ÉTS
Le génie pour l'industrie

IRTF Network Management Research Group
55th meeting

Introducing FlexNGIA: A Flexible Internet Architecture for the Next-Generation Tactile Internet

Mohamed Faten Zhani
École de technologie supérieure (ÉTS Montreal)
Canada

Montreal, Canada, 27 July 2019

Le génie pour l'industrie

Outline

- A Glance into the Future
- Limitations of Today's Internet
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
- Use cases/intent
- Conclusion

M. F. Zhani, H. ElBakoury, "*FlexNGIA: A Flexible Internet Architecture for the Next-Generation Tactile Internet*," ArXiV 1905.07137, May 17, 2019
https://arxiv.org/abs/1905.07137

Requirements & Characteristics

- Future Applications: Telepresence, VR/AR, Holoportation
- Requirements:
  - High processing power: real-time processing
  - High bandwidth (e.g., VR (16K, 240 fps) ➔ 31.85 Gbps)
  - Ultra-low Latency: 1ms to 20ms
  - Multi-flow synchronization
  - High availability

Requirements & Characteristics

- Octopus-like applications: huge number of flows for each application
- Changing requirements : requirements can change over time

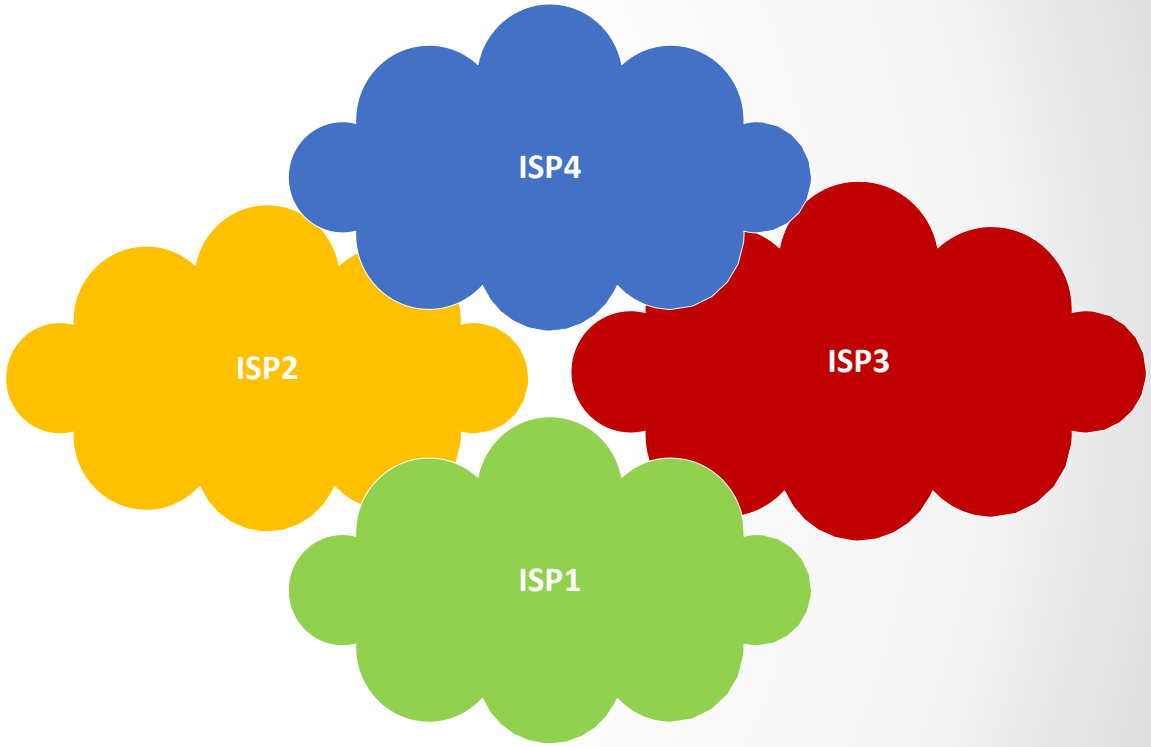
ÉTS
Le génie pour l'industrie

Outline

- A Glance into the Future
- Limitations of Today's Internet
  - Internet Infrastructure and Services
  - Network Stack Layers and Headers
  - Sources of Latency
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
- Use cases/intents
- Conclusion

Internet Infrastructure and Services

- A network of networks
- Offered service: "Best effort" data delivery.. no more
- No control over the infrastructure
  - ➔No control over the end-to-end path and quality of service
  - ➔No performance guarantees



Transport Layer Protocols

Many modern protocols like SCTP and QUIC but let's focus first on TCP..

- One-size-fits-all service offering: TCP offers reliability, data retransmission, congestion and flow control
- Blind Congestion control
- The two end points limitation:
  - High retransmission delays (~ 3x e2e delay)
  - Transport and network layers are not aware which flows belong to the same application

Network Layer Protocols

- Not aware of the applications
  - The application composition (in terms of flows)
  - Performance requirements of each of these flows and how these requirement change over time

  ➔Drop packets « blindly »
- No collaboration with the transport layer
  - Do not provide explicit feedback or support to transport layer (maybe ECN is interesting but it is not enough)
  - Do not help with other transport services (e.g., reliability)

Network Stack header

Problems with current headers:

- Do not provide additional informations about objects/sensors, flows belonging to the same application, applications' requirements, etc.
- Not flexible enough: It is not easy to incorporate meta-data and commands

ÉTS
Le génie pour l'industrie

Outline

- A Glance into the Future
- Limitations of Today's Internet
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
  - Future Internet Infrastructure and services
  - Business Model
  - Management Framework
  - Network Protocol Stack/Functions
  - Stack Headers
- Use cases/intents
- Conclusion

FlexNGIA:a Flexible Internet Architecture for the Next-Generation Tactile Internet

FlexNGIA

Computing resources

- In-Network Computing: any function anywhere

Business model

- Multiple source destination Service Function Chains
- Stringent performance requirements

Cross-layer Design (Transport+Network)

- Breaking the end-to-end paradigm
- In-network advanced transport functions
- Better congestion control
- Stringent performance and reliability guarantees

Application-Aware Network Management

- Advanced functions tailored to applications
- App-aware traffic engineering

Flexible headers

- Tailored to the application

Future Internet Infrastructure and Services

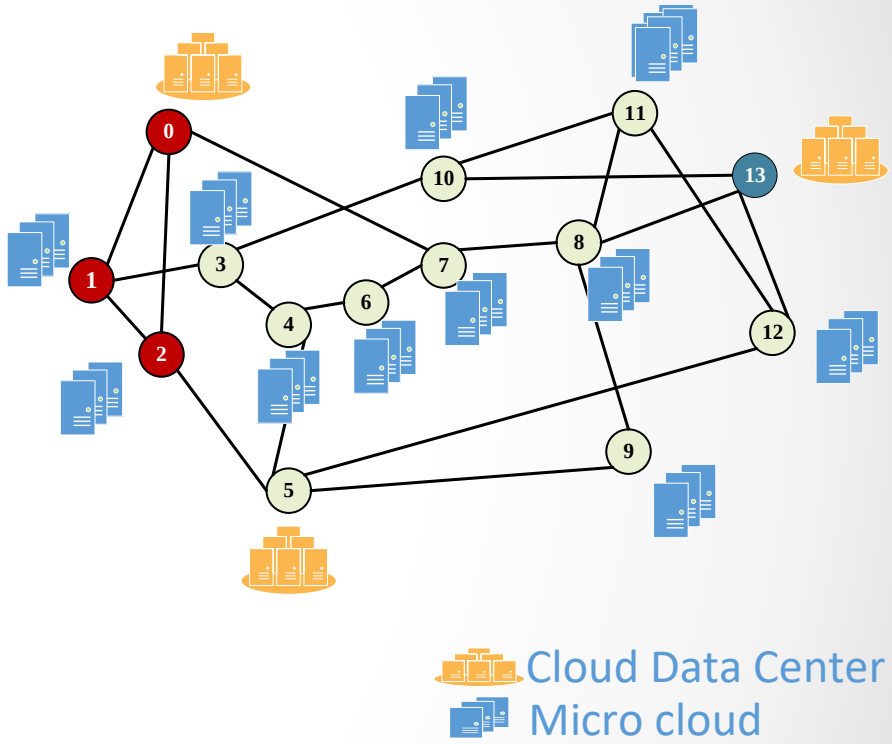
How a network will look like?

- Computing resources are everywhere: Available at the edge and at the core of the network
- Commodity servers but also dedicated hardware, FPGA, GPU, NPU, etc.

➔ In-Network computing

➔ Reduce steering delay

➔ Full Programmability: Any function could be provisioned anywhere (virtual machines/containers)



Future Internet Infrastructure and Services

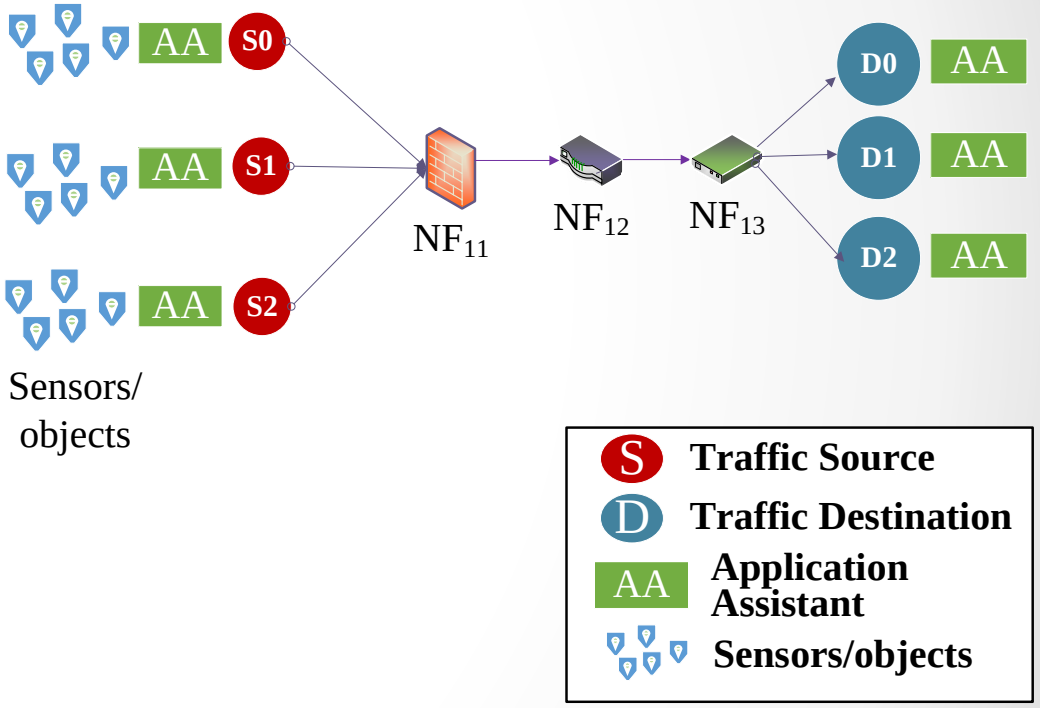
How does Future Internet look like?

- Still a network of networks..
- What is new?
  - More services: Service Function chains
    ➔More advanced functions
    ➔More than just delivery
  - Stringent performance guarantees

Future Internet Infrastructure and Services

Application-Driven Intent:
Service Function Chain (SFC)

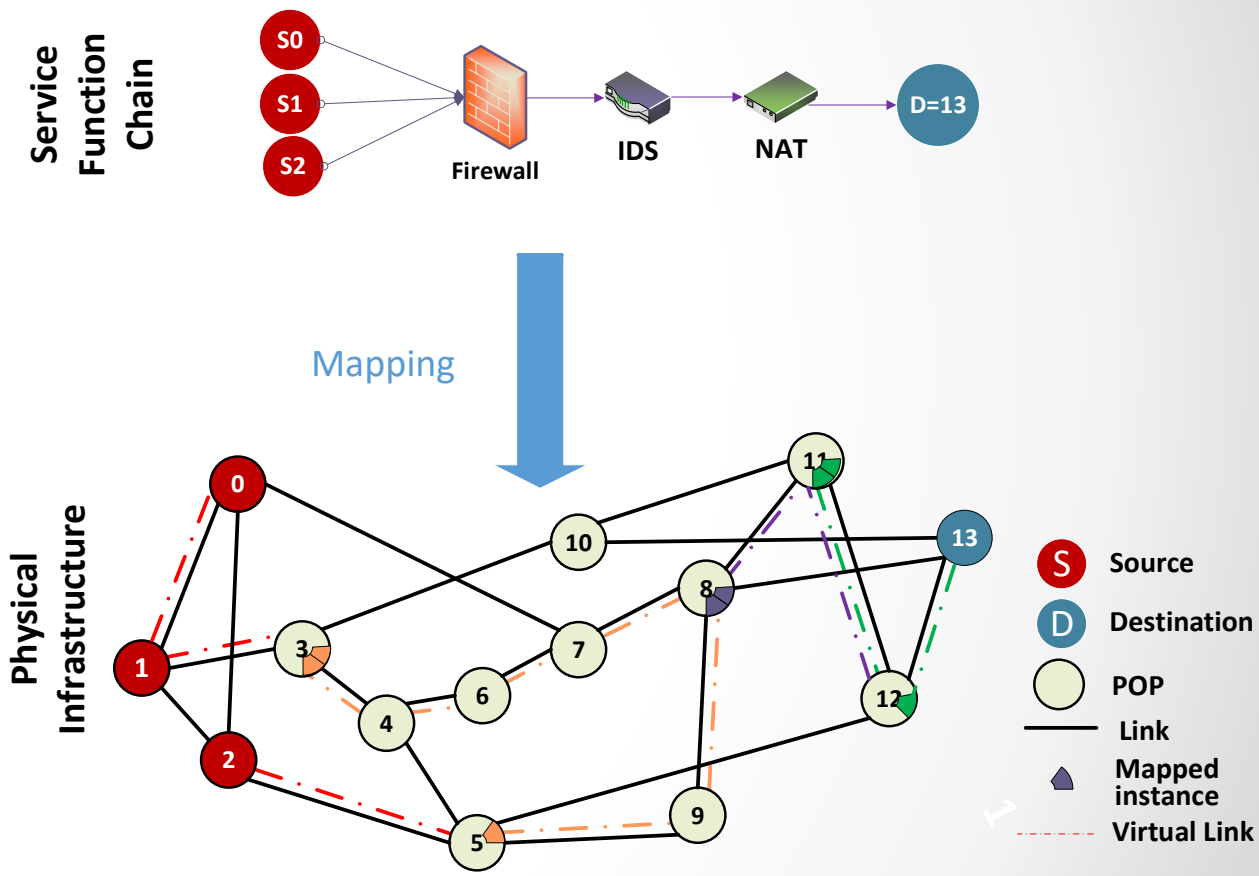
- Multiple connected network functions
- Multiple sources and destinations
- Made out from Network Functions
- Defines, for each network function, the type, software, input/output packet format, expected processing delay, buffer size
- Defines performance requirements (e.g., throughput, packet loss, end-to-end delay, jitter)



Business Model

Network Operators

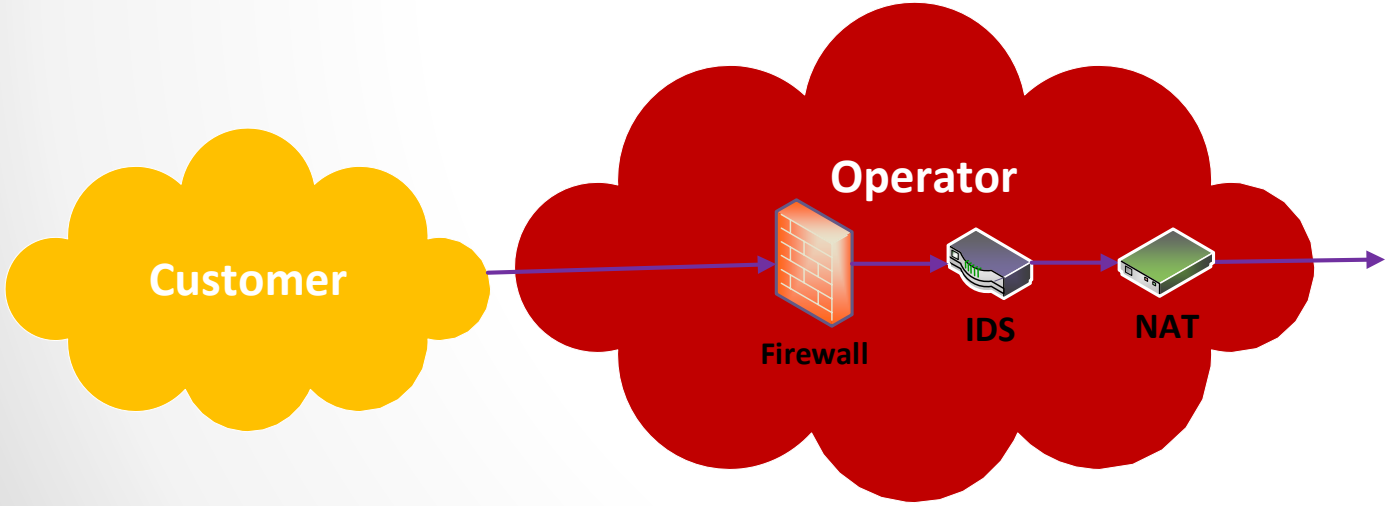
- Own and manage the physical infrastructure (i.e., one network)
- Deploy platforms and software required to run network functions
- The service could be simply data delivery or a SFC
- Provision and manage SFCs



Business Model (cont)

Customers

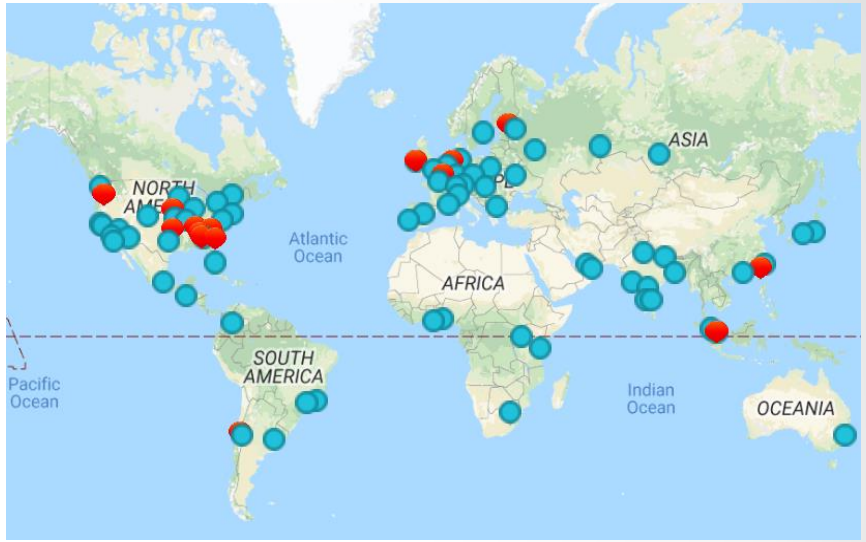
- Could be other network operators, companies or Institutions
- Define the required SFC and Identify the chain sources/destinations
- Rely on the operator to provision and manage the SFC and satisfy SLA

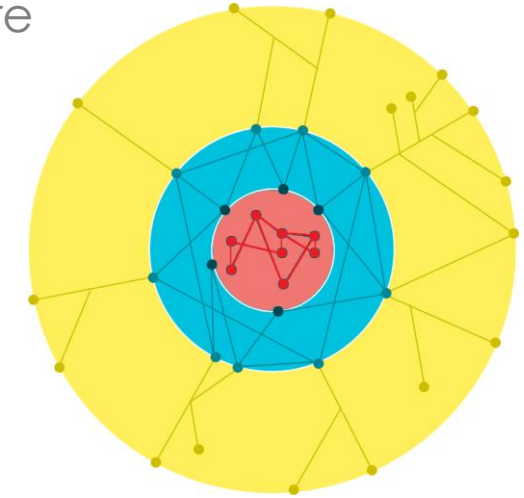


- SFC composition
- SLA requirements for the SFC
  - Bandwidth
  - End-to-end delay
  - Reliability, availability
- SLA requirements for each NFs
  - Processing power
  - Packet format(s)
  - Packet drop criteria…

Business Model (cont)

- Example of potential Network Operators:
  - ISPs (e.g., AT&T or Bell Canada) and web-scale companies (e.g., Google, Facebook, Amazon)
  - Example: Google Cloud Platform
    - World wide global Infrastructure
    - Software defined platform
    - Full control over the infrastructure





- 15 Data centers
- 100 Points of Presence (PoPs)
- 1000+ Edge nodes

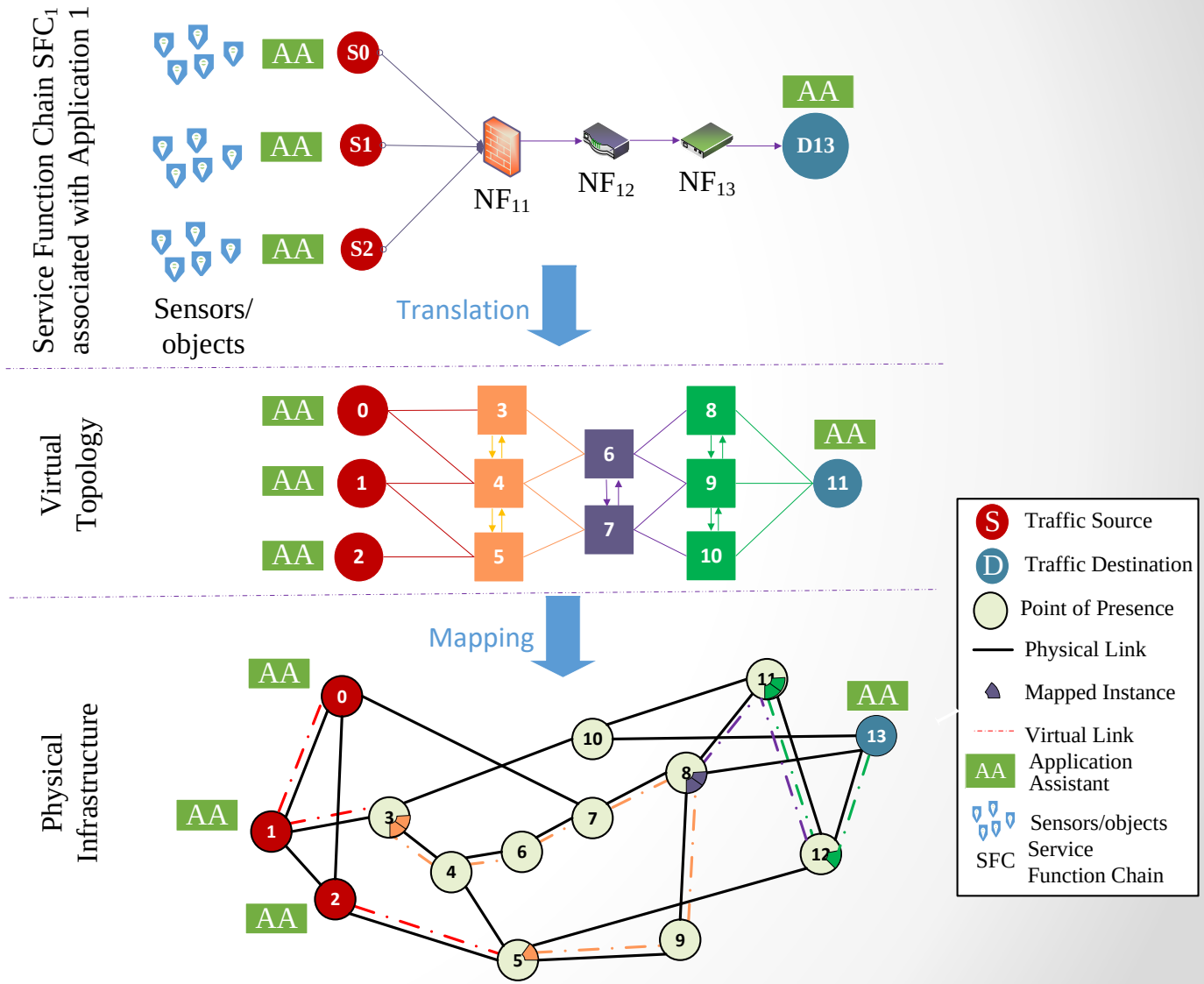
M. F. Zhani, H. ElBakoury - FlexNGIA 2019 (https://arxiv.org/abs/1905.07137)

Source: cloud.google.com

Resource Management Framework

Resource Allocation

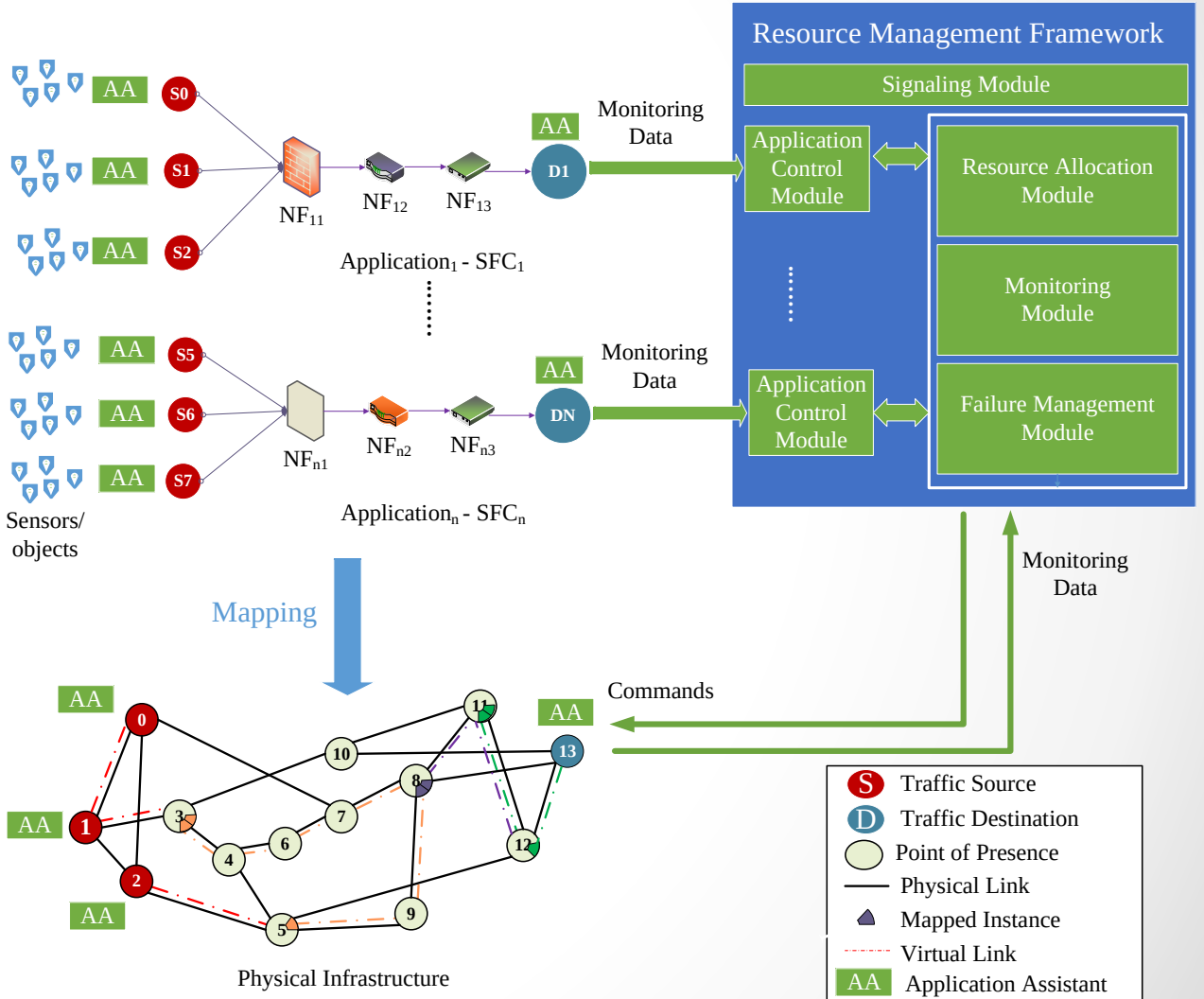
- The Service Function Chain (SFC) is defined by the application designer
- 2-step resource allocation:
  - Translation: the SFC is translated into a virtual topology
  - Mapping: virtual topology are mappa



Resource Management Framework

Main components:

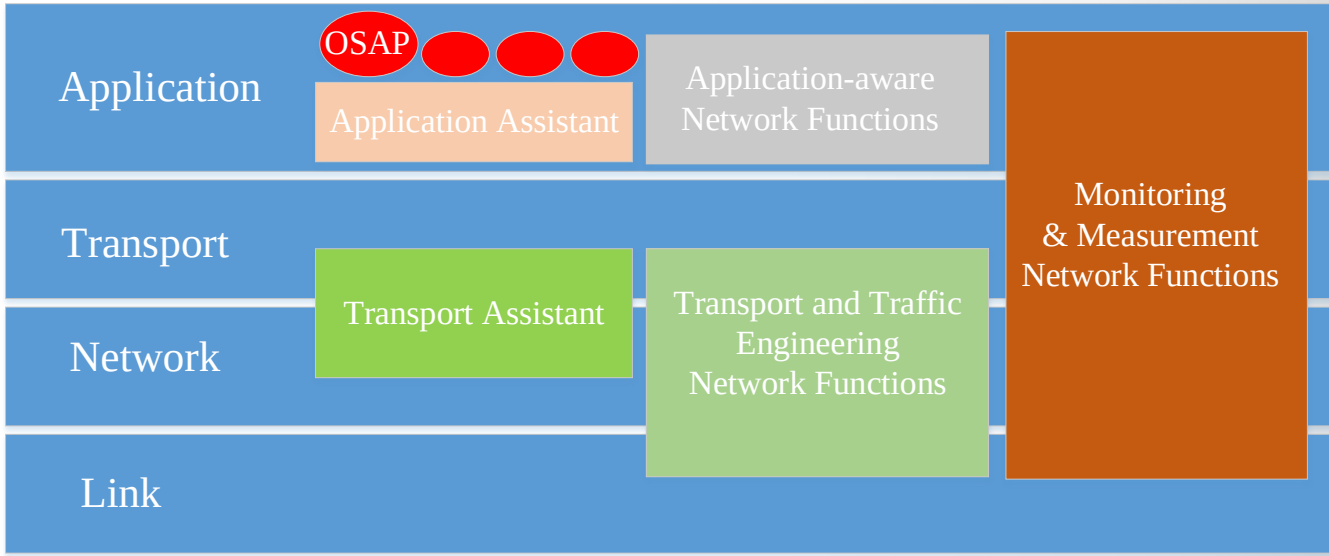
- Signaling module
- Application Control Module
- Ressource allocation Module



NB: For simplicity, the figure shows only the mapping of the chain SFC_1 associated to Application 1

Network Protocol Stack/Functions

- Basic Network Functions (e.g., packet forwarding)
- Advanced Network Functions:
  - Could operate at any layer
  - Only limited by our imagination
  - Examples: packet grouping, caching and retransmission, data processing (e.g., image/video cropping, compression, rendering, ML), application-aware flow multiplexing (e.g., incorporating/merging data)



➔ Functions could break the end-to-end principle

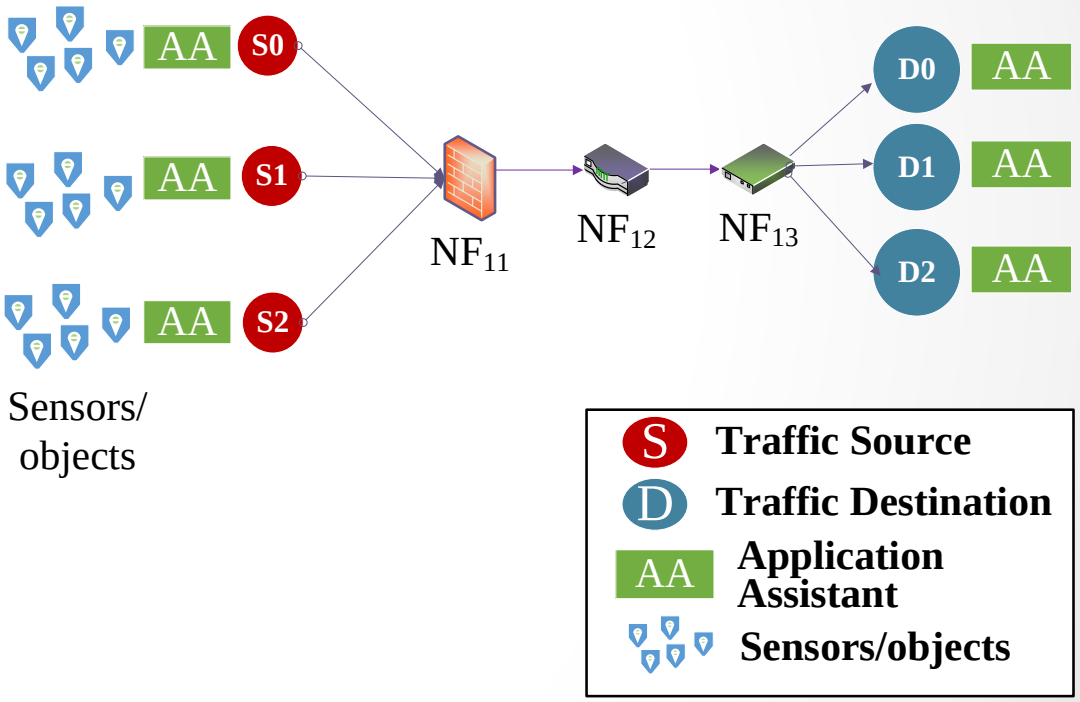
➔ SDN++: SDN should go beyond configuring forwarding rules and should provide the ability to dynamically configure these new functions

Network Protocol Stack/Functions Application Assistant

Application Assistant (AA)

- One AA at each end-point
- Interfaces with objects/sensors
- Measures the application performance and user QoE
- Identifies the applications' requirements at run-time
- Adds additional metadata To be used by subsequent Network Functions

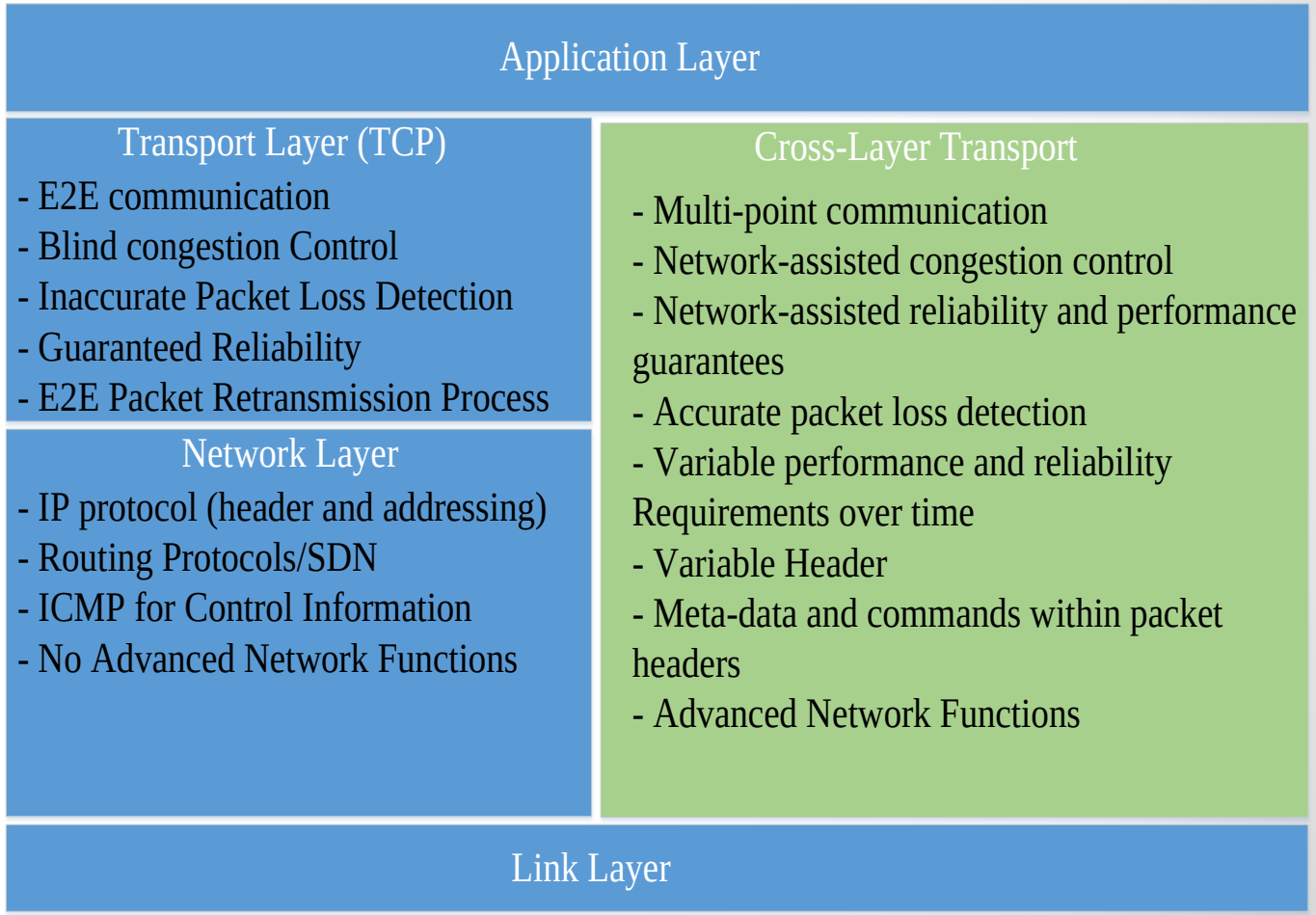
➔ Application-Aware Network Services



Network Protocol Stack/Functions
Transport Assistant

Transport Assistant (TA)

- A cross-layer Network Function
- Combines services of the transport and network layers
- Manages all the flows of the same application
- Implements Transport/Network functions (e.g., congestion control, packet loss detection, packet cache and retransmission, routing)
- One or multiple TA could be provisioned in the same SFC



ÉTS
Le génie pour l'industrie

Network Stack Headers

- Signaling packets
  - Instantiate an application
  - Convey application requirements
- Data packets: carry data
  - Layer 2 header: contains mainly the application id used for packet forwarding (similar to VLANs)
  - Upper layers:
    - Fully flexible header format (customizable meta-data and commands)
    - Defined depending on the application
    - Network functions should be aware of the expected format

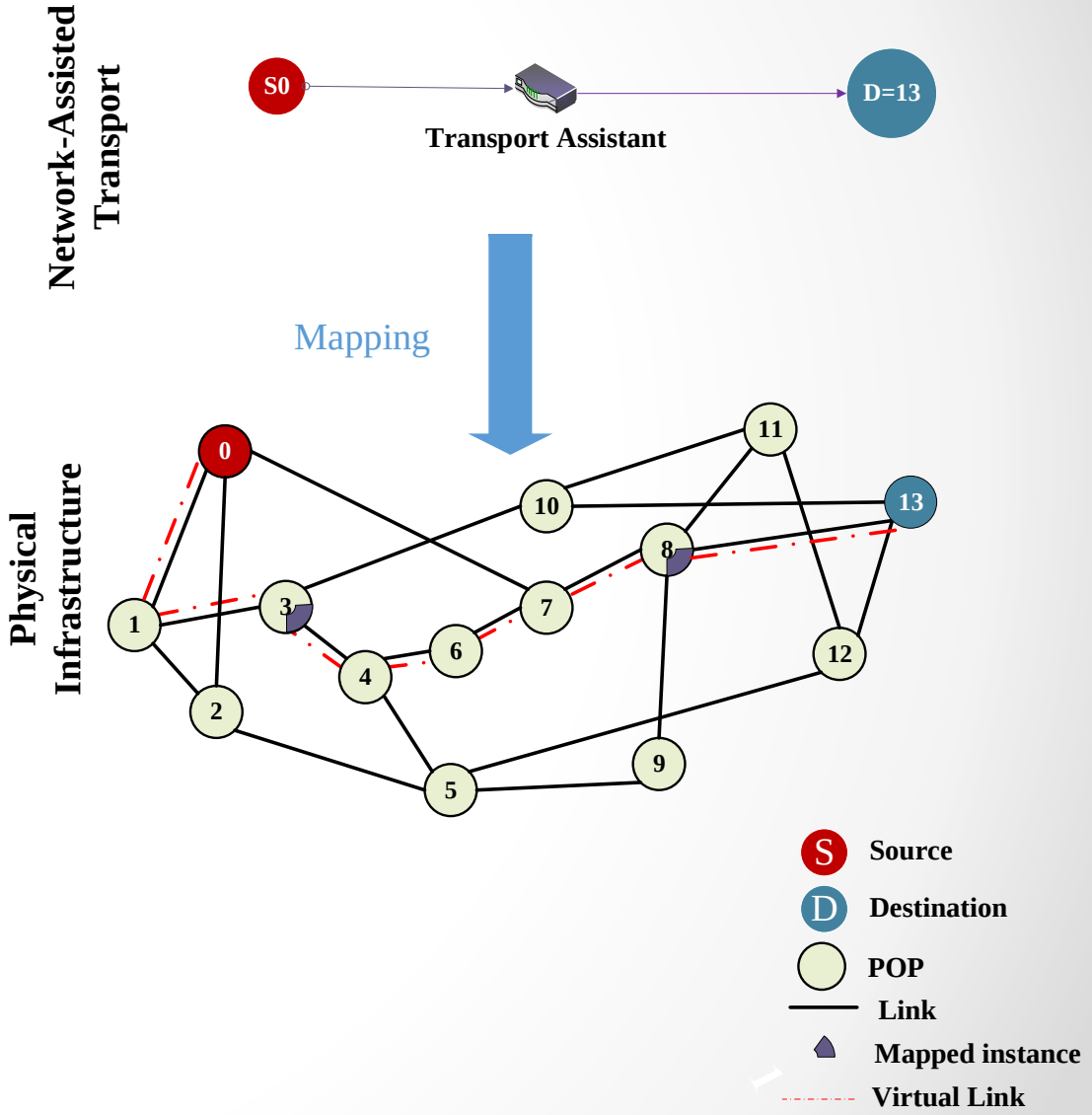
ÉTS
Le génie pour l'industrie

Outline

- A Glance into the Future
- Limitations of Today's Internet
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
- Use Cases scenarios/intents
- Conclusion

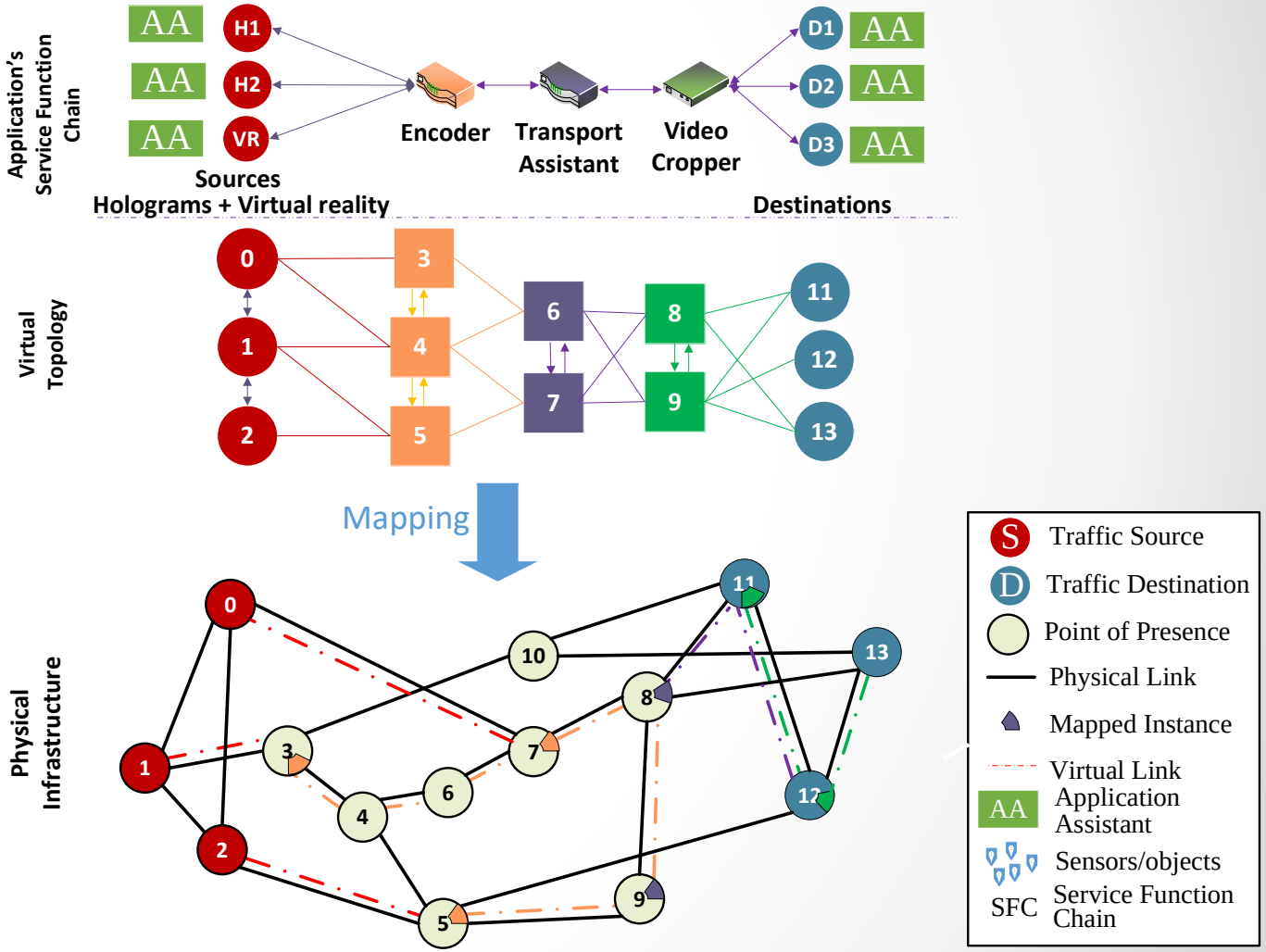
Network-Assisted Data Transport

- Goal
  - Minimize retransmission delay
  - Improved congestion control
- Solution: service chain with a "transport Assistant" function
- Service of the Transport Assistant:
  - Caching and retransmissting packets
  - Detecting packet loss
  - Congestion control: adjusting rate, dropping packets, compression



Mixed Virtual Reality and Holograms

- Users are exploring a virtual reality environment with several human holograms and objects
- Challenges
  - How many intermediate functions?
  - What kind of functions?
  - How the traffic should steered from the flow sources?
  - How many instances for each function?
  - Where to place them?
- Example of deployement
  - Encoder: encode and compress video
  - Transport manager: congestion control
  - Video cropper: crop 3D objects



ÉTS
Le génie pour l'industrie

Outline

- A Glance into the Future
- Limitations of Today's Internet
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
- Use Cases
- Conclusion

Conclusion

FlexNGIA

Computing resources	Business model	Cross-layer Design (Transport+Network)	Application-Aware Network Management	Flexible headers
- In-Network Computing: any function anywhere	- Multiple source destination Service Function Chains - Stringent performance requirements	- Breaking the end-to-end paradigm - In-network advanced transport functions - Better congestion control - Stringent performance and reliability guarantees	- Advanced functions tailored to applications - App-aware traffic engineering	- Tailored to the application

Looking for More Details?

- M. F. Zhani, H. ElBakoury, "*FlexNGIA: A Flexible Internet Architecture for the Next-Generation Tactile Internet*," ArXiV 1905.07137, May 17, 2019 https://arxiv.org/abs/1905.07137

Le génie pour l'industrie

Thank You

Questions