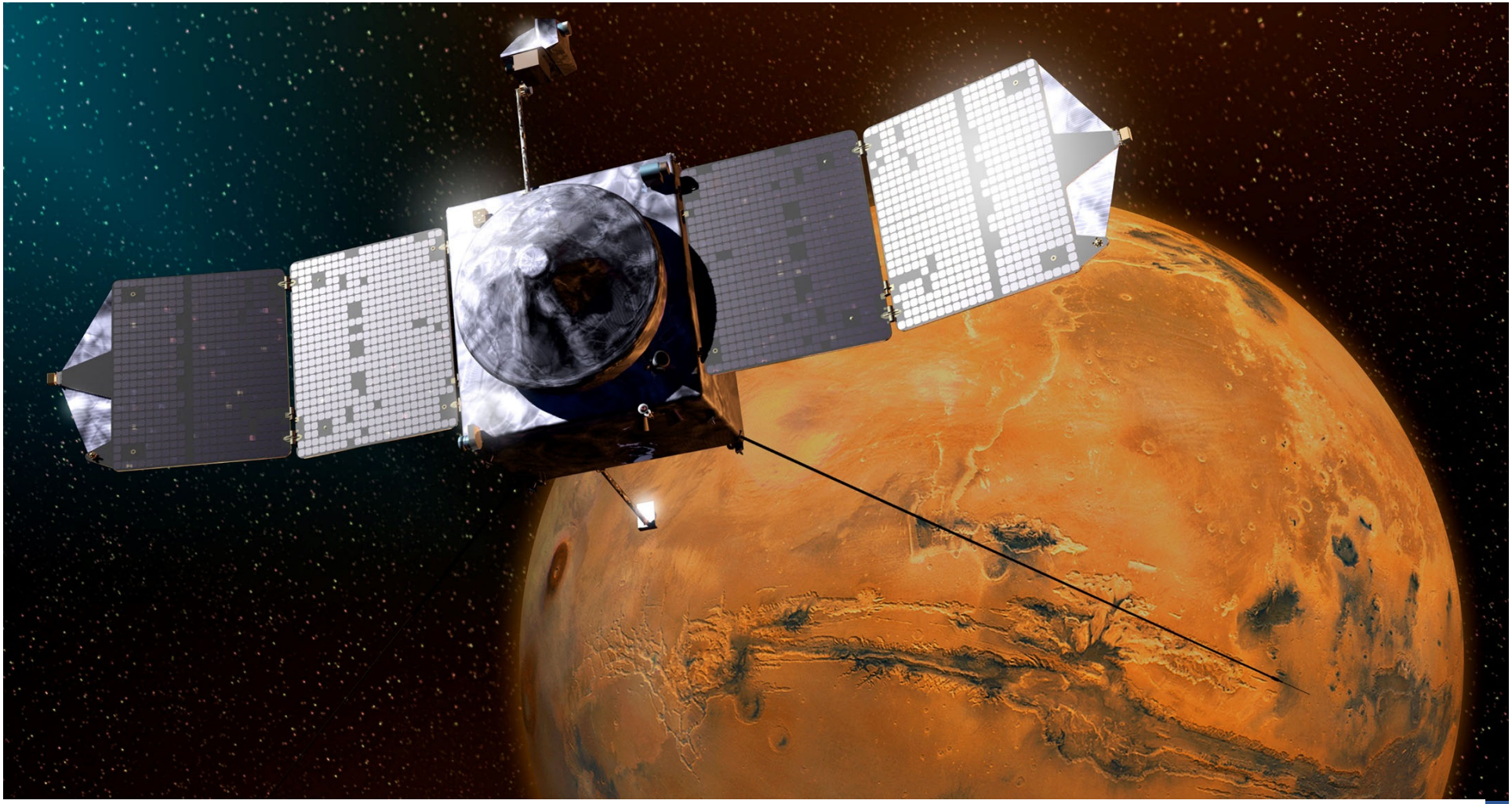


FEDERATED LEARNING BASED INTRUSION DETECTION SYSTEM FOR SATELLITE COMMUNICATION

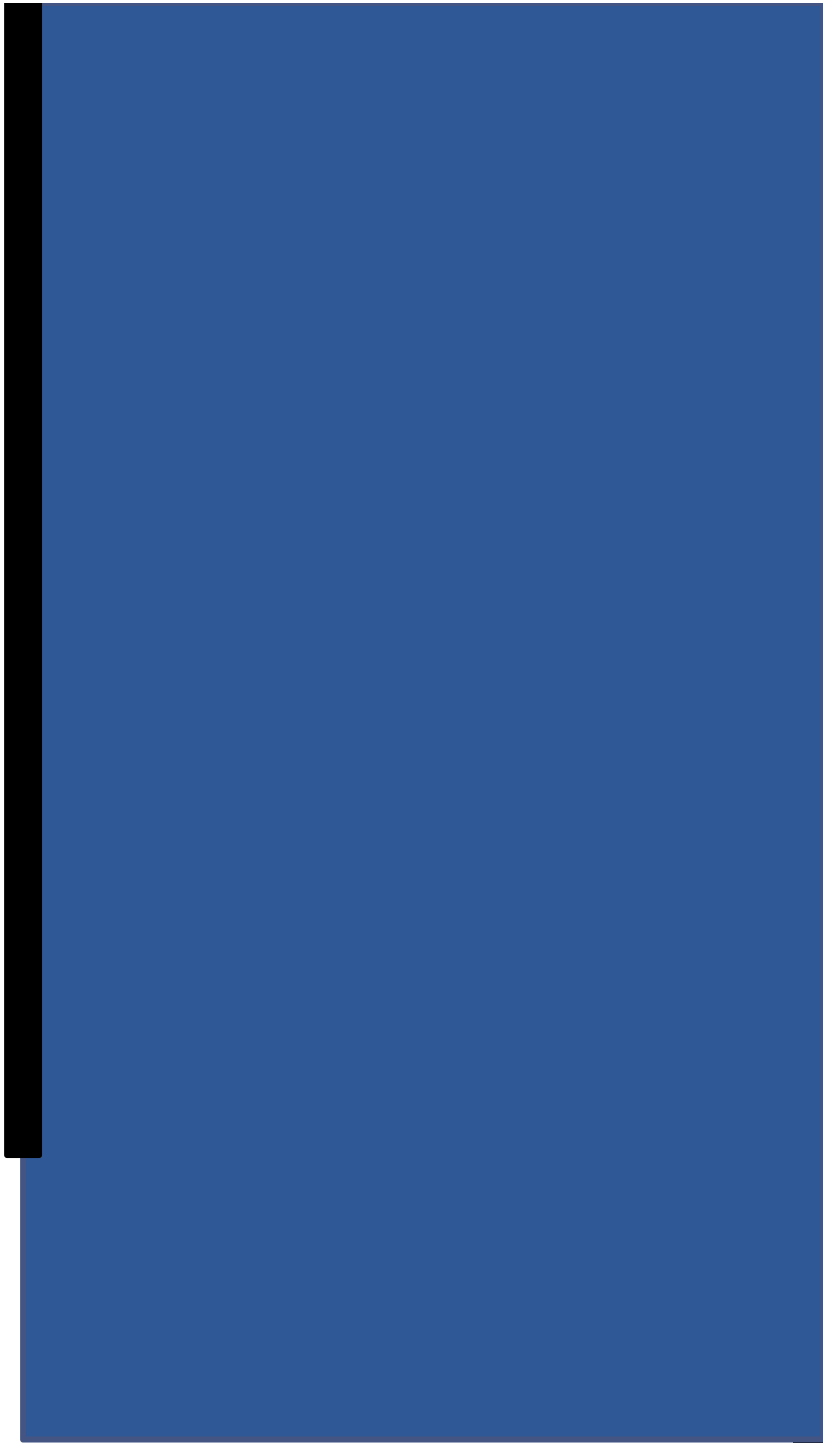
PREPARED BY:

RYHAN UDDIN

IMAGE SOURCE : NASA.GOV

OUTLINE

- **Introduction**
- **Challenges**
- **Primary objective**
- **Related Works**
- **Core Elements**
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- **Our Experiment**
- **Data preprocessing**
- **Evaluation**
- **Future Work**
- **Conclusion**

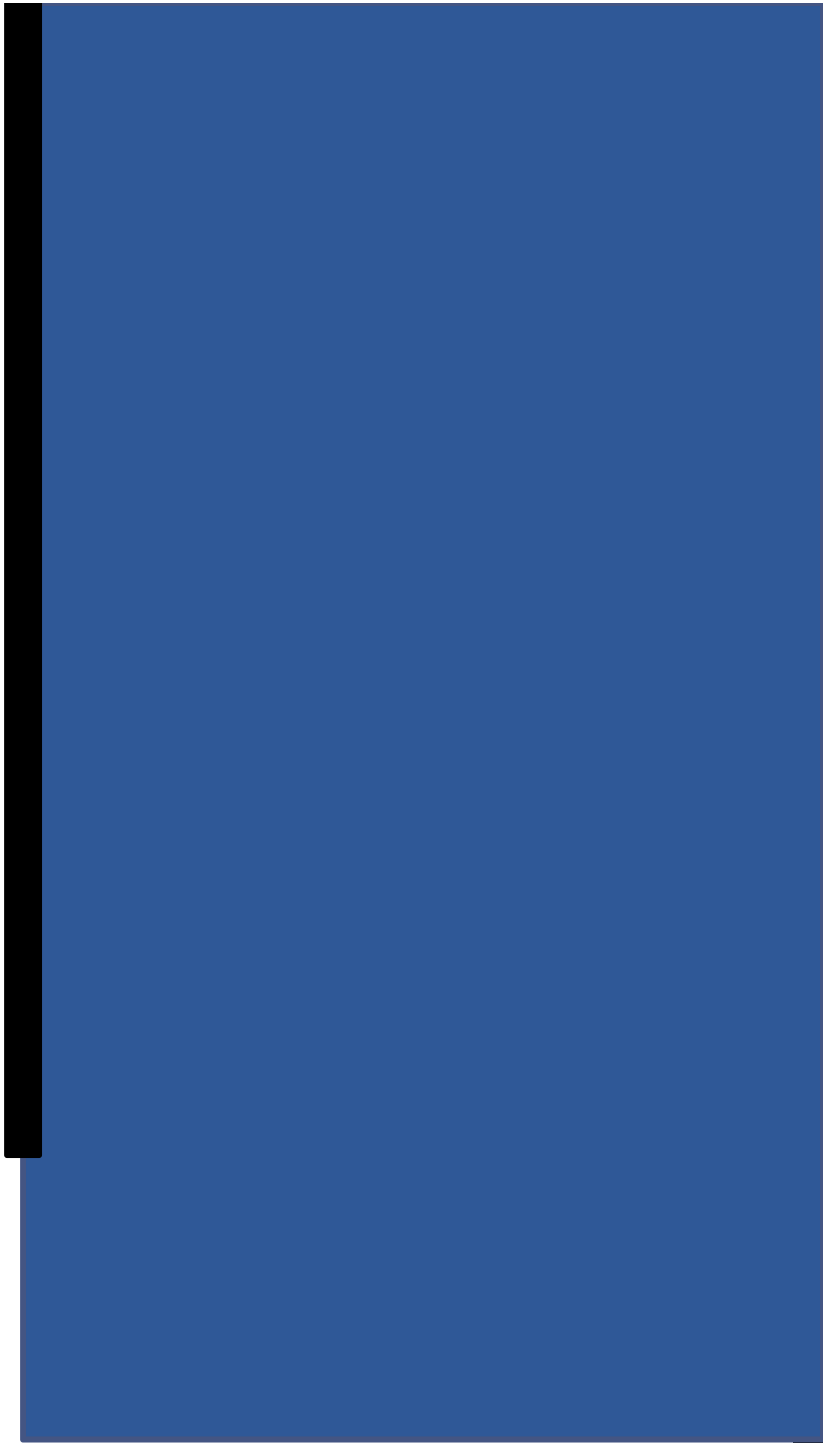


INTRODUCTION

- **Humans have always been fascinated with Space exploration**
- **It spawned massive ventures such as Apollo 11 (1969), International space station (1988), James Webb space telescope (2021)**
- **Recently NASA has made successful touchdown of Mars with Perseverance rover in the first quarter of 2021**
- **SpaceX‘s recent projects has re-ignited space exploration goals as its taking massive initiatives with the goal to colonize mars by 2050**
- **Numerous projects are now undergoing to establish interplanetary communication networks**

OUTLINE

- Introduction
- **Challenges**
- Primary objective
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- Evaluation
- Future Work
- Conclusion



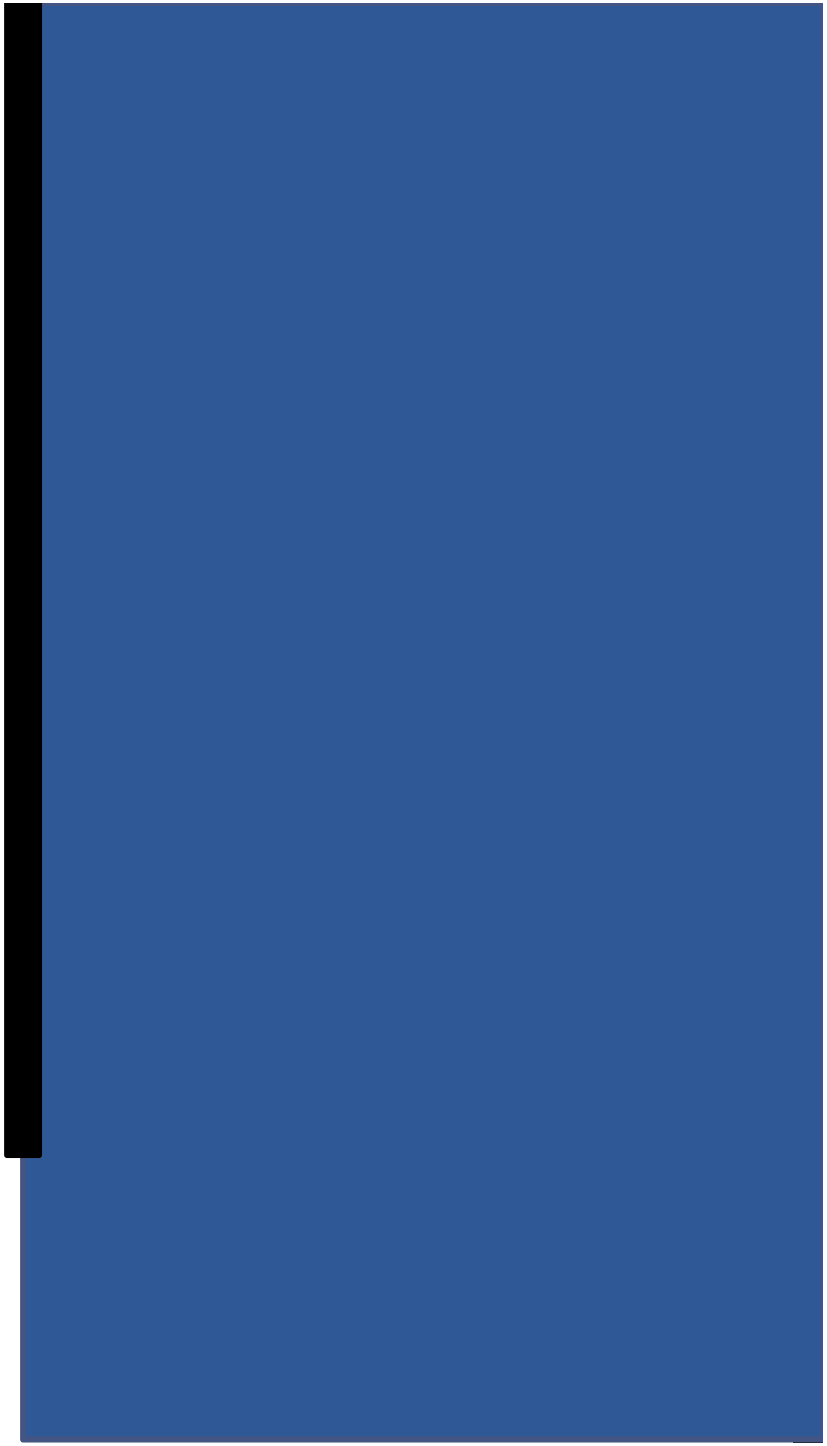
CHALLENGES

- Interplanetary communication networks are mostly based on traditional network backbones that are rigid, expensive and vertically integrated
- Systems are being targeted by ever increasing malicious threats that has seen an annual increase of 131% on online entities [1]
- Often these aggressors are masking their source IP or point of origin
- Centralized data aggregation is posing issues such as data privacy
- It is imperative to construct a strong security backbone to secure communication networks

[1] S. Cook, "20+ ddos attack statistics and facts for 2018-2023," Comparitech, 12-Feb-2023. [Online]. Available: https://www.comparitech.com/blog/information-security/ddos-statistics-facts/. [Accessed: 24-Mar-2023].

OUTLINE

- Introduction
- Challenges
- **Primary objective**
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- Evaluation
- Future Work
- Conclusion



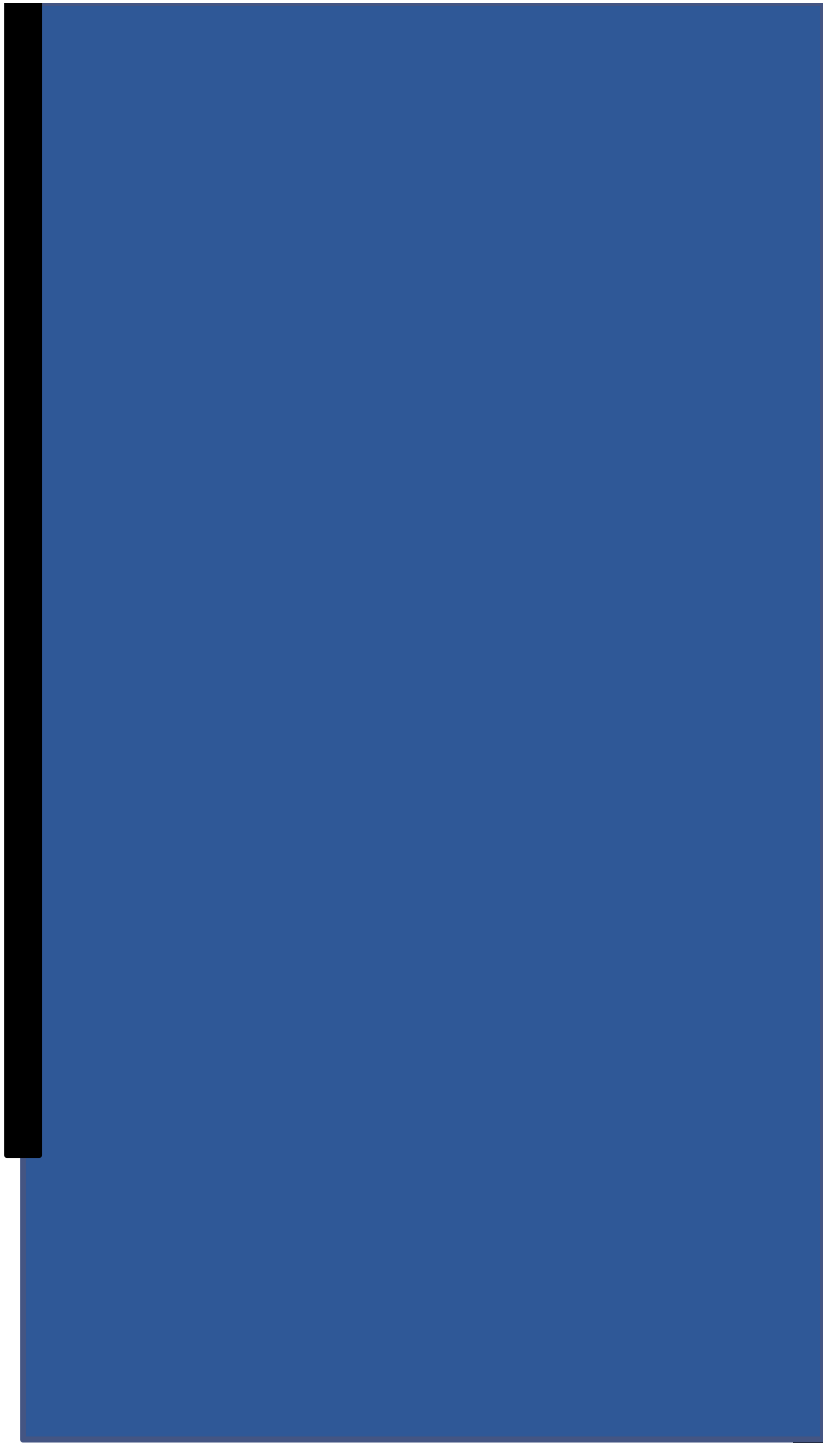
PRIMARY OBJECTIVE

Our framework intends to implement the following:

- **A Flexible and cost effective Framework (SDN)**
- **Intrusion detection through utilization of flow patterns (IDS)**
- **Preservation techniques to counter data breach (FL)**
- **Train model using satellite dataset (STIN)**

OUTLINE

- Introduction
- Challenges
- Primary objective
- **Related Works**
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- Evaluation
- Future Work
- Conclusion



RELATED WORKS

2020

K. Li, H. Zhou, Z. Tu, W. Wang, and H. Zhang, “Distributed network intrusion detection system in satellite-Terrestrial Integrated Networks using federated learning,” IEEE Access, vol. 8, pp. 214852–214865, 2020.

2022

Amarudin, R. Ferdiana, and Widyawan, “New approach of ensemble method to improve performance of ids using S-sdn classifier,” 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), 2022.

2023

N. Moustafa, I. A. Khan, M. Hassanin, D. Ormrod, D. Pi, I. Razzak, and J. Slay, “DFSat: Deep Federated Learning for Identifying Cyber Threats in IOT-based satellite networks,” IEEE Transactions on Industrial Informatics, pp. 1–8, 2023.

RELATED WORKS

- **Distributed network intrusion detection system in satellite-Terrestrial Integrated Networks using federated learning [2]**
- **Employed distributed IDS for satellite-Terrestrial networks using Convolutional Neural Network (CNN) along with FL**
- **Focus:**
  - Satellite-Terrestrial network
  - Horizontal FL method
  - Network topology optimization
  - Botnet, web-based and backdoor attacks and DDoS
- **Limitations:**
  - 12% packet loss on NIDS and data aggregation delay
  - Terrestrial network continued to occupy the interface to data monitoring while waiting for satellite node data

[2] K. Li, H. Zhou, Z. Tu, W. Wang, and H. Zhang, "Distributed network intrusion detection system in satellite-Terrestrial Integrated Networks using federated learning," IEEE Access, vol. 8, pp. 214852–214865, 2020.

RELATED WORKS (CONTD.)

- **New approach of ensemble method to improve performance of IDS using S-SDN classifier [3]**
- **False positive detection with S-SDN model based ensemble Learning using SVM and Naïve Bayes**
- **Focus:**
  - Ensemble classifier
  - UNSW-NB15 dataset
  - Fuzzers, exploits, backdoor, reconnaissance, DoS etc.
- **Limitations:**
  - More time consuming than single classifier
  - Centralized training therefore prone to data breach

[3] Amarudin, R. Ferdiana, and Widyawan, "New approach of ensemble method to improve performance of ids using S-sdn classifier," 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), 2022.

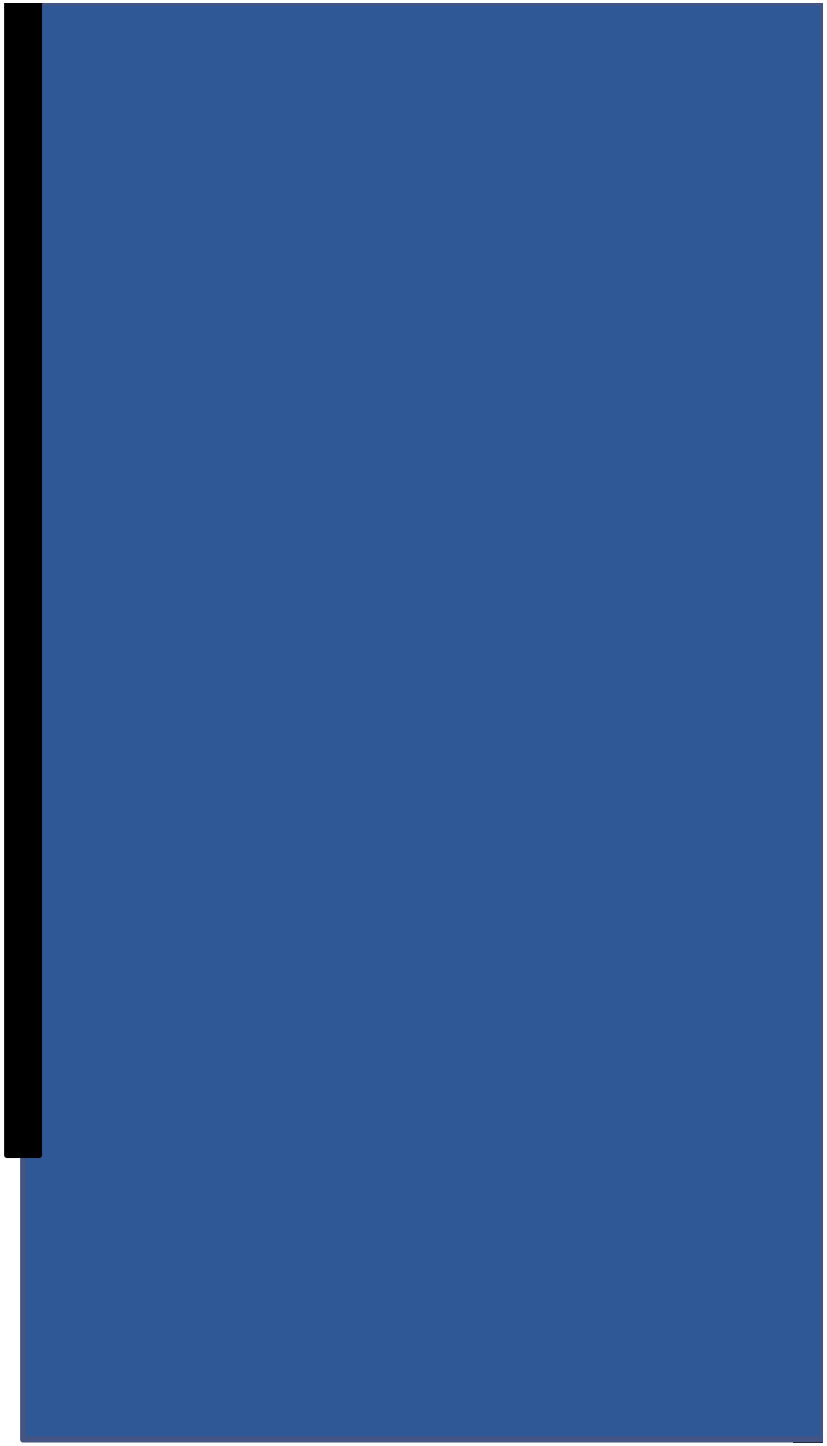
RELATED WORKS (CONTD.)

- **DFSat: Deep Federated Learning for Identifying Cyber Threats in IOT-based satellite networks [4]**
- **Deep learning framework in conjunction with a bi-directional LSTM**
- **Focus:**
  - Used FL to secure DFSat's parameter
  - ToN IoT and UNSW-NB 15 datasets
  - Reconnaissance, fuzzers, and DoS
- **Limitations:**
  - Uses a traditional network therefore dependent on proprietary hardware
  - Takes time for model training due to slow data aggregation

[4] N. Moustafa, I. A. Khan, M. Hassanin, D. Ormrod, D. Pi, I. Razzak, and J. Slay, "DFSat: Deep Federated Learning for Identifying Cyber Threats in IOT-based satellite networks," IEEE Transactions on Industrial Informatics, pp. 1–8, 2023.

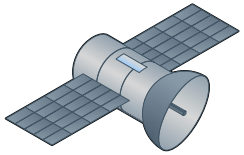
OUTLINE

- Introduction
- Challenges
- Primary objective
- Related Works
- **Core Elements**
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- Evaluation
- Future Work
- Conclusion



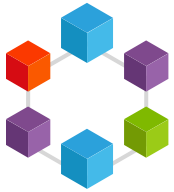
CORE ELEMENTS

Software Defined Network



Satellite dataset (STIN)

Intrusion Detection System



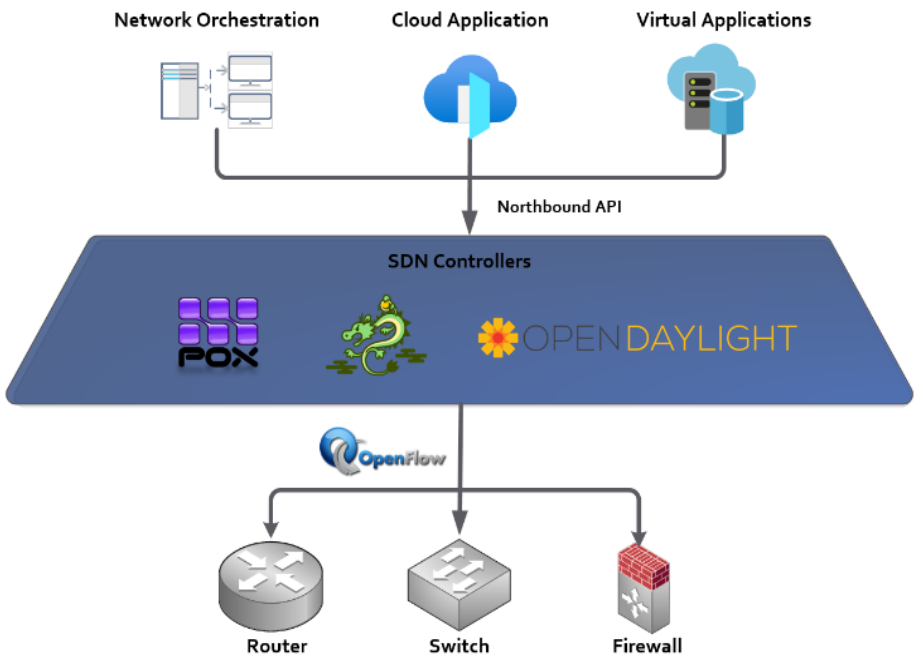
Federated Learning Technique

WHAT IS SDN

Despite the widespread adaptation of the traditional IP network, it is complex, Rigid and vertically integrated. That is where SDN comes into play.

SDN is the emerging paradigm that -

- **Separates Control and Data Layer**
- **Provides Programmability**
- **Gives Open Framework**
- **Scalable and cost effective**



SDN CONTROLLERS

The Programmable Abstraction of SDN is possible through the supervision of a controller that acts like a brain for the network.

STIN DATASET

- **Satellite-Terrestrial integrated network (STIN) combined a terrestrial dataset named TER20 with a satellite dataset named SAT20.**
- **These data sets were derived from 177,244 terrestrial network data and 132,320 satellite network data.**
- **Dimension: 82,320 rows and 31 columns.**
- **Contains critical features such as Flow duration, Flow numbers, Packet length, Packet bytes etc.**

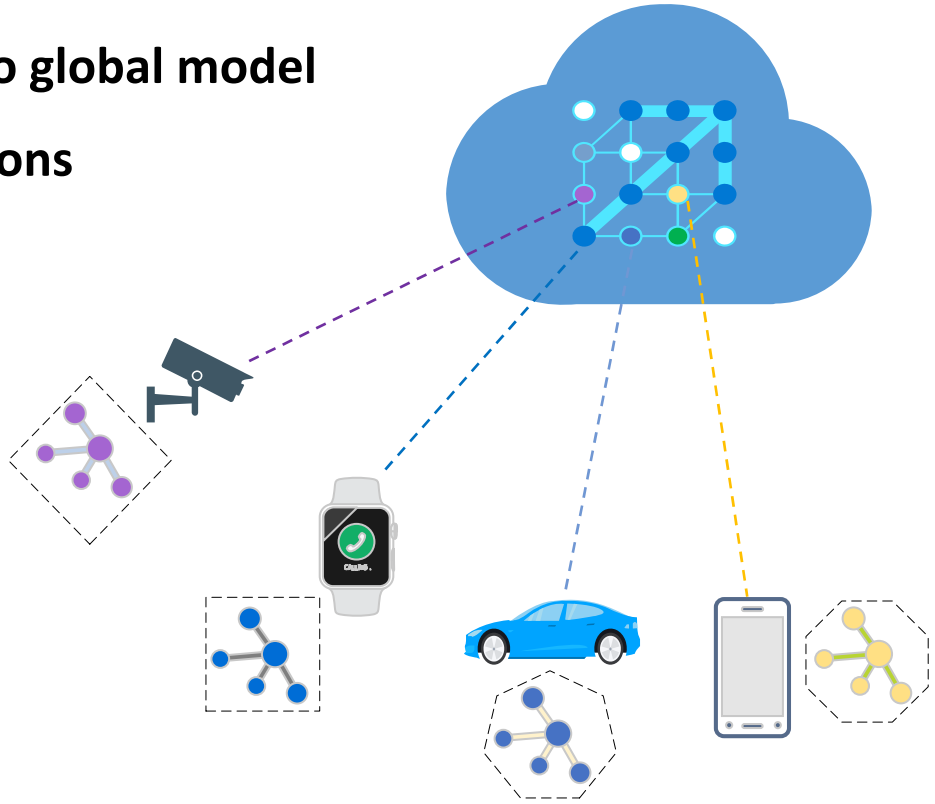
#	Name	Description
1	fl_dur	Flow duration
2	fw_pk	Total packets in the forward direction
3	l_fw_pkt	Total length of forward packets
4	l_bw_pkt	Total length of backward packets
5	pkt_len_min	Minimum length of a flow
6	pkt_len_max	Maximum length of a flow
7	pkt_len_std	Standard deviation length of a flow
8	fl_byt_s	Packet bytes transmitted per second
9	bw_iat_tot	Total time between of two backward packets
10	bw_iat_min	Minimum time between of two backward packets
11	fw_hdr_len	Number of bytes used in forward packet header
12	bw_pkt_s	Number of backward packets per second
13	syn_cnt	Number of packets with SYN
14	urg_cnt	Number of packets with URG
15	bw_win_byt	Number of backward bytes in the initial window

INTRUSION DETECTION SYSTEM

- The Ryu controller embodies the IDS itself
- Controller script deploys topology and instantiates IDS monitoring
- Extracts ongoing traffic from gateway switches
- Utilizes flow properties to identify traffic patterns
- Flags network traffic if there is a flood

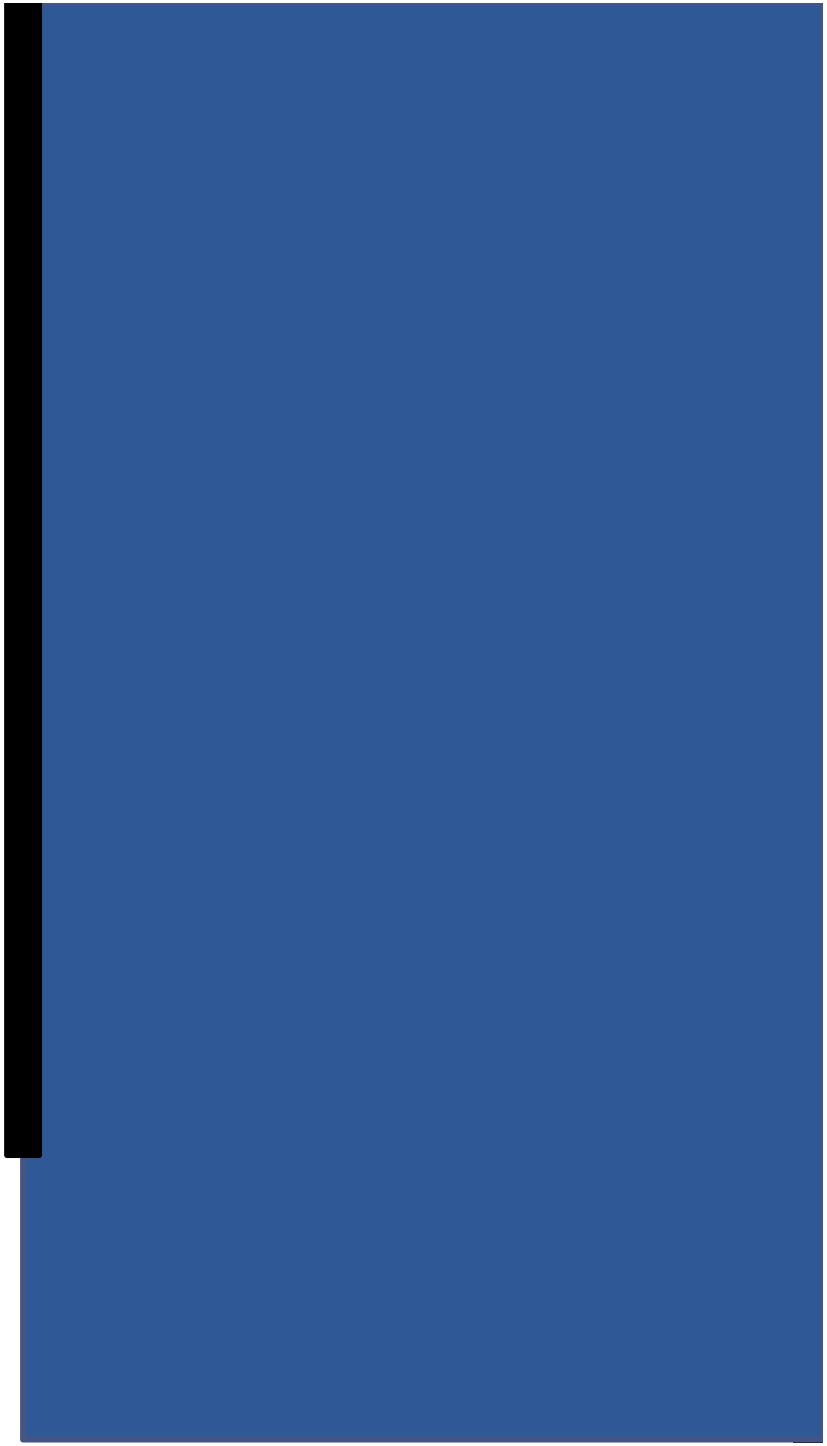
WHAT IS FEDERATED LEARNING

- Communication efficient learning of deep networks from decentralized data
- Models segmented into local model and global model
- Local models train locally learning from data on local devices
- Those Abridged models are sent to global model
- Global model then makes predictions

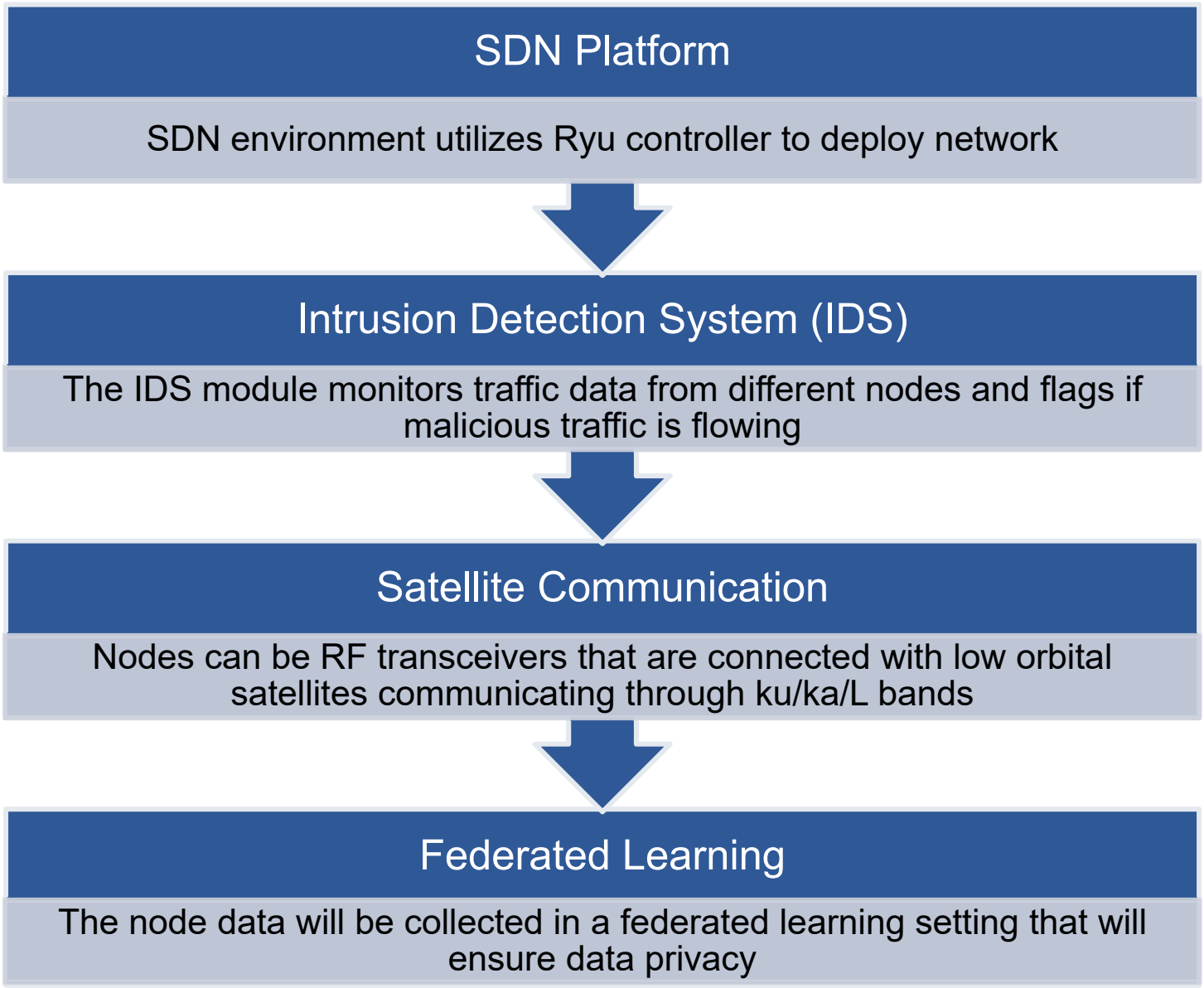


OUTLINE

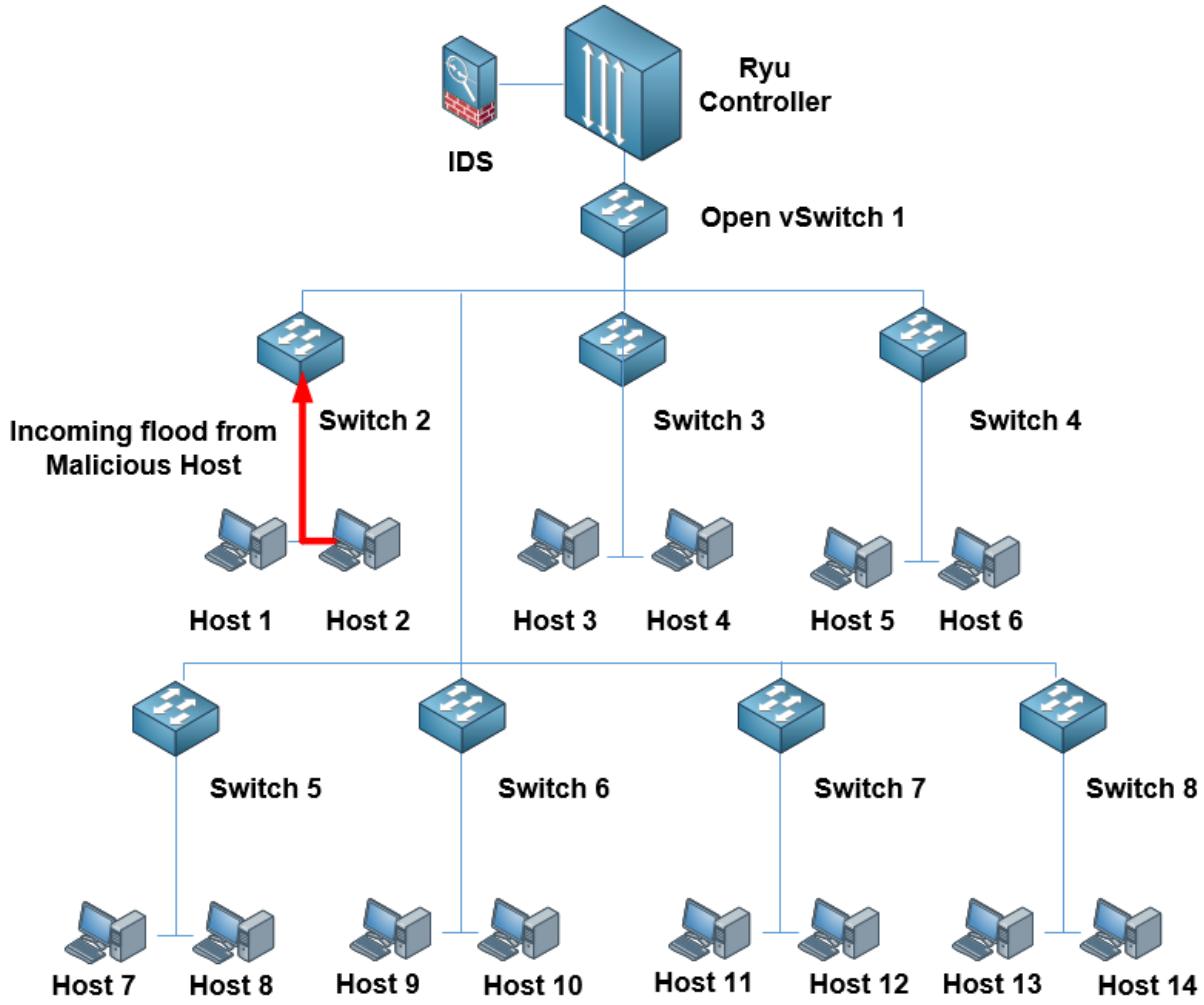
- Introduction
- Challenges
- Primary objective
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- **Our Experiment**
- Data preprocessing
- Evaluation
- Future Work
- Conclusion



WORK FLOW

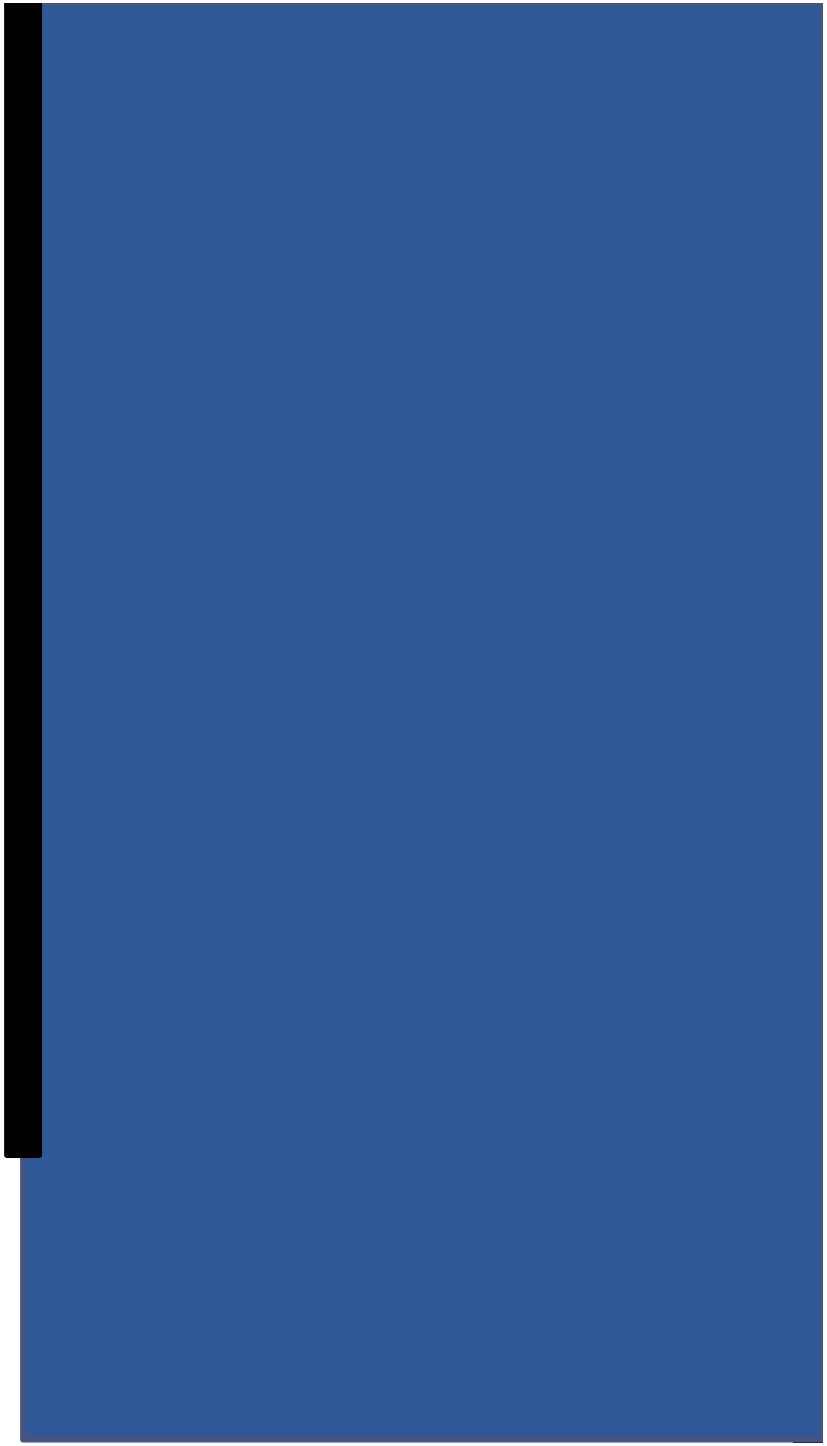

SDN Platform
SDN environment utilizes Ryu controller to deploy network
Intrusion Detection System (IDS)
The IDS module monitors traffic data from different nodes and flags if malicious traffic is flowing
Satellite Communication
Nodes can be RF transceivers that are connected with low orbital satellites communicating through ku/ka/L bands
Federated Learning
The node data will be collected in a federated learning setting that will ensure data privacy

NETWORK TOPOLOGY



OUTLINE

- Introduction
- Challenges
- Primary objective
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- **Data preprocessing**
- Evaluation
- Future Work
- Conclusion



DATA PREPROCESSING

Standard deviation, σ

$$= \sqrt{\frac{\sum_{i=1}^{n}(P_n - \overline{P})^2}{n}}$$

For all flows (in T period):

P_n **= Number of nth packets**

$\overline{P}$ **= Mean of total packets**

n = number of flow entries

Standard deviation, σ

$$= \sqrt{\frac{\sum_{i=1}^{n}(B_n - \overline{B})^2}{n}}$$

For all flows (in T period):

B_n **= Number of nth bytes**

$\overline{B}$ **= Mean of total bytes**

n = number of flow entries

FLOW UTILIZATION

- Previously we have used source and destination IP for the regulator where single IP flows are appended as intermediate flows.
- For IDS we have utilized the flow properties such as flow count, flow size etc.
- We count the standard deviation of flows from individual switches, $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(F_n - \overline{F})^2}{n}}$
- Each switches have individual flow numbers, for example:

  S1 flow numbers are F1 = 3

  S2 flow numbers are F2 = 3

  F3 = 2

  F4 = 2

  F5 = 3

  F6 = 3

  F7 = 2

  F8 = 3

Total flow count = 21

Mean flow $\overline{F}$ = 21 / 8 = 2.625, Since, n = 8

FLOW UTILIZATION (CONTD.)

- From here we count the standard deviation of the flow for the network, for this example: it is 0.4845
- Therefore, the average flow of the network (about 68% of the flows according to normal distribution) should be about 2.625 ± 0.4845.
- However, in case of a flood this number goes very high. We consider a scenario where the flow value from the switch is a factor of 10.
- A switch with 30 flows yields an average flow of 2.625 ± 9.689 which is about the triple of a normal flow scenario which clearly declares the network as under attack state.

FEDERATED LEARNING PROCESS

Node Side:

- **After OpenMined framework is initialized the nodes are connected using AWS OpenGrid network that uses public IPs**
- **For our experiment, we have used lookback IP for the pairing using PSK**
- **SAT20 data frame was imported with some preliminary pre-processing**
- **The data frame is converted into Torch Tensor objects**
- **Those Objects are stored in the OpenMined pandas data store**
- **Objects are split into data and target pointers with searchable attribute**
- **Tags are added for the objects with contents descriptions**
- **Request handler action is set as “accept” for authorization.**

FEDERATED LEARNING PROCESS (CONTD.)

Controller Side:

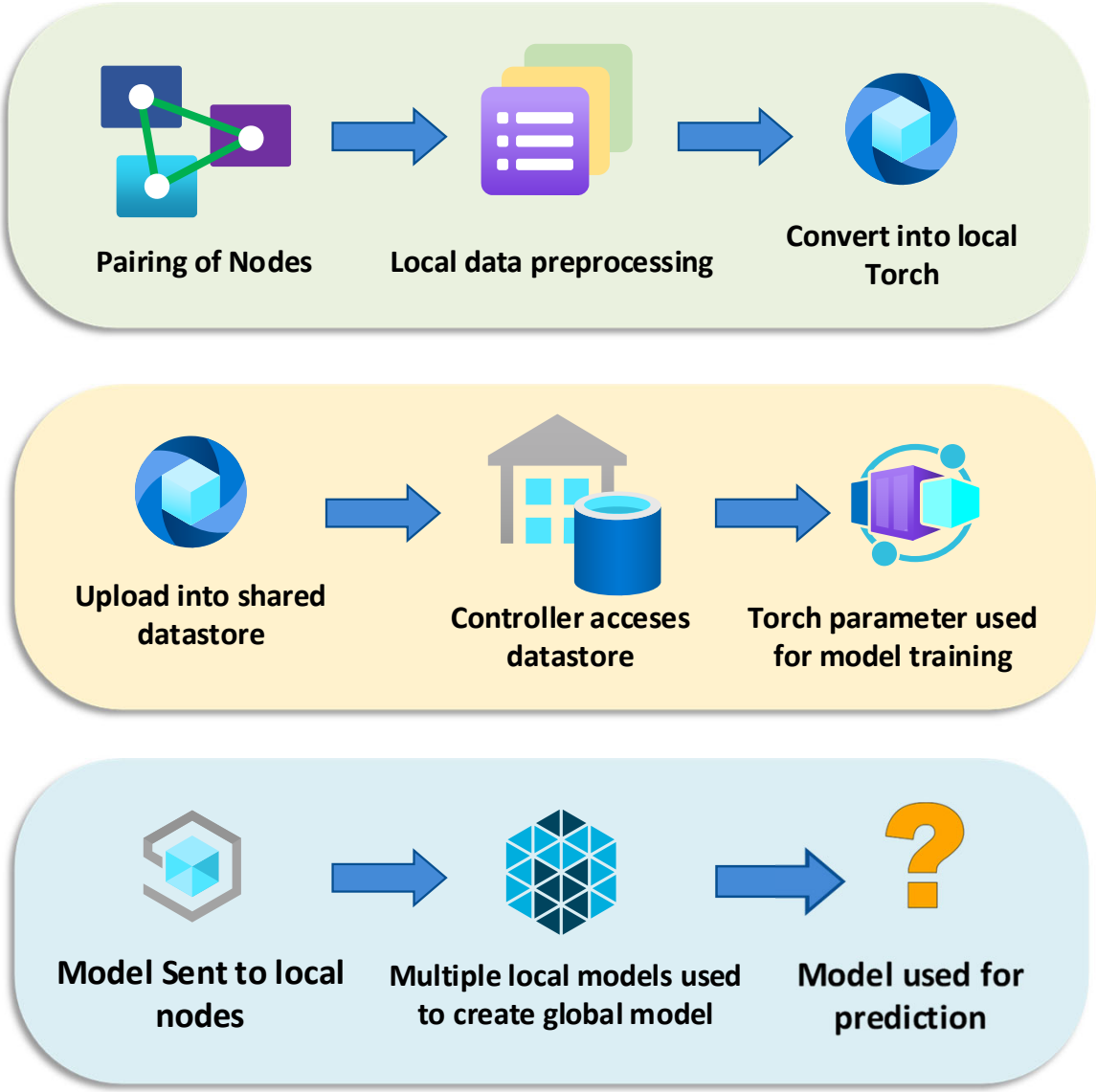
- **Pandas data store is accessed that contains the uploaded node objects (only the data pointers) with included tags**
- **Now we design our DL architecture with 31 input dimension and 82,320 samples**
- **we have used two hidden layers one having 20 neurons and the other with 30**
- **For our forward pass we have used relu (layer 1) and for the output layer we have used log_softmax activation function**
- **After this we create the local models, which are then sent to each nodes as remote models**
- **We have set remote model parameters which has individual list pointers**
- **Then we used Adam optimizer for the optimization of the remote models**

FEDERATED LEARNING PROCESS (CONTD.)

Controller Side:

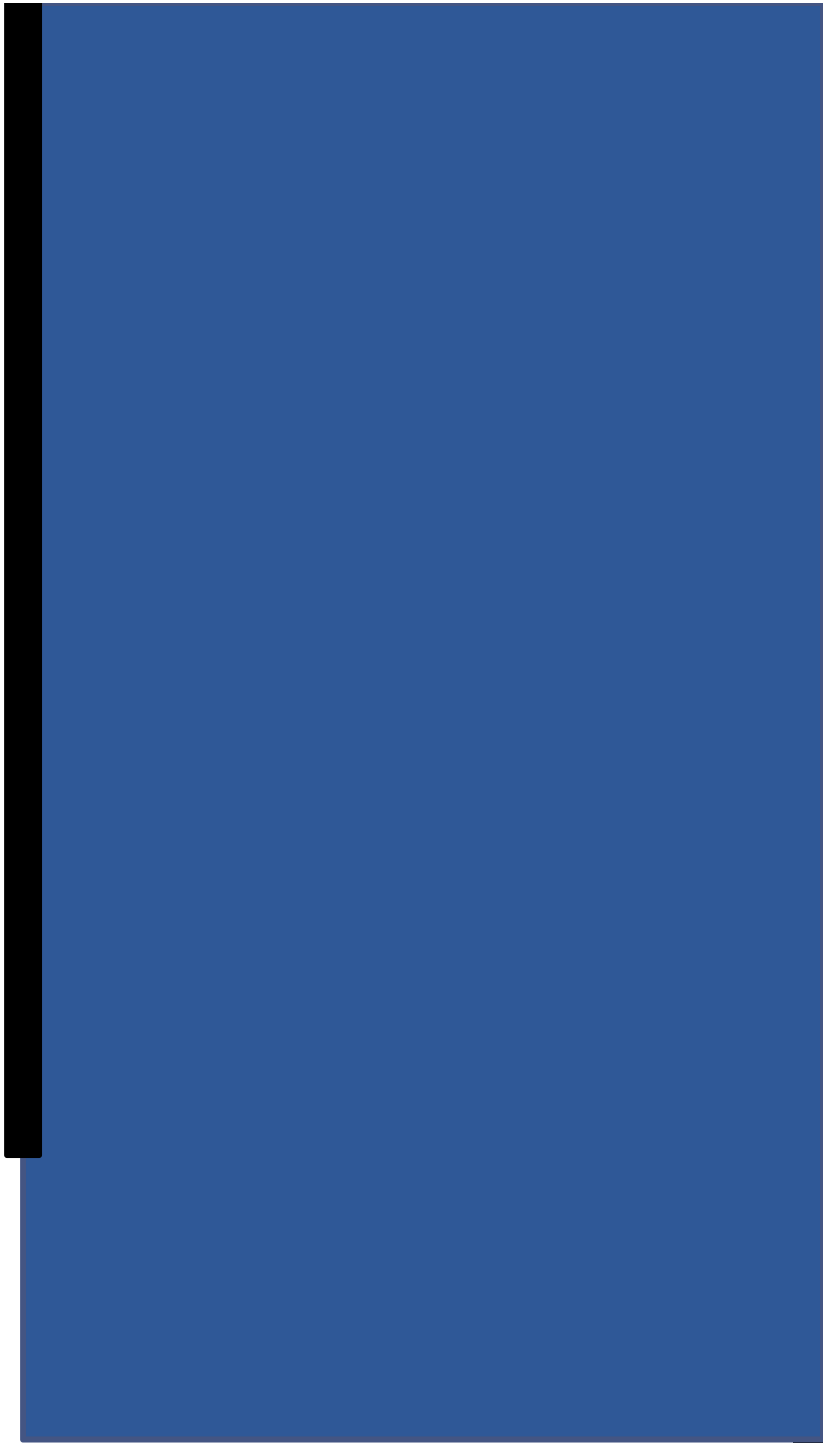
- **Afterwards we trained our loss function which includes parameters such as the remote model, optim, data pointer, target pointer etc.**
- **For our loss function we have used negative log likelihood loss (nll_loss)**
- **Then we track the training losses in each stages (for 100 epochs)**
- **Then we train the remote model which we recall by using get model function**
- **This gets the trained remote model and all the associated layers that we have designed (in this case 2 layers)**
- **Now that we have our model ready we can test it with our test dataset**
- **We import the SAT20 test dataset and split into X and y**
- **We then test the data and compare it to the ground truth to check how accurately is it predicting the outputs**

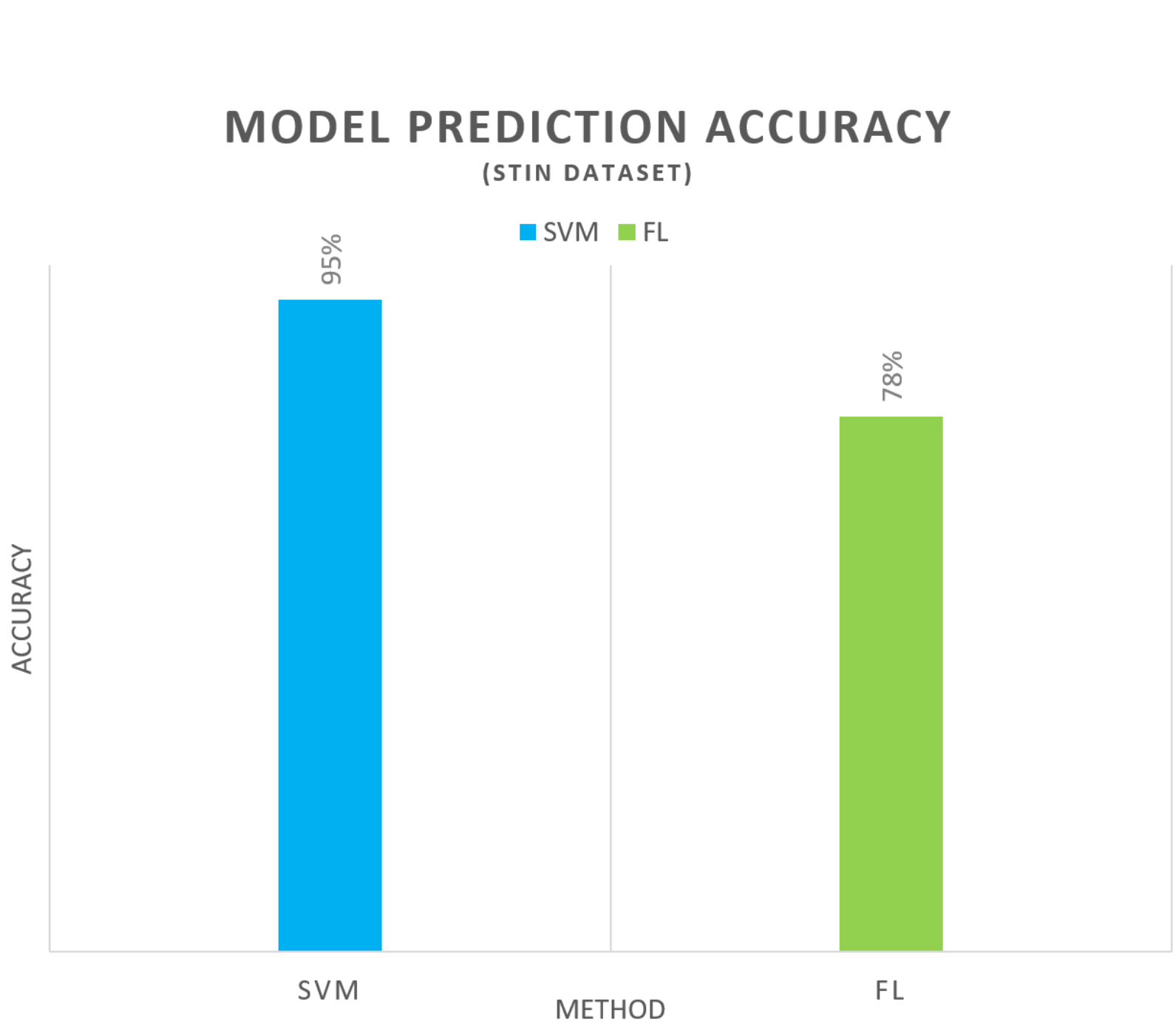
FEDERATED LEARNING WORK FLOW



OUTLINE

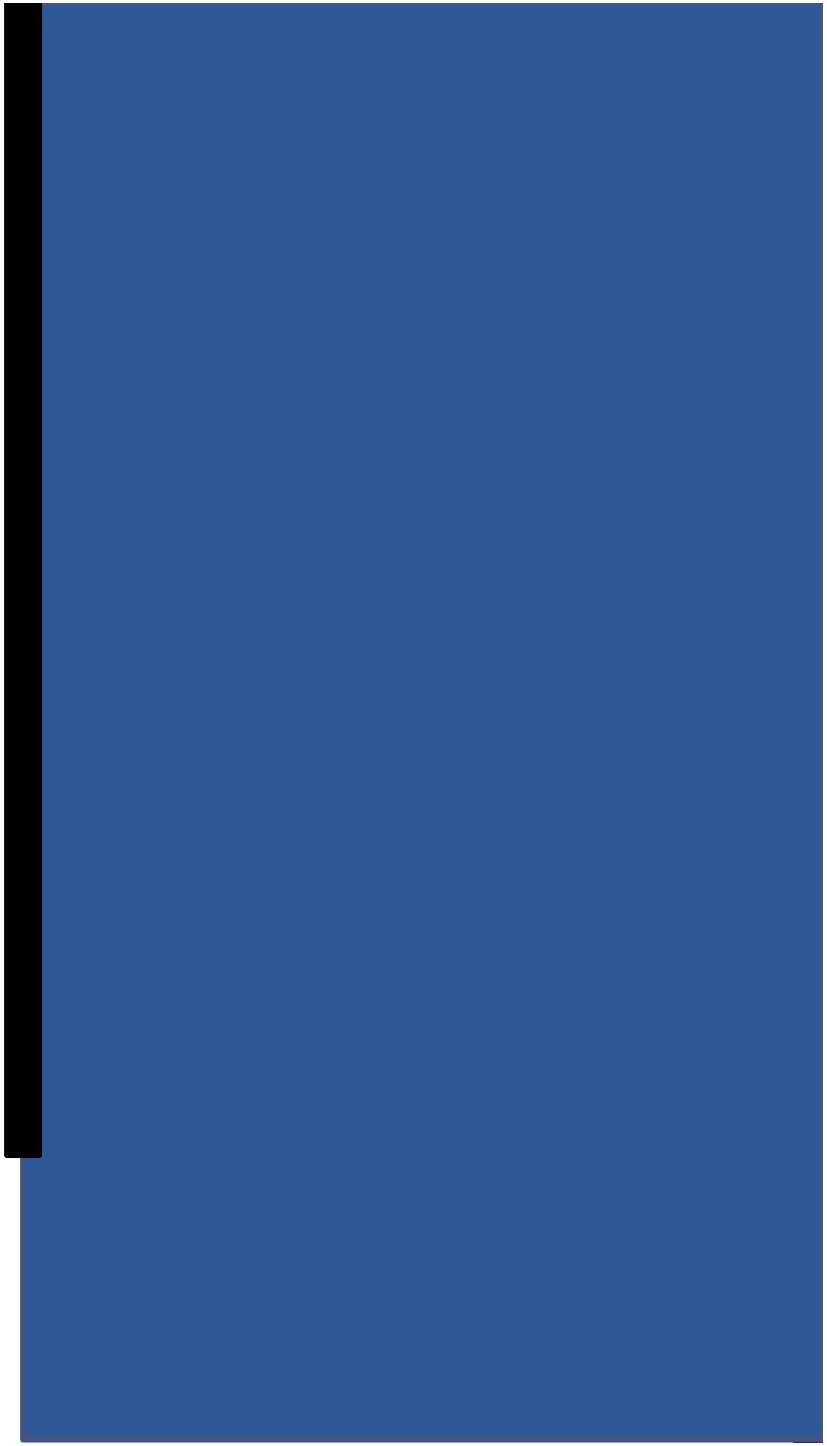
- Introduction
- Challenges
- Primary objective
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- **Evaluation**
- Future Work
- Conclusion




MODEL PREDICTION ACCURACY
(STIN DATASET)
SVM
FL
95%
78%
ACCURACY
SVM
FL
METHOD

OUTLINE

- Introduction
- Challenges
- Primary objective
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- Evaluation
- **Future Work**
- Conclusion

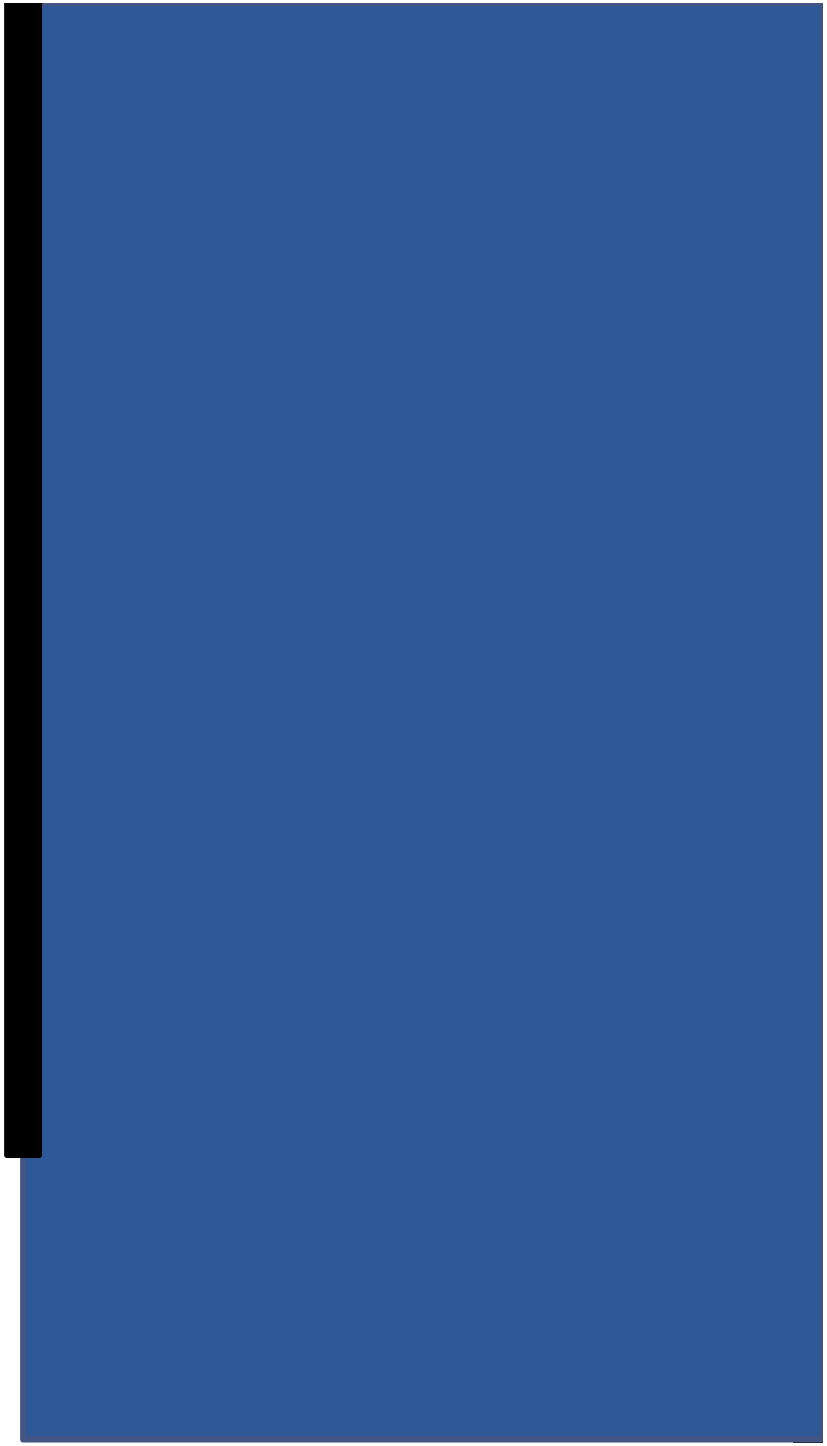


FUTURE WORK

- System adaptation of unlabeled data associated with real systems
- Various FL techniques can be combined (FedSGD, FedDANE, FedPro, etc.)
- Testing with different SDN controllers (POX, Floodlight, ODL, etc.)
- Incorporate TensorFlow based FL technique
- Explore other types of intrusions

OUTLINE

- Introduction
- Challenges
- Primary objective
- Related Works
- Core Elements
  - SDN
  - Satellite dataset
  - IDS
  - Federated Learning
- Our Experiment
- Data preprocessing
- Evaluation
- Future Work
- **Conclusion**



CONCLUSION

- It is imperative to develop adaptive interplanetary communication infrastructure by ensuring security and data privacy
- SDN can offer the flexibility and scalability to the network infrastructure
- An IDS can be used to identify intrusive traffic utilizing traffic patterns
- FL offers an added layer of data privacy by avoiding direct data sharing between neighboring nodes
- However, this data privacy is availed at the cost of model training duration and additional training pre-processes

REFERENCES

1. S. Cook, “20+ ddos attack statistics and facts for 2018-2023,” Comparitech, 12-Feb-2023. [Online]. Available: https://www.comparitech.com/blog/information-security/ddos-statistics-facts/. [Accessed: 24-Mar-2023].
2. K. Li, H. Zhou, Z. Tu, W. Wang, and H. Zhang, “Distributed network intrusion detection system in satellite-Terrestrial Integrated Networks using federated learning,” IEEE Access, vol. 8, pp. 214852–214865, 2020.
3. Amarudin, R. Ferdiana, and Widyawan, “New approach of ensemble method to improve performance of ids using S-sdn classifier,” 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), 2022.
4. “Federated learning,” OpenMined Blog. [Online]. Available: https://blog.openmined.org/tag/federated-learning/. [Accessed: 01-Apr-2023].
5. Snort Intergration — Ryu 4.34 documentation, Ryu.readthedocs.io, https://ryu.readthedocs.io/en/latest/snort_integrate.html. [Accessed: 17-Apr-2023].

THANK YOU