Encrypted Traffic Classification: From Small Data to Scalable Production

Methods for Detecting Cyber Attacks

Raz Cohen, Dael Hacohen Waingarten, Shir Bismuth

ARIEL UNIVERSITY

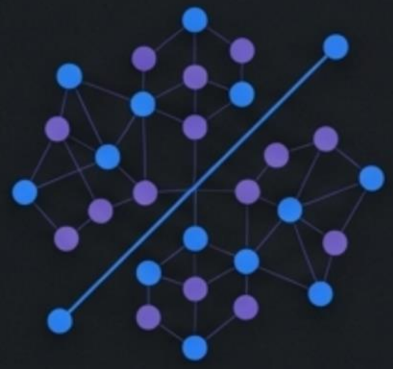
The 'Small Data' Challenge: From Big Data to Smart Data

Identifying 128 applications from sparse samples required a strategic shift from 'Big Data' volume to 'Smart Data' intelligence.

The Challenge: High Sparsity, High Cardinality

- **128 Applications:** Massive class imbalance due to a 'long tail' effect with few training examples per class.
- **5 Attribution Types:** Limited instances to differentiate nuanced behaviors like real-time audio vs. video.
- **Encrypted Payloads:** Traditional port-based or deep-packet inspection methods are rendered obsolete.

Our Strategy: Engineering Resilience



- **Class-Weight Balancing:** Configured models with `class_weight='balanced'` to force equal attention to rare classes.
- **Domain-Driven Feature Selection:** Engineered features based on network protocol theory to capture stable signals and reduce noise.

Feature Engineering: Uncovering the Behavioral DNA of Encrypted Flows

Application Identification (128 Classes)

Key Insight: Focused on the unencrypted metadata of the TLS negotiation phase.

Top Features

`handshake_avg` & `handshake_std`
Measures timing and size patterns of the initial TLS handshake. The negotiation protocol itself is a unique signature.

`fwd_bwd_pkts_diff`
Captures flow asymmetry, distinguishing upload-heavy vs. download-heavy applications.

Attribution (5 Classes)

Key Insight: Created custom signatures to model spatio-temporal usage patterns.

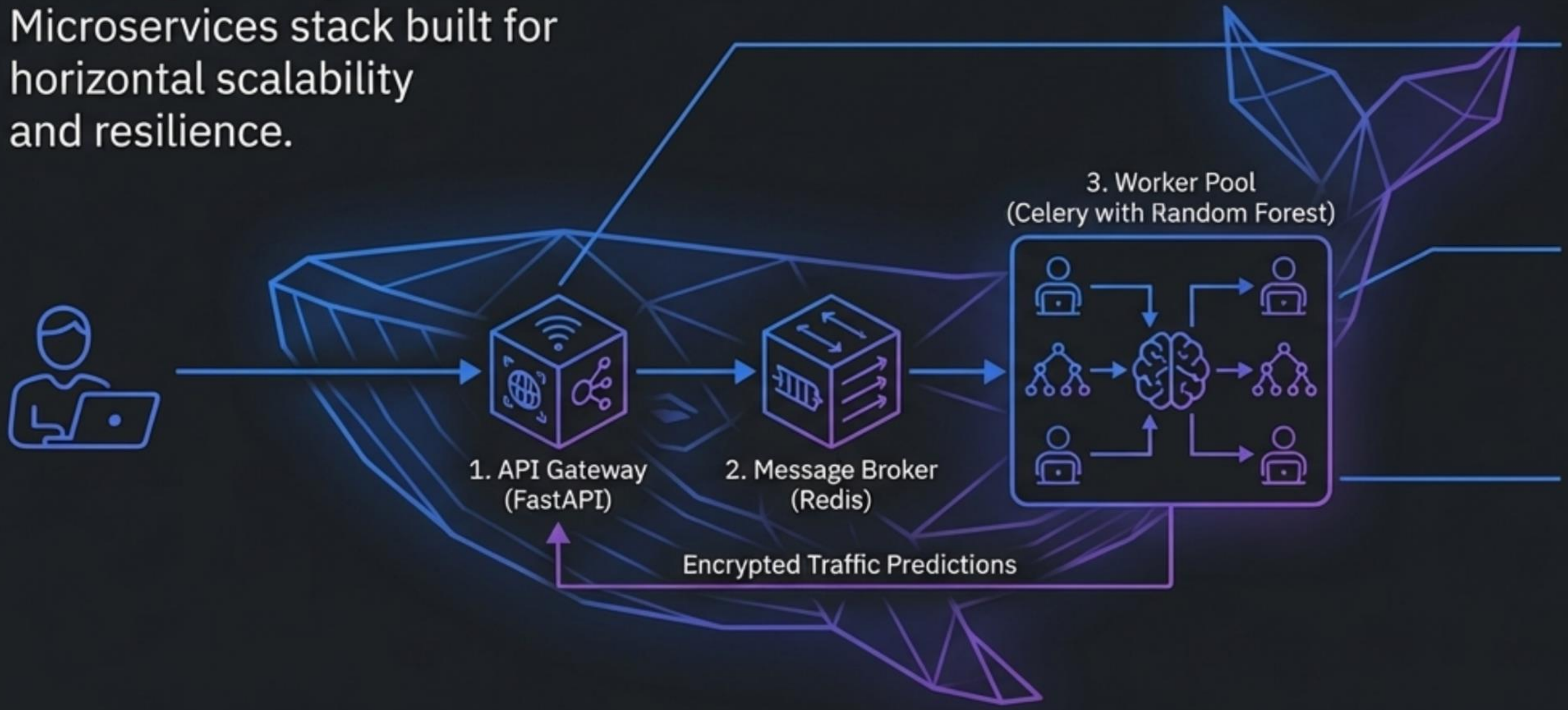
Custom Signatures

`chat_signature`
A metric combining `silence_ratio` and `burstiness` to identify messaging patterns (long silences, short data bursts).

`stream_signature`
Highlights high-volume, high-throughput, low-silence traffic typical of VOD or large file downloads.

System Architecture: Engineered for the Hexabyte Era

A decoupled, asynchronous Microservices stack built for horizontal scalability and resilience.



1. API Gateway (FastAPI):
High-performance, non-blocking entry point for receiving traffic data.

2. Message Broker (Redis):
In-memory task queue buffers requests, preventing overload and enabling asynchronous processing.

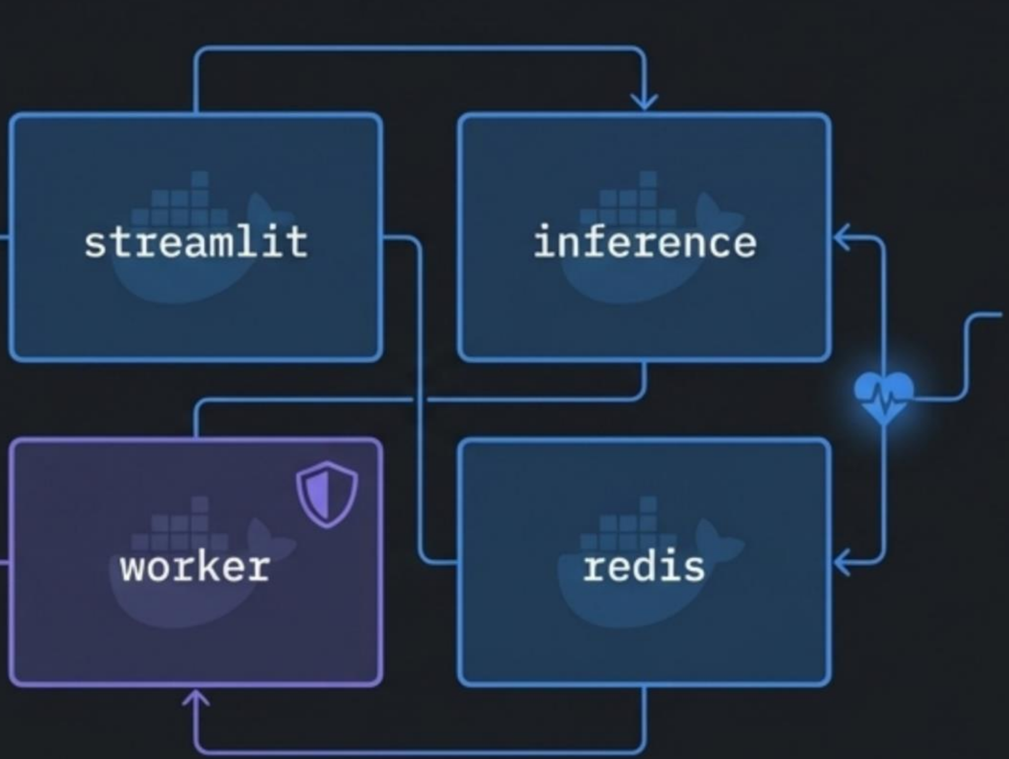
3. Worker Pool (Celery with Random Forest):
Stateless 'heavy lifters.' Pre-load models for millisecond inference. Scale horizontally by adding more worker containers.

This design ensures the UI remains responsive and inference can scale to handle massive telecom traffic without changing core code.

Secure & Reliable Deployment: A Containerized Approach

Multi-Container Orchestration

- We use Docker Compose to manage four specialized services: `streamlit`, `inference`, `worker`, and `redis`.
- This separation optimizes image sizes and allows independent scaling of the ML workload.



Production-Grade Reliability

- **Health Checks: System ensures the Redis broker is 'Healthy' before starting dependent services, preventing startup failures.
- **Environment Consistency: Docker guarantees an identical environment from development to production, eliminating 'it works on my machine'.

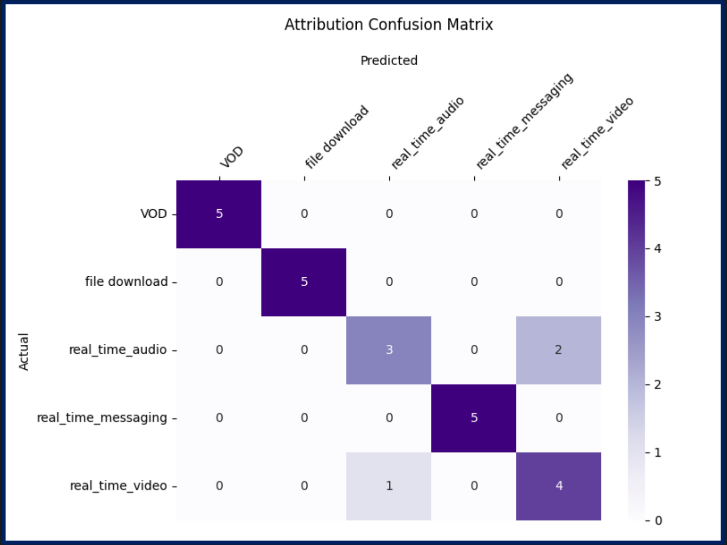
Embedded Cybersecurity Practices

- **Non-Root Execution: Containers run under a dedicated `appuser`, preventing root access to the host.
- **Minimal Attack Surface: We use `python:3.11-slim` base images to reduce vulnerabilities and storage overhead.

Results & Impact: Depth, Accuracy, and Scale

Attribution Classification (5 Classes)

88% Accuracy



Actual \ Predicted	VOD	file download	real_time_audio	real_time_messaging	real_time_video
VOD	5	0	0	0	0
file download	0	5	0	0	0
real_time_audio	0	0	3	0	2
real_time_messaging	0	0	0	5	0
real_time_video	0	0	1	0	4

- Perfect classification for VOD, File Download, and Messaging due to distinct behavioral signatures.
- Primary challenge: confusion between `real_time_audio` and `real_time_video` due to similar low-latency UDP patterns.

Application Identification (128 Classes)

51.7% Accuracy

66x Better Than a Random Guess

(0.78% baseline)

- Model excels at identifying apps with unique TLS `handshake` dynamics, proving our 'Behavioral DNA' approach.
- Example apps with high F1-scores: "brainfund` (1.00), `vimeo` (1.00), `ndtv` (1.00), `wprdc.org` (1.00)

A production-ready solution that successfully balances classification depth with the demands on massive scale