Managing end-to-end resource reservation

Luis Almeida, Moris Behnam, Paulo Pedreiras
Application trends

- **Largely distributed real-time interaction**
  - Remote interactions, interactive multimedia, time-sensitive cloud services

- **Complex structure**
  - High heterogeneity (functionality, requirements, resources…)
  - Variable composition (versions, modes, connections…)

- **Need to be robust with respect to**
  - Topology changes (node crashes, reconfigurations…)
  - Changes in available resources (energy, bandwidth…)
  - Denial-of-service (malfunctioning nodes, malicious actions…)
  - Intrusion (unauthorized accesses or actions…)

**How to design these systems ??**
Real-time capable platforms needed

- **Amenable to modeling of timing behavior**
  - Bounded and computable delays
    → real-time guarantees
- **While supporting multiple and varying**
  - applications, users, operating conditions, ...
- **And being resource efficient**
  - bandwidth, energy
Focusing on the network...

- **Real-Time communication technologies**
  - well developed for (static) DES
  - focused on latency and isolation
- **General purpose communication technologies**
  - well developed for large networks (Cloud/Internet)
  - essentially best-effort (particularly in access networks)
  - focused on openness, scalability and throughput

Unifying effort needed, towards scalable, open and efficient real-time communication
The real-time enabled cloud

Hierarchy of reservations (virtual channels)

Physical world

Node interfaces

Physical world

Traffic segregation and isolation
Related frameworks

- **Hierarchical Scheduling Framework**
  - Generation of efficient interfaces
  - Enforcement protocols in realistic settings
Related frameworks

- **Global resource management frameworks**
  - FRESCOR: Framework for Real-time Embedded Systems
  - (H-)QRAM: (Hierarchical) QoS-based Resource Allocation Model

- **End-to-end resource reservations in networks**
  - RSVP: Resource Reservation Protocol (RSVP)
  - SRP: Stream Reservation Protocol (IEEE 802.1Q-2011)
  - TT-Ethernet, Profinet-IRT, Ethernet POWERLINK
  - AFDX, Ethernet IP, …

*Generalized limited scalability and absence of slack sharing*
Open problems

- Formulating resource requirements
  - Defining adequate interfaces
- Expressing adaptivity
  - In requirements and interfaces
- Scalable adaptive reservations
  - Adapting entangled reservations
- Requirements feasibility
  - Per resource and in whole
- Global admission control
  - And reservations enforcement
- Track and distribute slack

RT-enabled cloud