Node to node communication in Vantage6

Renee Witsenburg
A data scientist has (a) research question(s) that needs data from multiple data sources

Copying the datasets and centralizing them

Disadvantages:

- Transport
- Loss of control
- Privacy
- Security
- Law
Personal Health Train

- Patient data (possibly enriched with personal health data) stored in data stations
- Data stations can be clustered together in larger data stations
- Algorithms/trains are moved to the data stations by researchers

- Data owners have access to their own data and can set access permissions
“Coronary artery disease: risk estimations and interventions for prevention and Early detection – a personal health train project”

- Carrier project
Vantage6

Infrastructure to do federated learning

- Vantage6 server
- Vantage6 nodes
- Collaboration
- Organisation
- Registries

Source: https://vantage6.ai/blog-index/about-secure-multi-party-computation/
Research questions

How can Vantage6 edge nodes work together efficiently, without the interposition of the central server?

- What are the issues in the current Vantage6 project that create a bottleneck between working nodes?
- Which infrastructures could be implemented to make the nodes work together?
- Is it possible to tamper with, read or intercept data or the model?
- How are malicious attempts to corrupt the data or the model detected?
Test setup Vantage6

- Orleans (145.100.104.120)
  - Guest-01
    - Vantage6 Server
      - (145.100.111.50)
    - Docker Host
    - server_vserver_1
  - Guest-02
    - Vantage6 Node
      - (145.100.111.51)
    - Docker Host
    - vantage6-dummy-user
  - Guest-03
    - Vantage6 Node
      - (145.100.111.52)
    - Docker Host
    - vantage6-dummy-user
  - Guest-04
    - Vantage6 Node
      - (145.100.111.53)
    - Docker Host
    - vantage6-dummy-user
    - algorithm
Network traffic inside the Vantage6 infrastructure
Nodes working together

Secure sum

dk-Secure sum
Node to node communication

- Node to node communication with the Vantage6 server in the middle
- Node to node communication without the Vantage6 server in the middle
  - Peer-to-peer
  - VPN
  - Overlay
Proof of Concept (Overlay)

Docker Swarm: Vantage6 server is docker swarm manager

Create overlay network on the Vantage6 server and share the swarm IP + token with the Vantage6 nodes in the collaboration

Vantage6 nodes in de collaboration connect to the swarm

Algorithm containers connect to the overlay network
Proof of concept
Proof of concept

“All swarm service management traffic is encrypted by default, using the AES algorithm in GCM mode. Manager nodes in the swarm rotate the key used to encrypt gossip data every 12 hours.”

“To encrypt application data as well, add --opt encrypted when creating the overlay network. This enables IPSEC encryption at the level of the vxlan.”

With overlay encryption enabled, Docker creates IPSEC tunnels between nodes. Keys rotate every 12 hours.
Discussion

Only researched overlay networks

Main focus on node to node communication and not security

All nodes in the infrastructure need to be in the swarm to connect to the overlay network

Setup can be recreated without a swarm but with a key/value container
Conclusion

Nodes can interact with each other through an overlay network

Vantage6 still has a lot of bugs and missing features
Future work

Other federated learning frameworks like PyGrid an PySyft

Other node to node communication performance

Developing Vantage6 and Docker-py further
Questions?