DelphiFL: An Investigation into a More Robust Federated Learning Model Using Method Chaining and Zero-Trust

PRESENTERS: **Jonathan** Flores (BSU) and **Hailey** Whipple (Utah Tech University)

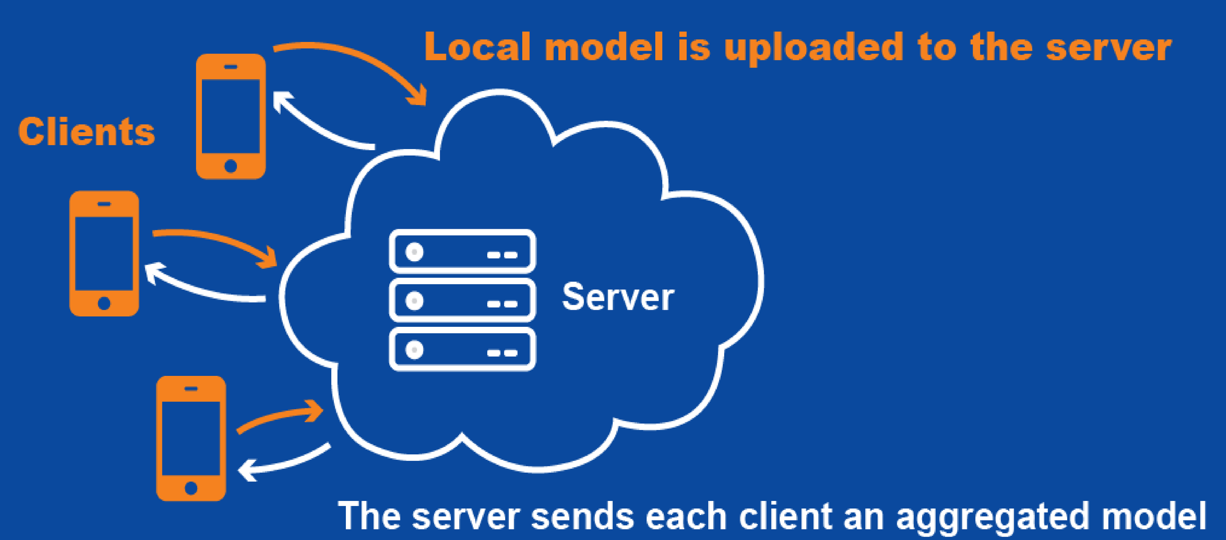
BOISE STATE UNIVERSITY

HOW TO DEFEND MACHINE LEARNING?

Background:

Federated Learning (FL)

- A distributed machine learning paradigm



Purpose

- Improve the security and privacy of sending data over a network

Obstacles

- FL is reliant on user contributions
- Poisoning attacks "poison" or change the raw data before the local model trains on it

DelphiFL:

DelphiFL is an aggregation method which relies on a process known as method chaining.

Method Chaining

- The process of taking existing security measures and chaining them together to yield a more robust result

Structure

- Bulyan defense using Krum and Trimmed Mean
  - Krum: Select a representative distribution using mean and error
  - Trimmed Mean: Cut top and bottom outliers out

DelphiFL V1

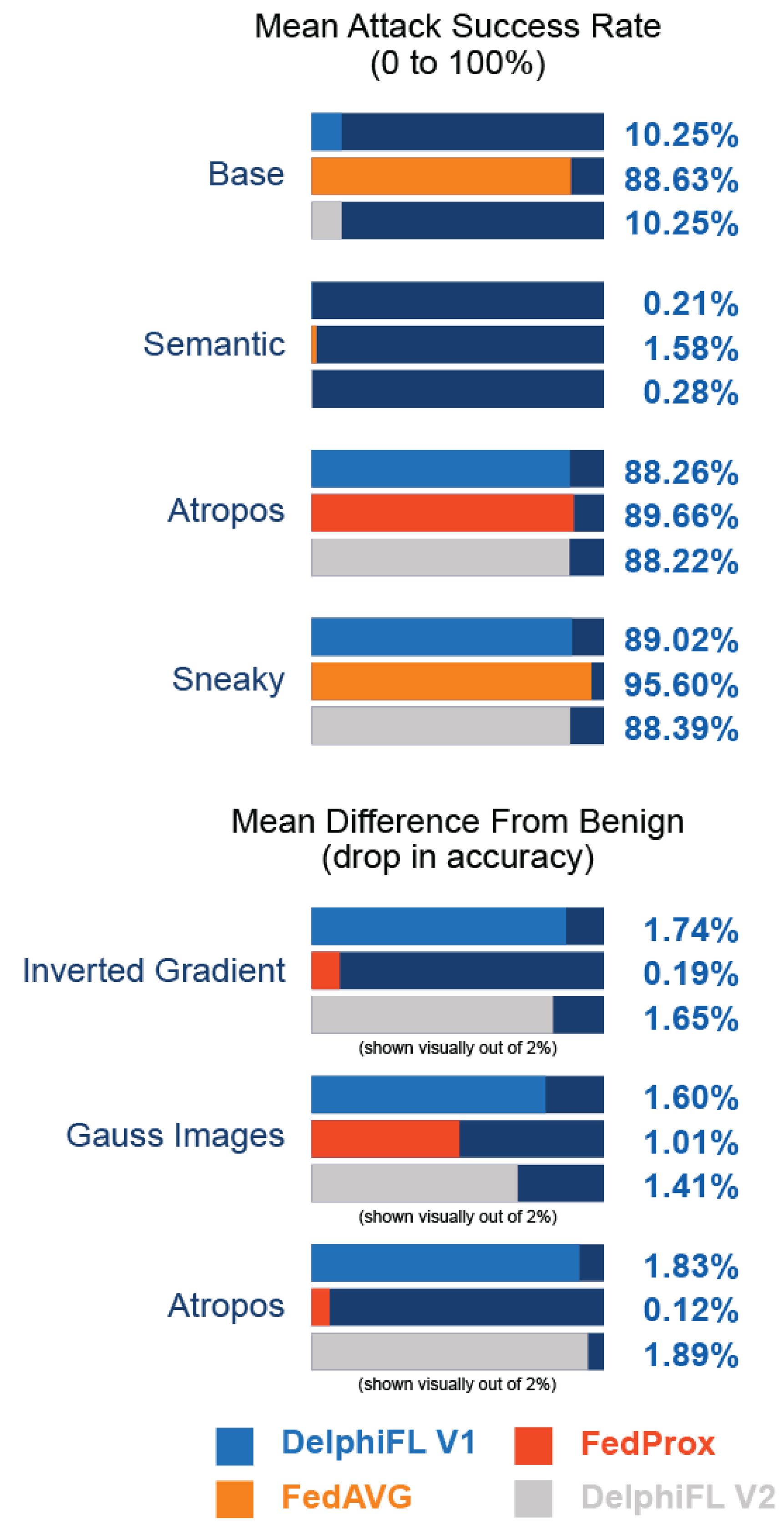
- Incorporates zero-trust policy

DelphiFL V2

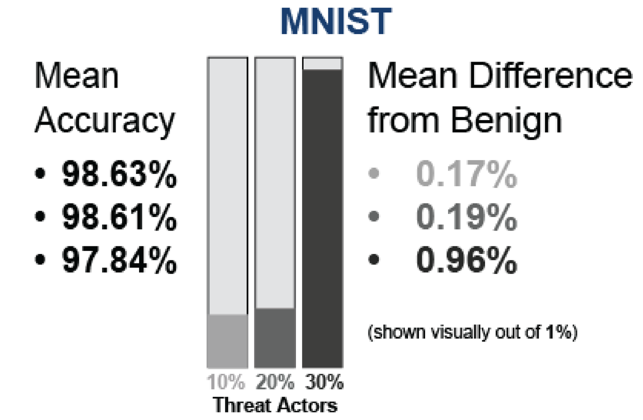
- An attempt to further combat a specific backdoor attack (Sneaky Backdoor) using Robust Learning Rate (RLR) methods, which alter the learning rate of the optimizer with respect to the uniformity of the data.

Data:

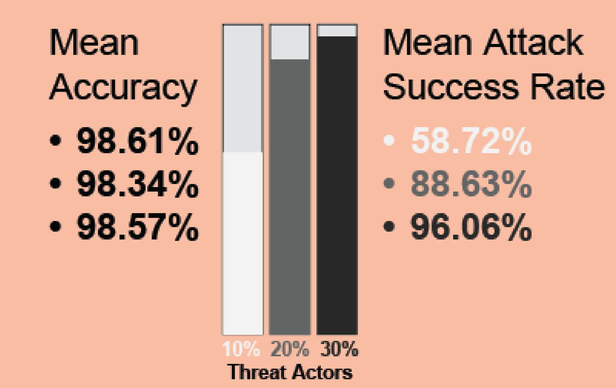
MNIST Dataset with 2 Threat Actors



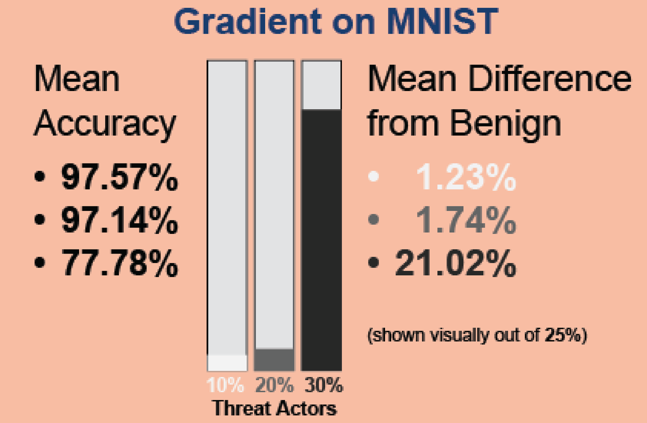
FedProx with Inverted Gradient on MNIST



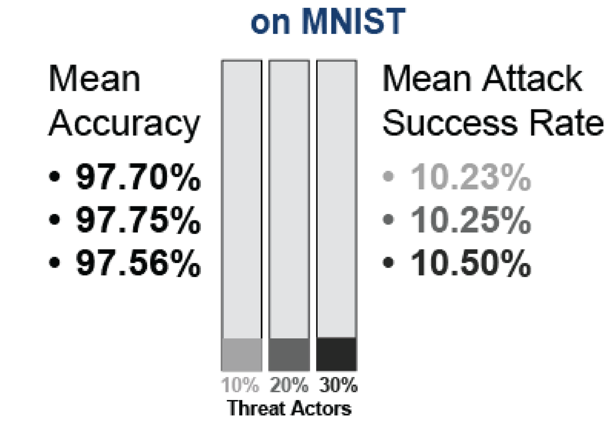
FedAVG with Base Backdoor on MNIST



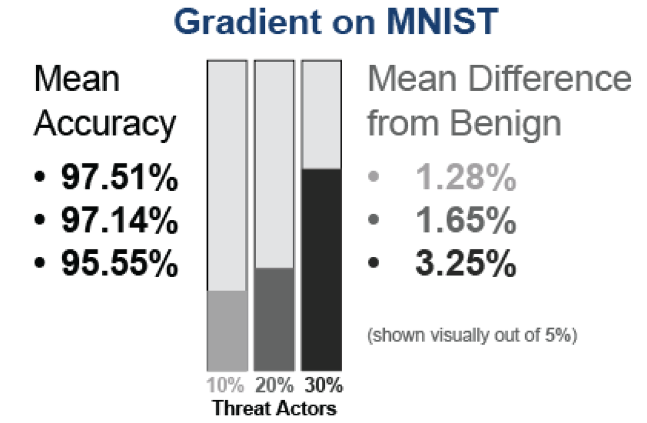
DelphiFL V1 with Inverted Gradient on MNIST



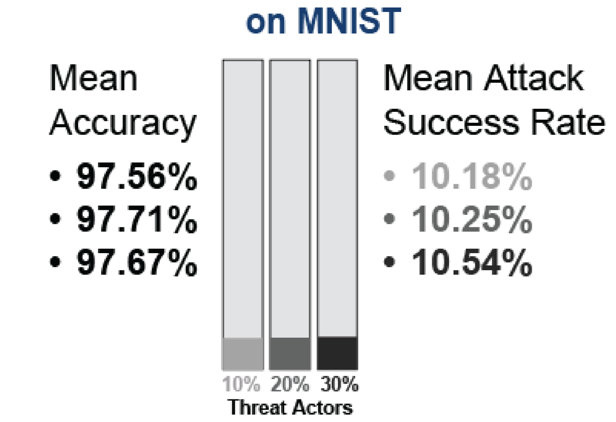
DelphiFL V1 with Base Backdoor on MNIST



DelphiFL V2 with Inverted Gradient on MNIST



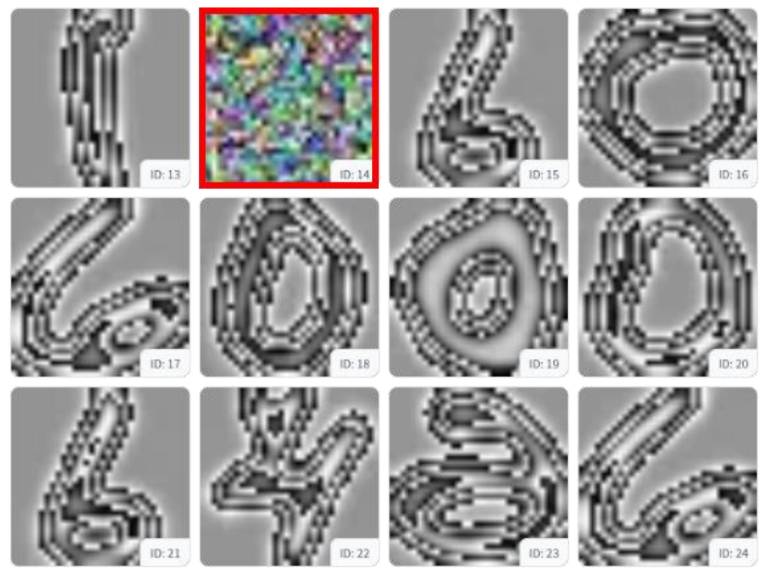
DelphiFL V2 with Base Backdoor on MNIST



MNIST Dataset with Base Backdoor

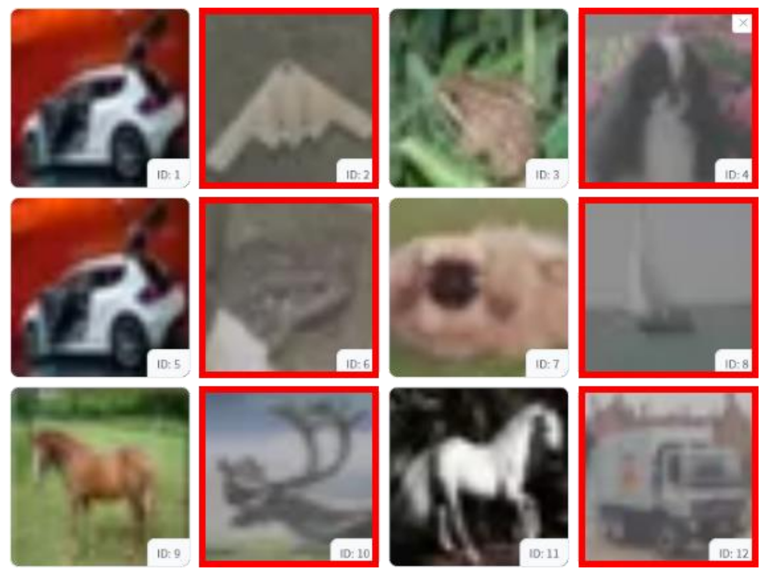
ID	Label
1	4
2	2
3	4
4	2
5	4
6	2
7	1
8	2
9	4
10	2
11	0
12	2

USPS Dataset with Sneaky Backdoor



ID	Label
13	1
14	2
15	6
16	0
17	6
18	0
19	0
20	0
21	6
22	4
23	2
24	6

CIFAR10 Dataset with Gauss Images



ID	Label
1	Automobile
2	Dog
3	Frog
4	Cat
5	Automobile
6	Truck
7	Dog
8	Cat
9	Horse
10	Automobile
11	Horse
12	Frog

If you're interested in seeing the simulation and the work we have done, scan the QR Code for more results.

Methods:

We used a program that was created by the MARS group at Wuhan University meant to simulate a FL scenario (1). Using their simulation for our research, we created our own method with the following variables:

- **Control Variables**
  - 100 rounds of training and testing
  - Random seed set to 123
- **Independent Variables**
  - **Datasets: MNIST** (Training Size: 60,000, Testing Size: 10,000), **USPS** (7,291, 2,007), **CIFAR10** (50,000, 10,000)
  - **Number of Threat Actors:** 10%, 20%, 30%
  - **Methods: DelphiFL V1**, **DelphiFL V2**, **FedAVG** (Using means to combine gradients of uniform size), **FedProx** (FedAVG for datasets of variable size)
  - **Total Clients:** 10
- **Dependent Variables**
  - Mean Attack Success Rate
  - Mean Difference From Benign

Many attacks were also tested, including those developed by our partner group (Red Team) in an adversarial approach.

Conclusion:

DelphiFLV1 is a more robust, "unified federation method" than existing methods, and V2 outperformed V1 in certain scenarios. For future work, we hope to create our own local method for V3 and include internet packet data (PCAP) into the zero-trust policy.

Cloud Computing Security and Privacy REU

Erin Kendall (Transylvania University), Adam Crayton (BSU), Hao Chen (Ph.D., BSU), and Zavareh Bozorgasl (BSU)

References:

1. Huang, W., Ye, M., Shi, Z., Wan, G., Li, H., Du, B., & Yang, Q. (2024). Federated learning for generalization, robustness, fairness: A survey and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–20. https://doi.org/10.1109/tpami.2024.3418862
2. Ponte, A., Trizna, D., Demetrio, L., Biggio, B., Ogbu, I. T., & Roli, F. (2024, May 23). SLIFER: Investigating performance and robustness of malware detection pipelines. *arXiv.org*. https://arxiv.org/abs/2405.14478

Supported by National Science Foundation Award #2244596