Virtualization techniques for redesigning mobile backhaul networks: challenges and issues

Fabrice Guillemin
Orange Labs, IMT/IMT/OLN/CNC/NCA

MobiArch 2015 – September 7, 2015

Outline

1 Introduction

2 Convergent gateway: design principles and implementation

3 Further issues related to convergent gateways

Introduction: Fixed/mobile convergence

- Fixed/mobile convergence is a recurrent issue in the design of networks (at least for the past 25 years …)
- Architecture of mobile and fixed access networks has been inherited from principles of
  - the 80's for fixed networks
  - the 90's for cellular networks
- Mobile networks are greatly centralized: all traffic converges to a few concentration points:

  Packet Gateways, PGW
- Architecture of cellular networks designed by 3GPP and that of ADSL/FTTH fixed networks by Broadband Forum

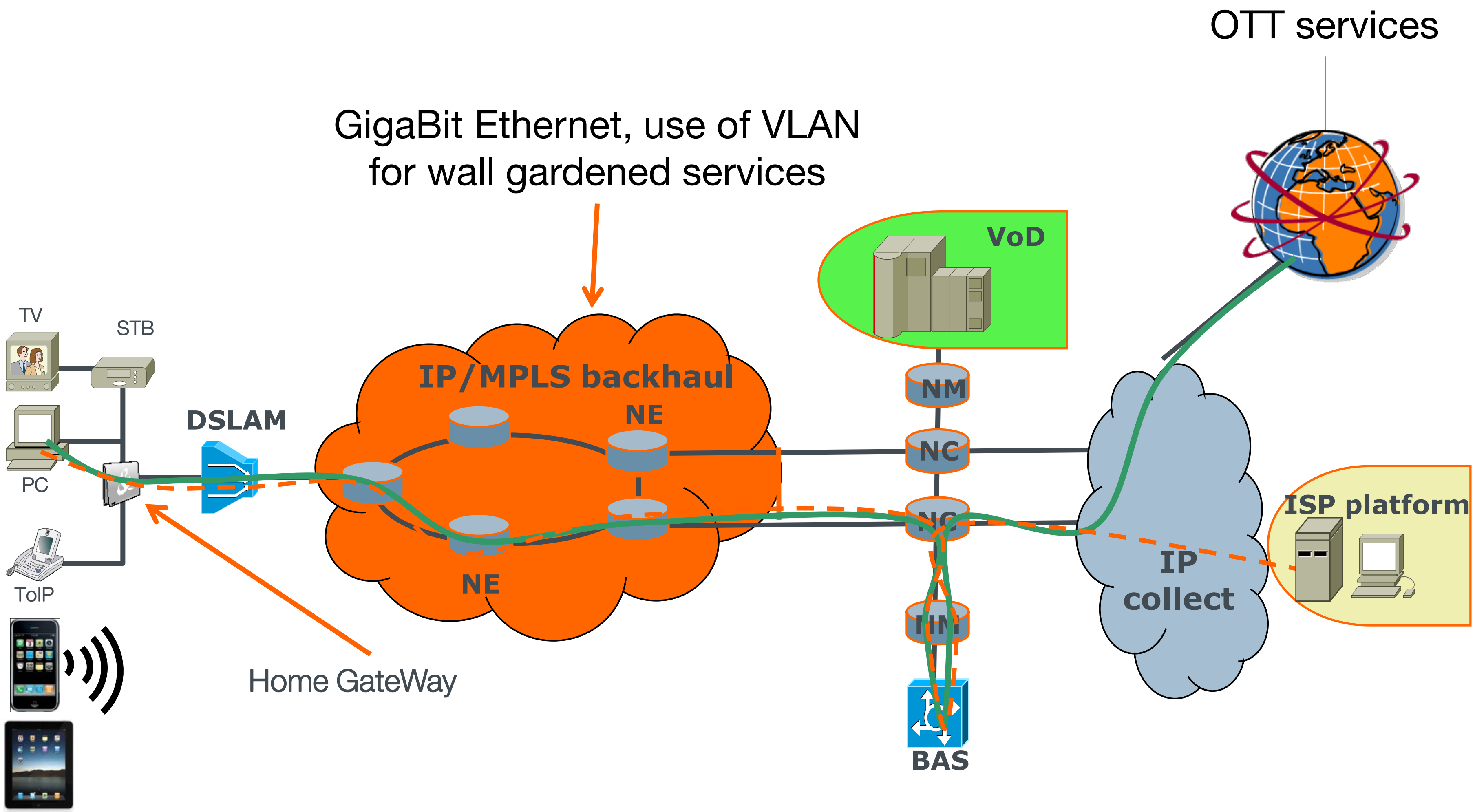
Introduction: 5G and beyond

- Challenges for networks:
  - Customers usage is rapidly changing (data traffic explosion, content services, streaming on cellular networks, user generated content, P2P, …)

    Dramatic increase of traffic
  - Densification of the radio coverage (3G/4G, WiFi, small/femto/macro cells)
  - IP has become the common convergence layer (even if L4/L3 functions are not in force, information is already packetized according to IP principles at the access) – no more circuit radio

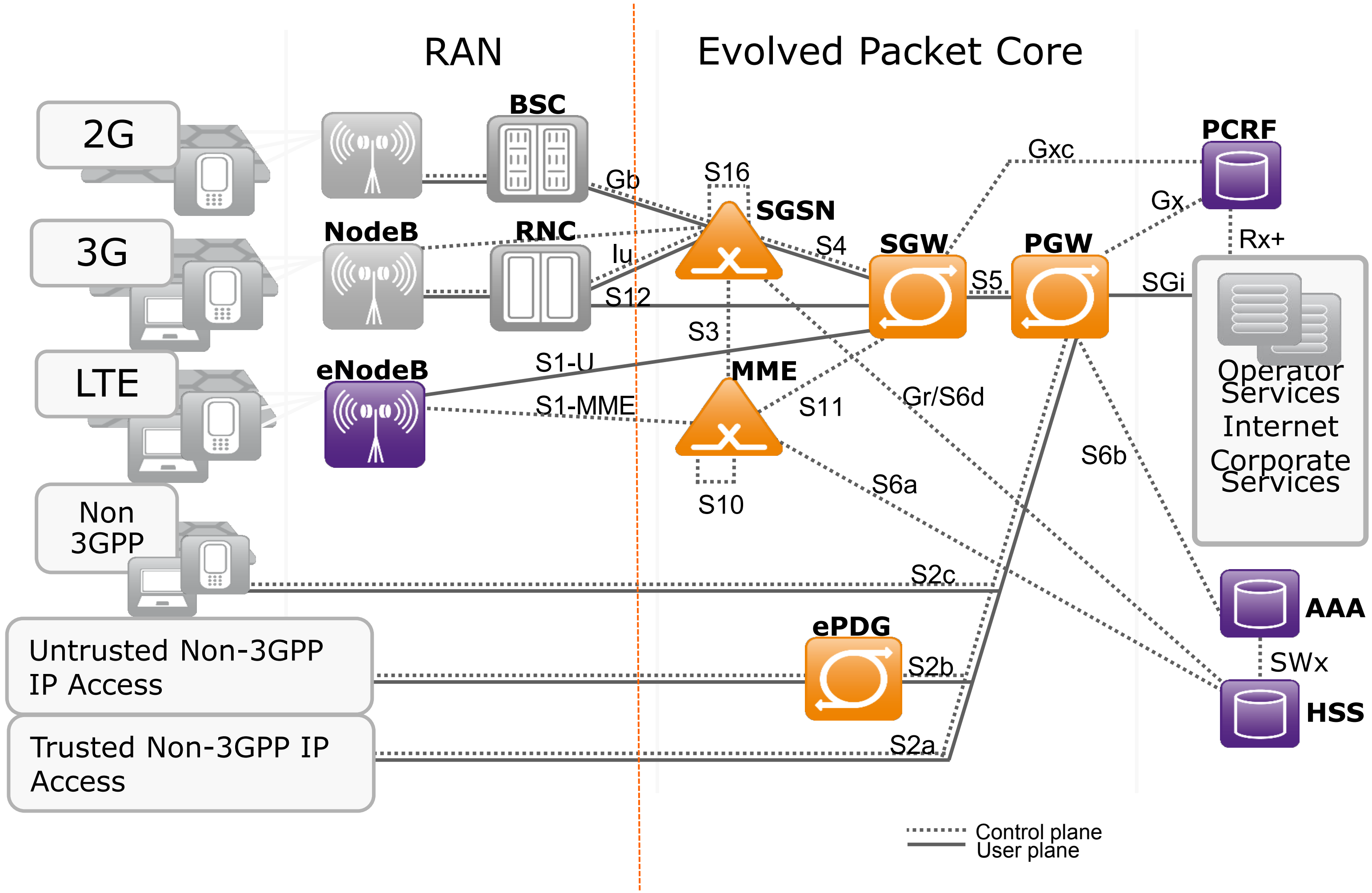
Key challenge: design of backhaul and mobile core networks for 5G (enabling true convergence !)

Architecture of fixed networks



IP functions actually begin above the BNG (or BAS)

Architecture of cellular networks



Only a few PGW (less than 10 for the Orange France network)

Rationale for introducing a Convergent Gateway (CGW)

- Three main drivers:
  - change of usage with smartphones and tablets (studies show that users are massively using WiFi in domestic networks)
  - densification of radio network (WiFi via fixed access or hot spots, femto, small, macro cells)
  - explosion of traffic

- Usage
  - massive usage by customers of the wireless technology to connect to the Internet (mobile and WiFi)
  - customers wish to access the Internet without taking care of the access technology (WiFi, 3G/4G, wired)
  - The customer is surrounded by many access networks but can hardly switch from one network to another (different authentication, addressing, customer's profile, routing, charging, …)

Limitations due separate architectures

- In terms of usage:
  - each access network has its own addressing system and allocates IP addresses via PPP or DHCP
    - NAT, local DHCP,… at the home gateway
    - NAT, (DHCP), … at the PGW
  - Addressing systems and policies are different

Is IPv6 sufficient to overcome this problem? Unique IPv6 address per device? What about mobility, roaming, etc . ?

  - No unified authentication procedure (SIM based, subscriber line based, etc.)
  - IP address changes when switching from one access network to another
  - Applications need to adapt themselves in order to be resilient to IP address changes (issues with VPN, conversational and real time constrained applications, …)
  - Many applications are resilient to IP address changing (e.g., adaptive streaming) but **for real time applications (voice) this is still an open issue**

Cost of separate architectures

- Customers generate more and more mobile traffic
  - to cope with traffic explosion, centralized PGWs and some links of the network have to be upgraded to carry traffic
- When a user switch from one network to the other, complex signaling procedures are run (notably radio bearer (de)activation)
  - because of centralized control, huge amount of signaling traffic can appear all over the network

Bearer activation is a source of latency

- Centralized functions are vulnerable points in the architecture (Single Point of Failure, need for resilience at extra cost)
- Separate AAA mechanisms need to be maintained

Challenge: AAA without smart cards for unified access?

Technological trends

- L2/L1 concentration higher in the network
  - larger distance between ONUs and OLT in optical access
  - BBU hostelling in mobile networks (coordinated radio resource management)
  - development of new convergence points in the network: Next Generation PoP (NGPoP)
  - colocation of OLT and BBU hostels in NGPoPs enables a global view of resources
- Distribution of some IP features lower in the network:
  - distributed CDN to make information available closer to the end user
  - Fog computing (distribution of data centers)

Need for a new type of network element realizing the convergence between fixed and mobile access: The convergent gateway

CGW:
Design principles and implementation
(implementation in progress by BCOM)

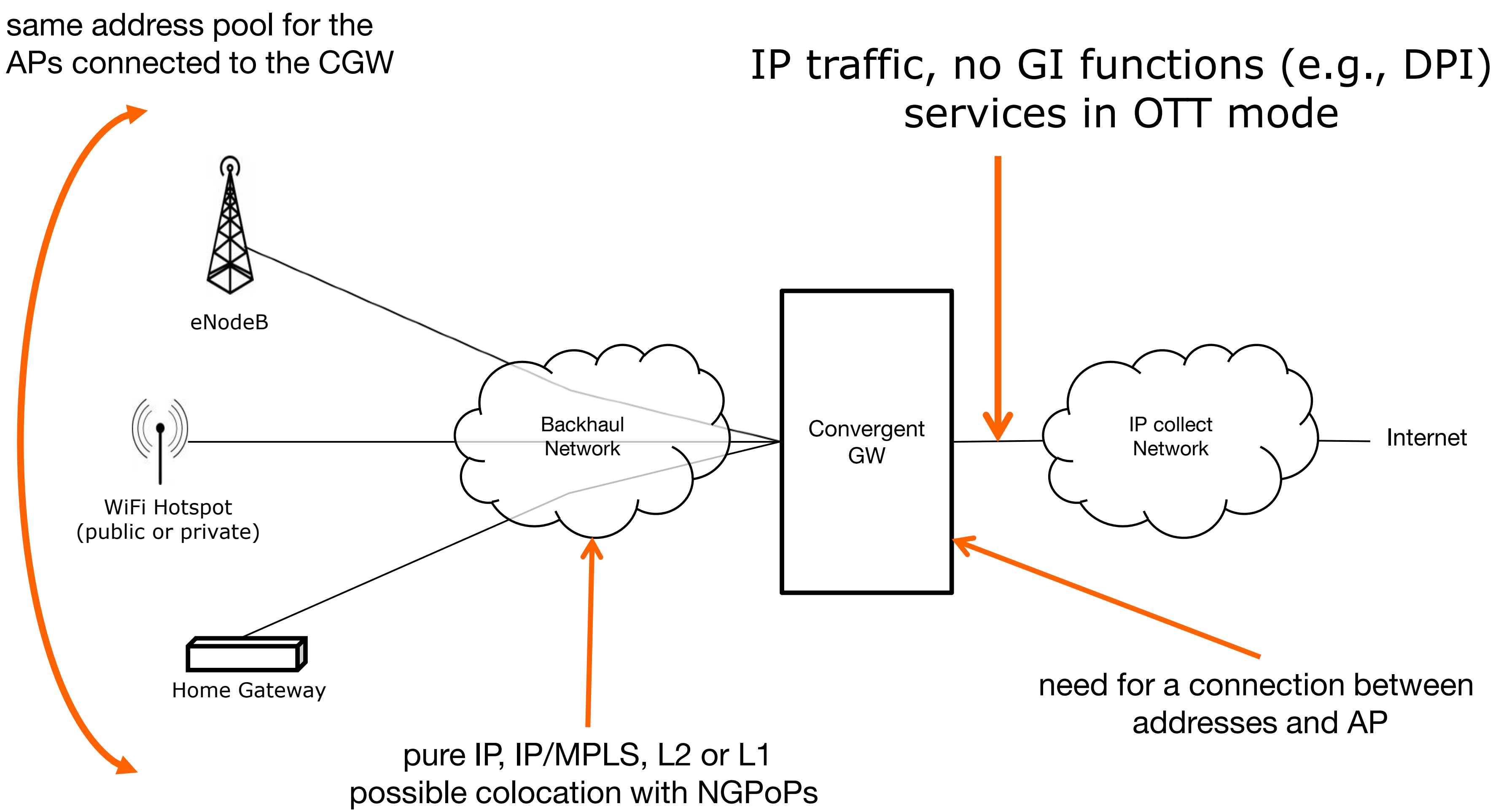
CGW principles

- Convergence of various types of access at Network's Edge: fixed, WiFi, cellular
- Authentication for cellular and fixed access

Open issue: Need for a convergent AAA (done by BCOM for cellular/WiFi by using 3GPP standards)

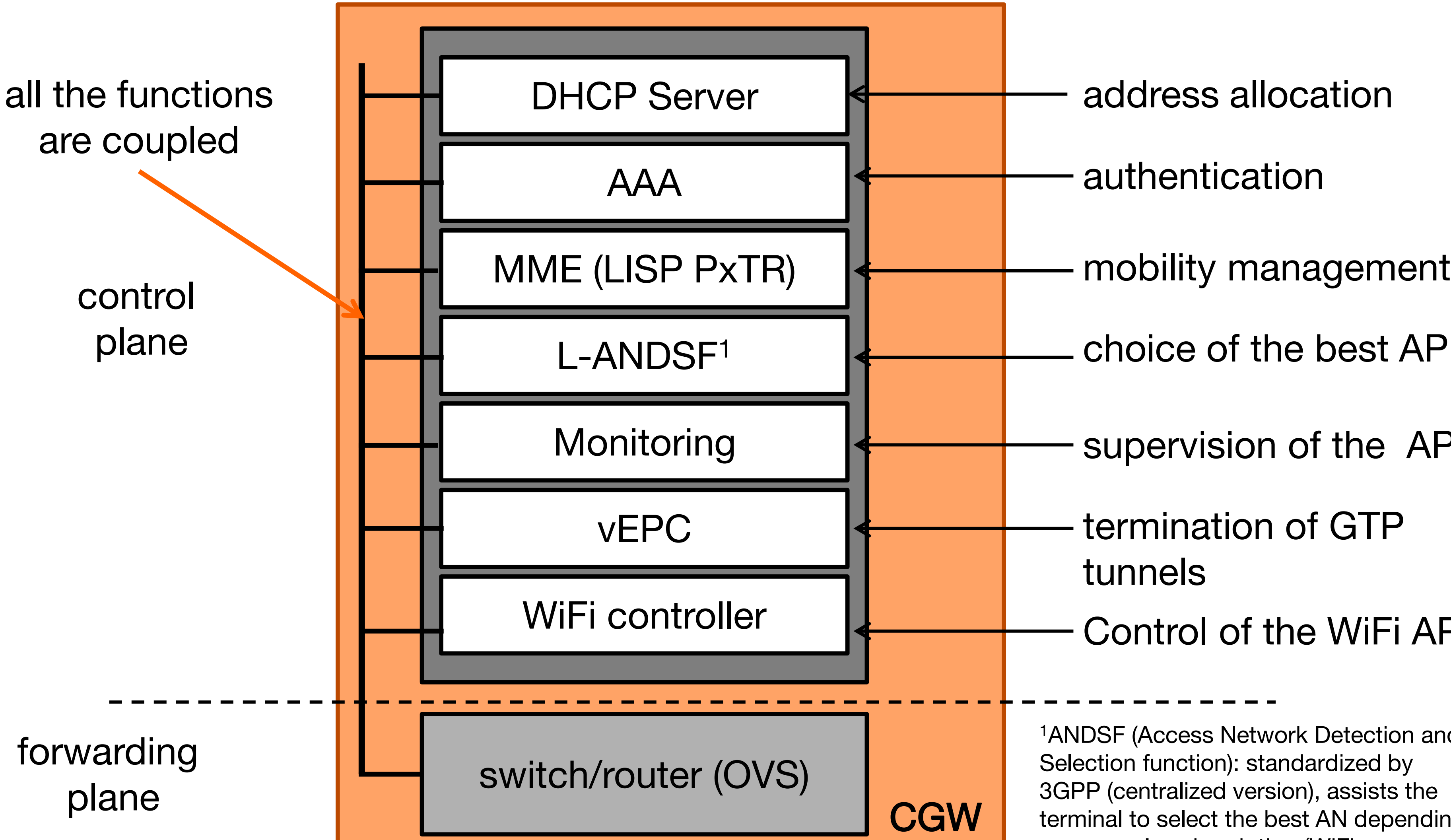
- IP functions at the CGW: routing, quality of service, mobility
- Information (in the form of IP packet) is delivered to the IP collect network above the CGW
  - no need for centralized platforms
- Dynamic instantiation on COTS servers/switches through Networks Functions Virtualization: flexible deployment, configuration, upgrades, scaling
- SDN used for flexible traffic control at Network's Edge

Location of the CGW



The functional blocks of the CGW

all the functions are hosted by virtual machines (NFV) and are based on open source software



[1] ANDSF (Access Network Detection and Selection function): standardized by 3GPP (centralized version), assists the terminal to select the best AN depending upon user's subscription (WiFi offloading)

Implementation issues

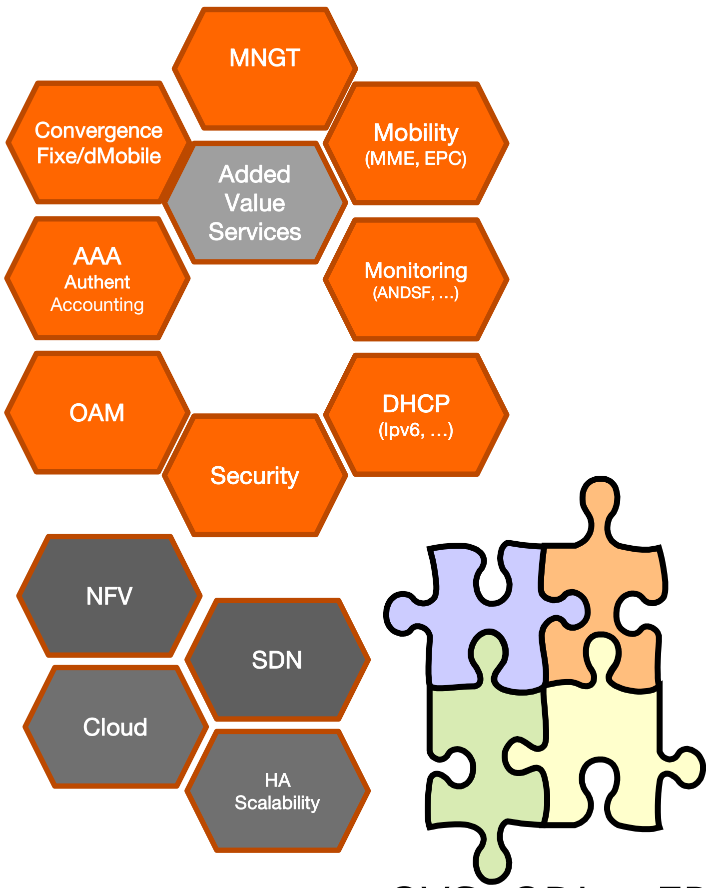
- Mobility management (MME):
  - intra CGW MME
    - the need for MME depends on the addressing scheme used for allocating addresses to terminals connected to APs
    - either addresses depend on APs and in that case addresses may change when one terminal is moving from one AP to the other ➔ possible need for MME (LISP tested in lab, SDN solutions under study)
    - or addresses are common to all AP; in that case, need for maintaining a map between the address and the AP for forwarding information (**convergence function**)
  - inter CGW MME
    - when one terminal moves from one CGW coverage to another one
    - possible use of LISP principles if the terminal communicate the identifier of the CGW; if a NAT is implemented in CGW (EID = IPv6 address of the terminal, and local ID = IP address of the CGW to maintain in the map resolver)

alternative: do nothing in the network and let terminals manage mobility

Implementation issues (cont'd)

- vEPC:
  - need for open source vEPC stacks
  - challenge with the current vEPC:
    - cellular traffic is traversing the VM hosting the vEPC by Open Air Interface – same issue with OpenCAPWAP for WiFi
    - break the link between GTP-U and GTP-C
      - GPT-U should be handled by the forwarding function only
      - GTP-C should be hosted by separate VMs
    - address allocation should be done by the global DCHP server
  - More challenging: no more GTP tunnels
    - use the same principles as for fixed networks without PPP
    - use of DHCP for address allocation, flat IP in the backhaul, unified AAA between fixed and wireless networks
- Monitoring
  - Is it possible to do monitoring in Virtual Machines? Can (soft) switches do port mirroring at high speed (> 1Gbit/s)?
  - What is the trade-off between processing capacities and the level of traffic analysis?

Further issues



OVS, ODL, vEPC, ...

- Charging and accounting:
  - The distribution of the mobile core makes it difficult to count traffic/user
  - reinvent charging? End of capped offers?
- CGW instantiated in (mini) data centers
  - possibility to couple with added value services:
    - WebRTC TURN servers to enable WebRTC services
    - CDN servers co-located with the CGW
    - ....
- Quality of service and traffic management
  - monitoring functions and Local-ANDSF to perform traffic and QoS management
  - possible API for QoS negotiation
  - how to manage QoS without radio information? Couple traffic management with BBU hostels?

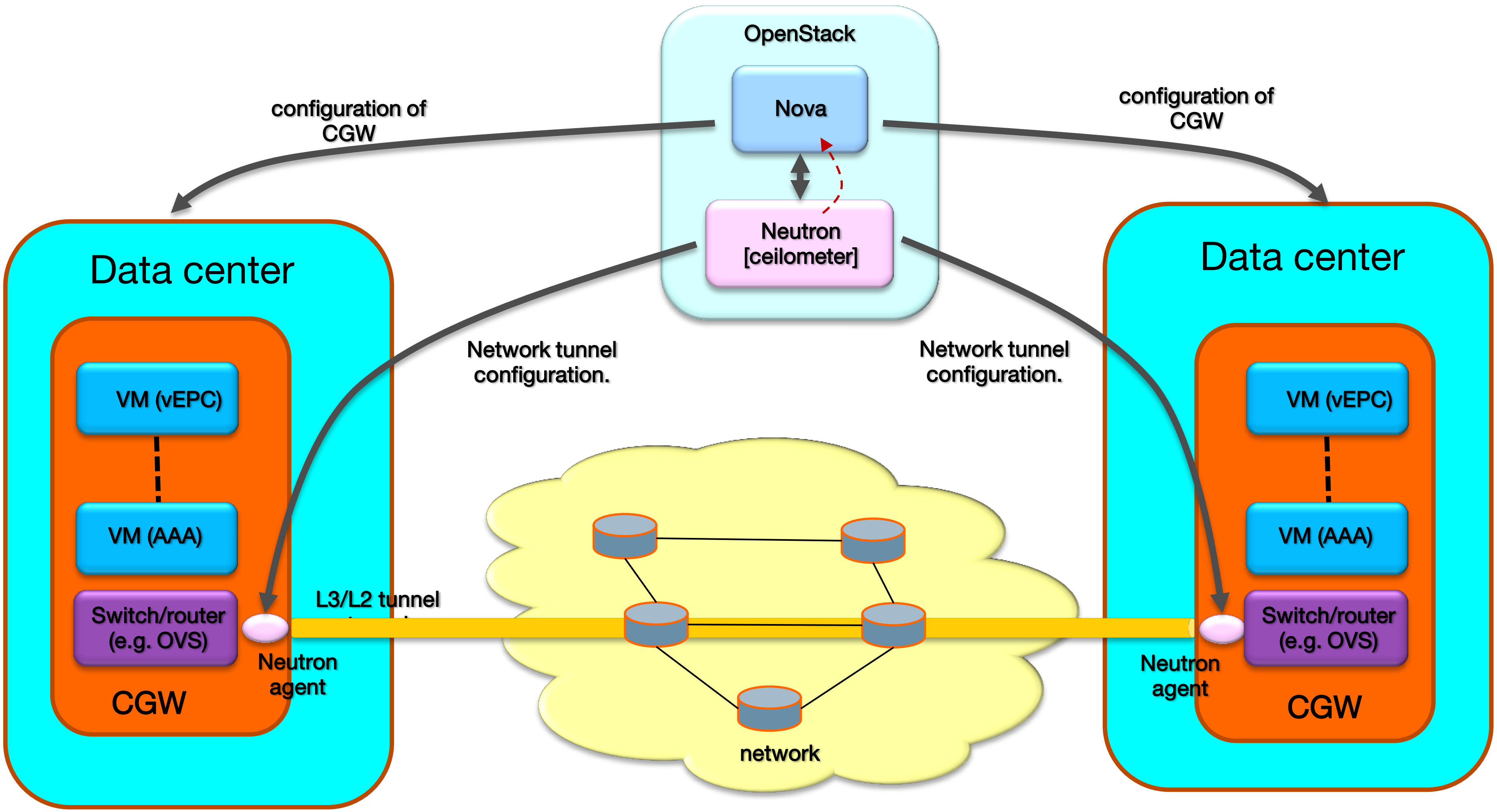
Further issues related to a CGW

CGW “on the fly”

- CGWs should be instantiated on the fly on data centers (OpenStack) distributed at network edges (fog computing)
  - there is hence a need for a tool capable of configuring such elements (typically OpenStack)
  - But CGWs should be interconnected, sometimes with bandwidth constraints
  - OpenStack is not sufficient by itself, there is a need for a tool able to configure the network in order to interconnect CGW (e.g., OpenDaylight), typically when used to backhaul an enterprise network
- OpenDayLight and Openstack have been developed for given purposes, there is a need for a tool with a global view of the network in terms of storage, computing and bandwidth

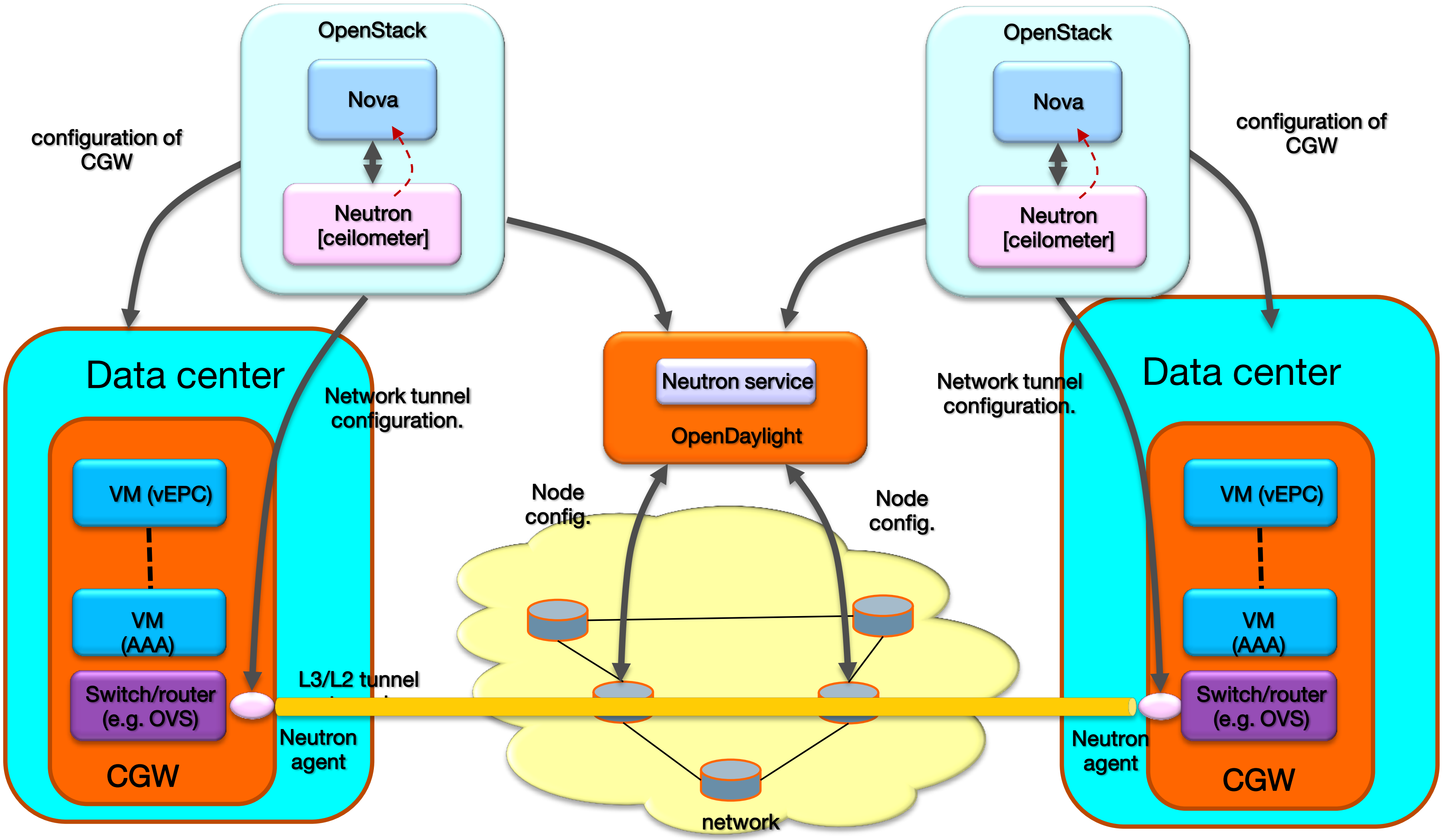
➔ GlobalOS under study at OrangeLabs

Configuration of a CGW (OpenStack – centralized view)

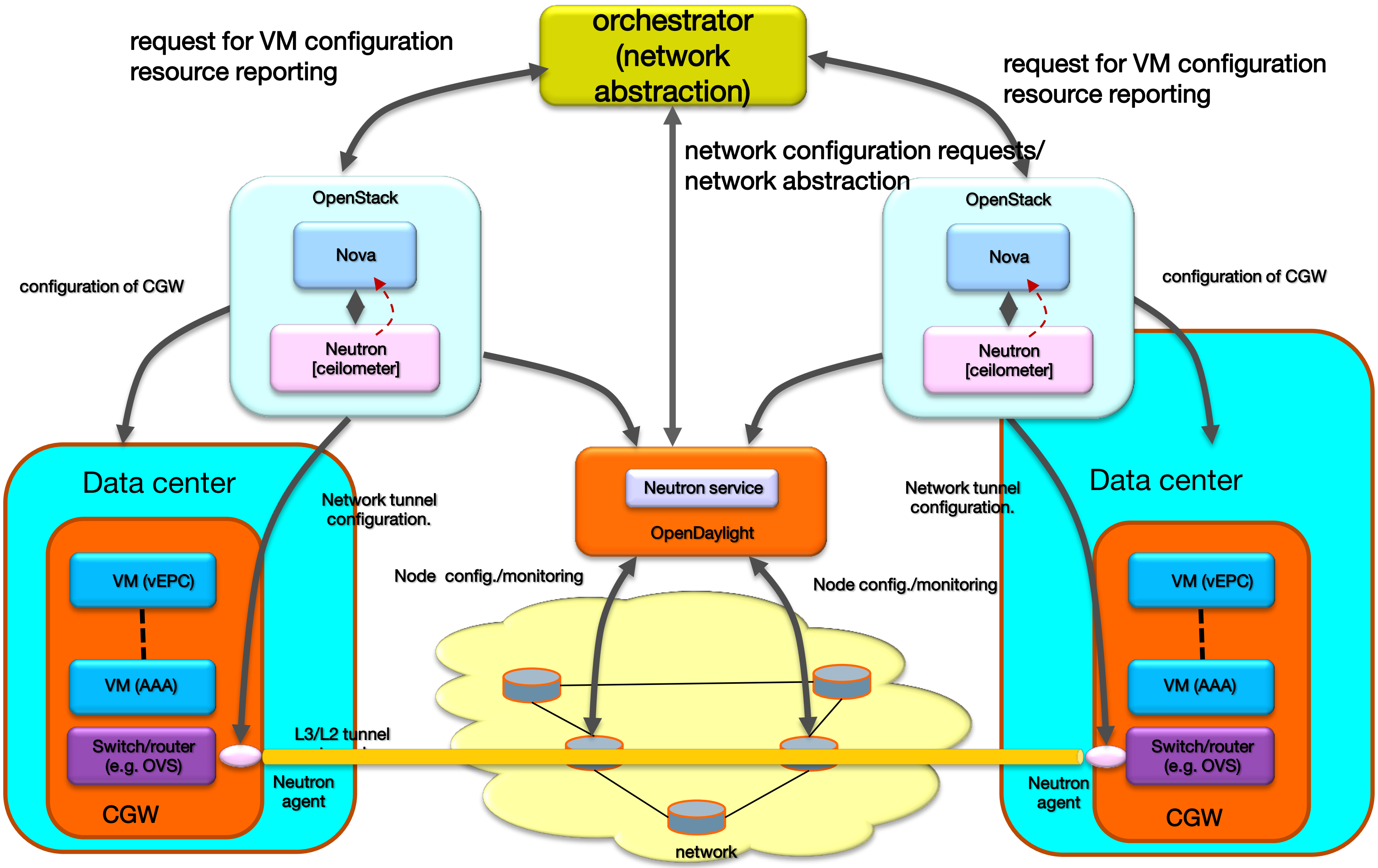


Openstack can control and configure a CGW from the edge (nothing in the network)

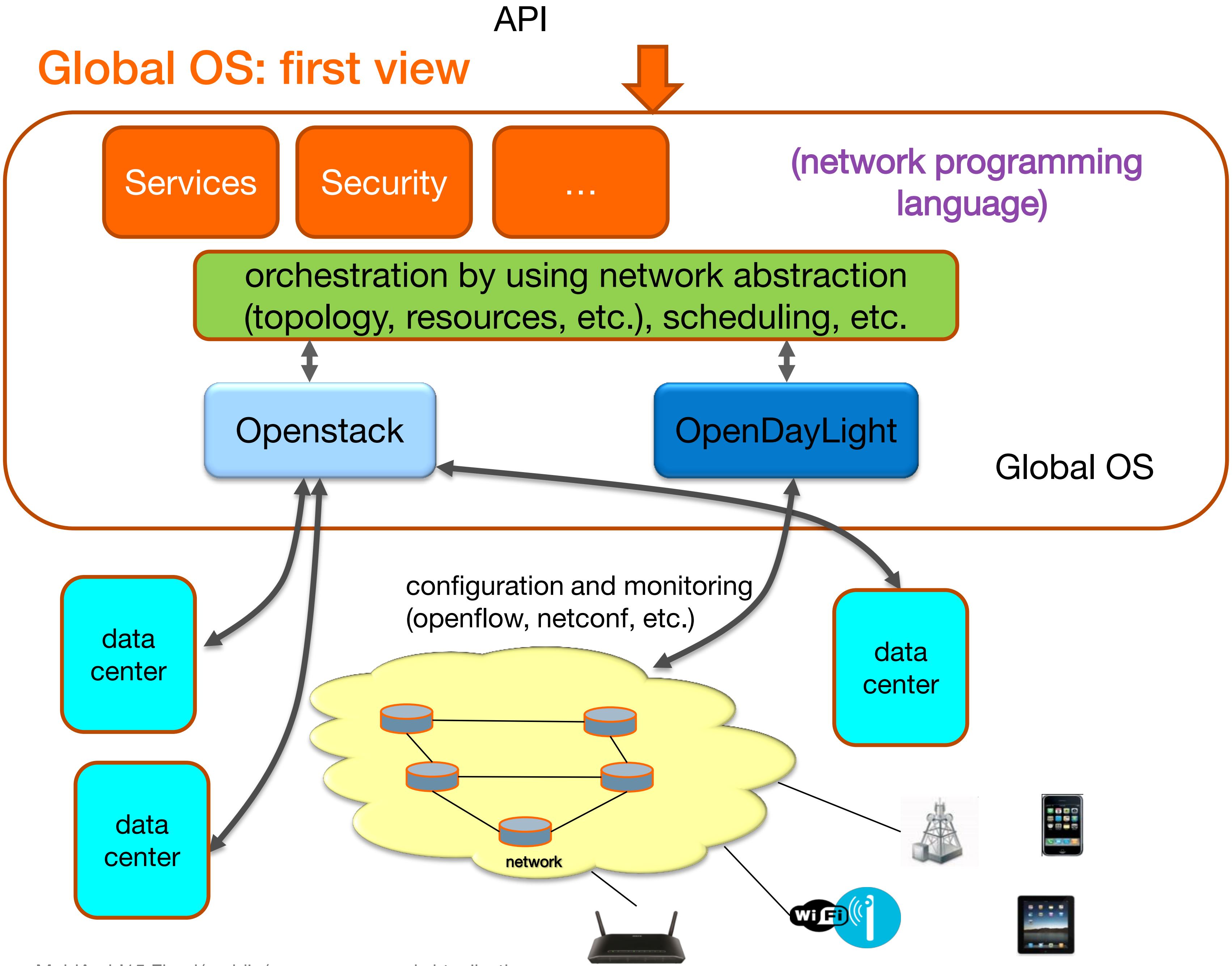
Configuration of a CGW (OpenStack + ODL)



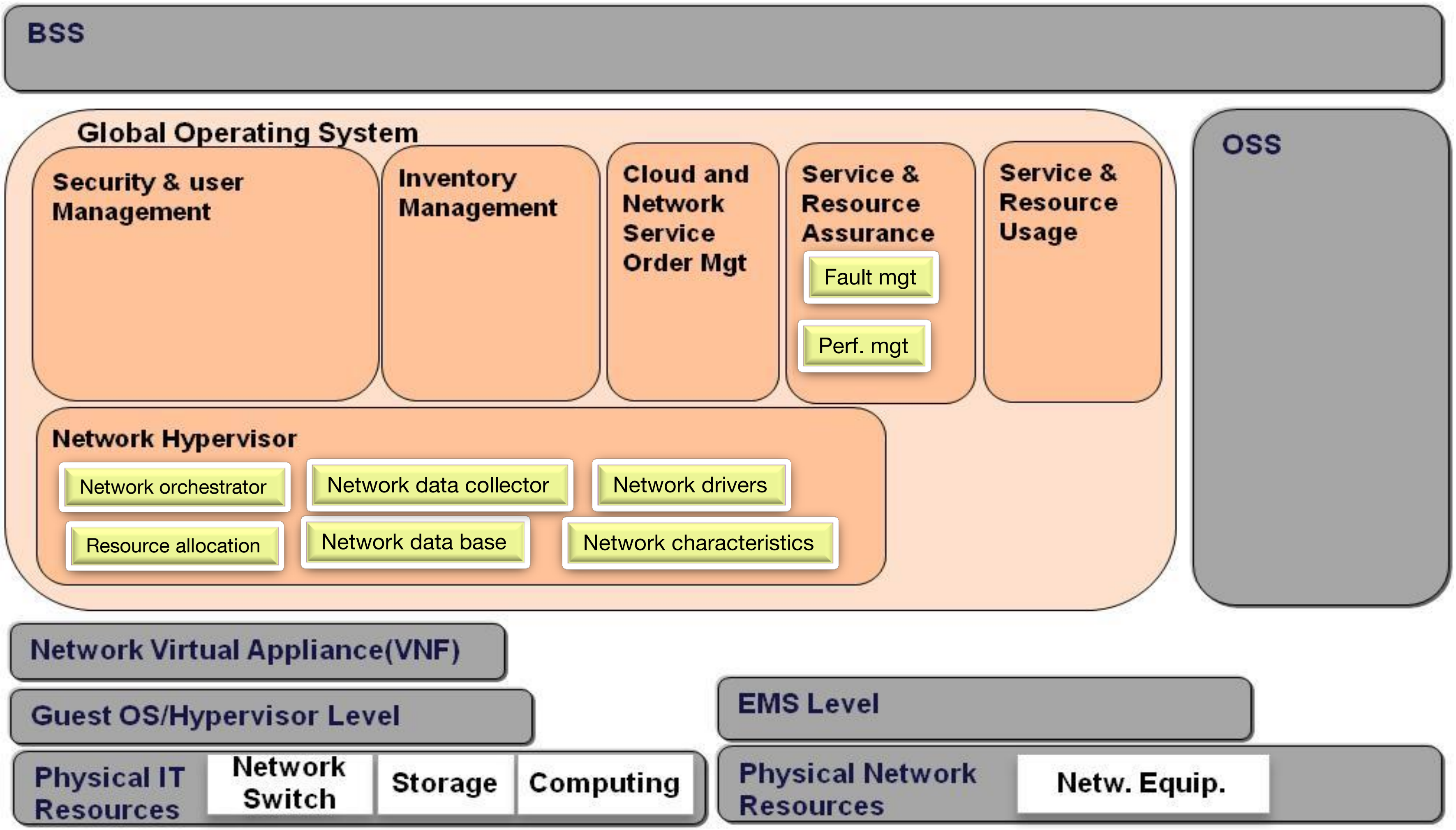
Configuration of CGW (orchestration)



Global OS: first view



GlobalOS framework



Network programming language

- Very active domain of research
- Many languages have been proposed so far
  - static approach
    - NetKAT (Kleen Algebra with Test): express network procedures into an equational system (for formally proving properties of procedures)
    - NICE
    - MERLIN
  - dynamic approach
    - Kinetic
    - VeryFlow
- Most languages are packet based and do not include resource allocation aspects

Conclusion

- Virtualization techniques offer new possibilities for networking
  - instantiation of VMs on the fly for specific tasks (in particular NFV)
  - Convergent gateways: package of VMs realizing control for fixed/mobile convergence (vEPC, AAA, DHCP, convergence functions, etc.)
- Convergent gateways raise many issues
  - addressing, AAA, and mobility management
  - new charging scheme (more difficult to count traffic per user, capped offers vs. usage/network conditions)
  - monitoring: Is it possible to instantiate network probes on VM instead of dedicated hardware?
- Convergent gateways can be included in a more general framework: GlobalOS
  - a global OS acts as the OS of a computer but at a network scale
  - need for abstraction of network resources
  - resource management, language for network configuration, etc.

Thank you