

Terastream

Implementation of All IP New Architecture

B. Zaluški, B. Rajtar*, H. Habjanić, M. Baranek, N. Šlibar, R. Petračić, T. Sukser

Croatian Telecom, Zagreb, Croatia

* Deutsche Telekom, Bonn, Germany

Abstract: In this paper we describe Terastream all IP architecture including market and economic drivers, IPv6 addressing scheme, routing and tunneling protocols, interfaces and optical integration, front end data centers, home networking and performance aspects initial findings. We point out standardization efforts where we are contributing and present initial findings after actual implementation of pilot network.

I. INTRODUCTION

Work on new IP architecture started in 2011 within Deutsche Telekom. In 2012 pilot Terastream network was implemented by Croatian Telecom. In 2013 we will continue research activities and provide additional contributions to standardization work performed by IETF and other standardization organizations.

This paper is describing architecture, implementation and initial findings during implementation. New research activities are only mentioned where appropriate.

II. MARKET AND ECONOMIC DRIVERS

Over the top providers, cable operators, mobile operators and fixed operators are in fierce competition. They are facing massive IP traffic growth driven by improved broadband access technologies, new Internet services and new business models. Mobile and Fixed operators are moving from legacy multi-layer networks with too much complexity.

In order to reduce complexity and enable service innovation, more features, faster provisioning resulting in revenue increase — operators need simplification and one truly converged de-layered network to enable cloud era economics.

With Internet traffic growth forecasts increasing from conservative 10 times to more than 50 times in next ten years operators are facing major changes. Ongoing migration from IPv4 to IPv6 is adding to complexity. In order to get revenue and cost balance right, new service delivery model based on simplified IPv6 network architecture is needed.

III. ARCHITECTURE OVERVIEW

Traditional models produce IP connectivity using separate networks for telecom services and other providers in a way that is too costly to provide for traffic growth. In

order to enable more economical IP connectivity Terastream relies on following design principles:

- Reduce the amount of technologies used
- Use IPv6¹ only for all internal functions and services
- Size the network to handle IP traffic without packet losses
- Integrate optical and IP network as much as possible
- Avoid internal interfaces
- Use one network for all services

Architecture is using two types of routers. One we call R2 - they deal with peering, data center connectivity and are meshed among them self to produce telecom network. They have dual stack capability IPv4/IPv6, huge number of 100GE ports to connect so called R1 routers in a horseshoe and number of 10GE ports for directly connecting front end data center computers.

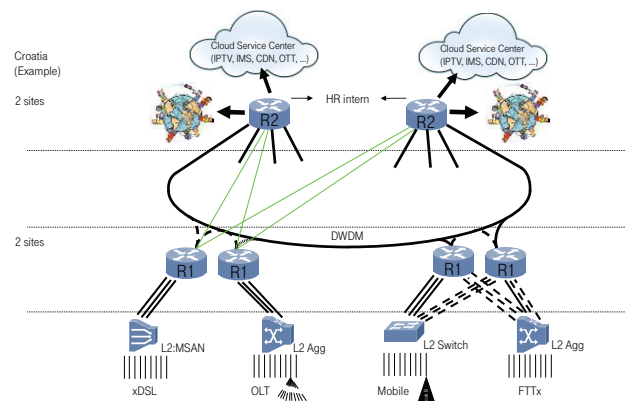


Figure 1. Terastream Architecture Overview

In order to benefit from x86 architecture, network services are implemented using off the shelf servers connected directly to R2 routers. Services like L2 and L3 virtual private networks or IPv4 through DS-Lite² tunnel termination require throughput and low latency offered by front end data center infrastructure. Store and forward services like mail or messaging are implemented in regular data centers.

IV. IPV6 ASPECTS, ROUTING AND TUNNELING

IPv6 address uses 128 bits which enables huge address space with near 3.4×10^{38} available addresses. This solves shortage of IPv4 address and its address translation mechanism (except in some special cases).

There are different address scopes in IPv6: local, global addresses and Unique Local Addresses - ULA. ULA addresses are equivalent of IPv4 private addresses.

Addresses can be generated automatically by the client without preserving state using *SLAAC* – *Stateless Autoconfiguration* which relies on RA (*Router Advertisement*) messages from the router. DHCPv6 protocol can also be used for better control.

Service providers used to build hybrid networks using Ethernet in aggregation part and IP/MPLS in the core part.

TeraStream is pure IP network which lowers the number of used protocols and simplifies network management.

More specifically TeraStream is pure IPv6 network because all other protocols are tunneled (if needed). That includes IPv4 protocol which becomes only one of the network services. Tunneling is done between the Home Gateway and the network AFTR (Address Family Translation) service that sits in the Data Center (using DS-Lite or Lightweight 4over6) as shown on Figure 2.

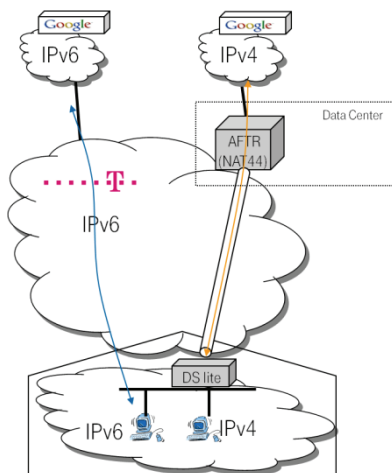


Figure 2. IPv4 as a service over IPv6 using DS-Lite

Routing in TeraStream is done using familiar IS-IS and BGP (Border Gateway Protocol) protocols. IS-IS protocol is used only for exchanging loopbacks and interface addresses. Whole network is one IS-IS Level 2 domain. All customer prefixes are exchanged using BGP with R1 routers being RR clients of R2 routers. R2 routers can have external BGP peerings.

In TeraStream concept IPv6 address bits (Figure 3.) have special meaning attached to them:

- Defining if address is public, belongs to infrastructure or service (so called PIE bits),
- Defining service type (so called SSS bits, they have values for Internet, IPTV or other service)
- Defining user location (router, interface, VLAN)

Examples:	Source PIESSS	Destination PIESSS
User -> IMS	000110	011110
IMS -> User	011110	000110
User -> User (best effort)	X00001	X00001
User -> Internet (best effort)	100001	XXXXXX
Internet -> User (best effort)	XXXXXX	100001
Lan-Lan service	010101	010101

R Registry 32 bit, Registry/IANA assigned
P Public 1 bit, 0-traffic internal to Telekom, 1-traffic external to Telekom
I Infrastructure 1 bit, 0-end user traffic, 1-infrastructure packet
E Endpoint/Service 1 bit, 0-network endpoint, 1-service
S Service type 3 bits, 0-res, 1-internet, 2-res, 3-res, 4-video, 5-L2 service, 6-voice, 7-mgmt
a RI Area 5 bits, Indicates what RI that the address is delegated from, max 32 RI
p RI User 13 bits, User Identifier
u User subnet, Delegated to user

Figure 3. TeraStream IPv6 addressing PIESSS examples

Each customer gets /56 prefix for each of the used services (Internet, IPTV, etc.). Concept enables not only connection of the HGW but also end devices to the TeraStream network.

Idea is to also enable more complex topologies so the home network can have connections to multiple service providers or just be multihomed.

Multi-functional end devices can get multiple address prefixes, one for each service. Current IPv6 standards use longest prefix match rules to choose the right source prefix for the destination. As such approach has many limits suggestion is to use so called “coloring” mechanism which will “color” the prefix given to the customer. Each device or application will then be able to choose the right “color” to reach the desired service prefix (that will also be colored). Policies could be preconfigured or downloaded from network management server.

V. IP AND OPTICAL INTEGRATION

Key idea is integration of 100G transponders in routers without classical Dense Wavelength Division Multiplexing – DWDM – components: (ROADM, MUX, DEMUX) using only splitters and amplifiers. This is now possible using coherent tunable transponders which can tune its transmitters and receivers to any of 80 optical channels.

With this approach we can avoid double optical/electrical conversion and reduce number of interfaces as shown on Figure 4 with separate transponders and on Figure 5 with integrated transponders.

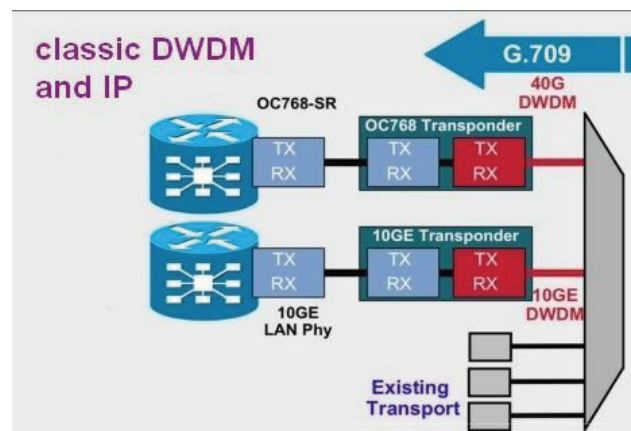


Figure 4. Classic concept with DWDM and IP separated ports

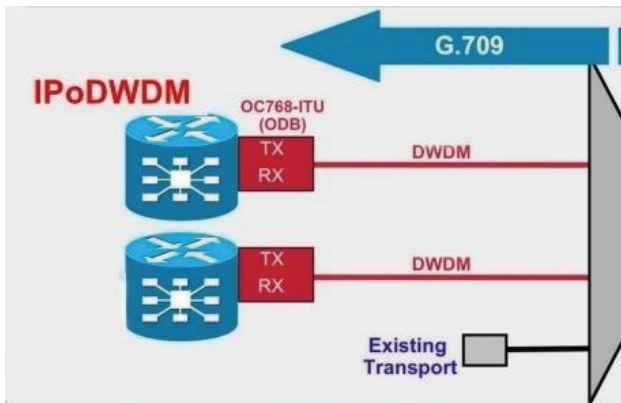


Figure 5. IPoDWDM with integrated transponders

Client ports in router and transponder are not needed since transponder is integrated in router card. This concept reduces CAPEX costs and provide improved resilience of the network (less OPEX cost).

Each R1 (max 8) site has a splitter combination to add / drop channels and is connected to two different R2 routers to enable services in case of fiber interruption. Optical connectivity is shown on Figure 6.

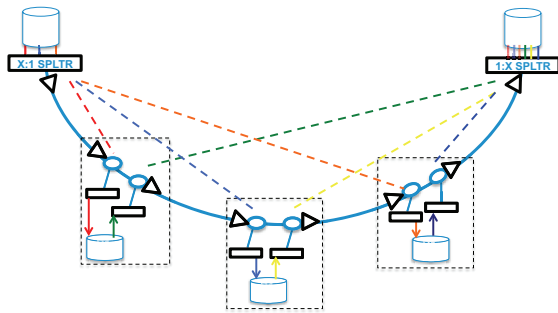


Figure 6. TeraStream optical connectivity

All circuits are 100Gb/s³ Ethernet point-to-point utilizing 50GHz spacing. Typical overall distance of the horseshoe is 600Km.

The solution must also be adapted to the Core network of meshed R2 sites (not depicted).

Each R1 site is equipped with 8 optical channels. We start with utilized two 100Gb/s channels per R1 site.

The DWDM network leverage a “Drop and Waste” architecture where every wavelength that is dropped at an R1 site will be wasted or not available to upstream nodes as shown on Figure 7.

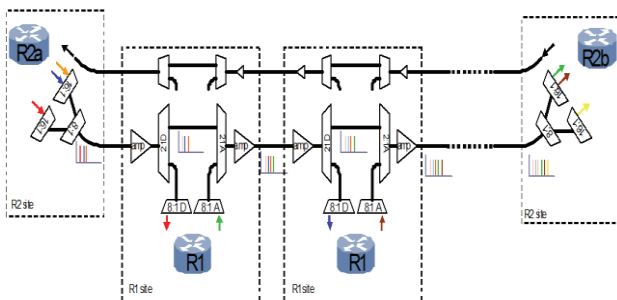


Figure 7. TeraStream “Drop & Waste” architecture with splitters

Important fact in this architecture is that every optical channel is available at every transponder due to the usage of splitters and not filters as in classical DWDM system. With previous transponder technology this architecture would be impossible to implement.

Coherent technology enables us to simplify optical network using only splitters and amplifiers from classical DWDM systems eliminating need for expensive ROADMs, MUX/DEMUX and other components.

This concept is applicable in presented “Horseshoe” topology with 600km span limit.

VI. FRONT END DATA CENTER

Data center directly attached to the TeraStream network is called front end data center. The name was derived from its function. Such data center primarily runs front end (customer facing) services. Data centers traditionally offer compute, storage and network resources, while here it is a platform for hosting:

- basic network service
- voice (VOIP), television (IPTV)
- content delivery network

The front end data center is designed in form of cloud, which enables high flexibility, agility and scalability for the services hosted in it. By using standardized cloud platform⁴ with well-defined API (Application Programming Interface) it is possible to decouple hardware from software, so that servers from different vendors can be used in the same data center more easily. This approach would enable easier replacement of hardware in future.

Services which are running in the data center are delivered in form of virtual machines, where each virtual machine would contain the software for the service. Each service can run its own operating system which is the best for it, reducing the compatibility issues.

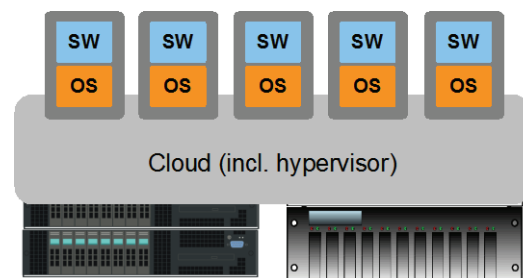


Figure 8. decoupling hardware from software using hypervisor

Having a single data center might be beneficial for easier management, but in order to bring services closer to customers, it is necessary to have the data center at each R2 site. Management of multiple data centers, each running lots of services, is a nightmare and requires a lot of people involved.

By using orchestration tools, it is possible to reduce human errors, improve service delivery times and reduce service outages. The most important functions of the orchestrator:

- configuration of the whole data center from the scratch,
- start and configuration basic data center and network services (DNS, NTP, orchestrator, logging facilities),
- deployment and configuration of a service (as a virtual machine or set of virtual machines),
- tearing down a service (removal from the data center),
- detection and correction of service, virtual machine or physical server failures,
- communication with other network management systems in order to have entire end-to-end overview,
- reporting of data center usage and trend prediction.

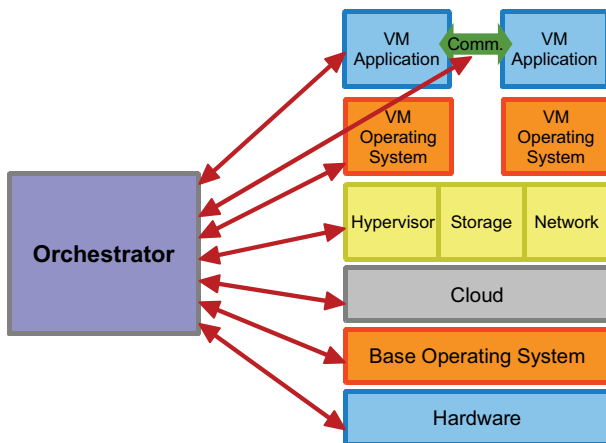


Figure 9. Orchestrator interface

Orchestrator acts on different layers as shown on Figure 9: hardware, base operating system on physical servers, cloud software (including hypervisor, storage and network resources), virtual machines as virtual appliances, operating system inside virtual machines, applications inside virtual machines, and helper for establishing communication between various virtual appliances.

VII. HOME NETWORKING

With the introduction of IPv6 in the home, customers now receive 2^{72} IP addresses they can use for networking in the home. Comparing to the IPv4 addressing space, this means the customers can connect more devices at their home than they are existing devices on the Internet today reachable via IPv4. This allows for all sorts of opportunities such as home automation, connecting all home appliances into one inter-connected network, etc.

The addressing space for IPv6 is a public addressing space which means the customer is directly available from the Internet and is “on” the Internet rather than “connected” to the Internet. Unlike IPv4, where the lack of addresses implies that each customer has only one public IPv4 address, the whole customer IPv6 addressing

space is public. NAT – Network Address Translation from the single public IPv4 address to the private addressing space (typically 192.168.1.0/24) used for IPv4 home networking brings a lot of limitations regarding home accessibility from the Internet and is non-existent with IPv6.

The typical scenario of Internet connectivity uses a Home Gateway as a central point of home administration and a point of interconnection. However, the TeraStream architecture allows that the customer connects himself directly to the R1 router and uses only low-end switches to build their own home network. However, this is possible only with devices which are IPv6 capable.

The Home Gateway is used for three primary reasons:

- Firewall used for the protection of the home against malicious attacks from the Internet;
- Access point for wireless connectivity.
- IPv4 connectivity.

Since nowadays a lot of existing equipment are IPv4-only, the Home Gateway is used as a endpoint for tunneling customer IPv4 traffic. This is accomplished using DS-Lite, a protocol that encapsulates customer IPv4 traffic (from the private address space) and sends it to the data center where NAT to the public IPv4 address is performed. The traffic is then forwarded to its final destination on the IPv4 Internet.

VIII. CONCLUSION

TeraStream pilot in Croatian telecom was implemented in less than three months. It proved number of new IPv6 technologies including DS-Lite — by using front end data center it was shown that it is possible to implement network functions like DS-Lite on standard PC hardware in cloud environment — as well as that it is possible to build 100GE coherent drop & waste optical system with integrated optics managed by routers on IPv6 only network.

ACKNOWLEDGMENT

We would like to thank colleagues from Deutsche Telekom that made simplification and improvements of traditional IP architecture possible in TeraStream architecture drafted number of early design documents that started pilot in Croatia: Peter Lothberg, Axel Clauberg, Rainer Schatzmayr, Guenter Honisch and many others.

REFERENCES

- [1] Hinden, RM. "RFC 4291 - IP Version 6 Addressing Architecture." 2006.
- [2] Durand, Alain et al. "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion." draft-ietf-software-dual-stack-lite-04 (work in progress) (2010).
- [3] Winzer, Peter. "Beyond 100G ethernet." Communications Magazine, IEEE 48.7 (2010): 26-30.
- [4] 26-30.Computing, Rackspace Cloud. "OpenStack Open Source Cloud Computing Software." 2010-12-10. <http://openstack.org> (2012).