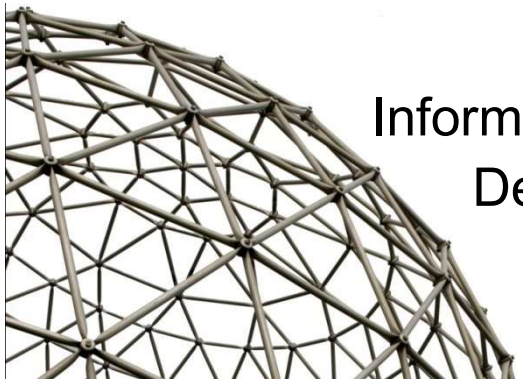
UCL DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING
INFORMATION AND COMMUNICATION ENGINEERING GROUP

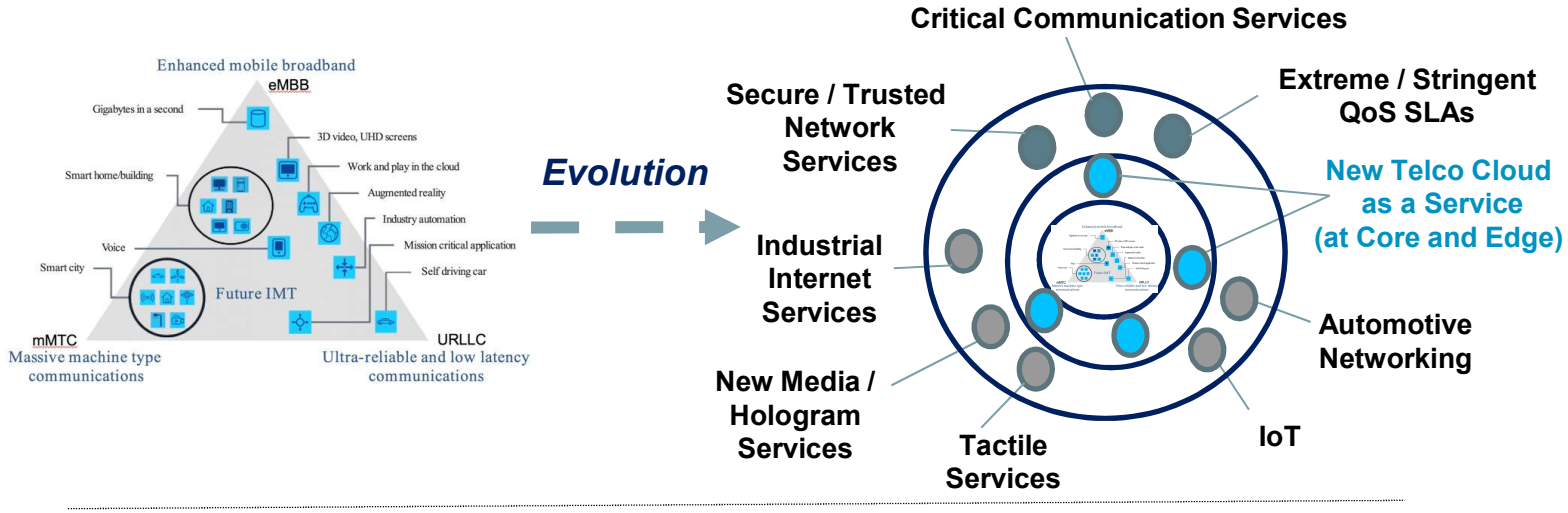
Towards Service- and Application-Aware Ultra Fast Networking

Prof. George Pavlou
together with
Prof. M. Rio, Dr I. Psaras, Prof. A. Galis, Dr D. Griffin,
Dr S. Clayman, Dr M. Charalambides, Dr O. Ascigil, Dr F. Tusa



http://www.ee.ucl.ac.uk/~gpavlou/
Information and Communication Engineering Group
Dept. of Electronic & Electrical Engineering
University College London, UK

Network Services Evolution



Present Digital Network Services & 2020 Network Services

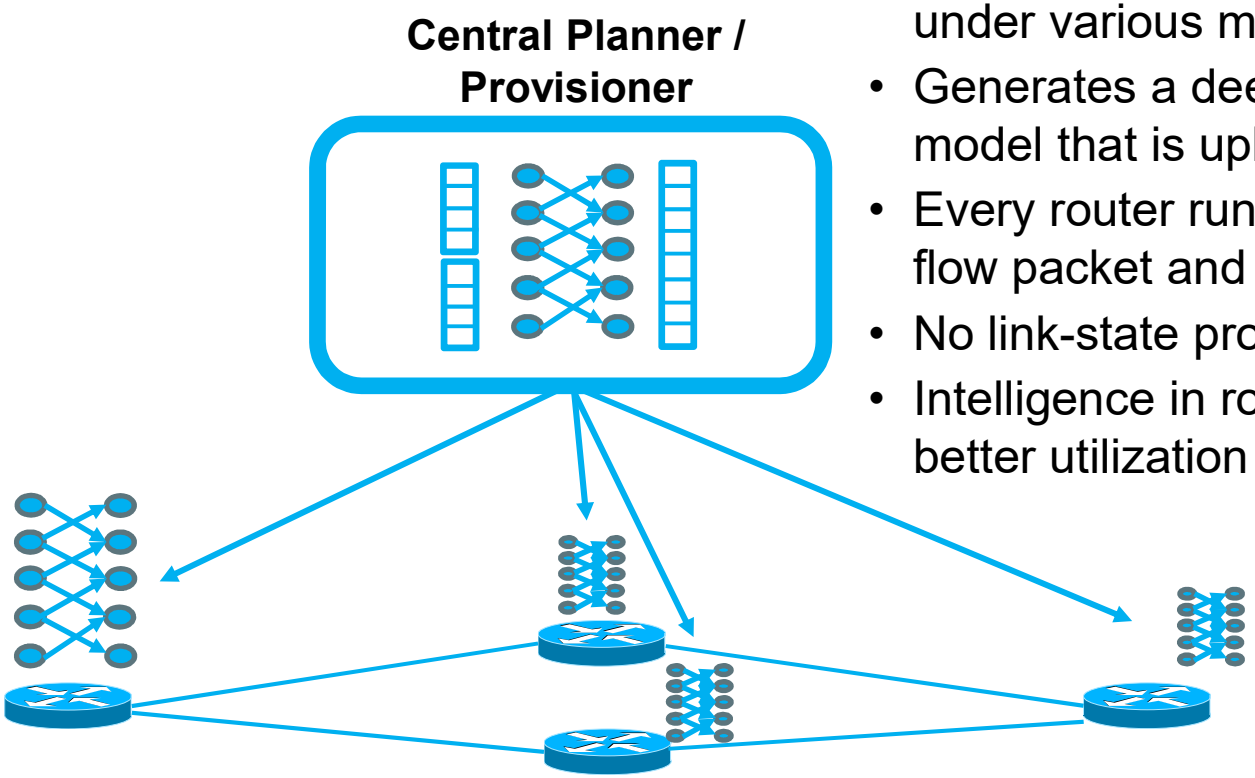
2030 and Beyond Network Services

- Evolution of network services towards: very low latency, very high bandwidth, very high reliability / resilience, trustworthiness and privacy
- In-network computation a key aspect, evolution of telecom clouds to support computation for third parties in both the core and edge
- Network programmability a key architectural principle / design choice to enable soft network evolution as requirements change

Research Areas Addressed at UCL EEE

- A number of research areas that contribute collectively in this direction are addressed at UCL Dept. of Electronic & Electrical Engineering
  - Neural routing and network-based congestion control (zero queuing) through deep reinforcement learning
  - Intent-based networking for declarative closed-loop network configuration, supported by adaptive line-speed traffic monitoring
  - Service routing without DNS-based resolution, with service addresses encoded in IPv6 packet headers – backwards compatible approach
  - New IP framework and protocol with meta-information in the header to trigger appropriate processing in routers for high-precision services
  - Big packet protocol as part of the new IP framework, with prioritized payload dropping for continuous media flow under adverse conditions

Neural Routing

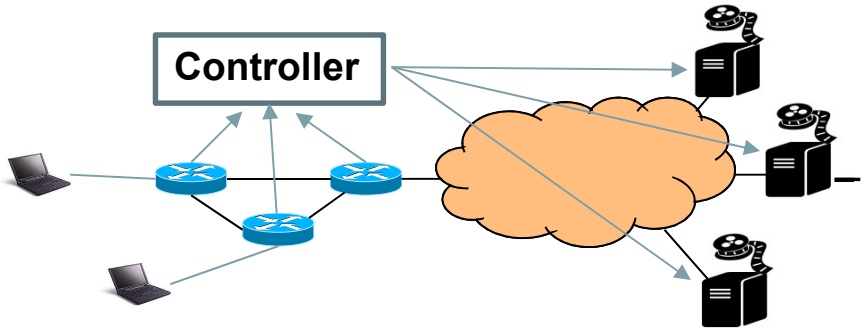


- Central Planner simulates the network under various measured traffic conditions
- Generates a deep reinforcement learning model that is uploaded to the routers
- Every router runs the model for the first flow packet and creates forwarding state
- No link-state protocol / MPLS in routers
- Intelligence in routing decisions, resilience, better utilization / load-balancing

Network-based Congestion Control

Motivation:

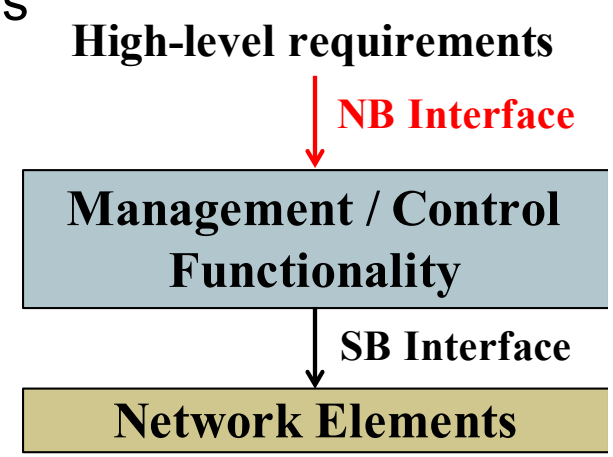
- "Zero queuing"
- Different utility functions
- Plug-in fairness



- Applied to eye-ball / edge domains
  - Elephant flow traffic is reported / monitored
  - Controller uses machine learning to produce optimal sending rates for each sender
  - Reinforcement learning algorithms based on Asynchronous Actor-Critic Agents (A3C)
  - Enforces the rates either by modifying TCP window size on the fly or through a control channel (requires transport modifications)
  - Encouraging first results

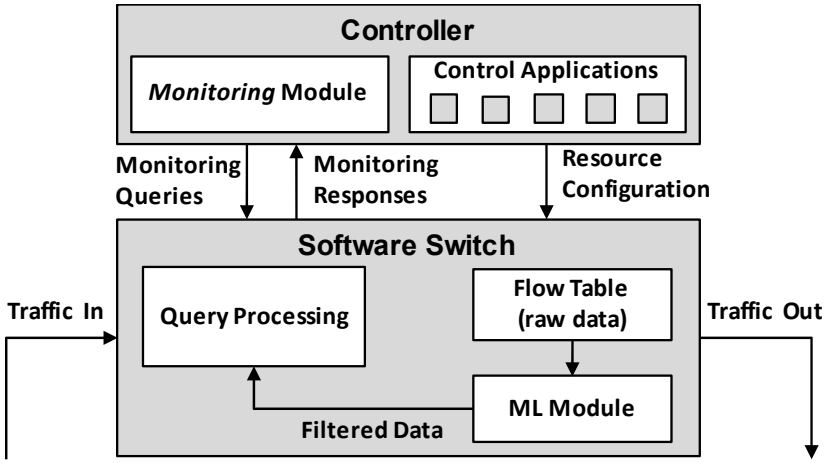
Intent-based Networking

- "Don't tell me what to do, tell me what you want"
  - Overcome configuration complexity using declarative directives
  - Decomposition of requirements to configuration commands
- Functionality realized through SDN North-Bound Interface
  - Automatically associates requirements to mgmt. functions, resources and commodities
  - Simplifies configuration process and avoids configuration inconsistencies
  - Allows frequent changes to high-level requirements
  - Facilitates faster service deployment
  - Smartness and adaptability in network operations and management
  - Key issues: conflict detection / resolution, coexistence with hard-coded applications



Adaptive Line-Speed Monitoring

- Management & control applications require timely and accurate monitoring data for closed-loop enforcement of intents
- Adaptive line-speed monitoring of software-based networks
  - Scalable (#devices, network diameter)
  - Programmable (measurement task, frequency, granularity)
  - Adaptive (new requirements, network dynamics)
- Efficient monitoring techniques at high packet rates
  - Process large amounts of state information in very tight timeframes
  - Reduce query processing time and memory consumption
  - Machine Learning (ML) to filter out unwanted data and classify flow entries according to monitoring task



Service Routing

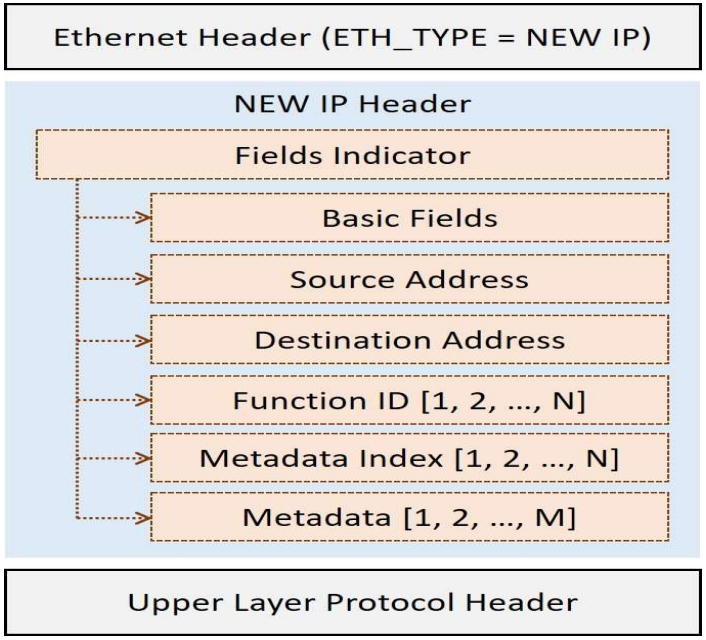
- DNS resolution overhead is unacceptable for latency-critical services, such as e.g. AR/VR, that require responses from service logic deployed close to the users
  - DNS resolution itself is much greater than the required service response time, which could be less than a few 10s of milliseconds
  - Also applicable to services/applications that use in-network computation in dynamic Edge/Fog Computing environments
- Service routing: route packets based on service names
  - No DNS resolution at all
  - The network performs *late-binding* of service names to locations
  - Key issue: find an incrementally deployable solution using current IP

Service Routing using IPv6

- Applications hash the first name part (e.g. www.AR.com) to obtain the IPv6 prefix and hash the suffix to obtain the serviceID
  - Service routers (SRs) map IPv6 prefix and serviceID to next hop – for legacy (i.e. non-service) routers, longest prefix match will still work
- Each service has service information maintained in SRs
  - Maximum capacity, current load
  - Ultimate goal is to route requests so as to jointly optimize network and service resources (latency, bandwidth, service capacity, load balance)
- Issues currently researched on:
  - Routing information scalability, forwarding strategies, service information metrics and other

New IP Framework for High Precision Services

- Key idea: include meta-information in variable length header to support programmability re how to process/forward packets in routers
  - ➢ Service-aware packet processing in routers considered together with network conditions could be used to support KPIs for future high precision services
  - ➢ Huawei/UCL short paper to appear in NOMS'2020*

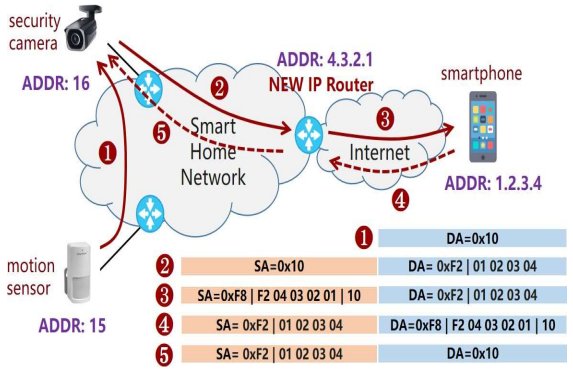


- **Variable IP address length** to seamlessly support communication across different network domains, e.g. fixed Internet and IoT
- **Semantic IP address definition** to identify both physical and virtual entities
- **User-extensible IP header** allowing services / applications to specify functions to be performed on packets in routers
- **Flexibility in programming IP's header**

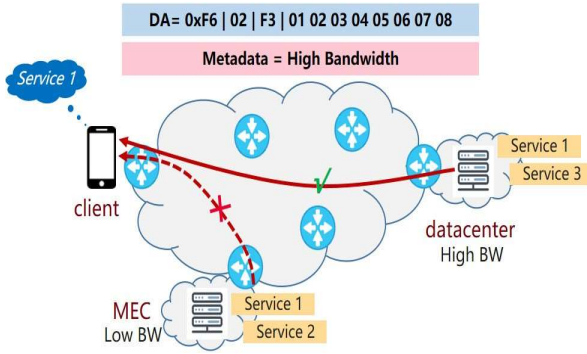
* Z. Chen, G. Li, Z. Lou, S. Jiang, A. Galis, "*New IP Framework and Protocol for Future Applications*", short paper to appear in IEEE/IFIP NOMS, 20-24 April 2020, https://noms2020.ieee-noms.org

Key Benefits vs Today's IP

- Service-aware packet processing based on network conditions for high precision services
- Service routing for faster service access without the DNS (see also slides 8/9)
- Dynamic provisioning of services and content given they can be accessed by name – service- / content-based routing
- Simpler implementation of network nodes in constrained environments due to variable length addresses



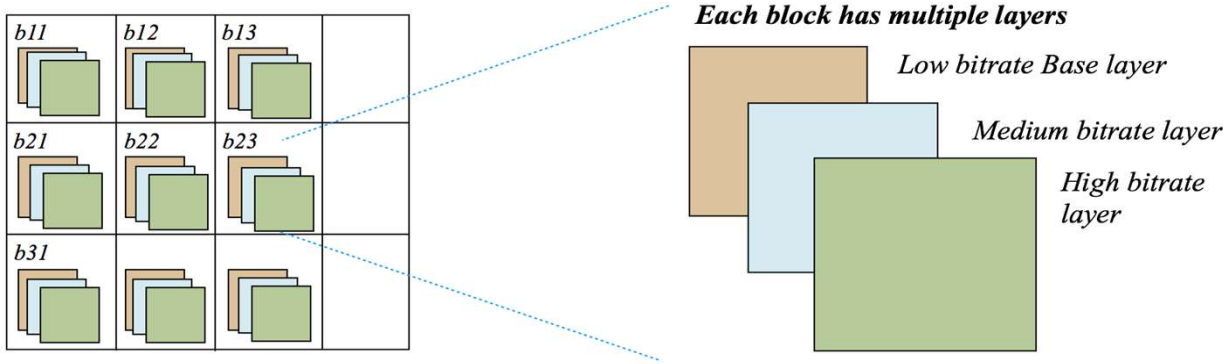
Home/Edge Automation



Routing Based on Service Needs

Media Big Packet Protocol (BPP)

- Specific part of the new IP framework, parts of media payload (“chunks”) are dropped in adverse conditions based on meta-data to avoid stalling
- Presented our work in Network 2030 in Geneva showing that BBP needs specific encoder/decoder and cannot use existing MPEG-DASH



- For future video services, BPP will need a scaling multi-layer encoder to embed multiple encodings of a frame into the same Macro Block
- We are currently looking at application-BPP interaction for VR/holographic services, putting different parts of the visible space into different chunks
  - Also at a feedback receiver->sender loop to take into account network state

Summary

- A number of areas addressed at UCL EEE that collectively contribute towards service-aware ultra fast future networks
  - AI-based Intelligent solutions to traditional problems, such as routing, congestion control and high-speed monitoring
  - Radical re-thinking of operations and management in a programmable manner through high-level intents
  - Service-based routing in a backwards compatible manner
  - Next-generation high-precision IP that supports programmability regarding the way in which service-aware packets are treated
- Network programmability a key issue and design choice for graceful future network evolution

More Information

- For more information on those areas you may contact the following people
  - Neural routing and network-based congestion control:
    Prof. Miguel Rio miguel.rio@ucl.ac.uk
  - Intent-based networking and adaptive line-speed monitoring:
    Prof. George Pavlou g.pavlou@ucl.ac.uk
  - Service-based Routing:
    Dr Yiannis Psaras i.psaras@ucl.ac.uk
  - High precision new IP:
    Prof. Alex Galis a.galis@ucl.ac.uk
  - Big Packet Protocol for Media:
    Dr Stuart Clayman s.clayman@ucl.ac.uk