On-Device Network Intrusion Detection for Resource-Constrained Devices of the Internet of Things

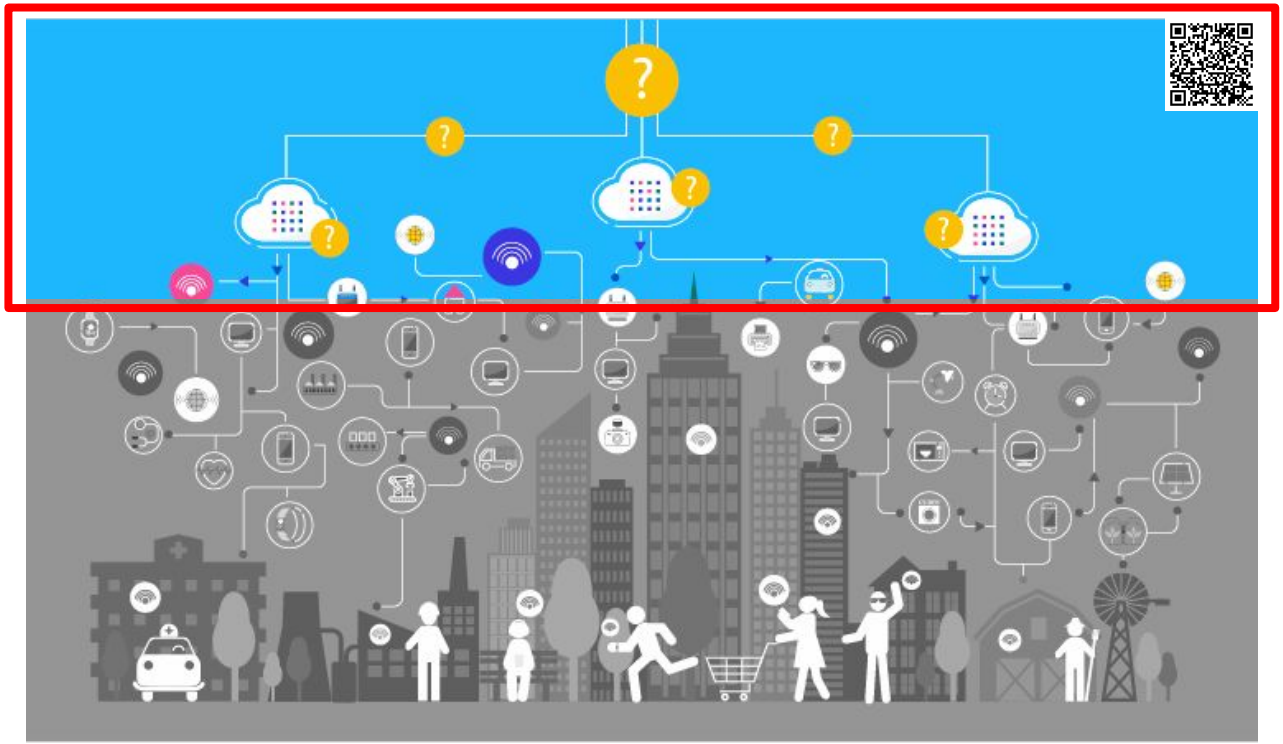
Jefferson Cavalcante[1,2], Tiago Barros[2], Neuman Souza[1]

[1]Universidade Federal do Ceará (UFC)
[2]Centro de Estudos e Sistemas Avançados do Recife (CESAR)

UNIVERSIDADE FEDERAL DO CEARÁ

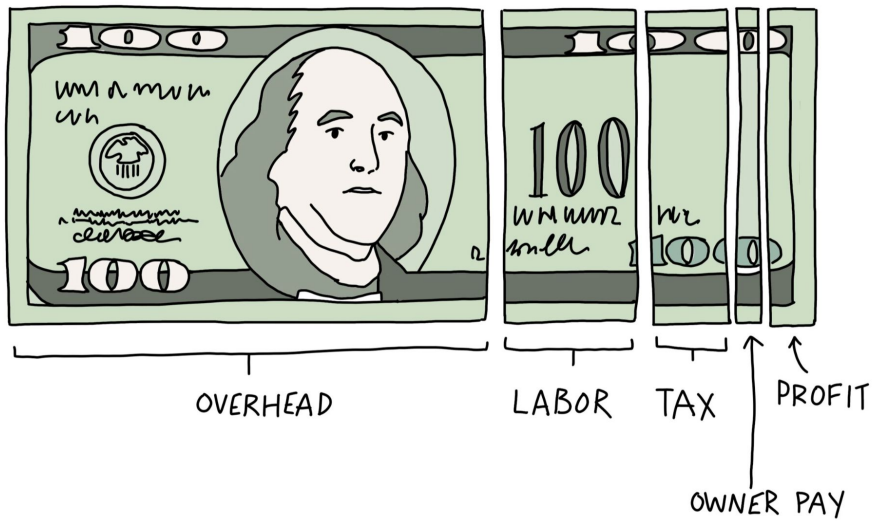
Motivação

Problemas



1 - Infraestrutura

Problemas



2 - Custo computacional

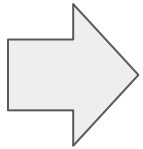
Desafios

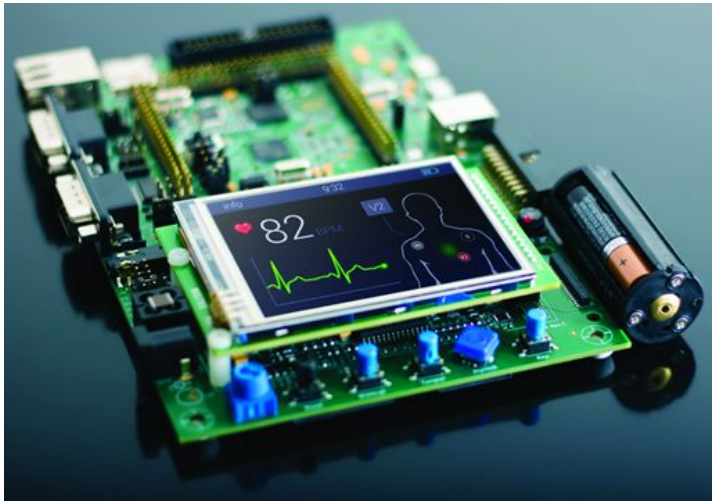
1. Dados de network intrusion
2. Selecionar microcontrolador
3. Treinar modelos de ML
4. Embarcar modelos
5. Otimizar para hardware

Desafios

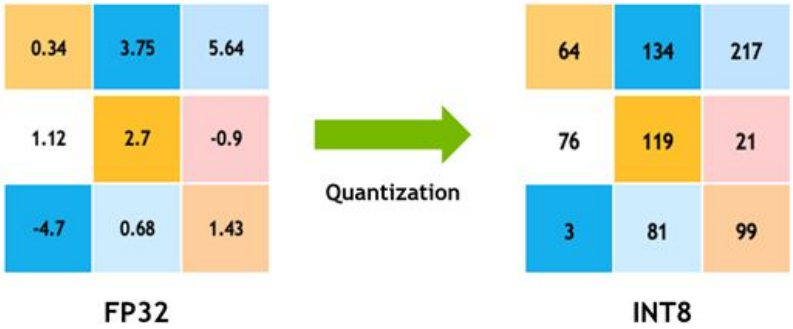
1. Dados de network intrusion
2. Selecionar microcontrolador
3. Treinar modelos de ML
4. Embarcar modelos
5. Otimizar para hardware

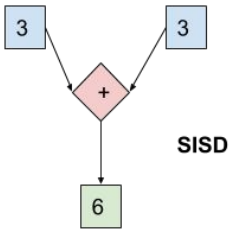
Desafios

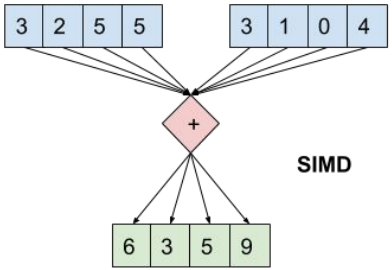




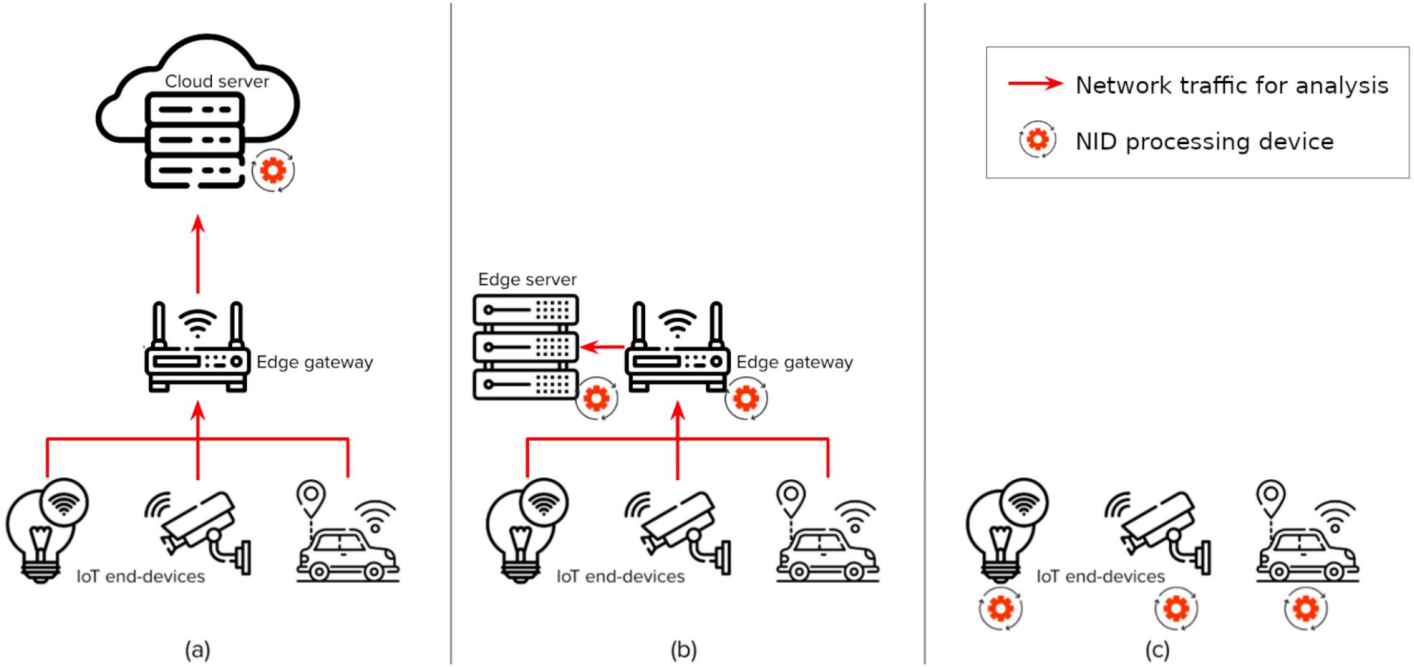
Desafios


0.34
3.75
5.64
1.12
2.7
-0.9
-4.7
0.68
1.43
Quantization
64
134
217
76
119
21
3
81
99
FP32
INT8

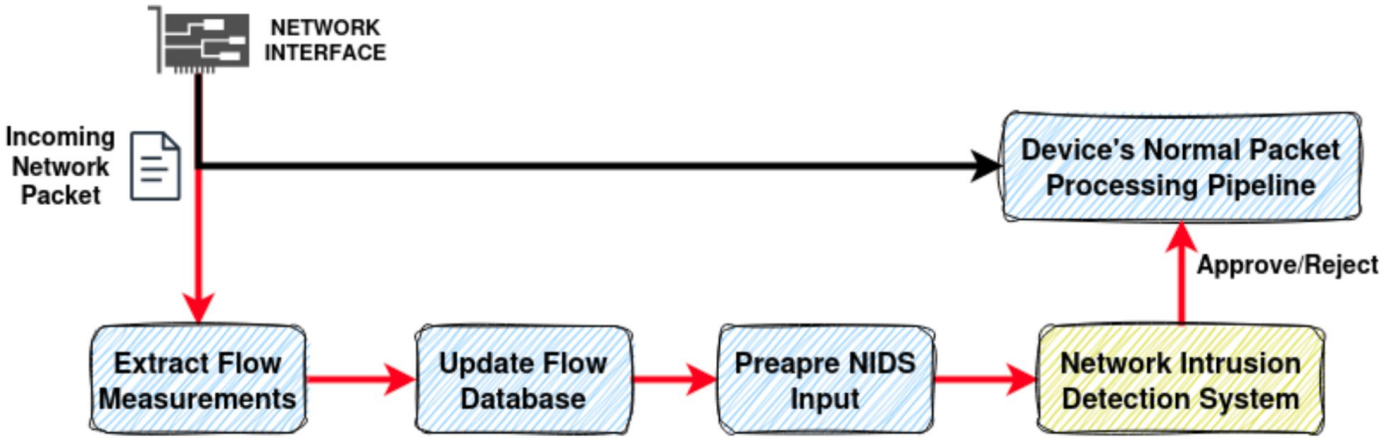

3
3
+
6
SISD


3 2 5 5
3 1 0 4
+
6 3 5 9
SIMD

Solução Proposta



Solução Proposta



Experimentos

LUFlow Dataset: 20 dias de monitoramento (2021/2022)

- Dispositivos IoT
- Honeypots

Recente, abrangente e realístico

Table 1. LUFlow features selected

Features	Descriptions
bytes_in	Cumulative number of bytes received
bytes_out	Cumulative number of bytes sent
dest_port	Flow's receiver port number
entropy	Flow's data entropy in bits per byte
num_pkts_out	Cumulative number of packets sent
num_pkts_in	Cumulative number of packets received
proto	Protocol number
src_port	Sender's port number
total_entropy	Entropy from all data fields of the flow in bytes
avg_ipt	Average of the flow's inter-packet transmission time

Experimentos

ESP32-S3

Designed for AIoT applications

2.4 GHz Wi-Fi and Bluetooth 5 (LE)
Powerful AI acceleration
Reliable security features

DSP para aceleração via hardware
Registradores de 128 bits
Single Instruction Multiple Data
API em C
Tensorflow Lite

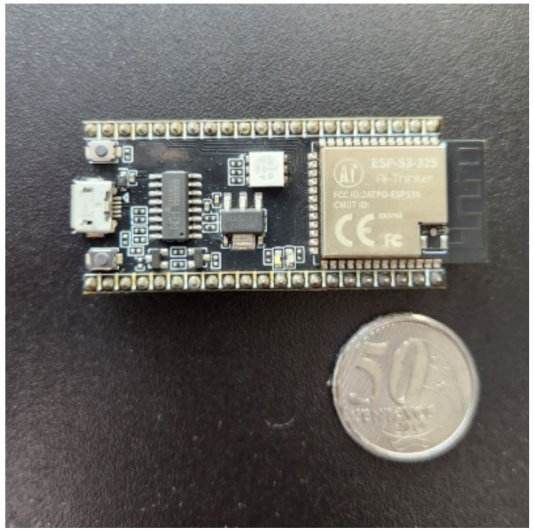
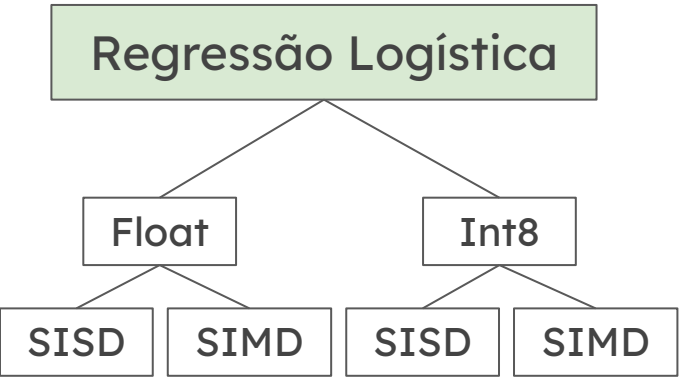
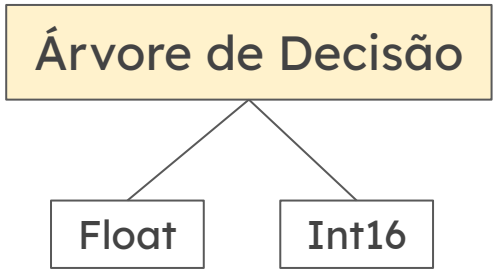
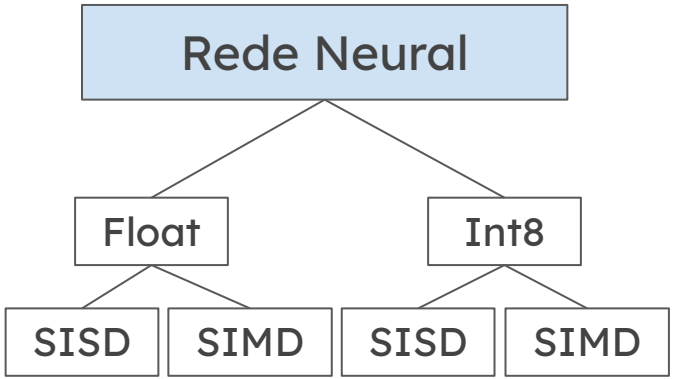


Figure 3. IoT device with an ESP32S3 chip.

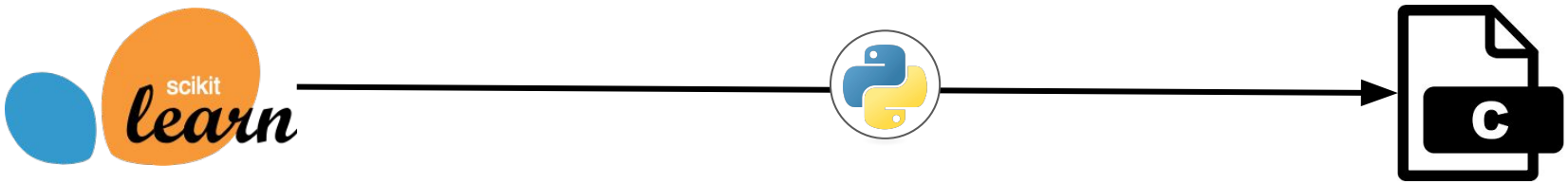
Experimentos







Experimentos



Regressão Logística
Árvore de Decisão

Conversor sklearn -> C
Quantização

DSP API

Experimentos

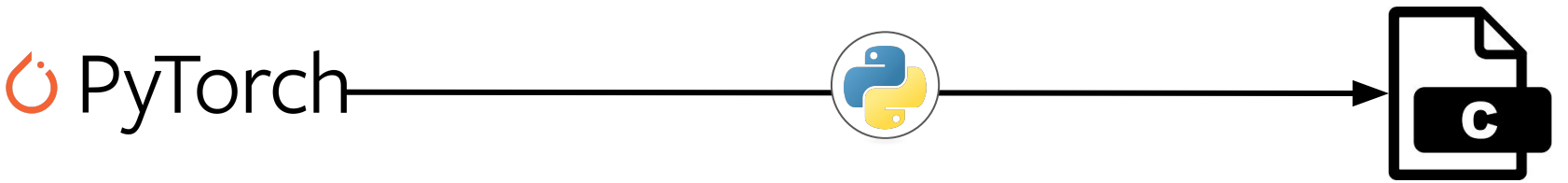
Rede Neural

ESP-DL
Conversor e
quantização

ESP-DL
API

Experimentos

Experimentos



Rede Neural + quantização

Conversor Pytorch -> C

DSP API

Avaliação

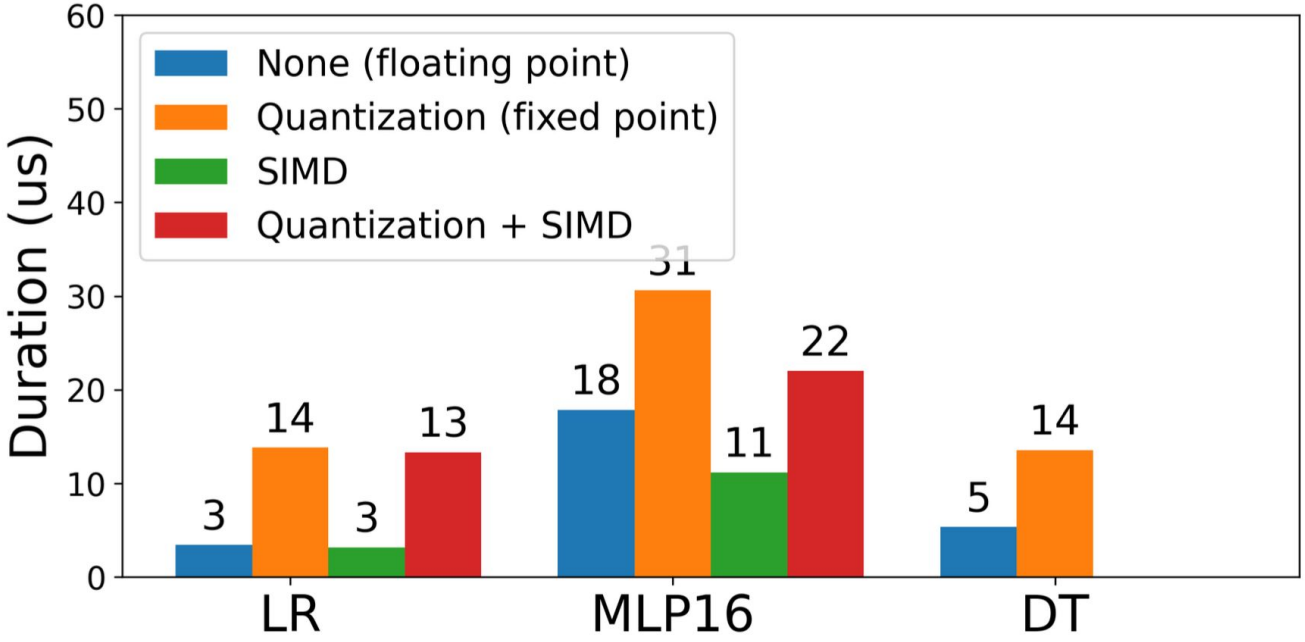


Figure 4. Average inference time.

Avaliação

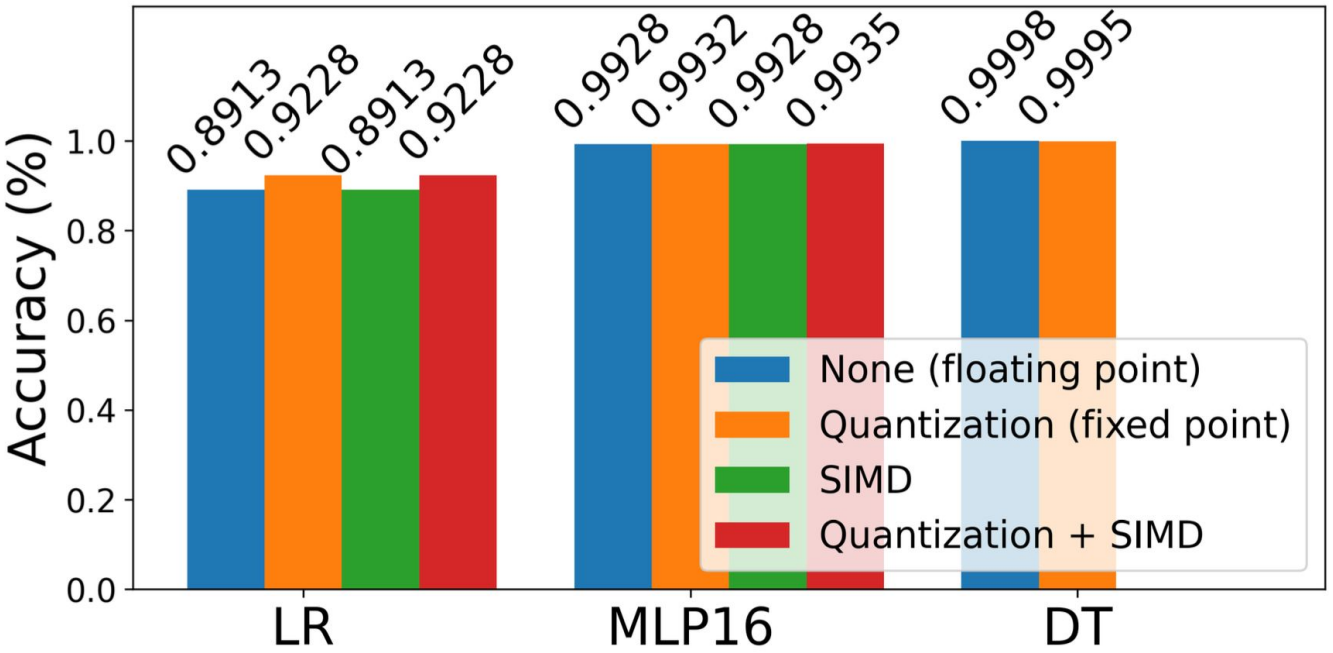


Figure 5. Accuracy.

Avaliação

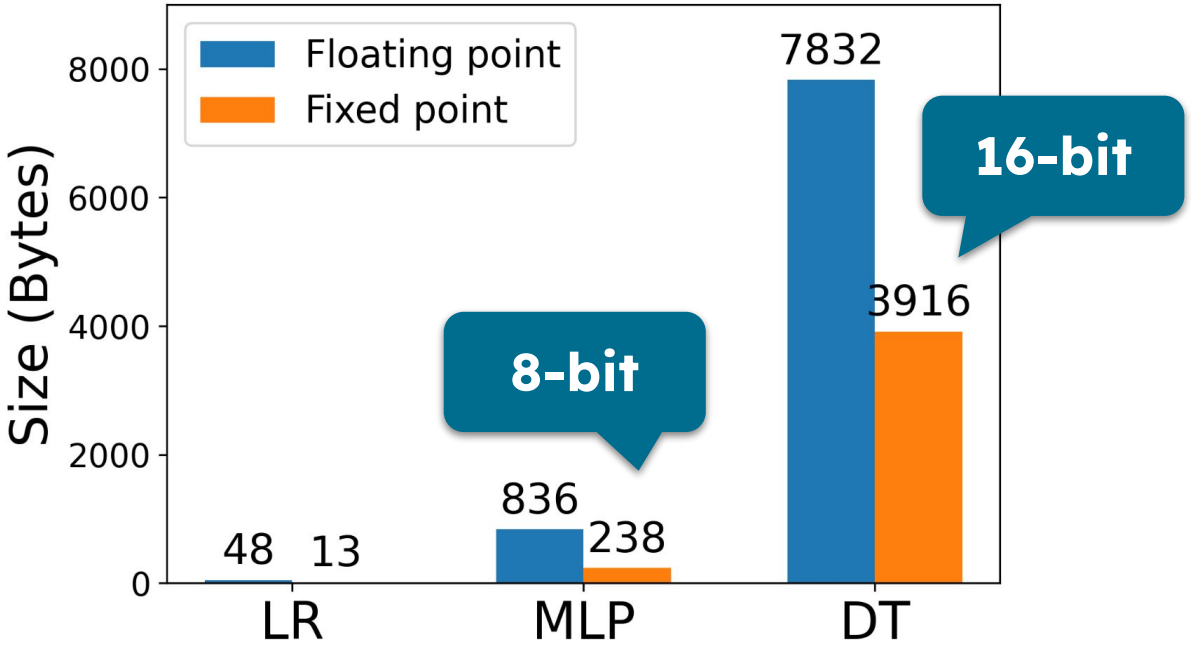
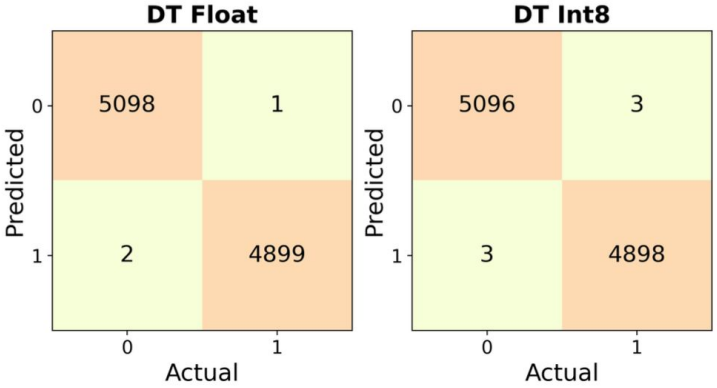
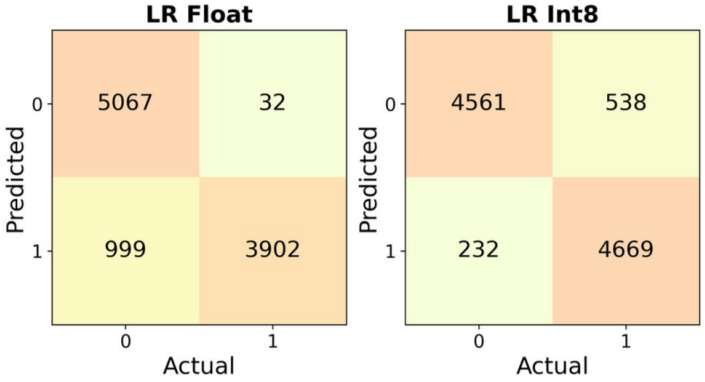
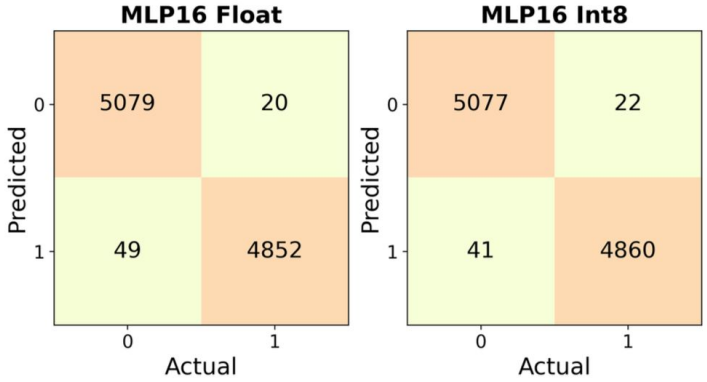


Figure 6. Model sizes based on the number of parameters.

Avaliação


DT Float
Predicted
Actual
0
1
5098
1
2
4899
DT Int8
Predicted
Actual
0
1
5096
3
3
4898


LR Float
Predicted
Actual
0
1
5067
32
999
3902
LR Int8
Predicted
Actual
0
1
4561
538
232
4669


MLP16 Float
Predicted
Actual
0
1
5079
20
49
4852
MLP16 Int8
Predicted
Actual
0
1
5077
22
41
4860

Considerações finais

- Viabilidade de Network Intrusion Detection on-device
- Árvores de decisão parecem boa escolha
  - Bom tradeoff tamanho/latência/acurácia
- Quantização e normalização são gargalos para modelos enxutos
  - Buscar evitar essa necessidade
  - Árvores de decisão não requerem normalização
- Converter de frameworks python para código nativo com aceleração por hardware nem sempre é fácil

Trabalhos futuros

- Fundir datasets para aumentar capacidade de generalização
- Acelerar árvores de decisão
- Suportar mais arquiteturas de microcontroladores

Obrigado!

Jefferson Cavalcante
jrac@cesar.org.br

Tiago Barros
tgfb@cesar.school

Jose Neuman
neuman@ufc.br

C.E.S.A.R

UNIVERSIDADE FEDERAL DO CEARÁ

Patrocinadores do SBSeg 2024!

nic.br

cgi.br

Google

Tempest

CAPES

SiDi

FAPESP

zscaler™

BugHunt

CNPq

C.E.S.A.R

FACULDADE IBPTECH