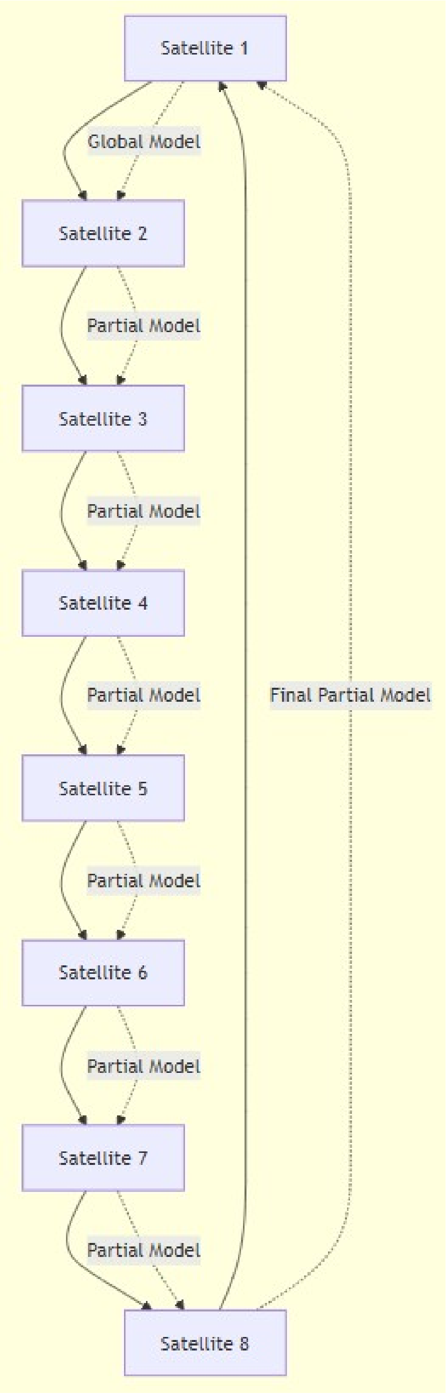

Figure 1: In-orbit model forwarding and aggregation.

Figure 1 illustrates the process of forwarding and aggregating models in orbit. The process starts with Satellite 1 sending its model to Satellite 2, which aggregates it with its own. This aggregated model is then passed on to Satellite 3, and so on until it reaches Satellite 8. Finally, Satellite 8 sends the final aggregated model back to Satellite 1, which then distributes it to all other satellites.

To detect anomalies in user behavior, potentially indicating unauthorized access attempts, we employ the adaptive sliding window method [22]. This method dynamically adjusts the data analysis window size based on the current situation, improving anomaly detection accuracy.

The adaptive learning algorithm used in this study involves a combination of logistic regression and support vector machines (SVM) to classify user behaviour based on the collected data. The key feature of this algorithm is its adaptability to changing user patterns, which is achieved through an adaptive sliding window mechanism. Here, we provide a more detailed technical description of how this algorithm is implemented in practice:

1. Data Preprocessing: Data from user behaviour, such as mouse movement data, is collected and normalized to ensure consistency. Features are standardized to have a mean of zero and a unit variance, which is crucial for the performance of machine learning models like logistic regression and SVM.
2. Sliding Window Implementation: The adaptive sliding window is responsible for dynamically adjusting the amount of data used for model training and anomaly detection. Initially, a default window size is set. The window size increases during stable behaviour to reduce computational load and decreases when unusual behaviour is detected to allow more focused analysis. This helps in improving anomaly detection accuracy.
3. Model Training: The logistic regression model is used to establish a linear relationship between the features and the classification labels (e.g., normal or anomalous behaviour). For more complex, non-linear relationships, an SVM with a radial basis function (RBF) kernel is employed. The models are trained incrementally, with the sliding window providing the latest batch of data for continuous learning. This allows the system to adapt quickly to new user patterns.
4. Anomaly Detection Process: The detection process begins by comparing the current user behaviour against historical patterns stored in the model. If the deviation exceeds a predefined threshold, the system classifies the behaviour as anomalous. The threshold is dynamically adjusted based on recent observations to reduce false positives.
5. Algorithm Workflow:

Initialization: Initialize model parameters (weights for logistic regression and hyperplane for SVM). Set initial window size and rejection threshold.

Data Processing Loop: use a continuous sequence in which incoming data samples are processed in real-time. For each new data sample, the data is normalized and standardized before being incorporated into the sliding window. Depending on the observed behavior, the window size may be adjusted—reduced for focused analysis when anomalies are detected or increased during stable periods to lower computational demands. The models (logistic regression and SVM) are then retrained with the latest data, and the sample is classified as either normal or anomalous.

6. Model Updating: As new data arrives, the model is updated using an online learning approach. This incremental updating ensures that the model remains up to date without requiring a full retraining, which is computationally expensive for LEO satellites.
7. Communication Efficiency: Given the resource-constrained environment of LEO satellites, the training is optimized to minimize communication overhead. Only essential model updates are transmitted between satellites, reducing the demand on communication bandwidth while maintaining model accuracy.

For instance, the window size is reduced when suspicious activity is detected for more detailed analysis and increased during normal operation to reduce computational load. The criterion for adjusting the window size can be the deviation of current user behavior parameters from historical averages. This method's effectiveness was demonstrated in [22], where it was successfully applied for recognizing hand gestures using a data glove.

Table 1
Main parameters of the system

Parameter	Value
Number of satellites	8
Data rate	1 Mbps
Training set size	1000 samples per satellite
Adaptive sliding window parameters	Rejection threshold: 0.8, Window size: 10-100 samples

Table 1 presents the key parameters used in the simulation of the proposed system. One potential limitation of this approach is the reliance on reliable communication between satellites for on-orbit model aggregation. Communication loss or satellite failure may disrupt this process. To mitigate this, backup communication channels and data recovery mechanisms can be employed.

This reliance presents a significant challenge, as the dynamic and often unpredictable conditions of LEO satellite environments can lead to frequent interruptions or degradations in communication quality. For instance, electromagnetic interference, satellite positioning errors, or unexpected hardware failures can all contribute to communication breakdowns, making it challenging to maintain a continuous and reliable connection between satellites. Such disruptions could directly impact the aggregation process, leading to incomplete or inconsistent models that degrade system performance.

To mitigate these risks, it is crucial to develop robust redundancy mechanisms, such as backup communication channels that automatically engage in the event of primary communication failure. Additionally, employing distributed data storage and model checkpointing techniques could help in preserving the intermediate states of the model, ensuring that any progress made before a failure is not lost. These enhancements would not only improve system resilience but also facilitate quicker recovery, thus reducing the potential impact of communication disruptions on overall system performance.

3. Results and discussion

The performance of the proposed FedSecure approach was evaluated using Intersection over Union (IoU) and Dice Coefficient metrics, common for assessing semantic image segmentation quality [16]. IoU measures the overlap between predicted and actual masks, while the Dice Coefficient considers both overlap and region size. The model was trained using stochastic gradient descent with mini-batches of size 4 and a learning rate of ζ=0.00008.

However, it's crucial to acknowledge that these simulations might not fully represent real-world LEO satellite conditions. The system could be influenced by external factors like radiation [18, 25], interference in quantum channels [19, 26], instability of the radiation source [19, 27], and the

Earth's gravitational field [20, 28]. Thus, these simulation results should be considered preliminary and require further validation with real-world data, such as data collected from actual LEO satellite communication systems.

Moreover, while the preliminary simulations provide valuable insights into the feasibility and efficiency of the proposed approach, they may not fully account for the complexities present in real-world LEO satellite environments. To ensure the reliability and robustness of the system, further experiments using actual satellite communication data are necessary. These experiments will help address potential discrepancies caused by environmental factors like electromagnetic interference, unpredictable network latencies, and hardware limitations that are challenging to replicate in simulations. Verification with real-world data will be crucial to refining the proposed methods and ensuring their practical application in mission-critical satellite operations.

Table 2

Comparison of FedSecure with other approaches

Approach	Convergence time (hours)	Accuracy (%)	Computational cost (ms)
FedSecure	3	88.76	<9
FedISL [4]	4	82.76	-
FedHAP [7]	15	-	-
FedSpace [9]	96	-	-

As shown in Table 2, FedSecure demonstrates faster convergence and higher accuracy compared to other approaches. Factors Affecting Performance. Several factors influence FedSecure's performance:

Number of satellites: Increasing the number of satellites in the constellation accelerates model convergence but increases communication overhead. Optimization methods from [10, 23] can be used to address this.

Data rate: Limited bandwidth between LEO satellites and the ground station can hinder the training process. Increasing data rate improves convergence speed but demands more satellite processing resources. Figure 2 illustrates the impact of data rate on the accuracy and computation time of the proposed approach.

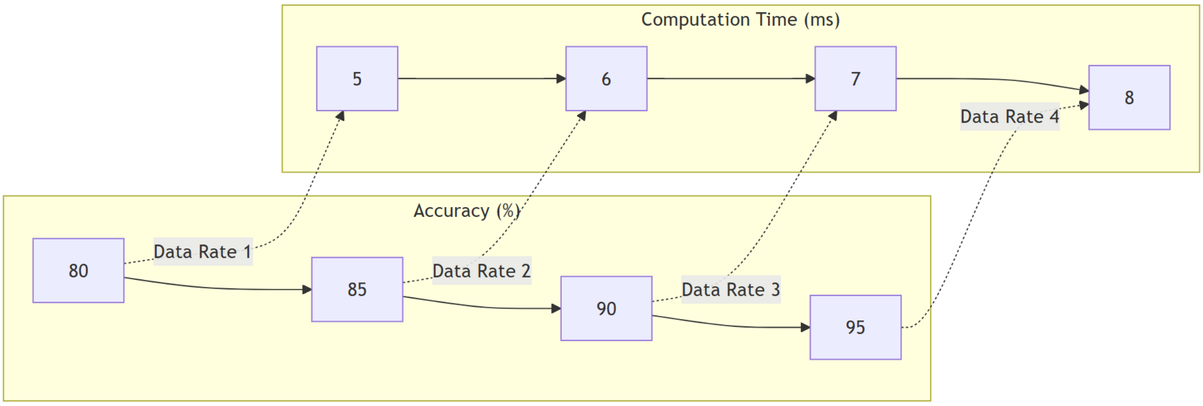

Figure 2: Accuracy and computation time for different data rates.

Figure 2 shows that increasing the data rate leads to higher accuracy but also increased computation time. Training set size: Larger training sets on each satellite enhance classification accuracy but increase the computational burden on onboard systems. A balance between training set size and satellite capabilities is essential. Adaptive sliding window parameters: Optimal parameters for the adaptive sliding window method, such as rejection threshold and window size, depend on the specific application and data characteristics.

Limitations of Preliminary Simulations and Optimization Paths. Preliminary simulations were conducted using the Balabit Mouse Challenge dataset, containing user mouse movement data [22, 24], to simulate user behavior and evaluate the adaptive sliding window method for anomaly detection. To address the reliance on inter-satellite communication, backup communication channels and data recovery mechanisms can be used.

Table 3
Comparison of FedSecure with other security approaches

Approach	Protocol Type	Protection against man-in-the-middle attacks	Transmission Error Resistance
FedSecure	Adaptive Machine Learning with Functional Encryption	High	High
BB84 [19]	Quantum Key Distribution	Medium	Medium
CCMEA [11]	Multiple Encryption with Control Code	Low	Low

Table 3 shows that FedSecure offers better protection against man-in-the-middle attacks and higher resistance to transmission errors compared to other security approaches. This advantage stems from the use of functional encryption and decentralized key generation, which significantly increase the complexity for attackers attempting to intercept and decrypt data. The high resilience to transmission errors is achieved by employing quantum key distribution (QKD), enabling the detection and correction of errors occurring during data transmission over a quantum channel.

4. Conclusion

This paper presents a novel approach to federated learning for LEO satellites, using adaptive machine learning with logistic regression and support vector machines, combined with cryptographic protocols. The primary objective was to develop a lightweight, efficient encryption method that ensures high data security without significantly impacting the computational resources of LEO satellite systems. Our approach employs the adaptive sliding window method, enabling dynamic adaptation to user behavior changes and effective detection of anomalies that may indicate unauthorized access attempts. Functional encryption and decentralized key generation enhance the system's resilience against various attacks. Preliminary simulations using mouse movement data show that our approach achieves high anomaly detection accuracy (over 80%) with minimal computation time (less than 9 ms). This result suggests that FedSecure can be an effective tool for data security in LEO satellite communication systems, particularly in resource-constrained environments.

Future research will involve experiments with real-world data collected from LEO satellites to evaluate the effectiveness and security of this method under realistic conditions. We will investigate various functional encryption and decentralized key generation methods and different federated learning architectures. Scalability and fault tolerance of the system, considering limited resources and the dynamic nature of LEO satellite networks, will be a particular focus. Further studies will assess the impact of environmental factors, such as radiation exposure, on the proposed method's performance. This research is expected to contribute to the development of more advanced data encryption methods for satellite communication systems, enhancing their security and reliability. This is particularly crucial for mission-critical applications like flight control, telemetry, and data transmission, where data security is paramount.

Acknowledgements

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Project No. AP19677508).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in AISTATS, 2017, pp. 1273–1282.

[2] H. Chen, M. Xiao, and Z. Pang, “Satellite-based computing networks with federated learning,” IEEE Wireless Communications, vol. 29, no. 1, pp. 78–84, 2022.

[3] Seitbattalov, Z.Y., Atanov, S.K., Moldabayeva, Z.S. An Intelligent Decision Support System for Aircraft Landing Based on the Runway Surface SIST 2021 - 2021 IEEE International Conference on Smart Information Systems and Technologies, 2021, 9466000

[4] N. Razmi et al., “On-board federated learning for dense LEO constellations,” in IEEE ICC, May 2022.

[5] M. Elmahallawy and T. Luo, “Optimizing federated learning in LEO satellite constellations via intra-plane model propagation and sink satellite scheduling,” in 2023 IEEE International Conference on Communications (ICC). IEEE, 2023.

[6] Utegen, A.S., Moldamurat, K., Ainur, M., Amandykuly, A.G., Brimzhanova, S.S.Development and modeling of intelligent control system of cruise missile based on fuzzy logic Proceedings - 2021 16th International Conference on Electronics Computer and Computation, ICECCO 2021, 2021

[7] M. Elmahallawy and T. Luo, “FedHAP: Fast federated learning for LEO constellations using collaborative HAPs,” in 14th IEEE International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2022, pp. 888–893.

[8] M. Elmahallawy and T. Luo, “AsyncFLEO: Asynchronous federated learning for LEO satellite constellations with high-altitude platforms,” in 2022 IEEE International Conference on Big Data (BigData), 2022, pp. 5478–5487.

[9] Brimzhanova, S., Atanov, S., Moldamurat, K., Brimzhanova, K., Seitmetova, A. An intelligent testing system development based on the shingle algorithm for assessing humanities students'

academic achievements Education and Information Technologies, 2022, 27(8), страницы 10785–10807

[10] M. Elmahallawy, T. Luo, and M. I. Ibrahem, “Secure and efficient federated learning in leo constellations using decentralized key generation and on-orbit model aggregation,” in GLOBECOM 2023-2023 IEEE Global Communications Conference. IEEE, 2023, pp. 5727–5732.

[11] J. Liu et al., “Control code multiple encryption algorithm on satellite-to-ground communication,” Mobile Networks and Applications, vol. 24, pp. 1955–1974, 2019.

[12] J. Ma et al., “Privacy-preserving federated learning based on multi-key homomorphic encryption,” International Journal of Intelligent Systems, vol. 37, no. 9, pp. 5880–5901, 2022.

[13] Khuralay Moldamurat1, Sabyrzhan Atanov, Kairat Akhmetov, Makhabbat Bakyt, Niyaz Belgibekov, Assel Zhumabayeva, Yuriy Shabayev, Improved unmanned aerial vehicle control for efficient obstacle detection and data protection, IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 13, No. 3, September 2024, pp. 3576-3587, ISSN: 2252-8938, DOI: http://doi.org/10.11591/ijai.v13.i3.pp3576-3587

[14] S. Kim et al., “Function-hiding inner product encryption is practical,” in Security and Cryptography for Networks: 11th International Conference, SCN. Springer, 2018, pp. 544–562.

[15] F. Hao and P. Zielinski, “The power of anonymous veto in public discussion,” Transactions on Computational Science IV: Special Issue on Security in Computing, pp. 41–52, 2009.

[16] I. Demir et al., “Deepglobe 2018: A challenge to parse the earth through satellite images,” in Proceedings of the IEEE CVPR Workshops, 2018, pp. 172–181.

[17] H. Kaen and J. Kaen, “Secure global satellite network,” US Patent 20,220,094,431, Mar. 24, 2022.

[18] I. DSouza et al., “Repeated radiation damage and thermal annealing of avalanche photodiodes,” EPJ Quantum Technology, vol. 8, no. 1, p. 13, 2021.

[19] J. S. Sidhu et al., “Finite key performance of satellite quantum key distribution under practical constraints,” Communications Physics, vol. 6, no. 1, p. 210, 2023.

[20] Khuralay Moldamurat, Yerzhan Seitkulov, Sabyrzhan Atanov, Makhabbat Bakyt, Banu Yergaliyeva, Enhancing Cryptographic Protection, Authentication, And Authorization In Cellular Networks: A Comprehensive Research Study, International Journal of Electrical and Computer Engineering (IJECE), Vol. 14, No. 1, February 2024, pp.479-487.

[21] X.-L. Pang et al., “Experimental quantum-enhanced cryptographic remote control,” Scientific reports, vol. 9, no. 1, pp. 1–5, 2019.

[22] G. Luzhnica et al., “A sliding window approach to natural hand gesture recognition using a custom data glove,” in 2016 IEEE Symposium on 3D User Interfaces (3DUI), IEEE, 2016, pp. 81–90

[23] Makhabbat B., Khuralay M., Assem K., Adil M. and Dina S., “Integration of Cryptography and Navigation Systems in Unmanned Military Mobile Robots: A Review of Current Trends and Perspectives” CEUR Workshop Proceedings, Volume 36802024, 8th International Conference on Digital Technologies in Education, Science and Industry, DTESI 2023, Almaty 6 December 2023 through 7 December 2023, Code 199323.

[24] M. Nasr, R. Shokri, and A. Houmansadr, “Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning,” in IEEE symposium on security and privacy (SP). IEEE, 2019, pp. 739–753.

[25] J. Lin, J. Xu, Y. Li, and Z. Xu, “Federated learning with dynamic aggregation based on connection density at satellites and ground stations,” in 2022 IEEE International Conference on Satellite Computing (Satellite). IEEE, 2022, pp. 31–36.

[26] Bakyt, M., Moldamurat, Kh., Satybaldina, D.Zh., Yurkov, N.K., MODELING INFORMATION SECURITY THREATS FOR THE TERRESTRIAL SEGMENT OF SPACE COMMUNICATIONS, CEUR Workshop Proceedings, Volume 33822022 7th International

Conference on Digital Technologies in Education, Science and Industry, DTESI 2022, Almaty 20 October 2022 through 21 October 2022, Code 188290.

[27] J. So et al., “Fedspace: An efficient federated learning framework at satellites and ground stations,” arXiv preprint arXiv:2202.01267, 2022.

[28] M. Abdalla et al., “Multi-input functional encryption for inner products: Function-hiding realizations and constructions without pairings,” in Advances in Cryptology–CRYPTO: 38th Annual International Cryptology Conference. Springer, 2018.

[29] T. Liu et al., “Satellite-based continuous-variable quantum key distribution under the earth’s gravitational field,” Quantum Information Processing, vol. 21, no. 11, pp. 1–17, 2022.