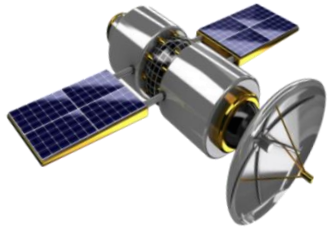
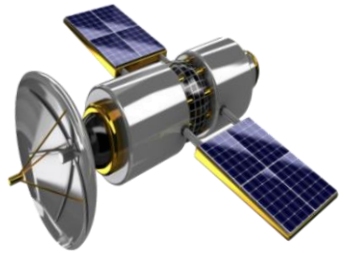




Does Space AI Security Matters? Unlocking Privacy-Preserving Federated LEO Satellite Learning for Border Threat Detection

Oct. 08, 2024, Richland, Washington

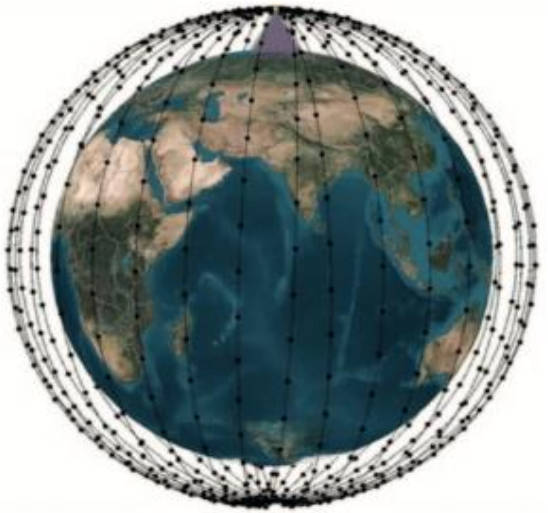
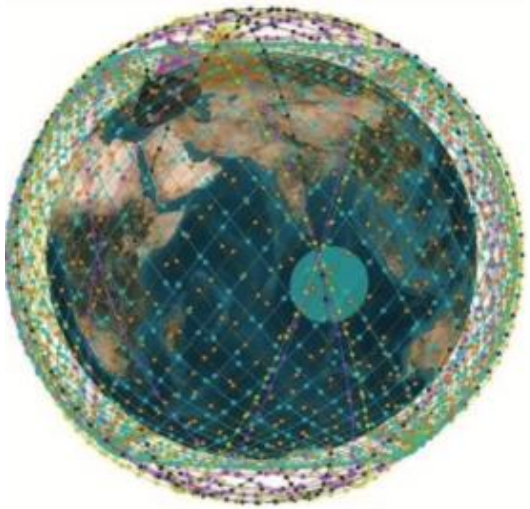
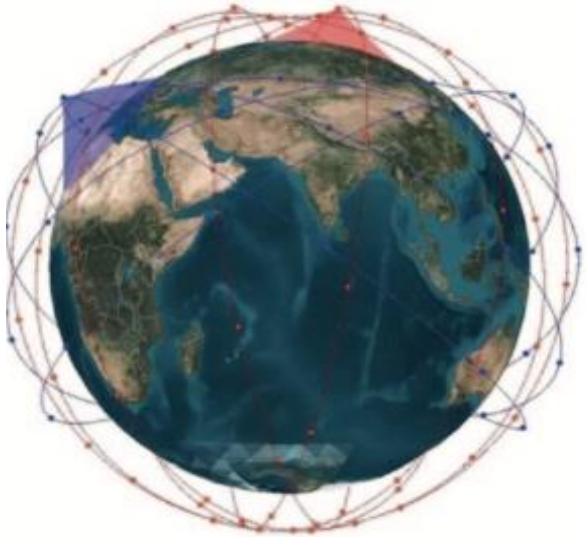
Presenter: Mohamed Elmahallawy

Assistant Professor, Computer Science & Cybersecurity Program

WASHINGTON STATE - Tri-Cities UNIVERSITY

Low Earth Orbit (LEO) Satellite Networks

- The **recent surge** of **interest** and **investment** in large-scale LEO satellites

- **Industry**

SPACEX STARLINK

amazon Project Kuiper

ONEWEB EUTELSAT

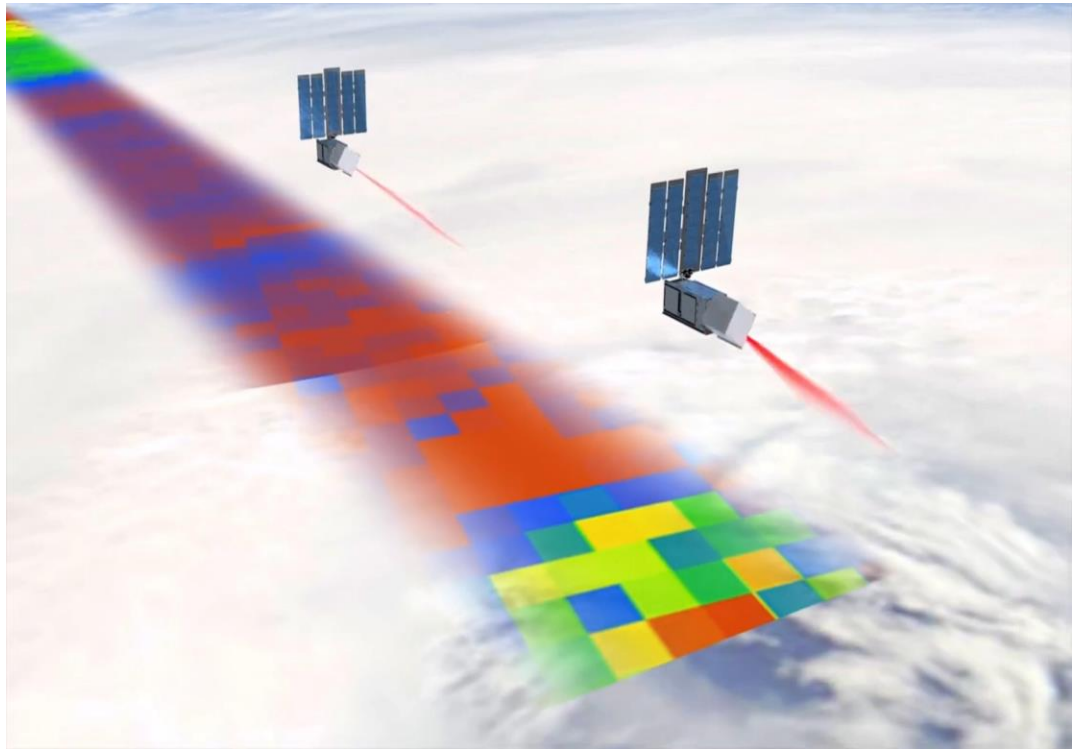
- **Government**

NASA

esa European Space Agency

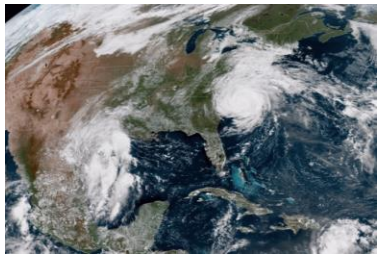
Low Earth Orbit (LEO) Satellite Networks

- Enable novel applications empowered by machine learning, such as:



Border Monitoring

Disaster Detection

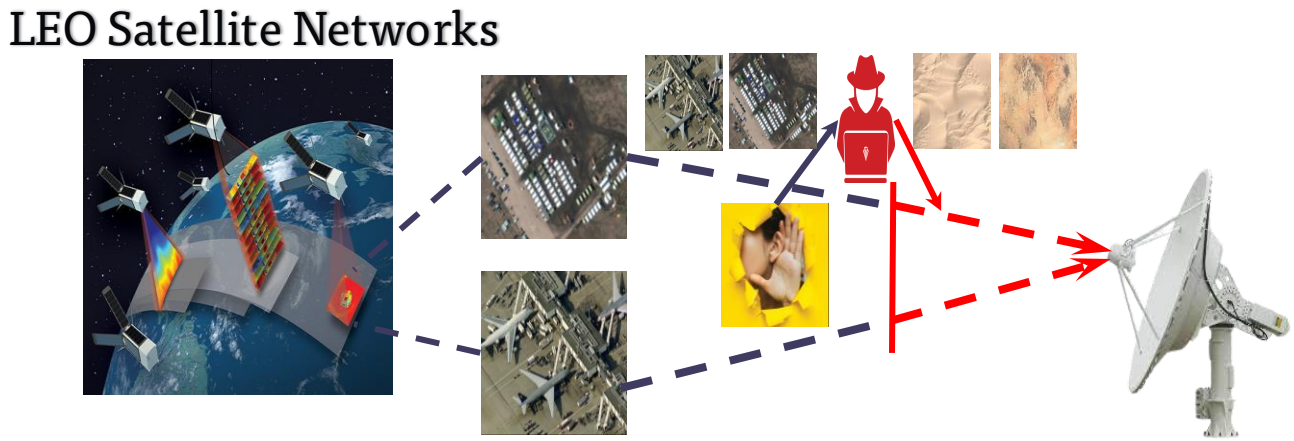


5 petabytes of image data per day (2019)!

[1] P. Wang, H. Li, B. Chen, and S. Zhang, "Enhancing earth observation throughput using inter-satellite communication," IEEE Transactions on Wireless Communications, vol. 21, no. 10, pp. 7990–8006, 2022

Low Earth Orbit (LEO) Satellite Networks

- Conventional **centralized** learning requires satellites to download their **high-resolution** images to a ground station (GS)
- This is **impractical** because:
  - **Privacy and Security Challenges:**
    - ✓ **Raw data** transmission
  - **Efficiency Challenges:**
    - ✓ **Limited Bandwidth** (available 50~500MB vs. 5 petabytes satellite data!)
    - ✓ **Sporadic and irregular** visibility to the GS (a few times a day and each last in 5 minutes)

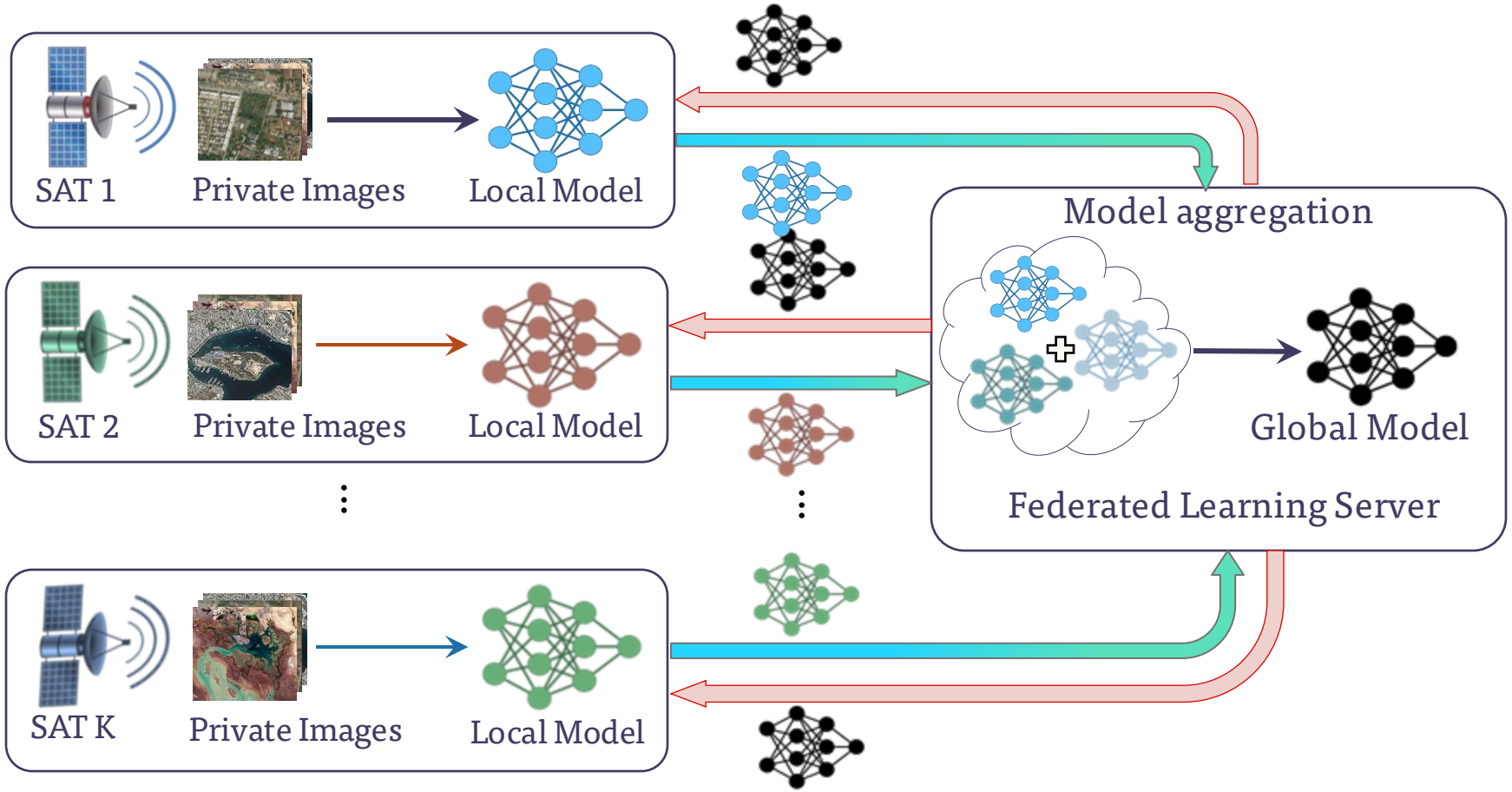


Consequence of delay

Four–hour delay in downloading images resulted in the **burning** of **140,000 acres** and loss of **56 lives** [2]

[2] https://news.berkeley.edu/2018/11/15/new-satellite-view-of-camp-fire-as-it-burned-through-paradise/

Introducing Federated Learning (FL) into LEO Networks



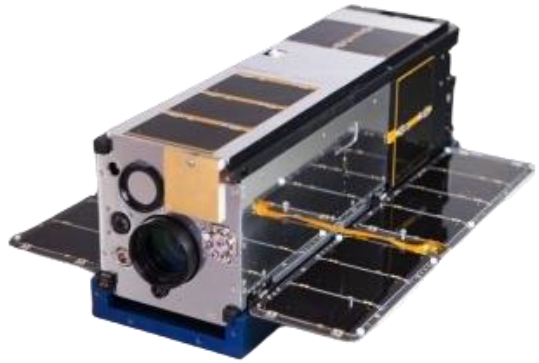
Challenges in FL-LEO Network

✖ Privacy and Security Threats

- Communication over **insecure** channel, making satellites' model vulnerable to various attacks such as **model inversion** and **membership** inference

✖ Limited Computation and Storage

- LEO satellites **cannot** train large-scale ML models onboard **!!!**



LEO satellite (10x10x30 cm)

✖ Sporadic and Irregular Visibility Pattern

- **Iterative** nature of the **FL process** causes the global model to take several **days** or even **weeks** to converge.

Privacy vs. Security

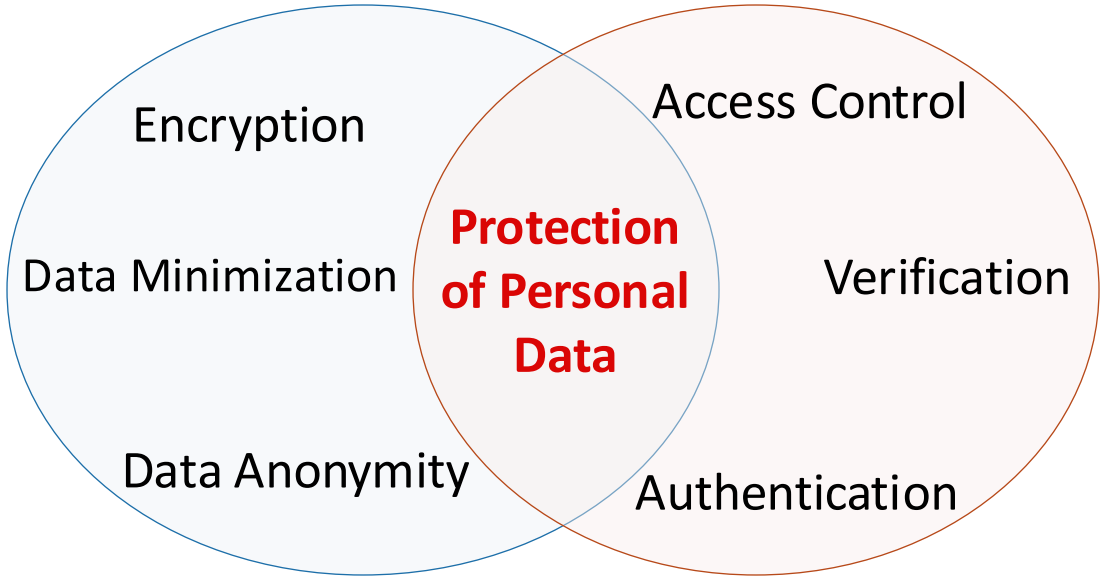
- **Privacy and Security are intertwined, but they are not same**

Privacy

Safeguarding users' personal data
"Data Control"

Security

Preventing unauthorized access of personal data
"Data Integrity "



Authentication vs. Verification

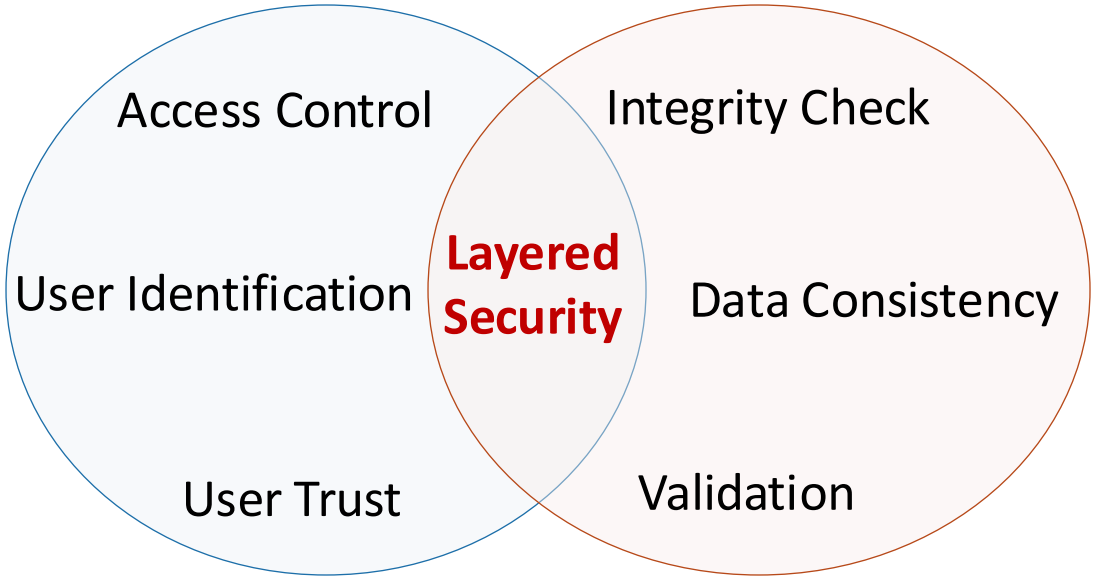
- **Authentication & verification are intertwined, but they serve different purposes**

Authentication

It is used to confirm a user's identity
"who are you"

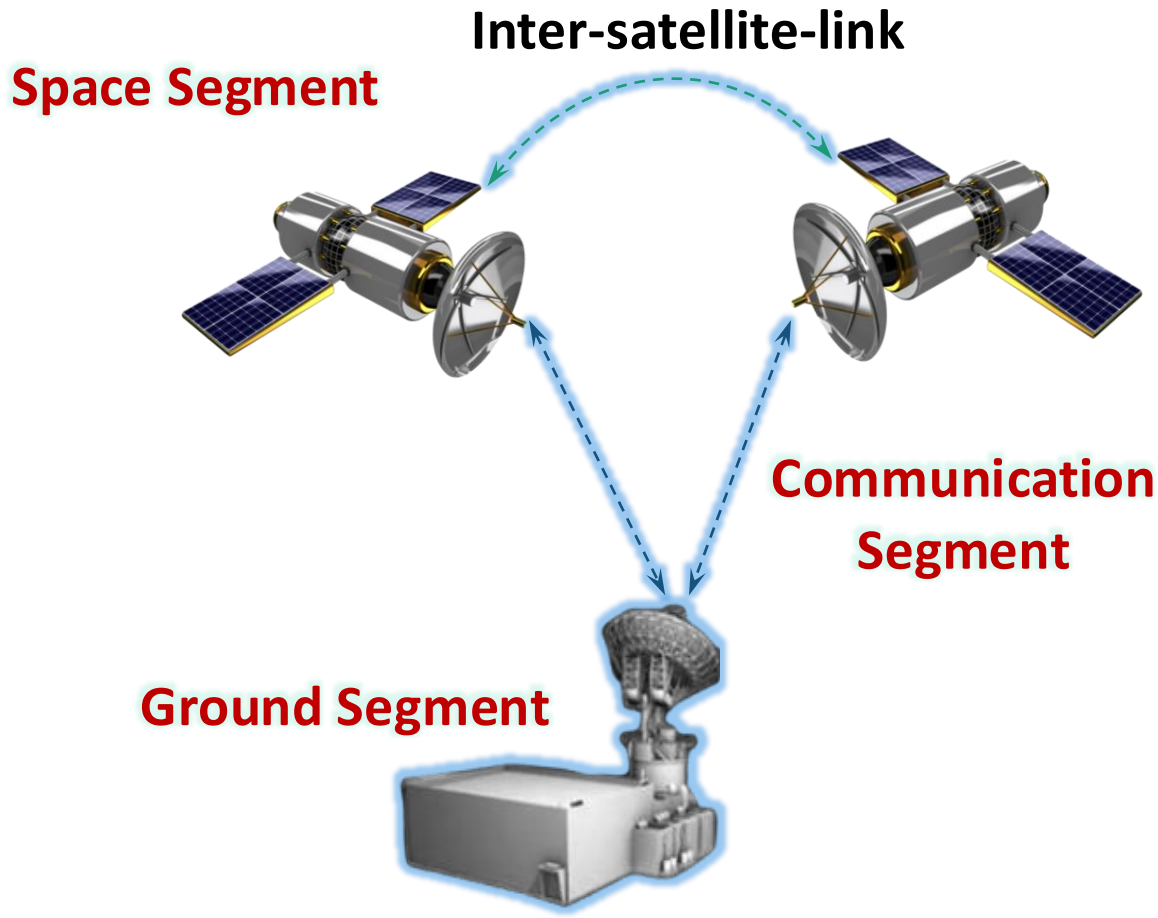
Verification

It checks whether a particular action is true
"Is this what it claims to be?"

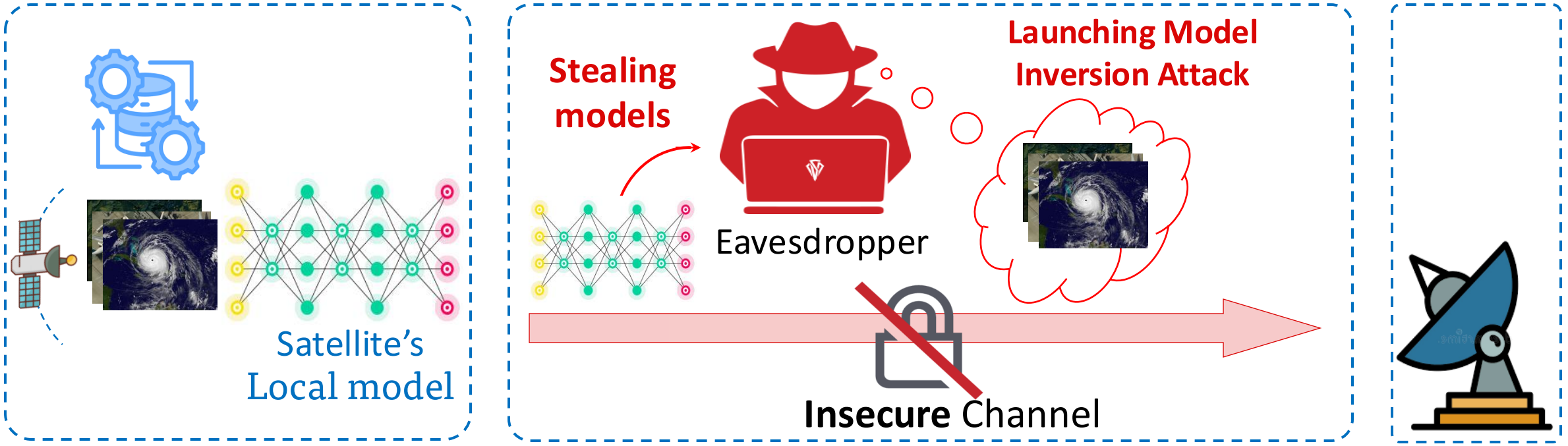


Threats in FL-LEO Network

- **Space Segment**
  - Command intrusion
  - Payload control
  - Poison
  - ……
- **Communication Link Segment**
  - Data Stealing
  - Data manipulating
  - Replay
  - ……
- **Ground Segment**
  - Hacking
  - Malware
  - Colluding
  - …



Threats in FL-LEO Network

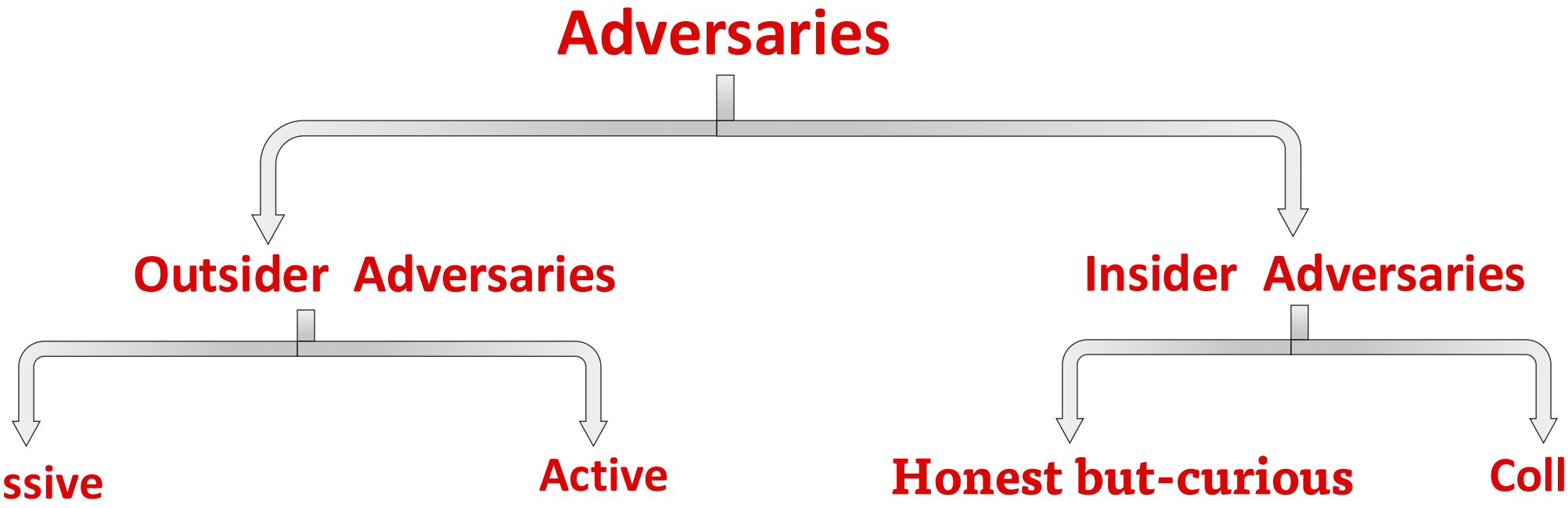


On-board Training

Transmitting Models Over the Air

Ground Station

Adversaries in FL-LEO Network



- **Steal** local or global models for malicious purposes
- **Reply** or **Impersonation** attacks

- **Manipulate** model parameters
- Introduce **bias** into the model
- Hinder the **convergence** process
- E.g., Model **integrity** attacks

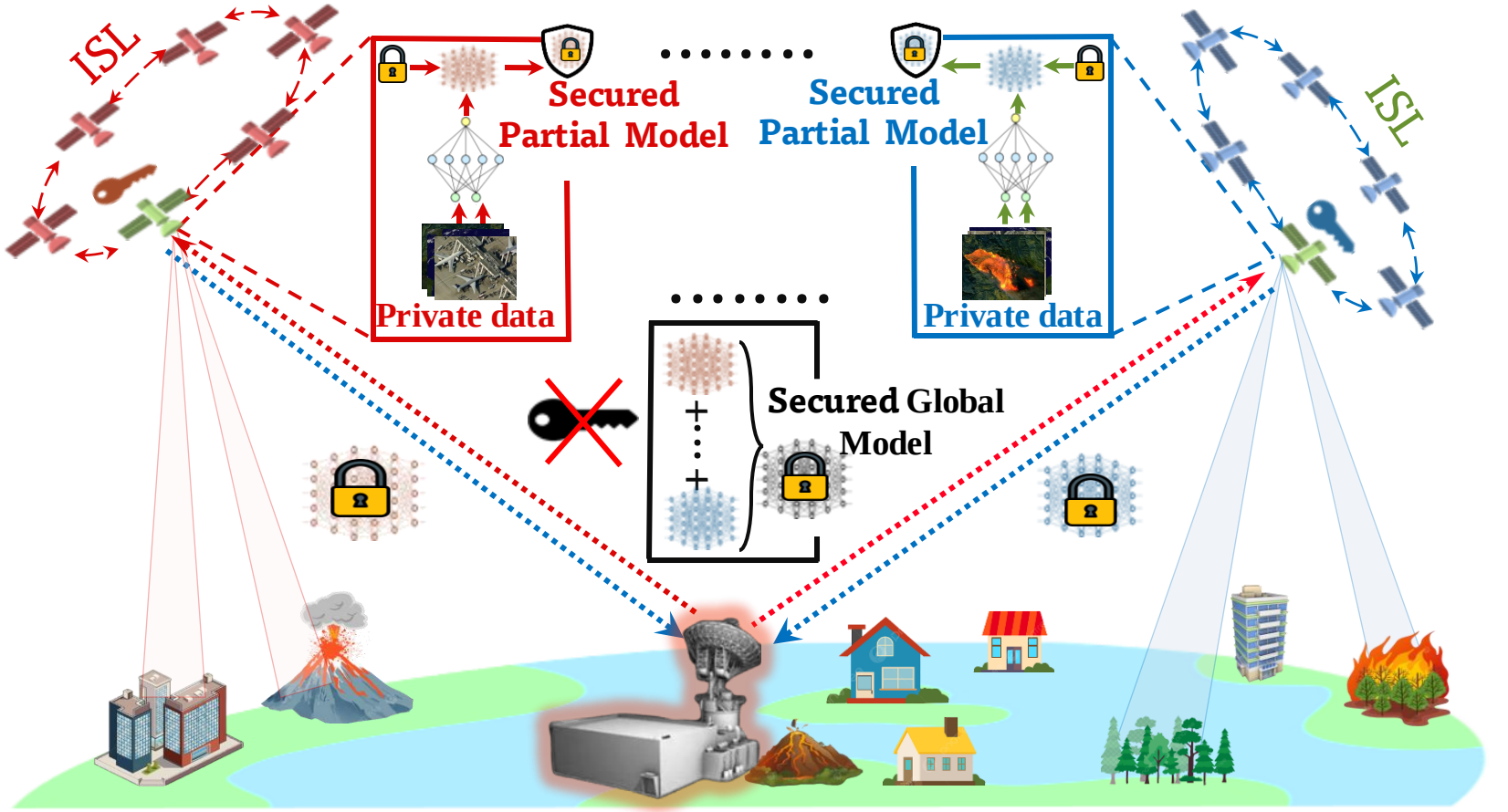
- Following the FL **honesty**
- Server has **curiosity**
- To **learn** information about satellite's raw data.

- Satellites from a vendor may **collude** with the server
- To train **biased** models or **steal** of other vendors' models

1. Secure Aggregation

WASHINGTON STATE - Tri-Cities
UNIVERSITY

Secure Aggregation



Methods

➢ Tackling Privacy Concerns

✓ We encrypt satellite models using a lightweight cryptography

✓ Anonymous veto protocol is used to generate public and private keys in a decentralized manner.

➢ Tackling Security Concerns

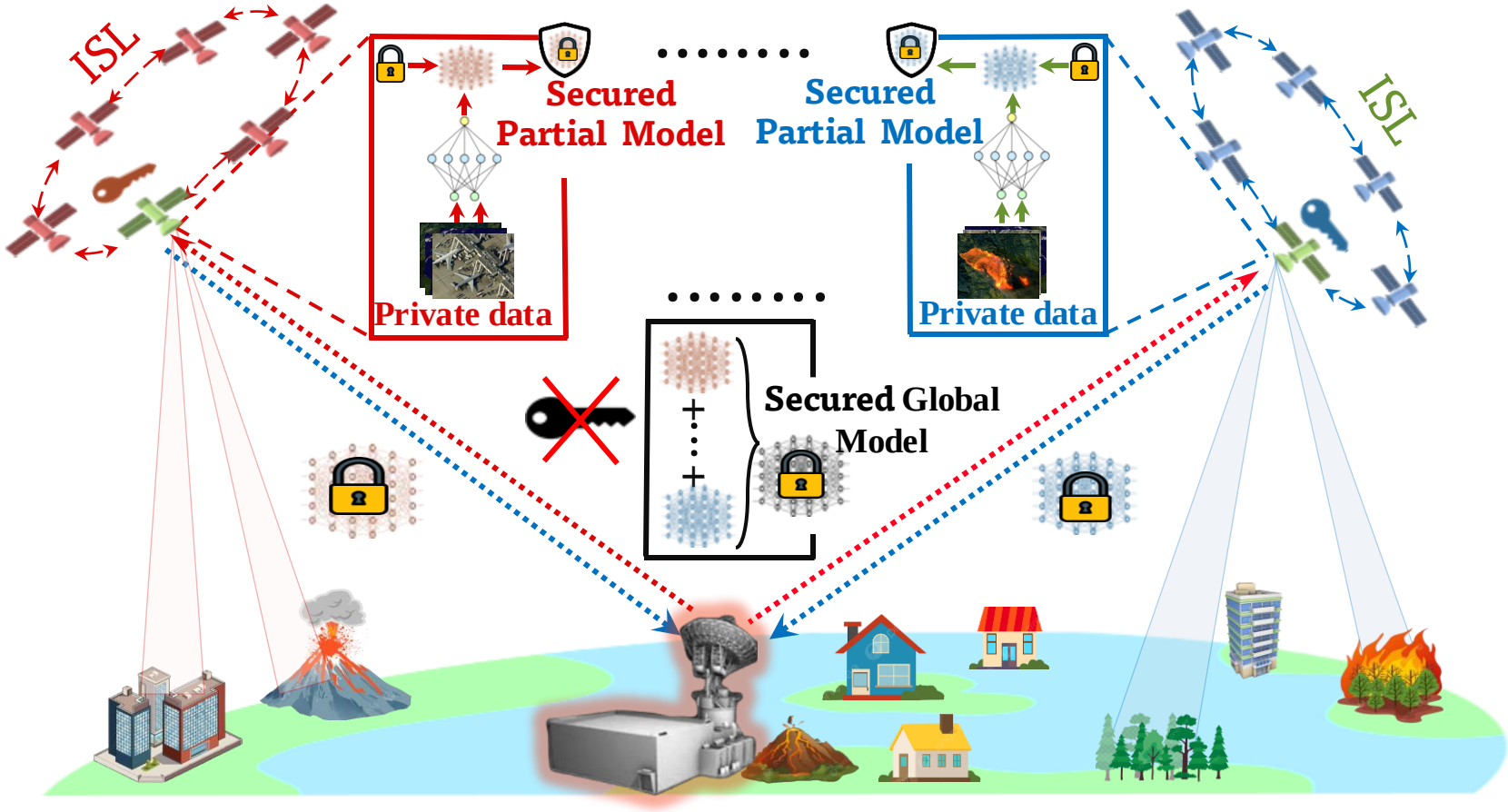
❑ Before Aggregation

GS employs a secure model verification mechanism that guarantees the integrity of models

❑ During Aggregation

GS employs an inner product functional encryption scheme to provide secure aggregation for all satellites' local models.

Secure Aggregation



Advantages

- ✓ Not require trusted key distribution center (**KDC**) or any **secure channel** to distribute the keys among satellites.
- ✓ The max the $\mathcal{AS}$ can do is to obtain the **aggregation sum** of satellites models, **revealing** nothing about their **private date**.
- ✓ The **integrity** of satellites' models is protected against model any **model integrity attacks**.
- ✓ **Less computation** and **communication** overheads.

Secure Aggregation (Results)

Privacy and Security Comparison with the state-of-the-art

- We compare our approach, **SecFLEO**, with Threshold Homomorphic Encryption (**THE**) [2].

A. Computation Overhead

- The computation is **measured** by calculating the **time** required to **encrypt** the parameters of ML models.
- SecFLEO takes less than **1 msec**, whereas THE takes more than **22.3 msec**, indicating a substantial **23x** difference.

B. Communication Overhead

- The communication is measured by analyzing the **size of message** exchanges between the AS and LEO satellites.
- THE involves a much larger models (39x), limiting its scalability substantially.

TABLE V
COMPARISON OF COMMUNICATION COST IN SECFLEO AND THE IN EACH ROUND.

Model-Dataset	Communication Overhead		Computation Overhead (FLOPS)	
	THE	**SecFLEO**	THE	**SecFLEO**
CNN–MNSIT	641.37 MB	16.70 MB	359.5 G	14.86 G
CNN–CIFAR-10	1,169.90 MB	30.47 MB	982.33 G	42.71 G
CNN–CIFAR-100	11,366.49 MB	296 MB	1227.28 G	53.36 G
DeepLabV+–DeepGlobe	39,080.75 MB	1,017.73 MB	2761.48 G	115.18G

[2] J. Park and H. Lim, "Privacy-preserving federated learning using homomorphic encryption," Applied Sciences, vol. 12, no. 2, 2022

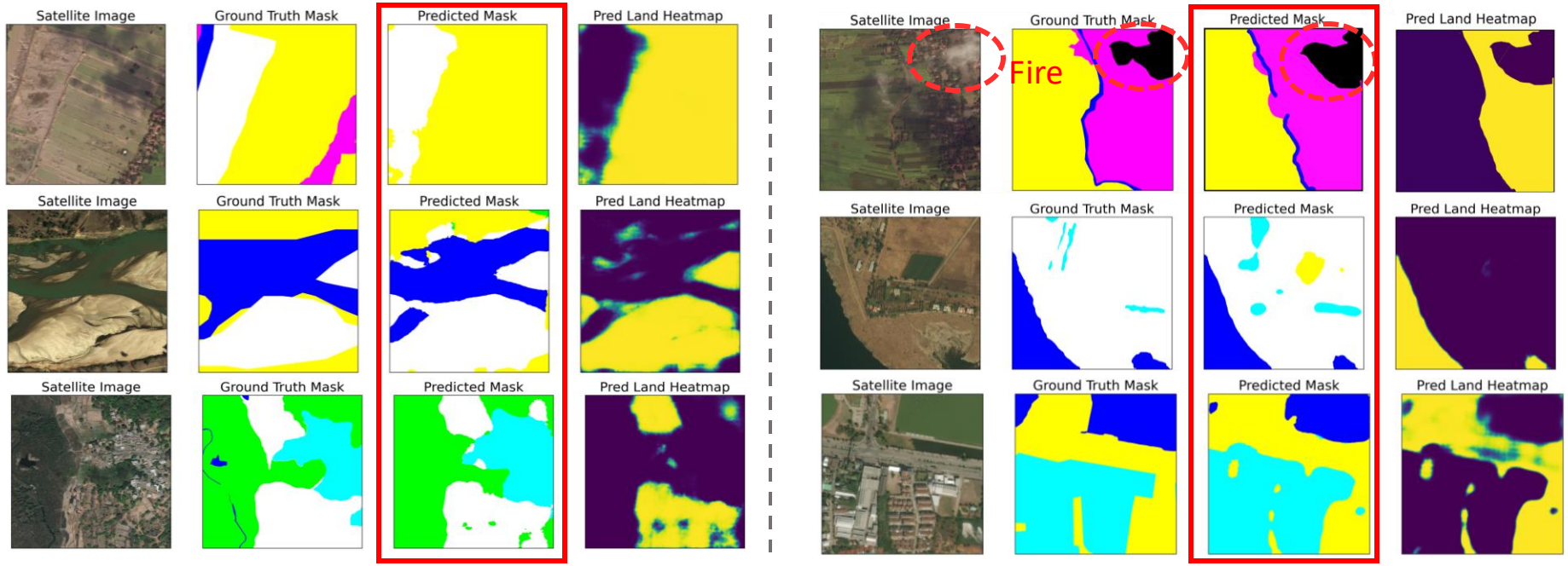
Secure Aggregation (Results)

Performance Comparison with the state-of-the-art

- We evaluate SecFLEO under **realistic** LEO satellite **network**, and satellites dataset, **DeepGlobe**

C. Semantic Segmentation tasks (DeepGlobe)

- Predicating 7 classes represented in various colors: Urban, Agriculture, Rangeland, Forest, Water, Barren, and Cloud/Fire



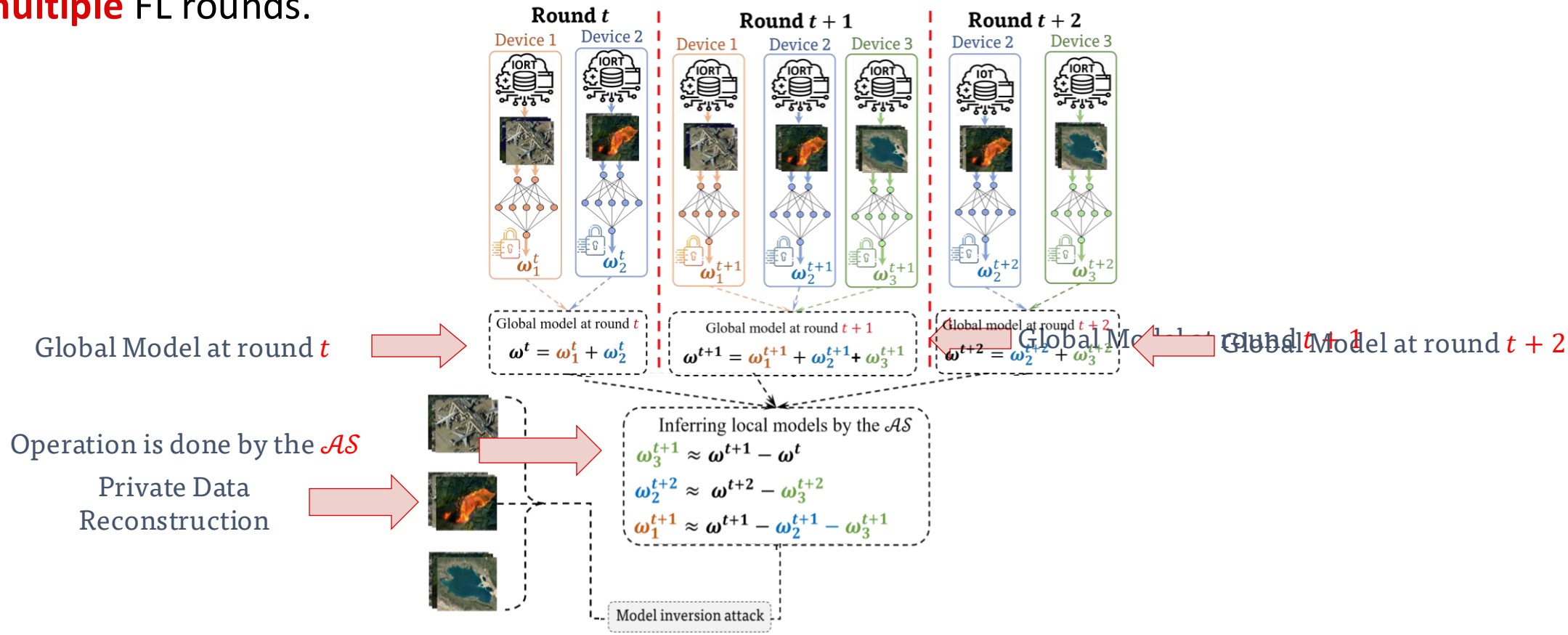
Sample of satellite images with truth masks, predicted masks, and heatmaps of SecFLEO *after only **3 hours** of convergence*

1. Long-Term Privacy

WASHINGTON STATE - Tri-Cities
UNIVERSITY

Secure Aggregation Concern

- We uncover that secure aggregation approaches are **myopic** and may **fail** against **strategic** intruders.
- It protects only each **single FL round**, overlooking the risk of strategic privacy threats across **multiple** FL rounds.



Long-Term Privacy (LTP) Persevering

- To address the LTP leakage, we propose **L**ong **T**erm **P**rivacy-preserving asynchronous **F**ederated learning for **L**ow **E**arth **O**rbit satellite networks (LTP-FLEO)
- LTP-FLEO consists of **three** main components:

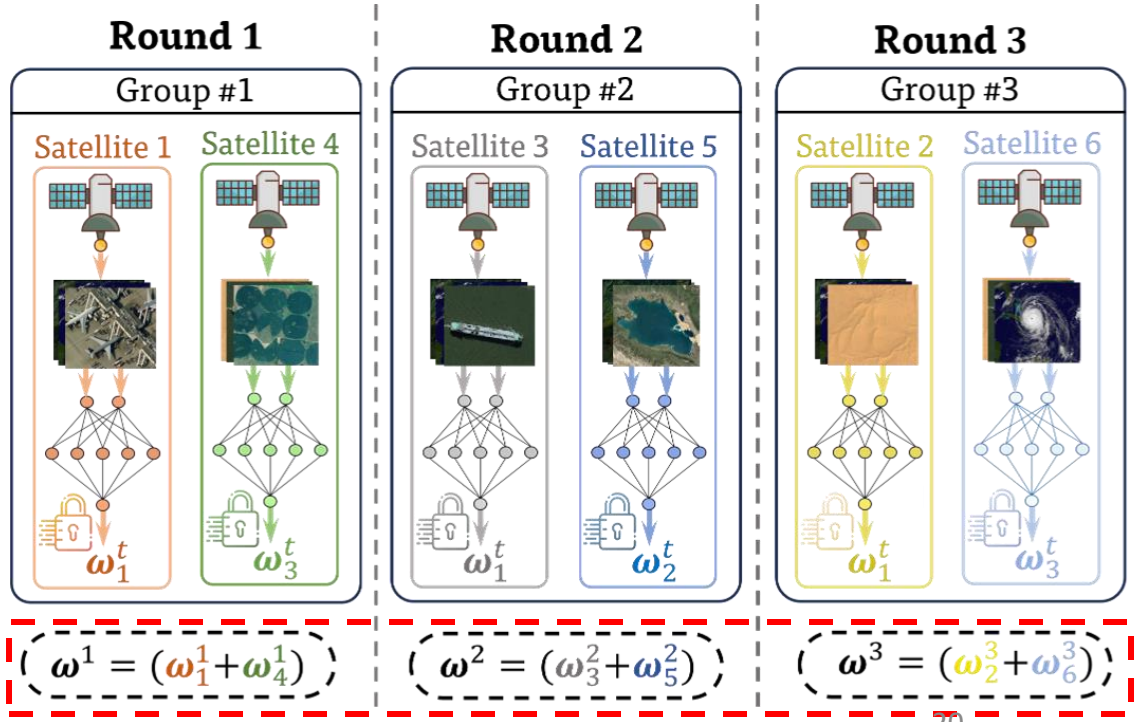
Privacy-aware satellite grouping

Partition LEO networks into groups, where satellites within the same groups should **participate together**, or **not participate** at all.

$$G^* = \arg\min_{G\in\mathcal{G}} \max_{k\in G} p_k^t$$

p_k^t: Probability of satellite k to be visible in round t

The maximum the **GS** can do is to obtain the **partial sum** of satellite models, **unable** to decode any **individual** model.



Long-Term Privacy (LTP) Persevering

- To address the LTP leakage, we propose **L**ong **T**erm **P**rivacy-preserving asynchronous **F**ederated learning for **L**ow **E**arth **O**rbit satellite networks (LTP-FLEO)
- LTP-FLEO consists of **three** main components:

Privacy-aware satellite grouping → Model age handling

Balance the impact of **stale** and **fresh** satellites' local models on the global model

Participation frequency for each group G

$$f_G^t \triangleq \sum_{m=1}^{t-1} \mathbb{1}\left\{s_G^m = 1\right\}, \ \forall G \in \mathcal{G}'$$

Binary indicator for the participation in each group in round m

Introduce a **tolerance factor** α, such that $t - \alpha \leq f_G^t \leq t - 1$

The GS **chooses** only those groups G that **meet** the condition to participate in round t.

Long-Term Privacy (LTP) Persevering

- To address the LTP leakage, we propose **L**ong **T**erm **P**rivacy-preserving asynchronous **F**ederated learning for **L**ow **E**arth **O**rbit satellite networks (LTP-FLEO)
- LTP-FLEO consists of **three** main components:

Privacy-aware satellite grouping → Model age handling → Fair global aggregation

- Prevent a **biased global model** by ensuring different orbits of satellites contribute to the global model **equally**

Fairness gap

$$F = \max_{G\in\mathcal{G}'} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left\{s_G^t = 1\right\}\right] - \min_{G\in\mathcal{G}'} \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}\left\{s_G^t = 1\right\}\right]$$

- **Small** weight for **outdated** models
- **Larger** weight for **fresh**/regularly updated models, such that

$$\beta_G^t = \frac{\gamma_G^t}{\sum_{G\in\mathcal{G}''}\gamma_G^t}, \quad \gamma_G^t = \frac{f_G^t}{\sum_{G\in\mathcal{G}''} f_G^t} \times \frac{|D_G|}{|D_{\mathcal{G}''}|} \longrightarrow \boldsymbol{w}^{t+1} = \sum_{G\in\mathcal{G}''}\beta_G^t \sum_{k\in G}\boldsymbol{w}_k^t$$

Long-Term Privacy (Results)

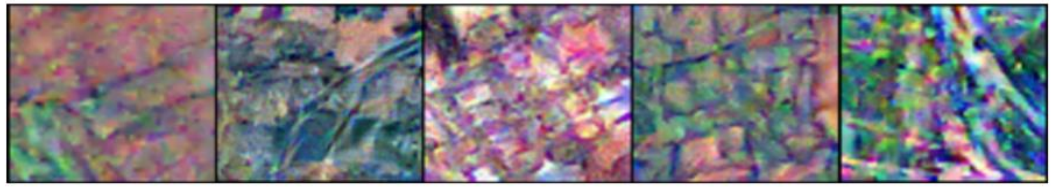
- Results of LPAFO with different **group sizes** (L = 2, 4, 6 - number of satellites per group) on the EuroSat dataset for classification tasks under **model inversion attack**

Table 2: Comparison of Image Reconstruction Performance: LPAFO vs. FedSecure under various metrics.

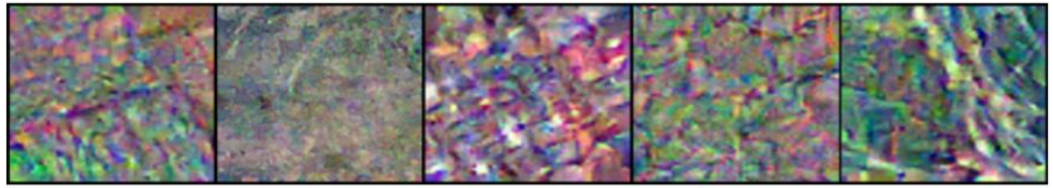
Metric	FedSecure	LPAFO		
	$L = 1$	$L = 2$	$L = 4$	$L = 6$
PSNR ↓	10.68	5.28	3.03	-1.98
MSE ↑	0.1076	0.528	0.898	1.577
FMSE ↑	0.1558	1.063	1.926	3.308
Resolution Acc %	97.65	56.97	36.16	22.96

(a) Ground Truth samples of EuroSat dataset.

(b) Reconstructed samples of EuroSat dataset under FedSecure.



(c) Reconstructed samples of EuroSat dataset under LPAFO when L=2.



(d) Reconstructed samples of EuroSat dataset under LPAFO when L=4.

(e) Reconstructed samples of EuroSat dataset under LPAFO when L=6.

Long-Term Privacy (Results)

Classification tasks (EuroSat)

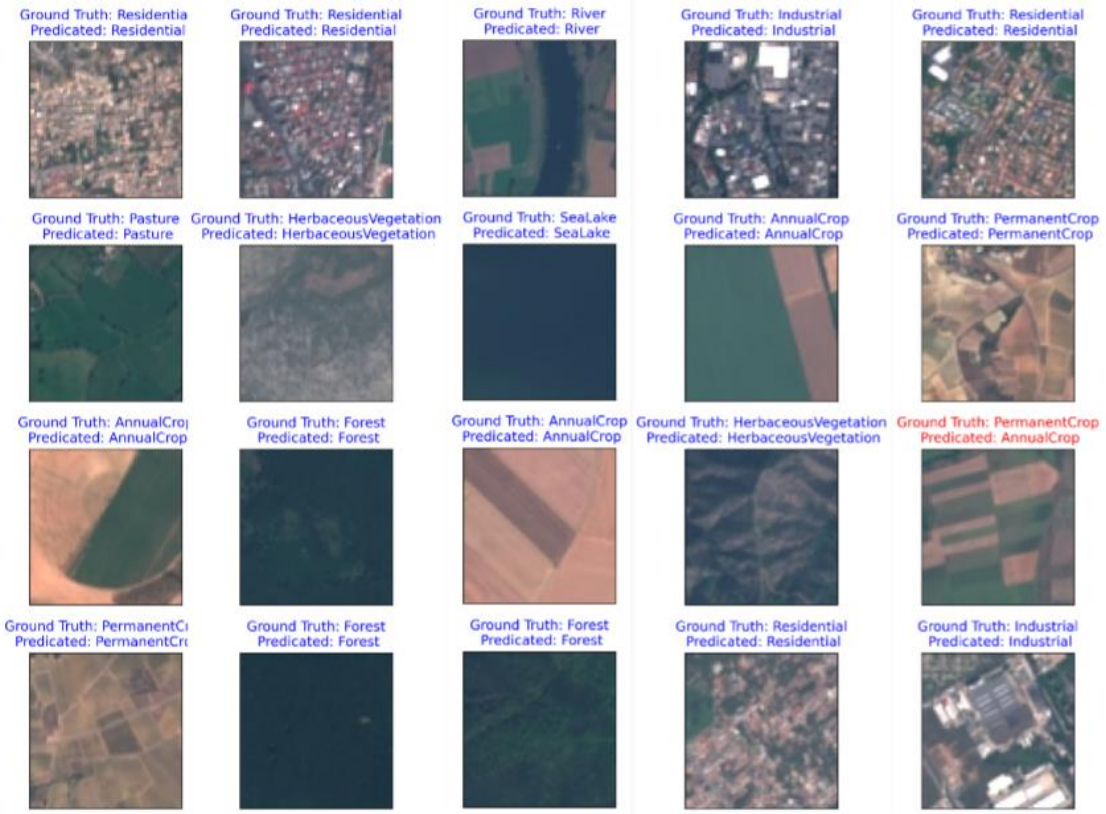
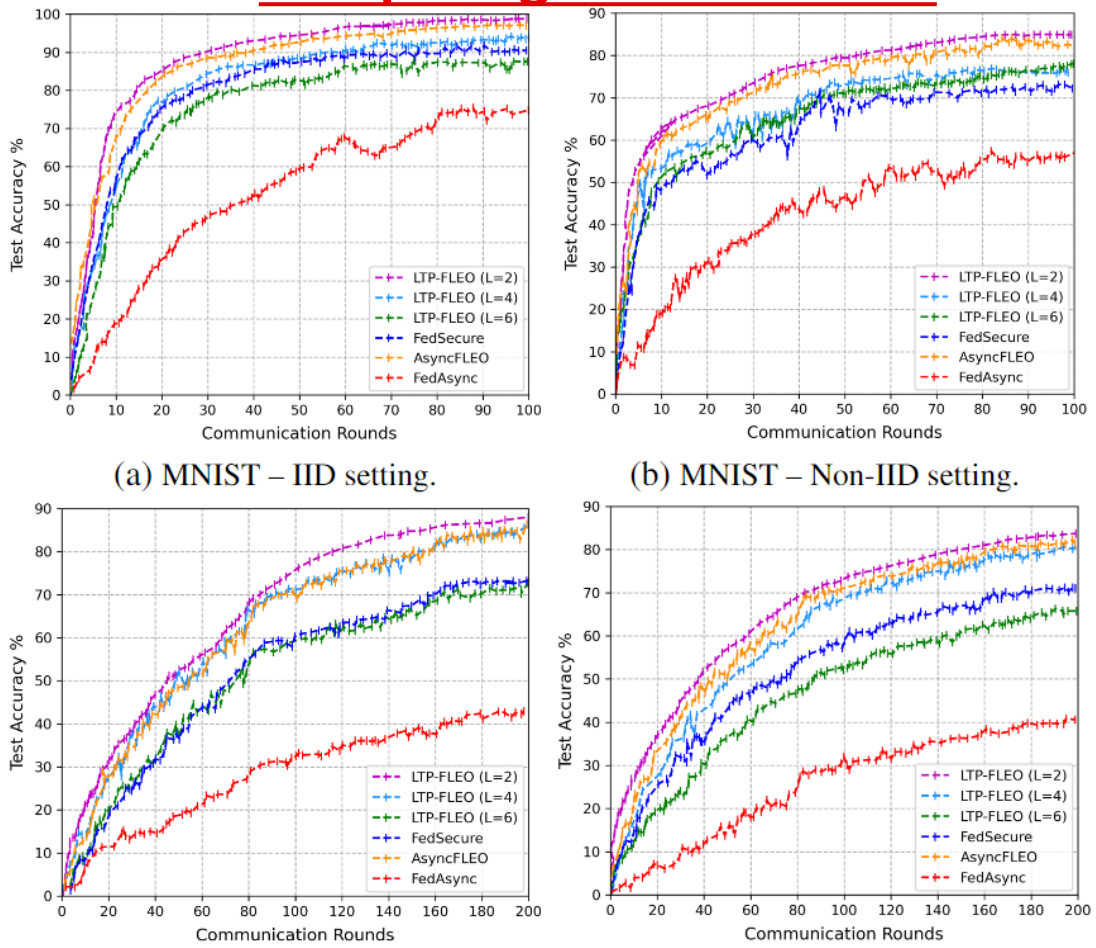


Fig. 6: Twenty randomly selected images from a Eurostat test set of 5400 samples, illustrating the predicted vs. ground truth labels. Blue and red color represent correct and incorrect predictions, respectively.

Comparing with the SOTA



(a) MNIST – IID setting.

(b) MNIST – Non-IID setting.

(c) CIFAR-10 – IID setting.

(d) CIFAR-10 – non-IID setting.

Figure 5. Performance comparison with baselines on MNIST and CIFAR-10 datasets under IID and non-IID data distributions.

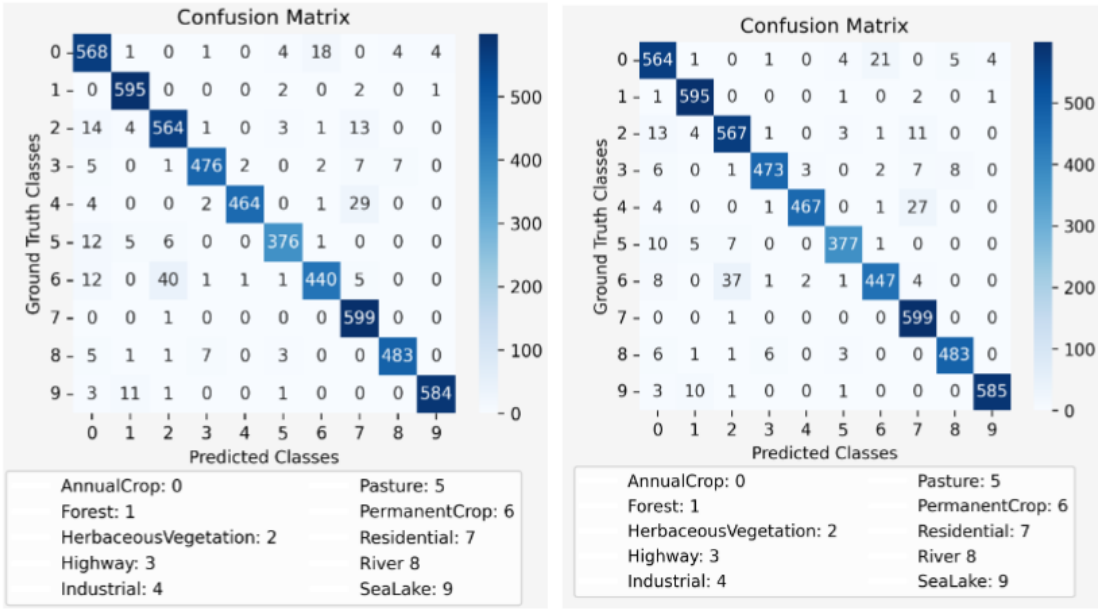
Long-Term Privacy (Results)

Individual Class Accuracy for EuroSat Dataset

TABLE IV: Evaluation of our approach on EuroSat Dataset.

Class \ Metric	# of images	Near-polar constellation (85°)				Inclined constellation (45°)			
		ACC(%)	PC (%)	RC (%)	F1 (%)	ACC(%)	PC (%)	RC (%)	F1(%)
AnnualCrop	600	97.63	92.06	94.67	93.34	98.39	91.71	94.0	92.84
Forest	600	98.19	96.73	98.67	97.69	99.52	96.59	99.17	97.86
HerbaceousVegetation	600	99.32	93.05	93.67	93.36	98.5	92.12	94.5	93.33
Highway	500	99.61	97.74	95.0	96.35	99.31	97.93	94.60	96.24
Industrial	500	98.89	98.95	93.80	96.30	99.30	98.94	93.40	96.09
Pasture	400	99.23	96.42	94.25	95.32	99.33	96.67	94.25	95.44
PermanentCrop	500	97.46	95.14	90.0	92.50	98.54	94.50	89.40	91.88
Residential	600	99.01	91.45	99.83	95.46	99.04	92.15	99.83	95.84
River	500	99.56	98.17	96.60	97.38	99.44	97.37	96.60	96.99
SeaLake	600	99.11	98.66	98.0	98.33	99.63	99.15	97.50	98.32

Better visibility **Less visibility**

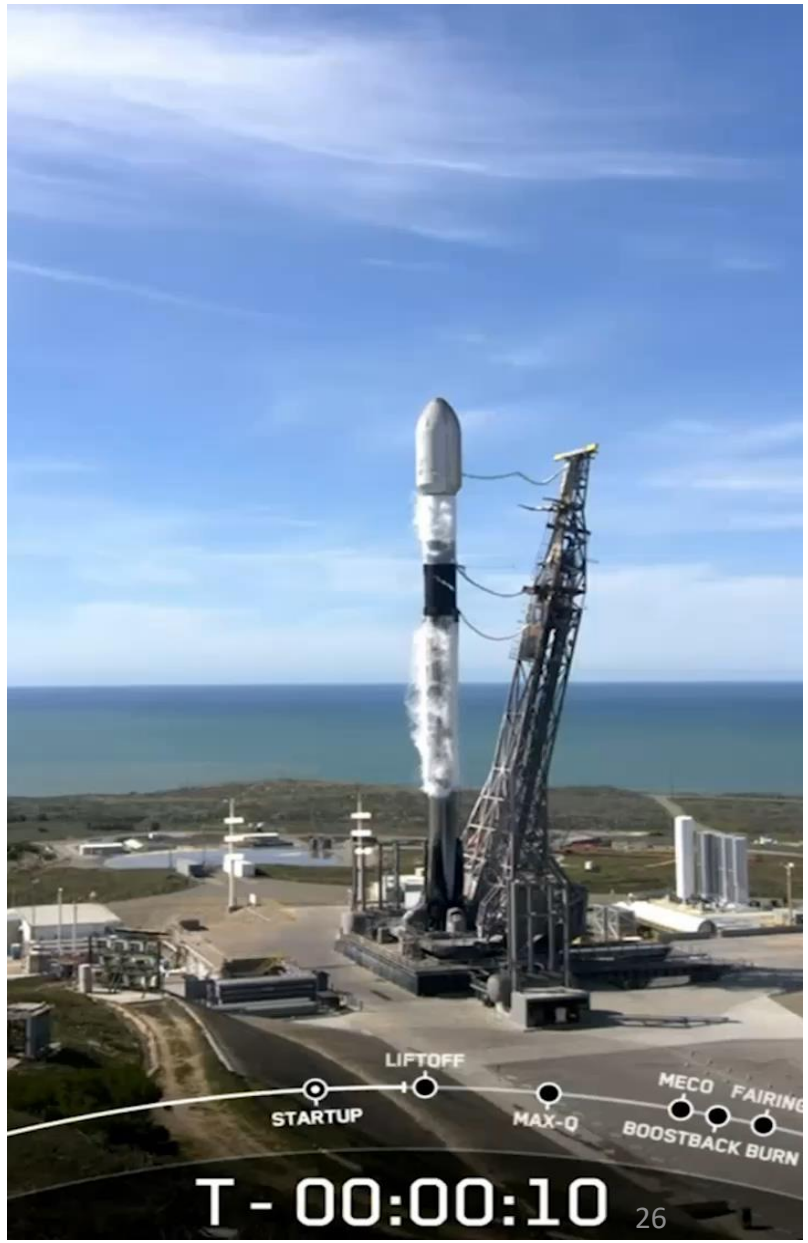


(a) Near-polar constellation (85°). (b) Inclined constellation (45°).

Fig. 5: Confusion matrix that compares 10 predicted and ground-truth classes for 5400 test images.

From Lab to Space

- **Recent Launch (March 6, 2024)**
  - SpaceX launched an LEO satellite developed by **Missouri University of Science and Technology**, tasked with capturing Earth images.
  - Equipped with Raspberry Pi
- **Future Launch (Feb. 2026)**
  - Currently working on building two other LEO satellites: *Missouri Rolla* (**MR**) and *Missouri Rolla Second* (**MRS**).
  - MR and MRS will be equipped with Jetson Nano to train ML models onboard



- AI-driven **satellite edge computing (SEC)** is coming.
- **Security** and **privacy** risks can be addressed through **decentralized secure aggregation** and **privacy-aware schemes**.
- LEO satellites can effectively **run FL** despite their limited computational capability by enabling satellites to train **lightweight models**
- **Convergence** speed of FL-LEO can be **accelerated** while ensuring **security**, converged **accuracy**, and **fair** aggregation

Funded by: NSF

Acknowledgement: NASA

Thank You

Questions?