I2INF for Data Center Netoworking——Usecases and Framework

Kehan Yao (China Mobile)

Wenfei Wu (Peking University)

Jaehoon Paul Jeong (Sungkyunkwan University)

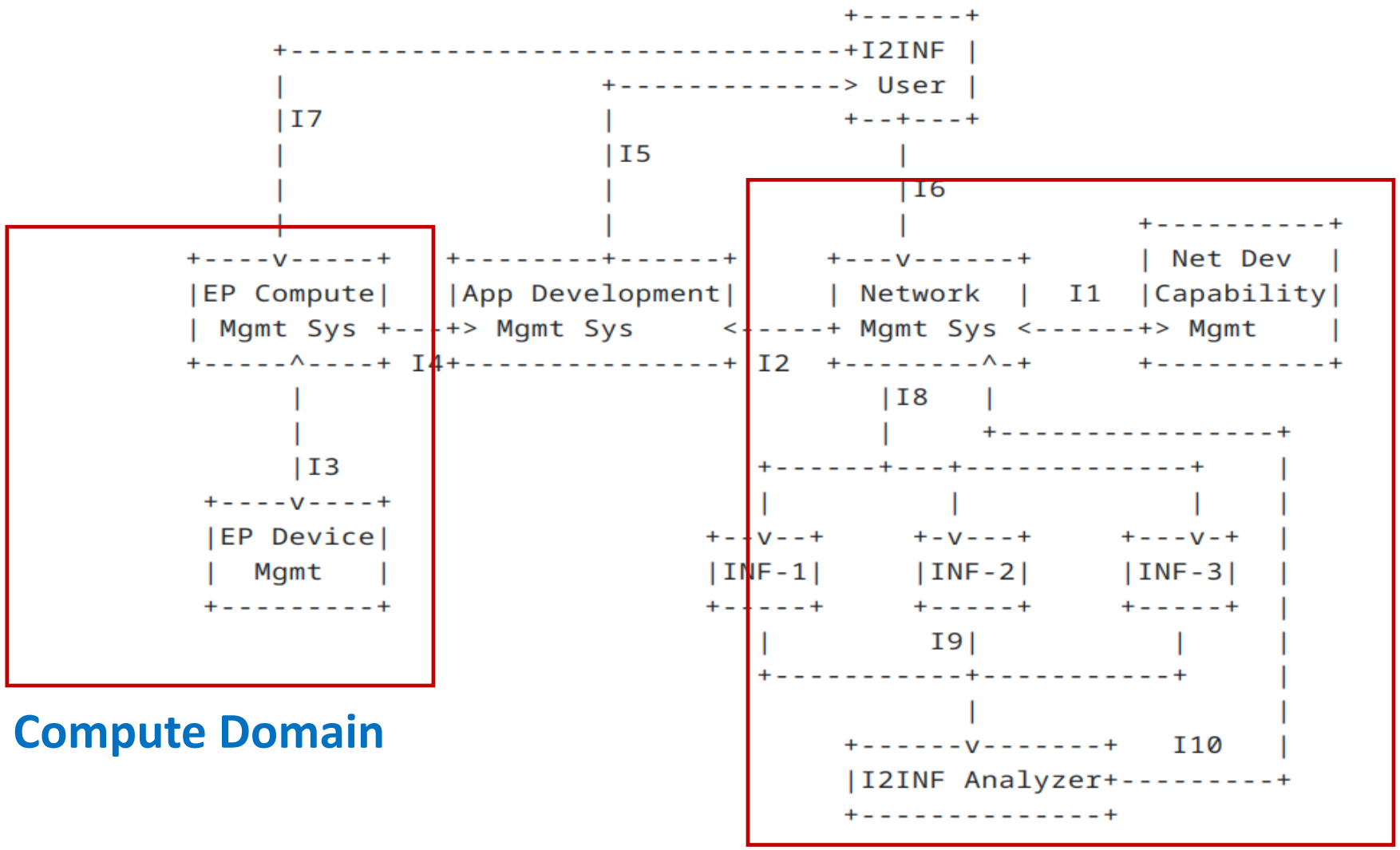
IETF 121, Dublin

draft-ywj-opsawg-i2inf-data-center-networking-00

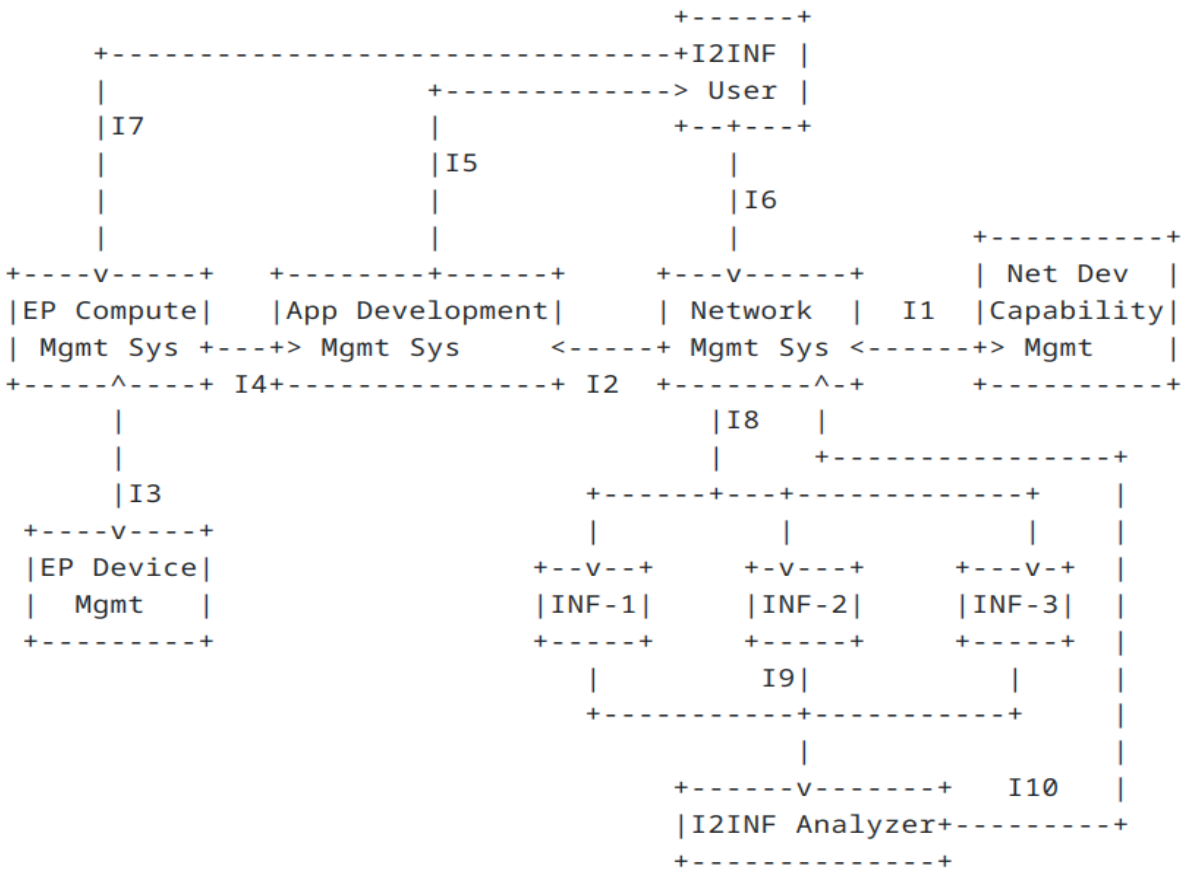
I2INF Data Center Networking Applications

- Distributed Machine learning:
  - In-network aggregation for collective communication
  - In-network inference
- Big Data Analysis:
  - K-V aggregation
- Operations offloading for common distributed applications:
  - In-network caching
  - In-network lock management

I2INF Data Center Networking Framework

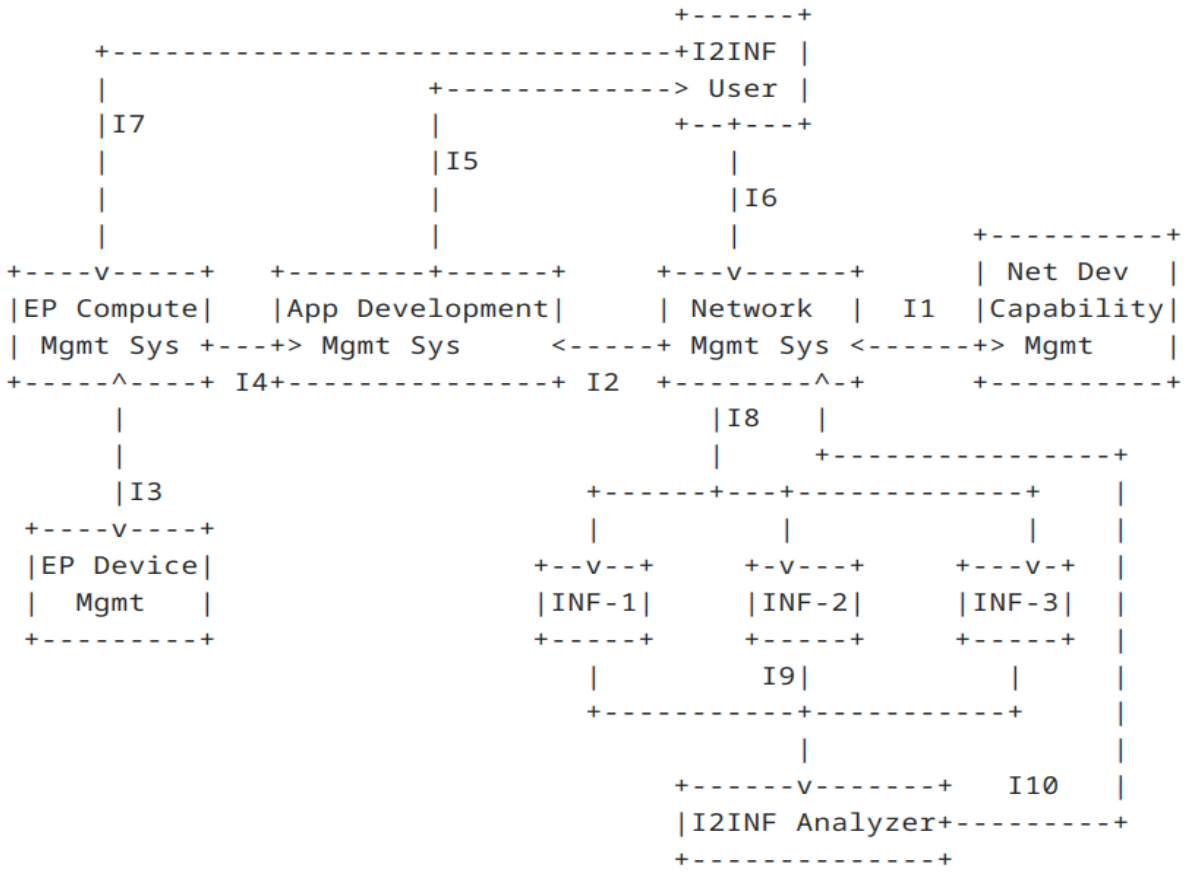


I2INF Data Center Networking Interfaces



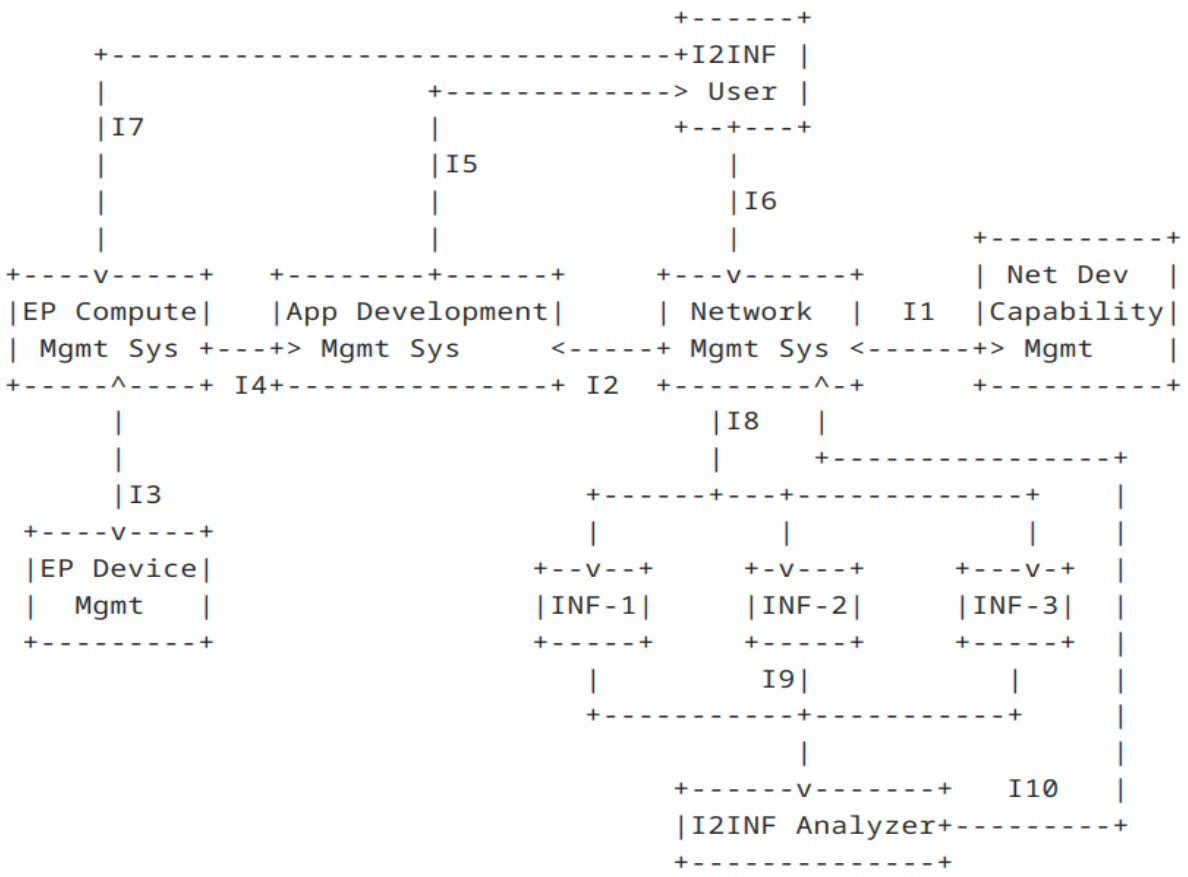
Interfaces	Meanings
I1	Registration interface between network device capability management system and network management system.
I2	In-network computing capabilities exposure interface

I2INF Data Center Networking Interfaces



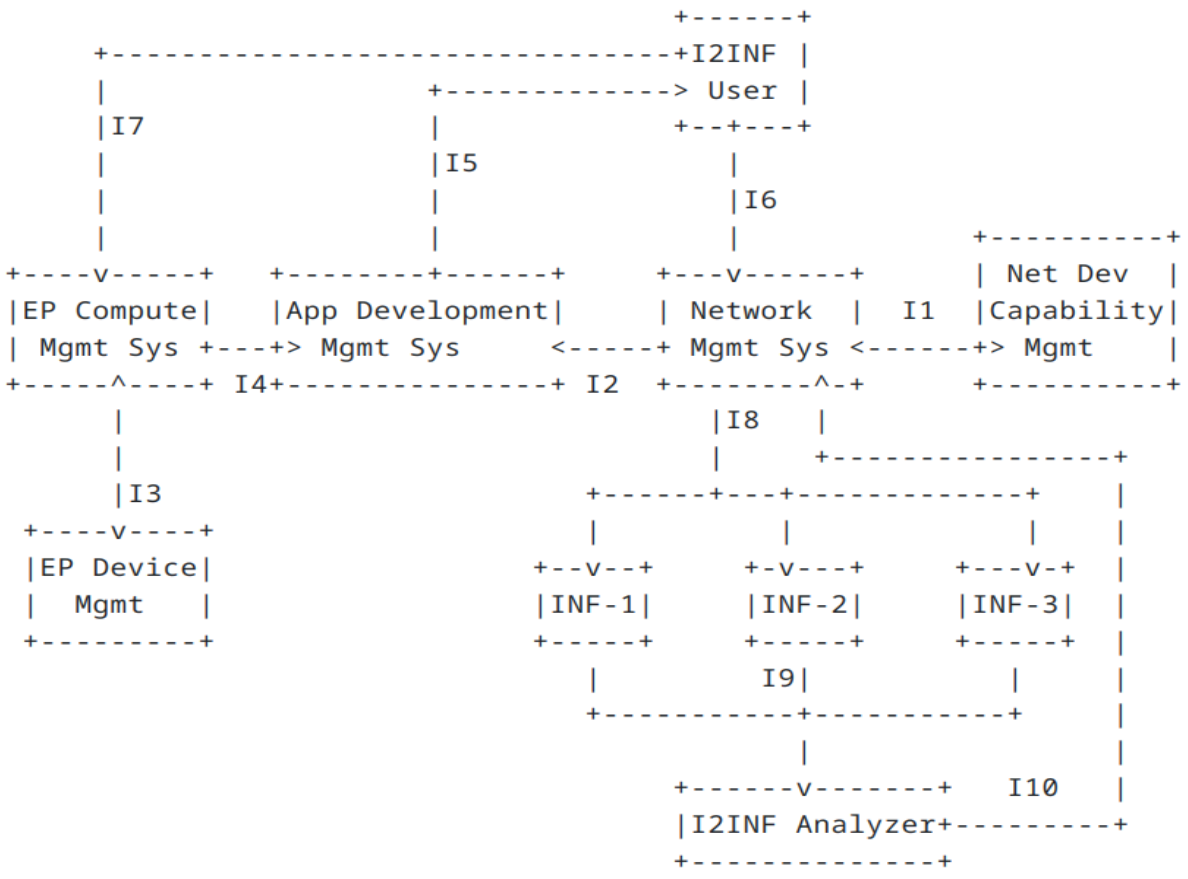
Interfaces	Meanings
I3	Registration interface of endpoint compute capabilities.
I4	Endpoint computing capabilities exposure interface

I2INF Data Center Networking Interfaces



Interfaces	Meanings
I5	Notification interface from application development management system to the I2INF user.
I6	Configuration interface from task management platforms to network management system.
I7	Configuration interface from task management platforms to the endpoint compute management system.

I2INF Data Center Networking Interfaces



Interfaces	Meanings
I8	Configuration interface for INFs with a low-level policy that is translated from the high-level policy.
I9	Monitoring interface via which monitoring data is collected from INFs to I2INF analyzer.
I10	Analytics interface via which policyreconfigurations or feedback information is delivered from I2INF analyzer to the network management system.

Related Documents

- I-Ds:
  - draft-ywj-opsawg-i2inf-data-center-networking
  - draft-yao-tsvwg-cco-requirement-and-analysis
  - draft-yao-tsvwg-cco-problem-statement-and-usecases

- Paper from co-authors:
  - ClickINC: In-network Computing as a Service in Heterogeneous Programmable Data-center Networks, https://dl.acm.org/doi/10.1145/3603269.3604835 **SIGCOMM 2023**

- Any comments?