Response Letter

Manuscript No: 1570932011
Semantic Communication in Satellite-borne Edge Cloud Network for Computation Offloading

We highly appreciate the constructive comments and invaluable suggestions from reviewers. We have carefully addressed all the review comments and improved the quality of this manuscript accordingly. Please find below our detailed response to all the review comments:

• Response to the comments from Review 1

This is a very solid work on satellite edge offloading. Besides the theoretical analysis, the work shines due to the extensive experimental analysis that has been conducted. This allows to see the benefits and drawbacks of the different alternatives and understand the effect of every parameter in the problem. Also, the system model is very broad and general, encompassing any possible communication between the entities.

Response:
Thanks for the kind words.

In the following I highlight some aspects of the paper, along with some corrections that may be addressed:

■ Comment 1:
In the last contribution of the paper, last bullet: remove "etc".

Response:
Thanks for your comment. To address this comment, we have removed the "etc" in the last bullet of the last contribution of the paper.

■ Comment 2:
•Typo, two stops: "...via Ka-band backhaul links to provide cloud service for users."

Response:
Thank you for your comment. Based on this comment, we have deleted the extra stop in this sentence.

■ Comment 3:
Style typo in eq. (4): f_Cloud may be f_{Cloud}?

Response:
Thank you for your comment. Based on this comment, we have corrected *f_Cloud* and replaced it with *f_{Cloud}*.

■ Comment 4:
Please consider writing eq. (6) as a full fraction.

Response:
Thank you for your comment. Based on this comment, we have rewritten the Eq. (6) as a full fraction.

■ Comment 5:
After eq. (12), what does "special" antenna array mean?

Response:
Thank you for your comment. The “special” antenna array means the terrestrial-station-terminal (TST) antennas operating in the Ka-band have good directivity [1,2]. TST can select multiple satellites that have enough angular separations among each other to transmit information. This ensures the off-axis antenna gain is lower and the interference is tolerable.

Based on this comment and to clarify, we removed the expression of “special” antenna array and explained that the antennas of TSTs provide proper directivity.

■ Comment 6:
Typo: "The transmission delay of user c when the task IS transmitted...".

Response:
Thank you for your comment. Based on this comment, we have added “is” in this sentence.

■ Comment 7:
Typo after eq. (13): F \in C, C may be calligraphic.

Response:
Thank you for your comment. Based on this comment, we corrected C to calligraphic format.

■ Comment 8:
Typo in the following line: "...data. We...", remove the stop.

Response:
Thank you for your comment. To address this comment, we have removed the stop after “… data”.

■ Comment 9:
Use the "log" notation in eq. (15) for consistency.

Response:
Thank you for your comment. Based on this comment, we have used the "log" notation

throughout the paper.

■ Comment 10:

In Section II.E or III, introduce a little the concept of federated learning for non-experts.

Response:

Thank you for your comment. Based on this comment, we have added a brief introduction to federated learning in Subsection II.E.

■ Comment 11:

Typo: "The federated model then MUST be sent...".

Response:

Thank you for your comment. Based on this comment, we have added "must" to this sentence.

■ Comment 12:

Please, rewrite: "Since more training rounds and the more important parameters should have higher privacy sensitivity".

Response:

Thank you for your comment.

We have rephrased and further explained this sentence in the following. "By increasing the number of training epochs, the parameters of the training model become closer to the final trained model. Therefore, before training concludes, the version of the model obtained from additional training epochs holds greater significance than the model from earlier training epochs. Furthermore, the more critical parameters should exhibit higher privacy sensitivity. We denoted the privacy leakage for TST b's encoder training by …"

■ Comment 13:

Problem (27) is not explained. Please explain the cost and the constraints.

Response:

Thank you for your comment. The aim of problem (27) is to select the appropriate satellite to perform update training of the semantic coder. In this formulation, we jointly incorporate TSTs' training delay and energy consumption. The considered cost in problem (27) is therefore set as the training transmission and propagation delay plus the total energy consumption of transmission from TSTs to a satellite. We further considered the weighting factors to balance latency and energy consumption. The constraints for this optimization problem include the maximum acceptable service time due to the mobility of the satellite, the maximum tolerable service interruption delay, and the satellite selection factor.

Based on this comment and for further clarification, we have provided a more detailed description of the problem's cost and constraints before and after the formulation of the problem

(27).

■ Comment 14:
Revise Algorithm 2, it has many typos and incorrect indices.

Response:
Thank you for your comment. Based on this comment, we have thoroughly proofread Algorithms 1 and 2, and corrected the typos and grammar issues.

■ Comment 15:
In "The objective of PSFed during training...", change "objective" by "optimization problem".

Response:
Thank you for your comment. Based on this comment, we have revised the “objective” in "The objective of PSFed during training..." to “optimization problem”.

■ Comment 16:
In eq. (30a), remove "arg".

Response:
Thank you for your comment. Based on this comment, we have removed the “arg” in Eq. (30a)

■ Comment 17:
Problem (30) is not explained. Please explain the cost and the constraints.

Response:
Thank you for your comment. The objective of the problem (30) is to maximise training accuracy, i.e., minimize the training loss function, using our proposed PSFed. Therefore, the cost function is a loss value formulated according to our PSFed based on Section II-E (System model-Semantic coder updating)

The constraints in this optimization problem are the time constraints due to satellite mobility and service interruption tolerance. Further, we consider the privacy leakage in this problem and add that as a constraint.

To address this comment and avoid confusion, we have provided a more detailed description of the problem's cost and constraints before and after the formulation of the problem (30).

■ Comment 18:
The definition of M_{b,r} should be moved to problem (27).

Response:
Thank you for your comment. We strongly agree with and appreciate your suggestion, we have

removed the definition of M_{b,r} to problem (27).

■ Comment 19:

Typo in gamma variables after eq. (35): a bracket is missing.

Response:

Thank you for your comment. To address this comment, we have added the bracket after gamma variables.

■ Comment 20:

In eq. (36a), should Phi_1 be Phi_{c1}?

Response:

Thank you for your comment. Based on this comment, we have replaced *Phi_1* with *Phi_{c1}*.

■ Comment 21:

In eq. (36e), one of the variables x should be power.

Response:

Thank you for your comment. To address this comment, we corrected Eq. (36e) where one variable is *x* and the other is power.

■ Comment 22:

Typo: "γ3cand γ4c", space missing.

Response:

Thank you for your comment. Based on this comment, we have added the space between "*γ3*" and "and".

■ Comment 23:

"local computing capability and transmission power etc." Remove "etc." Again after eq. (38).

Response:

Thank you for your comment. Based on this comment, we have removed the "etc." in this sentence.

■ Comment 24:

In IV.D: why eq. (36) can be considered 5A?

Response:

Thank you for your comment. Here, *5A* means the decision problem in Eq. (36) can be considered as (5 times A) independent subproblems. Further, *5* is five offloading decision subproblems, i.e., 1) local computing; 2) offloading the tasks to SEC directly; 3) offloading the tasks to SEC via the TST; 4) offloading the tasks to the terrestrial cloud only via the satellite;

and 5) offloading the tasks to the terrestrial cloud via the TST and the satellite, where A is A satellite selection subproblems. This is because we assume that there are A satellites that can be selected for TST b.

To clarify, we provided a more detailed and further description of the meaning of $5A$.

■ Comment 25:

Problem (39) has a closed-form expression. Please, study the convexity of the problem and present the solution.

Response:

Thank you for your comment. Based on this suggestion, we have analysed the convexly of problem (39) and provided the solution.

■ Comment 26:

"The dual function is...". No, this is the "dual problem".

Response:

Thank you for your comment. Based on this comment, we have revised the “dual function” to “dual problem”.

■ Comment 27:

It may be appropriate to point out that eq. (45) is complementary slackness.

Response:

Thank you for your comment. Based on this comment, we have pointed out that Eq. (45) is complementary slackness.

■ Comment 28:

"The computation capabilities of SEC on satellite a...", "a" should be a variable.

Response:

Thank you for your comment. To address this problem, we substituted “a” with a variable format.

■ Comment 29:

References to Fig. (6) and (7) are swept. Please, fix it.

Response:

Thank you for your comment. This is due to our inadvertent reversal of the order of presentation of Fig. 6 and Fig. 7. To address and for further clarification, we have amended the order of Fig. 6 and Fig. 7.

[1] Z. Song, Y. Hao, Y. Liu and X. Sun, "Energy-Efficient Multiaccess Edge Computing for

Terrestrial-Satellite Internet of Things," IEEE Internet of Things Journal, vol. 8, no. 18, pp. 14202-14218, 15 Sept.15, 2021.
[2] R. Deng, B. Di, S. Chen, S. Sun and L. Song, "Ultra-dense LEO satellite offloading for terrestrial networks: How much to pay the satellite operator?", IEEE Trans. Wireless Commun., vol. 19, no. 10, pp. 6240-6254, Oct. 2020.

- # Response to the comments from Review 2

■ Comment 1:
Advantages:
++ Innovative Concept: The central idea of integrating SemCom with SEC is both intriguing and innovative. Such novel integrations have the potential to pave the way for advanced research in the domain.
++ Logical Flow: The paper presents its ideas with a logical progression, moving from the challenges of the current system to the proposed solutions, making it easier for readers to follow the authors' line of thought.
+ Well-Structured: The paper's structure is commendable. Each section has a defined purpose and collectively they form a coherent narrative.
+ Clarity in Presentation: Despite the complexity of the topic, the paper does an admirable job of presenting the information with clarity. The use of figures, tables, and other visual aids further enhances this clarity.
+ Robust Methodology: While there are areas for improvement, the methodology's attempt to merge SemCom with Federated Learning in SEC networks is commendable. Such an integration, when refined, could offer unique solutions to existing challenges in edge environments.
+ In-depth Analysis: The paper delves deep into the intricacies of the proposed integration. From latency to energy consumption and privacy, it presents a well-rounded view of potential benefits and challenges.

In conclusion, while there are areas of the paper that require further refinement, the underlying concept is strong, and the presentation is organized and structured. With focused revisions addressing the provided feedback, this paper has the potential to make a significant contribution to the field.

Response:
Thanks for the kind words.

■ Comment 2:
weakness:
The introduction could benefit from some streamlining for clarity. At times, it felt a bit dense, especially for readers not deeply familiar with the topic. Can consider using subheadings or paragraph breaks to differentiate key topics, e.g., "Background", "Challenges", etc. When transitioning between different topics, transitional sentences can make the narrative smoother.

Response:
Thank you for your comment. Based on this comment, we have added subheadings in "Introduction". Furthermore, we have revised the transition between different topics to improve the readability of the text.

■ Comment 3:
While the paper does a good job introducing MEC and the associated challenges, a brief description of what "SemCom" (Semantic Communication) actually entails might be beneficial for readers not familiar with the concept.

Response:
Thank you for your comment. Based on this comment, we have added an extra paragraph in “Introduction” to present the concept of semantic communication.

■ Comment 4:
Ensure that abbreviations are consistently used once they are defined. For instance, after introducing "satellite-borne edge cloud (SEC)", always use "SEC" afterward.

Response:
Thanks for your comment. Based on this comment, we checked the paper thoroughly and made sure that all abbreviations were consistently used as defined.

■ Comment 5:
When discussing the challenges like latency, energy consumption, and privacy, it might be helpful to provide real-world examples or scenarios where these challenges manifest. This can help readers understand the practical implications.

Response:
Thank you for your comment. Based on this comment, we have presented real-world examples or scenarios for each challenge in the “Introduction”.

■ Comment 6:
The term "novel" is indeed used frequently, particularly in the later part of the introduction where the authors outline their contributions. However, the true novelty isn't thoroughly contrasted against existing methods. One reason is for the lack of related work discussion. While the introduction mentions several related works and their contributions, it's crucial to provide clear delineation points that make the proposed approach "novel". For instance, how does the SemCom-SEC framework differ from other frameworks that integrate semantic communication and edge computing? For readers, especially reviewers, the novel aspects should be clearly and succinctly highlighted. The introduction would benefit from a segment (a paragraph or a set of bullet points) dedicated to pinpointing the novelty of the presented work against existing solutions. Any claims of novelty should be substantiated either in the introduction or in the sections that follow. The paper should have comparisons, possibly both qualitative and quantitative, against existing methods.

Response:
Thank you for your insightful comment. The SemCom-SEC framework differs from existing methods as it integrates semantic communication and edge computing. The existing works' are limited in terms of integrating SemCom with edge computing frameworks. For instance, Qin et. al [1] proposed a general SemCom framework involving users and terrestrial base station edge cloud. In [2], the SemCom framework only involves users and terrestrial base station edge cloud. The difference is that the users in [2] need to provide information to the base stations for semantic extraction. These methods, however, involve only the user and the edge cloud. In SEC offloading scenarios, the SemCom for offloading framework needs the participation of the users, terrestrial-station-terminal, satellites, and terrestrial clouds. Therefore, the existing frameworks are not directly applicable to SEC offloading scenarios. The main contribution of our work is to propose the SemCom-SEC framework that enables the participation of these entities. Furthermore, the proposed framework incorporates the characteristics of SEC networks such as satellite mobility.

Based on this comment and for further clarification, we provided specific approaches and the design of the relevant references. We also added a new paragraph to summarise the challenges faced by the existing works considering the offloading scenario. We further highlighted the proposed novel details in each bullet point.

■ Comment 7:
The phrasing in certain parts, like "As mentioned previously,", could be avoided. Instead, direct referencing to the previous section or subsection would be clearer.

Response:
Thank you for your comment. Based on this comment, we removed the phrase "as mentioned previously" throughout the paper and have replaced it with references to specific chapters.

■ Comment 8:
It's essential to clearly state any assumptions made. For instance, why is OFDMA used in the user-TST link? A brief explanation or citation supporting this choice would be beneficial.

Response:
Thank you for your comment. To address this comment, we explained the advantages of OFDMA on the user-TST link and added more references in Section II. In addition, we made sure that in the revised paper all the hypotheses were explained.

■ Comment 9:
The motivation should be clearly described. E.g., Why SemCom Specifically?: While the potential advantages of using SemCom are mentioned (such as reduced communication costs), the paper might benefit from a more in-depth discussion on why SemCom is particularly suited for this problem over other methods or technologies. Clearer Problem Statement: While the challenges in integrating SemCom in SEC are outlined, a more explicit problem statement might help. What

exactly are the limitations of current SEC networks, and how does SemCom address these limitations? How does this integration impact the end-users?

Response:
Thank you for your comment.

In most cases, offloading massive computing tasks to SEC requires an extremely high transmission rate hence large throughput. Therefore, SEC computation offloading is challenged by the reliability of transmission and intrinsic limitation of accessible radio spectrum. We argue that it is essential to develop techniques to significantly improve the spectrum efficiency of SEC offloading systems while maintaining quality of service in offloading.

To the best of our knowledge, semantic communication is the most promising technology to exceed the Shannon limit. This is because it employs machine learning to extract the meaning of the transmission data and only transmits the meaning of the data, hence, the size of the transmitted data is greatly reduced. Furthermore, the received messages restored by machine learning are also more robust, i.e., messages are more reliable, than conventional communications [3]. Therefore, semantic communication provided unparalleled advantages in the considered scenario in terms of spectral efficiency and robustness to path loss. This enables users to offload more and more accurate information to the edge cloud within a limited time.

Based on this comment, we have highlighted the motivation of integrating SemCom in SEC networks in the "Introduction".

■ Comment 10:
The methodology's attempt to merge SemCom with Federated Learning (FL) in Satellite Edge Computing (SEC) networks is commendable. This integration is ambitious and, if effective, could present a novel approach to enhancing communication in edge environments. However, I still have some comments about the methodology part:

a. Scalability: How scalable is the proposed CTPS mechanism, especially when considering a massive number of users or large-scale satellite networks?

b. Complexity: The methodology combines both SemCom and FL in SEC networks. While this is ambitious and could lead to significant advancements, the complexity might make it hard for real-world implementation without thorough experimental validation. It would be better to provide some analysis on the complexity.

c. Game Theory Application: Given the use of the Rubinstein bargaining game, how will the system handle players/users who do not act rationally or make unpredictable decisions?

Response:
Thank you for your comment.

In terms of scalability, CTPS is designed for multiple users. Increasing the number of users does not affect the number of calculations of CTPS. Furthermore, although increasing the number of satellites linearly increases the number of CPTS calculations, at any given time instance, there exist only limited satellites available in the sky per user [4].

Regarding the complexity, we strongly agree with the reviewer's comment. Nevertheless, semantic communication is considered and demonstrated as an important enabling technology for next-generation wireless networks [3]. Hence, it can be also used in satellite-air-terrestrial networks which is one of the key application scenarios for next-generation wireless communications. In terms of actual complexity, the convergence rate of FL is $O(1/T)$ [5], where T is the training times of participants. Therefore, FL can be used to train & update semantic coders more quickly and accurately. Furthermore, the proposed method reduces the space complexity of FL for satellite-air-terrestrial networks due to reducing the communication cost. Since, CTP can be also used in large-scale satellite networks, we argue that the integration of these technologies is not only feasible but also necessary.

Finally, regarding the considered game model and reasonability of the players' decisions, we simply assume that all the users are trustworthy. We also note that there is extensive literature investigating the identification of anomalous users. Therefore, although we appreciate the relevance of this issue, we believe this falls outside of the scope of this work.

Based on this comment, we discussed the potential of our proposed CTPS in large-scale satellite networks in Section IV-D. Furthermore, we added the adopted security assumption during the Rubinstein bargaining game in Section IV-B.

■ Comment 11:

One of my prevailing concerns is anchored in the simulation and methodology sections, which, in turn, raises foundational questions about the entire paper. The system model and problem formulations intricately explore federated learning while seemingly sidestepping the distinct challenges and opportunities presented by SemCom. As a result, the integration of SemCom into the proposed framework feels superficial, both in theoretical discussion and in the evaluations. We cannot know the actual improvement of using this strategy regarding the data size compression, latency, and accuracy after encoding-decoding. In the context of research, especially when introducing a novel integration such as the one between SemCom and SEC, it's crucial to provide a comprehensive evaluation that touches upon all major aspects of the proposed solution. Although the rainfall experiments show the robustness and efficiency of SemCom under changing environmental conditions, it doesn't directly address the specific advantages of SemCom in terms of data size, latency, or accuracy after encoding/decoding. It mainly highlights the robustness of SemCom in adverse weather conditions, which is not the complete picture. And specific details regarding which semantic encoder or decoder model they used, or how it was implemented, are not explicitly mentioned.

Here are some of the key areas that appear to be missing or not elaborated upon sufficiently in the provided text:

a. Semantic Compression Evaluation: SemCom, at its core, is about the semantic compression of data. There should be a detailed examination of how the proposed methods compress data, what the compression ratios are, and the impact of this compression on end-to-end system performance.

b. Latency after Encoding-Decoding: It would be valuable to understand the added latency introduced by the encoding-decoding process, especially since the proposed system focuses on computational offloading.

c. Quality of Reconstruction: While they do use metrics like PSNR for assessing quality, a more thorough examination of the encoding and decoding effects in terms of how well the decoded data matches the original and the impact of this process on the overall system performance would have been beneficial. Providing some visual examples before/after semantic encoding decoding is necessary.

d. Comparison with Non-Semantic Methods: For any evaluation to be meaningful, it's important to benchmark against traditional non-semantic methods to demonstrate the efficacy of using SemCom.

Although I understand it may be hard to cover all these aspects, the simulation shouldn't come across as solely centered on federated learning. Such an approach inadvertently undermines the paper's foundational motivations. Incorporating even rudimentary evaluations in this area could remarkably elevate the study's merit.

Response:

Thank you for your comment. Semantic communication (SemCom) is a relatively new communication paradigm. The main objective of SemCom is to significantly improve data compression rates hence increasing spectrum efficiency and exceeding the Shannon capacity. In general, there exist two important research challenges in this area including SemCom transceiver design for end-to-end communication and SemCom systems design for networks [3].

SemCom transceiver design for end-to-end communication aims to develop a pervasive SemCom transceiver for different content transmission. Ideally, such a transceiver can be used in any communication system similar to a conventional communication transceiver. Here the objective is to demonstrate the advantages of the designed SemCom transceivers over conventional communication transceivers. This advantage is often constituted not only in more efficient spectrum utilization but also in improving transmission robustness and efficiency.

At the network level, the objective design of SemCom systems is to investigate the distinct challenges and opportunities of SemCom transceiver applications in different networks and connectivity scenarios. This is mainly because SemCom transceivers have been shown to have advantages over conventional transceivers. Nevertheless, SemCom is based on machine learning (ML) technology which converts part of the communication load into the computation load. Therefore, the network resource allocation architecture is fundamentally different. The reallocation of various networks' resources is therefore one of the main challenges in SemCom studies [3]. Furthermore, SemCom also requires real-time updating of machine learning models. Therefore, designing novel real-time distributed learning approaches in different networking scenarios is also considered a challenge in SemCom research [1], [2].

In this paper we adopted the second approach and based on the research objectives of SemCom network system design, assuming that SemCom is actually in service, we considered issues such as the additional latency introduced by SemCom and various satellite access methods. To do this we proposed an SEC network resource allocation scheme. Furthermore, in case SemCom coders (i.e., transceivers) need updating, we considered relevant issues such as satellite mobility, and terrestrial-station-terminal quality of service. The special challenge of SemCom coder updating, e.g., encoder privacy, is also considered. To do this, PSFed is proposed for SemCom coders updating in SEC networks. Simulation results demonstrate the effectiveness of our proposed approaches in addressing distinct challenges of SemCom in SEC networks.

In this paper we adopted the second approach and based on the research objectives of SemCom network system design, assuming that SemCom is actually in service, we considered issues such as the additional latency introduced by SemCom and various satellite access methods. To do this we proposed an SEC network resource allocation scheme. Furthermore, in case SemCom coders (i.e., transceivers) need updating, we considered relevant issues such as satellite mobility, and terrestrial-station-terminal quality of service. To do this, PSFed is proposed for SemCom coders updating in SEC networks. Simulation results demonstrate the effectiveness of our proposed approaches in addressing distinct challenges of SemCom in SEC networks.

The SemCom coder model we utilised is based on a classical SemCom coder/transceiver model, i.e., [6]. Similar to other SemCom coder studies, this coder improves end-to-end transmission performance. We intercepted some of the simulation results from [6] as follows:

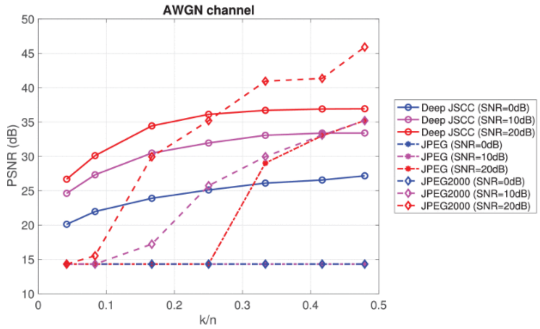


Fig. 1 Performance comparison of SemCom coder with conventional coder in different compression ratio k/n [6].

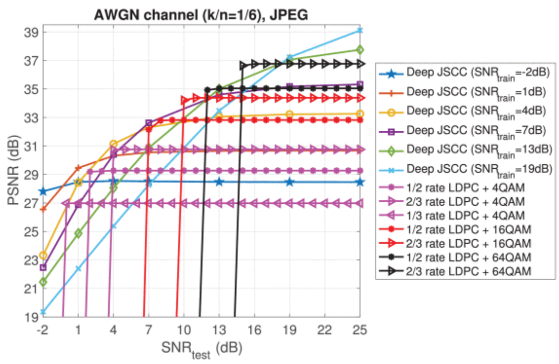


Fig. 2 Performance comparison of SemCom coder with conventional coder in different signal-to-noise ratio (SNR) [6].

As it is seen, SemCom has an absolute advantage over conventional communication in terms of compression ratio, and accuracy (PSNR) in end-to-end transmission.

(a) FedAvg (b) PSFed (c) FedRep

Fig. 3 Visual examples of PSFed and FedRep after semantic coding.

We also strongly agree with you that a thorough examination of the effects of encoding and decoding is essential. We therefore used the widely recognised picture evaluation method PSNR for this purpose. We considered visual examples, but the results were not intuitive (Fig. 3). The image of Fig. 3 is from the CIFAR 10 dataset and is a 32*32 pixel blur. The label of this image is "ship". We can observe that the clarity of FedAvg and PSFed is essentially the same. But the stern of the boat in FedRep appears blurred (in the red box). This is because SemCom not only greatly compresses the transmitted bits but also greatly improves the quality of the transmission. The visual examples of models based on different SemCom techniques are not perceptive to the human eye. In our previous work [6], instead, we proposed to use proven machine learning image recognition models to evaluate the recognition accuracy of images before and after transmission/offloading via SemCom. This can be utilised to assess the total impact of different semantic communication technologies on our considered offloading systems. It further allows evaluation of the image transmission quality from the perspective of the machine as this paper is on data offloading in the IoT environments. The numerical results are shown in Fig.4. The accuracy here is the proportion of the received object/image recognition accuracy to the pre-transmission image recognition accuracy. It is seen that for the offloading scenario, PSFed achieves similar accuracy to FedAvg while reducing communication load. Further, the accuracy of PSFed is much better than FedRep. We, therefore, argue that using the approach in [7] provides an intuitive demonstration of the impact of SemCom on the quality of transmission/offloading in an offloading scenario.

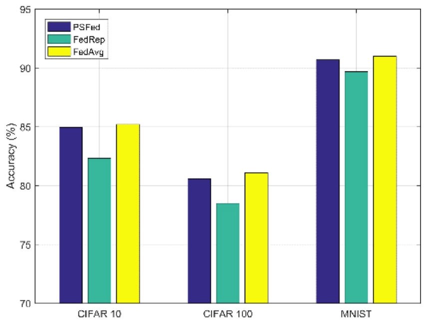


Fig. 4 Image recognition accuracy

To address this comment, we highlighted the difference between SemCom transceiver design for end-to-end communication study and SemCom systems design for networks study in “Section II-A”. We also highlighted the advantage of SemCom over conventional communication in Section I. In addition, based on item *c* in this comment, we provided image recognition accuracy before/after semantic encoding decoding to further demonstrate the efficiency of our proposed PSFed. Regarding item *b* in this comment, the extra delay introduced by SemCom has been considered in our formulation and the CTPS method was proposed to optimize the network resource.

Regarding items a and d in this comment, we appreciate that these comments aim to provide a comprehensive assessment of SemCom. We also note that the submitted paper is solely focused on the unique challenges of the SemCom system for SEC networks. The pervasive SemCom model employed in this paper has been demonstrated to be advantageous regarding items *a* and *d*. We, therefore, leave investigating *a* and *d* to future works.

■ Comment 12:

The presentation of the figures lacks visual coherence and alignment. Some figures appear excessively large and lack detail, making them less informative. The limited information in the figures and their captions does not convey results clearly and efficiently to readers. Figures 5-12, in particular, are quite basic; it might be beneficial to group some of them together or align multiple figures on the same line rather than dedicating an entire line to each. Enhancing figure captions and the content within the figures can not only improve clarity but also free up space for more critical sections of the paper.

Response:

Thank you for your comment. Based on this comment, we have improved the presentation of the figures in the revised version of the paper, see, Fig. 5 - Fig. 12.

[1] Z. Qin, G. Y. Li and H. Ye, "Federated Learning and Wireless Communications," IEEE Wireless Communications, vol. 28, no. 5, pp. 134-140, October 2021.

[2] G. Shi, Y. Xiao, Y. Li and X. Xie, "From Semantic Communication to Semantic-Aware Networking: Model, Architecture, and Open Problems," IEEE Communications Magazine, vol. 59, no. 8, pp. 44-50, August 2021.

[3] Qin, Z., Tao, X., Lu, J., & Li, G. Y. (2021). Semantic communications: Principles and challenges. arXiv preprint arXiv:2201.01389.

[4] R. Deng, B. Di, S. Chen, S. Sun and L. Song, "Ultra-Dense LEO Satellite Offloading for Terrestrial Networks: How Much to Pay the Satellite Operator?," IEEE Transactions on Wireless Communications, vol. 19, no. 10, pp. 6240-6254, Oct. 2020.

[5] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, “On the convergence of FedAvg on non-IID data,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2020, pp. 1–26

[6] E. Bourtsoulatze, D. Burth Kurka and D. Gündüz, "Deep Joint Source-Channel Coding for Wireless Image Transmission," IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 3, pp. 567-579, Sept. 2019.

[7] G. Zheng, Q, Ni, K. Navaie, H. Pervaiz, and C. Zarakovitis. "A Distributed Learning Architecture for Semantic Communication in Autonomous Driving Networks for Task Offloading", IEEE Communication Magazine.

• Response to the comments from Review 3

■ Comment 1:

The paper is well written and easy to read. Some revision is needed in Section II.E "In generally", Fig.2 Purn instead of "Purning" and Fig.6 and Fig.7 are swapped.

Response:
Thank you for your kind words. To address this comment, we corrected the above typos and amended the order of Figs. 6 and 7 are also swapped. In addition, we have thoroughly proofread the manuscript and corrected the typos and grammar issues.

- Comment 2:

The authors may also consider the extension of the used model to include collaborative computing of the nearby users with free resources. The satellite availability (due to movement) time to the user as well as the base station need to be taken into account and an assumption that every transmission period of the base station is less that the satellite appearance time.

Response:
Thank you for your insightful comment. We have discussed the satellite availability time in both Section III and Section IV. However, due to our oversight, it was not explicitly represented mathematically.

To address this comment, we added more discussion on satellite availability time for users and TSTs in Section II, Sections III and IV. Furthermore, we added the constraint of satellite availability time mathematically in Problems (27), (30) and (36).

Regarding extending the used model, we believe including collaborative computing with terrestrial free resources is a great idea providing further opportunities for the development of satellite edge computing. This submitted paper is however based on a common satellite edge offloading scenario without free terrestrial resources, same as [1], [2], and [3]. For instance, in a desert, ocean or disaster zone, users are not likely to be supported by additional terrestrial free services. Hence, we leave investigating the impact of the nearby free terrestrial resources on satellite edge offloading based on semantic communication to future works.

- Comment 3:

The privacy leakage model in Section III needs more justification and explanation rather than just refereeing to the reference.

Response:
Thank you for your comment. Previous studies (see, e.g. [4]) show that when reconstructing an ML model increasing the number of parameters increases the accuracy of the model following a logarithmic function. In SemCom, the accuracy of the SemCom coder represents the accuracy of the received data. Therefore, the privacy of the coder model/parameter is closely tied to the accuracy. We can adopt a general parameter privacy leakage metric as in [5] and assess model parameter leakage by

$$\Phi_b(\theta_b) = \chi log_2(1 + e^{1-\frac{N_b+1}{n_b}}), \quad (25)$$

where χ is the weight parameter, N_b is the total number of parameters at the encoder model and n_b is the number of transmitted parameters. In practice Φ_b adopts a value in [0,1], where Φ_b =0 indicates that there is no privacy leakage, while a Φ_b =1 indicates fully compromised privacy where the same information can be decoded from the leaked model as the original model.

By increasing the number of training epochs the parameters of the training model become closer to the final trained model. Therefore, the model obtained from more training epochs is more important relative to the model obtained from previous training epochs before training is finished. Therefore, the private information contained in the parameters is increased over time. More important parameters bear higher sensitivity in terms of privacy. Therefore, we rewrite the privacy leakage for TST b's encoder training as:

$$\Phi_b(\theta_b) = \sum_{r=1}^{R} W_r \, \chi log_2(1 + e^{1-\frac{\sum_i^{N_b} I_i n_{b,i}+1}{\sum_{j=1}^{n_b} I_j n_{b,j}}}), \tag{26}$$

where r is the communication rounds and R denotes the total number of communication rounds (epochs). Also, W_r is a weight representing the model importance in training round r. Similarly, I_i is a weight parameter denoting the importance transmitted parameter i.

Based on this comment and for further clarifications, we added our justifications and explained the privacy leakage model in Section III.

■ Comment 4:

The authors claims that as the FedRep converges much slowly that PSFed, the total communication resources can be considered the same. This is part of the novelty introduced by this paper and more accurate calculation is needed in that aspect by generate a table or figures that combine results from Fig.4 and Fig.5.

Response:

Thank you for your comment. The advantage of convergence speed is difficult to specifically quantify with dataset discrepancies and with the random nature of training. Nevertheless, the convergence speed of our PSFed is superior to FedRep in any of the datasets. Moreover, it can be observed in Fig. 5 that FedRep saves ten communication rounds of computational resources relative to our PSFed. Combining these two, in the paper, we argue that the required communication resource for training in the two methods is "similar", not the "same". Importantly, our PSFed achieves significant improvement in accuracy compared to FedRep for similar communication resource consumption.

To address this comment and to clarify, we highlighted the cost of required communication resource in the two approaches are "similar", not the "same" and we presented our argument. In addition, we added discussions of our method and FedRep in terms of convergence speed, communication cost and accuracy, combining the results in Fig. 5, Fig. 6, and Fig. 7.

■ Comment 5:

Averaging the result over 50 simulation might be not sufficient and more simulations are needed.

Response:

Thank you for your comment. We strongly agree with you and appreciate your suggestion. Based on this comment, we repeated the simulation 200 times.

[1] Q. Tang, Z. Fei, B. Li and Z. Han, "Computation Offloading in LEO Satellite Networks With Hybrid Cloud and Edge Computing," IEEE Internet of Things Journal, vol. 8, no. 11, pp. 9164-9176, 1 June1, 2021.

[2] Z. Song, Y. Hao, Y. Liu and X. Sun, "Energy-Efficient Multiaccess Edge Computing for Terrestrial-Satellite Internet of Things," IEEE Internet of Things Journal, vol. 8, no. 18, pp. 14202-14218, 15 Sept.15, 2021.

[3] S. S. Hassan, Y. M. Park, Y. K. Tun, W. Saad, Z. Han and C. S. Hong, "Satellite-Based ITS Data Offloading & Computation in 6G Networks: A Cooperative Multi-Agent Proximal Policy Optimization DRL With Attention Approach," IEEE Transactions on Mobile Computing.

[4] Z. Chen, T. -B. Xu, C. Du, C. -L. Liu and H. He, "Dynamical Channel Pruning by Conditional Accuracy Change for Deep Neural Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 799-813, Feb. 2021.

[5] R. Xing, Z. Su and Y. Wang, "Intrusion Detection in Autonomous Vehicular Networks: A Trust Assessment and Q-learning Approach," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 2019, pp. 79-83.

• Response to the comments from Review 4

■ Comment 1:

It is not obvious why on the last paragraph of page 3, authors has mentioned that semantic communication is necessary to be updated due to the mobility of the users. Why the information transmitted depends on the mobility of the users.

Response:

Thank you for your comment. In this paper, we considered an offloading scenario. The user transmits tasks that require offloading to a satellite or terrestrial cloud for faster processing. These tasks may include scene perception or augmented reality functions. For example, offloading scene perception tasks involves transmitting images or videos of the scene to the edge cloud for processing. Since scenes change as the user moves, the semantic coder is content-oriented and needs to be updated with different content. Consequently, changes in the images and videos of the scenes affect the transmitted content. As a result, the mobility of users leads to the updating of multiple content-oriented semantic coders.

Based on this comment and to clarify, we revised the presentation of page 3’s last paragraph. In addition, we explained why user mobility affects changes in transmitted information and updates to semantic coders.

■ Comment 2:

What is d in (5)?

Response:

Thank you for your comment. The variable d, i.e., d_0 in Eq. (5) represents a "sub-carrier d_0" in the link of user c to TST b. In Eq. (5), $B_{d_0}^{cb}, p_{c,d_0}^{cb}$ and g_{c,d_0}^{cb} are bandwidth, transmission power and the channel gain on this sub-carrier d_0 in the link of user c to TST b.

Based on this comment and to clarify, we have reformulated the introduction of the parameters in this paragraph and highlighted what is d_0 in Eq. (5).

■ Comment 3:

In equation (9) and (12), how is the channel gain defined? and how is the channel gain distinguished from path loss?

Response:

Thank you for your insightful comment. We view channel gain as a factor influencing transmission power variation due to channel characteristics. Path loss is indeed a significant component that contributes to the channel gain. However, we aim to better illustrate to our readers the benefits of semantic communication under various path loss scenarios and mathematically establish why it offers advantages compared to conventional communication. Therefore in Eq. (9) and Eq. (12), similar to [1], we separated the path loss. The channel gain here is the gain with the path losses removed.

To address this comment, we added the definition of channel gain after Eq. (9) and explained why we list the path loss separately rather than integrating it into channel gain in the formula.

■ Comment 4:

Authors should give the reference to (12).

Response:

Thank you for your comment. We mistakenly wrote semantic communication processing time from Eq. (13) into Eq. (12). However, this mistake does not affect our subsequent derivations and simulations. Here, Eq. (12) is similar to Eq. (9) and is still based on the conventional Shannon transmission theorem. We thus have not provided relevant references.

Based on this comment and to clarify, we deleted the incorrectly written, i.e., semantic communication processing time in Eq. (12). Further, we have thoroughly proofread the manuscript, and corrected such mistakes.

■ Comment 5:

The delay in (13) is the transmission delay of all offloaded users to satellite a and it can not say that it is the transmission delay of user c. Authors needs to clarify this.

Response:
Thank you for your comment. The transmission delay in Eq. (13) is the transmission delay of all users.

To address this comment, we have clarified the definition of Eq. (13).

■ Comment 6:
Some texts are not complete say "Since the computation task calculation result is often much smaller than the offloaded data." in page 4.

Response:
Thank you for your comment.

To address this comment, we replaced the full stop with a comma and rephrased the sentence as: “Since the computation task calculation result is often much smaller than the offloaded data, we thus ignore the backhaul transmission delay links similar to [23] and [24]”.

■ Comment 7:
Maximization in (27a) is taken on which parameter?

Response:
Thank you for your comment. We can express the Eq. (27a) as:

$$\min_{x_a} \sum_{a=1}^{A} x_a \left(\alpha \max\left\{\frac{M_{b,r}}{R_b^{ba}} + \frac{2h^{ba}}{c_l} \middle| b \in \mathcal{B}\right\} + \sum_{b=1}^{B} \beta\, p_b^{ba} \frac{M_{b,r}}{R_b^{ba}}\right), \tag{27a}$$

where $\max\{\frac{M_{b,r}}{R_b^{ba}} + \frac{2h^{ba}}{c_l} | b \in \mathcal{B}\}$ is the training transmission delay ($\frac{M_{b,r}}{R_b^{ba}}$) and propagation delay ($\frac{2h^{ba}}{c_l}$), identified by the terrestrial-station-terminal (TST) with the longest transmission and propagation time. Moreover, $\sum_{b=1}^{B} p_b^{ba} \frac{M_{b,r}}{R_b^{ba}}$ is the total energy consumption of transmission from TSTs to a satellite. Further, α and β are weight parameters to balance the importance and unit of latency and energy consumption. The entire Eq. (27a) is taking into account TSTs' training delay and energy consumption jointly.

To address this comment and for clarification, we have presented the significance of each parameter in detail and highlighted the design and the meaning of Eq. (27a).

■ Comment 8:
Authors should explain constraint (27b).

Response:

Thank you for your comment. We can express the Eq. (27b) as:

$$\sum_{r=1}^{R} \frac{M_{b,r}}{R^{ba}} < t_b', \forall\, b \tag{27b}$$

where $M_{b,r}$ is the coder model size in communication round r and t_b' is the maximum tolerable service interruption time. Moreover, R^{ba} is the transmission rate from TST b to satellite a and R is the total training communication rounds. The constraint Eq. (27d) denotes the transmission time of the TST for training the semantic model to be less than the maximum tolerable service interruption time.

To address this comment, we have explained constraint (27b) in detail.

■ Comment 9:

It seems that processing choice 4 in page 7 wants to offload data to the cloud through satellite. This is while they have mentioned that this processing is for directly offloading date from user to the cloud. Authors should clarify this.

Response:

Thank you for your insightful comment. Choice 4 is to offload the user's task to the terrestrial cloud through the satellite.

Based on this comment, we revised the presentation of choice 4 and clarified that this choice is to offload data to the cloud through satellite.

■ Comment 10:

In equation (34), why propagation from satellite to the cloud is not considered?

Response:

Thank you for your comment. In Eq. (34), t_c^{proC} is the propagation delay for user c who chooses to offload the task to the terrestrial cloud. We defined it in Eq. (14).

$$t_c^{proC} = \frac{4h}{c_l}, \tag{14}$$

where c_l is the speed of light and h is the distance between users and satellite a. Due to the mobility of satellites, the distance from the satellite to the terrestrial cloud is approximated by h. In addition, here, we consider 4 propagation processes, i.e., user-satellite, satellite-cloud, cloud-satellite and satellite-cloud. Therefore, we already consider the propagation delay from satellite to the cloud.

Based on this comment, we highlighted the propagation delay process considered in Eq. (14).

■ Comment 11:

What is t_a^{cloud} which has been used in (35)? It is not defined before.

Response:

Thank you for your comment. Here, *t_a^{cloud}* is the transmission delay between satellite and cloud. It has been defined at the end of Section II-C.

Based on this comment and for further clarification, we have added a reminder of the definition of *t_a^{cloud}* after Eq. (35).

■ Comment 12:
What is the value that has been considered for the energy factor ε in the simulation result section?

Response:
Thank you for your comment. The energy factor ε is set as 10^{-26} as in [2]. To address this comment, we have added this in Subsection V-A "simulation setting".

■ Comment 13:
It is suggested that authors explain in detail FedRep and FeAvg schemes as used for the benchmarks.

Response:
Thank you for your comment. Based on this comment, we have added detailed explanations of FedAvg and FedRep in Section II-E and simulation settings.

■ Comment 14:
In Fig. 8, why CTPS and "CTPS without game" perform nearly the same and what is the conclusion from this observation?

Response:
Thank you for your comment. This is because in case the number of users is small, the TST can satisfy all the number of sub-carriers requested by users. In cases where the number of users is small, the proposed CTPS thus maintains almost the same processing cost as "CTPS without game". By increasing the number of users, TST becomes unable to satisfy all the requests and CTPS starts to show its advantage in reducing the cost. We expect this advantage to increase by further increasing the number of users. This is because the optimal reallocation of resources through our design game scheme increases the efficiency of network resource utilisation.

Based on this comment and for further clarification, we highlighted the relationship between CTPS and "CTPS without game" and provided the conclusion of the observation.

[1] K. Tekbıyık, G. K. Kurt and H. Yanikomeroglu, "Energy-Efficient RIS-Assisted Satellites for IoT Networks," IEEE Internet of Things Journal, vol. 9, no. 16, pp. 14891-14899, 15 Aug.15, 2022.
[2] Z. Song, Y. Hao, Y. Liu and X. Sun, "Energy-Efficient Multiaccess Edge Computing for Terrestrial-Satellite Internet of Things," IEEE Internet of Things Journal, vol. 8, no. 18, pp. 14202-14218, 15 Sept.15, 2021.

Semantic Communication in Satellite-borne Edge Cloud Network for Computation Offloading

Guhan Zheng, Qiang Ni, *Senior Member, IEEE*, Keivan Navaie, *Senior Member, IEEE*, and Haris Pervaiz, *Member, IEEE*

Abstract

The low earth orbit (LEO) satellite-borne edge cloud (SEC) and machine learning (ML) based semantic communication (SemCom) are both enabling technologies for 6G systems facilitating computation offloading. Nevertheless, integrating SemCom into the SEC networks for user computation offloading introduces semantic coder updating requirements as well as additional semantic extraction costs. Offloading user computation in SEC networks via SemCom also results in new functional challenges considering, e.g., latency, energy, and privacy. In this paper, we present a novel SemCom-assisted SEC (SemCom-SEC) framework for computation offloading of resource-limited users. We then propose an adaptive pruning-split federated learning (PSFed) method for updating the semantic coder in SemCom-SEC. We further show that the proposed method guarantees training convergence speed and accuracy. This method also improves the privacy of the semantic coder while reducing training delay and energy consumption. In the case of trained semantic coders in service, for the users processing computational tasks, the main objective is to minimise the users' delay and energy consumption, subject to sustaining users' privacy and fairness amongst them. This problem is then formulated as an incomplete information mixed integer nonlinear programming (MINLP). A new computational task processing scheduling (CTPS) mechanism is also proposed based on the Rubinstein bargaining game. Simulation results demonstrate the proposed PSFed and game theoretical CTPS mechanism outperforms the baseline solutions reducing delay and energy consumption while enhancing users' privacy.

Index Terms

Satellite-borne edge cloud, SemCom, computation offloading, delay, energy consumption, privacy.

I. Introduction

A. Background

MULTI-ACCESS edge computing (MEC) is emerging as one of the key techniques for next-generation wireless communication systems [2]. MEC enables the development of Internet of Things (IoT) applications and improves network performance and quality of service (QoS) [3]. MEC brings cloud services closer to the users at the network edge, e.g., base stations (BSs), and roadside units (RSUs) providing them with abundant computational resources. Therefore, users can offload their computationally intensive tasks to the MEC for faster processing.

Nevertheless, users located in remote areas or disaster zones might not be able to connect to terrestrial edge cloud network infrastructures. Alternatively, such under-served users may offload their computationally intensive tasks to remote core cloud servers via Geosynchronous Equatorial Orbit (GEO) or Medium Earth Orbit (MEO) satellites. In addition to the costs, the corresponding propagation latency to and from the satellite platforms however impedes the delay requirements of these users. Using Low Earth Orbit (LEO) satellites can partly address this issue by providing lower propagation latency as their orbits are much closer to the ground compared to GEO and MEO satellites. Comparing to GEO and MEO, constellations of LEO satellites also provide low-cost, high-throughput services and extensive radio coverage. To further reduce the propagation delay, the satellite-borne edge cloud (SEC) setting was proposed, where the offloaded processing is conducted on board the LEO satellite, hence reducing the propagation delay by a factor of 2 [4], [5].

Adopting SEC for users in remote areas or disaster zones has been recently investigated in [6] and [7]. The authors in [6], and [7] mainly focused on developing offloading decisions that minimise offloading delay or energy consumption for cases where users have direct radio links to the satellites. (e.g., in C-Band). An alternative access scenario is proposed in [8], where the user transmits to the SEC indirectly through an intermediary terrestrial-station-terminal (TST). In this approach, the user transmission to the TST is on a C-band radio link and TST communicates to the SEC through a K-band radio link. Wang et al. [9] also proposed a dual-edge cloud network, where the edge servers are placed on both BSs and LEO satellites. In this

G. Zheng, Q. Ni and K. Navaie are with the School of Computing and Communications, Lancaster University, LA1 4WA, UK (Email: g.zheng2, q.ni, k.navaie@lancaster.ac.uk).

H. Pervaiz is with the School of Computer Science and Electronic Engineering (CSEE), University of Essex, CO4 3SQ, UK (Email: haris.pervaiz@essex.ac.uk).

Part of this work was presented at the IEEE ICC 2022 DDINS workshop [1].

approach, a BS acts as a TST to assist users with computation offloading to the SEC. Similarly, [10] proposed an energy-efficient strategy for terrestrial users to offload computing tasks to the SEC via TSTs. Tang et al. [11] further investigated the impact of the core cloud on users' offloading decisions. They then proposed a minimal energy consumption computing offloading decision method, where users access SEC directly.

B. Challenges: SEC for user offloading

The approaches mentioned above frequently confine their investigations to a singular connectivity scenario between users and the SEC. In essence, by concentrating solely on specific performance aspects, such as energy consumption or latency, potential privacy concerns and associated risks to users are disregarded. This poses inherent risks to users. For example, prioritizing latency without considering energy consumption and privacy may lead to a user in the desert swiftly losing the ability to communicate, with this information potentially accessible by a third party. To address this issue, in this paper, we investigate SEC incorporating various access modalities, task processing entities, latency, energy consumption, and privacy of users.

Moreover, in the majority of instances, offloading substantial computing tasks to the SEC demands an exceptionally high transmission rate and substantial throughput. Consequently, alongside considerations of latency, energy efficiency, and data privacy, the computation offloading to SEC encounters a fundamental constraint—the inherent limitation of accessible radio spectrum. Hence, it is imperative to devise techniques that markedly enhance the spectrum efficiency of these systems, all the while upholding the quality of service (QoS) in the offloading process. A promising approach to address this issue is semantic communication (SemCom) based on machine learning (ML) [12].

SemCom leverages ML techniques for information transmission. A goal-oriented semantic encoder, powered by ML, selectively extracts semantic information from the transmitted or offloaded content. Rather than transmitting raw data, only the essential semantic information is conveyed, later decoded by the ML-based semantic decoder. This approach significantly enhances spectrum efficiency by balancing the communication load against the computational load through machine learning. Moreover, it mitigates the impact of unstable radio links, such as variable path loss due to weather conditions commonly observed in high-frequency satellite links. SemCom thus plays a pivotal role in the significant enhancement of the performance and speed of offloading. The integration of SemCom and SEC for computation offloading presents a promising solution to address the challenges of task offloading in the next generation of wireless communications.

C. Challenges: SemCom for SEC

Integrating SemCom and SEC for computation offloading requires a carefully designed architecture. Such an architecture needs to consider various possible task-processing entities (satellites and terrestrial cloud) and various user access methods (direct and indirect) to the SEC network. Furthermore, goal-oriented ML-based SemCom coders need to be updated in real-time according to new transmission content [13].

In the SEC network, updating the semantic coder presents several emerging challenges, e.g., mobility of SEC, low tolerance of service interruption and energy consumption, and privacy. However, the existing distributed learning frameworks designed for SemComs in generic networks (e.g., [14]–[16]) do not seamlessly translate to the SEC network. For instance, Xie and Qin [14] introduced a pruned lite ML model tailored for distributed semantic coders. Their approach focuses on refining trained models rather than updating goal-oriented coders. Similarly, Qin et al. [16] proposed a general SemCom framework involving users and terrestrial base station edge clouds. In [15], the SemCom framework also includes users and terrestrial base station edge clouds, with the distinction that users in [15] must provide information to base stations for semantic extraction. However, these frameworks suffer from prolonged service interruptions, increased energy consumption, and heightened privacy risks within SEC networks. Furthermore, these methods only engage users and the edge cloud. In SEC offloading scenarios, the SemCom for offloading framework necessitates the active participation of all parties including users, terrestrial-station-terminal, satellites, and terrestrial clouds. The aforementioned research underscores the critical need to develop efficient distributed learning methods for updating semantic coders in SemCom SEC networks.

In addition to the above, SemCom alters the transmission paradigm of SEC networks by increasing the computational load while reducing the communication load. Users are therefore required to develop optimal computational task strategies in case trained semantic coders are utilised for computation offloading. Such strategies need to be developed taking into account not only scenarios specific to SemCom in the SEC, but also operational factors that have not been considered in the existing SEC offloading research. Such factors include using both access modalities, the task processing entities, latency, energy consumption and privacy.

D. Contributions

To tackle the above-mentioned challenges, in this paper, we propose a novel SemCom-assisted SEC (SemCom-SEC) framework for terrestrial users' computation offloading. In our proposed method, we split the SemCom service into in-maintenance (i.e., semantic coders need updating) and in-service (i.e., trained semantic coders are utilised for computation offloading) scenarios. For the in-maintenance scenario, we investigate real-time updating of deployed semantic coders in SemCom-SEC.

A pruning-split federated learning (PSFed) approach is then proposed to update semantic coders considering offloading QoS while privacy-preserving. For the in-service scenario, we study the computational task processing challenge of terrestrial users in the new SemCom paradigm. We then propose a new computational task processing scheduling (CTPS) mechanism based on the Rubinstein bargaining game to minimise the users' processing delay and energy consumption while preserving users' privacy. The main contributions of this paper are summarised as follows:

- We integrate the SemCom and SEC networks and propose a novel SemCom-SEC framework enabling task offloading for under-served users. Diverging from current SemCom frameworks, which exclusively factor in users and terrestrial edge clouds, the envisioned framework extends its reach by deploying semantic coders on both the TSTs and satellites. Furthermore, SemCom-SEC accommodates a variety of user task-processing approaches and access modalities. Computational tasks for users can occur locally, at SEC, or in the core cloud server. Additionally, users possess the flexibility to access LEO satellites either directly or through the semantic encoder-equipped TST.

- We then propose a PSFed approach for semantic coder updating for the SemCom-SEC framework enabling computation offloading. PSFed adaptively "splits" and "prunes" the semantic coders for federated aggregation subject to various users' personalised conditions. In contrast to the conventional "split" and "prunes" models, the semantic coder model components remain intact after updating. PSFed reduces the consumption of training communication resources and improves the privacy of the trained encoder while enhancing the training convergence speed and model accuracy.

- We introduce an innovative CTPS mechanism, distinct from previous studies that only address partial performance considerations. Our approach takes a comprehensive stance, jointly addressing user privacy, delay, energy consumption, and fairness to tackle the novel challenge of incomplete information task processing scheduling in SemCom-SEC. The CTPS operates in two steps: firstly, a game-theoretic model is crafted to transform this mixed-integer nonlinear programming (MINLP) problem from incomplete information, stemming from privacy concerns, into a complete information problem. In the second step, the converted complete information MINLP problem is decomposed and solved through the application of the Lagrangian dual decomposition method.

The rest of the paper is organised as the following. Section II presents the system model of the proposed SemCom-SEC framework. In Section III and Section IV, we then investigate the unique challenges and corresponding solutions for SemCom in-maintenance and in-service scenarios, respectively. The performance of the proposed PSFed and CTPS are then evaluated and analysed by simulations in Section V. Finally, conclusions are drawn in Section VI.

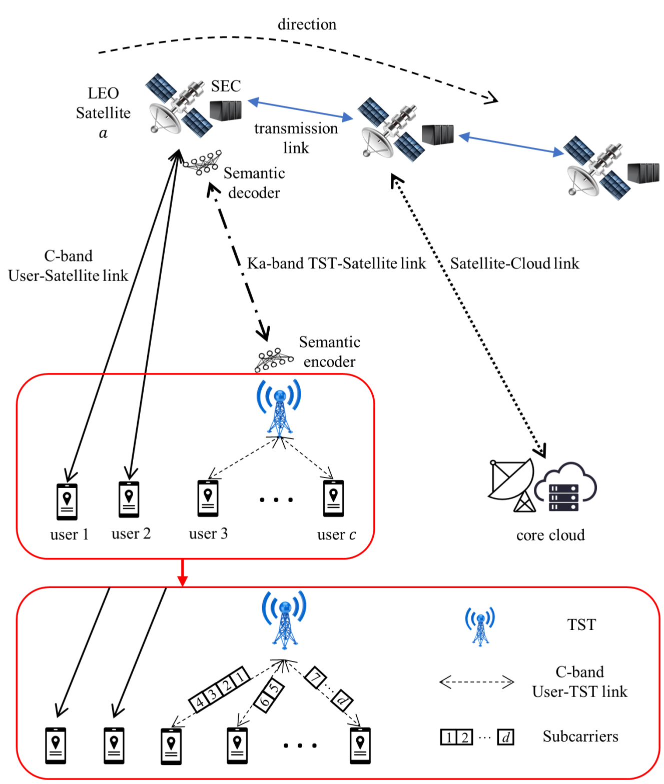


Fig. 1: The proposed SemCom-SEC framework.

II. System Model

In this section, the system model of the proposed SemCom-SEC is introduced. We then provide the computing, communication, path loss and semantic coder training model.

A. System description

Consider the SemCom-SEC (Fig.1), where terrestrial users are located in areas without having access to terrestrial edge service. Users can offload computation-intensive tasks to LEO SEC. In practice, an LEO satellite constellation is similar to a cellular network operating above the ground [17]. whereas the space cellular network is on the move, while ground users are relatively stationary.

We consider both types of approaches for users to access the SEC for computation offloading [8]. Users can communicate with LEO satellites directly through a C-band user-satellite radio link. Furthermore, they are also allowed to indirectly access the SEC through a TST via a C-band link to TST, and a Ka-band link between TST and SEC. The terrestrial C-band user-TST link spectrum resources are utilised in an orthogonal frequency division multiple access (OFDMA) setting to optimise the utilisation of terrestrial radio resources [10].

To improve the spectrum efficiency and QoS of SEC networks, semantic coders are deployed on the TSTs and LEO satellites for transmitting offloaded tasks over Ka-band. This is due to TSTs being primarily responsible for transmitting significant amounts of tasks to satellites and requiring extremely high spectral efficiency. Furthermore, their service area is fixed and the content to assist in task offloading (e.g., scene perception task, augmented reality task) only minimally varies. The mobility of the users causes the fact that the offloading content is often variable. For instance, the content of the transmission when offloading a scene perception task varies depending on the scene. The content-oriented semantic coders need to be constantly updated as the user moves. We thus consider factors such as utilisation, and reliability, for which goal-oriented SemCom is most appropriate for the TST-satellite link in SEC networks. Moreover, due to the dynamic nature of the system and the limited storage resources of LEO satellites, it is not viable to store semantic decoders for all TSTs on the route. The semantic coders are therefore stored on the TST. Similarly, for economic and satellite storage resources considerations, at least the trained decoder of TSTs should be the same for the same transmission task. The TST delivers the related semantic decoders to the corresponding satellite when it needs to perform SemCom. Furthermore, LEO satellites can alternatively connect to the cloud servers on the terrestrial network via Ka-band backhaul links to provide cloud service for users.

In this model, a user may process indivisible computational tasks in either of the following five scenarios: 1) computing locally; 2) offloading the tasks to SEC over the user-satellite link; 3) offloading the tasks to the SEC via TST; 4) offloading the tasks to terrestrial cloud over the user-satellite link; 5) offloading the tasks to the terrestrial cloud via TST-satellite link.

B. Computiong models

Denote the set of LEO satellites as $\mathcal{A} = \{1, 2, ..., a, ..., A\}$ and set of TSTs as $\mathcal{B} = \{1, 2, ..., b, ..., B\}$. A TST b is on the terrestrial and provides service to C users within the coverage as a small cell in which the set of users in TST b's service range is denoted by $\mathcal{C} = \{1, 2, ..., c, ..., C\}$. We consider each terrestrial user c to have indivisible computational sensitive tasks with the size in bits of $m_c \in \{m_1, m_2, ..., m_c, ..., m_C\}$, and the CPU cycles needed to execute one bit of tasks is δ. The local computation task latency of the user c can be given by

$$t_c^{LC} = \frac{\delta m_c}{f_c}, \tag{1}$$

where f_c is user c's CPU-cycle frequency with the unit cycles/s. The energy required to calculate locally is hence expressed as [1]:

$$E_c^{LC} = p_c^{LC} t_c^{LC} = \varepsilon f_c^3 \frac{\delta m_c}{f_c} = \varepsilon \delta m_c f_c^2, \tag{2}$$

where $p_c^{LC} = \varepsilon f_c^3$ is the power needed to be computing locally and ε is the energy factor related to the electronics [18].

Similarly, if user c chooses to offload the tasks to SEC or the terrestrial cloud, the computational latency can be obtained by

$$t_c^{SEC} = \frac{\delta m_c}{f_a}, \tag{3}$$

$$t_c^{Cloud} = \frac{\delta m_c}{f_{Cloud}}, \tag{4}$$

where f_a and f_{Cloud} are the CPU-cycle frequency of the LEO satellite a being offloaded to and terrestrial cloud, respectively. Similar to [11] and [19], we assume that all LEO satellites have similar computing capabilities.

C. Communication models

There are two options for each user to access LEO satellites, i.e., directly access the LEO satellite or via a semantic encoder deployed on the TST. The total bandwidth of the C-band user-TST link is divided into D_0 orthogonal sub-carriers based on OFDMA manner [10]. The transmission rate of the user c to the TST b on a sub-carrier d_0 in this link is

$$r_{c,d}^{cb} = B_{d_0}^{cb} \log_2(1 + \frac{p_{c,d_0}^{cb} g_{c,d_0}^{cb}}{\sigma_0^2}), \tag{5}$$

where $B_{d_0}^{cb}$, p_{c,d_0}^{cb} and g_{c,d_0}^{cb} are bandwidth, transmission power and the channel gain on sub-carrier d_0 in the user-TST link, separately. Further, in (5), σ_0^2 is the noise power in this link. Hence, the transmission delay from user c to TST b is

$$t_c^{cb} = \frac{m_c}{\sum_{d_0=1}^{D_0} x_{d_0}^{cb} r_{c,d_0}^{cb}}, \tag{6}$$

where $x_{d_0}^{cb} \in 0, 1$ is the allocation indicator of user-TST over the C-band. In the case of a sub-carrier d_0 in C-band is allocated to user c to offload the tasks, $x_{d_0}^{cb} = 1$; otherwise, $x_{d_0}^{cb} = 0$. Therefore, the transmission energy is

$$E_c^{cb} = t_c^{cb} \sum_{d_0=1}^{D_0} x_{d_0}^{cb} p_{c,d_0}^{cb}. \tag{7}$$

If user c chooses to access satellite a directly, due to the ultra-long propagation distance, the propagation delay is not negligible and the round-trip propagation delay is

$$t_c^{proa} = \frac{2h}{c_l}, \tag{8}$$

where h is the distance between user c and satellite a, c_l is the speed of light. We assume the users in the same TST, this TST and terrestrial cloud have the same distance to the satellite a. Moreover, path loss should be considered when transmitting over long distances. We are not concentrating on the path loss in the user-TST link because they communicate in a small cell range and haven't got a significant impact on the transmission delay. The transmission rate from the user c to satellite a thus can be denoted by

$$R_c^{ca} = B_c^{ca} \log_2(1 + \frac{p_c^{ca} g_c^{ca}}{\sigma_0^2 PL_c^{ca}}), \tag{9}$$

where B_c^{ca}, p_c^{ca} and g_c^{ca} are bandwidth, transmission power, and channel gain from the user c to satellite a, respectively. Furthermore, PL_c^{ca} is the path loss. Note that the path loss affects the channel hence the channel gain. Nevertheless, to better demonstrate the advantages of SemCom, similar to [20], we present the path loss separately in the formula to facilitate subsequent analysis. Normally, the path loss PL for the satellite channels mainly consists of free-space path loss PL_f and atmospheric (rainfall) loss PL_r [20]. Hence, we assume the total path loss $PL = PL_f + PL_r$. We will specify these losses later. We then have the transmission delay and energy consumption when user c accesses the SEC a directly, which are given by

$$t_c^{ca} = \frac{m_c}{R_c^{ca}}, \tag{10}$$

$$E_c^{ca} = t_c^{ca} p_c^{ca}. \tag{11}$$

In contrast to users, the transmission process from TST b to satellite a integrates SemCom. It thus increases the computing delay while significantly decreasing the data required to be transmitted. The transmission rate of TST can be expressed as:

$$R_b^{ba} = B_b^{ba} \log_2(1 + \frac{p_b^{ba} g_b^{ba}}{\sigma_0^2 PL_b^{ba}}), \tag{12}$$

where B_b^{ba}, PL_b^{ba}, P_b^{ba} and g_b^{ba} are bandwidth, path loss, transmission power and the channel gain in TST b-satellite a link, respectively. In addition, since antennas of TSTs have good directivity, they can communicate with multiple LEO satellites via Ka-band and the corresponding interference can be ignored [10], [21], [22]. Therefore, the transmission delay of all users' tasks are transmitted from TST b to satellite a is

$$t_c^{ba} = \frac{\sum_{j=1}^{F} \psi m_j}{R_b^{ba}} + \frac{\sum_{j=1}^{F} m_j}{R_{SemCom}^{ba}}, \tag{13}$$

where F is the number of users allocated to offloading the task to satellite a and $F \in \mathcal{C}$. Furthermore, ψ is the compression ratio and the R_{SemCom}^{ba} is the rate of semantic extraction and semantic parsing, i.e., computing delay during data transmission.

Since the computation task calculation result is often much smaller than the offloaded data, it is reasonable to ignore the backhaul transmission delay (see also [23] and [24]. Moreover, estimating the number of subcarriers provided by satellite a to user c is difficult due to the large number of satellite service users. We assume that the satellite transmits user data to the ground cloud with a constant transmission rate R_c^a similar to [11]. The transmission delay between satellite and cloud t_a^{Cloud} thus equals m_c/R_c^a. Due to the mobility of satellites, the distance from the satellite to the terrestrial cloud is difficult to precisely inform users, we thus use h to estimate the distance between the satellite and the terrestrial cloud. The propagation delay where user c chooses to offload to the terrestrial cloud is

$$t_c^{proC} = 2t_c^{proa} = \frac{4h}{c_l}. \tag{14}$$

D. Path loss model

As mentioned in Section II-C, the path loss for the terrestrial-satellite channel is mainly free-space path loss PL_f and atmospheric (rainfall) loss PL_r. Free-space path loss is a basic power loss that increases depending on the communication distance. In dB, PL_f is [25]

$$PL_f(\text{dB}) = 92.44 + 20\log(h) + 20\log(f), \tag{15}$$

where h is the communication distance unit in km, and f is the operating frequency with the unit of GHz.

Atmospheric loss is a type of signal absorption and scattering due to meteorological causes, i.e., mainly related to rainfall. The rain attenuation is described by [26]

$$PL_r(dB) = \xi L_E, \tag{16}$$

where ξ is the frequency-dependent parameter unit in dB/km and L_E is the effective path length unit in km. We first introduce the calculation method of ξ as:

$$\xi = k(R_{0.001})^v, \tag{17}$$

where $R_{0.001}$ is the rainfall rate, unit in mm/h. Further, k and v are coefficients given as:

$$k = [k_H + k_V + (k_H - k_V)cos^2(\omega)cos(2\tau)]/2, \tag{18}$$

$$v = [k_H v_H + k_V v_V + (k_H v_H - k_V v_V)cos^2(\omega)cos(2\tau)]/2, \tag{19}$$

where $\tau = \pi/4$ for circular polarization and ω is the elevation angle between terrestrial transmitter and satellite. Moreover, k_H, k_V, v_H, and v_V are coefficients related to operating frequency f and can be found out the specific value from [27]

L_E, is therefore

$$L_E = L_R v_{0.001}, \tag{20}$$

where L_R is the distance parameter related to rainfall height and $v_{0.001}$ is the adjustment factor. We have

$$v_{0.001} = \frac{1}{1 + \sqrt{sin(\omega)}(\frac{31(1-e^{-(\frac{\omega}{1+\chi})})\sqrt{LR^{\xi}}}{f^2} - 0.45)}, \tag{21}$$

where χ equals 36-—latitude— in the case of latitude less than 36^o, or equals 0. In most scenarios

$$L_R = \frac{h_R - h_s}{sin(\omega)} \tag{22}$$

where h_R is the rain height relative to the mean sea level and h_s is the altitude of the terrestrial transmitter, all units in km.

E. Semantic coder training model

In general distributed learning frameworks based on FedAvg [28], the training process requires multiple distributed participants and a federated aggregation node. Participants train their ML models locally and upload them to the federated aggregation node at fixed communication rounds. The federated aggregation node aggregates all the models and then returns the aggregated model to the participants for further training. This enables participants to update the model without sharing private training data. The goal of FL is to collaboratively train a global coder model among multiple TSTs while keeping TSTs' local data private. We set the $X_b = \{x_{in}^b\}_{b=1}^{s_b}$ as the data set of the TST b, where x_{in}^b is the in-th input sample and s_b is the size of the data set. The objective of FedAvg can be denoted by

$$\min_{\Theta} \frac{1}{B} \sum_{b=1}^{B} L_b(\theta_b), \tag{23}$$

where θ_b is the coder model parameter of the TST b and $\Theta = \theta_1, \theta_2, ..., \theta_b$. Further, $L_b(\theta_b)$ is the loss function of the TST b trained by X_b. We utilise the mean squared error (MSE) loss as the loss function in this paper. We have

$$L_b(\theta_b) = \frac{1}{s_b} \sum_{in=1}^{s_b} L_{MSE}(\theta_b; x_{b,in}, \widehat{x_{b,in}}), \tag{24}$$

where $\widehat{x_{b,in}}$ is the fitting output and L_{MSE} is the MSE loss.

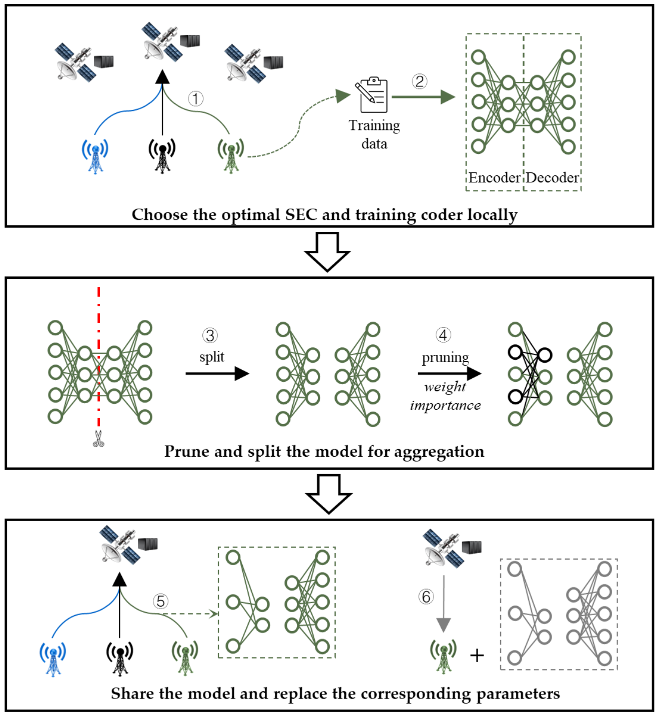


Fig. 2: The schematic of the proposed PSFed in one communication round. The workflow contains the following 6 steps: ① TSTs choose optimal SEC for federated aggregation jointly; ② local training on private data; ③ the TST's coder model is split into the encoder and decoder part; ④ the TSTs prune the encoder model according to parameter importance; ⑤ each TST uploads the model for federated aggregation; ⑥ the TSTs download the personalised models and replace the corresponding parameters.

III. Updating the Semantic Coders

Employing general FL frameworks for SemComs, TSTs need to upload encoder and decoder models to the SEC to implement federated aggregation after one communication round of training. Therefore, the federated model must be sent back to TSTs for the next communication round of training. However, uploading and downloading all coder models by TSTs would cause long-term interruptions of the offloading-assisted service, significant energy consumption, and lead to privacy leakage of entire coder models. Previous studies, e.g. [29] show that when reconstructing an ML model, increasing the number of parameters increases the accuracy of the model following a logarithmic function. In SemCom, the accuracy of the SemCom coder represents the accuracy of the received data. Therefore, the privacy of the coder model/parameter is closely tied to the accuracy. We can adopt a general parameter privacy leakage metric as in [30] and assess model parameter leakage by

$$\Theta_b(\theta_b) = \chi \log_2(1 + e^{1-\frac{N_b+1}{n_b}}), \tag{25}$$

where χ is the weight parameter, N_b is the total number of parameters at the encoder model and n_b is the number of parameters transmitted. In practice, Θ_b adopts a value in [0,1], where $\Theta_b = 0$ indicates that there is no privacy leakage, while a $\Theta_b = 1$ indicates fully compromised privacy where the same information can be decoded from the leaked model as the original model.

By increasing the number of training epochs the parameters of the training model become closer to the final trained model. Therefore, the model obtained from more training epochs is more important relative to the model obtained from previous training epochs before training is finished. In other words, the private information contained in the parameters is increased over time. More important parameters bear higher sensitivity in terms of privacy. Therefore, we rewrite the privacy leakage for TST b's encoder training as:

$$\Theta_b(\theta_b) = \sum_{r=1}^{R} W_r \chi \log_2(1 + e^{1-\frac{\sum_i^{N_b} I_i n_{b,i}+1}{\sum_i^{N_b} I_i n_{b,i}}}), \tag{26}$$

where r is the communication rounds and R is the total rounds. Also, W_r is the model importance weight of training round r. Similarly, I_i is a weight parameter denoting the importance transmitted parameter i.

In the proposed PSFed (Fig. 2), the goal is to collaboratively train semantic coder models among multiple TSTs while reducing network service interruptions, and energy consumption, and decreasing the degree of privacy leakage. Due to the high mobility of satellites, we note that all TSTs are not always within the same satellite service area. TSTs are therefore required to select the most appropriate satellite for each model aggregation round from the multiple satellites based on real-time circumstances. Taking into account TSTs' training delay and energy consumption jointly, the satellite selection algorithm is denoted by

$$\min_{x_a} \sum_{a=1}^{A} x_a (\alpha \max\{\frac{M_{b,r}}{R_b^{ba}} + \frac{2h^{ba}}{c_l} | b \in \mathcal{B}\} + \sum_{b=1}^{B} \beta p_b^{ba} \frac{M_{b,r}}{R_b^{ba}}), \tag{27a}$$

$$s.t. \quad \sum_{a=1}^{A} x_a = 1, \forall b \tag{27b}$$

$$x_a = \{0, 1\}, \tag{27c}$$

$$\sum_{r=1}^{R} \frac{M_{b,r}}{R_b^{ba}} \leq t_b^{'}, \forall b \tag{27d}$$

$$\max\{\frac{M_{b,r}}{R_b^{ba}} + \frac{2h^{ba}}{c_l} | b \in \mathcal{B}\} < t_a^{'}, \forall a \tag{27e}$$

where $\max\{\frac{M_{b,r}}{R_b^{ba}} + \frac{2h^{ba}}{c_l} | b \in \mathcal{B}\}$ is the training transmission and propagation delay, identified by the TST with the longest transmission and propagation time. Here, A is the number of accessible satellites of all TSTs, and h^{ba} is the distance between TST b and satellite a. Further, $\sum_{b=1}^{B} \beta p_b^{ba} \frac{M_{b,r}}{R_b^{ba}}$ is the total energy consumption of transmission from TSTs to a satellite. In (27a), α and β are weight parameters to balance the importance and unit of latency and energy consumption. Furthermore, p_b^{ba} is the transmission power of TST b to satellite a, and x_a is the federated decision for all TSTs. Constraint (27d) ensures that the transmission time of the TST for training the semantic model remains less than the maximum tolerable service interruption time. Also, $M_{b,r}$ is the coder model size in communication round r, $t_b^{'}$ is the maximum tolerable service interruption time and $t_a^{'}$ is the maximum service time of the satellite a in this region. The optimization problem in (27) is a simple 0,1 linear programming and hence can be easily solved.

During training in each communication round, we split the coder model into an encoder and a decoder. Only the decoder model needs entire federated aggregation. This is due to LEO satellites having limited storage capacity, it is not practical to use individual decoder models for each task of each TST. The semantic coders are therefore stored on the TST. For economic considerations, we argue that TSTs require a shared decoder model to be used. We then encourage TSTs to assess the importance of the encoder parameters during the local training phase. Inspired by continual learning [31], changes in parameters with different importance have a different impact on the output results. We thus evaluate parameter importance according to the implications of parameter changes on the loss function. We express the change in the loss by

$$L_b(\theta_b + \delta) - L_b(\theta_b) \approx \sum_{i=1}^{s_b} g_{b,i} \delta_{b,i}, \tag{28}$$

where g_i is the gradient and δ_i is the update of parameter i during this parameter assessment period of the TST b. Setting $g_i = \frac{\partial L_b}{\partial \theta_{b,i}}$ during online training, the parameter importance weight is

$$I_i = -\frac{\partial L_b}{\partial \theta_{b,i}} \delta_{b,i}. \tag{29}$$

Subsequently, to reduce the training communication cost, we prune the encoder models uploaded by TSTs according to parameter importance. Parameters with high importance contain most of the valid information [32] and therefore can provide further valid information to the aggregated model than lower-important parameters. The lower-importance parameters are thus encouraged to be pruned. The pruning here differs from the conventional ML studies. It is not the deletion of the training model parameters, but the non-transmission of the pruned parameters for federated aggregation. The corresponding SEC generates a

Algorithm 1 PSFed

Input: dataset $\{X_1, X_2, ..., X_b\}$, model size $\{M_1$,M_2,...,$M_b\}$ and total communication rounds R
Output: trained coder models $\{\theta_1, \theta_2, ..., \theta_b\}$
Initialize: the TSTs' model parameters and the importance weight of parameters **SECs:**

for each communication round $r \in R$:
  $Y_b^{r+1}, \theta_b^{r+1} \longleftarrow TST\ update(\theta_b^r)$
  Update $\{\theta_{b,1}, \theta_{b,2}, ..., \theta_{b,N_b}\}$ according to Y_b^{r+1} and θ_b^{r+1}
end for

TSTs:

TST b receives θ_b from the SEC
TSTs choose the optimal SEC for federated aggregation
for each TST in parallel:
  for each local training epoch:
    Loss $\longleftarrow \frac{1}{s_b} \sum_{in=1}^{s_b} L_{MSE}(\theta_b; x_{b,in}, \widehat{x_{b,in}})$
  end for
  foreach encoder parameter i:
    $I_i = -\frac{\partial L_b}{\partial \theta_{b,i}} \delta_{b,i}$
  end for
  Splitting coder model and pruning encoder model based on I_i in the case of satisfying:
  $\begin{cases} \sum_{r=1}^{R} \frac{M_{b,r}}{R_b^{ba}} \leq t_b^{'} \\ \Theta_b(\theta_b^r) \leq \Theta_b^{'} \end{cases}$
  Obtain θ_b^r to be shared
  return: θ_b^r
end for

global encoder model and a global decoder model based on the federated aggregation of the number of the received parameters. Once TST receives the global decoder model and personalised pruned global encoder model, it merely substitutes the local decoder and substitutes important parameters of the local encoder. It trains the individual local coder again based on the personal encoder model and the global decoder model in the next communication round of training.

Furthermore, the closer to the completion of the training, the higher the importance of the parameters. To further reduce the privacy leakage degree, our proposed PSFed progressively increases the pruning ratio according to the number of communication rounds. This is until the coder model is split and only the decoder model is federated aggregated. The more important privacy training models are thus kept local.

The objective of PSFed during training is to minimise the training loss, therefore,

$$\min_{\Theta, Y} \quad \sum_{b=1}^{B} L_b(y_b^1 \theta_{b,1}, y_b^2 \theta_{b,2}, ..., y_b^n \theta_{b,N_b}), \tag{30a}$$

$$s.t. \quad \sum_{r=1}^{R} \frac{M_{b,r}}{R_b^{ba}} \leq t_b^{'}, \forall b \tag{30b}$$

$$\max\{\frac{M_{b,r}}{R_b^{ba}} + \frac{2h^{ba}}{c_l} | b \in \mathcal{B}\} < t_a^{'}, \forall a \tag{30c}$$

$$\sum_{r=1}^{R} W_r \chi \log_2(1 + e^{1 - \frac{\sum_i^{N_b} I_i n_{b,i} + 1}{\sum_i^{N_b} I_i n_{b,i}}}) \leq \Theta_b^{'}, \forall b \tag{30d}$$

where $y_b^n \in [0, 1]$ is the aggregation weight vector of parameter i in TST b. It acts similar to the weighted average in FedAvg. Since each TST uploads a different number and location of parameters in the same model, the proportion of each parameter that is weighted is different. The y_b^n for various parameters also different and $Y = y_1, y_2, ..., y_b$. Further, $\Theta_b^{'}$ is privacy leakage consideration and $\Theta_b^{'}$ is the maximum tolerable leakage The procedure of the PSFed is demonstrated in Algorithm 1.

IV. The Semantic Coders in Service

In this section, the problem of users' computational task processing schedule for SemCom-SEC is presented first. We then detail the proposed CTPS.

A. Computational task processing

In service offloading decision-making, we consider the SemCom-SEC with C users severed by one TST b in A satellite coverage. Each user has five task processing choices, 1) local computing; 2) offloading the tasks to SEC directly; 3) offloading the tasks to SEC via the TST; 4) offloading the tasks to the terrestrial cloud only via the satellite; 5) offloading the tasks to the terrestrial cloud via the TST and the satellite. We firstly list the user c's cost functions in terms of processing delay and energy consumption for each option in order as follows based on Section II:

$$\Phi_{c1} = \alpha t_c^{LC} + \beta E_c^{LC}, \tag{31}$$

$$\Phi_{c2} = \alpha(t_c^{proa} + t_c^{ca} + t_c^{SEC}) + \beta E_c^{ca}, \tag{32}$$

$$\Phi_{c3} = \alpha(t_c^{proa} + t_c^{cb} + t_c^{ba} + t_c^{SEC}) + \beta E_c^{cb}, \tag{33}$$

$$\Phi_{c4} = \alpha(t_c^{proa} + t_c^{ca} + t_c^{Cloud} + t_a^{Cloud}) + \beta E_c^{ca}, \tag{34}$$

$$\Phi_{c5} = \alpha(t_c^{proC} + t_c^{cb} + t_c^{ba} + t_c^{Cloud} + t_a^{Cloud}) + \beta E_c^{cb}, \tag{35}$$

where Φ_c is the actual processing cost when the user c sizing a task. It is related to user task processing decisions, the transmission power, and the number of subcarriers allocated. In the above, t_a^{Cloud} is the transmission delay between satellite and cloud as mentioned in Section II-C. We also utilise $\gamma_c = \{0, 1\}$ to represent the offloading decision of user c and $\gamma_c \in \{\gamma_{1c}, \gamma_{2c}, \gamma_{3c}, \gamma_{4c}\}$. If user c chooses one processing strategy, the indicator for the corresponding strategy equals 1, otherwise equals 0. We argue that the optimal decision for a user is to minimise the latency and energy consumption of the processing tasks. Mathematically, the optimisation task processing strategy problem of user c thus can be formulated as a MINLP problem:

$$\min_{\gamma_c, f_c, p_{c,d_0}^{cb}, m_{c,d_0}, p_c^{ca}} \quad \sum_{a=1}^{A} \Phi_c = (1 - \gamma_{1c} - \gamma_{2c} - \gamma_{3c} - \gamma_{4c})\Phi_{c1} + \gamma_{1c}\Phi_{c2} + \gamma_{2c}\Phi_{c3} + \gamma_{3c}\Phi_{c4} + \gamma_{4c}\Phi_{c5}, \tag{36a}$$

$$s.t. \quad f_{cloud} \geq f_a \geq f_{c,max} \geq 0, \tag{36b}$$

$$\gamma_{1c}, \gamma_{2c}, \gamma_{3c}, \gamma_{4c} \in \{0, 1\}, \tag{36c}$$

$$\gamma_{1c} + \gamma_{2c} + \gamma_{3c} + \gamma_{4c} \leq 1, \tag{36d}$$

$$\sum_{d_0=1}^{D_0} x_{d_0}^{cb} p_{c,d_0}^{cb} \leq P_{c,max}, \tag{36e}$$

$$P_c^{ca} \leq P_{c,max}, \tag{36f}$$

$$x_{d_0}^{cb} \in \{0, 1\}, \tag{36g}$$

$$\sum_{d_0=1}^{D_0} x_{d_0}^{cb} \leq D_0, \tag{36h}$$

$$t^* < t_a^{'}. \tag{36i}$$

The constraint (36b) guarantees that edge and cloud have strong computing capability that is not less than users' maximum computing capability $f_{c,max}$. Constraints (36c) and (36d) show the relationship between $\gamma_{1c}, \gamma_{2c}, \gamma_{3c}$ and γ_{4c}. In constraints (36e) and (36f), $P_{c,max}$ is the maximum available transmission power of user c to TSTs or satellites. The constraint (36g) denotes the subcarrier allocation indicator. The constraint (36h) means that the number of allocated subcarriers should not exceed the total number of sub-carriers. The constraint (36i) is to ensure the optimal decision's transmission time t^* is less than the time $t_a^{'}$ available to access satellite a.

The problem in (36) is an MINLP problem with incomplete information due to privacy concerns. This is because users need the allocation of subcarriers to make decisions. Nevertheless, such information is relevant to decisions and privacy information (e.g., local computing capability and transmission power) from other users. This MINLP problem thus is computationally complex and hard to solve.

B. CTPS

In this paper, we propose a CTPS mechanism (see, Fig. 3) to minimise the delay and energy consumption of users to process computational tasks, while privacy-preserving and equitable. We assume all the participants are trustworthy It is divided into two steps. Firstly, it converts the optimisation task processing strategy problem with privacy considerations into a complete information problem based on the Rubinstein bargaining model [33] equitably. Subsequently, users develop the optimisation task processing strategies by solving the complete information MINLP problem of Eq. (36). We detail our CTPS mechanism as follows.

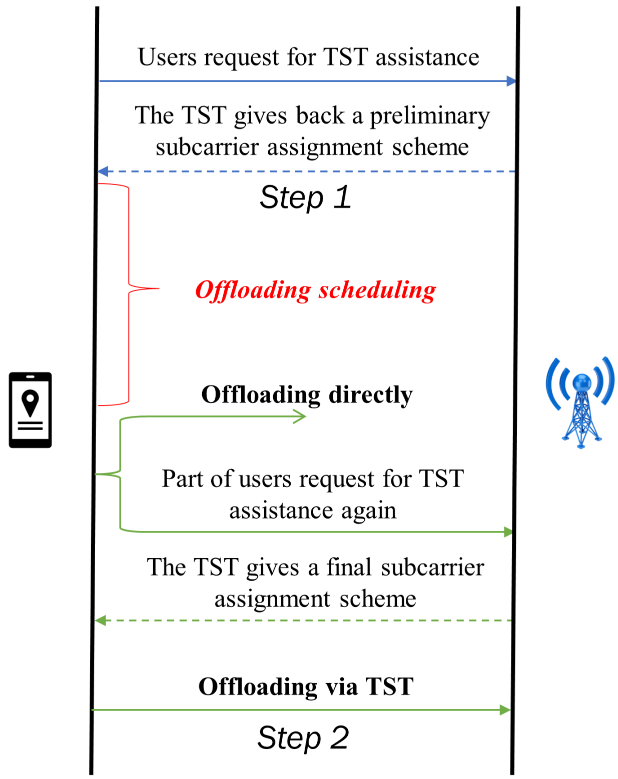


Fig. 3: Proposed CTPS mechanism.

C. First step of the CTPS mechanism

We enable users to communicate/bargain with TST several times so that subcarriers are allocated fairly without privacy leakage based on the Rubinstein bargaining game. TST acts as the bidder and the user has the option to continue the game or leave the game. The gaming process is limited to two periods. In the first period, the users send the offloading request to the TST. Upon receiving users' offloading requests, without loss of generality and fairness, TST allocates the number of C-band sub-carriers based on the size of the tasks offloaded by users. Further, the transmission delay of the TST to the satellite and semantic extraction delay are also notified via this communication.

To achieve the game-perfect equilibrium, the cost function for user c to assess to continue participating in the game can be denoted by

$$\mu_c^{'} = \epsilon\iota\Phi_c^{'},\ \ \Phi_c^{'} = \{\Phi_{c3}, \Phi_{c5}\}, \tag{37}$$

where $\iota \in (0, 1)$ is the bargaining discount factor that represents the revenue loss value for the second-period communication due to the bargaining process being time and energy-consuming. Further, $\epsilon \geq 1$ is the weight parameter to evaluate the further possible benefit by applying offloading again via the TST b, i.e., remaining engaged in the game. This is attributable to some users abandoning their requests for TST offloading due to not being allocated a satisfactory number of C-band subcarriers. The actual number of subscribers should eventually be greater than or equal to this allocation. Simultaneously, the strategies of various users also affect the user-satellite link interference for different users. In order to estimate the influence of interference, pricing is a frequently utilised method in the game theory employed studies [34]. We hence rewrite the part of the cost function for user c considering interference pricing as:

$$\mu_c^{''} = \Phi_c^{''} + \alpha\varrho m_c\varpi,\ \ \Phi_c^{''} = \{\Phi_{c2}, \Phi_{c4}\}, \tag{38}$$

where ϱ is the factor for the interference related to the number of users, transmission power, and channel gain. Further, $\varpi \in [0, 1]$ is the proportion to denote the anticipation rate of not performing local computing users, thus predicting the fraction of time in which interference is received.

Finally, the incomplete information MINLP problem is converted to a complete information MINLP problem. Users thus could develop the optimal processing decision based on allocated subcarriers and the calculation frequency or transmitting power in the second step.

D. Second step of the CTPS mechanism

In the second step, users make the decision based on the complete information MINLP problem of Eq. (36) to minimise the latency and energy consumption of the processing tasks. The maximum number of satellites expected to be accessible at the same time is extremely limited [22]. The decision problem Eq. (36) can be considered as $5 \cdot A$ independent subproblems, where 5 is five offloading decision subproblems and A is A satellite selection subproblems. In case of the local computing, the best user c's CPU-cycle frequency f_c is only related to local computing costs. We thus can express the f_c optimisation subproblem as:

$$\min_{f_c} \quad \Phi_{c1} = \alpha \frac{\delta m_c}{f_c} + \beta \varepsilon \delta m_c f_c^2, \tag{39a}$$

$$s.t. \quad (36b). \tag{39b}$$

We can express the first-order derivative of (39a) as: $-\alpha \frac{\delta m_c}{f_c^2} + 2\beta\varepsilon\delta m_c f_c$. Eq. (39a) monotonically increases in the constraint (39b), hence $f_c = f_{c,max}$.

In addition, in case the user needs to employ TSTs, the user needs to derive the optimal subcarrier task allocation strategy m_{c,d_0} and subcarrier transmission power p^{cb}_{c,d_0}. To model and optimise the transmission power, in CTPS, we assume each subcarrier in the same link accomplishes the transmission tasks at the same time for fully using spectrum resources in a synchronous manner based on previous studies [23], [35]. As the allocated subcarrier for user c is known, we set η to denote the number of allocated subcarriers. We can simplify the optimisation problem associated with TST as:

$$\min_{m_{c,d_0}, p^{cb}_{c,d_0}} \quad \sum_{d_0=1}^{D_0} \left(\frac{\alpha x^{cb}_{d_0} m_{c,d_0}}{\eta r^{cb}_{c,d_0}} + \frac{\beta p^{cb}_{c,d_0} x^{cb}_{d_0} m_{c,d_0}}{r^{cb}_{c,d_0}} \right), \tag{40a}$$

$$s.t. \quad (36e), (36g), (36h), \tag{40b}$$

$$\sum_{d_0=1}^{D_0} x^{cb}_{d_0} m_{c,d_0} = m_c. \tag{40c}$$

We only need to consider the situation that $x^{cb}_{d_0} = 1$. By relaxing constraints, we have the Lagrangian function for Eq. (40a) as:

$$\begin{aligned} L = & \sum_{d_0=1}^{D_0} x^{cb}_{d_0} \left(\frac{\alpha m_{c,d_0}}{\eta r^{cb}_{c,d_0}} + \frac{\beta p^{cb}_{c,d_0} m_{c,d_0}}{r^{cb}_{c,d_0}} \right) \\ & + \varphi \left(\sum_{d_0=1}^{D_0} x^{cb}_{d_0} p^{cb}_{c,d_0} - P_{c,max} \right) + \lambda \left(m_c - \sum_{d_0=1}^{D_0} x^{cb}_{d_0} m_{c,d_0} \right), \end{aligned} \tag{41}$$

where φ and λ are the Lagrangian multipliers. The dual problem thus is $\min_{m_{c,d_0}, p^{cb}_{c,d_0}} L$. Then, we can observe that Eq. (41) can be further decomposed into D_0 independent subproblems, and the actual objective function in each d_0 subproblem can be denoted by

$$\min_{m_{c,d_0}, p^{cb}_{c,d_0}} \quad L_{d_0} = \frac{\alpha m_{c,d_0}}{\eta r^{cb}_{c,d_0}} + \frac{\beta p^{cb}_{c,d_0} m_{c,d_0}}{r^{cb}_{c,d_0}} + \varphi p^{cb}_{c,d_0} + \lambda m_{c,d_0}. \tag{42}$$

For simplicity, we define

$$H_{d_0} = \frac{\alpha}{\eta r^{cb}_{c,d_0}} + \frac{\beta p^{cb}_{c,d_0}}{r^{cb}_{c,d_0}}. \tag{43}$$

According to Karush-Kuhn-Tucker conditions, taking the partial derivatives of L_{d_0} with respect to p^{cb}_{c,d_0} and m_{c,d_0}, respectively. We have

$$\begin{cases} \dfrac{\partial L_{d_0}}{\partial p^{cb}_{c,d_0}} = m_{c,d_0} \dfrac{\partial H_{d_0}}{\partial p^{cb}_{c,d_0}} + \varphi = 0 & \text{(44a)} \\ \dfrac{\partial L_{d_0}}{\partial m_{c,d_0}} = H_{d_0} - \lambda = 0 & \text{(44b)} \\ \varphi(\displaystyle\sum_{d_0=1}^{D_0} x^{cb}_{d_0} p^{cb}_{c,d_0} - P_{c,max}) = 0. & \text{(44c)} \end{cases}$$

Thus, we have

$$\begin{cases} \varphi = 0, \displaystyle\sum_{d_0=1}^{D_0} x^{cb}_{d_0} p^{cb}_{c,d_0} \leq P_{c,max}, & \text{(45a)} \\ \varphi > 0, \displaystyle\sum_{d_0=1}^{D_0} x^{cb}_{d_0} p^{cb}_{c,d_0} = P_{c,max}, & \text{(45b)} \end{cases}$$

where (45) is complementary slackness. For (45a), p^{cb}_{c,d_0} can be directly solved by (44) causing $m_{c,d_0} \neq 0$. After deriving the optimal p^{cb}_{c,d_0}, m_{c,d_0} can be easily solved as all subcarriers have the same subcarrier completion time. Only if the solution $\sum_{d_0=1}^{D_0} p^{cb}_{c,d_0} = P_{c,max}$, we need to consider Eq. (45b). In that case, the Lagrangian multipliers can be obtained by the sub-gradient method and further achieve the optimal p^{cb}_{c,d_0}, m_{c,d_0}. Moreover, as we utilise the Lagrangian dual decomposition method, the solution may have a duality gap. However, this gap should approach zero and can be ignored in practical systems as the number of subcarriers D_0 is large enough [10].

Therefore, users can make the decision based on the computation cost of various alternatives, without compromising privacy. Throughout the CTPS, the user is only communicated externally about the size of the tasks being processed. It also needs to be known by TST during the offloading process. Hence the CTPS protect the privacy of computing power, transmit power, etc. Further, the computational complexity is linearly related to D_0 and A, whereas both D_0 and A are finite. CTPS thus can be used in large-scale satellite networks. The CTPS and offloading decision process is summarised as Algorithm 2.

V. Simulation Results

A. Simulation setting

In this section, we evaluate the performance of the present PSFed and CTPS. In the simulations, if not specifically mentioned, we set the parameters as follows. The LEO satellites' coverage radius is 280 km and the vertical altitude is 780km based on the Iridium satellite system [36]. The frequencies of the C-band and the Ka-band are 4.5 GHz and 30 GHz separately based on 3GPP specifications [37]. We assume the number of C-band subcarriers is 128, the maximum transmission power of users is 23 dBm and the transmit power of each TST is 30 dBm [10]. The offloading task is assumed an image recognition task and the semantic coder is considered an autoencoder based on the convolutional autoencoder (CAE) similar to [38].

Communication rounds for the proposed PSFed to aggregate the semantic encoder are 20 rounds. The coder settings are listed in Table I. Furthermore, we set the number of CPU cycles for computing one bit δ as 120 cycles/bit, which is from the real applications [18]. We assume all users have the same CPU frequency f_c, and set it as 0.5×10^9 cycles/s. The computation capabilities of SEC on satellite a and the cloud server are 3×10^9 cycles/s and 10×10^9 cycles/s, respectively [11]. The energy factor ε is set as 10^{-26} [10].

Moreover, we assume weight parameters of latency and energy consumption are set as $\alpha = 0.5$ and $\beta = 0.5$, and weight parameters in bargain process ι and ϵ are all considered as 1. In addition, the atmospheric loss is adopted, and the related coefficients are shown in Table II [27]. The simulation parameters are also listed in Table III.

B. Performance evaluation of PSFed

Fig. 4 illustrates the convergence speed of the different frameworks under different transmission tasks. The TSTs' images are from CIFAR 10 [39], CIFAR 100 [40] and MNIST [41] image datasets and TSTs perform federated aggregation after every five local epochs. Based on the feasibility in SEC networks, we compare the proposed PSFed with the generalised learning approach for SemCom [15], [16], i.e., FL frameworks based on the FedAvg [28].

Algorithm 2 CTPS

Input: Tasks m_c generation

Output: The computation offloading and resource allocation result $\gamma_c, f_C, p^{cb}_{c,d_0}, m_{c,d_0}, x^{cb}_{d_0}$

1: Initialize the optimal TST transmission power p^{ba}_b
2: Obtain necessary information $x^{cb}_{d_0}$ after first period game
3: Obtain the necessary information $x^{cb}_{d_0}$ after first period game
4: Calculate optimally f_c
5: Relax Eq. (40)
6: **if** $\varphi = 0$:
7: $\quad p^{cb}_{c,d_0} \longleftarrow \frac{\partial H_{d_0}}{\partial p^{cb}_{c,d_0}}$
8: $\quad m_{c,d_0} \longleftarrow \frac{m_c p^{cb}_{c,d_0}}{\sum_{d_0=1}^{D_0} x^{cb}_{d_0} p^{cb}_{c,d_0}}$
9: **else**:
10: $\quad p^{cb}_{c,d_0} \longleftarrow$ Eq. (44)
11: $\quad m_{c,d_0} \longleftarrow \frac{m_c p^{cb}_{c,d_0}}{P_{c,max}}$
12: **end if**
13: Find the maximum Φ_c and derive γ_c
14: **if** $\gamma_{c3} + \gamma_{c5} = 1$:
15: $\quad$ Obtain the necessary information $x^{cb}_{d_0}$ after the second period game
16: $\quad$ Obtain updated p^{cb}_{c,d_0} and m^{cb}_{c,d_0}
17: **end if**
18: Find the maximum Φ_c and derive γ_c

TABLE I: The setting of the CAE

Encoder	Neuron num	Decoder	Neuron num
Conv+ReLU	512	transConv+ReLU	10
Conv+ReLU	256	transConv+ReLU	32
Conv+ReLU	128	transConv+ReLU	64
Conv+ReLU	64	transConv+ReLU	128
Conv+ReLU	32	transConv+ReLU	256
Conv+Sigmod	10	transConv+Sigmod	512

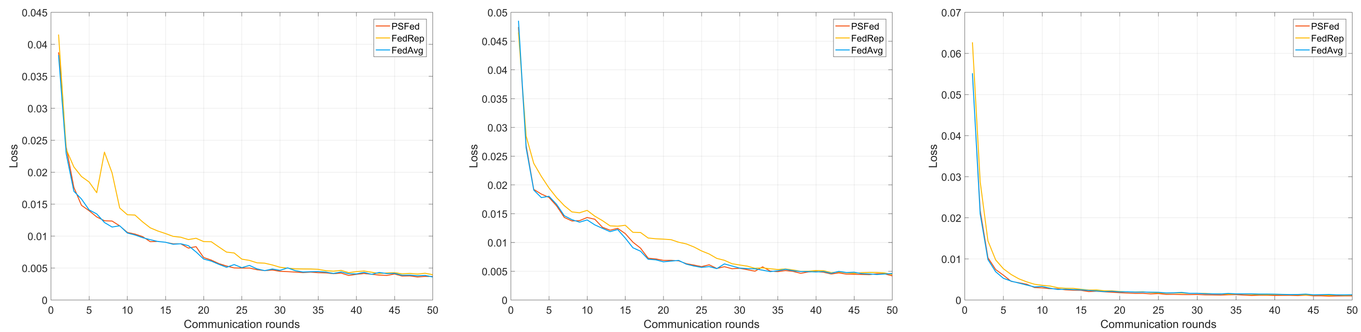
TABLE II: Rainfall coefficients

C-band	Value	Ka-band	Value
k_H	0.0001340	k_H	0.2403
k_V	0.0002347	k_V	0.2291
v_H	1.6948	v_H	0.9485
v_V	1.3987	v_V	0.9129

TABLE III: Simulation parameters

Parameters	Default values
The coverage radius of LEO satellites	280 km
Ka-band carrier frequency	30 GHZ
C-band carrier frequency	4.5GHZ
Number of C-band subcarriers	128
The maximum transmit power of each user	23dBm
Transmit power of TST	30 dBm
h	780km
δ	120
ε	10^{-26}
f_c	0.5×10^9 cycles/s
f_a	3×10^9 cycles/s
f_{Cloud}	10×10^9 cycles/s
α, β	0.5
ι, ϵ	1

Based on the existing FL methods that are potentially for SEC SemCom, FedRep [42] is also compared to demonstrate the effectiveness of our PSFed. The FedRep is based on the Fedavg but only aggregates part of the training model during each communication round. We set it to only aggregate SemCom decoder to adapt the SemCom-SEC. Moreover, we set the training sample to 5000 images per TST to reflect the differences between the frameworks more effectively. It can be observed that our PSFed achieves similar convergence rates to the FedAvg and is much better than the FedRep, regardless of the dataset. This is because our method aggregates important weights in the early stages of training and therefore accelerates convergence similarly to the FedAvg with all parameters aggregated.



(a) CIFAR 10 dataset (b) CIFAR 100 dataset (c) MNIST dataset

Fig. 4: Convergence speed of various learning algorithms with different datasets.

In Fig. 5, we compare the total communication cost of PSFed, FedRep and FedAvg during training. We assume that each neuron transmitted consumes the same amount of communication resources. The communication cost is therefore defined as the number of neurons transmitted during communication. It is seen that the PSFed expenses are approximately the same communication cost as the FedAvg in the early stages of training. The growth then gradually slows down and increases at the same magnitude as the FedAvg after round 20. This is because the PSFed gradually decreases the number of weights aggregated by the encoder model.

It is also seen that in round 20, the number of aggregated weights for the encoder model is 0, the same as the FedRep, only the decoder model is aggregated. Therefore, the PSFed only consumes additional communication resources for the importance weight aggregation than the FedRep. Considering that the FedRep converges much more slowly than the proposed PSFed, the total communication resource consumption can be considered to be similar. However, in comparison to the FedAvg, the communication consumption of our PSFed decreases by 40.50% in round 50.

C. Performance evaluation CTPS

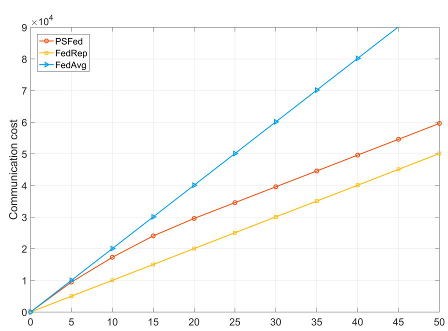


Fig. 5: Communication cost of various learning approaches.

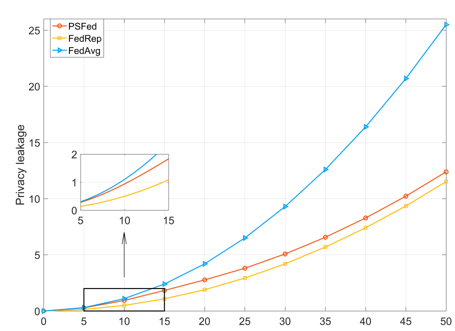


Fig. 6: Privacy leakage of various learning approaches.

We evaluate the total model privacy leakage during training in Fig. 6 according to Eq. (26). We assume that the model in each communication round has the same importance and that each neuron is of equal importance. It can be observed that PSFed is initially similar to FedAvg leakage and subsequently follows the same growth trend as FedRep. This is equally due to the number of PSFed decreasing importance weight aggregations. After training, both the PSFed and the FedRep encoder models are saved locally. It is foreseeable that if the importance of each round of communication changes, the PSFed would be extremely close to the FedRep in terms of total privacy leakage. In addition, the privacy leakage of PSFed should widen the gap with FedAvg, even though the privacy leakage of our PSFed already decreases by 51.43% in round 50 in comparison to FedAvg in the same importance.

In Fig. 7, the accuracy of the different frameworks under different transmission tasks is shown. We evaluate the accuracy utilising Peak Signal-to-Noise Ratio (PSNR), a general metric for evaluating image transmission in SemCom [38]. We have

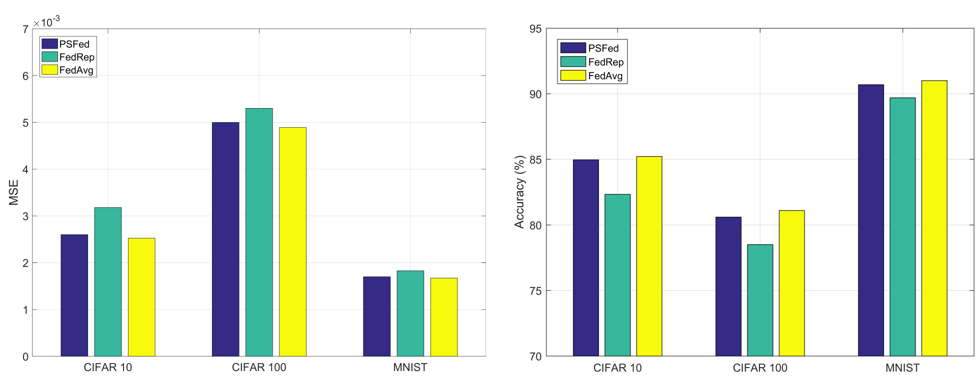
$$PSNR = 10 \log \frac{MAX^2}{MSE} (dB), \tag{46}$$

where MAX is the maximum value for a pixel and MSE is the mean squared deviation. Since different datasets have different MAX, we assume that the learning method with the smaller MSE has a higher accuracy. It is seen that the FedRep is significantly the least accurate with different datasets trained. The accuracy of PSFed is similar to FedAvg but slightly FedAvg higher. Because encoder models of both PSFed and FedRep are kept at the TST that are not aggregated when training is completed. Some aggregation information thus is lacking. However, the average training accuracy of the PSfed decreased by only 0.33% relative to the FedAvg due to the important weight aggregation acting as pre-training. Compared to the FedAvg, the accuracy loss of the PSfed deems acceptable given the significant communication cost and privacy concerns of the former.

Fig 7(b) further demonstrates the effect of image transmission accuracy on offloading via different approaches. We employed commonly used ML models for image recognition to identify the accuracy of images before/after transmission. The accuracy here is the proportion of the received object/image recognition accuracy to the pre-transmission image recognition accuracy. It can be seen that with the same trend as Fig. 7(b) FedRep has the significantly lowest accuracy while our method is similar to FedAvg but slightly lower. Figs. 5, 6, and 7 collectively suggest that PSFed achieves the fastest convergence rate, the lowest communication cost, and a high accuracy rate.

Fig. 8 illustrates the impact of users in one TST coverage on the total cost. As users are not always able to offload tasks via the TST, the proposed CTPS is compared with the local computing, offloading to the SEC directly, offloading to the cloud directly and CTPS without the game. The task size for each user is randomly generated over a range of 5 kb-300kb and subjected to 200 times replications of the simulation. Fig.8 shows that the total cost grows with the number of users. This is because raising the number of users increases the corresponding number of computing tasks and thus the total cost of users. The total cost of the proposed CTPS always keeps the total cost to the minimum and the advantage increases as the number of users increases. In addition, in cases where the number of users is small, the proposed CTPS thus maintains almost the same processing cost as ”CTPS without game”. By increasing the number of users, TST becomes unable to satisfy all the requests and CTPS starts to show its advantage in reducing the cost. We expect this advantage to increase by further increasing the number of users. This is because the optimal reallocation of resources through our design game scheme increases the efficiency of network resource utilisation.

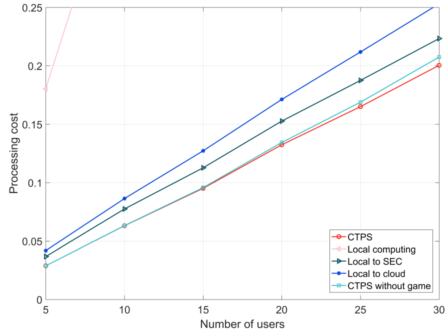
In Fig. 9, we show the offloading and computing cost of a single user versus the size of generating tasks. It is observed that the cost increases with the data size for all schemes. Our proposed mechanism always has a lower cost compared to the other three approaches. In case the data size is small (10 kb), our CPTS choose local computing as the optimal option. As the data size grows, the local computing latency and energy consumption increase, and CTPS chooses other minimum cost strategies, i.e., offload tasks to the SEC via the TST. After 250kb, the optimal value of our mechanism fluctuates. This is due to the data size being large enough, and the best strategy changes to offload tasks to the cloud via a TST. Therefore, the processing of the single-user tasks can be performed efficiently via our proposed processing strategy.



(a) MSE (b) Recognise accuracy matches after coding

Fig. 7: Accuracy of various learning algorithms with different datasets.

Fig. 10 demonstrates the importance of integrating SemCom into SEC networks in future communication environments. We set the user and the TST to maintain the same status to transmit to LEO satellites in different rainfall environments. It can be observed that as the rainfall probability increases, the task transmission cost of TST without SemCom exhibits a significant increase. Because the Ka-band frequency is extremely high and is strongly influenced by rainfall-induced path loss. In contrast,



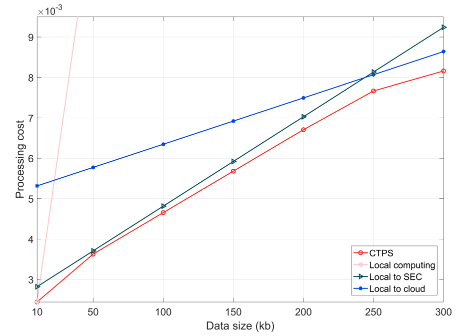


Fig. 8: Communication cost of various learning approaches.

Fig. 9: Privacy leakage of various learning approaches.

the processing costs for users transmitting via C-band are only slightly increasing. Since the C-band frequency is smaller than the Ka-band frequency and thus tolerates less path loss. Nevertheless, the TST configuration with the semantic encoder spends the least processing cost. Furthermore, the processing cost did not increase significantly with the increase in rainfall rate. This is because the latency of semantic extraction is not affected by the environment. The improved spectrum efficiency also reduces the impact of rainfall-induced path loss. Therefore, the integration of SemCom in SEC networks is necessary.

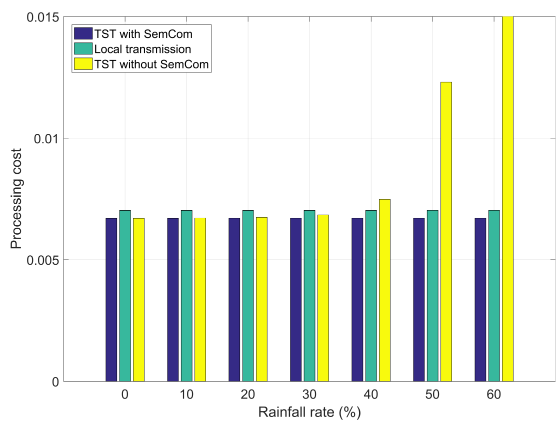


Fig. 10: The usefulness of SemCom in the network.

In Fig. 11, the influence of α and β on user strategies are investigated and the data size is from 5kb to 300kb simulated 50 times. The energy consumption weight β is always set as 0.5. We list the proportion of users that do not choose to offload via TST. It can be noticed that as the number of users increases, the unwillingness to offload increases due to the reduced number of subcarriers being allocated to them. However, users are always more reluctant to offload via TST in case the delay is more important (i.e., bigger α). These provide a criterion for the appropriate α and β to be chosen.

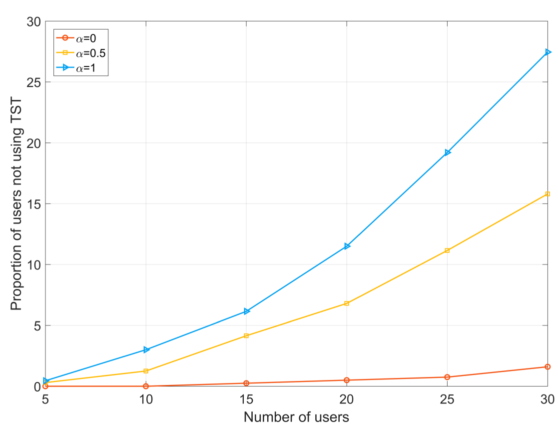


Fig. 11: Impact of α and β on strategy developing.

VI. Conclusion

In this paper, we investigated the integration of SemCom and SEC networks for terrestrial resource-limited users' computation offloading. We further proposed a novel SemCom-SEC framework for computation offloading. In addition, we examined the challenges that SemCom confronts in the proposed framework. For analysis, we then considered the challenges in two different scenarios. For the in-maintenance SemCom service, we proposed PSFed for the semantic coder update challenge. In the in-service SemCom service, we presented a game theoretical CTPS mechanism for task processing decision challenges of users. Compared with the general learning approach for semantic coder updating in SEC networks, simulation studies indicate that, on average, the proposed PSFed saves 40.50% of communication resources and further reduces privacy risk by 51.43%. Nevertheless, the training accuracy and convergence speed of PSFed and the general learning approach almost remain the same.

References

[1] G. Zheng, Q. Ni, K. Navaie, H. Pervaiz and C. Zarakovitis, "Efficient Pruning-Split LSTM Machine Learning Algorithm for Terrestrial-Satellite Edge Network," *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*, Seoul, Korea, Republic of, 2022, pp. 307-311.

[2] P. Rahimi, C. Chrysostomou, H. Pervaiz, V. Vassiliou and Q. Ni, "Joint Radio Resource Allocation and Beamforming Optimization for Industrial Internet of Things in Software-Defined Networking-Based Virtual Fog-Radio Access Network 5G-and-Beyond Wireless Environments," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 4198-4209, June 2022.

[3] Y. Xiao, G. Shi, Y. Li, W. Saad and H. V. Poor, "Toward Self-Learning Edge Intelligence in 6G," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 34-40, December 2020.

[4] Z. Zhang, W. Zhang and F. Tseng, "Satellite Mobile Edge Computing: Improving QoS of High-Speed Satellite-Terrestrial Networks Using Edge Computing Techniques," *IEEE Network*, vol. 33, no. 1, pp. 70-76, Jan/Feb 2019.

[5] E. C. Strinati, S. Barbarossa, T. Choi, A. Pietrabissa, A. Giuseppi, E. De Santis, J. Vidal, Z. Becvar, T. Haustein, N. Cassiau, F. Costanzo, J. Kim, and I. Kim, "6G in the sky: On-demand intelligence at the edge of 3D networks (Invited paper)," *ETRI J.*, vol. 42, no. 5, pp. 643–657, Oct. 2020.

[6] Y. Wang, J. Yang, X. Guo and Z. Qu, "A game-theoretic approach to computation offloading in satellite edge computing", *IEEE Access*, vol. 8, pp. 12510-12520, 2019.

[7] Z. Zhang, W. Zhang and F.-H. Tseng, "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques", *IEEE Netw.*, vol. 33, no. 1, pp. 70-76, Jan./Feb. 2019..

[8] B. Di, L. Song, Y. Li and H. V. Poor, "Ultra-Dense LEO: Integration of Satellite Access Networks into 5G and Beyond," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 62-69, April 2019.

[9] Y. Wang, J. Zhang, X. Zhang, P. Wang and L. Liu, "A computation offloading strategy in satellite terrestrial networks with double edge computing", *Proc. IEEE Conf. Commun. Syst.*, pp. 450-455, Dec. 2018.

[10] Z. Song, Y. Hao, Y. Liu and X. Sun, "Energy-Efficient Multiaccess Edge Computing for Terrestrial-Satellite Internet of Things," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14202-14218, 15 Sept.15, 2021.

[11] Q. Tang, Z. Fei, B. Li and Z. Han, "Computation Offloading in LEO Satellite Networks With Hybrid Cloud and Edge Computing," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9164-9176, 1 June1, 2021.

[12] Q. Lan et al., "What is Semantic Communication? A View on Conveying Meaning in the Era of Machine Intelligence," *Journal of Communications and Information Networks*, vol. 6, no. 4, pp. 336-371, Dec. 2021.

[13] Z. Qin, X. Tao, J. Lu, and G. Y. Li, "Semantic communications: Principles and challenges," Jun. 2022, [online] Available: http:// arXiv:2201.01389.

[14] H. Xie and Z. Qin, "A Lite Distributed Semantic Communication System for Internet of Things," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 142-153, Jan. 2021

[15] G. Shi, Y. Xiao, Y. Li and X. Xie, "From Semantic Communication to Semantic-Aware Networking: Model, Architecture, and Open Problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44-50, August 2021.

[16] Z. Qin, G. Y. Li and H. Ye, "Federated Learning and Wireless Communications," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 134-140, October 2021.

[17] L. D. Earley, "Communication in Challenging Environments: Application of LEO/MEO Satellite Constellation to Emerging Aviation Networks," *2021 Integrated Communications Navigation and Surveillance Conference (ICNS)*, 2021, pp. 1-8.

[18] A. P. Miettinen and J. K. Nurminen, "Energy efficiency of mobile clients in cloud computing", *Proc. USENIX HotCloud*, pp. 4-11, Jun. 2010.

[19] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang and X. S. Shen, "Software Defined Space-Air-Ground Integrated Vehicular Networks: Challenges and Solutions," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 101-109, July 2017.

[20] K. Tekbıyık, G. K. Kurt and H. Yanikomeroglu, "Energy-Efficient RIS-Assisted Satellites for IoT Networks," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14891-14899, 15 Aug.15, 2022.

[21] J. Du, C. Jiang, H. Zhang, Y. Ren and M. Guizani, "Auction design and analysis for SDN-based traffic offloading in hybrid satellite-terrestrial networks", *IEEE J. Sel. Areas Commun.*, vol. 36, no. 10, pp. 2202-2217, Oct. 2018.

[22] R. Deng, B. Di, S. Chen, S. Sun and L. Song, "Ultra-dense LEO satellite offloading for terrestrial networks: How much to pay the satellite operator?", *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6240-6254, Oct. 2020.

[23] F. Wang, J. Xu, and Z. Ding, "Multi-antenna NOMA for computation offloading in multiuser mobile edge computing systems," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2450–2463, Mar. 2019.

[24] Y. Wu, K. Ni, C. Zhang, L. P. Qian, and D. H. K. Tsang, "NOMAassisted multi-access mobile edge computing: A joint optimization of computation offloading and time allocation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12244–12258, Dec. 2018.

[25] S. Fu, J. Gao, and L. Zhao, "Integrated resource management for terrestrial-satellite systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3256–3266, Mar. 2020.

[26] ITU-R, "Propagation data and prediction methods required for the design of Earth-space telecommunication systems," International Telecommunication Union (ITU), Recommendation P.618-13, 12 2017. [Online]. Available: https://www.itu.int/dms pubrec/itu-r/rec/p/ R-REC-P.618-13-201712-I!!PDF-E.pdf

[27] ITU-R, "Specific attenuation model for rain for use in prediction methods," International Telecommunication Union (ITU), Recommendation P.838-3, 03 2005. [Online]. Available: https://www.itu.int/dms pubrec/itu-r/rec/p/R-REC-P.838-3-200503-I!!PDF-E.pdf

[28] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, vol. 54, 2017, pp. 1273–1282.

[29] Z. Chen, T. -B. Xu, C. Du, C. -L. Liu and H. He, "Dynamical Channel Pruning by Conditional Accuracy Change for Deep Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 799-813, Feb. 2021.

[30] R. Xing, Z. Su and Y. Wang, "Intrusion Detection in Autonomous Vehicular Networks: A Trust Assessment and Q-learning Approach," *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Paris, France, 2019, pp. 79-83.

[31] F. Zenke, B. Poole and S. Ganguli, "Continual learning through synaptic intelligence", *Proc. 34th Int. Conf. Mach. Learn.*, pp. 3987-3995, 2017.

[32] X. Ma, J. Zhang, S. Guo and W. Xu, "Layer-wised model aggregation for personalized federated learning", *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 10092-10101, 2022.

[33] A. Rubinstein, "Perfect equilibrium in a bargaining model," *Econometrica: Journal of the Econometric Society*, pp. 97–109, 1982.

[34] R. Deng, B. Di and L. Song, "Pricing Mechanism Design for Data Offloading in Ultra-Dense LEO-Based Satellite-Terrestrial Networks," *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1-6.

[35] Y. Pan, M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei, "Energy efficient NOMA-based mobile edge computing offloading," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 310–313, Feb. 2019.

[36] K. Maine, C. Devieux and P. Swan, "Overview of iridium satellite network", *Proc. WESCON Conf. Rec. Microelectron. Commun. Technol. Producing Quality Products Mobile Portable Power Emerg.*, pp. 483, Nov. 1995.

[37] 3GPP TR, 38.811 (V0.3.0), "Study on New Radio (NR) to Support non Terrestrial Networks (Release 15)", Dec. 2017.

[38] E. Bourtsoulatze, D. Burth Kurka and D. Gündüz, "Deep Joint Source-Channel Coding for Wireless Image Transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567-579, Sept. 2019.

[39] A. Krizhevsky, V. Nair, and G. Hinton, "Cifar-10 (Canadian Institute for Advanced Research)." [Online]. Available: http://www.cs.toronto.edu/ kriz/cifar.html

[40] A. Krizhevsky, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-2009, 2009

[41] Y. LeCun, C. Cortes, and C. J. Burges. The MNIST database of handwritten digits, 1998. URL http://yann.lecun.com/exdb/mnist/.

[42] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "Exploiting shared representations for personalized federated learning," arXiv preprint arXiv:2102.07078, 2021.