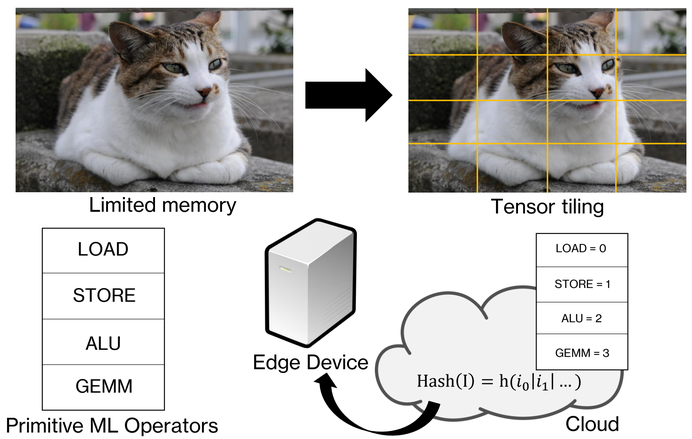
TrustZone as a Secure Tensor Processor

PURDUE UNIVERSITY
Electrical and Computer Engineering

Heejin Park, Noah Curran, Felix Xiaozhu Lin
Purdue University

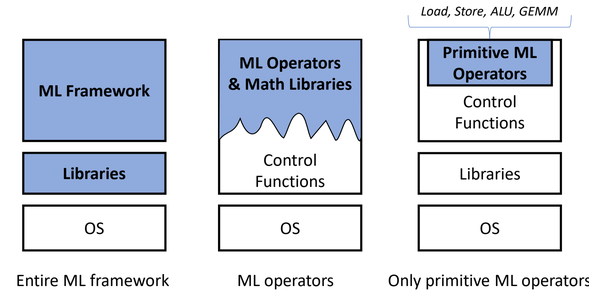
Challenges Towards Trustworthy Inference at the Edge

- Bringing inference to the edge can allow the large amounts of data to process in a timely manner
- Inference on the edge has some constraints:
  - Limited memory (<10MB)
  - Minimize Trusted Computing Base (TCB)
  - Verification of results

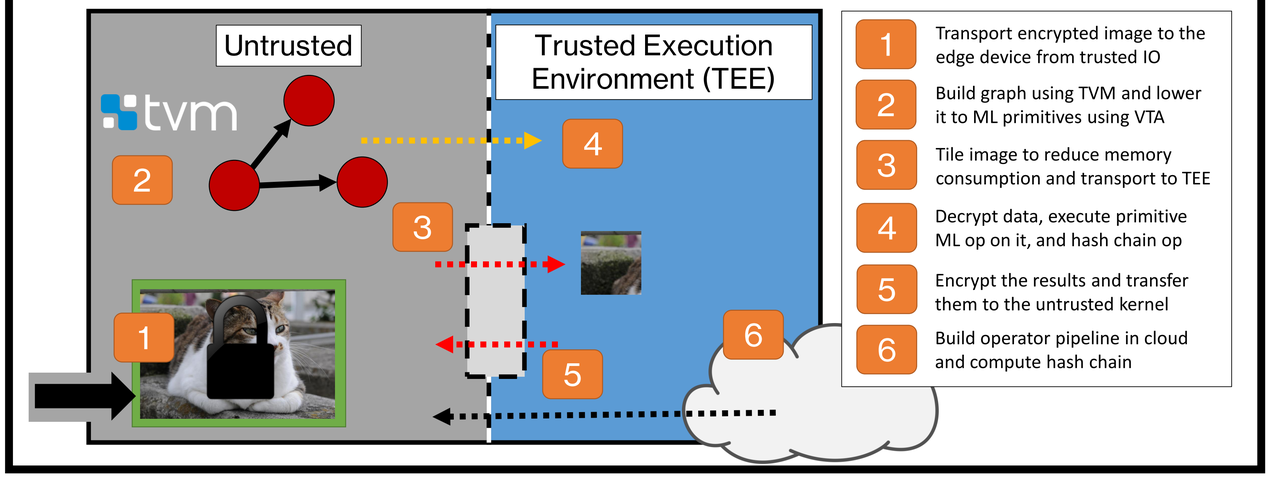


Where to Put Security Boundary in ML Software Stack

- Large software stack grants several opportunities for attacks
- Vulnerabilities allow a third party to compromise private user data, such as voice and video recordings
- Emerging trend to standardize the lowest level of the ML software stack (i.e., TVM)
- TCB size is lower if the security boundary is lower in the ML software stack

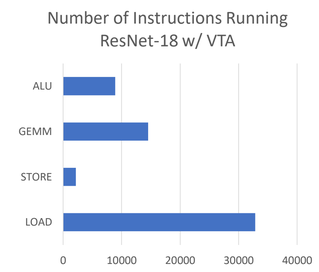


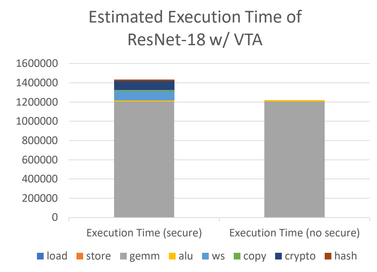
The Secure Tensor Processor Architecture



1. Transport encrypted image to the edge device from trusted IO
2. Build graph using TVM and lower it to ML primitives using VTA
3. Tile image to reduce memory consumption and transport to TEE
4. Decrypt data, execute primitive ML op on it, and hash chain op
5. Encrypt the results and transfer them to the untrusted kernel
6. Build operator pipeline in cloud and compute hash chain

Estimations of Overhead





- Estimations from ResNet-18
- The overhead for the various security features is around 0.2s, or about 14% increase in time

Future Work

- Refining the details for validating the Neural Network computation using hash chains
- Each primitive ML operator will have a unique identifier for the purposes of hash chain
- Cloud will remain agnostic of the input data to the graph

References

T. Moreau, T. Chen, L. Vega, J. Roesch, E. Yan, L. Zheng, and J. Fromm. A hardware–software blueprint for flexible deep learning specialization. In *IEEE Micro*, vol. 39, no. 5, pp. 8-16, Sept.-Oct. 2019.

H. Park, S. Zhai, L. Lu, and F. X. Lin. StreamBox-TZ: Secure stream analytics at the edge with trustzone. In *Proceedings of the 2019 USENIX Conference on Usenix Annual Technical Conference (ATC)*, 2019.

http://xsel.rocks