Le génie pour l'industrie

2nd Visions for Future Communications Summit
Technologies and Services Towards 6G

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

Mohamed Faten Zhani

École de technologie supérieure (ÉTS Montreal)

Canada

Lisbon, Portugal, November 28, 2019

Outline

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

- Keynote at ACM SIGCOMM 2019 Workshop on Networking for Emerging Applications and Technologies (NEAT 2019)
- 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

ÉTS
Le génie pour l'industrie

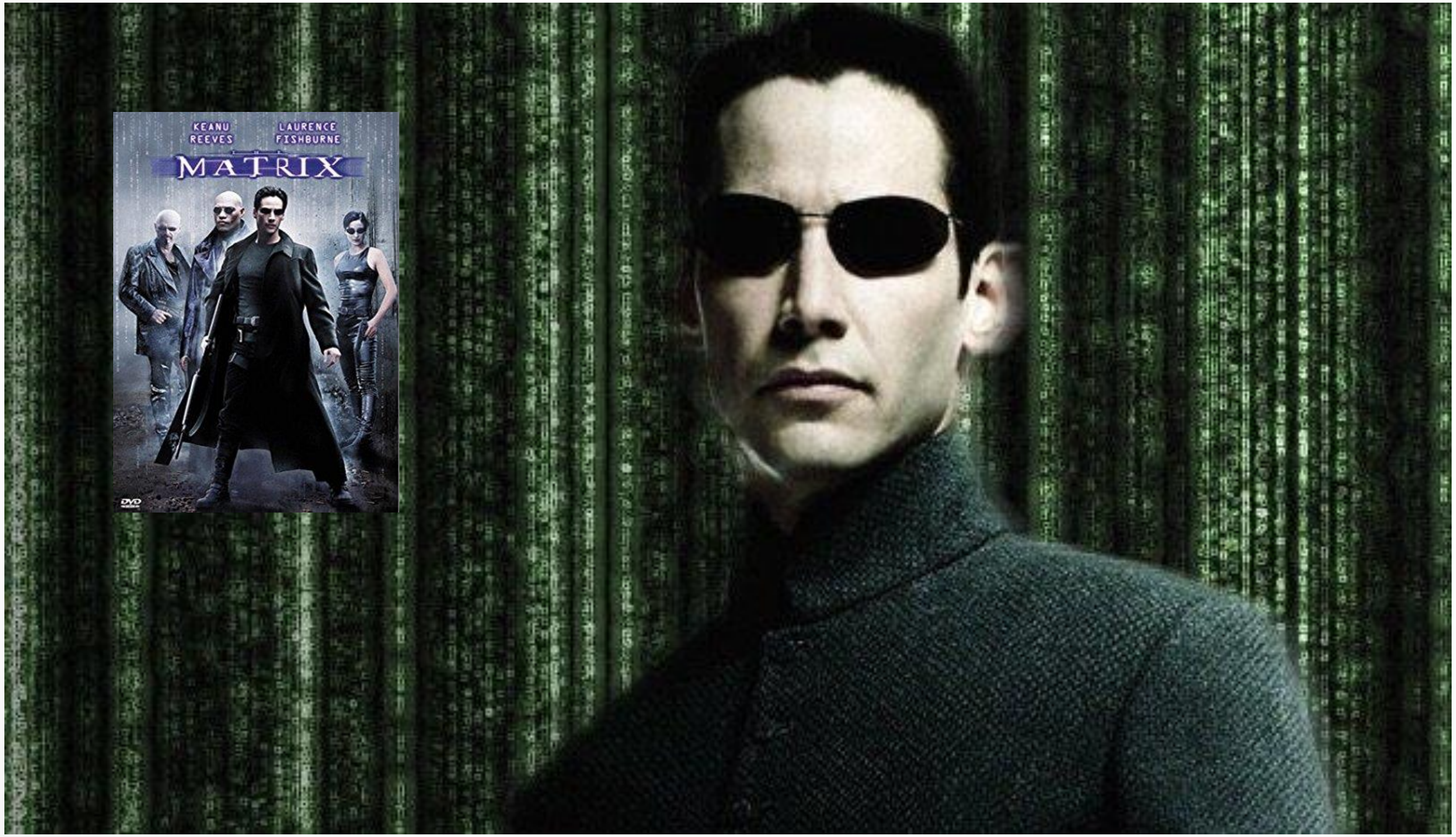
A Glance into the Future

Future Applications

- Telepresence
- Virtual Reality
- Augmented Reality
- Holoportation
- Haptics
- …

Loading…

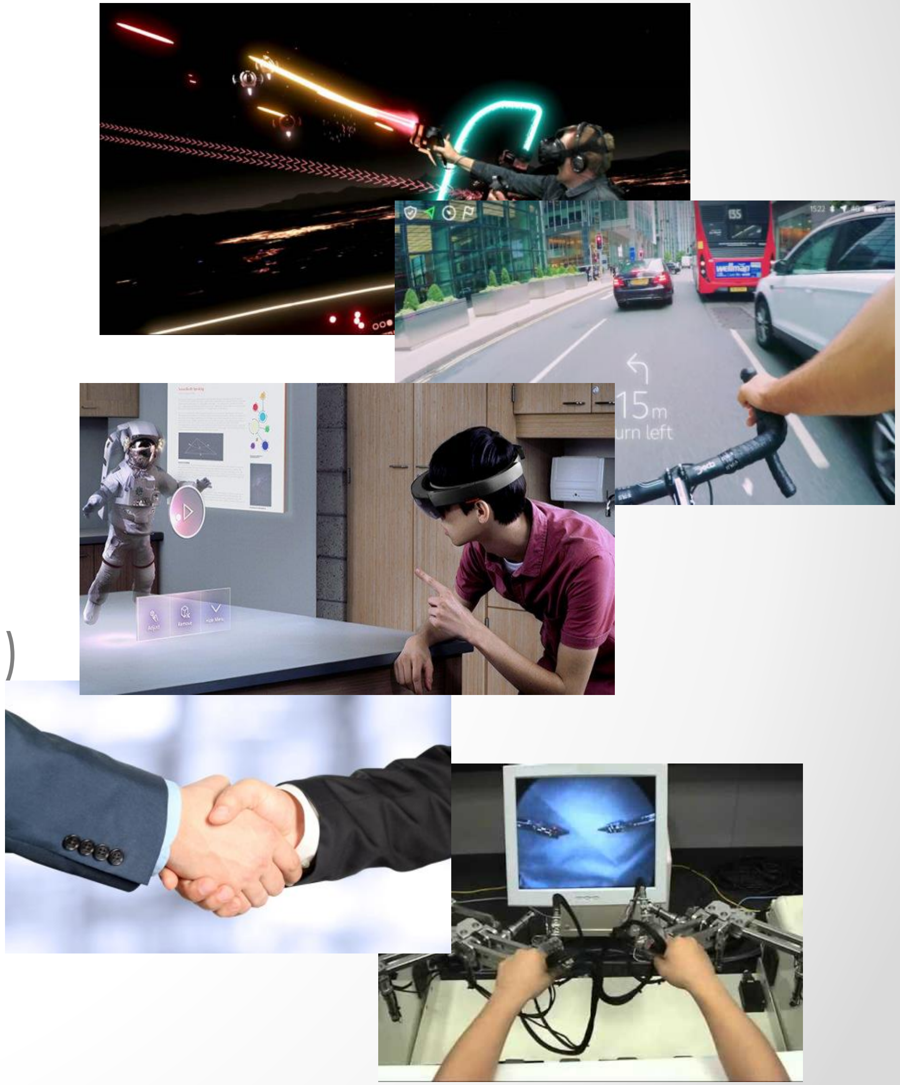
Welcome to the Matrix



Future Applications Requirements & Characteristics

Le génie pour l'industrie

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



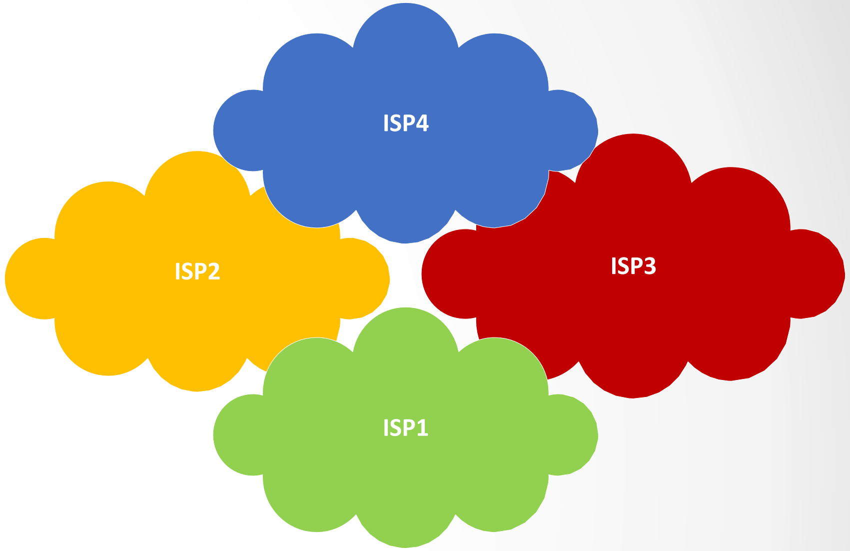
M. F. Zhani, H. ElBakoury - FlexNGIA 2019 (https://arxiv.org/abs/1905.07137)

Outline

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

Le génie pour l'industrie

Today's Internet Limitations



- A network of networks
  - No control over E2E performance
  - Offered service: "Best effort"→ data delivery.. no more
- Transport Layer Protocols
  - One-size-fits-all service offering: TCP offers reliability, packet retransmission, congestion and flow control
  - Blind congestion control
  - The two end points limitation
  - Same limitations are inherited by QUIC and SCTP

Today's Internet Limitations (2)

- Network layer protocols and services
  - Not aware of the applications characteristics and requirements (which flow belongs to which app?, priorities?)
  - No collaboration with upper layers (transport, application)
    - The network knows better about the congestion (e.g., location, severity)
    - The network could help with reliability (e.g., retransmission)…
- Packet header
  - Not easy to include metadata or commands
  - Whose is going to use such metadata/commands?

Outline

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

Le génie pour l'industrie

FlexNGIA

FlexNGIA

Computing resources

- In-Network Computing: any function anywhere

Business model

- Network providers offer Service Function Chains/meshes
- 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 Services

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

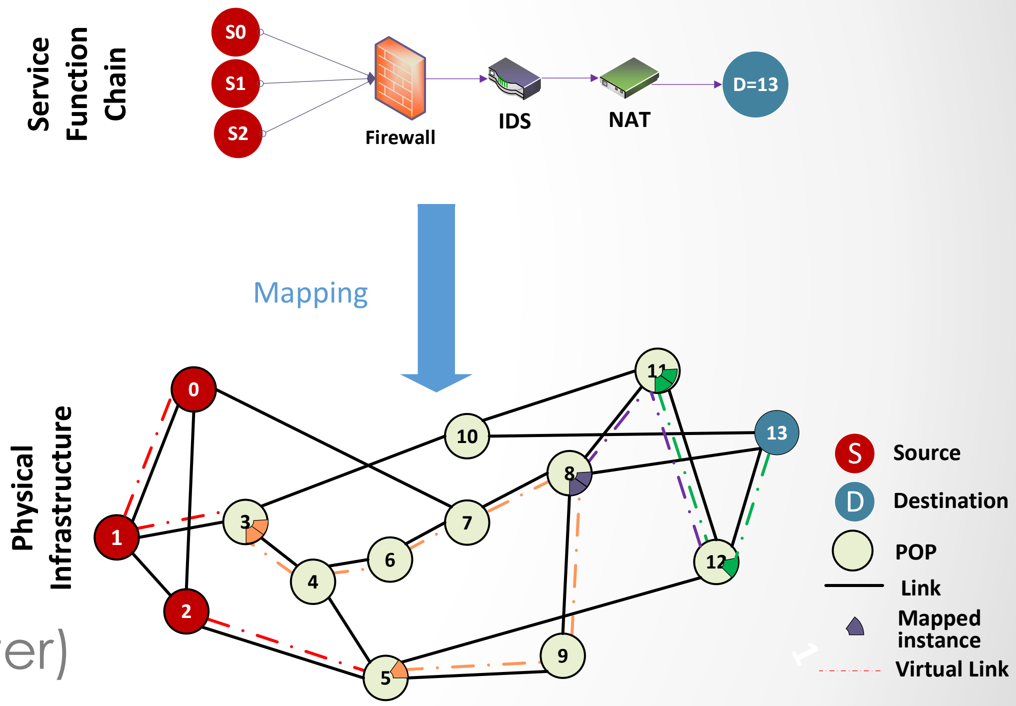
Flexible headers

- Tailored to the application

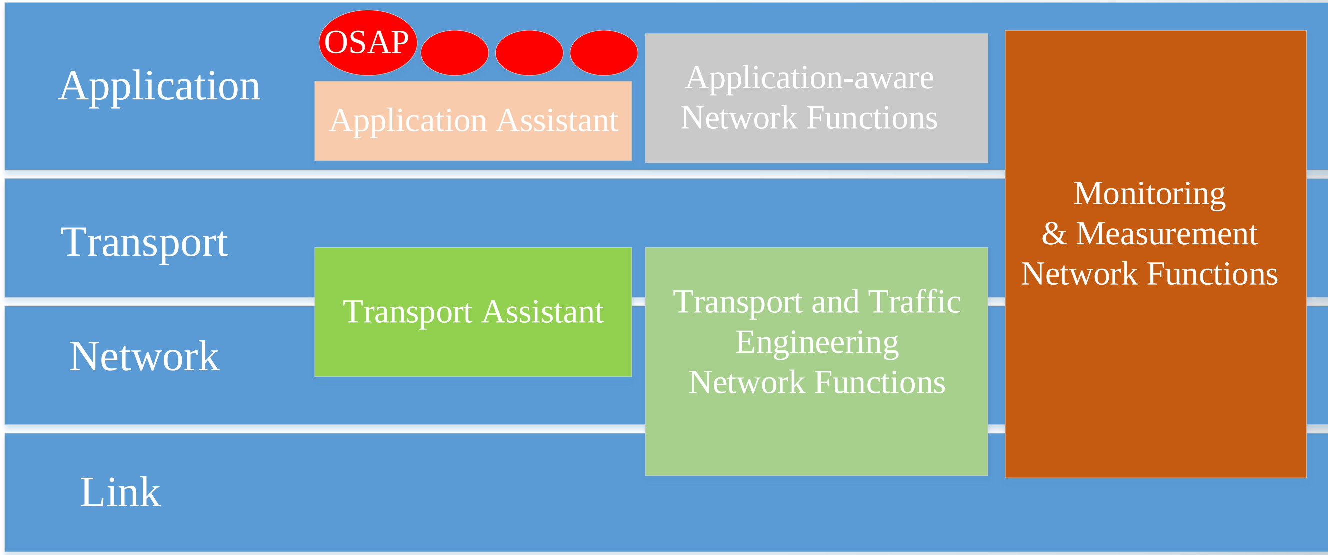
Future Internet Services

Service Function Chain/Mesh (SFC):

- Multiple sources and destinations
- Define your Network Functions: type, software, behavior, input/output packet format, expected processing delay, buffer size
- Define communication protocols (layer 3 and above)
- Define performance requirements (e.g., throughput, packet loss, end-to-end delay, jitter)



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

Outline

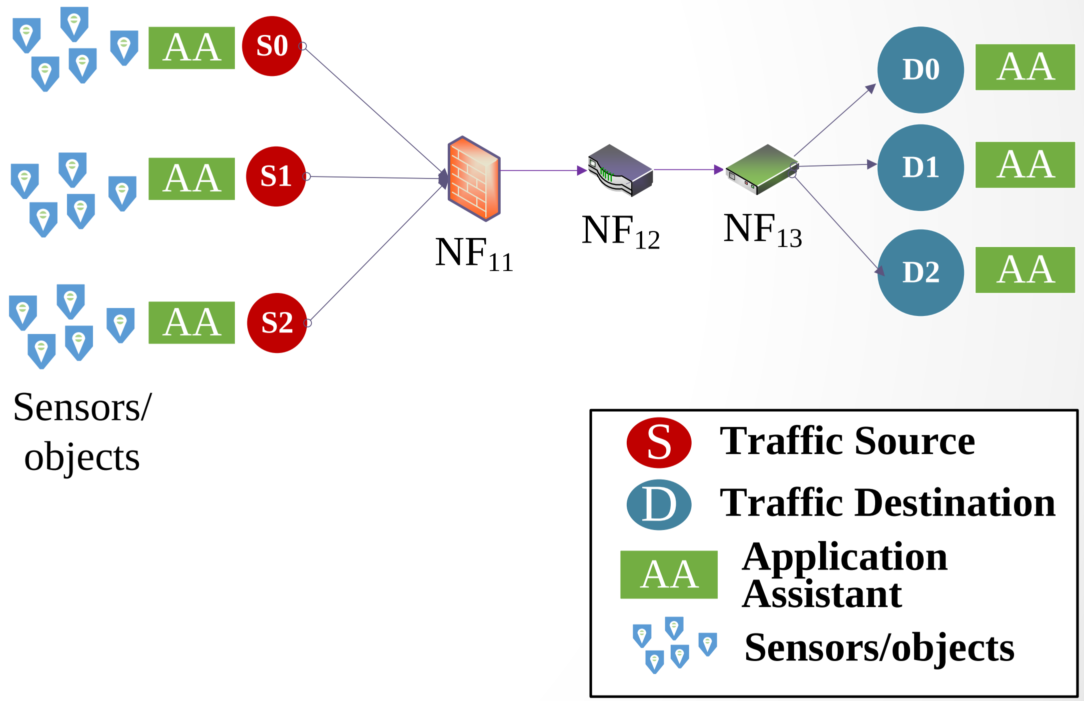
- A Glance into the Future
- Limitations of Today's Internet
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
- Use Cases
- Research Challenges

Application-Aware Network Services

Application Assistant (AA)

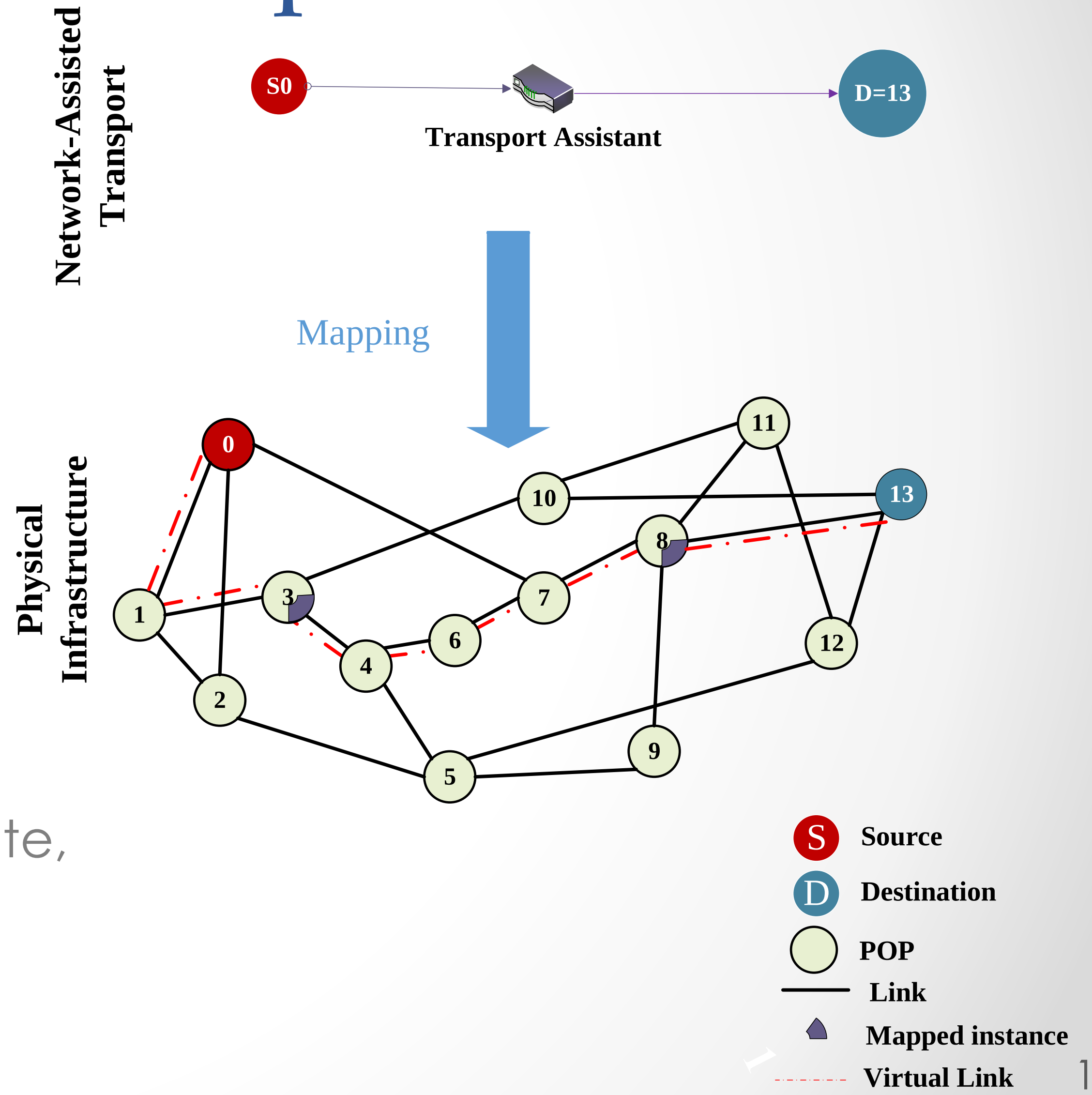
- One AA at each end-point
- Interfaces with objects/sensors
- Measures the application performance and user QoE
- Identifies the applications' requirements at run-time
- Adds additional metadata to be used by subsequent Network Functions

➔ Application-Aware Network Services



Network-Assisted Reliable Data Transport

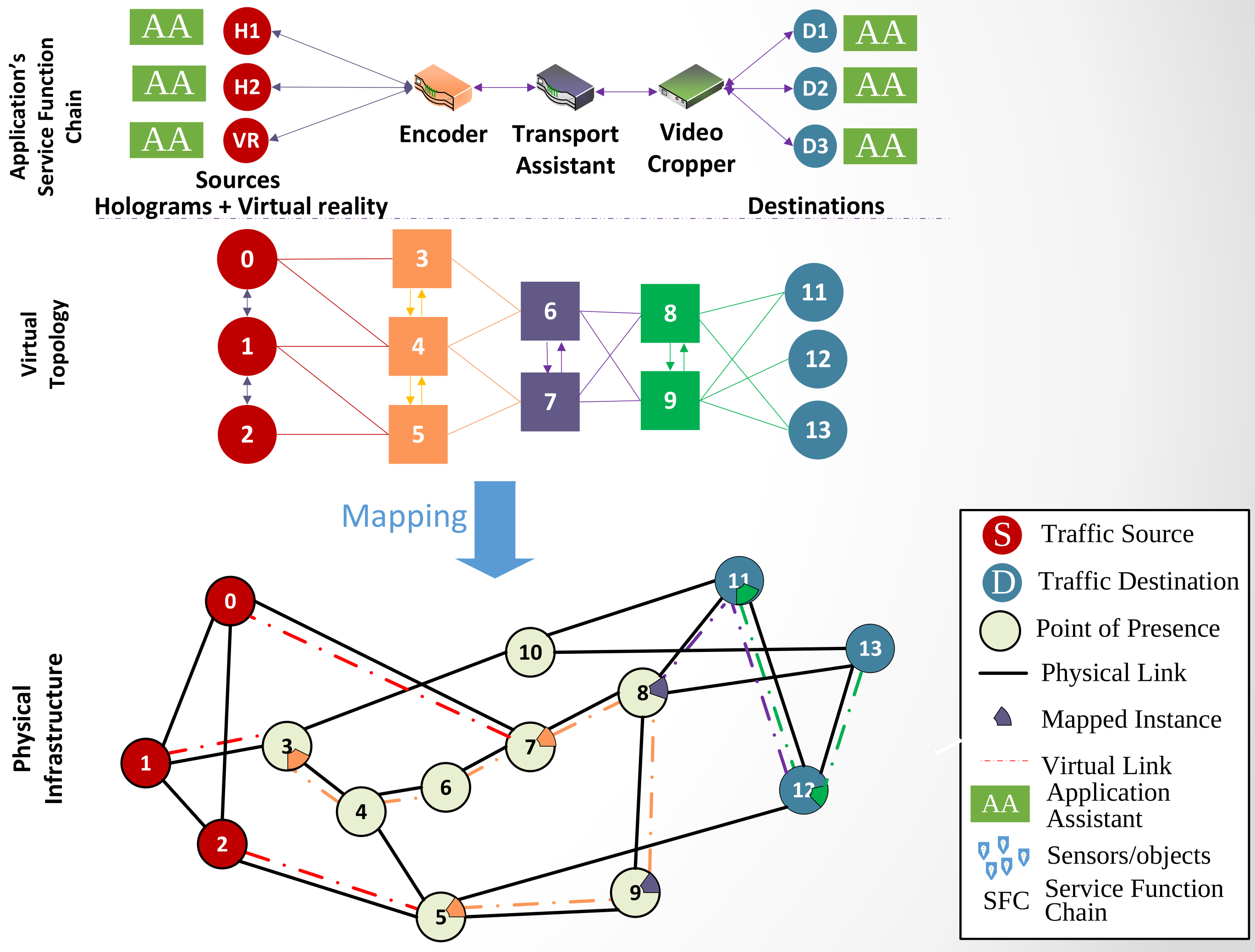
- Goal
  - Minimize retransmission delay
  - Improved congestion control
- Solution: service chain with a "transport Assistant" function
- Services of the Transport Assistant:
  - Combines transport and network layers
  - Cache and retransmit packets
  - Detect packet loss
  - Routing and congestion control: adjusting rate, dropping packets



Le génie pour l'industrie

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

Research Challenges

- Designing Service Function Meshes tailored to applications
- High-performance softwarized functions
- Signaling
- Slicing/Resource Allocation
- Fault-tolerance and Failure Management
- High-precision and fine-grained monitoring and measurements
- SDN++
- Distributed cross-layer transport protocol (socket, caching, communication)
- Security and Privacy

Looking for More Details?

- Mohamed Faten Zhani, https://profs.etsmtl.ca/mfzhani/
- FlexNGIA Project: 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
- Keynote at ACM SIGCOMM 2019 Workshop on Networking for Emerging Applications and Technologies (NEAT 2019)

École de technolgie supérieure (ÉTS Montreal)
University of Quebec, Canada

Thank You

Questions

Le génie pour l'industrie

Le génie pour l'industrie

Le génie pour l'industrie

ACM SIGCOMM 2019 Workshop on Networking for Emerging Applications and Technologies (NEAT 2019)

FlexNGIA
A Fully Flexible Novel Architecture for the Next-Generation Tactile Internet

Mohamed Faten Zhani

École de technologie supérieure (ÉTS Montreal)

University of Quebec

Canada

Beijing, China, 19 August 2019

Outline

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

Le génie pour l'industrie

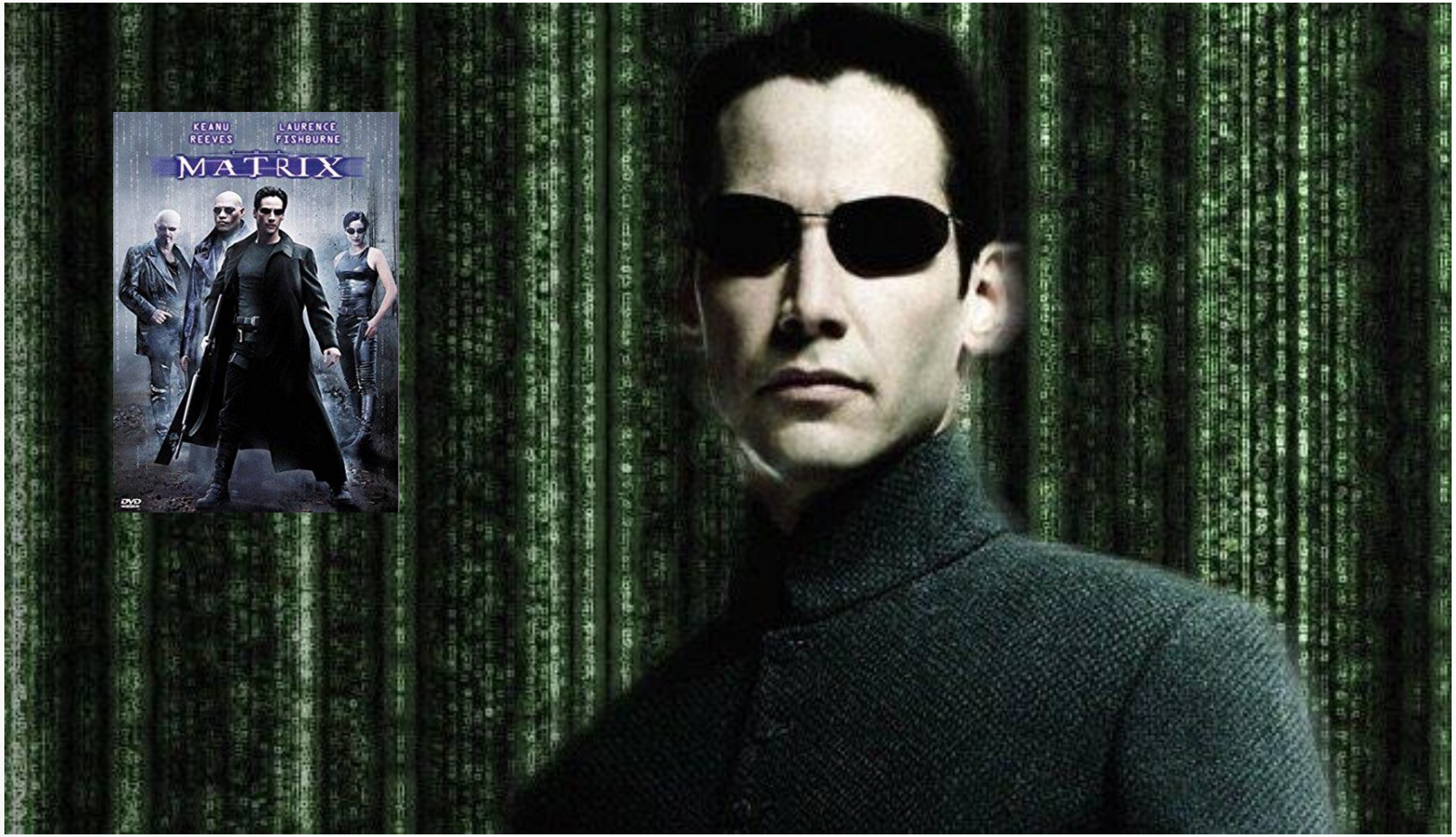
A Glance into the Future

Future Applications

- Telepresence
- Virtual Reality
- Augmented Reality
- Holoportation
- Haptics
- …

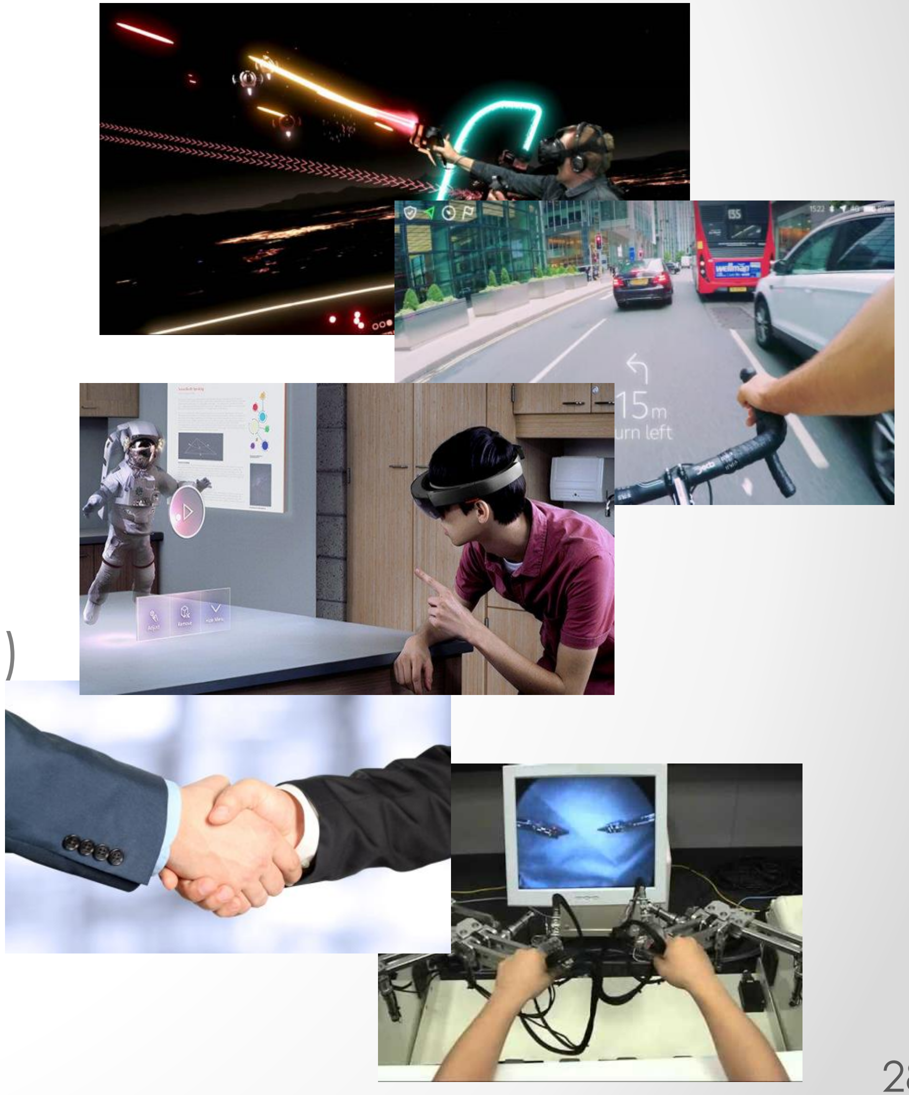
Loading the Matrix…

Welcome to the Matrix



Future Applications Requirements & Characteristics

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



Outline

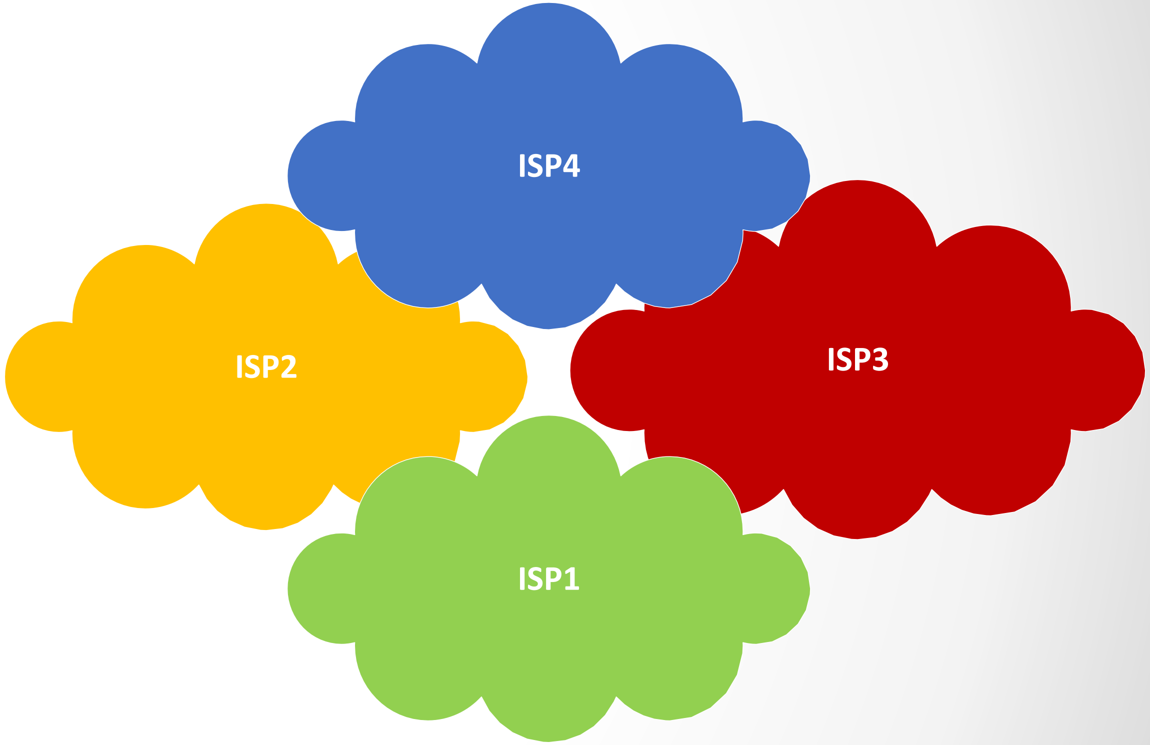
- A Glance into the Future
- Limitations of Today's Internet
  - Internet Infrastructure and Services
  - Network Stack Layers and Headers
- FlexNGIA: Fully-Flexible Next-Generation Internet Architecture
- Use cases
- Conclusion

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



Le génie pour l'industrie

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

Le génie pour l'industrie

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

Le génie pour l'industrie

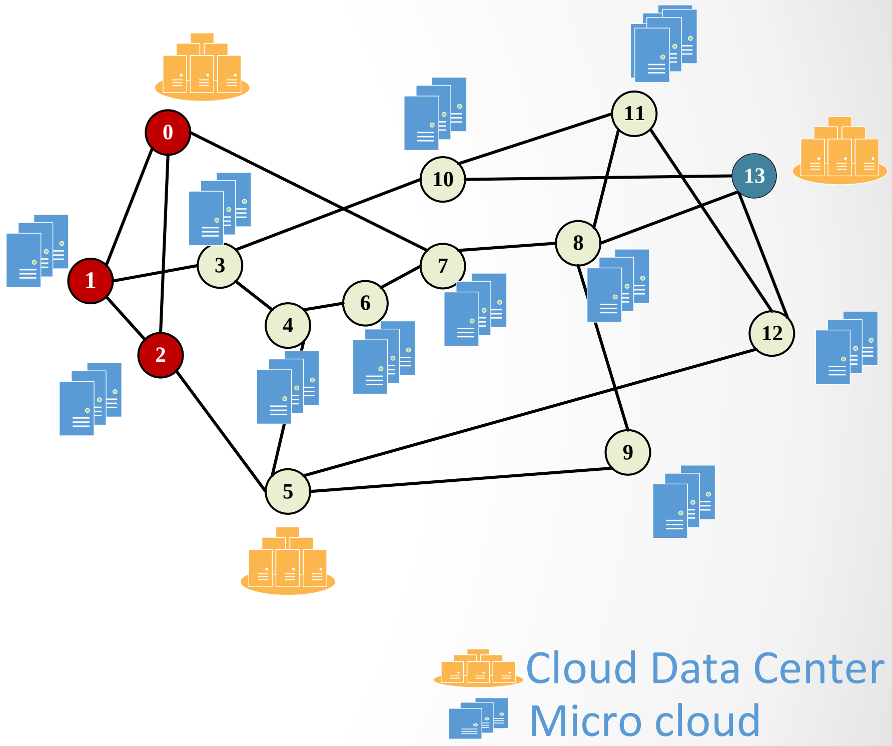
Outline

- 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
- Conclusion

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

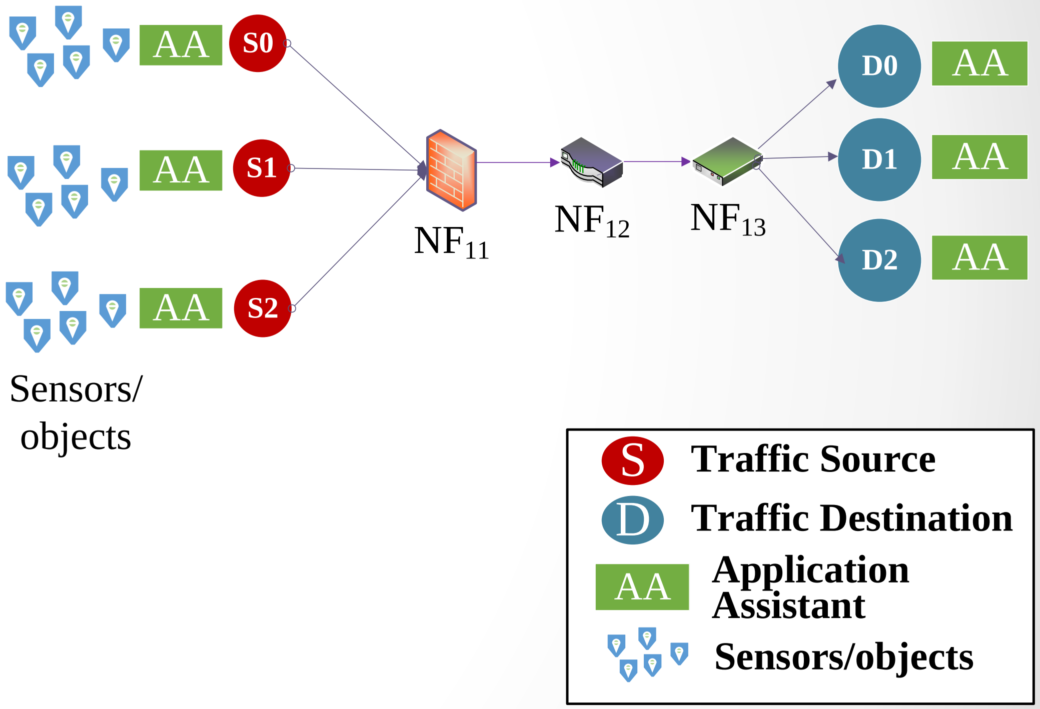
How does Future Internet look like?

- 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

Service Function Chain (SFC)

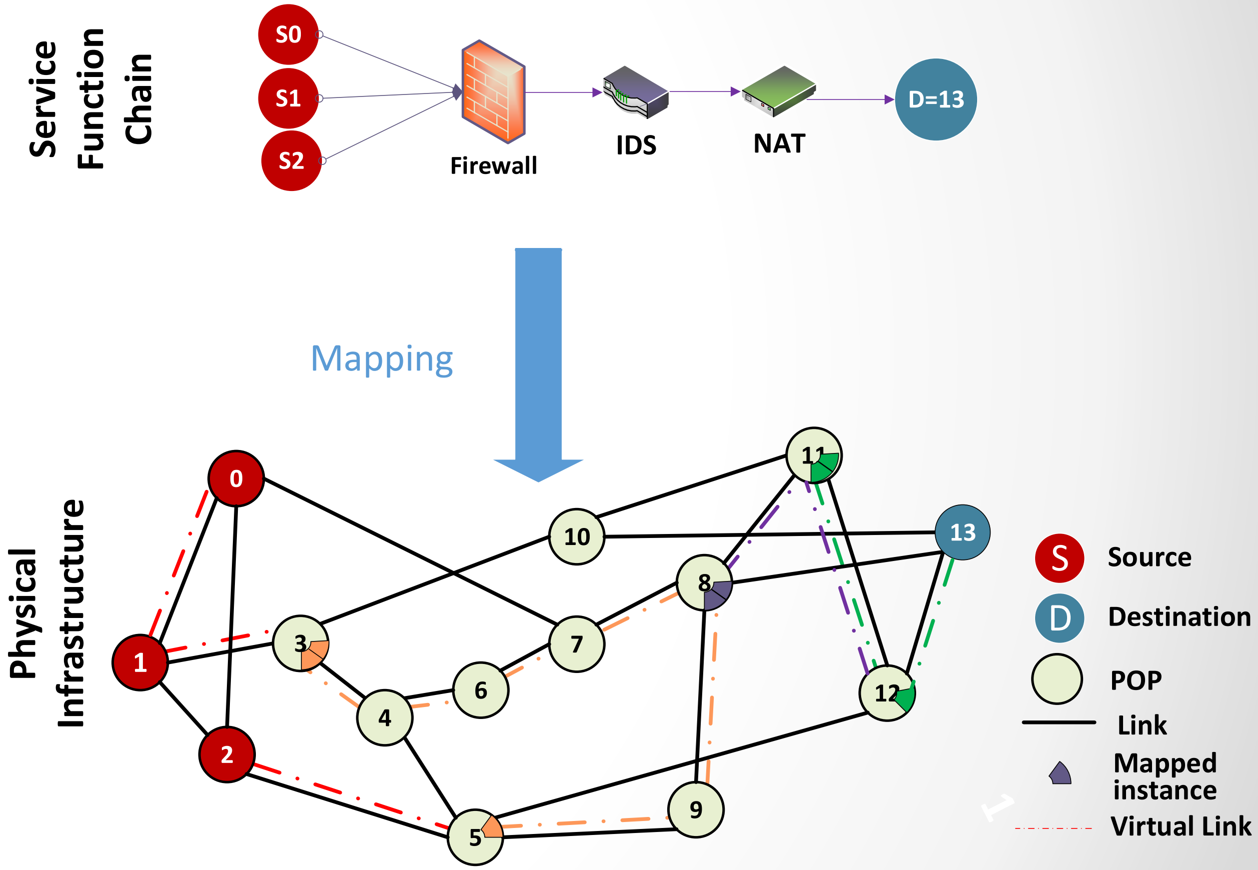
- Multiple connected network functions
- Multiple sources and destinations
- Made out from Network Functions
- Defines, for each network function, the type, software, input/output packet format, expected processing delay, buffer size
- Defines performance requirements (e.g., throughput, packet loss,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



Business Model

Network Operators

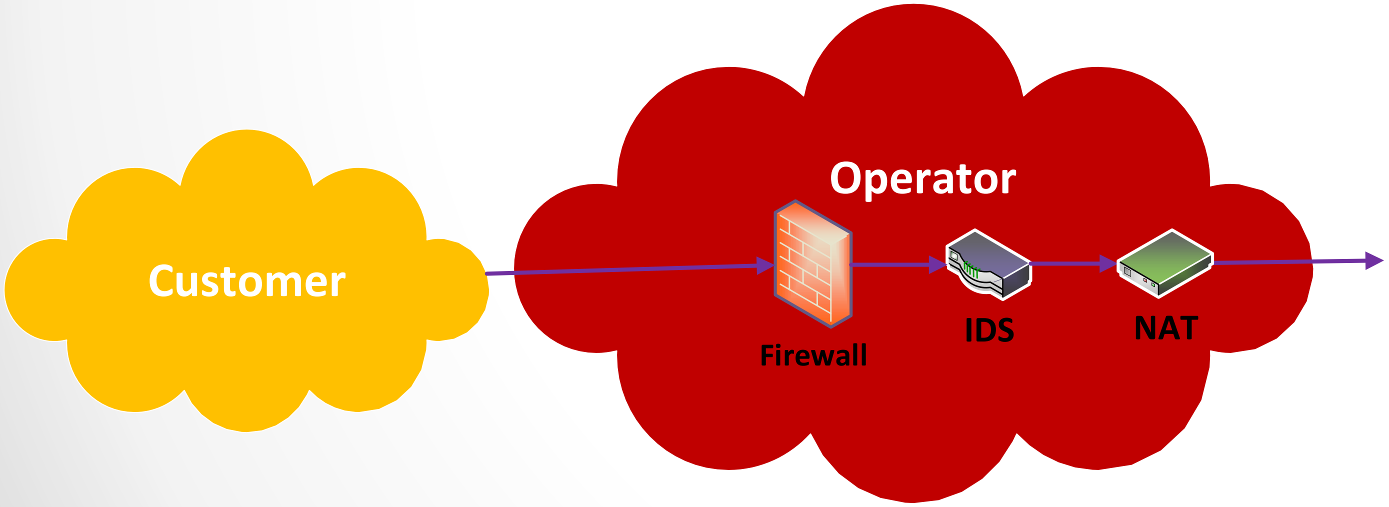
- Own and manage the physical infrastructure (i.e., one network)
- Deploy platforms and software required to run network functions
- The service could be simply data delivery or a SFC
- Provision and manage SFCs



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

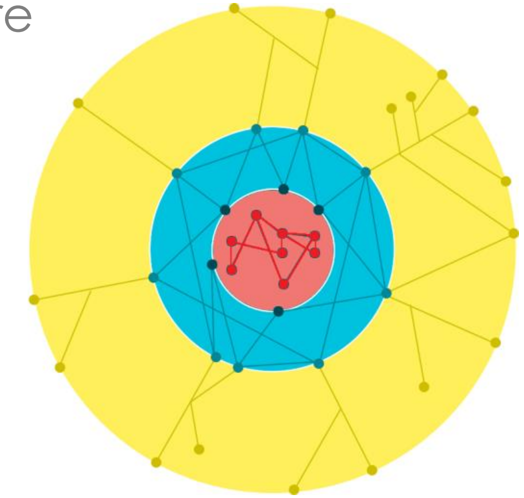


- SFC composition
- SLA requirements for the SFC
  - Bandwidth
  - End-to-end delay
  - Reliability, availability
- SLA requirements for each NFs
  - Processing power
  - Packet format(s)
  - Packet drop criteria…

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

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



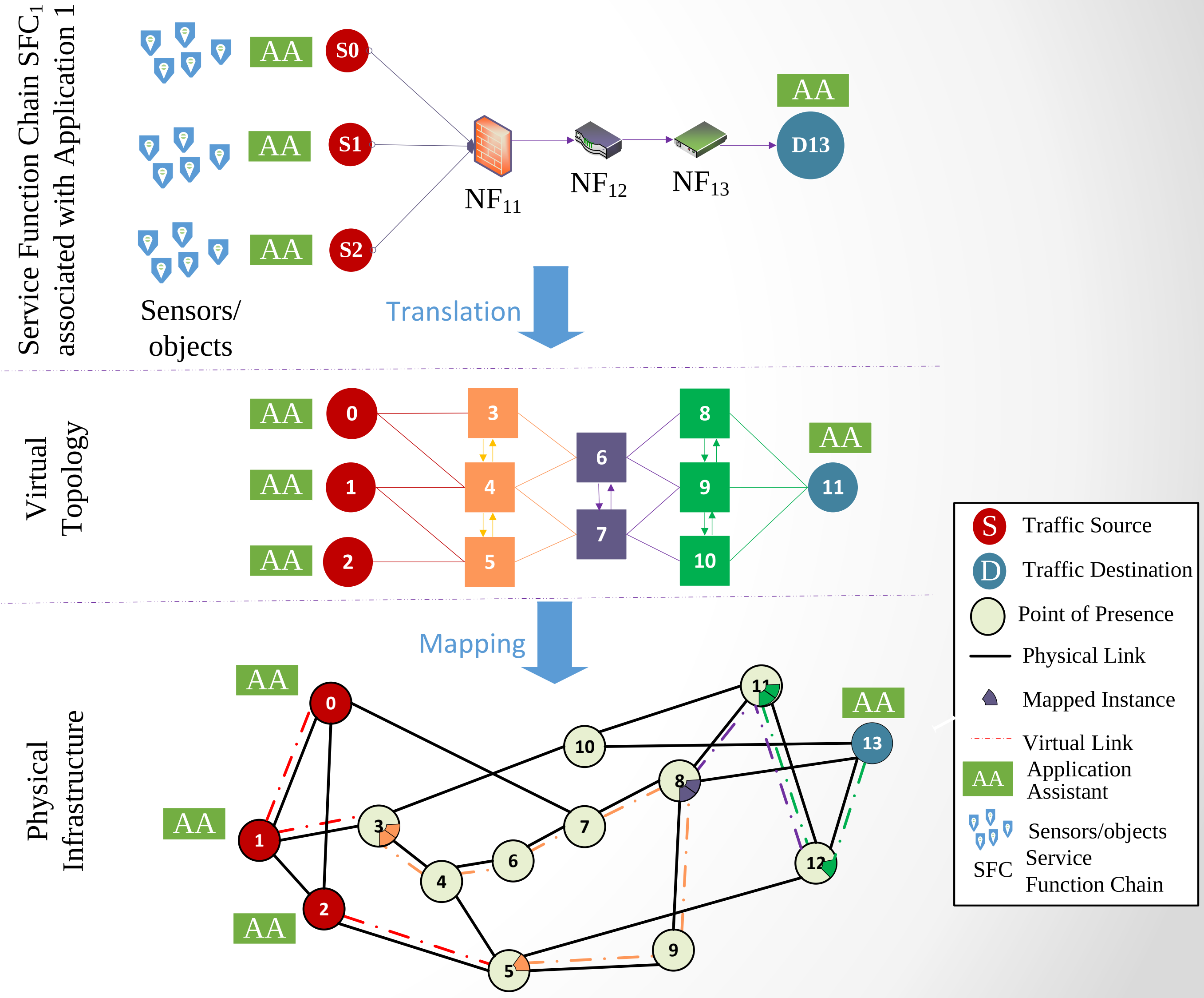
15 Data centers
100 Points of Presence (PoPs)
1000+ Edge nodes

Source: cloud.google.com

Resource Management Framework

Resource Allocation

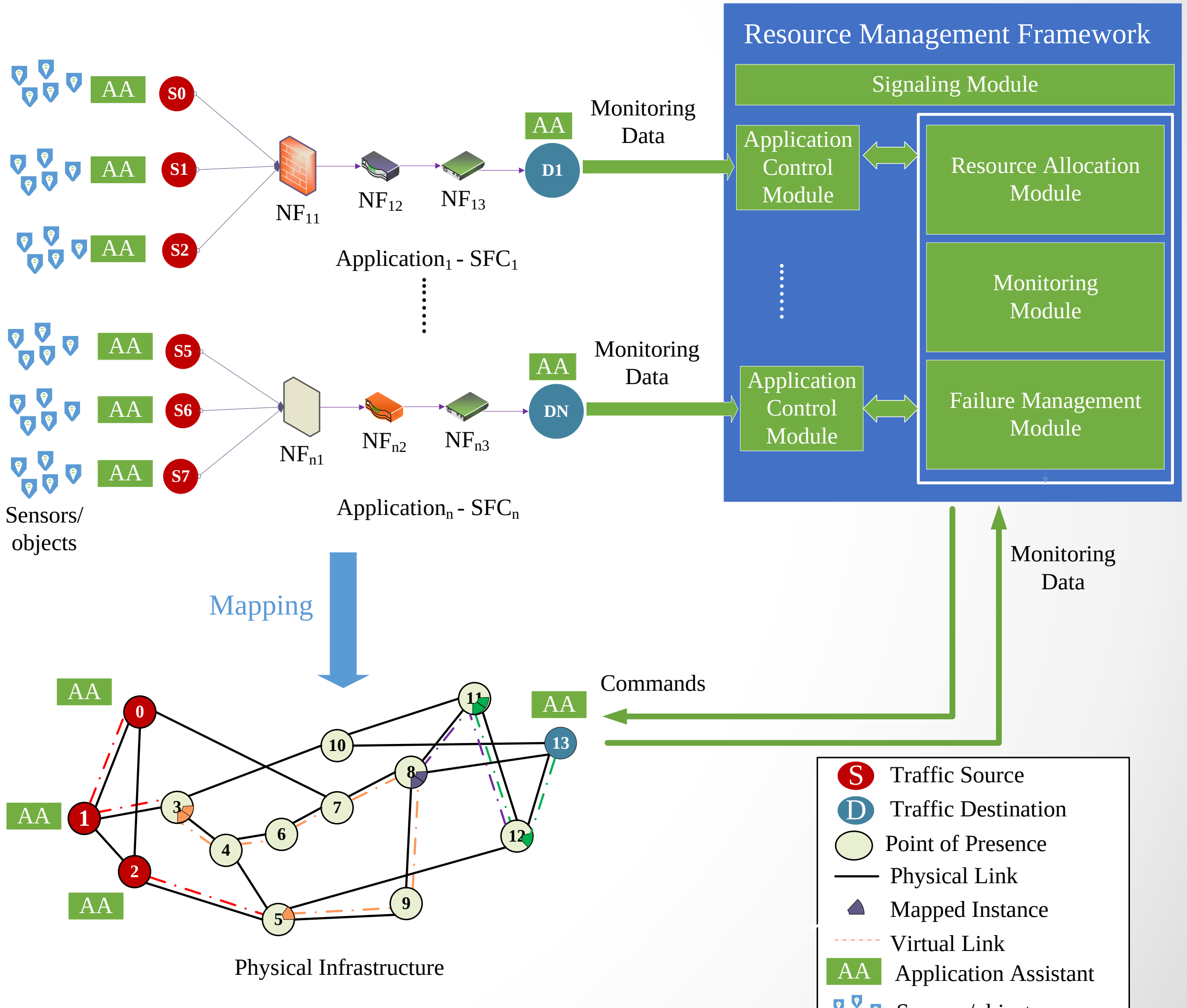
- The Service Function Chain (SFC) is defined by the application designer
- 2-step resource allocation:
  - Translation: the SFC is translated into a virtual topology
  - Mapping: virtual topology are mappa



Resource Management Framework

Main components:

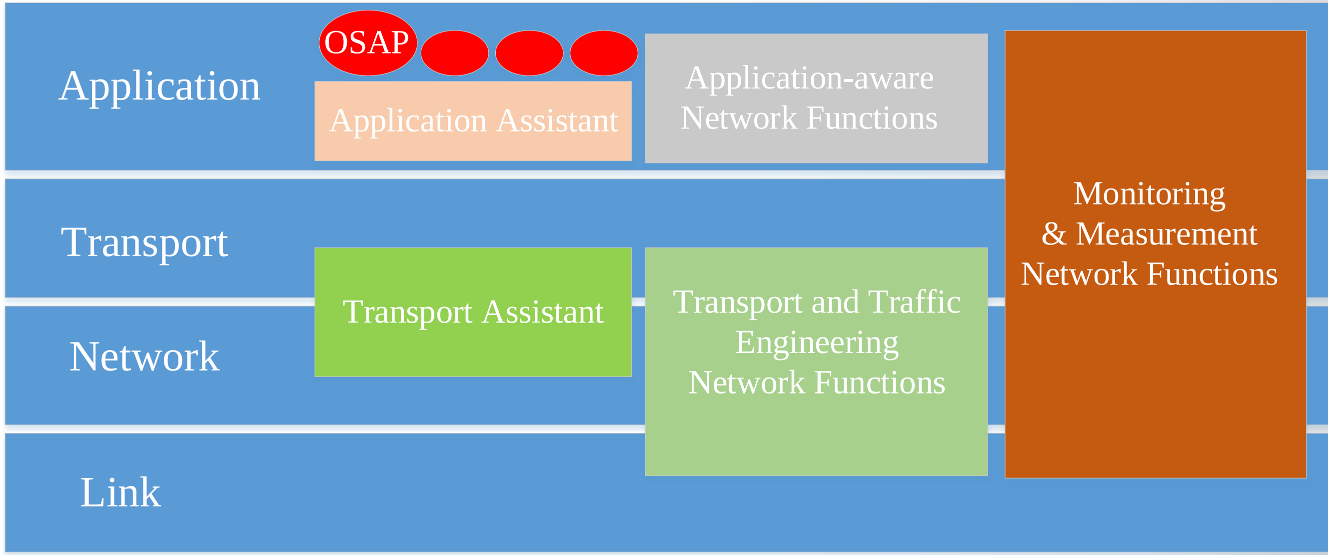
- Signaling module
- Application Control Module
- Ressource allocation Module



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

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

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

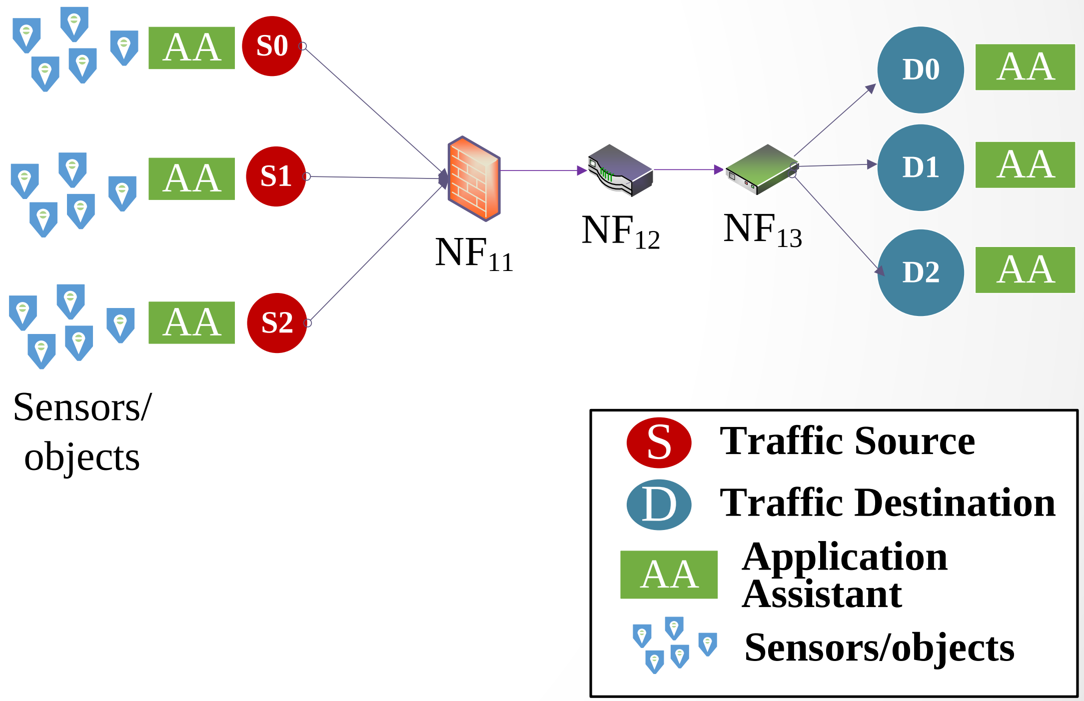
M. F. Zhani, H. ElBakoury - FlexNGIA 2019 (https://arxiv.org/abs/1905.07137)

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



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

Network Protocol Stack/Functions Transport Assistant

Le génie pour l'industrie

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

Application Layer

Transport Layer (TCP)
- E2E communication
- Blind congestion Control
- Inaccurate Packet Loss Detection
- Guaranteed Reliability
- E2E Packet Retransmission Process

Network Layer
- IP protocol (header and addressing)
- Routing Protocols/SDN
- ICMP for Control Information
- No Advanced Network Functions

Cross-Layer Transport
- Multi-point communication
- Network-assisted congestion control
- Network-assisted reliability and performance guarantees
- Accurate packet loss detection
- Variable performance and reliability Requirements over time
- Variable Header
- Meta-data and commands within packet headers
- Advanced Network Functions

Link Layer

Network Stack Headers

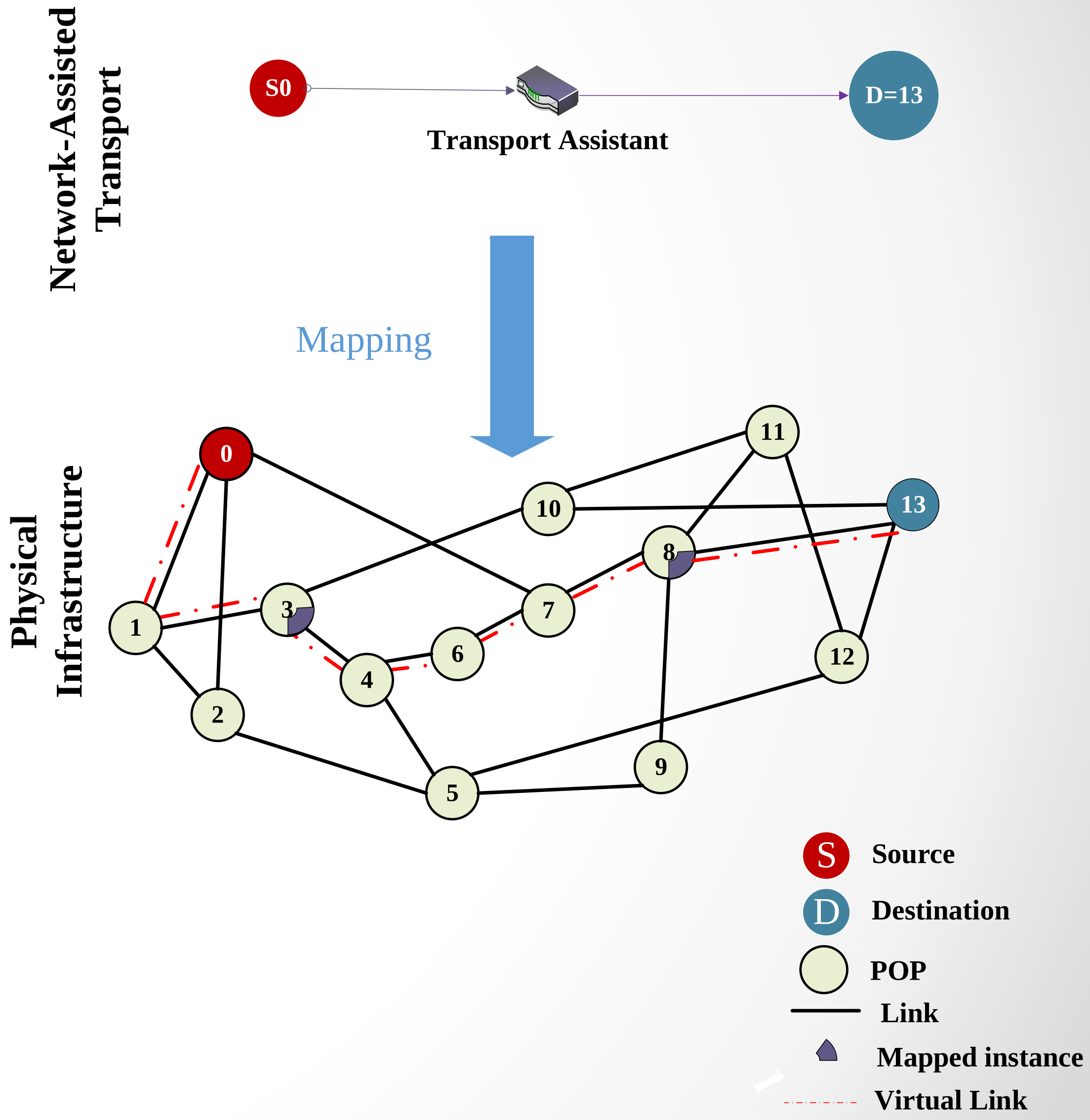
- Signaling packets
  - Instantiate an application
  - Convey application requirements
- Data packets: carry data
  - Layer 2 header: contains mainly the application id used for packet forwarding (similar to VLANs)
  - Upper layers:
    - Fully flexible header format (customizable meta-data and commands)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

Outline

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

Network-Assisted Data Transport

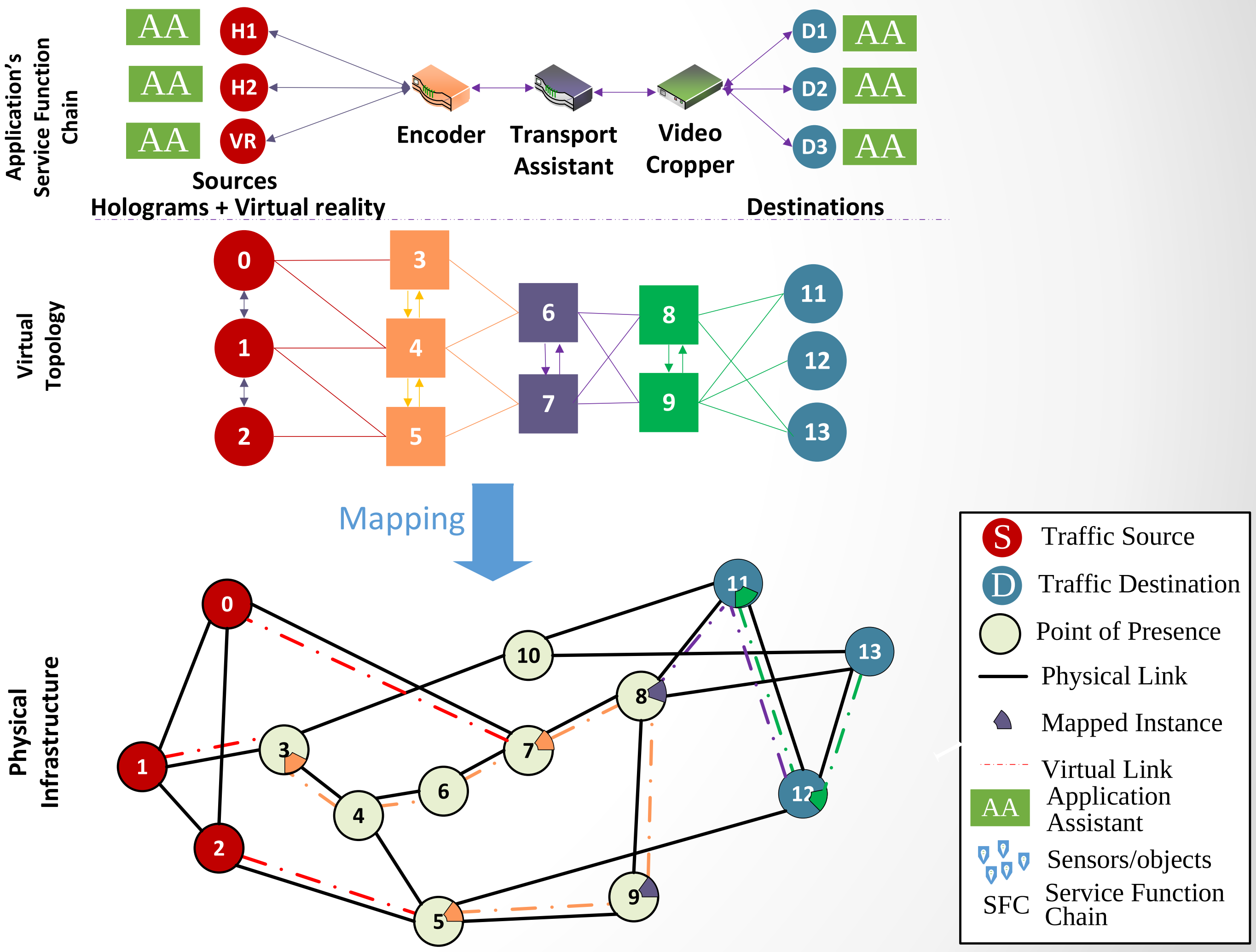
- Goal
  - Minimize retransmission delay
  - Improved congestion control
- Solution: service chain with a "transport Assistant" function
- Service of the Transport Assistant:
  - Caching and retransmissting packets
  - Detecting packet loss
  - Congestion control: adjusting rate, dropping packets, compression



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

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



Outline

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

Le génie pour l'industrie

Conclusion

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

Looking for More Details?

- Mohamed Faten Zhani, https://profs.etsmtl.ca/mfzhani/
- FlexNGIA Project: 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

École de technolgie supérieure (ÉTS Montreal)
University of Quebec, Canada

Thank You

Questions

Research Challenges

- Designing Service Function Chains tailored to applications
- High-performance softwarized functions
- Signaling
- Resource Allocation
- Fault-tolerance and Failure Management
- High-Precision and Fine-grained Monitoring and Measurements
- SDN++
- Distributed Cross-Layer Transport Protocol (sockets, caching, communication)
- Security and Privacy

Details are available in the paper (https://arxiv.org/abs/1905.07137)

Transport Layer Protocols (cont)

QUIC

- Transport over UDP
- Multi-streaming:
  - Every stream is a **reliable** bidirectional bytestream
  - Multiplexed streams between **two endpoints**
  - Stream prioritization
- Flow-control and congestion control very similar to TCP
- Endpoints use Explicit Congestion Notification (ECN)

SCTP

- Basically, a TCP++
- Multi-streaming
- Unordered delivery is possible
- Flow control and congestion control similar to TCP

Transport Layer Protocols (cont)

What are the limitations of SCTP and QUIC?

- E2E communication: multiple flows (streams) of the same application may connect more than two end-points
- A blind congestion control
- No support from the network: the network knows better about its state

→ Can better locate and manage congestion
→ Predict and detect more efficiently congestions/failures/problems…
→ Can retransmit faster
→ Can provide better garantees in terms of delay and packet loss

Network Protocol Stack/Functions Transport Assistant (cont)

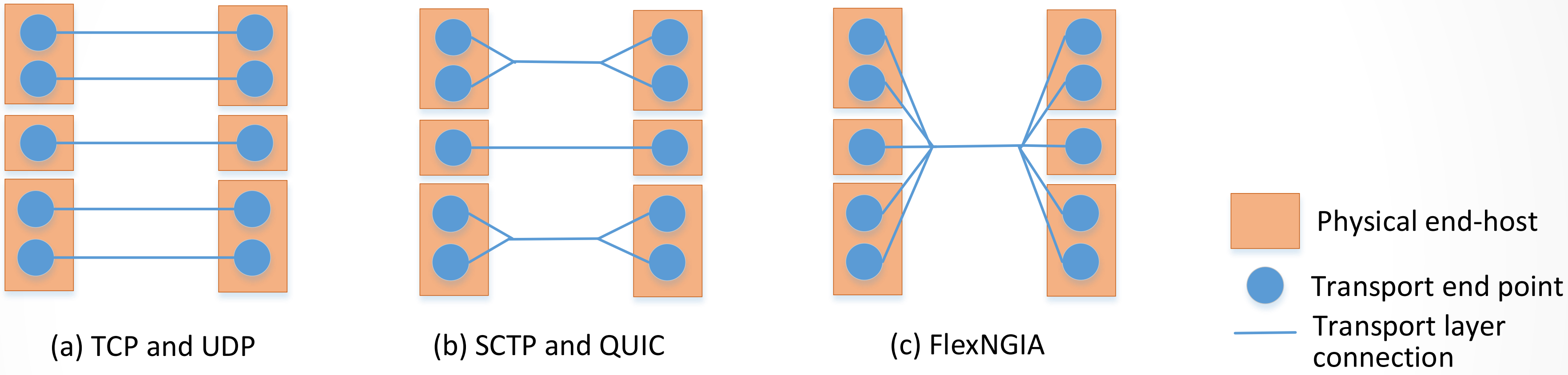


Illustration of how one single network application might be seen at the transport Layer

- Transport Assistants manage all these flows while taking into account that they all belong to the same application
- TAs monitor these flows, divide the total bandwidth allocated for the application among them.