MITHRIL SECURITY

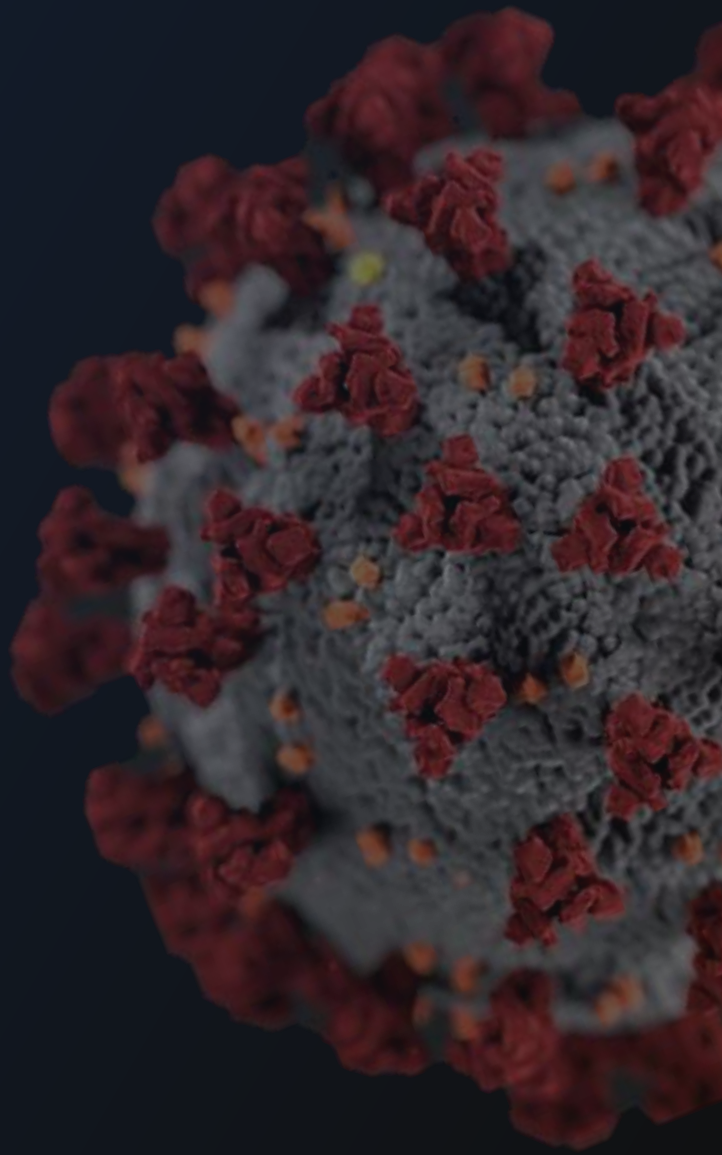
How to reconcile AI and privacy

Daniel HUYNH - CEO

June 24th, 2022

Use case

Use AI models to diagnose patients with COVID from Chest X Ray

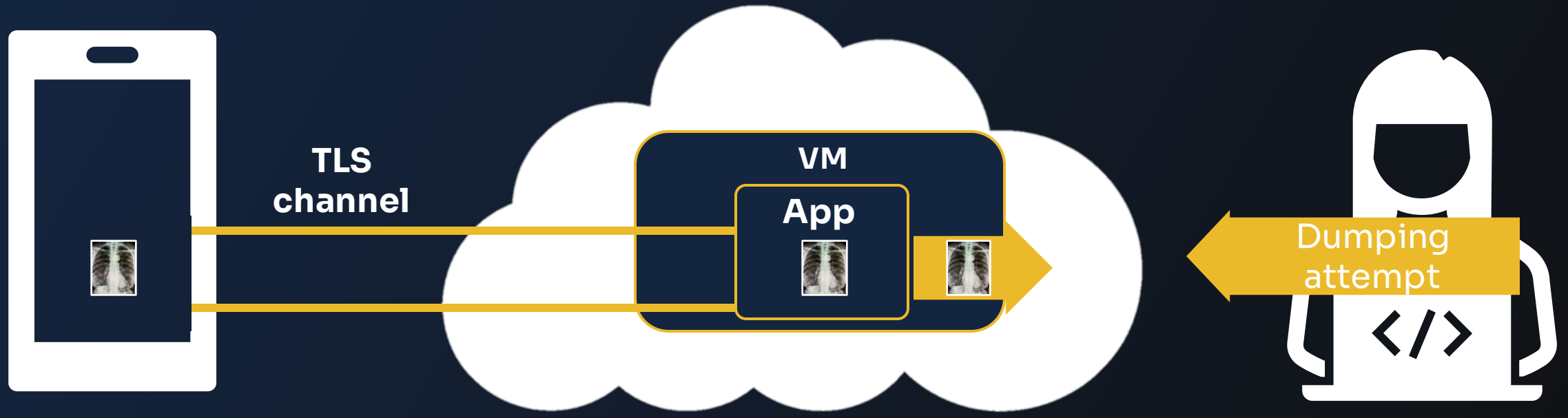


Current AI assistant workflow

User device

Cloud solution

Malicious insider



User Medical scan to be encrypted

Data is decrypted

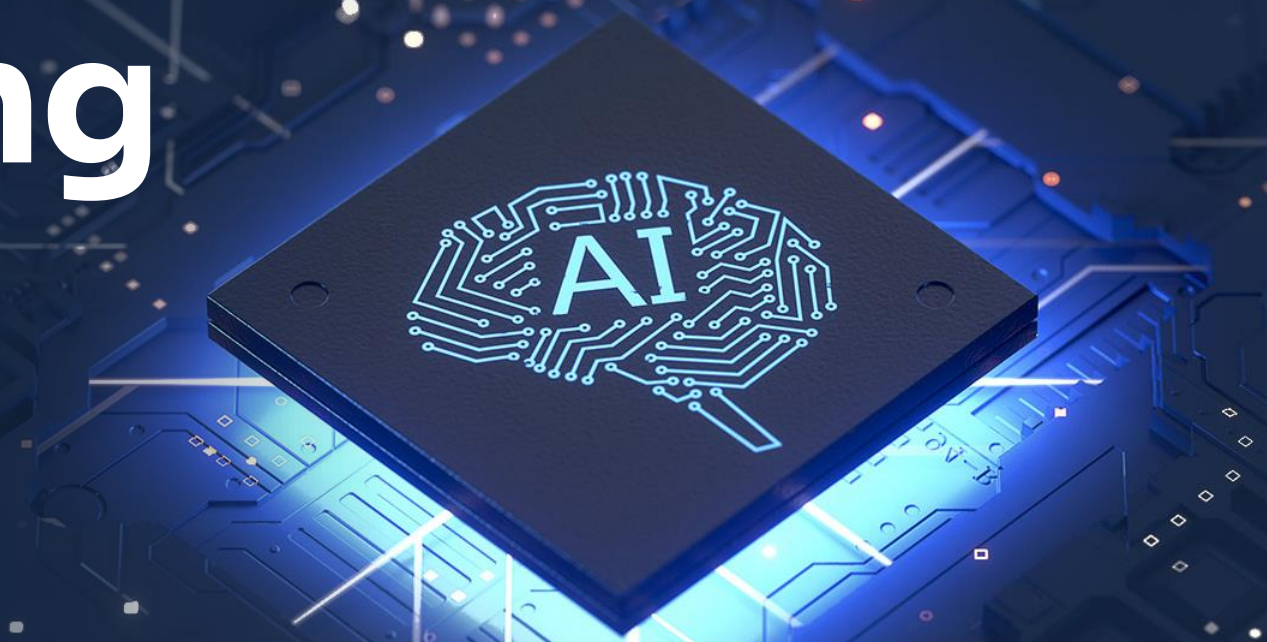
Malicious insider could access data in clear

Mithril Security

MITHRIL SECURITY

Introducing BlindAI

an Open-source and secure solution to deploy models with secure enclaves

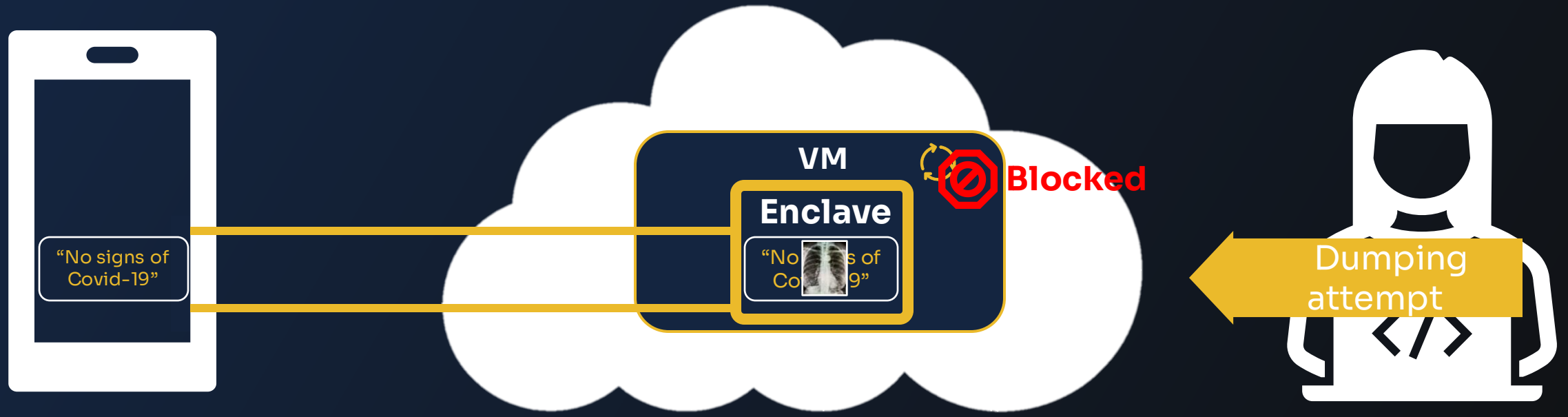


Privacy-friendly AI with BlindAI

User device

Cloud solution

Malicious insider



User can decrypt it to see the end result

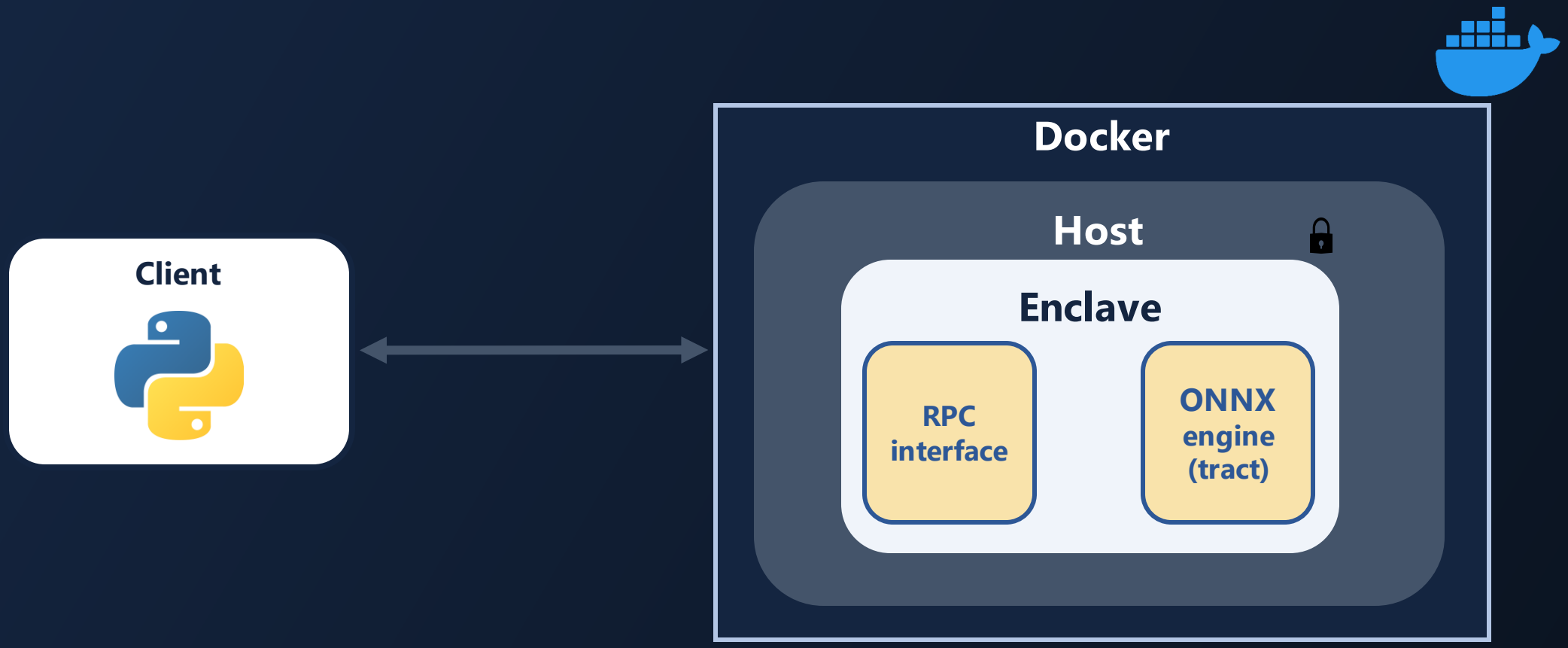
Encrypted result is sent back using the same secure channel

Enclave protects data from outside thanks to hardware isolation and encryption

Data leakage attempt fails

Mithril Security

Our architecture



Easy to deploy by engineers

1 Launch server

```
docker run \
    -p 50051:50051 \
    -p 50052:50052 \
    --device /dev/sgx/enclave \
    --device /dev/sgx/provision \
    mithrilsecuritysas/blindai-server:latest
/root/start.sh $PCCS_API_KEY
```

2 Upload model

```
from blindai.client import BlindAiClient,
ModelDatumType

# Launch client
client = BlindAiClient()

client.connect_server(
    addr="localhost",
    policy="policy.toml",
    certificate="host_server.pem"
)

client.upload_model(model="./distilbert-
base-uncased.onnx", shape=(1, 8),
dtype=ModelDatumType.I64)
```

3 Get prediction

```
from blindai.client import BlindAiClient
from transformers import DistilBertTokenizer

# Load the client
client = BlindAiClient()
client.connect_server(
    addr="localhost",
    policy="policy.toml",
    certificate="host_server.pem",
)
# Prepare the inputs
sentence = "I love AI and privacy!"
inputs = tokenizer(sentence, padding =
"max_length", max_length = 8)["input_ids"]

# Get prediction
response = client.run_model(inputs)
```

Out of the box coverage of various use cases

Baggage screening (YOLOv5)

Deploy dangerous items detection AI for airports

Hospital automation (BERT)

Provide Cloud based AI text analysis to automate administrative tasks

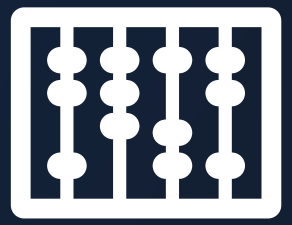
Facial recognition (ResNet)

Airport facial recognition for identification

Optimized inference speed for scalability

Model name	Inference time outside enclave	Inference time inside enclave	Hardware
BERT	49.682ms	58.023ms (+17%)	Intel(R) Xeon(R) Gold 6334 CPU @ 3.60GHz (Ice Lake)
Wav2vec2	627.148ms	755.621ms (+20%)	Intel(R) Xeon(R) Gold 6334 CPU @ 3.60GHz (Ice Lake)
Facenet	44.749ms	46.300ms (+3%)	Intel(R) Xeon(R) Gold 6334 CPU @ 3.60GHz (Ice Lake)
YOLOv5	340.762ms	348.461ms (+2%)	Intel(R) Xeon(R) Gold 6334 CPU @ 3.60GHz (Ice Lake)

Minimal codebase for security



Just enough operators to run most models

Reinforced operators for side channel resistance

Get started with secure AI now

Contact us

contact@mithrilsecurity.io

Try BlindAI

github.com/mithril-security/blindai

Annex

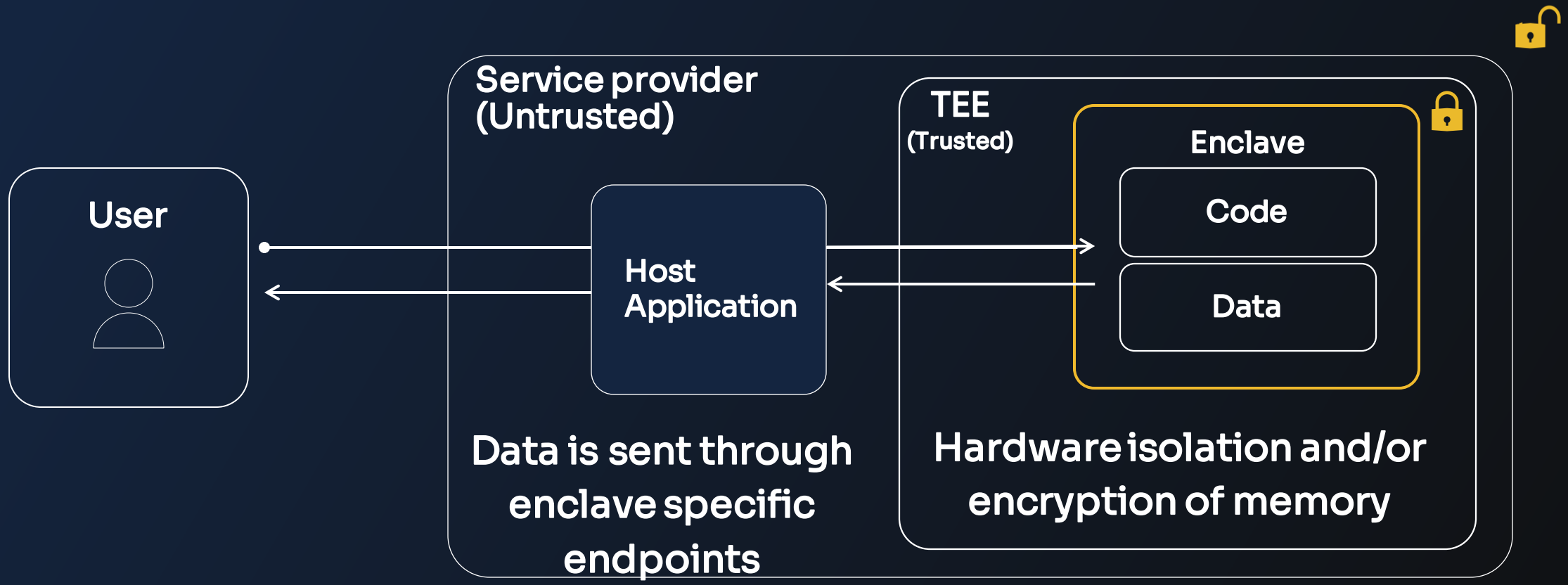
Introducing Confidential Computing

"Confidential Computing protects data in use by performing computation in a **hardware-based Trusted Execution Environment**."

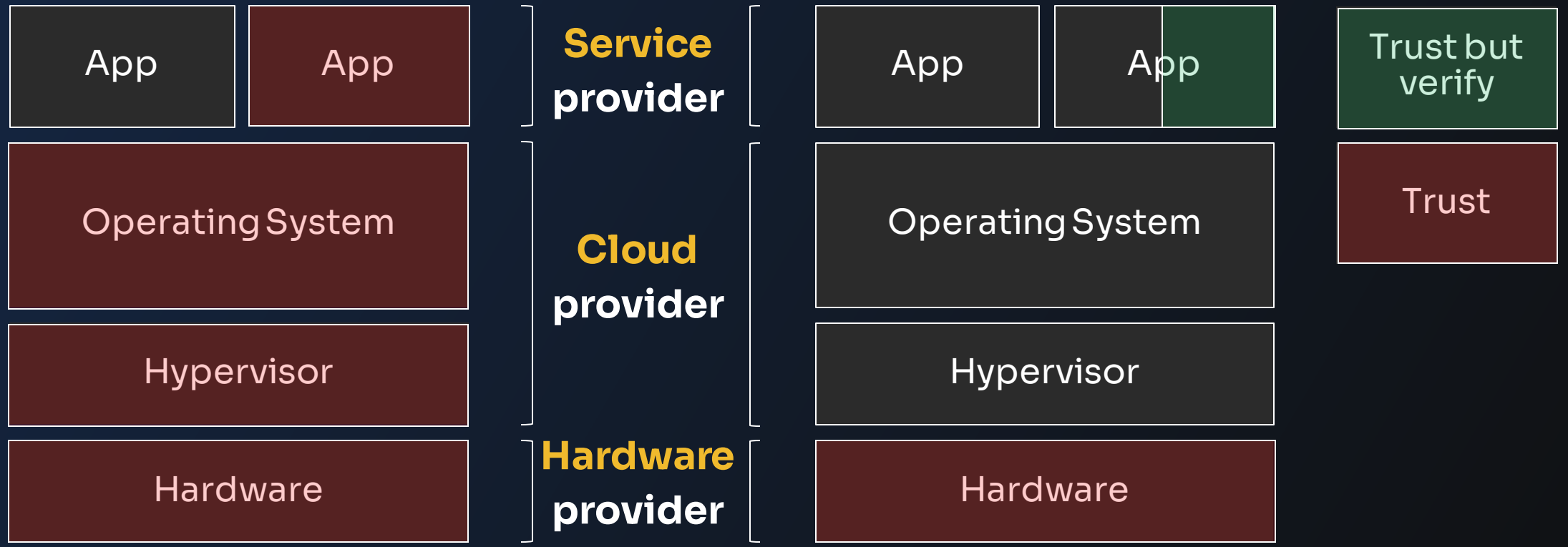
DATA PRIVACY
Data is **protected** even during analyses

CODE ATTESTATION
Providing cryptographic **proofs** of what code is executed

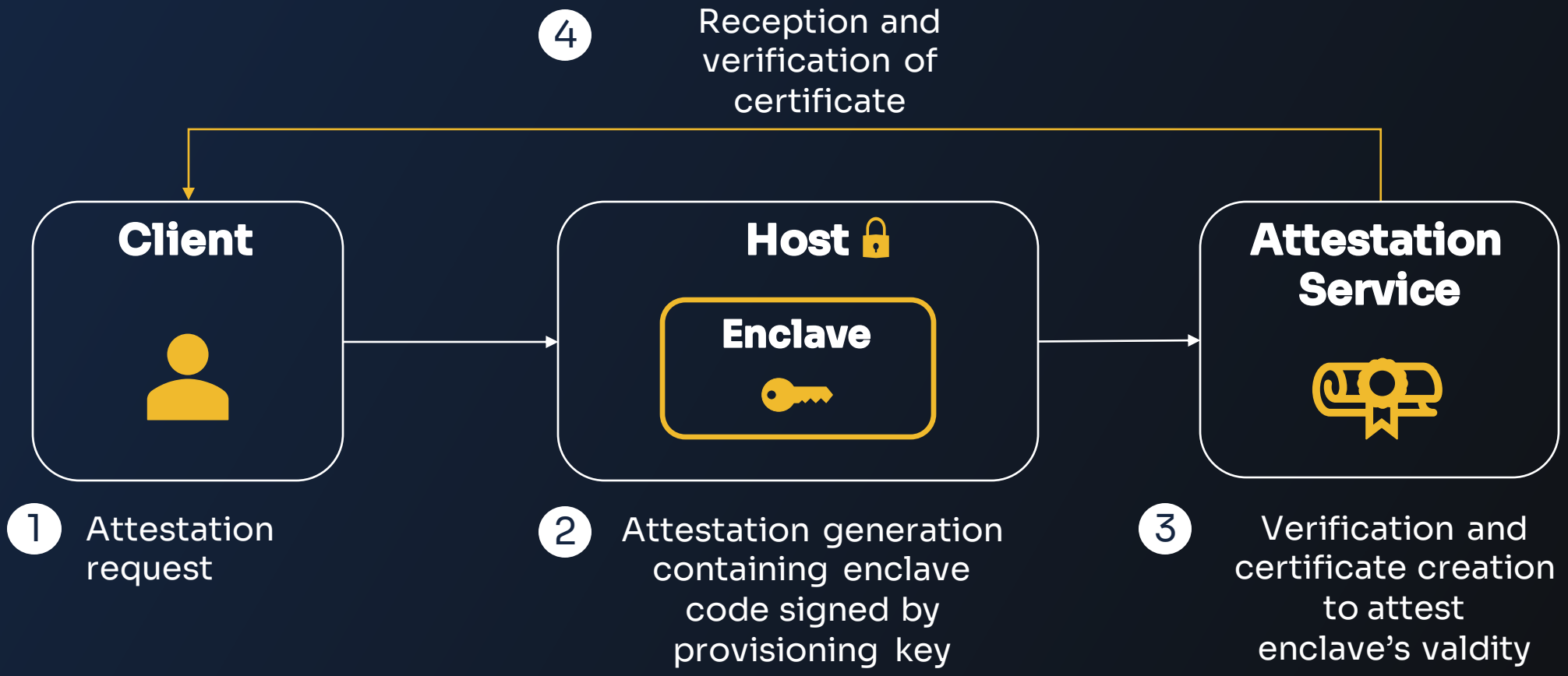
Data protection during analysis



A shift in trust model



Remote attestation for code integrity



Example with Remote attestation-TLS