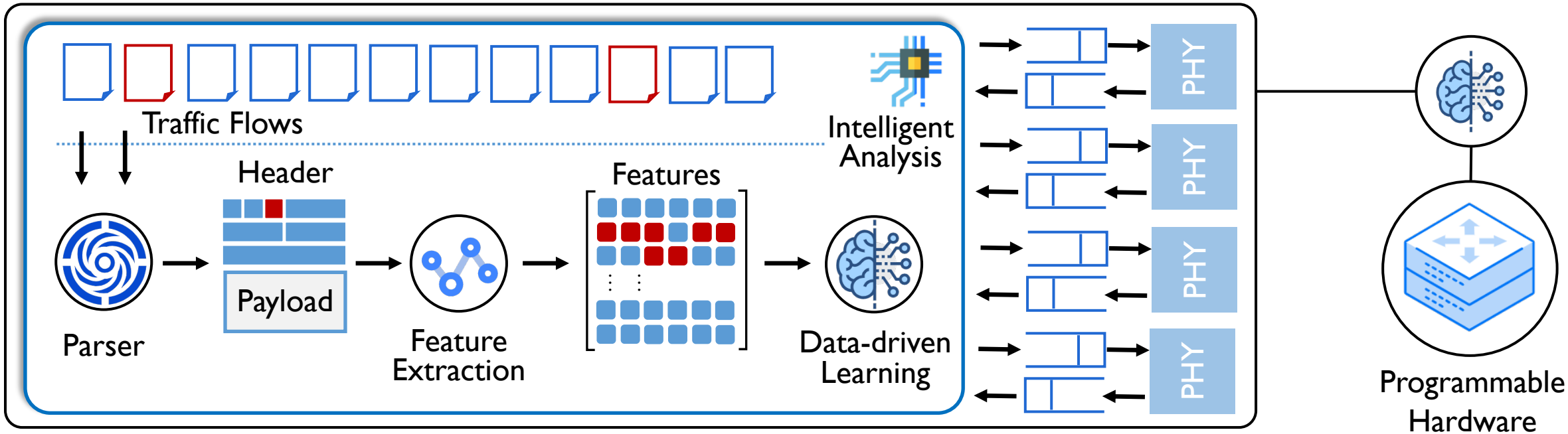
IETF 125 Hackathon

Pegasus

Pegasus: Deep Learning Framework on the Dataplane

Intelligent Dataplane (IDP)

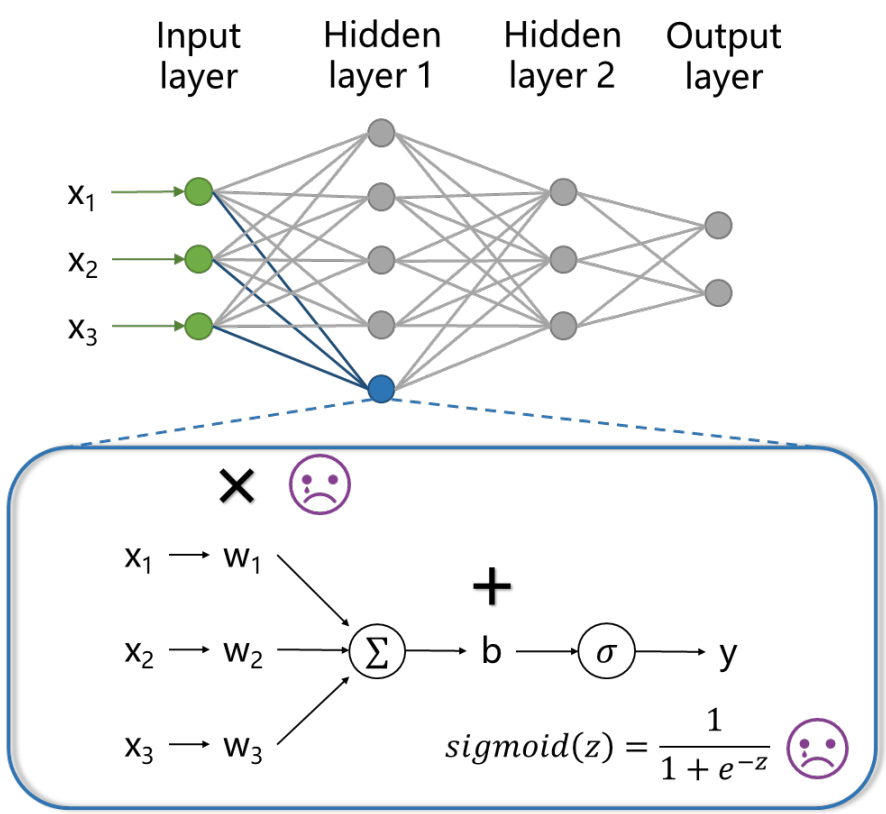
The Challenge: Exploding traffic diversity and encrypted threats in 5G/6G & Cloud ecosystems.
The Solution: Embedding AI inference directly into the data plane.
The Vision: Enabling Intelligent Dataplane for real-time, autonomous traffic analysis.

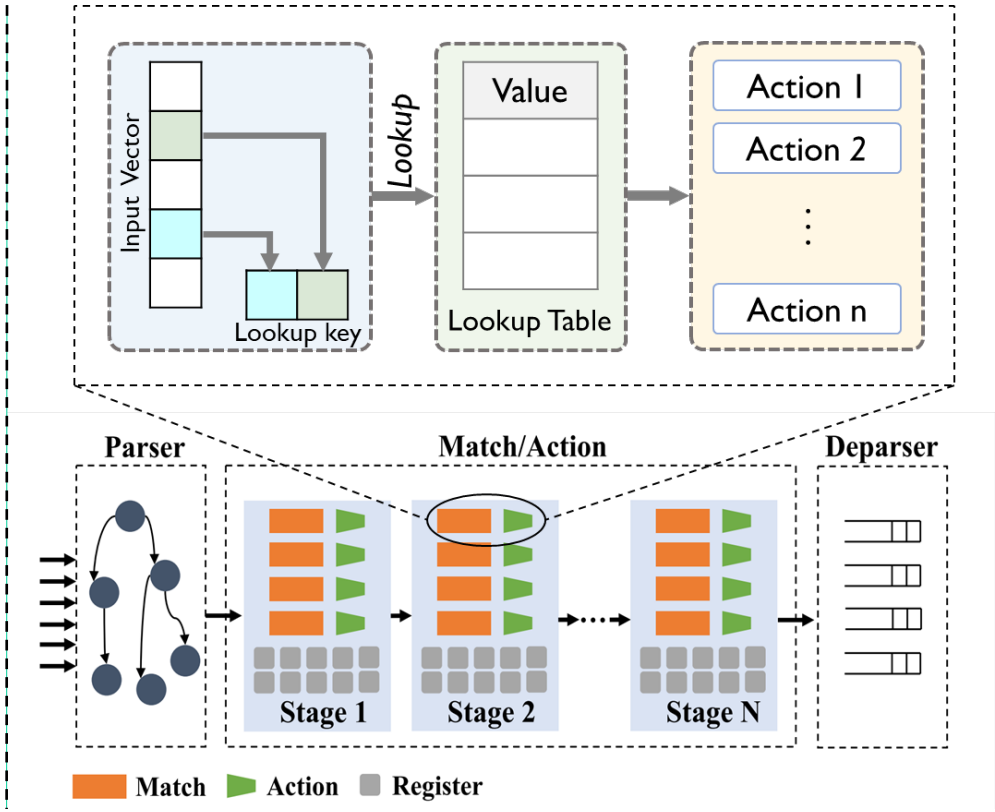


Line-speed Intelligence Flexibility Efficiency

Challenge

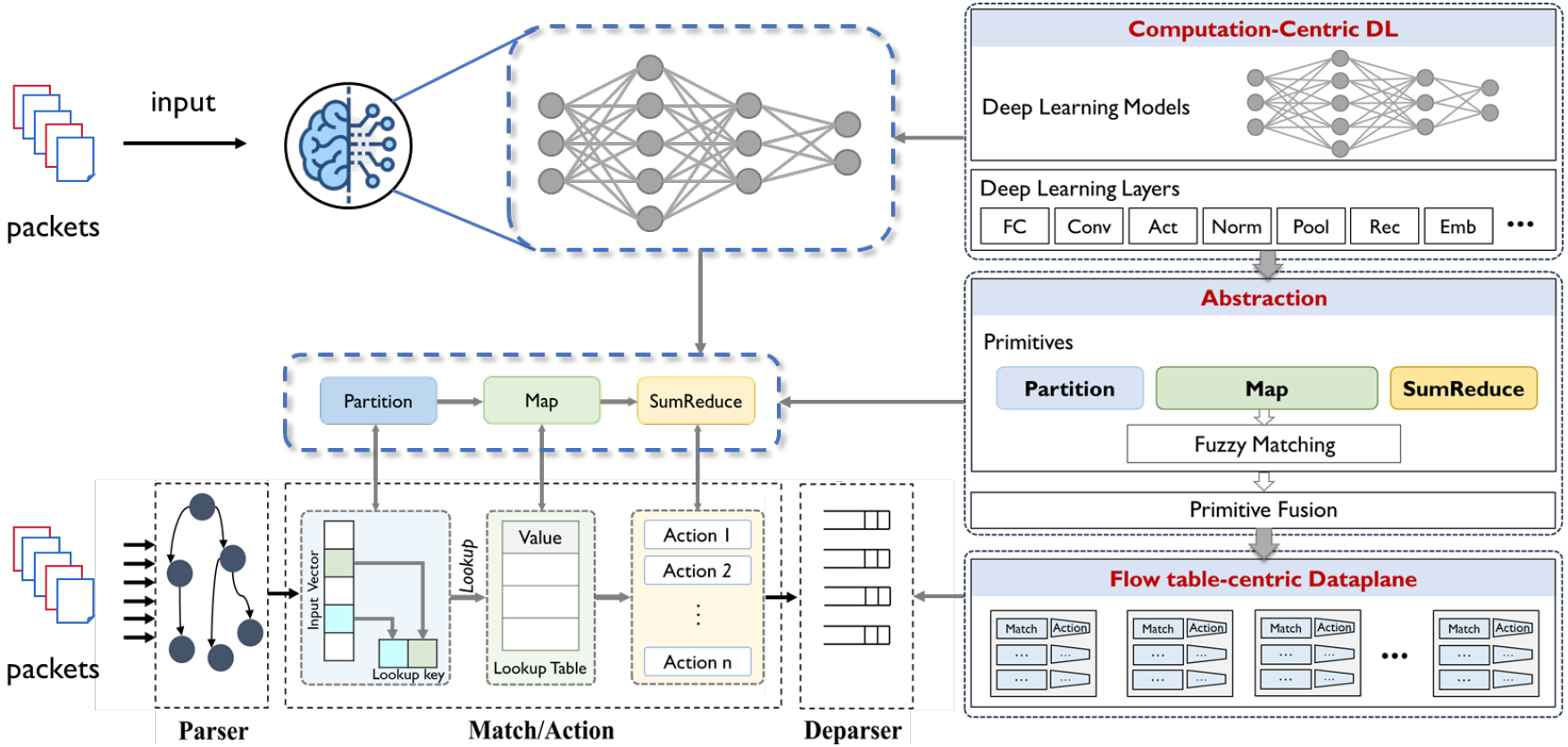
The Conflict: DL models demand intensive matrix operations and high memory buffers.
The Constraint: Data planes are optimized for simple flow tables and restricted resources.
The Core Mismatch: Flow-based Abstraction vs. Computation-Centric Deep Learning.





Method

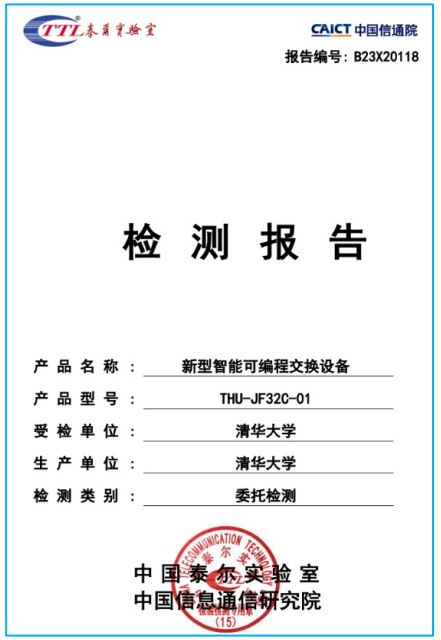
Concept: Translating computation-centric DL into flow-centric data plane primitives.
Mechanism: Decomposing complex models into a sequence of hardware-native primitives.
Optimization: High-efficiency deployment via Fuzzy Matching and Primitive Fusion.



Conclusion

Realizing the Pegasus Framework

- **Proven Prototype**: Successfully implemented on our self-developed hardware platform.
- **Pegasus Framework**: First general-purpose, scalable DL inference framework for line-rate dataplane.
- **Performance**: Achieving high-throughput inference with minimal accuracy trade-offs.



CAICT 中国信通院

报告编号：B23X20118

检 测 报 告

产 品 名 称 ： 新型智能可编程交换设备

产 品 型 号 ： THU-JF32C-01

受 检 单 位 ： 清华大学

生 产 单 位 ： 清华大学

检 测 类 别 ： 委托检测

中 国 泰 尔 实 验 室

中国信息通信研究院

Future Work

Beyond a Single Node

- **Distributed Intelligence**: Scaling from single-device inference to multi-node collaborative intelligent data planes.
- **Standardization**: Leading the effort to define industry standards for Intelligent Dataplane architectures.
- **Ecosystem Growth**: Advancing robust, automated deployment pipelines for next-gen 6G networks.

Source code: https://github.com/afireswallow/Pegasus

Email: afireswallow@gmail.com

Thanks! Questions?