MEVICO
Mobile Networks Evolution for Individual Communications Experience

CELTIC
Telecommunication Solutions

Project Number:	**CELTIC / CP7-011**
Project Title:	Mobile Networks Evolution for Individual Communications Experience – MEVICO
Document Type:	D (Deliverable)

Document Identifier:	D3.2
Document Title:	**Innovative Solutions for Mobile Backhaul**
Source Activity:	WP3
Main Editor:	Jose Costa-Requena
Authors:	Pekka Wainio, Tuomas Taipale, Nicklas Beijar, Eyal Ezri, Carmi Zisapel, Mounir Kellil
Status / Version:	Final draft submitted to QEG/0.9
Date Last changes:	08.11.2012
File Name:	D3_2_v0.9.doc

Abstract:

This document describes the technologies proposed for the transport layer in mobile networks. The different technology options are presented together with validation results showing the proposed performance improvements.

Document History:	
27.2.2012	Document with draft TOC created
10.3.2012	Aalto contributions
5.6.2012	Integrated CEA contribution, added RAD topics
24.9.2012	Integrated RAD contribution
24.10.2012	Updated Aalto contribution (on CES)
30.10.2012	NSN contribution, CEA updates
1.11.2012	NSN contribution updated, intro improved, Some editorial clean up
6.11.2912	Document reviewed at WP3 and ready for QEG evaluation

Table of Contents

Authors

Partner	Name	Phone / Fax / e-mail
AALTO University	Jose Costa-Requena	
		Phone: +358505770142 e-mail: jose.costa@aalto.fi
Aalto University	Nicklas Beijar	
		Phone: +358 50 400 6184 e-mail: nicklas.beijar@aalto.fi
NSN	Tuomas Taipale	
		Phone: +358407561478 e-mail: tuomas.taipale@aalto.fi
NSN	Pekka Wainio	
		Phone: +358405811211 e-mail: pekka.wainio@nsn.com
CEA	Mounir Kellil	
		Phone: e-mail: mounir.kellil@cea.fr
RAD	Eyal Ezri	
		Phone: e-mail: eyal_e@rad.com
RAD	Carmi Zisapel	
		Phone: e-mail: carmi_z@rad.com

Executive Summary

The purpose of the document is to present all the technologies proposed to address the challenges and requirements for transport network. The key features and validation results are included to indicate the expected performance improvements with the proposed technologies.

List of acronyms and abbreviations

ALG	Application Layer Gateway
ARP	Address Resolution Protocol
BRAWE	Broadband multi-antenna radios for millimeter wave frequency bands;
BTS	Base Station (refers to any RAT)
CA/AA	Certification Authority / Attribute Authority
CE	Customer Edge
CES	Customer Edge Switch
CETP	Customer Edge Traversal Protocol
CMOS	Complementary Metal Oxide Semiconductor
DHCP	Dynamic Host Configuration Protocol
DHT	Distributed Hash Table
DNS	Domain Name System
DPI	Deep Packet Inspection
DSCP	Differentiated Services Code Point
EAP-AKA	Extensible Authentication Protocol Method for UMTS Authorization and Key Agreement
eNB	Evolved NodeB
EPC	Evolved Packet Core
FAP	Femto Access Point
FDD	Frequency Division Duplexing
FTP	File Transfer Protocol
FQDN	Fully Qualified Domain Name
HeNB	Home eNodeB
HIP	Host Identity Protocol
HTTP	Hypertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
ID	Identifier
IKE	Internet Key Exchange
IP	Internet Protocol
KVM	Kernel-based Virtual Machine
LTCC	Low temperature co-fired ceramic
LTE	Long Term Evolution
MAC	Media Access Control
mmW	Millimeter Wave
MPLS	MultiProtocol Label Switching
NAT	Network Address Translator
NGMN	Next Generation Mobile Networks
O&M	Operation and Maintenance
OAM	Operations, Administration and Maintenance
PCP	Priority Code Point
PE	Provider Edge
PRGW	Private Realm Gateway
RAT	Radio Access Technology
RLOC	Routing Locator
SA	Security Association
SDK	Software Development Kit

SDP	Service Description Protocol
SeGW	Security Gateway
SIP	Session Initiation Protocol
SON	Self-Organizing Network
SPT	Spanning Tree
SQL	Structured Query Language
STP	Spanning Tree Protocol
STUN	Session Traversal Utilities for NAT
SW	Software
T2T	Trust-to-Trust Protocol
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TLV	Type-Length-Value
TLS	Transport Layer Security
TRILL	Transparent Interconnection of Lots of Links
TURN	Traversal Using Relay NAT
UDP	User Datagram Protocol
VLANID	Virtual Local Area Network Identifier
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
WMN	Wireless Mesh Network

1. Introduction

Mobile data volumes have surged during past few years and are expected to grow exponentially in the near future. Behind this trend there is a mixture of factors including the new smart phone era and generally a cheaper access to mobile broadband. Along with this development, mobile network technologies have been evolving quickly to handle the massive mobile data volume growth. The latest mobile communications standard, LTE-Advanced (Long Term Evolution), is able to offer data speeds from hundreds of megabits all the way to the gigabit mark, which can be regarded as a genuine broadband mobile access. This is mainly achieved with more advanced radio interface technologies and by streamlining and simplifying especially the core part of the system, the whole system converging towards an all-IP (Internet Protocol) network.

At the same time the mobile operator’s revenue growth is declining due to e.g. flat free pricing models. This will lead to a cost pressure especially to transport network, which is a major cost item for mobile operators. The mobile backhaul network must be updated to accommodate this expected traffic growth along with the ever increasing demand for lower cost and optimized transport solutions.

The focus of Work Package 3 is on the evolution of the transport networks towards packet-based infrastructures and new architectures as well as for support of large and evolving set of services and applications requiring more bandwidth and having specific QoS requirements.

The purpose of the document is to present all the new transport related technology innovations proposed to address the challenges and requirements set for transport network planned to serve the future mobile communication systems like LTE-Advanced and beyond. The key features and validation results of the proposed technologies are included to indicate the expected performance improvements achieved with these technologies.

2. Wireless Mesh access backhaul for small cell base stations

In this chapter we describe a mobile backhaul networking concept especially targeted for LTE and LTE-Advanced small cells deployments with inherent self-healing, self-optimization and self-configuration capabilities based on mesh topology utilizing millimeter wave radio links.

2.1 Introduction

The concept of heterogeneous network deployments has been introduced in LTE-Advanced specifications. In heterogeneous networks the coverage area of a macro cell base station is complemented with smaller coverage base stations, i.e. small cells, to better target high data rate demand hot spot areas such as city centers by offloading some of the macro data traffic to these small cells. Evidently small cells are the most prominent answer to future capacity and coverage shortage problems.

However, the concept of small cells itself introduces a set of challenges. Accurate frequency and other network planning are required, since poorly deployed small cells can in the worst case destroy capacity through interference. Also, the coordination between macro cell and small cells need to be efficiently carried out in order to minimize excess signaling and other control traffic. One of the most outstanding issue with small cell deployments is the backhaul solution i.e. how to efficiently transport the small cell access traffic to the core network. Since every small cell base station needs to have a high capacity last hop backhaul connection, the sheer amount of backhaul units will increase heavily. In addition, as the deployment places for small cells and subsequently backhaul units move closer to street levels, factors that have not traditionally affected telecommunications equipment will have larger impact. These include, among other things aesthetics (the small cell equipment including backhaul unit should blend into the street-scape because small cells will generally be located in and around the areas where people meet, eat, entertain and gather), increased change of signal blocking (due to e.g. tall vehicles and trees) and increased pole sway (lamp post assembly vs. traditional broadcast masts). Even so, the small cell backhauling solution should still be able to fulfill the LTE-Advanced requirements with decent Quality of Service, availability, capacity etc., yet the installation and operational costs should be as low as possible.

At the moment, there does not seem to be any general consensus of how exactly the future small cell backhaul should be implemented. Industry forums such as Small Cell Forum and NGMN (Next Generation Mobile Networks) help to move the industry forward by clarifying consensus around the operators' requirements for small cells and have among other things published extensive requirements reports for future small cell backhaul systems, including requirements on backhaul capacity, connectivity, resiliency, Quality of Service, synchronization and security. However, the requirements obviously do not take a stand on how the different features should be implemented in practice. Therefore, there is a need for investigating smart and flexible transport solutions for small cells.

An innovative new solution called Wireless Mesh Network (WMN) has been developed as part of MEVICO Work Package 3 to tackle these challenges. The main research partners have been Nokia Siemens Networks (NSN) and VTT Technical Research Centre of Finland in co-operation with Aalto University during the concept validation. The WMN system consists of a set of nodes partially meshed with each other forming an independent transport sub-network. The WMN backhaul nodes are connected to each other with directional point-to-point wireless links. However, other types of communications media, such as fiber, are also supported.

Generally, there exists a lot of research on wireless mesh networks for different applications. The most popular standardized mesh technologies are the IEEE 802.11s WLAN mesh (Wireless Local Area Network), IEEE 802.15.5 WPAN mesh (Wireless Personal Area Network) and IEEE 802.16 WiMAX mesh (Worldwide Interoperability for Microwave Access) variants. They all offer quite complete mesh solutions for different scope, but the applicability for backhaul on the other hand is not optimal in most cases. For example, WPAN solutions have a variety of mobility management features which are unnecessary features in a static small cell deployment and WiMAX meshes provide some applicable features such as centrally controlled scheduling features. In spite of the vast amount of proposed solutions and research material for wireless mesh networking, present state-of-the-art solutions seem to only tackle one or only a few problem areas of mesh networking (e.g. Media Access Control (MAC) scheduling, protection techniques, Quality of Service etc.) leaving the system level procedures omitted or vaguely defined. In addition, the proposed solutions are not directly applicable to fulfill the small cell backhaul specific requirements without extensive modifications.

The mesh network solution presented in this document aims to tackle the above small cell backhaul challenges with extensive inherent SON (Self-Organizing Network) capabilities, i.e. self-healing, self-optimization and self-configuration, build into to the meshing protocols. A detailed introduction to the solution is presented in the next chapters.

2.2 Detailed description of the proposed solution

The WMN (Wireless Mesh Network) backhaul solution is a novel concept solution targeted for next generations' small cell, ultra high capacity mobile base station first mile access backhaul. Essentially the WMN system is a highly transparent sub network offering connectivity between desired end points (e.g. a set of base stations and an aggregation transport network gateway) with advanced and smart self-optimization, self-healing and self-configuration capabilities offering rich mix of traffic engineering and configuration features but requiring little or no OAM intervention.

The WMN sub network consists of a set of wireless mesh backhaul nodes in the range from 20 to 200 in partial mesh topology. The WMN backhaul elements are connected to each other with directional point-to-point links. The WMN nodes offer a Layer 2 transport service for the client systems, e.g. small cell base stations or fixed broadband equipment. The WMN sub network is connected to external transport networks through special gateway elements, and all traffic coming in and out of the WMN sub network will traverse through these gateways.

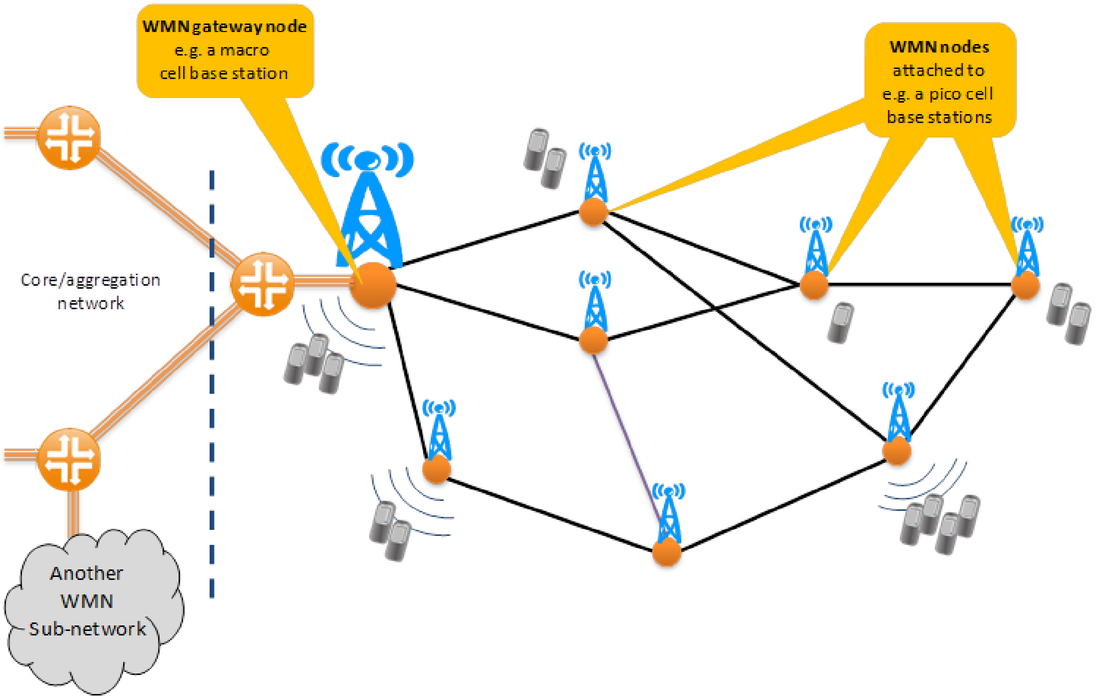


Figure 1. Wireless Mesh Network example

The WMN employs a comprehensive and automated resiliency scheme that aims to reduce the overall impact of link and equipment failures on the WMN subnetwork caused for example by rain outages or other line-of-sight blockages.

In addition the WMN system offers a wide array of self-optimization features, for example a flexible Quality of Service scheme, congestion control and management mechanisms as well as extensive load balancing and traffic management features. These mechanisms allow a highly flexible control and steering of the traffic flows inside the mesh network thus enabling automated dynamic QoS aware optimization of the traffic at any given time so that the entire transport capacity (all routes) of the mesh network can be optimally utilized.

Finally, the WMN system offers a set of self-configuration mechanisms, automating everything spanning from smart adjustment of the point-to-point wireless links to complete network startup.

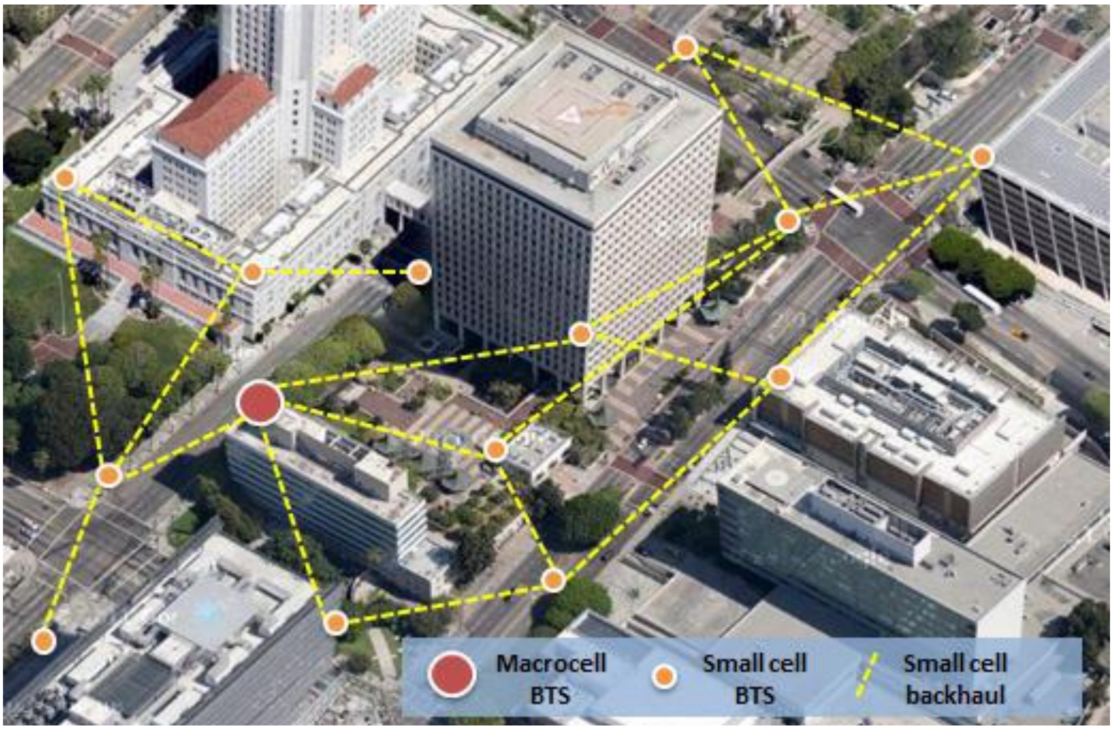


Figure 2. Example deployment for small cell backhaul. The backhaul hop lengths vary from 200 to 300 meters.

In typical mobile backhaul use case the WMN nodes are expected to be tightly integrated to the base station site solution. One deployment example is given in the figure above.

2.3 Description of validation system

The goal of the validation was to prove the feasibility of the WMN concept for backhauling LTE and LTE-Advanced base stations. The main aim was to proof that the WMN concept is able to perform satisfactorily under likely traffic, load and network situations on a system level. Especially the target was to show that

- transparent Ethernet transport service from client nodes to the WMN gateways are provided with fixed network quality in variable network conditions,
- the transport capacity of the network can be optimally shared between the base stations connected to the WMN subnetwork,
- the backhaul performance requirements of LTE and LTE-Advanced are fulfilled over 3-5 wireless hops and
- resiliency to different network failure situations can be provided.

Due to the advanced nature and novelty of the WMN system concept, the functionality and feasibility of the whole concept system required practical prototype testing. Although simulation-based solutions can offer valid and perhaps a wider variety of results and experiments the implementation of a suitable simulator can be rather challenging. Present network simulator tools only include the most common standardized and used networking elements to date. Adding new networking protocols and elements can be highly laborious e.g. due to different programming language interdependencies.

The WMN concept was targeted to offer multi gigabit data throughput rates. Thus it was deemed necessary to utilize network processor-based platforms in validation since they combine the flexibility of a general-purpose processor without sacrificing any of the packet processing capabilities of dedicated chips manufactured specially for packet forwarding. Commercial off-the-shelf networking devices such as Juniper or Cisco would offer the needed hardware accelerated processing capabilities but in the end could not be utilized due to the closed software environments. In the end, the Octeon network processor from Cavium Networks was chosen for as the network processor platform.

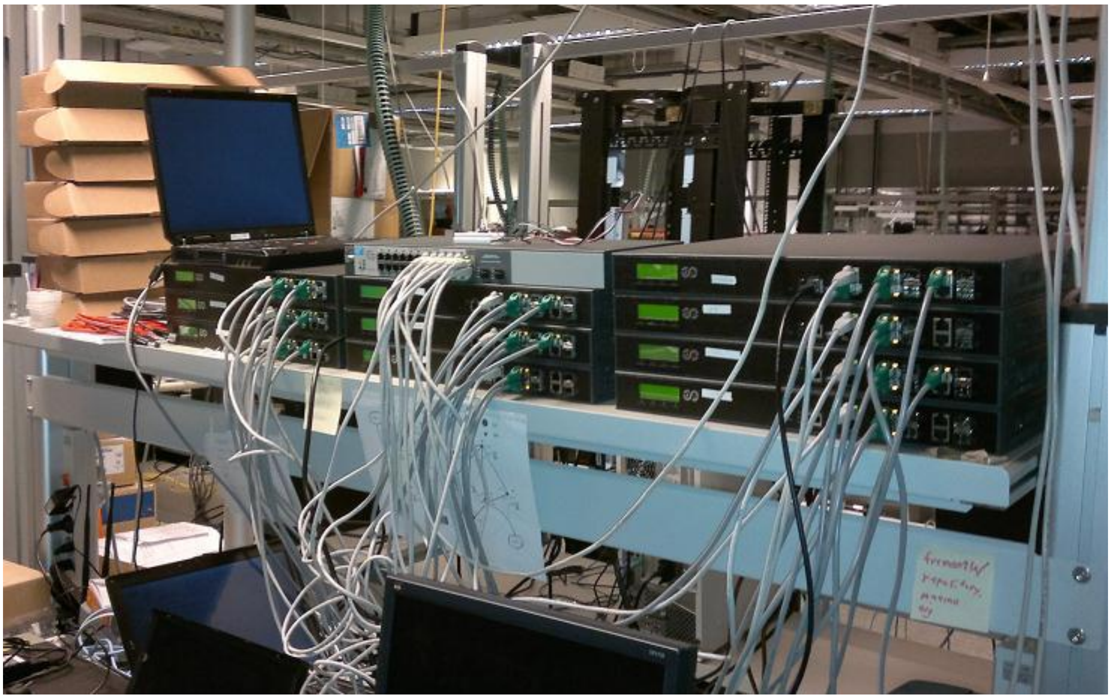


Figure 3. Ten node WMN system set up in the Nokia Siemens Networks Mobile Backhaul Research laboratory

The final WMN proof-of-concept system was built based on Lanner MR-730 networking appliance units with four core Octeon network processors each running the experimental wireless mesh protocol software. The Lanner network processing units were connected together to form the partial mesh network. Several different network topologies and configurations up to 20 WMN nodes were created and tested during the validation.

The wireless connections between different WMN nodes were mostly emulated as a set of Ethernet cable connections between the Lanner MR-730 units. In addition to verify the operation with a real wireless link, two hops in the test topology were implemented with an experimental electrically beam steerable E-Band radio systems provided by Finnish publicly funded research project called BRAWE (Broadband multi-antenna radios for millimeter wave frequency bands). The BRAWE millimeter wave (mmW) radio system prototype is a combination of research efforts by VTT and Aalto University's Department of Micro and Nano-technology and Department of Radio Science and Engineering. Advanced research on e.g. CMOS (Complementary Metal Oxide Semiconductor) transistor technology for millimeter wave applications, LTCC (Low Temperature Co-fired Ceramic) chip packaging and lens antennae technologies culminated in the creation of one prototype system working in the 80 GHz millimeter wave band. The BRAWE radio system integrated into the WMN proof-of-concept consisted of two static transmitters without beam steering capabilities and one receiver unit that employed the experimental mmW lens antenna and beam steering capabilities.

The actual protocol functionality of the WMN system was implemented in the form of prototype software developed using Cavium SDK (Software Development Kit) and Linux. The structure of the software follows the basic principles of any routing software and hardware combination. There is a separate data plane that is running on one of the Octeon network processor cores and a control plane that is running on another core.

The practical concept and WMN protocol software development along with the parallel validation work took place over the period of twelve months in total. Overall, the process was incremental and consisted of a set of phased functionality milestones, all adding up towards the final version of the prototype protocol software. The functionality phases included mechanisms such as basic routing and scheduling, link break protection and traffic management. The general aim in all phases was to verify the correct operation of the mechanisms under likely real life traffic and network situations. Thus for example the basic routing testing included testing with a varying traffic profile, spanning from best effort file transfer to more demanding real time traffic and with different virtual connections active.

Video downloading from external server was used as the demonstration use case scenario to show the system performance and the main features of the concept. The demonstration scenarios included various congestion and network fault situations. Test data generators like Spirent were used when applicable to verify the end-to-end system functionality and measure performance.

2.4 Main results

Overall, the system validation process progressed quite smoothly and the proof-of-concept environment performed exemplarily. Complex and fatal bugs were non-existent, and the few bugs that bothered the validation process in the beginning were eventually sorted out to originate from the platform restrictions, having nothing to do with the concepts of the WMN system. Also, it is fair to mention that the prototype protocol was extremely high quality software and had relatively small amount of mainly minor bugs. This is largely thanks to precise and accurate specification documentation by NSN and experienced programming carried out by VTT.

Basic connectivity and networking through the WMN system was implemented with the novel spanning tree and scheduling principles. Overall, both mechanisms worked as originally specified. Incoming data on the ingress of the WMN system gets mapped correctly according to virtual connection mapping tables defined in the software configuration. In case of incorrect or undefined VLAN tagged data in the ingress, the data packets were simply dropped. With the properly working routing scheme, the prototype protocol can be easily amended with functionalities utilizing the spanning tree-based routing. The scheme also allows possible traffic engineering features as certain VLANID mappings can be configured to follow suitable spanning trees or in turn special spanning trees can be created to overlay the network in a particular way.

As the protocol implementation is wholly software based, timing accuracy with the scheduling principle can be off occasionally, though with further software optimization and hardware accelerated processing of the main functionalities, the targeted millisecond scheduling times and microsecond level recovery times should be easily achievable.

As part of the basic connectivity testing the BRAWE radio system was successfully integrated to the WMN system. The hardware used along with BRAWE was not entirely optimal for the task, thus there were some deviations from the planned concept in terms of e.g. time scales. Generally though, the successful integration of the BRAWE radio system was a significant milestone in the concept verification process. It meant that the novel concepts of network-wide scheduling principle and shared resources are functional and feasible technologies and that they can be implemented in practice with real radio hardware. The basic transport mechanisms developed for the WMN system are thus entirely feasible and realizable in practice.

Self-healing testing included basic path protection and fast reroute verification. All in all, link break detection, link break signaling and the end-to-end path re-selection features work as specified. Data can be forwarded between a source and destination pair as long as there is a spanning tree connection between them. The path re-selection reroutes traffic traversing a breaking spanning tree but other traffic is not affected in any way. Moreover, the fast rerouting mechanism provides a clear improvement in the packet forwarding capabilities of the WMN system in link break situations. The positive impact of the mechanism is likely to only grow with higher data rates and larger topologies as the failure notification messages take naturally more time to propagate in multi-hop topologies.

Self-optimization testing included verification of the dynamic traffic management features of the WMN system (i.e. congestion control, route self-optimization, traffic and load balancing mechanisms) with different priority class traffic flows under varying network load situations. Proper traffic and load management requires the combined and correct functionality of Quality of Service tagging, congestion detection and control and a priority-based traffic flow handling. Overall, the self-optimizing features were found to be working correctly. The software can distinguish different priority traffic flows and detect different levels of congestion and based on these, make load balancing decisions.

Finally, the performance of the software was tested in terms of throughput and latency. Overall, the software performs really well with nearly maximum capacities over long times. The throughput values hover just under the gigabit mark. In terms of latency, packets are naturally queued due to the scheduling principle. However, on average and with the planned schedule timing, the delay induced by the WMN system should be small enough for the target levels of LTE-Advanced, for example (few ms for S1 data and sub milliseconds for X2 traffic).

The completed prototype protocol SW included a working and verified routing and scheduling scheme, extensive resiliency and Quality of Service feature sets as well as successfully integrated 80 GHz radio link with electronically steerable beams. In addition, as the resiliency and Quality of Service schemes are highly automated, the protocol and the system fulfill the self-healing and self-optimization features of the SON portfolio. In principle, the demonstrator environment running the newest version of the prototype protocol would be capable of proper mobile transport as such.

Public demonstrations

During the validation process, a few possibilities to showcase the demonstrator system publicly emerged. The first public demonstration was in January 2012 as part of MEVICO project mid-term review. The demonstration platform included the WMN demonstration system and the BRAWE radio system. The demonstrated features include the resiliency and Quality of Service schemes. The demonstration cases were similar to the ones used during the validation process utilizing a high and a low priority video streams traversing through the WMN topology.

The second public demonstration was given in conjunction with the annual Celtic-Plus Event in February 2012 in Stockholm, Sweden. The functional setup built at the exhibition hall is presented in the figure below. The demonstrated features were mainly the same as in MEVICO mid-term review demo, though the Quality of Service scheme was amended with congestion control and load management mechanisms. All in all, both public demonstrations were successful and the WMN concept aroused interest within visitors from telecommunication operators, vendors, standardization bodies and research organizations.

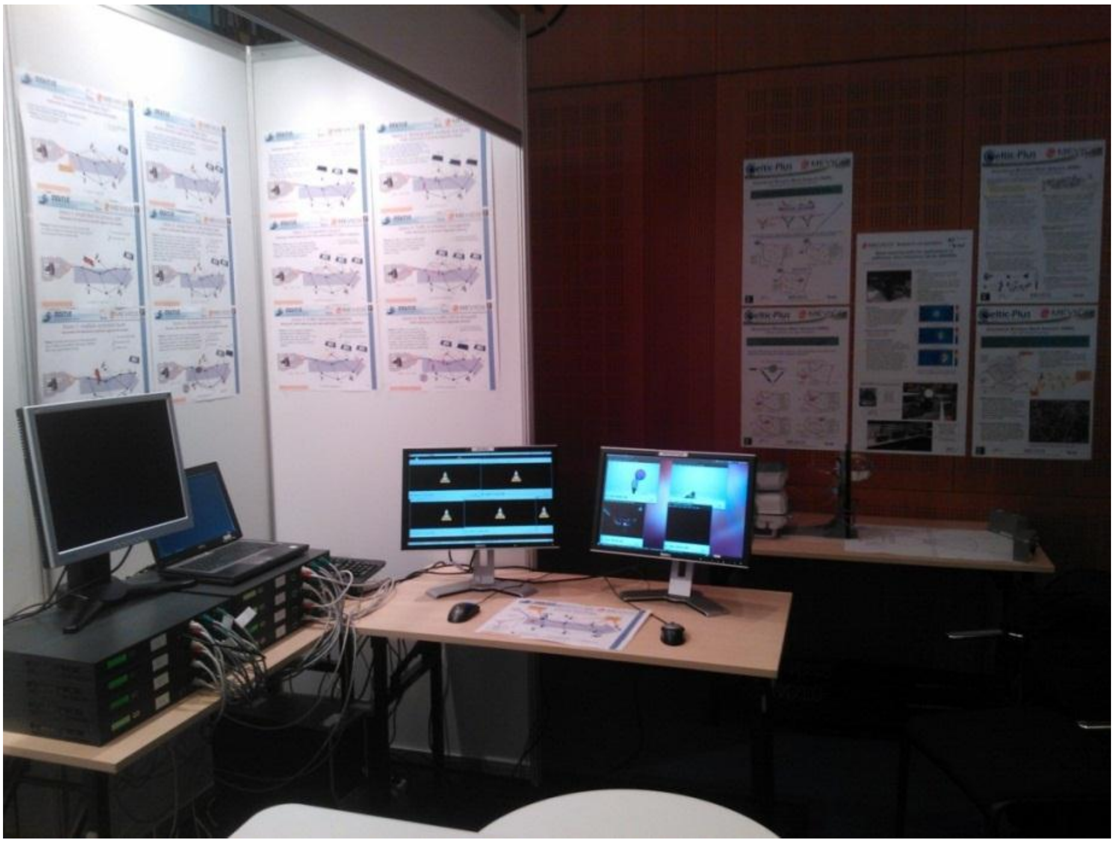


Figure 4. The demonstration setup at the Celtic-Plus Event 2012. BRAWE prototype on right at the background

Key Performance Indicator details and measurements:

The following key performance indicators were assessed during the proof-of-concept validation:

KPI 1.1 Throughput gain in 3GPP access and backhaul

The maximum possible throughput of the validation system (1 Gbps) was verified through one WMN node/eNB with 1000 byte packet length. With smaller packets the performance of the demo software implementation will deteriorate slightly. In traffic congestion situations lower QoS classes are dropped in favor of guaranteeing maximum throughput for high priority traffic classes.

KPI1.2 Backhaul and RAN influence on E-E delay

Only delay measurements with relative values were able to measure due to the restrictions of the validation system. Measured average value 1.5 – 2.5 ms over 3-5 hops translates in theory to around

500 us with shorter scheduling cycle. Note this delay is for access network only. Aggregation and core transport delays needs to be added to traffic which is not handled locally within the WMN access network.

KPI 1.3 Recovery time from link failures or congestions and/or OPEX reduction in the in the backhaul or core transport network layer

Only recovery time with Ethernet cable connections was able to be measured. Ethernet port failure detection time was noted to be the dominating factor. On average 0,45 s failure detection time was measured, the protection switching or rerouting time being negligible in comparison (few us). With radio links, the targeted hitless protection switching and μs level of the total recovery time inside the WMN system is expected to be reached.

KPI 1.4 Efficient load distribution in the backhaul and in the core

Several load balancing scenarios were demonstrated and verified.

KPI 2.2 Capacity aggregation and E2E QoE sustainment

In WMN system the throughput gain under varying network load situations is achieved through balancing and optimizing the available transport resources (transport link and routes) between the client systems by using the developed traffic management (congestion control, traffic balancing and route self-optimization) schemes. All mechanisms are QoS aware meaning that higher priority traffic flows are always favored over lower priority traffic flows. The implications are that basically any real time or delay constrained traffic can be forwarded with quite deterministic delay bounds and sustainable QoS through multi-hop WMN networks.

2.5 Further work

Naturally all necessary product features and technology building blocks required for the final backhaul product have not been implemented in this research Proof-of-Concept system. These include for example high performance mmW TDD radio link system replacing the experimental BRAWE system, full network synchronization, transport security solution and enhanced fault monitoring and reporting to network OAM.

Examples of potential improvement areas spawned during validation phase are further enhanced resiliency scheme in form of e.g. rapid path discovery and enhanced traffic and load management capabilities for the best effort traffic classes to improve network utilization even more. It would be also interesting to test more hierarchical topology solutions to address scalability issues.

As future research work, we propose to further improve the solution with E2E SON features and optimization in coordination with RAN SON capabilities.

2.6 Conclusions

The objective behind the assembly of the proof-of-concept system was to verify the functionality and feasibility of the WMN access backhaul network concept. Based on the validation results it can be concluded that the developed WMN concept is able to fulfill the key requirements regarding bandwidth, latency, QoS, easy management and operability set for future small cell backhaul networks. The different functionalities, including the mesh network algorithms and the wireless millimeter wave link, were proven to work satisfactorily together even with slightly suboptimal hardware.

Furthermore, the validation process proofed that the WMN concept performs satisfactorily under likely traffic, load and network situations, on a system level. With some further development and enhancement, the WMN system concept displays extreme potential for a state-of-the-art access backhaul transport technology and is therefore a viable backhauling solution future LTE and LTE-Advanced small cell deployments.

2.7 List of publications

Tuomas Taipale, Feasibility of wireless mesh for LTE-Advanced small cell access backhaul, Master’s Thesis, Aalto University School of Electrical Engineering, September 2012.

3. L2 Routing and Mobility Based on TRILL

3.1 Introduction

The goal with 2 mobility consist of localizing the handover process in the access network and specifically in Layer 2 i.e. Ethernet. The expected benefit would be to reduce S1 signaling since the addressing updates are performed faster which reduces latency (and possibly reduces packet loss) when transferring the session within the access network. The L2 mobility supports higher number of handovers in small cells scenarios where handover process increases. The technology proposed for implementing L2 mobility is named TRILL (Transparent Interconnect of Lots of Links; http://tools.ietf.org/html/rfc6326).

TRILL leverages IS-IS routing protocol to achieve Ethernet shortest path frame routing with arbitrary topologies. In this research item the goal is to utilize TRILL extended with DHT to deploy mobility in the network edges. The goal is to combine the advantages of bridging and routing and fully distributed mobility mechanism implemented in the Link layer (i.e. Ethernet). In order to increase the available throughput we consider that is necessary to move towards lower layer switching and minimize processing per packet. In case of mobility, every handover and transaction requires several transactions with mobility management servers normally in a centralized location (i.e. HSS). Instead, the proposal is to optimized routing/Traffic Engineering and mobility transactions to happen in the edges for energy efficiency. Therefore, by moving mobility management to lower layer and handle it in the network edge, we avoid waste transmission (unused bits, unnecessary data) and avoid long data paths. The proposed solution as depicted in next figure the edge nodes will handle the mobility related transactions.

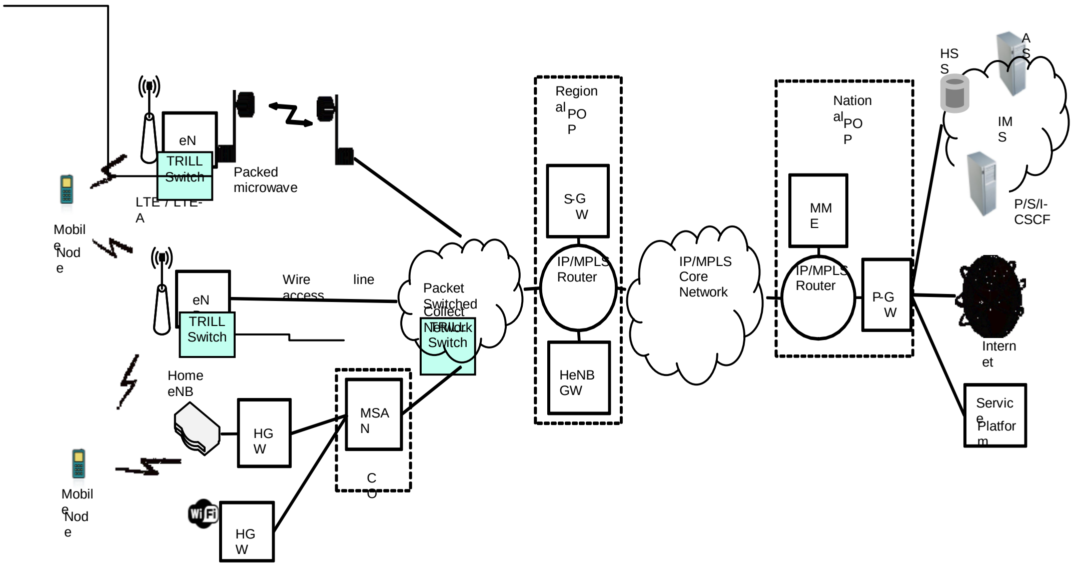


Figure 5.Ethernet mobility with TRILL.

3.2 Detailed description of the proposed solution

The proposed solution will utilize TRILL extended with DHT for efficient sharing of routing updated between the switches serving the eNodeBs where the handover is performed. The current work will deliver a L2 mobility system that will cover the cases where the UE moves inside the same mobility domain, i.e., within the same MME. With the proposed solution we can enable "micro mobility" performed with standard Ethernet switches where only few components e.g. eNodeB switch or Customer Edge Switch need to be enhanced with TRILL+DHT functionality. This allows UEs moving around in the network, connecting to different eNBs without their session being interrupted (i.e., the TCP session).

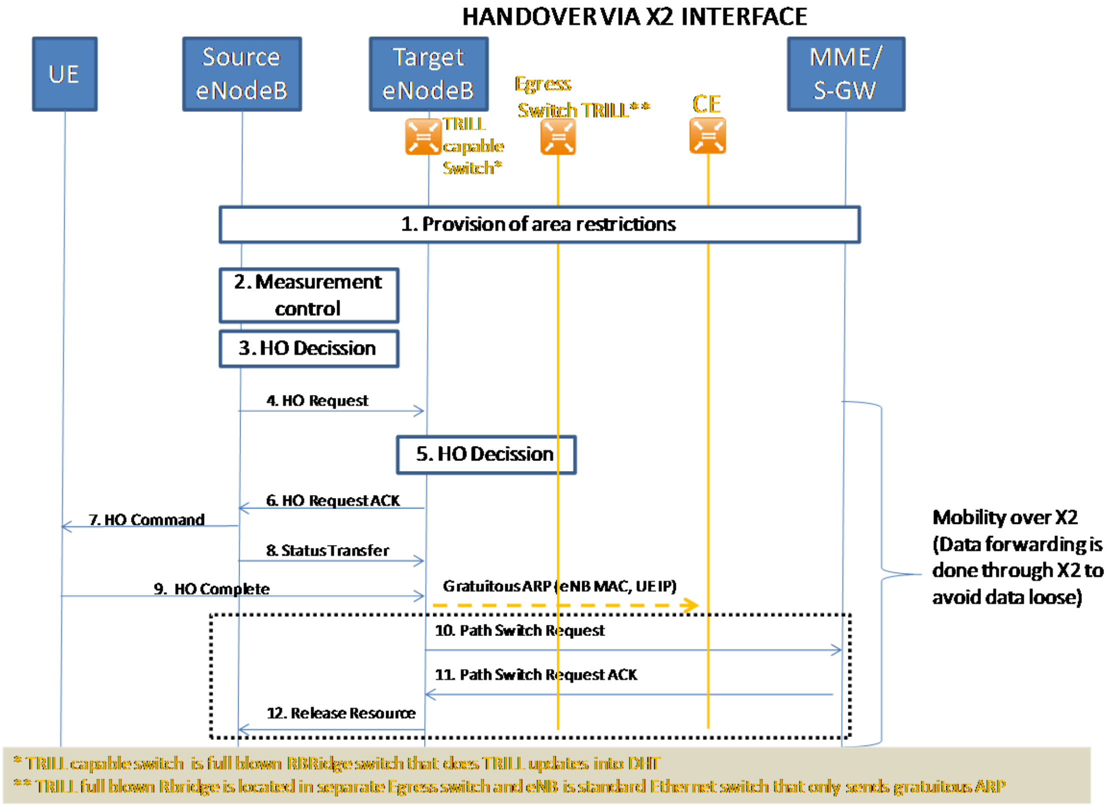

HANDOVER VIA X2 INTERFACE
UE
Source eNodeB
Target eNodeB
Egress Switch TRILL**
CE
MME/ S-GW
TRILL capable Switch*
1. Provision of area restrictions
2. Measurement control
3. HO Decission
4. HO Request
5. HO Decission
6. HO Request ACK
7. HO Command
8. Status Transfer
9. HO Complete
Gratuitous ARP (eNB MAC, UE IP)
10. Path Switch Request
11. Path Switch Request ACK
12. Release Resource
Mobility over X2 (Data forwarding is done through X2 to avoid data loose)
* TRILL capable switch is full blown RBRidge switch that does TRILL updates into DHT
** TRILL full blown Rbridge is located in separate Egress switch and eNB is standard Ethernet switch that only sends gratuitous ARP

Figure 6. Handover through X2 interface.

3.3 Description of validation system development.

UE mobility in 3GPP networks creates signaling load between the core network and the eNBs involved in the handover operation. The signaling complexity and end-to-end path length also affects the UE handover delay.

Our proposal implements intra-domain Ethernet based mobility directly on top of the link layer so that the 3GPP network core does not need to be involved when a UE moves between two eNBs. This reduces the signaling between the 3GPP network core and the eNBs as well as lowers the UE handover delay.

The design works by keeping track of UE-IP, eNB MAC address pairs to identify the location of each UE in a 3GPP network domain. When a UE moves to a different eNB, the eNB MAC address associated with the UE-IP is updated without further changes in anchor, UE-IP or the network core.

We demonstrated the underlying technology and the handover delay with two separate demonstrations. The benefits of the underlying technology will be demonstrated by a network with multiple mobile nodes, where the signaling load for the mobility events are measured. The handover delay is demonstrated using streaming media, where the client receiving the media stream moves around in the network. The quality of the media stream is visually observable in the demonstration during the mobility events. In both cases, we compare our design with a suitable baseline technology with an identical testing environment and test case.

Key Performance Indicator details and measurements:

We measure the Key Performance Indicator (KPI) values based on two separate network topologies:

1. A centralized transport architecture acting as the baseline topology, with handover operations carried over the S1 interface between the Core network and the eNBs.

2. A flat transport architecture, where public Internet is reached from the Access network, and the handover operations are carried over the X2 interface between the eNBs.

Figure 7 shows the intended network topology for the centralized transport architecture. The UEs communicate through theAccess, Aggregation and Core networks to the End point in the public Internet. We introduce two-way delay between the Internet and Core network, the Core network and the Aggregation network, and the Aggregation network and the access network edgefor each Access network segment.For the centralized transport architecture, we base our link delay modeling roughly on the *3GPP (section 6.1.7.2 of TS 23.203 at QCI characteristics)* standard.Concretely, both the End Point – Internet, and the Internet – Core network links have 50ms of mean delay, the Core network – Aggregation network has 10ms of mean delay, and finally, the Aggregation network – Access edge device has a mean delay of 25ms. Each delay is modeled as a normally distributed random variable with a standard deviation of 10% of it's mean.

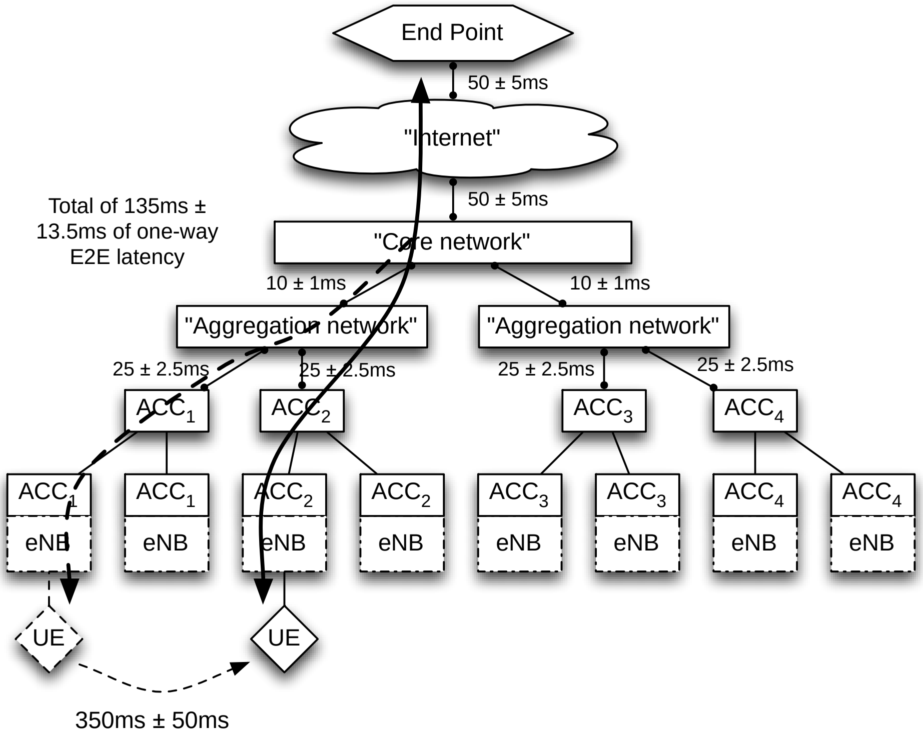


Figure 7Centralized transport architecture topology

Figure 8 presents the network topology for the flat transport architecture. Here, the UE communicates with the End point in the public Internet directly through the Access network, because each edge device in the Access network segment has a connection to a gateway device that directly connects to the public Internet.For the flat transport architecture, we introduce two-way delay only on the gateway – public Internet, and the Internet – End Point links, similar to the centralized transport architecture topology. In addition, because the flat transport architecture is expected to be an option in the future, we have also reduced the public Internet link delay from the 50ms per link in the centralized transport architecture, to 15ms per link in the flat transport architecture. Note that for the flat transport architecture, the Aggregation network and Core network are unused in the KPI measurements for both the baseline and the TRILL/DHT test cases.

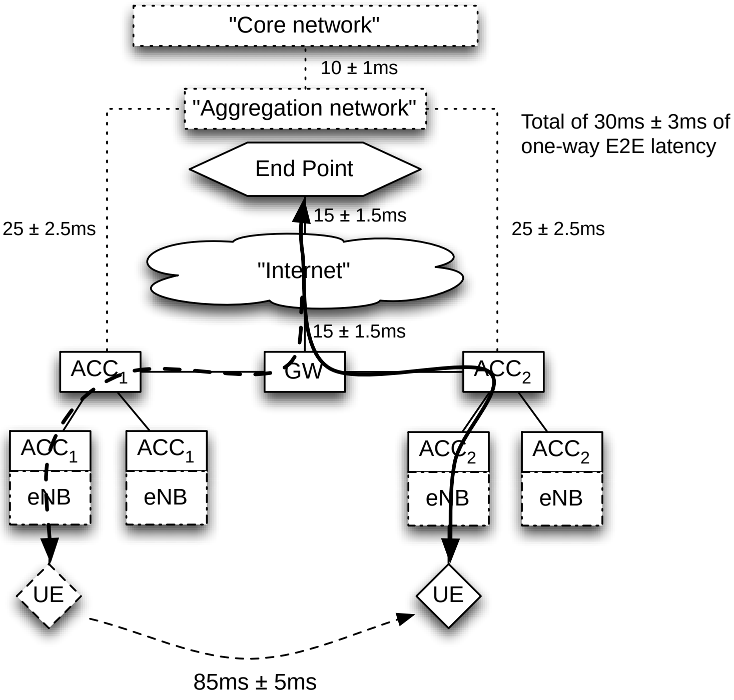


Figure 8 Flat transport architecture topology

Each access network node is emulated using a separate virtual machine, and the Internet, Core network, and Aggregation network are represented by a single virtual machine each on the path that performs the two-way delay on the links. The UE and the End point are modeled as separate virtual machines running various applications to perform the measurements required for the KPIs, and to demonstrate the platform in action. The TRILL/DHT „location store" is placed on the Aggregation network virtual node in the centralized transport architecture, and in the gateway virtual node in the flat transport architecture.

Handover related KPI measurements are modeled by attaching the UE to a different eNB with a delay.Packet buffering during the emulated handover process is performed by the first-hop switch of the old eNB. Once the handover delay has elapsed, a signaling message release the packets on the old eNB to be forwarded to the new location by the switches in the network.

KPI 1.1: Throughput gain in mobile access

The primary focus of our design is to reduce the S1 related signaling in the transport architcture. As such, our DHT extension to RBridges brings minimal benefits to throughput in the access network. However, RBridges have several significant throughput benefits over STP-based switches, that are supported by our extension without any modifications. The expected throughput gain for our overall design is facilitated by two separate things:

1. The Ethernet frame forwarding in access networks is performed by RBridges (TRILL), which uses shortest path forwarding instead of a spanning tree. TRILL has provisions for unicast Equal Cost Multi Pathing (ECMP), as long as the physical topology supports it. Finally, multicast forwarding is based on multiple bidirectional distribution trees, rooted on TRILL devices in the access network.
2. Intra-domain handover process is accelerated by our DHT extension toTRILL, as less eNB <->Core signaling (via S1 protocol) is required. In addition, our design allows efficient eNB to eNB communication over the Access network.

KPI 1.2, 3.2: Backhaul and RAN influence on E-E delay. E-E delay between UEand content.

The end-to-end delay measurements for the KPIsdepend on the delays we choose for the simulated "Core - Internet", "Aggregation - Core", and "STPRoot – Aggregation" links.The baseline without any delays on the links is expected to be in the single digit milliseconds or lower.

In the centralized transport architecture, the end-to-end connection suffer from all three modeled delays on the path, as public Internet access is through the core network. In the flat transport architecture, public Internet access is moved to the access network, so the end-to-end connection suffer from only Internet related delays.

End-to-end delay measurement test case:

The testing was performed by taking the average of 1000 ping RTT results from the network, using the UE as the source of the ping, and the End Point as the destination. For the centralized transport architecture in Figure 7, the packet passes through the access, aggregation, and the core networks before travelling through the Internet to the end point. This results in 135ms of one-way delay, for a round-trip time of approximately 270ms. For the flat transport architecture in Figure 8, packets sent and received by the UE pass through the Access network directly to the public Internet, and to the End Point. This results in 30ms of one-way delay, for a total round-trip time of approximately 60ms.

KPI 1.3: Reliability and failure recovery time(response time to link failures, bootstrap time)

TRILL is a newly standardized link layer protocol, and as such does not provide rapid link failure recovery as part of the standard. TRILL uses a modified IS-IS as the link state protocol, which is used to respond to link and node failures in the network.

The IS-IS specification reacts to link failures upon not receiving a message from the neighbor in three consecutive heartbeat intervals. At minimum, the specification allows the heartbeat transmit interval to be set to one second, thus the minimum response time to non-local link failures is approximately 3 seconds.

Our DHT extension to TRILL responds to link and node failures in conjunction with the link state protocol operation. The information content held in TRILL nodes by our DHT extension is automatically repaired during link state protocol convergence when TRILL nodes fail.

Bootstrap time of TRILL nodes and our DHT extension is likely to be of minor consequence, as forwarding nodes in the access network rarely break. The typical bootstrap phase (ready to receive and transmit user traffic) of an RBridge when powered on is roughly 30 seconds.

TRILL OAM and improved link failure detection features in general arecurrently being standardized in several separate drafts:

1. Requirements for Operations, Administration and Maintenance (OAM) inTRILL*(http://tools.ietf.org/html/draft-ietf-trill-oam-req)*
2. Routing Bridges (RBridges): Operations, Administration, and Maintenance (OAM) Support*(http://tools.ietf.org/html/draft-ietf-trill-rbridge-oam)*
3. TRILL (Transparent Interconnetion of Lots of Links):Bidirectional Forwarding Detection (BFD) Support*(http://tools.ietf.org/html/draft-ietf-trill-rbridge-bfd)*
4. TRILL: RBridge Channel Support *(http://tools.ietf.org/html/draft-ietf-trill-rbridge-channel)*

Our current implementation of the RBridges base specification has no support for the advanced reliability features presented in the drafts above.

KPI 2.3: Handover delay, Service interruption delay due to handover.

Handover delay will be modeled as a delay in the attachment process, when a UE moves between two emulated eNBs. During attachment process, frames destined to the UE are buffered in the old eNB (actually in the first hop switch). Upon attachment completion, buffered frames will be forwarded to the UE through the access network.

In the centralized transport architecture, the handover delay and service interruption delay are affected by the link latency between the eNBs and the Core network when using the standard S1 based signaling mechanism. The expected value is a sum of the cumulative network delay and the normative handover delays specified in the 3GPP standard or other authoritative sources.

The flat transport architecture uses X2 for eNB – eNB handover signaling, and moves the relevant network entities from the Core network to Access network. This bypasses the network latency caused by the aggregation and the core networks. The expected value for the handover delay in a flat transport architecture is a sum of the normative X2 interface delay during handover, and the minimal delay caused by the network signaling of our DHT extension.

By implementing TRILL and our DHT extension in either transport architecture, we bypass the S1 signaling completely, and keep the mobility signaling in the access network. Expected value for the handover delay in this case is a sum of the normative handover delay and a single digit milliseconds delay caused by our extension updating the DHT location information for the UE, and propagating it to the necessary entities in the access network.

KPI 2.5 Handover related signaling load on the network

Our solution creates signaling load in the network whenever a UE moves between two eNBs. Attachment to the new eNB generates a signaling message (e.g., gratuitous ARP message), informing the TRILL/DHT node where the new eNB is connected, that an end-host has arrived behind the node.

The signaling message is intercepted, and location (and layer 3 addressing) information is updated in the access network via one (or two) DHT signaling primitive(s) messages used by our DHT extension. The destination of the update information is the TRILL/DHT server responsible for storing the information. If the server notices a change in the information, (e.g., the location has changed), the updated information is propagated in another DHT signaling primitive message to the TRILL/DHT node that the old eNB is connected to.

After receiving the updated information, the old eNB will forward all frames received for the UE to the new eNB, and in turn update the information on the TRILL/DHT node that originated the frame with the incorrect location information.

In summary, a handover by a UE causes at minimum two unicast signaling messages in the access network, typically at least three. Additional signaling is also required to reactively propagate the information to network nodes that are communicating with the UE.

Handover delay measurement test case:

The handover delay measurements were performed by moving the UE in the network between the two eNBs (a mobility event) in Figure 7 and Figure 8, while a stream of ICMP echo packets (ping) were sent from the End Point in the Internet to the UE and back. The interval of the mobility events was set to 5 seconds, and 100 mobility events were recorded for each test. The performed handover was seamless, i.e., no lost packets were observed during the mobility event, however packet reordering was observed.

The centralized transport architecture modeled the simulated handover delay as 350ms, with a standard deviation of 50ms. In addition, the baseline test case added a single RTT of delay (70ms with a standard deviation of 7ms) to the handover delay to model the signaling delay related to the operation of the S1 protocol between the eNBs and the Core network. For our TRILL/DHT extensions, we added an additional delay of 50ms with a standard deviation of 5ms to model link latencies between the two separate Access networks where the eNBs are located and the Aggregation network.

The reported average handover delay of the mobility events was calculated by collecting the round-trip times of the first buffered ICMP echo packet, with the average end-to-end delay removed from each round-trip time. During the testing, the Internet End Point was sending ICMP echo packets with an interval of 275ms to the UE in the Access network.

The flat transport architecture tests model the X2 interface to signal handovers directly between the eNBs, bypassing the Aggregation and Core network links completely. Thus, in both the baseline, and our TRILL/DHT extensions case, the handover was modeled by a 85ms delay, with a standard deviation of 5ms. No additional delays were added to the handover. The reported average handover delay of the mobility events was calculated as with the centralized transport architecture, however during the testing, ICMP echo packets were sent with an interval of 80ms from the End Point in the Internet to the UE in the Access Network.

3.4 Future work

The usage of TRILL+DHT to handle mobility within the same Location Routing Area improves the handover delay and the signaling overhead since mobility between adjacent eNodeBs is handled at Ethernet level.

Future work consists of implementing similar mobility functionality in SW Defined Networks (SDN). Therefore, adding the mobility into SDN controller facilitate the deployment of cost effective mobile infrastructure that can cope with the expected traffic demands.

3.5 Conclusions

The validation results shows following main results.

E2E delay influence of RAN acces:

Centralized transport architecture: Baseline: 274ms, TRILL/DHT 274ms

Flat transport architecture: Baseline: 63.3ms, TRILL/DHT 63.2ms"

Handover delay:

Centralized transport architecture: Baseline: 393ms, TRILL/DHT: 370ms

Flat transport architecture: Baseline: 98ms, TRILL/DHT: 101ms

Handover signaling:

a) 1 signalling message from new eNB to TRILL/DHT location store for the UE

b) 1 signalling message from TRILL/DHT location store for the UE to the old eNB

c) [0..N] signalling messages from old eNB to any TRILL/DHT node sending traffic to the UE through the old eNB (stale location information on the TRILL/DHT node)"

4. Customer Edge Security

4.1 Introduction

One of the most critical problems of the current Internet is that the amount of available IPv4 addresses is no longer able to meet the demand. The requirement of new IP addresses is particularly high in mobile networks. Address reuse is hindered as a higher portion of the handsets are connected to the Internet almost constantly. It is becoming apparent that all operators will not be able to provide public addresses to all their customers. The shortage of public addresses leads to the deployment of network address translators (NATs), since the deployment of longer-term solutions, such as IPv6 and HIP, is difficult and slow because of the changes needed in the user equipment and applications. Consequently, it is expected that mobile users will be connected to the Internet via a NAT.

In order to run a publicly accessible server, a host must be able to accept inbound connections. Several types of applications, including conversational applications (instant messaging, voice over IP, etc), collaborative, distributed and peer-to-peer applications require inbound connections. The presence of a NAT or firewall makes accepting inbound connections difficult, especially if the NAT is operated by the service provider and the user has no ability to create static mappings. Application developers have solved the problem by developing a set of methods to traverse NATs. NAT traversal, however, consumes considerable amount of energy on cellular devices, as each application separately need to keep their NAT mapping valid by sending traffic, which prevents the device from going into longer periods of sleep mode. NAT traversal weakens security because of the uncontrolled way NATs and Firewalls are bypassed. Furthermore, NAT traversal requires that network layer functionality is implemented in the applications, making applications more complex and heavy.

There is also a growing need to protect the operator's and user's networks from attacks and unwanted traffic (port scanning, spam). Firewalls improve security by separating the user network from the public Internet. Currently, many operators have disabled inbound traffic completely in cellular networks for security reasons. When the customer is charged based on traffic, it is difficult to motivate why the customer should pay for unwanted traffic. However, NAT traversal mechanisms also can enable inbound traffic in that case. As mobile connections are replacing fixed connections, there is no reason for the mobile network to be more restrictive than fixed networks. Moreover, phones are a natural platform for a large group of conversational applications, which lay in the background waiting for an inbound call or message. We seek a way to enable inbound traffic for applications that require inbound traffic based on policies. Instead of applications utilizing insecure NAT traversal methods that consume energy, the network must provide controlled ways for applications to define the type of traffic expected.

Customer Edge Switching (CES) aims to increase the scalability and the security of the network by separating the user network from the public network. The user and public networks have separate addressing, routing and transport. The CES device operates at the border of trust between the networks and forwards traffic based on policies. It replaces the current NATs and replaces/complements the current firewalls. Contrary to NATs, applications do not need to be aware about the CES. Instead, CES provides an interface that looks like a global IP network to the host. Thus, the host can accept inbound connections as in the case it had a global IP address, provided that these connections are accepted by the user's policy. The host continues using IPv4 or IPv6, as supported by the private network. The applications need no modifications.

When both endpoints are in CES enabled networks, the concept provides additional features, including multihoming, mobility, connection monitoring, return routability checking and advanced security methods. This is realized using the Customer Edge Traversal Protocol (CETP), which is used for signaling between peering CES devices. As a further result of the separation, multiple types of public networks based on different technology can coexist. Thus, new technologies such as IPv6, routed End-to-end Ethernet and completely new (e.g. clean-slate) routing paradigms can be introduced in gradually the public network without modifications to hosts. The used transport technology is selected based on the support of the communicating CES devices and the existence of an end-to-end transport connection based on the given technology.

In the case where only one of the endpoints is in a CES enabled network, the CES falls back to a NAT-like operation mode for outbound connections. To handle inbound connections in this situation, a novel server-side NAT concept has been developed. The concept is based on a circular pool of proxy addresses. A server in a CES enabled network can serve an unlimited number of clients in the public IP network. The number of concurrent connections and servers that can share a single (or a small block of) public IP addresses is practically unlimited, whereas the limitation rather comes from the rate of incoming connections per second.

CES separates between the name and the address. The Fully Qualified Domain Name (FQDN) is considered as a global user identifier. The application uses the FQDN to indicate the destination, which can be a service or a host. The address is dependent on the specific network technology. In addition to the FQDN, users and applications can identify themselves using other types of identifiers. This identifier (ID) can be of various types, including session specific identifiers, hash of domain name or an operator assured identifier.

4.2 Detailed description

The following subsections focus on the main development areas that have been performed in MEVICO.

4.2.1 Customer Edge Switching in the EPC

While the fundamental CES concept already existed before MEVICO, the concept has been further enhanced and validated within the MEVICO project. Especially the aim has been to apply the CES concept to the cellular network scenario.

In contrast to the fixed environment, where the CES is maintained by the customer, the cellular scenario requires the operator to maintain the CES on the user’s behalf. The policies thus reflect both the operator’s and the customer’s views. The user may be provided with an interface to configure policies in a simplified way. The user may for example allow or disallow traffic from certain IDs. On the other hand, the operator may require that the source address of the communication is legitimate. Input from a reputation system (e.g. calculated based on reported attacks and suspicious traffic) may affect the policies. For example, a source with a bad history or an unidentified source may have limited access.

CES separates between the control and data planes. The CES data planes are located at any element connecting the operator network to the public Internet, i.e. in the PGW and in local breakouts. There are no restrictions in the location of the control plane. This allows for architectures where the control and data planes are collocated or where a single control plane controls several data planes. However, planning must consider that the first packets of a flow must be processed by the control plane in order to decide whether traffic is admitted and to create forwarding state. In a centralized architecture, it is reasonable to integrate the CES completely as a part of the PGW.

When the CES communicates with another CES, it uses the CETP protocol. For traffic between CES enabled operators, a dedicated inter-operator network between PGWs can be applied. The user traffic, being tunneled in CETP, can be carried over any transport, including IPv4 or IPv6. In case of a dedicated inter-operator network, also future schemes such as routed Ethernet (TRILL) can be used. Transport alternatives also include encrypted connections, e.g. Transport Layer Security (TLS). The choice of transport is independent of the user transport (e.g. IPv4).

CES does not affect the transport between the UE and the PGW. Neither does it require changes to user devices, user applications, and other network elements. However, CES affects addressing, as only private addresses are allocated to users. Each user can be provided a full private IP network that is completely separated from all other users and from the public network. Thus, CES provides a full available IP address space on each GTP tunnel, which improves both scalability and security.

CES uses a Diameter interface in order to obtain identify information and policy rules from the HSS and PCRF, respectively. The Diameter application for CES is left for future development.

4.2.2 Customer Edge Traversal Protocol and Policy Control

The Customer Edge Traversal Protocol (CETP) (earlier named T2P) is a protocol operating between CES devices. The protocol has two main objectives:

1. To tunnel data packets with minimal overhead while transporting the IDs of the source and the destination.
2. To improve security and reliability by allowing CES devices to exchange control information and agree on policies.

Consequently, the CETP protocol has a data plane and a control plane. The control plane transports control information elements between two CES devices, while the data plane transports a tunneled packet. Each protocol message has a header that identifies the IDs of the communicating endpoints. All information is represented as data elements in a TLV (type, length, value) format to allow flexibility and future expandability. The protocol provides mechanisms for implementing extensions and signaling processing of unrecognized information elements.

The CEPT protocol implements the following security methods:

- Return routability checks. The purpose of return routability checking is to detect packets with spoofed source addresses and prevent them from being forwarded into the network.
- Policy control. A trust domain can define a policy about what information from the peer is required before establishing communication. The policy also defines what information is disclosed to the peer. Packet admission can be modified based on the trust level of a given identity or based on information about detected or suspected attack attempts.
- Attack reporting. When an attack is detected, the trust domain can report the problem to the concerned party. For example, in a reflector attack, the reflector can be informed about a spoofed source address and tighten its policy.
- Signatures. A signature can be calculated over the control information to prevent identity theft and man-in-the-middle modification of information.
- Identity Certification Authority. The protocol provides the address of a certification authority (such as a HSS), with which the correctness of the identity can be checked.
- Selection of ID types and revocation of IDs. A CES can require a certain type of ID to be used. A given ID can be also revoked, e.g. if identity theft has been detected or if it has been used a long time.
- Controlled removal of expired state information. Removal of expired state information is signaled and synchronized between CES devices.
- Postponing state creation. Since CES maintains state information about the active sessions, a possible attack is to drain the resources of the CES by creating bogus state. To avoid this, the protocol can postpone state creation until the identity of the source is confirmed.

In order to facilitate for multihoming, reliability and the use of parallel technologies in the public network, the protocol provides a set of public addresses called Routing Locators (RLOCs). The public addresses can be based on different technologies, including IPv4, IPv6, Ethernet and, in the future, other routing architectures. The addresses are ordered according to preference and order values, which allow ordering different technologies and CES devices according to priority and to balance the load between CES devices.

CETP can be transported on top of various protocol layers, including IPv4, IPv6, Ethernet, UDP and SCTP.

A CETP message consists of a mandatory protocol header, an optional control plane section and an optional data plane section.

The protocol header specifies the protocol version, indicates whether a control plane section is included, and provides the header length and payload length. The protocol header also specifies the source ID and the destination ID. Each ID is defined using the type, the length and the actual ID value. The ID can be of various types, including random IDs generated locally by the CES, locally certified IDs, mobile operator assured IDs and user certificates obtained from a mobile operator certification authority.

The control plane section is a list of TLV (type, length, value) elements. The type field is divided into subfields: a 2-bit group, a 7-bit code and a 2-bit operation. The group defines the high-level type of TLV element while the code defines the detailed type of TLV element within the group. The operation bits (named Q and R) specify one of four possible operations: query, response, reliable response and acknowledgement. Each TLV can be thought as a message within the message. Furthermore, the type field contains compatibility bits indicating how an unrecognized TLV type should be handled and extension bits reserved for future use. The length field can be 7 or 15 bits long, depending on the value of the first bit. The aim is to minimize the message length while accommodating for large values such as certificates. The definition of value depends on the TLV type.

The data plane contains a tunneled data packet. Currently two types of encapsulation is defined: a IPv4 packet with a compressed header and a full Ethernet frame allowing for tunneling of raw IPv4, IPv6 and other packets. Normally, an IPv4 packet is transported using the compressed header, but if options are present or if fragmentation is used, the IP packet is transported as a full Ethernet frame. Future encapsulation will be defined for IPv6.

The TLV type number space covers IDs, encapsulated payload, reachability information, and control information. The fact that the number space is shared between the control and data planes allows referring to certain information between the planes, e.g. for requesting given types of information.

The reachability of a destination is given by Routing Locators (RLOCs). A RLOC is an address on the public network of a given type. Each RLOC is routed to an interface on a CES device. RLOC types include IPv4 addresses, IPv6 addresses, and MAC addresses. For each RLOC, a preference and order is given, which allows ordering RLOCs according to priority and to split the load between several CES devices, in a way similar to DNS NAPTR records. The following TLV elements are specified for reachability information:

- IPv4 reachability information: Provides public IPv4 addresses and the order and preference for each address.
- IPv6 reachability information: Provides public IPv6 addresses and the order and preference for each address.
- Ethernet reachability information: Provides public MAC addresses and the order and preference for each address.

The reachability information provided by CETP matches the reachability information stored in DNS in NAPTR records. While the source obtains the set of RLOCs for the destination from DNS, the destination needs to use CETP to obtain alternative RLOCs from the source. The reachability information can also be used to modify and update reachability information during ongoing session, e.g. to distribute load or signal unavailable RLOCs.

The following TLV control elements have currently been specified (in the reachability and control groups):

- Timeout of state information: Gives the timeout value for state information.
- Cookie: Transports state information in the message instead of creating connection state information in the inbound CES.
- Address of certification authority: Provides an address used for assurance queries.
- Fully qualified domain name (FQDN): The FQDN of the communicating endpoint. This is matched with the information in DNS in a return routability check.
- Header signature: The signature calculated over the control TLVs.
- Unexpected message report: Allows the inbound CES to report about messages that are not related to ongoing connections in order to stop reflector attacks.
- Backoff: Reports error conditions that require the connection to be aborted.

4.2.3 Interworking Between CES Enabled and Legacy Networks

The CES concept has primarily been designed for scenarios where both communicating endpoints are behind a CES device. This scenario enables all features of the CES concept, including the use of multiple technologies in the public network, complete separation of the customer and operator networks, using identities to identify the users and using the CETP protocol to provide enhanced security. However, it is expected that in several cases one of the endpoints will be in a network that is not CES enabled. This is especially true while CES is being deployed. Not even in the most positive outcome, the CES technology will cover all existing networks. Consequently, interworking mechanisms are needed for scenarios where one of the endpoints is behind a CES and the other uses legacy IP network without the support of a CES device.

With the terms interworking scenario and CES-Legacy scenario, we refers to the situation where one of the endpoint is in a CES enabled work but the other is not. The CES-Legacy scenario is divided into two sub-scenarios depending of the direction of the traffic:

1. CES-to-legacy: The initiator of the data connection is behind a CES but the destination is not. The connection is called an outbound connection.
2. Legacy-to-CES: The destination of the data connection is behind a CES but the initiator of the connection is not. The connection is called an inbound connection.

Interworking is enabled by integrating a new module called Private Realm Gateway (PRGW) into the CES. The PRGW performs address translation operations between a private and a public realm for outbound and inbound connections. In the case of outbound connections, the operation is similar to a NAT. The PRGW acts as a DNS proxy and a default gateway for private hosts accessing the public network. For inbound connections, it incorporates a novel algorithm called the Circular Pool of Public Addresses, or Circular Pool for short.

The PRGW allocates public addresses that represent the private host toward the public host. For that purpose, the PRGW maintains a pool of public addresses. The PRGW requires that the CES contains a DNS server that operates as an authoritative name-server for the given zone of authority.

The PRGW stores state information for each CES-Legacy connection in the Connection Table. The state entry contains the local IP and port (A:iPA), the outbound IP and port (R1:oPA), the remote IP and port (E1:oE1), and the protocol. Each entry has a timer used to delete expired entries, i.e. entries for which no traffic has been forwarded within a given time. For inbound connections, another table of temporary state is used before the connection state has been created. This so called Waiting State maps a reserved public address to a private host. Entries in this table also have a short timeout (a few seconds), corresponding to a maximum roundtrip delay.

For a new outbound connection, the PRGW creates connection state similarly to NATs. Since the public address pool contains several addresses, a public address is allocated from the pool either randomly or according to a predefined scheme. The ports used by the original connection are preserved if possible, otherwise port translation is performed.

For an incoming connection, the Circular Pool algorithm is used. The algorithm works by pairing the DNS query with the traffic of the connection. Pairing is needed because the DNS query contains the destination FQDN, but not the sender's address (because of recursive queries or iterative queries by a separate resolver). The data packets, on the other hand, contain the sender's address. In order to create state, both are needed. During the pairing, the temporary Waiting State is used. There can be one Waiting State for each address in the public pool.

The following procedure is used for the pairing. When a DNS query is received from a public host, Waiting State is created and a new address is allocated from the public pool. The allocated address is sent to the querying host in the DNS reply. The created Waiting State contains the mapping from the allocated address to the private host's FQDN as specified in the DNS query. When the first packet is received to the allocated address (which does not match with any active connection), the address of the public host is known and the final connection state can be created. The Waiting State is removed and the public address becomes available for other new inbound connections. If no data packet is received within a timeout, the Waiting Sate expires, freeing the public address for another inbound connection.

Once the connection has been established, further packets will match with the connection state based on the remote IP address and the public address. Consequently, there is no limitation on how many public and private hosts can share a public address. The only limitation is the number of connections that can be simultaneously in the Waiting State, which is one per public address.

4.2.4 Protocol Compatibility Evaluation of CES

Customer Edge Switching changes the philosophy from end-to-end communication to trust-to-trust communication. When CES is used, an application cannot directly address a destination by its IP address. This is because all destinations reside in private networks. Instead, the sender specifies the destination using a Fully Qualified Domain Name (FQDN). In the CES concept, the FQDN is considered as a global name in contrast to the IP address which is used for routing in each network separately. Thus, the role of IP addresses changes from having a global scope into having a local scope. IP addresses, both IPv4 and IPv6, can be used inside customer networks and provider networks, but addressing and routing is operating only within these domains. IP addresses cannot be used across domains.

As a matter of fact, end-to-end communication has not in many years been a reality. Various middleboxes, such as firewalls and NATs, have been existing a long time and these break the end-to-end principle. Several protocols are broken by middleboxes. Today's solution to the problem is to use NAT traversal, i.e. to adapt the application to the network and include network layer functionality into the application. Also CES breaks the end-to-end principle similarly to NATs. However, CES enables a large group of protocols to operate without any NAT traversal mechanisms. Thus, most applications will work in a private network in the same way as if the network was public.

The compatibility of applications is an important topic, since one of the primary requirements is that the host and applications must not be modified. To verify that most client-server applications work with CES, we performed a protocol compatibility evaluation. The aim is to identify which protocols are not natively working with CES and to find the reasons why a protocol is not working. This helped to form rules what is required for a protocol to work with CES. The result of the protocol tests are presented in later sections while the rules are summarized in the following.

When CES is used, applications must identify destinations using FQDNs and perform a DNS lookup to the FQDN before sending packets. The IP address returned by the DNS query is a locally valid proxy address generated by the CES. The proxy address directs the traffic to the CES device, which forwards the traffic to the correct destination.

This gives a set of rules for protocols and applications in order to be compatible with CES.

1. A protocol must address the endpoint using a FQDN. The protocol cannot transport IP addresses between domains.
2. An application must perform a DNS lookup on the FQDN before sending traffic to a given destination.

As a consequence, applications are not allowed to signal IP addresses to their peers, which use the address for sending traffic. The address sent by the source is not a valid address in the destination’s network. Instead, an application can send the FQDN to a peer, which performs a DNS lookup on the FQDN to obtain an IP address for the communication.

A few applications do not fulfill these requirements. An application may start the communication by sending traffic to an IP address directly without performing a DNS query first. In practice, this scenario is rather uncommon, since users mostly use domain names to specify destinations. The requirement to perform a DNS query is specific to the CES concept.

Some peer-to-peer software may locally memorize the IP addresses of the peers between sessions and reuse these in later sessions. A more serious problem is the inability to use IP addresses globally. Some applications use in-band signaling to send their IP address to a peer device and expect the peer to send traffic to this address. This is typically used in applications where the control connection is separated from the data connection, e.g. in FTP and SIP. This problem is common both to ordinary NATs and CES.

4.2.5 Application Layer Gateways

Protocols that are not natively working with CES are handled by an Application Layer Gateway (ALG), which modifies protocol messages on the application layer. We have chosen the approach of adapting the CES to the application rather than adapting the application to the CES. This is to avoid the multitude of disadvantages caused by todays NAT traversal mechanisms utilizing the latter approach. Once the incompatible protocols are identified, the objective is to develop solutions for enabling communications with applications and protocols that are not as such compatible with the CES environment. Our intention is to show that these protocols can still be used in a CES enabled network. As part of the validation, ALGs were implemented for two protocols: SIP and FTP.

SIP is fundamentally a client-server protocol but since messages can be forwarded between multiple servers, the protocol is more like peer-to-peer from a practical viewpoint. The media connection uses a route different from the control connection. The SIP ALG needs to modify the IP addresses and port numbers in SIP messages. The SIP ALG also, in some situations, must create new connection mappings for media and media control connections according to the information in the SDP body.

FTP is a client-server protocol. The requirement from ALGs is due to the separate data connections the IP addresses of which are signaled in the control connections. Private IP addresses as seen by one host is not valid in the network of the other host. The problem for CES is the same as for NATs in general, therefore FTP ALGs are well-known. A FTP ALG modifies the IP addresses and port numbers conveyed in control messages and creates new mappings for the data connections. Since FTP uses messages in text format, the IP address translation causes changes in the packet size. Therefore, the ALG must adjust the acknowledgement numbers and sequence numbers in TCP headers correspondingly. The ALG must maintain state information about the cumulative difference in packet size that is applied to the acknowledgement and sequence numbers.

4.3 Validation system

As a new concept, the CES needs to be comprehensively tested. In order to test and validate the CES approach a prototype has been implemented. The prototype consists of a data plane and a control plane. The data plane forwards packets without the involvement of the control plane. We have developed two data planes: a fast data plane in C and a more comprehensive data plane in Python. The former uses libpcap for packet capturing and the latter uses Scapy. The data plane and control plane communicates through a socket based proprietary protocol, which later could be replaced with extended OpenFlow or ForCES protocols. The control plane is involved in creation of forwarding state upon the first packets, serving DNS queries and managing security. The control plane contains a policy management module, a host register, and a connection state table. The CES integrates a DNS proxy, a DNS server and an ARP server. The prototype operates with an external DHCP server. The current prototype simulates the

HSS and the Diameter protocol with SQL databases and the SQL protocol. The development of a Diameter application for CES is left for future work. The control plane is developed in Python, which provides a good platform for testing various features. For the validation, the speed and flexibility in testing various solutions are more important than the processing speed, since once state is created packets are forwarded with the separate data plane. The control plane uses the Scapy library for packet processing. The prototype runs in a Linux environment, although it could easily be adapted to other platforms.

In order to validate the CETP protocol operation, the CETP protocol is implemented and integrated into the CES prototype. This adds new modules to the prototype: a CETP protocol parser/generator, a finite state machine (FSM) and a Policy Engine. The protocol is defined using the Domain Specific Language (DSL) in Scapy, which allows a protocol to be defined in a structured way by defining protocol fields and the relationship between them. Validation includes both the control and data plane operations of CETP. The existing CES prototype has earlier tunneled packets with ordinary IPv4-in-IPv4, IPv4-in-IPv6 and IP-over-Ethernet tunnels. As part of validation, CETP tunneling is implemented on top of IPv4, IPv6 and routed Ethernet (e.g. TRILL). The control plane validation includes testing the basic mechanisms of TLV element transfers based on queries and asynchronous responses. The mechanisms for requesting information as specified by policies are tested. Moreover, security mechanisms including postponed state creation and return routability checks are tested. Finally, the control plane testing includes implementing various TLVs and the underlying control mechanisms.

The PRGW has been implemented as part of our CES prototype. The testing scenario comprises a private host, a PRGW and a public host. For the purpose of validating the operation of the Circular Pool, a test program was developed for simulating the public host. The program generates connection requests according to a specified rate and a distribution. A connection consists of a DNS request and transported data. The test program records statistics about the success rate, the number of required DNS attempts and the delay. The DNS behavior can simulate different operating systems, which have their specific number of retries and delays between retries. In presented results, the maximum number of retransmission is set to 4, which is common in today’s operating system. The delay and packet loss parameters (and their distributions) of the network between the user and the CES are controlled by a network emulator. Using the statistics, the efficiency of the Circular Pool algorithm has been evaluated and the impact of the different parameters has been analyzed in a controlled manner.

In order to show that protocols that are not natively working with CES can be enabled in the CES environment, we implemented ALGs for selected protocols: SIP and FTP.

For protocol compatibility testing, applications were installed and tested on Windows and Linux (Ubuntu) platforms, depending on the platform supported by the applications.

Protocol compatibility with CES is tested in two scenarios:

1. CES-CES: Both endpoints are located in CES enabled networks.
2. CES-legacy: One endpoint (the source or the destination) is located in a CES enabled network and the other endpoint in a network that is not CES enabled.

Some of the tested protocols are proprietary and there is no publicly available server that can be installed in a private network. These require using the server maintained by the provider of the application, whereas only the CES-Legacy scenario could be tested. For applications allowing installation of a server on a private network, we tested also the CES-CES scenario.

4.4 Main results

4.4.1 Customer Edge Traversal Protocol

The test results show that it is possible to exchange packets in CETP encapsulation between two CES devices and to use IDs for identifying the hosts behind CESs. CETP has been transported over IPv4 and directly on top of Ethernet. Even though the CES normally participates in the MTU (Maximum Transmission Unit) detection method of TCP and thereby avoids fragmentation, some cases require fragmentation (UDP without the Don’t Fragment bit). CETP is able to fragment packets that are too large for the following link, whereas CETP falls back to the uncompressed generic encapsulation. CETP has been tested with multiple parallel RLOCs using different technology and with automatic selection of RLOC based on preference.

Testing with the control plane shows that the basic functions of TLV encoding work. Using CETP the inbound CES is able to query the FQDN from the outbound CES and use the information to reply to reverse DNS queries (PTR queries). This can be done proactively or as triggered by a reverse query. It has been verified that CES can use the Cookie mechanism to match the previous and the new connection in the case that an ID changes. ID changes have been tested with a frequency as high as one change per each message.

4.4.2 Private Realm Gateway

The prototype implementation validates that the concept works as expected.

The figure below shows the success ratio for a circular pool of 5 public addresses for different round-trip delays when the offered load varies from 10 to 70 new connection requests per second. We can see that the number of connections per second that reliably (near 100%) can be served depends on the delay. The round trip delay affects how long time a public address is in the Waiting State. For dimensioning purpose, it is feasible to use a high value for the delay. With 5 public addresses we can reliably serve 20 new connections per second. In terms of pool size, the Circular Pool scales linearly, e.g. 10 public addresses is sufficient to serve 40 new connections per second. We remind the reader that once the connection is established, the number of concurrent connections is not limited.

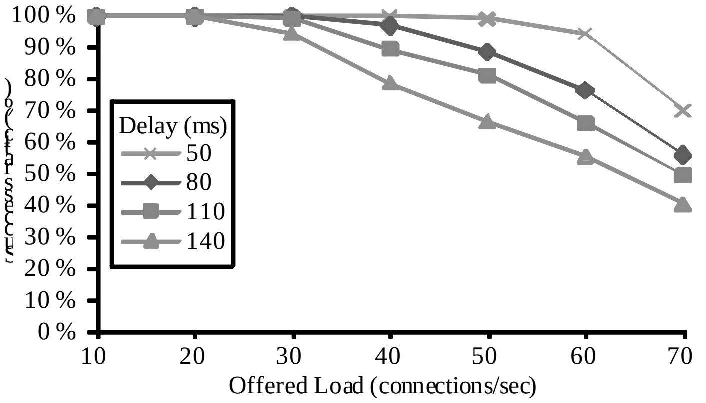


Figure 9. CES success ratio.

The scalability can also be analyzed theoretically with the Erlang-B formula. The analysis allows determining the number of users that can be served by a given pool size. The figure below shows the results for a blocking rate of 0.1% of the offered connection. According to the figure below, about 7.5 millions of users can be served with a C-class block of addresses. As Erlang-B does not consider retransmissions of DNS requests, the results indicate the lower bound for performance.

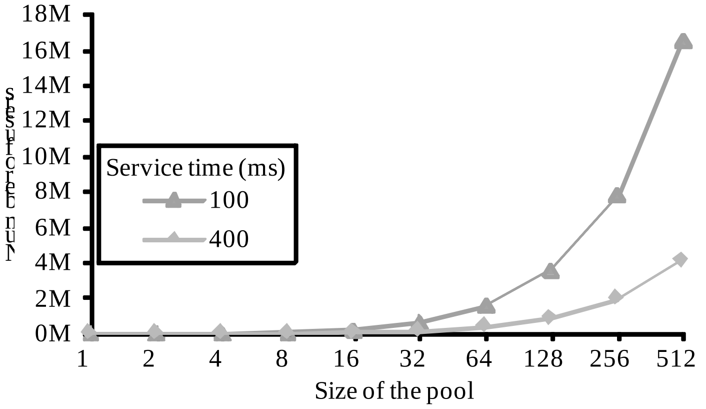


Figure 10. CES supported users.

Detailed performance and scalability results are available in the publication and Master’s thesis indicated below. With the implemented prototype we tested the PRGW with various protocols to ensure that the concept works correctly. The results are presented later in this document.

4.4.3 Protocol Compatibility

In the selection of protocols for compatibility testing we choose two sets of protocols. The first set consists of common web protocols forming the majority of the traffic. Web traffic uses client-server communication, where a client sends a request and the server replies with the requested information. This type of protocols is very simple and is expected to work with CES without problems. The second set of protocols establishes connections directly between users. This type of protocols is used for messaging, voice/video calls and file transfer between users. From a mobile perspective, these protocols can be considered as the most interesting ones as phones is principally used for inter-person communications. Because of the required inbound connectivity and the separate connections used for media, these are expected to be challenging for CES.

A given protocol is typically implemented in several applications. We selected the most common application(s) for each protocol. Some protocols are tightly linked to an official application. However, for these protocols also third-party applications are available.

The tested applications are the following:

Tested protocol	Tested applications	Tested scenarios
HTTP	Servers: Apache2, Tomcat Clients: Mozilla Firefox, Windows Internet Explorer, Chrome	CES-CES, CES-Legacy
HTTPS	Servers: Apache2, Tomcat Clients: Mozilla Firefox, Windows Internet Explorer, Chrome	CES-CES, CES-Legacy
SSH	Server: sshd Client: ssh	CES-CES, CES-Legacy
Internet Control Message Protocol (ICMP)	Ping, Traceroute	CES-CES, CES-Legacy
Session Initiation Protocol (SIP)	Servers: Kamailio, 3CX Clients: Ekiga, Twinkle, 3CX	CES-CES, CES-Legacy
File Transfer Protocol (FTP)	Server: vsftpd Client: ftp	CES-CES, CES-Legacy
Extensible Messaging and Presence Protocol (XMPP)	Google Talk, Empathy, Psi, Pidgin, Tkabber	CES-CES, CES-Legacy
Internet Relay Chat (IRC)	Empathy, Konversation, Xchat, IRSSI	CES-CES, CES-Legacy
Microsoft Notification Protocol (MSNP)	Windows Live Messenger, aMSN, Pidgin, Emesene	CES-Legacy
Skype	Skype	CES-Legacy
Oscar	AIM, ICQ, Pidgin, Empathy, Kopete	CES-Legacy
Yahoo! Messenger Protocol (YMSG)	Yahoo! Messenger, Pidgin, Empathy, Kopete	CES-Legacy

We also tested the use of communications services through web-based interfaces:

- www.ebuddy.com (for MSN, Yahoo, AIM, Google Talk, and ICQ)
- imo.im (for MSN, Skype, Yahoo Messenger, AIM, Google Talk, and ICQ)
- www.meebo.com (for MSN, Yahoo, and AIM)

Each test can give one of three outcomes:

1. The protocol works without problems
2. The protocol does not work but the problems can be solved using an ALG
3. The protocol does not work and an ALG cannot be implemented e.g. because of encryption

NAT traversal can be used to bypass CES devices. Therefore, test results may be affected if the application uses NAT traversal methods such as TURN. In that case, the protocol works because of NAT traversal, but using an ALG allows moving the responsibility for connectivity from the application to the network. The same protocol may succeed in one application and fail in another depending on the use of NAT traversal.

The results are presented in the table below:

Protocol	**Scenario**	**Operation**	**Result**	**Reason**
HTTP	CES-Legacy	Page retrieval	Success	Optimization with proxy
	CES-CES	Page retrieval	Success	
HTTPS	CES-Legacy	Page retrieval	Success	Optimization with proxy
	CES-CES	Page retrieval	Success	
SSH	CES-Legacy	Interactive	Success	
	CES-CES	Interactive	Success	
ICMP	CES-Legacy	Ping	Success	
	CES-CES	Ping	Success	
SIP	CES-Legacy	Calls	ALG required	Private IP used
	CES-CES	Calls	ALG required	
FTP	CES-Legacy	File transfer	ALG required	Private IP used
	CES-CES	File transfer	ALG required	Private IP used
IRC	CES-CES	Messaging	Success	
		File transfer	ALG required	Private IP used
MSN	CES-Legacy	Messaging	Success	
		File transfer	Success	
Skype	CES-Legacy	Messaging	Success	
		Calls	Success	
XMPP	CES-Legacy	Messaging	Application dependent	Private IP used
		File transfer	Application dependent	Private IP used
	CES-CES	Messaging	ALG required	Private IP used
		File transfer	ALG required	Private IP used
Oscar (AIM)	CES-Legacy	Messaging	Success	
	CES-Legacy	File transfer	Application dependent	Private IP used
Oscar (ICQ)	CES-Legacy	Messaging	Success	
	CES-Legacy	File transfer	Application dependent	Private IP used
	CES-Legacy	Calls	Application dependent	Private IP used
YMSG	CES-Legacy	Messaging	Success	
	CES-Legacy	File transfer	Application dependent	Private IP used
	CES-Legacy	Calls	Application dependent	Private IP used
HTTP access to messaging application	CES-Legacy	Messaging	Success	

HTTP works correctly in both directions for the CES-CES scenario and in outbound direction in the CES-Legacy scenario. For inbound HTTP connections in the CES-Legacy scenario, we propose using a proxy, which enables smooth loading of linked elements on a web page.

Some results depend on the application used for testing the protocol. In these cases, the official application (provided by the provider of the service) worked successfully because of the use of NAT traversal. Third-party applications were not working due to the lack of NAT traversal. For these protocols, ALGs are necessary in order to remove the need for NAT traversal and to enable all implementations of the protocol.

4.4.4 Application Layer Gateways

We successfully implemented ALGs for two selected protocols: SIP and FTP.

The SIP ALG was implemented in a stateless way that uses two types of algorithms: 1. adapting addresses and ports in the messages between private and public realms, and 2. replacing the addresses with the corresponding FQDN. The first type of algorithm is used in the CES-Legacy scenario. The CES-CES scenario could also use the first type of algorithms, but that would make the ALG more complex and require state information. Therefore we chose to use the second type in the CES-CES scenario, although we also implemented the first type. The second type of algorithm convey FQDNs in protocol fields instead of IP addresses, which is rather uncommon in today’s SIP applications. However, according to the SIP standard, FQDNs can be used in place of IP addresses in headers and in the SDP body, and our experiments showed that today’s SIP clients handle them correctly.

Scenario	Algorithm	ALG-specific state	Modifies addresses	Modifies ports	Creates mappings
CES-CES	1 (IP)	Yes	Yes	Yes	Yes
CES-CES	2 (FQDN)	No	Yes	No	No
CES-Legacy	1 (IP)	No	Yes	Yes	Yes

As the ALG in the CES-CES scenario replaces IP addresses with FQDNs, new DNS queries are performed by the hosts, which trigger the creation of new connection mappings indirectly. Therefore, the ALG does not need to create the mappings.

If the SIP client instead of using IP addresses uses FQDNs, no ALG would be required. This is the mode of operation we recommend in a CES enabled network.

Our SIP ALG has been tested in all possible scenarios formed from combinations of locating two SIP clients and two SIP servers in three networks (two private and one public). This includes also very unlikely scenarios. We found that all common and most unlikely scenarios work as expected. A few unlikely scenarios would need more complex ALGs than the implemented one.

Our FTP ALG was tested in all scenarios formed from combinations of locating the two endpoints in different networks (two private and one public). All tests were successful.

ALGs for other messaging applications can be implemented in a similar way. We developed a set of guidelines for developing ALGs for CES. Further details can be found in our publications.

4.5 Further work

As a rather new concept, CES provides many opportunities for further development.

Regarding CETP, the encoding format of a few TLVs is open (including the signature TLV). The specification work on the mandatory policy functions (those that have a central role in protocol operation) is ongoing. We also have ongoing work for integrating an existing Deep Packet Inspection (DPI) implementation into the security framework and utilize the reports on suspect traffic in CETP.

The fundamental work on providing connectivity between CES enabled networks and legacy IP networks has been finished. However, interworking with legacy networks will be part of all future development of the CES technology, including security mechanisms and mobility solutions.

Our philosophy is the application should not contain networking layer code in order to enable end-to-end communication. Instead, traversal of the edge is implemented by ALGs in the CES devices for protocols that need special handling. In order to be able to add new ALGs in a flexible and secure way, we need to define an interface between the ALG and the CES. This interface might allow the ALG to reside on a different device than the CES itself. Furthermore, the ALG could be developed by a third party. Therefore it must be run in a secure way. One option would be to run it in a separate sandbox.

The work on protocol compatibility testing could be continued with other types of protocols. Untested protocol categories include protocols used in peer-to-peer applications.

4.6 Conclusions

CES replaces the current NATs and firewalls. Using CES allows allocating private addresses to customers instead of public addresses. CES provides the user with connectivity that is very similar to using public addresses, and allows the user to accept inbound traffic. Thus, the need for public addresses can be significantly reduced without disadvantages to users. The shortage of public IPv4 addresses is one of the most important problems the Internet is currently facing.

In addition to address reuse, CES improves security. As CES devices can communicate with each other using CETP, the endpoints can agree on the required level of trust to enable communication. For example, a public web server could serve almost anyone without any identification. On the other hand, private users may require more information from the other party before providing access. As new communication paradigms, such as Internet of Things, becoming widespread, there is an increasing need for secure and reliable communication with particular needs. As various types of attacks and unwanted traffic become more common, there is a need to enable only expected traffic to entering the operators and customer’s networks.

In the development of CES, our work has been concentrated on a few main topics, including CETP, PRGW, ALGs and protocol compatibility.

CETP provides a tunnel protocol that carries identity information and offers bandwidth reduction with a compressed IP header. Furthermore, CETP provides control plane functionality used for negotiating the information required for accepting connections according the user-defined policies. CETP also enables negotiation of transport addresses, providing multihoming and tunneling over various types of transport. Our proof-of-concept implementation shows that the protocol is feasible. Testing of various combinations of policies shows that the combinations provide the expected outcomes.

Compared to a normal NAT, a PRGW enables inbound connections. This avoids the need for unreliable and insecure NAT traversal algorithms or fixed port forwarding. The PRGW allows a high number of connections sharing the same public address. An operator can allocate a small pool of public addresses for inbound traffic, which allows users to receive traffic on the standard ports used by protocols. Although the PRGW has been developed as a module for the CES, the solution is generic in nature and can be used to replace current NATs even in case the whole CES is not adapted. However, using PRGW without CES does not provide any security mechanisms and does not provide the benefits that CES does for traffic between two networks using PRGW. The PRGW allows CES to be introduced gradually, one network at a time, offering some of the benefits of CES. As the devices become more common, the full features of CES-to-CES communication become available.

CES shares many of the properties and problems of other middleboxes such as NATs and Firewalls. The current solution to connectivity problems is either to use NAT traversal (adapting the application to the network) or ALGs (adapting the network to the application). These solutions can be applies to CES as well. However, without ALGs and NAT traversal mechanisms, CES provides inbound connectivity to a large group of client-server applications that would not work with NATs. The requirement for this is that 1) the protocol does not transport IP addresses in protocol messages and 2) performs a DNS query to the FQDN of the destination before communication.

Our study showed that several communication protocols tend to convey the local private IP addresses to the peer in protocol messages. These fail in the presence of CES (and other middle boxes). Some of the applications solve the problem using NAT traversal mechanisms. We showed that we can avoid NAT traversal by using ALGs. By implementing ALGs for selected protocols we obtained general principles for developing ALGs. Using ALGs instead of NAT traversal enables faster connection setup (without the delay caused by trying various traversal methods) and avoids the bandwidth and delay overheads caused by third-party relays. The ALG-based solution simplifies application development and does not require the service provider (or network provider) to relay data (causing costs as well as responsibilities). Removing third-party elements such as STUN/TURN servers and relays, improves security by reducing the risk for man in the middle attacks. Using ALGs allows the network provider to better control the protocols used on the network and integrate with DPI.

We consider ALGs as a solution for the particular applications that do not natively work with CES. The preferred way is to implement applications and protocols using the FQDN for identifying hosts instead of the IP address. In addition to transparent operation with CES, this makes the application less dependent of the network layer and thus more future proof.

4.7 List of publications

The following publications have been submitted for review:

- Jesús Llorente Santos, Raimo A. Kantola, Nicklas Beijar and Petri Leppäaho. Implementing NAT Traversal with Private Realm Gateway. Submitted to IEEE International Conference on Communications (ICC), 9-13 Jun 2013.

  Abstract: A Network Address Translator (NAT) allows hosts in a private address space to communicate with servers in the public Internet. There is no accepted solution for an arbitrary host in the Internet to initiate a communication with a host located in a private address space despite the efforts to create one. This paper proposes to replace NATs with a new concept we call Private Realm Gateway (PRGW). Private Realm Gateway creates connection state based on incoming DNS queries towards the hosts in the private network. The state gives means for the private network operator to apply elaborate access control to packet flows arriving from the Internet to the private network. PRGW does not require changes in the hosts and the deployment can take place one network at a time. The paper shows that the PRGW is most applicable for connecting mobile and other wireless hosts to the Internet.

- Petri Leppäaho, Nicklas Beijar, Raimo Kantola, Jesús Llorente Santos. Traversal of the Customer Edge with NAT-Unfriendly Protocols. Submitted to IEEE International Conference on Communications (ICC), 9-13 Jun 2013.

  Abstract: Customer Edge Switching (CES) provides policy based reachability to hosts in a private network without the disadvantages caused by traditional mechanisms for traversing Network Address Translators (NAT). Although most protocols traverse the customer edge correctly, we identify a few protocols that require special processing because of the IP addresses carried in the user data. This paper first presents the results of protocol compatibility testing with CES and selects two protocols, SIP and FTP, for further study. The paper then reports the implementation of Application Layer Gateways for these two protocols and gives guidelines for processing other protocols. The solution enables transparent communication across address realms without keep-alive signaling and application layer code in end systems as required by the current recommended approach to NAT traversal. The proposed approach significantly cuts the session establishment delays typical in SIP and improves security. The presented work is a part of a larger project that proposes the Customer Edge Switching to replace NATs and form collaborative firewalls for protecting customer networks.

Part of the work has been implemented and published as Master's theses:

- Petri Leppäaho, Design of Application Layer Gateways for Collaborative Firewalls, Aalto University School of Electrical Engineering, May 2012.
- Jesús Llorente Santos, Private Realm Gateway, Aalto University School of Electrical Engineering, Work in progress.

The process of writing publications on CETP is currently ongoing. The following publications will be produced:

- A conference paper on CETP and the related policy and security methods.
- An IETF Internet Draft on CETP.
- A Master's thesis on implementing the CETP protocol and testing various policies.

Additionally, we plan to publish an IETF Internet Draft on the server-side NAT solution, since this concept can be utilized in a more generic way in the current Internet.

5. L2 Routing In Core Networks Based On Carrier Grade Ethernet with Centralized O&M

5.1 Introduction

In order to utilize Ethernet as the common transport for full scale deployment some improvements on the physical layer have been proposed such as running Ethernet over Synchronous Optical Networking (SONET) and Synchronous Digital Hierarchy (SDH). This allows the usage of high speed optical fibers to transmit Ethernet frame. Moreover, Ethernet has also tapped into the Multi Protocol Label Switching (MPLS) which directs data from one node to the next using path labels and not the historical network addresses.. MPLS has become one of the future carriers since it can utilized in the backbone with various technologies like, Ethernet, Frame relay, DSL etc. MPLS has a wider coverage and it can easily be used to link various VLANs over a wider geographical area.

A number of enhancements are proposed lately some of which have included VLANs, provide backbone bridging, double tagging, with the latter coming in to provide more control in VLANs.

One of those proposals to overcome some deficits in Ethernet, is the Provider Backbone Transport (PBT). PBT is actually a combination of some existing technologies and the reusage of Vlans, 802.1ad Q-in-Q double tagging and combining them with a subset of 802.1 ah Mac in MAC elements. This provides connection oriented transport which Ethernet is lacking.

5.2 Centralized Ethernet Routing

In this work we intend to utilize Ethernet routing with a central root element which takes care of detecting link breaks and calculating new routes.

We used Boson network simulator which is one of the latest Cisco applications used to simulate its soft and hardware devices. It uses a Cisco IOS command structure and a command line interface. It uses various engines like the router and e-router software technologies, virtual packet technologies from which it creates virtual packets that can be simulated to create routing tables hence providing an adequate network environment. The simulator provides various capabilities and a variety of devices that can be configured to provide adequate results depending on ones needs. It provides an environmental capacity of configuring a network with about 42 different router models and over 5 different switch models and other various devices with one of the most advanced software within the industry. It has an NMap topology design feature which uses a simple drag and drop feature allowing one to create any topology desired and also provides two different viewing styles i.e. telnet or console mode. It also provides device configuration windows where configuration commands can be entered and network traffic simulated. This particular package i.e. Boson netsim 8.0 has an advanced network simulator which has various components and allows configurations of various protocols like, RIP 1&2, IGRP, EIGRP, VLSM, OSPF (all areas), route redistributions, IS-IS, policy routing, BGP, VLANS, SPT configurations, PortFast, Uplink fast, routing on distribution switches, VLAN access control lists, troubleshooting based on gathered symptoms, Host Standby Routing Protocols (HSRP) and VRRP, traffic generator among others.

From this tool we utilized the improved NMAP tool to create various topologies, configuration environment which allowed configuration of for example VLANS, HSRP, STP, traffic generator. From these simulations we took echo request values or ping in terms of time (seconds) and also packets lost. In our simulation we took the minimum, maximum and average time from the various scenarios. We also got the number of bytes lost in terms of percentages.

5.3 Simulation scenarios

In our work, we looked at six different scenarios each having various components. In the first scenario we had a topology of 12 switches, two of them worked as the central with the remaining 10 working as operational and seven end devices. In this centralized scenario we measured various times it took to send and receive packets through the network, packet loss and convergence times as we shall see later. In the scenario 2 which was a distributed kind of topology we employed only operational switches which were 10 with the same number of end devices. We sent and received ping requests to the same end devices and measured the various times it took to reach various nodes and back together with packet loss and convergence times.

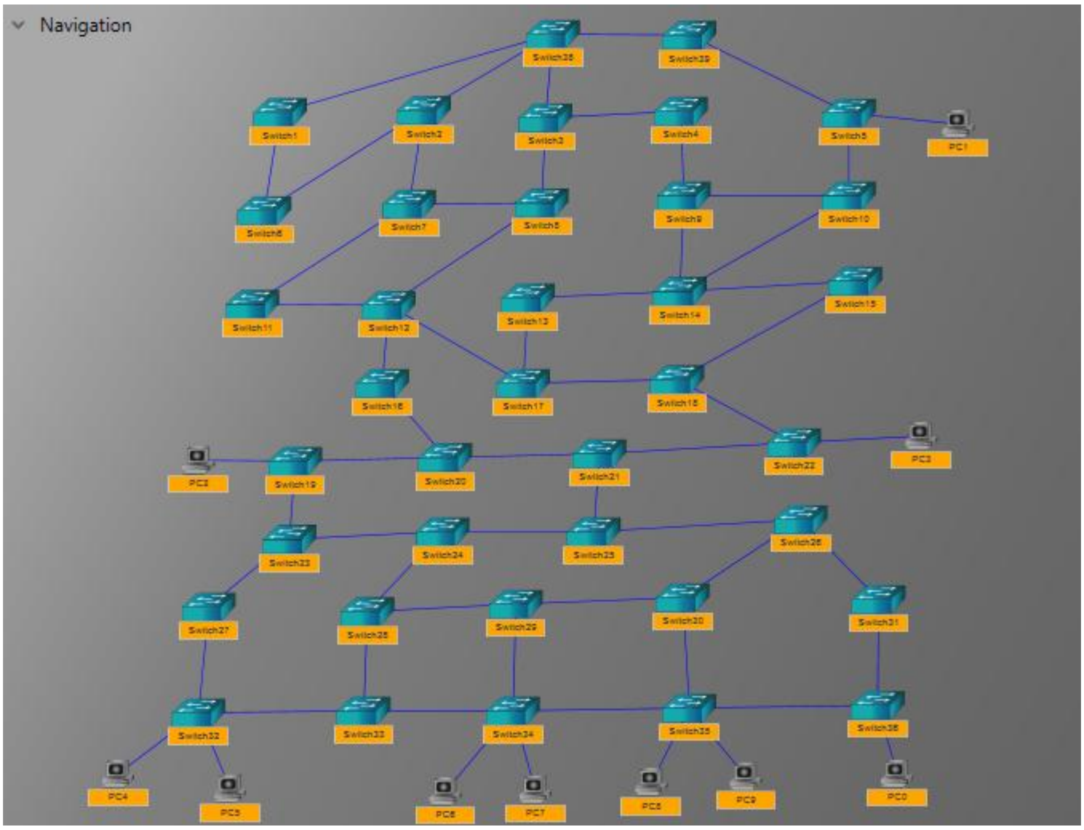


Figure 11. Distributed routing: 38 switches + 10 end devices.

In scenario 3, which was based on a centralize setting we used 22 switches 3 of which worked as the central command providing back up to each other in case of one getting off, the remaining 19 operational and 10 end devices. We used the same parameters in measuring the various times it took to send requests to and from various nodes through the network, packet loss and convergence times in cases of link breaks. In the fourth scenario was based on a distributed kind of routing and was similar to the third scenario but we only used the 19 operational switches and 10 end devices. The ping requests were similar to the ones sent in scenario 3 and were sent to the same nodes for easy comparison. In scenario 5, which was based on a centralized kind of routing we used 41 switches (3 central), 10 end devices, in this we continued with our analysis of the time it took to send and receive packets through the network and convergence times together with packets lost on the way. In scenario 6 we did the same but with only 38 switches because it was based on a distributed kind of routing. The central command was configured on a Hot Standy Routing Protocol base which is one of the best way for providing fail overs.

Scenario 1:

a) Centralized routing topology. 12 switches (2 central + 10 operational), 7 end devices.

From	To	Min time (ms)	Max time (ms)	Average
PC0	PC2	53	70	61
PC0	PC3	51	63	58
PC2	PC6	49	72	61
PC6	PC0	50	70	58

Scenario 2:

b) Distributed routing topology: 10 operational + 7 end devices.

From	To	Minimum time	Max time	Average
PC0	PC2	53	70	63
PC0	PC3	52	62	57
PC2	PC6	49	61	64
PC6	PC0	54	68	60

In scenario 1 where we looked at the centralised kind of Ethernet routing solution we used 10 switches i.e. 2 in central command and 8 operational together with 7 end devices. In scenario 2 which was a distributed kind of Ethernet solution we only used 8 operational switches with 7 end devices. The idea was to compare the centralised and distributed Ethernet routing in terms of delays or time taken to send and receive packets from different ends of our topology and also convergence times in cases of link breaks. So we sent various ping requests from various end devices to others throughout the network as shown in the tables above. We found out that the Centralised Ethernet routing was slightly better and faster in terms of time compared to the distributed Ethernet routing. The centralised solution had an average time of 59.5 ms compared to the distributed topology which had an average of 61 ms.

Using the same topologies, we also compared the convergence times between the two different Ethernet routings. The average time it took for the centralized routing to re-converge after a link break was 28.5 seconds compared to the 30 seconds it took for the distributed Ethernet routing. So despite the hassles and time it took to configure the centralized Ethernet routing, its much better in terms of delays and convergence times as compared to the distributed Ethernet routing making it a viable solution to smaller topologies like this one.

Scenario 3

c) Centralised routing topology with 22 devices 3 central + 19 operational + 10 end devices.

From	To	Mini time (ms)	Max time (ms)	Average
PC0	PC1	52	65	59
PC0	PC7	52	66	58
PC1	PC5	51	72	60
PC2	PC7	50	71	62
PC6	PC8	54	69	60
PC7	PC9	53	72	63
PC4	PC1	48	69	58

Scenario 4 table:

d) Distributed routing topology: 19 operational + 10 end devices

From	To	Min time	Max time	Average
PC0	PC1	50	70	63
PC0	PC7	54	63	59
PC1	PC5	57	71	61
PC2	PC7	48	71	61
PC6	PC8	57	70	65
PC7	PC9	49	71	61
PC4	PC1	48	54	51

In scenario 3 we used 22 switches of which 3 were in the central command and 19 operational switches and 10 end devices. In scenario 4 we used 19 operational switches and 10 end devices where we continued to compare both the centralised and distributed Ethernet routing. In our comparison we looked at the average times it took to send and receive packets throughout the topology and also the convergence times in case of a link break. In scenario 3, we sent out various ping requests as envisaged in the tables and found out that the average time it took for a ping request was 60 ms as compared to the distributed Ethernet routing which took 60.143 ms. So in this we concluded that when it comes to sizable topologies of about 20 switches centralised routing and distributed routing are not really different in terms of delays and packet loss and either can be used, however the difference comes in with the convergence times which were really different.

Centralized routing took a little longer to converge i.e. 39 seconds on average compared to the distributed routing which took 31 seconds. This showed a rather distinctive advantage to the centralised kind of routing. This together with the configuration hassles attributed to the centralised kind of routing made the distributed Ethernet routing a better choice in dealing with link breaks.

Scenario 5:

e) Centralised routing topology with 41 switches (3 central), 10 end devices

From	To	Min time (ms)	Max time (ms)	Average (ms)	Loss
PC0	PC1	61	150	92	25%
PC2	PC9	25	259	87	0
PC4	PC1	56	322	178	0
PC6	PC7	105	297	201	0
PC3	PC2	62	242	125	25%
PC9	PC4	71	494	274	25%

Scenario 6:

f) Distributed routing topology with 38 switches and 10 end devices.

From	To	Min time	Max time	Average
PC0	PC1	152	193	169
PC2	PC9	20	260	82
PC4	PC1	40	260	151
PC6	PC7	6	126	46
PC3	PC2	25	180	65
PC9	PC4	29	265	111

In scenario 5 we looked at 41 switches of which 38 were operational and 3 in the central command together with 10 end devices and in scenario 6 we looked at 38 switches with 10 end devices. In these also we continued our analysis in centring on delays and convergence times. Using ping requests from various end devices to other throughout the network as envisaged in the tables we got an average time of 159,5ms from the centralised Ethernet routing as compared to the distributed kind of routing which had 104 ms. This was a very big and clear difference between the two solutions.

When it came to looking at the convergence times our centralised solution took a still longer time of 46 seconds compared to the distributed solution which increased by only 2 seconds from 31 to 33 seconds. These two scenarios i.e. 5 and 6 clearly answered the scalability question we had as according to these scenarios a centralised solution becomes less and less optimal as the network enlarges compared to the distributed kind of solution which becomes more stable and has room for more.

Convergence times comparison between a centralized and distributed Ethernet routing.

	Centralized	Distributed
9 switches	28.5	30 seconds
18 switches	39	31 seconds
36 switches	46	33 seconds

5.4 Conclusions

The usage of centralized Ethernet routing shows improvements when considering large number of switches with fast connectivity. In this deployment the delays between distributed switches and the master are low so it compensates the delays of distributing routing information among the edge switches.

6. Automatic and Secure HIP-Based VPN Service

6.1 Introduction

Virtual Private Networks (VPNs) are popular in the wide area Internet to extend the private network domain to geographical distributed locations via an unsecured public network or to separate specific section of the network from public access. Several techniques have been defined to provide VPNs at different layers of the Open Systems Interconnection (OSI) model.

Host Identity Protocol has used as a technique to develop Laye3 VPNs such as IPsec VPNs and most recently Laye2 VPNs such as VPLS. The basic use case of the HIP based VPNs is to address the existing VPN/VPLS problems related to security, mobility and multihoming.

The Long Term Evolution (LTE) architecture proposes a flat all-IP backhaul network. 3rd Generation Partnership Project (3GPP) specified new security and traffic transport requirements of new LTE backhaul network. However, existing LTE backhaul traffic architectures are incapable of achieving these security requirements. In this research, we are focusing on how to provide the security features which are specified by 3GPP for the LTE backhaul networks.

On the other hand, various types of traffic will be transported by the LTE backhaul starting from evolved nodeBs (eNBs), such as S1-U traffic to the Service Gateway (SGW), S1-C traffic to the Mobility Management Entity (MME), X2-U and X2-C traffic to other eNBs etc . There are two crucial traffic transport issues identified due to these different traffics. First issue is to backhaul different traffics to the correct destination. Second problem is to provide different levels of Quality of Service (QoS), priority and fault management requirements for different traffic types. A VPN based backhaul traffic architecture is a promising solution to fix above issues.

Hence, we focus on developing secured VPN architectures not only to fulfill LTE backhaul security requirements but also to solve the above traffic transport problems. We are applying the HIP based VPN solution to mobile backhaul as a proposal. There are two HIP-based solutions proposed which are depending on underline backhaul network, in order to secure this VLAN based backhaul architecture.

1) HIP based VPLS for Layer 2 backhaul network

The backhaul network can be considered as a layer 2 network, when it has the layer 2 switches and equipment. In such a use case, we proposed to use HIP based VPLS scenario. Here, HIP used to create a secure VPLS overlaid on top of the untrusted backhaul network. This application of HIP differs from the traditional implementation of HIP within end hosts, because the payloads of the ESP-encrypted datagrams are not transport protocol data units (PDUs) instead are layer-2 frames. The access control decisions for the VPLS are taken by using the HI of the users. Therefore, separate authorization server (e.g. AAA) is needed to assist this task. For a new join request for a new user, existing VPLS user needs to verify the user by contacting the authorization server before granting the access to the new user.

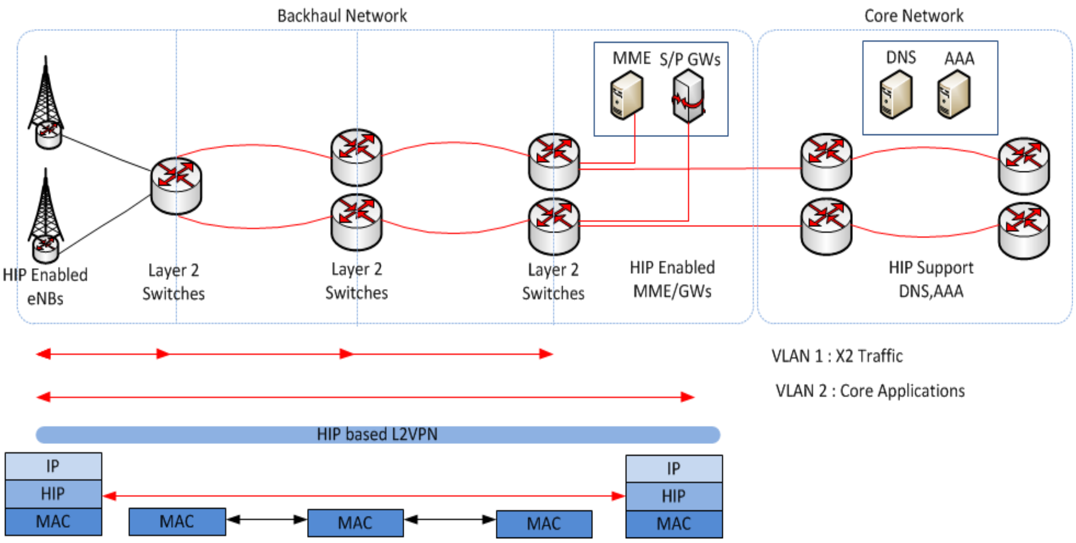


Figure 12. HIP based VPLS for Layer 2 backhaul network

Figure 12 illustrates the general protocol stack of the proposed Layer 2 VPN solution.

Several secured VPLS architectures are proposed during recent years and many of them do not provide a sufficient level of security. HIP (Host Identity Protocol) based Virtual private LAN service (HIPLS) is the first and only proposal which provides sufficient level of VPLS security. However, HIPLS has several issues such as lack of security plane scalability due to massive key requirements and lack of forwarding plane scalability due to inefficient broadcast mechanism. Hence, these HIPLS cannot be used in larger networks such as LTE backhaul.

We present a novel secured VPLS architecture based on HIP protocol by accounting above scalability issues. We proposes a session key based HIP VPLS (S-HIPLS) architecture which reduces the key storage complexity at a PE and the whole network while providing a higher degree of security features than other proposals. Additionally, it ensures the scalability and provides an efficient broadcast mechanism for the VPLS network.
The more details about the proposed solution is described in our journal article titles “A Scalable and Secured VPLS Architecture for Service Provider Networks”.

2) HIP based VPN for Layer 3 backhaul network

The backhaul network can be considered as a layer 3 network, when it consists of layer 3 routers and other equipment. In such a use case, we proposed to use HIP based VPN scenario. Here, HIP used to create a secure VPN overlaid on top of the untrusted IPv4/IPv6 backhaul network. This application of HIP almost similar to the traditional implementation of HIP within end hosts. Basically, the payloads of the ESP-encrypted datagrams are transport protocol data units (PDUs) as original HIP specification. Similar to the previous proposal, the access control for the VPN is checked by using the HI of the users. Therefore, separate authorization server (e.g. AAA) is needed to assist this task. The access control decisions for the VPN are taken by using the HI of the users. Therefore, separate authorization server (e.g. AAA) is needed to assist this task.

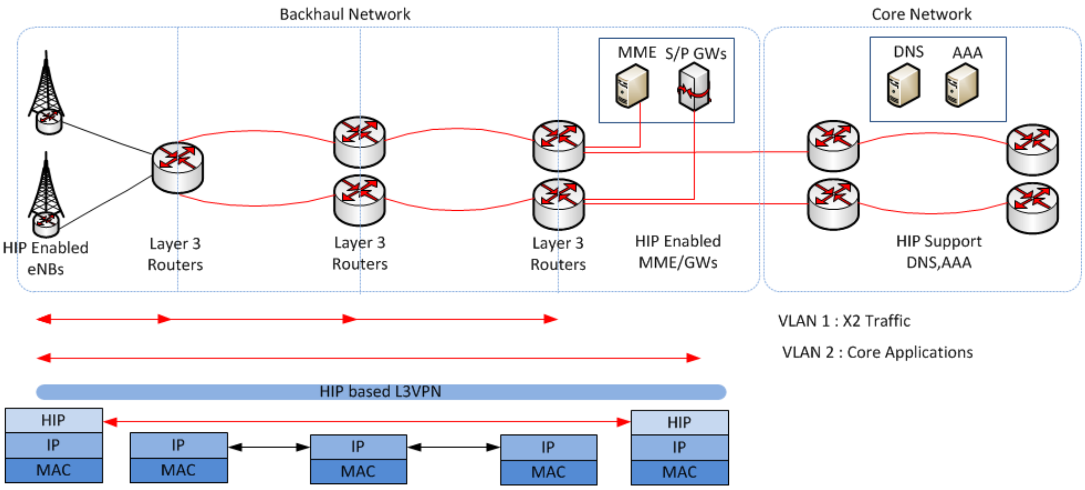


Figure 13. HIP based VPLS for Layer 3 backhaul network

Figure 13 illustrates the general protocol stack of the proposed Layer 2 VPN solution.

We propose two secured Virtual Private Network (VPN) architectures for LTE backhaul. Both architectures are layer 3 Internet Protocol security (IPsec) VPNs which are built using Internet Key exchange version 2 (IKEv2) and Host Identity Protocol (HIP). They are capable of fulfilling 3GPP security requirements such as user authentication, user authorization, payload encryption, privacy protection and IP based attack prevention. Furthermore, our proposals provide additional load balancing, automatic redundancy and best path routing capabilities in a fully or partly mesh backhaul network. Finally, we verified the advantages of the HIP based VPN solutions over other IPsec VPN solutions.
The more details about the proposed solution is described in our journal article titles “Secure Layer 3 VPN architectures for LTE Backhaul Networks” and conference paper on “Secured VPN models for LTE Backhaul Networks”

6.2 Main results

The mains results are summarized under each proposed solotion.

1. HIP based VPLS for Layer 3 backhaul network

   a. Layer 3 Secured traffic architecture of LTE Backhaul networks.
   b. Provide security plane by significant reducing the complexity of the key storage at a VPLS node, total key storage of the network
   c. Provide control plane scalability by significant the number of encryption per a broadcast frame..

2. HIP based VPN for Layer 3 backhaul network

   d. Layer 3 Secured traffic architecture of LTE Backhaul networks..
   e. Provide additional load balancing, automatic redundancy and best path routing capabilities in a fully or partly mesh backhaul networks.

6.3 Further work

This research serves a base for future studies such as study the impact of mobile backhaul nodes such as Mobile Femto Cells (MFCs) to the VPN architecture, develop a secure hierarchical VPN architecture for LTE backhaul and provide optimum load balancing mechanisms for multihomed nodes. Furthermore, we focus on extending our VPLS architecture for secured hierarchical VPLS networks and secured Virtual Private Multicast Services(VPMS).

6.4 Conclusions

We proposed Layer 2 HIP based VPLS and Layer 3 HIP based VPN architectures to secure the backhaul traffic. Both solutions provide the security features specified by 3GPP for the LTE backhaul. Namely, user authentication, user authorization, payload encryption, privacy protection and IP based attacks prevention.

6.5 List of publications

Madhusanka Liyanage, Andrei Gurtov, Secured VPN Models for LTE Backhaul Networks , to appear in Proc. of IEEE 76th Vehicular Technology Conference: VTC2012-Fall, Québec City, Canada September 2012.

Madhusanka Liyanage, Andrei Gurtov, Secure Layer 3 VPN architectures for LTE Backhaul Networks, submitted to IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Madhusanka Liyanage, Andrei Gurtov, A Scalable and Secured VPLS Architecture for Service Provider Networks, submitted to IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

7. Proxy HIP-based Secure Backhaul in Femtocell Technology

7.1 Introduction

The evolved communication technology introduces wide-spreading residential access points that enable mobile communication through the residential networks. The mobile networks can be widely spanned with the introduction of femtocells extending the operator network to subscriber residence. The home based FAPs enable access to cellular networks over the broadband connectivity. FAPs are 3G hot-spots to which the mobile users can connect over the same Global System for Mobile Communications (GSM) band. Even, FAPs may be WiFi enabled to support WiFi handsets. The Evolved Packet Core (EPC) architecture based on all-IP concept is adapted in femtocell technology. LTE focuses on the extensive use of subscriber installed FAPs for improved network coverage and high-speed connectivity. FAP establishes IPSec tunnels in either direction through the backhaul to protect the communication from attackers. It is realized that the connectivity between FAP and Secure GateWay (SeGW) is vulnerable to attacks since, both control and data traffic is carried over the unreliable broadband access or public Internet. Thus, protecting femtocell backhaul is a crucial requirement for secure communication.

The open access FAPs are somehow problematic, since the number of subscribers can be served simultaneously is limited. Increasing number of mobile nodes attached to FAP may degrade service quality or prevent desired subscribers accessing operator network. Therefore, access control is a critical requirement in femtocell technology. On the other hand, close access FAPs filter subscribers using Closed Subscriber Groups (CSG), though it may reduce the overall performance of the system. The existing femtocell architecture demands globally unique routable identity to be assigned on each connected device. In case of lacking IP addresses, mobile nodes that demand addresses to configure on it will not be served. For this reason, some operators implement address translation and address mapping in certain devices along the path. When it comes to mobility, IP addresses as identifiers result problems in user mobility. Therefore, identity, locator separation is highly demanded in mobile applications. HIP introduces a new identifier which obligates the rules of Domain Name Service (DNS). Thus, the change in IP address corresponds to the point of attachment may not affect transport layer associations.

In this report, we propose a modification to the existing protocol stack of the 3GPP femtocell architecture. We are more focused into mobility and security issues related to femtocell technology. This work proposes several enhancements to the femtocell technology such as, service registration, identity verification and node multi-homing.

7.2 Detailed description of the proposed solution

Femto Access Point Security

The femtocell security consists of FAP authentication and message encryption across the unreliable public network. Femtocell backhaul is vulnerable to any external attack since, there is no guarantee of security by the network provider. The femtocell security aspects are not yet standardized according to the 3GPP specifications. Thus, there are many ongoing research efforts to enable an end-to-end secure communication in femtocell technology. FAP authentication is a major consideration in femtocell security. In general, FAP authentication is performed using Extensible Authentication Protocol Method for Authentication and Key Agreement (EAP-AKA), certificate or as a combination of both. The 3GPP standard presumes validation and authentication to be performed sequentially. Thus, during the initial power-up, FAP gets authenticate to the core network.

If the certificate based authentication is used, the mutual authentication between the FAPand the core network is performed with X.509 certificate which is already configured at FAP and SeGW. Rather, Universal Integrated Circuit Card (UICC) that defines the identity of the secondary hosting party is used for the authentication. The FAP'sTrusted Environment (TrE) holds these credentials that are used to authenticate it to the core network. It is important to protect the certificate and any other data such as certificate revocation list during the operational lifetime and the time it is provisioned. Thus, a malicious user who attempts to manipulate the public key to impersonate the SeGW can be easily isolated. If EAP-AKA based authentication is used, the credential should be provisioned in the TrE of the FAP for non-3GPP access.

However, there is a high risk of compromising the authentication token via a brute force attack or a local physical intrusion. Further, a valid authentication token can be inserted into a manipulated FAP and can be used for harmful actions. The UMTS standard defines security in four domains such as network access security, network domain security, user domain security and application domain security. However, femtocells confront major security problems in locating a mobile user based on UICC and signaling messages, eavesdropping, DoS to User Equipment (UE) and core network and attacks on data integrity. The exposure of the core network to the Internet is the major vulnerability in this architecture. This inspires the intruders to execute Internet-based attack such as, node impersonation, DoS or Distributed DoS (DDoS). The exposure of a public IP to the Internet through which FAPs access the operator network is a potential point of failure in the femtocell architecture. For instance, it is well-known that many large companies have confronted DoS attacks. Distributed security mechanisms are more effective in detection of DDoS attacks since suppression mechanisms are most powerful close to the origin of the attack. However, the protection against such attacks demands the cooperation with Internet Service Provider (ISP) as well as the neighboring ISPs.

Femtocells' Mobility Issues

In the network layer, mobile nodes are identified by the IP address which is based on the actual topological location. In other words, IP address depicts both location and the identity of a particular mobile device. In general, overload nature of IP is a problem in IP domain. Mobility management becomes more crucial when the active sessions get interrupted by the change of point of attachment to the Internet. If IP addresses are only geographical locators, they identify the location of the mobile node but not the identity. Hence, there should be an additional technique to represent the identifier role which is relied at the transport protocol. In handover, upper layer protocols such as IPSec guarantees security though; it is only capable of applying and agreeing certain encryption standards between the nodes. This is somehow inefficient and unconvincing since; it does not help to mitigate Denial of Service (DoS) or node impersonation. Deployment of evolved mobile applications needs extensive support of security and mobility.

But, extended security may increase the communication overhead and processing power. Security a device can promise depends on signaling overhead and processing power of mobile device. The support of advance mobility and multi-homing scenarios such as simultaneous multi-access, network mobility, application mobility and session mobility together with seamless vertical handover are few challenges in existing femtocell architecture. Certain types of applications such as online games, movies and video calls demand high bit rate over the channel. The smooth handover between the femtocells carries a significant performance indication in terms of quality of service towards the mobile users including pedestrian and vehicular users. However, this handover scenario demands close investigating of the features inevitable to femtocells.

3GPP Specified Backhaul

The validation of FAP demands mutual authentication and initiate secure associations in either direction as a result of the authentication. An IP address is assigned to the FAP as a result of successful authentication and the secure backhaul connections are established in either direction for inbound and outbound traffic. These IPSec tunnels are established based on Internet Key Exchange version 2 (IKEv2). It provides layer-3 security and supports port and Network Address Translation (NAT). The following presents the femtocell architecture that consists of several major components such as Security Gateway (SeGW), Home Subscriber System (HSS), evolved NodeB (eNB), Packet Data Network gateway (PDN-GW) and Mobility Management Entity (MME).

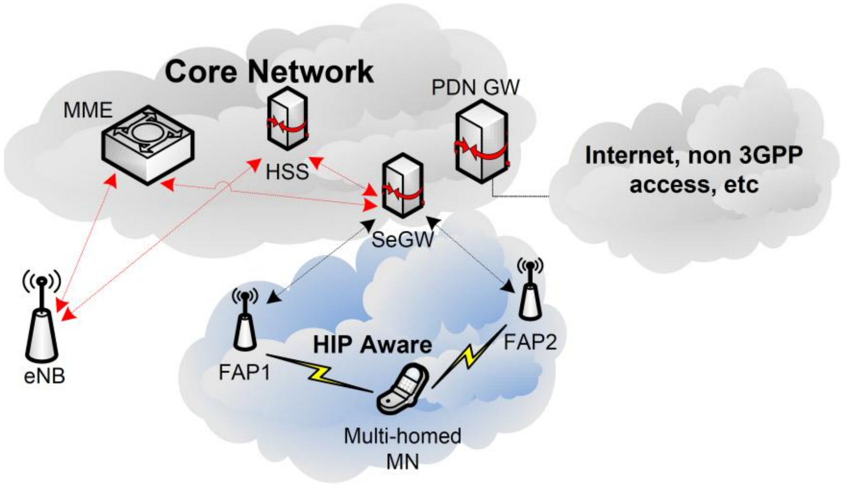


Figure 14. Architecture model for Home NodeB access network.

Acquiring an IP address FAP creates secure tunnel to SeGW. Separate tunnels through the backhaul can be established to exchange different type of traffics such as Operation Administration and Maintenance (OAM), validation and QoS information while primary tunnel is used to transmit bare traffic and signaling. When the peer node is not behind the same SeGW, the platform integrity should be verified alone the backhaul connection during the validation procedure. Hence, a separate network element should keep track on the state of validation of the platform integrity conjunction in the backhaul connection. Ultimately, this approach dictates an additional complexity keeping track of the states of each device platform integrity over the backhaul connection. Moreover, if a device is validated only at the authentication, the validity of platform integrity must be revised. Thus, an update policy for platform integrity validation procedure is executed in case of modification or termination of a backhaul connection. Further, this information is reflected to other devices which keep track of the platform integrity.

HIP Based Femtocell Backhaul Solution

This section presents a HIP based secure backhaul solution to handle mobility and security issues in 3GPP standardized femtocells technology. With the proposed HIP based solution, the IP addresses are no longer listed as identifies. Ultimately, it denotes the point of attachment of mobile node to the core network. However, IP address still performs network layer routing while separate name space is proposed to manage identity which does not change once it is configured. The Home Subscriber Servers (HSS) records the authentication information and subscription data correspond to each FAP and is retrieved whenever it is requested by the Authentication, Authorization and Accounting (AAA) server. The standard defines optional hosting party authentication which is based on the credentials stored in Hosting Party Module (HPM). However, it is out of our focus in this paper. HIP inherits several advanced mobility and security features including extended multi-homing support, middlebox traversal and address translation. In the following subsections we discuss how these features can be adapted in femtocell technology to improve security and to support mobility.

HIP-Based Secure Femtocells

The rapid growth of mobile communication revels mobility, not only to the nodes but also to the networks of many connected nodes. We present mobility in terms of node mobility and network mobility. There are three generic approaches of handling mobility signaling. The first approach assumes mobility signaling for each mobile node is handled individually by the node itself. This involves more signaling overhead, processing and long handover reaction time when the number of mobile nodes increases. The next approach is based on traffic tunneling where signaling traffic generated at the mobile node to the gateway is tunneled to a fixed gateway in the operator network.

This approach may not use the optimal path introducing an unexpected delay due to triangular routing. Introduction of IPv6 can resolve the problem of triangular routing which is a common issue with many Mobile IP (MIP) proposals. However, the tunneling overhead in the second approach may increase the packet size which results to lower the throughput. In the third approach, the mobile node delegates rights of mobility signaling to an associated gateway which may further delegate mobility signaling rights to a Local Rendezvous Server (LRVS) located in the core network. This proposal is a combination of above three approaches. In this approach, mobile devices and core network are assumed to be HIP aware. Moreover, specific Network Address Translation (NAT) mechanism which performs SPI mapping (SPINAT- Security Parameter Index multiplexed Network Address Translation) is adapted to hide node identities behind the NAT. SPINAT uses SPI value in ESP packets to demultiplex multiple IP addresses on the same IP address.

In the next subsection, we discuss SPINAT in detail. Here onwards, we assume FAPs are authenticated at the initial boot-up using the base exchange defined in HIP. The mobile node configures an IP address using whatever the available technique in place such as, manual configuration, DHCP or stateless auto-configuration. For instance, in stateless auto-configuration, the mobile node receives one or more prefixes correspond to its domain gateway or the associated FAP. The mobile node randomly selects an address out of the dedicated prefixes. Upon entering to the FAP domain, the mobile node acquires an IP address and triggers Security GateWay (SeGW) to run the base protocol. During the Base Exchange, the common keying materials are created and exchanged using Diffie-Hellman key exchange mechanism. Thus, the keys drawn from the keying material can be used to protect the signaling and data traffic.

For this reason, nobody except the mobile node and the SeGW can decrypt the communication. The Figure below presents the node registration and the handover from one femtocell to another. The HIP support over this use case is further explained in the coming paragraphs. In this case, the SeGW reads the cell information of the target femtocell and performs the access control for the non-CSG mobile nodes. For the CSG-capable mobile nodes, the access control shall be done by the core network and the result will be sent back to SeGW. If the target FAP is allowed access, the SeGW will then send the handover request to it. Since the SeGW only has the information of the connected FAPs, it is applicable only to the intra-GW femtocells. If the source and target femtocells belong to a different SeGW, the core network coordinated handover procedure should be invoked instead. By handling the handover procedure using the SeGW, the handover latency and the load of core network are reduced. However, new functionalities need to be added to the SeGW so that it is able to read and forward the handover request message.

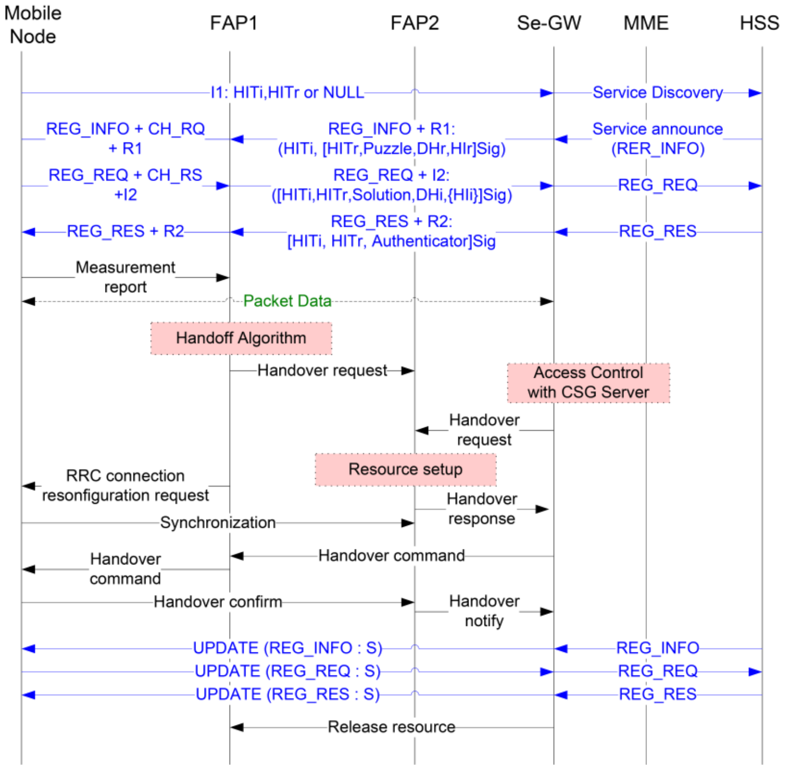


Figure 15. HIP based call flow in femtocell communication.

By nature, HIP enables end-to-end security. Thus, nobody except the end hosts can encrypt the communication even if the communication is eavesdropped. However, an attacker can still perform replay attack on the HIP hosts. In this proposal, we suggest a challenge, request based replay mitigation procedure which is presented in the above figure. The first step in the figure triggers the Base Exchange sending an I1 packet which includes the mobile node's HIT and the SeGW's HIT. The SeGW's IP address can be obtained from the DNS or repository service in place. If the opportunistic mode is used, the responder HIT field is kept null.

The I1 messages always pass by the FAP's associated Local Rendezvous Server (LRVS). This mechanism enables fast route updating when the mobile node moves with the correspond node. The generic DNS mechanism may not be a good solution to handle fast reroute updating in such situations. This proposal describes, how the HIP based passive service discovery can be used in femtocell technology. The Base Exchange authorizes the mobile node to exchange the service related information concatenated to the Base Exchange. The mobile nodes do not want to actively query the services since, it is not feasible to perform FAP discovery each time the nodes move. On receiving the I1 packet, the SeGW forwards it on upstream to the HSS. The HSS maintains records such as, subscriber profile database, service permissions, and preference settings. The HSS verify the conditions; query the records of the subscriber and services supported by the connected FAP.

Finally, the HSS creates a response (a service announcement packet) which includes the services supported by the operator. The service announcement packet includes the all parameters specified in the R1 packet. In the service announcement packet, the REG INFO parameter is mandatory containing the services provided by the core network. In addition to that it contains R1 packet parameters which allow the gateway to continue the base exchange. Thus, the mobile node can perform the service registration directly with I2 packet. The HIP REG INFO parameter in service announcement certainly contains the services provided by the operator. Other than that, it contains R1 parameters such as, SeGW's HIT, mobile node's HIT, cryptographic puzzle and SeGW's public key. The R1 parameters in the service announcement packet are signed by the SeGW using its public key. Upon receiving the R1 packet, the mobile node solves the puzzle and creates an I2 packet which includes mobile node's HIT,

SeGW's HIT, puzzle solution and mobile node's public key. This message is signed by the mobile node using its public key. The REG REQ parameter in I2 or UPDATE packet deliver the service(s) the mobile node is eligible. If the REG REQ parameter is in an UPDATE packet, the SeGW must not modify the content that are not listed in the parameter. On receiving I2 packet, the SeGW response to the mobile node with a R2 packet that includes SeGW's HIT, mobile node's HIT and few fields such as HMAC and HIP SIGNATURE. The HSS includes an REG RES parameter in its R2 or UPDATE packet only if a registration has successfully completed. By now, the secure association to the core network is established and the mobile node can start the communication. The FAP as a middle box has no mechanism to distinguish the legitimate nodes from the malicious nodes since, they are not aware of the encryption and integrity protection keys associated to the ESP secure association.

Attackers can eavesdrop the Base Exchange and grasp the SPI values of an existing association. Thus, fake ESP packets with valid SPI values can easily traverse through the FAP. For this reason, we propose a node authentication mechanism in to the HIP Base Exchange to enable identity verification of the sending node. This briefly outlines additional security measures for HIP-aware FAPs. There is a high risk of compromising a legitimate FAP by an unauthorized external user. Thus, the FAP may need to verify the identity of the mobile node during the Base Exchange. FAP adds CHALLENGE_REQUEST parameter to R1 message. Thus, the IP and HIP checksum must re-compute once again. This parameter includes an opaque blob of data to the unprotected part of the R1 packet. The opaque data field serves as nonce and puzzle seed value. The content in the CHALLENGE_REQUEST is to be copied unmodified to the CHALLENGE_RESPONSE parameter of the corresponding I2 packet. Otherwise, FAP may deny or degrade the service to the mobile node.

The same identity verification procedure can be applied with the UPDATE or NOTIFY messages as well. Apparently, the FAP can be protected from replay based attacks using this mechanism. After the Base Exchange, the mobile nodes are in a state to communicate. Upon entering to a new domain, the mobile nodes acquire an IP address and depreciate the previous address by sending an UPDATE or NOTIFY messages. The HIP associations can be refreshed by the UPDATE procedure when it is required. In the update message, the mobile node sends set of parameters to the SeGW including the LOCATOR parameter which contains the new IP address (es). However, before updating the association, the SeGW verifies the source address sending an ECHO_REQ which requests to echo back some nonce information. The peer can communicate with unverified address only for a short period of time since; it is controlled by the credit-based authorization. The same figure presents re-association or the HIP based update procedure. During the update, there is a possibility of an attacker attempting to impersonate the mobile node or the FAP. Thus, we recommend using the proposed challenge, requesting based identity verification in femtocell technology.

HIP-Based Secure Multi-homed Femtocells

In this section, we discuss the signaling flow of multi-homed FAPs. In other words, this is an attempt to address the network mobility scenario. For instances, in certain scenarios the mobile nodes do not move alone, but, as a part of a small network. Buses, trains, airplanes and Personal Area Networks (PANs) are few examples of network mobility scenarios. In other words, they can be assumed as mobile femtocells. The mobile nodes change their topological location with the FAP. Entering to a new domain, the FAP renew the IP configuration. And it updates the connected peers, associated LRVS, SeGW and DNS with the UPDATE_PROXY message. Afterwards, the previous set of locators can be depreciated. However, this update can be distinguished from an end-to-end update by the special message type UPDATE_PROXY. In processing perspective, the UPDATE_PROXY exchange is handled similar to the UPDATE exchange. In the figure below, we present the discussed network mobility scenario.

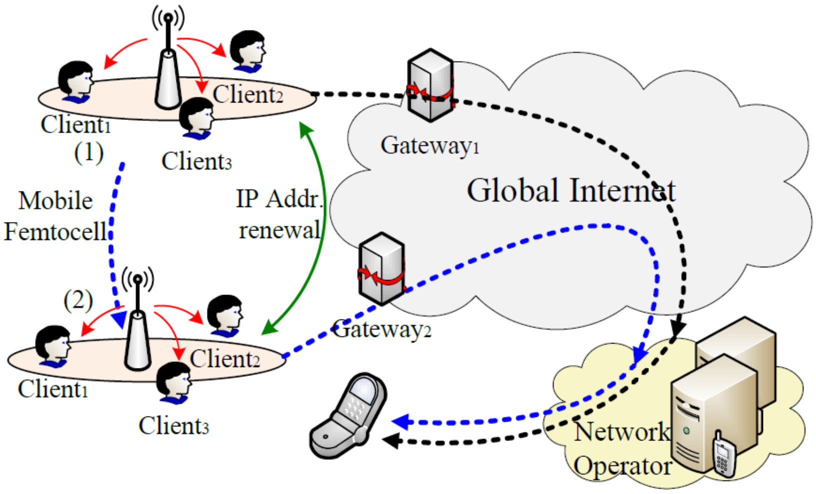


Figure 16. Mobile femtocell scenario.

It is impractical; the mobile nodes change their IP configuration each time the FAP update the location. Thus, it is possible to use rewriting mechanisms to rewrite prefixes in the packet headers when they pass by FAP. Thus, the nodes in the mobile femtocell can be configured using link-local subnet prefixes or unique local sub-network prefixes. The FAP rewrites it with globally routable prefixes before the packets are forwarded to upstream. In the nested network case, a FAP moves behind another FAP. Upon changing the attachment to the FAP, the mobile node trigger the UPDATE PROXY exchange to inform the associated peers, domain LRVS, SeGW and the DNS. On receiving packets at the FAP, it rewrites the packet header with a globally routable address. However, this approach does not allow the FAP to signal on behalf of the mobile node. Moving into a new network, it is recommended to trigger Base Exchange and generate new keying materials to prevent node impersonation.

7.3 Main results

In the existing femtocell architecture, the FAP establishes a secure IPSec tunnel between the SeGW and the FAP utilizing IKEv2. The Figure 17 presents the generic packet format in the femtocell communication and the HIP control and data packet formats. The local IP is the transport IP received from the access point, DHCP or any available mechanism after successful authentication. IPSec adds a tunnel header which is used to establish a IPSec tunnel utilizing IKEv2. The gateway inserts the remote IP address of the access point in the configuration payload during the IKEv2 exchange and establishes a secure IPSec tunnel with this address.

Data packets are transmitted encapsulating into UDP frames with destination port address set to 500 or 4500. SeGW allows the packets with source address set to FAP’s local IP address. Mobile nodes need to perform authentication twice when it is attached to a new service through new FAP. The below action points summarize the typical 3GPP standardized communication flow consequently.

- FAP authentication and authorization.
- Get the Local IP of the access point.
- Query DNS to obtain the gateway address.
- IPSec tunnel establishment.
    - Initiate the establishment of IPSec secure association with IKEV2.
    - Get the Remote IP from the IKEv2 configure payload field.
    - IPSec tunnel establishment.
    - IP in IP tunnel establishment over local IP.

The below action points summarize call flow procedure of the proposed protocol architecture consequently. However, this is already explained in more detail in the previous sections.

- FAP authentication and authorization.
- Query DNS to obtain the mobile node’s associated LRVS address.
- Initiate Base Exchange and service registration.
- IPSec tunnel establishment, service registration and identity verification.
    - Initiate the establishment of secure association by sending an I1 packet to the peer node.
    - Exchange the common keying material and generate session key.
    - Follow-up Base Exchange with service registration and identity verification.
    - ESP IPSec tunnel establishment.

Considering the device authentication (FAP authentication to core network) procedure defined in the 3GPP release 8, it was found that the EAP-AKA based authentication spends minimum 4 Round Trip Times (RTTs) between the FAP and the SeGW whereas, the certificate based authentication spends minimum 2 RTTs. Conversely, our approach spends same number of RTTs as certificate based authentication. Thus, compared to EAP-AKA our approach performs much better. Figure below, presents the control and data packet of 3GPP and HIP based femtocell solution.

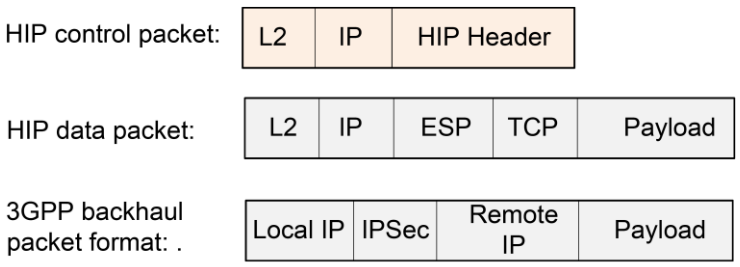


Figure 17.Proposed control/data packet header format and the 3GPP backhaul packet format.

The I1 packet is shown in the Figure 17: Proposed control/data packet header format and the 3GPP backhaul packet format. Base Exchange essentially passes through the LRVS. But, the remaining control packets of the Base Exchange bypass LRVS and establish end-to-end secure associations. After adding a new ESP header field, the data packets are provided confidentiality, data origin authentication, connectionless integrity, anti replay service and limited traffic flow confidentiality. Furthermore, the mobile node does not need to authenticate again and again even if, it reconfigure the association. And the same keying material can be used to encrypt the new association whereas, EAP-AKA, certificate-based or combined certificate-based authentication needs re-authentication.

7.4 Further work

HIP Enabled WLAN Femtocells

Nowadays many public and private places have wireless access which is standardized and matured over many years. The dual-mode mobile handsets empowered with technology convergence guarantees service continuity over different technology regions. To simplify this scenario, we can think of a mobile user who is entering to his home WLAN. When the mobile node has cellular and WLAN coverage, the user may prefer to use the WLAN since, it is cost effective and provides good coverage in home environment. Herein, we propose HIP to handle handover between different technologies. The mobile nodes discover FAP by the router advertisements and dynamically configure an IP address on its wireless interface. Thus, the previous address is depreciated.

However, the change in IP address does not affect the transport layer associations since; they are purely built on the HITs. The mobile node uses update exchange to inform the address reconfiguration to the peer nodes, LRVS, DNS and the SeGW. Upon completing the update exchange, the SeGW rewrite the packet header with the new address. Thus, the re-association does not affect the applications above. In mobile femtocell scenario, the FAP and the SeGW rewrite the packet header before it is forwarded. Otherwise, the packets are decrypted by the SeGW and forwarded upstream over the core network IPSec tunnels. After moving to the home WLAN, the same keying material can be used since, it is shared only between the mobile node and SeGW. If the keying materials are expired, the mobile node has to renew the association. Thus, the same signaling flow in the Figure 16 is applicable to the handover between cellular and WLAN networks assuming the FAP2 is Wi-Fi enabled.

7.5 Conclusions

In this work, we propose a modification to the existing protocol stack of the 3GPP femtocell architecture. We are more focused into mobility and security issues related to femtocell technology. This research work proposes several enhancements to the femtocell technology such as service registration, identity verification and node multihoming. Moreover, we could bring down the device authentication to 2 RTTs whereas; EAP-AKA spends 4 RTTs. Our proposal substantially improves the security by means of strong authentication and identity verification. Other than that, the protocol resists to DoS and Man-in-the-Middle attack by nature. The data is encapsulated into ESP packets to guarantee confidentiality, data origin authentication, connectionless integrity, anti-replay service and limited traffic flow confidentiality. In a nutshell, integrating all features above, this proposal can provide strong security and mobility support for femtocell networks.

7.6 List of publications

Suneth Namal, Andrei Gurtov, Mehdi Bennis, “Securing the backhaul for mobile and multi-homed femtocells”, Future Network & Mobile Summit (FutureNetw), 2011

8. Multiparty Overlay for PMIPv6 domains

8.1 Introduction

Supporting bandwidth-intensive applications for a large group of users in LTE-A like multimedia streaming requires an adapted transport service. The support of Multicast in LTE-A has been addressed through the 3GPP MBMS specification [MBMS2011]. Furthermore, PMIPv6 protocol, as an enabler for supporting mobile users in LTE-A, has been extended by the IETF to support multicasting. Although not officially adopted by 3GPP, multicast-PMIPv6 [RFC 6224] might be a good candidate for being integrated in LTE-A in addition/or as an alternative to MBMS/e-MBMS. Yet, so far, no known end-to-end multicast communication service is adopted by the 3G/4G operators because of its intrinsic deployment cost.

To address this problem, a short/mid-term approach will be considered for the MEVICO project. It consists in assuming a partial deployment of multicast (e.g., MBMS/e-MBMS, multicast-PMIPv6) in the LTE-A domain. This partial deployment will be compensated by the definition of an overlay transport service to ensure a full support of group (or multiparty) applications in the LTE-A domain. Furthermore, this service will accommodate user mobility considering both multicast-capable network segments and non multicast-capable network segments.

The support of group communications in PMIPv6 domains has already been standardized [RFC 6224]. Yet, such a support comes with a strict assumption requiring that all the MAGs of the PMIPv6 domain are multicast-capable. In the MEVICO architecture, this pre-requisite is not guaranteed in that some MAGs may not support multicast (for some reasons like the intrinsic cost of a short-term deployment of a full multicast infrastructure, a multi-operator scenario, etc.). Given this heterogeneity in terms of multicast capability in the network (in the PMIPv6 network domain, in particular), there is no means to setup a (common) multicast-based group communication session in the MEVICO architecture. The present work aims at overcoming this shortcoming by enabling end-to-end support of group communications in PMIPv6 networks where IP multicast is partially deployed in the whole PMIPv6 domain.

8.1.1 State-of-the-art

The document [RFC 6224] describes some techniques for deploying multicast listener functions in Proxy Mobile IPv6 domains without modifying mobility and multicast protocol standards. In the proposed solution, the LMA implements the function of the designated multicast router, MLD querier, and optionally an MLD proxy. According to MLD reports received from a MAG (on behalf of the MNs), the LMA manages multicast forwarding states at its corresponding downstream tunnel interfaces. Also, the MAG performs MLD proxy functions. On the other hand, the MN is only multicast-capable (group join/leave operations, as a standard receiver), and it does not support mobility operations.

The solution assumes that the MN performs standard multicast subscription operations, where the associated messages (MLD Report/Done) arrive at the MAG, which maintains group memberships. A MAG that does not support multicast operations will discard MN’s subscription message. Also, the MAG-to-LMA tunnel will contain all downstream links to MNs that share this specific LMA. In addition, MLD Report messages are aggregated by the MAG (as per [6224]) and then forwarded up the tunnel interface to the MN’s corresponding LMA. The traffic of the subscribed groups will arrive at the LMA, and the LMA will forward this traffic according to its group/source states.

Upon a handover, the MN does not send unsolicited MLD reports. Instead, if the MAG notices a new MN on a downstream access link, the MAG sends a MLD General Query. If it receives a report from the said MN, the MAG will processes it according to the proxy function (i.e., the report will be either ignored or accepted for further processing such as update states, and reports upstream if necessary).

This solution does not address the case where there is a partial deployment of multicast in the PMIPv6 domain (e.g., scenarios where some MAGs of the PMIPv6 domain do not support multicast operations). Also, other approaches like [Saada2012] [Hui2012] seek into optimizing mobility of multicast receivers in terms of handover latency and tunneling overhead at the MAG, yet these approaches do not solve the problem of partial deployment of multicast in the PMIPv6 domain.

8.2 Proposed Solution

The following figure gives an overview of the different PMIPv6 components involved in the multiparty overlay.

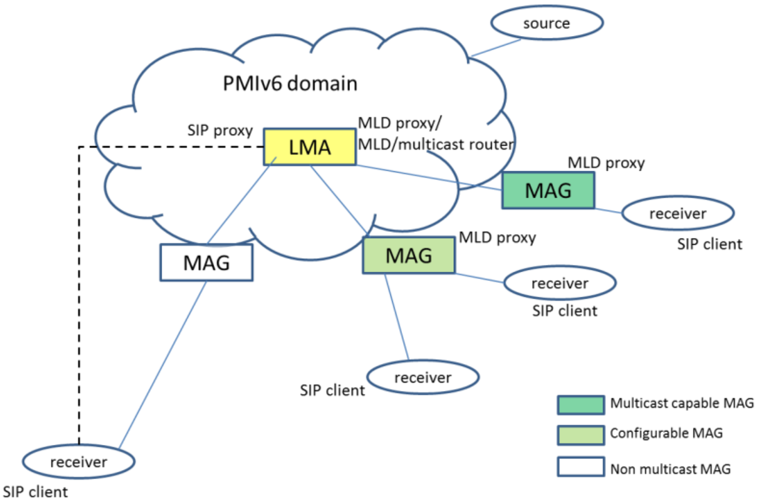


Figure 18. PMIPv6 components for the Multiparty Overlay

8.2.1 Assumptions

The solution leverages SIP protocol [RFC 3261] to enable the LMA storing information about active multicast groups. Therefore, it is assumed that a SIP framework is deployed over the PMIPv6 domain. In this framework the LMA plays the role of a SIP proxy. This proxy is of a tasteful type so as it can maintain states of active multicast groups and associated MNs. Also, the SIP server is present in the SIP framework, yet it will not be discussed in this document because it is not directly involved in the procedures described in this document.

In addition, the MN is assumed to perform standard PMIPv6 operations for network attachment and handover ([RFC 5213]) (cf. figure below).

```
 +------+               +------+               +------+
 | MN   |               | MAG  |               | LMA  |
 +------+               +------+               +------+
    |                      |                      |
MN Attached                |                      |
    |                      |                      |
    |       MN Attached Event from MN/Network     |
    |        (Acquire MN-Id and Profile)          |
    |                      |                      |
    |--- Rtr Sol --------->|                      |
    |                      |                      |
    |                      |--- PBU ------------->|
    |                      |                      |
    |                      |                  Accept PBU
    |                      | (Allocate MN-HNP(s), Setup BCE
    |                      |                   and Tunnel)
    |                      |                      |
    |                      |<------------- PBA ---|
    |                      |                      |
    |                 Accept PBA                  |
    |          (Set Up Tunnel and Routing)        |
    |                      |                      |
    |                      |==== Bi-Dir Tunnel ===|
    |                      |                      |
    |<--------- Rtr Adv ---|                      |
    |                      |                      |
 IP Address                |                      |
Configuration              |                      |
    |                      |                      |
```

Figure 19.Standard Mobile Node Attachment - Signaling Call Flow (copied form [RFC 5213])

Also, the MN is assumed to be a multicast-capable node (multicast receiver) as well as a SIP client. However, the MN is not aware of any possible multicast capability on the MAG. This avoids the modification of Neighbor Discovery protocol [RFC 4861].

Also, the LMA is assumed to be a multicast anchor point in that it implements the operations defined in RFC 6224 (MLD Querier, optinal MLD proxy, and multicast router). The LMA also manages a multicast subscription list (MSP) that has the following structure:

<mcast @>, <MN-ID>, <MN @>, <MAG @>, <MAG cap>

There may be multiple MAGs in the PMIPv6 domain. Some of these MAGs are multicast-capable whereas others are standard PMIPv6 MAGs. Also, some of these standard PMIPv6 MAGs may be configured with a mapping that associates a set of UDP port numbers to a set of multicast addresses. In view of that, three types of MAG are to be considered: non multicast capable MAG, configurable MAG, and multicast-capable MAG.

- *Non multicast capable MAG (NM):* in this option the MAG supports conventional PMIPv6 operations only, as pet RFC 5213.
- *Configurable MAG (CM):* a MAG that supports conventional PMIPv6 operations and stored a structure that maps an input port number to a multicast address. In addition, it is capable of forwarding multicast packets to the MNs attached to its network. The MAG’s input port number is known in advance by both the MAG and LMA.
- *Multicast capable MAG (MM):* a MAG that supports multicast operations as indicated in RFC 6224.

Also, in this solution it is assumed that the LMA knows the capability of each MAG in terms of multicast support. Also, it is assumed that each time the LMA receives a PBU message from a MAG it internally checks for MAG’s capability.

To store the information on the MAG capability in terms of multicast support, the standard PIMv6 binding cache structure may need to be modified so as each MAG address will be associated (e.g., will be pointing) to a MAG capability value NM, CM, or MM. Node Subscription

First, the MN attaches to the PMIPv6 network using the standard procedure defined in RFC 5213 (cf. figure 19). Once, the MN has configured its IP address, it initiates the group subscription procedure, by sending a SIP invite message to the LMA, which acts as a SIP proxy. The SIP invite message includes the multicast address of the desired group as well as the MN-ID.

When the LMA receives the SIP invite message from MN, it stores the MN-ID and multicast address in its multicast subscription list (MSL).

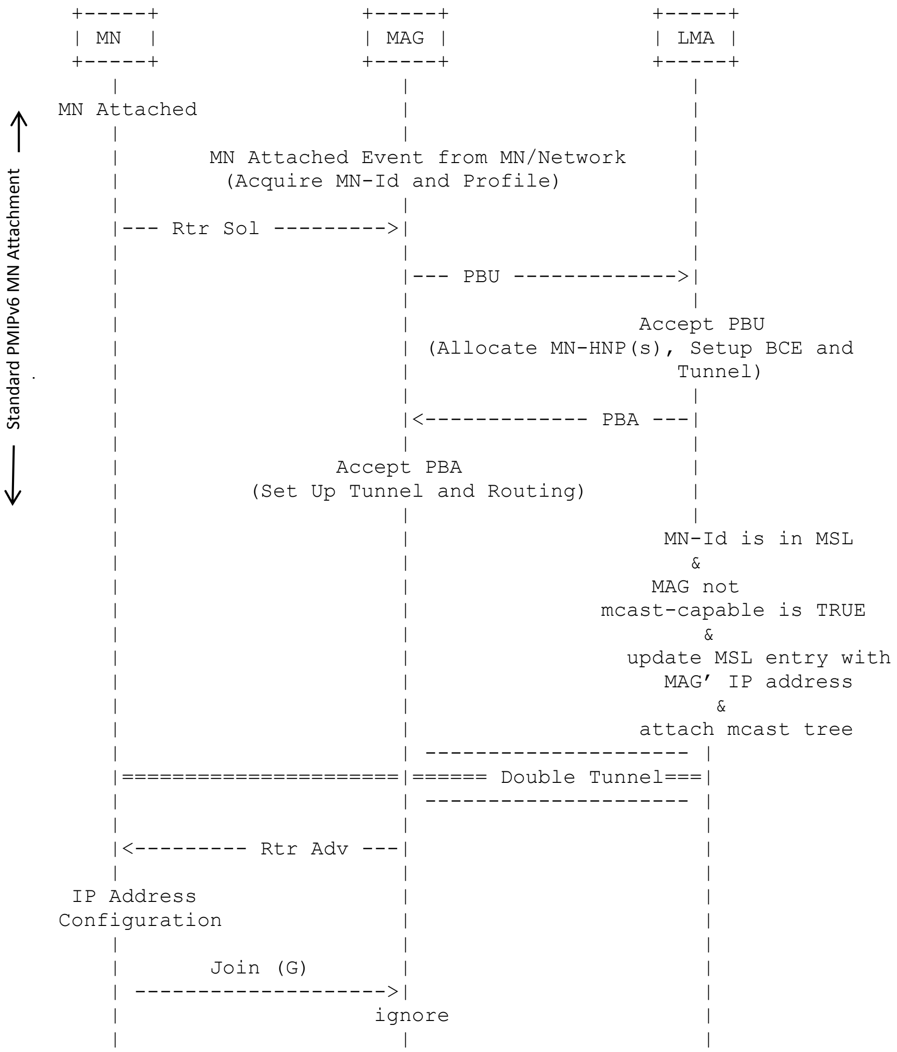
It is worth mentioning that since the LMA holds the couple “<MN-ID>, <MAG @>” in its binding cache as well, the present solution proposes that each time the LMA notices that a new MN has registered through a standard PMIPv6 procedure, it checks whether this MN is registered in the MSL list. If so, the LMA then checks if the said MSL entry includes the MN’s MAG address. If the MAG address does not exist, the LMA will add it in the MSL entry. If MN’s MAG address is already in the MSL list, no further operation is needed in the MSL entry.

In addition to its registration with the LMA, the MN sends an (unsolicited) MLD Report message on its link, without caring, though, whether the MAG is multicast-capable or not. These options are discussed in what follows.

8.2.1.1 Option 1 – Non multicast-capable MAG

When the LMA receives a PBU message from a MAG and notices that the said MAG is not multicast-capable, it checks whether the MN-ID included in the PBU message is registered in one of the entries of its multicast subscription list (MSL). If so, the LMA updates the associated MSL entry with the MAG’ IP address, and attaches to the multicast tree on behalf of the MN (cf. figure 20). In addition, the LMA establishes a double IP tunneling towards the MN (besides the LMA’s standard PMIPv6 tunnel). The inner tunnel will end at the MN’s IP address, and the outer tunnel will end at MAG’s IP address. This tunnel is explained in section 4.3. Of course, this double IP tunneling is transparent to the MAG, since the inner header includes MN’s IP address, while the outer tunnel (obviously) terminates at the MAG’s PMIPv6 tunnel interface.

If the MN-ID is not in the MSL list, no multicast-specific operation is needed on the LMA. On the other hand, when the MAG receives MN’s Report, it will ignore it.



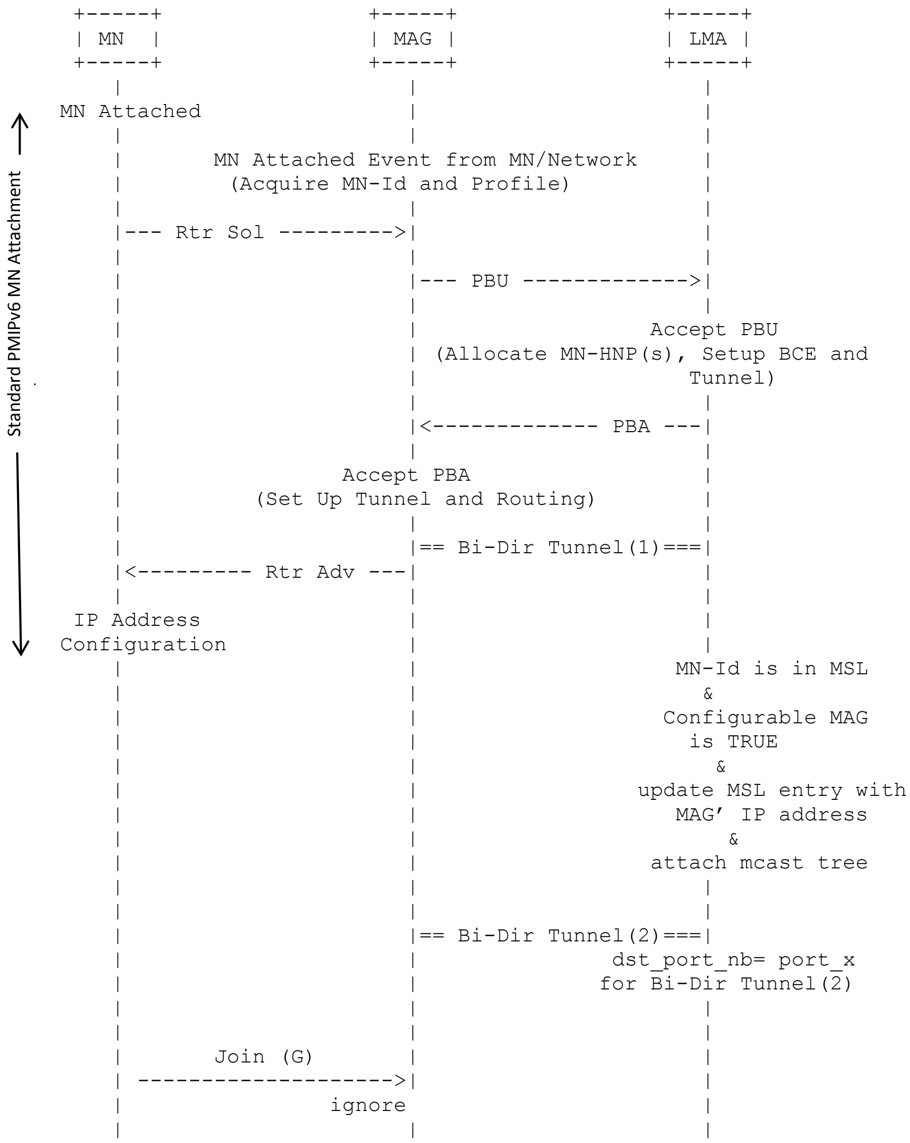
```
 +-----+                 +-----+                +-----+
 | MN  |                 | MAG |                | LMA |
 +-----+                 +-----+                +-----+
     |                      |                       |
MN Attached                 |                       |
     |                      |                       |
     |      MN Attached Event from MN/Network       |
     |        (Acquire MN-Id and Profile)           |
     |                      |                       |
     |--- Rtr Sol --------->|                       |
     |                      |                       |
     |                      |--- PBU -------------->|
     |                      |                       |
     |                      |                  Accept PBU
     |                      | (Allocate MN-HNP(s), Setup BCE and
     |                      |                     Tunnel)
     |                      |                       |
     |                      |<-------------- PBA ---|
     |                      |                       |
     |                 Accept PBA                   |
     |          (Set Up Tunnel and Routing)         |
     |                      |                       |
     |                      |                    MN-Id is in MSL
     |                      |                       &
     |                      |                   MAG not
     |                      |               mcast-capable is TRUE
     |                      |                        &
     |                      |                 update MSL entry with
     |                      |                    MAG’ IP address
     |                      |                         &
     |                      |                  attach mcast tree
     |                      | --------------------- |
     |======================|====== Double Tunnel===|
     |                      | --------------------- |
     |                      |                        |
     |<--------- Rtr Adv ---|                        |
     |                      |                        |
 IP Address                 |                        |
Configuration               |                        |
     |                      |                        |
     |      Join (G)        |                        |
     | -------------------->|                        |
     |                    ignore                     |
     |                      |                        |
```

Figure 20.PMIPv6-based MN’s Group Join Procedure – Case of Non-Multicast Capable MAG

8.2.1.2 Option 2 – Configurable MAG

When the LMA receives a PBU message from a MAG and notices that the said MAG is a configurable MAG, it checks whether the MN-ID included in the PBU message is registered in one of the entries of its multicast subscription list (MSL). If so, the LMA updates the associated MSP entry with the MAG’s IP address, and attaches to the multicast tree on behalf of the MN (cf. figure 21). In addition, the LMA establishes a unicast tunnel towards the MAG (besides the standard PMIPv6 tunnel) and sets a dedicated destination port number to the said tunnel. There may be one tunnel (and thus one port number) between LMA and MAG per multicast address or a common tunnel (and thus a common port number) between them for all the multicast addresses.

If the MN-ID is not in the MSL list, no multicast-specific operation is needed on the LMA. On the other hand, when the MAG receives MN’s MLD Report, it will ignore it.



```
 +-----+                +-----+                +-----+
 | MN  |                | MAG |                | LMA |
 +-----+                +-----+                +-----+
    |                      |                      |
MN Attached                |                      |
    |                      |                      |
    |       MN Attached Event from MN/Network     |
    |        (Acquire MN-Id and Profile)          |
    |                      |                      |
    |--- Rtr Sol --------->|                      |
    |                      |                      |
    |                      |--- PBU ------------->|
    |                      |                      |
    |                      |                  Accept PBU
    |                      | (Allocate MN-HNP(s), Setup BCE and
    |                      |                     Tunnel)
    |                      |                      |
    |                      |<------------- PBA ---|
    |                      |                      |
    |                 Accept PBA                  |
    |          (Set Up Tunnel and Routing)        |
    |                      |                      |
    |                      |== Bi-Dir Tunnel(1)===|
    |<--------- Rtr Adv ---|                      |
    |                      |                      |
 IP Address                |                      |
Configuration              |                      |
    |                      |                     MN-Id is in MSL
    |                      |                       &
    |                      |                    Configurable MAG
    |                      |                     is TRUE
    |                      |                       &
    |                      |                  update MSL entry with
    |                      |                     MAG’ IP address
    |                      |                        &
    |                      |                   attach mcast tree
    |                      |                      |
    |                      |                      |
    |                      |== Bi-Dir Tunnel(2)===|
    |                      |               dst_port_nb= port_x
    |                      |              for Bi-Dir Tunnel(2)
    |                      |                      |
    |                      |                      |
    |       Join (G)       |                      |
    | -------------------->|                      |
    |                ignore                       |
    |                      |                      |
```

Figure 21. PMIPv6-based MN’s Group Join Procedure – Case of Configurable MAG

8.2.1.3 Option 3– Multicast-capable MAG

When the LMA receives a PBU message from a MAG and notices that the said MAG is multicast-capable, it checks whether the MN-ID included in the PBU message is registered in one of the entries of its multicast subscription list (MSL). If so, the LMA updates the associated MSP entry with the MAG’ IP address, and carries on the rest of the MN registration operation according to RFC 6224 (i.e., receive and process aggregated MLD Reports (or Aggregated Join message) from the MAG, attach to the multicast tree on behalf of the MN, establish appropriate tunnel with the MAG).

If the MN-ID is not in the MSL list, no multicast-specific operation is needed on the LMA. On the other hand, when the MAG receives MN’s MLD Report, it will process it according to RFC 6224 specification (MLD Reports aggregation and forwarding to LMA).

8.2.2 Data Forwarding

When the LMA receives a multicast packet, it checks its MSL list for the entries that match the packet’s multicast address. For each matching entry, the LMA checks whether the associated MAG is multicast-capable or not.

8.2.2.1 Option 1 – Non multicast-capable MAG

If the MAG is not multicast-capable, the LMA will utilize a double-tunneling to send the multicast packet towards the MN. The resulting Packet to be forwarded by the LMA to MN via MAG will be as follows:

<---Outer IP hdr----->	<--Inner IP hdr -------->	<---------------------Mcast packet---------------------------->	
Src @= LMA @ Dst @ = MAG @	Src @ = mcast_src @ Dst @ = MN @	Src @ = mcast_src @ Dst @ = mcast @	Data payload

The outer IP header of the tunnel includes MAG's address as a destination. The inner IP header of the tunnel includes MN's IP address so as multicast packet forwarding from LMA to MN via the MAG will be transparent to the said MAG.

When the MAG receives the multicast packet from the LMA via its PMIPv6 tunnel interface, it removes the outer header and forwards the resulting packet to the MN (as per RFC 5213).

8.2.2.2 Option 2 – Configurable MAG

If the MAG is configurable, the LMA will tunnel the multicast packet towards the MAG using the dedicated UDP port number.

When the MAG receives the multicast packet from the LMA via a dedicated input UDP port, it removes the outer header and forwards the resulting (multicast) packet on its downstream interface.

8.2.2.3 Option 3 – Multicast Capable MAG

Data forwarding procedure in option 3 is similar to that of RFC 6224.

8.2.3 Node Handover

MN handover to a new network associated to a new MAG: nMAG is quite similar to that of the MN subscription phase (option 1, option 2 & option 3), excluding the SIP procedure, which of course is not needed for the handover phase. In addition, the LMA in the MN handover phase has to update the associated MSL entry with the nMAG' IP address (this update is done when the LMA checks the PBU's MN-ID in the MSL list). Furthermore, in the handover phase there may be no need to (re)build the multicast path (from the multicast tree) towards the LMA (since the said path is (normally) already built).

8.2.4 Node Unsubscription

When an MN wishes to leave a multicast group, it notifies the LMA via SIP Bye message and sends an MLD Exclude/Done on its link. The SIP message contains the multicast address of the group to be left by MN as well as the MN-ID.

When the LMA receives the SIP Bye message, it checks MN's ID and the multicast address against the MSL entries. If an entry is found, the LMA removes it and updates the multicast routing state accordingly (ex. the LMA removes the multicast forwarding state for the said multicast address if no more MNs are associated to this multicast address).

On the MAG's side, three options are to be distinguished.

8.2.4.1 Option 1 – Non multicast-capable MAG

When a non-multicast capable MAG receives an MLD Exclude/Done message, it ignores it.

8.2.4.2 Option 2 – Configurable MAG

When a configurable MAG receives an MLD Exclude/Done message, it will ignore it.

8.2.4.3 Option 3 – Multicast-Capable MAG

When a multicast-capable MAG receives an MLD Exclude/Done message, the rest of the procedure follows RFC 6224's operations (if necessary, update of the MAG's membership database and transmission of MLD Exclude/Done via the MAG-to-LMA interface towards the LMA, which then terminates multicast forwarding).

8.3 Validation system

System validation is out of the scope of this project.

8.4 Main results

In MEVICO project, a multiparty overlay has been proposed to provide an end-to-end transport service for group communications. The transport service leverages PMIPv6 functional components along with multicast capability of the network. This solution is particularly suitable for a PMIPv6 domain where multicast is partially deployed. Furthermore, this solution would not require any modification of the PMIPv6 or PMIPv6-multicast standards.

8.5 Further work

In a future work, the proposed solution will be implemented and evaluated.

8.6 Conclusion

The solution proposed in this section seeks into providing a good short/mid-term option for the deployment of an end-to-end multiparty transport service in PMIPv6 domains where multicast is partially/not widely deployed. To achieve this, a multiparty transport overlay is defined to cover both multicast-capable and non-multicast capable PMIPv6 domain segments. In addition, PMIPv6 protocol is leveraged to ensure end-to-end data delivery to a group of mobile nodes and provide dynamic management of the multiparty overlay. Also, SIP protocol is used for mobile node registration.

The solution requires a minor modification of PMIPv6 protocol (additional structures and operations on the LMA and, optionally, additional structures and operations on the MAG).

8.7 List of publications

An Internet Draft will be submitted to the IETF

9. Distributed policy control

9.1 Introduction

With the recent huge increase in data traffic, mobile operators are seeking for ways to optimize backhaul resources and to better control their user's traffic, allowing users and services differentiation.

The solution offered by the 3GPP covers QoS requirements for the traffic within the mobile system, that is, between UE and PDN GW.

Over the backhaul, the mobile's Qos may be mapped to IP transport layer Qos, (diffserv) but there are no specific 3GPP mapping recommendations for that.

The 3GPP Policy and Charging framework defines the following procedures:

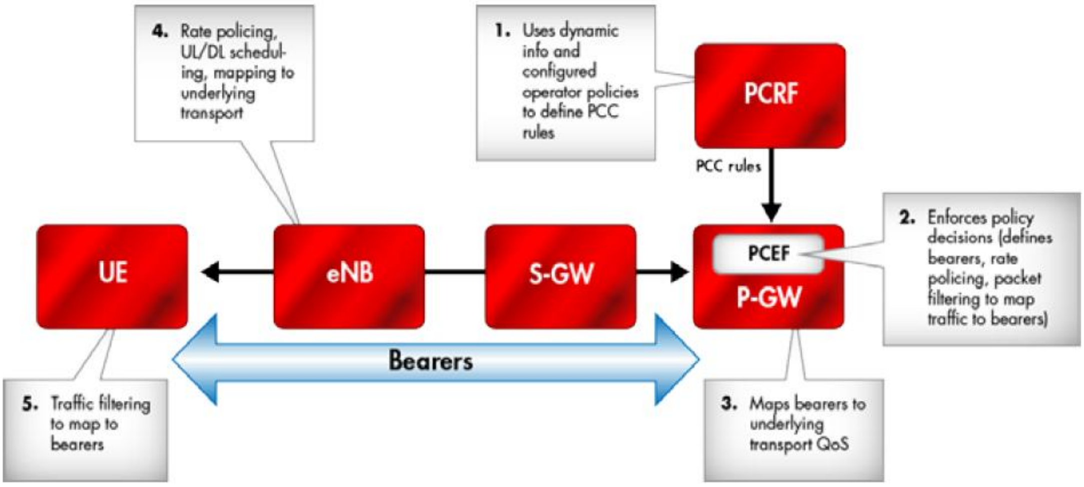

4. Rate policing, UL/DL scheduling, mapping to underlying transport
1. Uses dynamic info and configured operator policies to define PCC rules
PCRF
PCC rules
2. Enforces policy decisions (defines bearers, rate policing, packet filtering to map traffic to bearers)
UE
eNB
S-GW
PCEF
P-GW
Bearers
5. Traffic filtering to map to bearers
3. Maps bearers to underlying transport QoS

Figure 22. PCRF charging procedures.

- The PCRF makes dynamic and real-time decisions on policy rules and delivers it to the P-GW (PCEF)
- The P-GW performs traffic enforcement, such as traffic shaping, rate policing, as well as setting a bearer towards the UE.
- The eNB sets a radio bearer with specific Qos profile received from the P-GW over the air interface
- Between the eNB, S-GW and P-GW, the bearer Qos may be mapped to IP transport layer Qos.

As presented at the above diagram, the backhaul network is not a part of the policy control; moreover, the PCRF is not aware of any congestion in the backhaul network. However the backhaul network is typically engineered with oversubscription, so congestion might occur in the backhaul.

Additional limitations can result because of scalability issues and from the fact that the PCRF is not aware of the dynamic location of a user. This means that in case of congestion, the efficiency of the changes done by the PCRF can be lacking.

The network holds a lot of useful information about the user's quality of experience and by proper feedback could solve dynamically congestion events in the backhaul.

The Proposed Solution: General

To ensure end-to-end dynamic policy control and efficient backhaul bandwidth management, the backhaul equipment can alert the PCRF on possible congestion events before the backhaul network starts dropping packets randomly.

The proposed solution is to add a communication channel between the transport equipment (CSG) and the PCRF for reporting the case of possible congestion in the backhaul network.

This approach allows dynamic distributed policy enforcement, as described below:

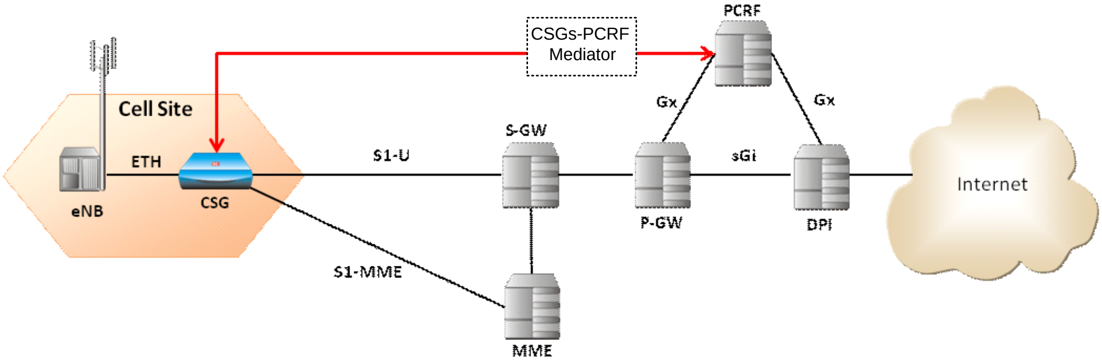


Figure 23. Dynamic distributed policy enforcement

We suggest using a Mediator between the CSGs and the PCRF.
This Mediator seems to be essential in order to control thousands of base stations and CSGs by a centralized PCRF.
The Mediator will do part of the analysis and will forward the results to the PCRF.
The PCRF in return can send updated policy per user per base station.

9.2 Detailed description

Today the policy enforcement functionality designed by 3GPP is typically based on a centralized approach, using the PGW and a dedicated DPI platform.

This approach has several limitations:

1. Scalability: The traffic volume on the Gi interface is expected to be very high and huge processing power is required in order to provide session based policy enforcement, keeping user's QoE.

2. Dynamicity: The P-GW and PCRF are not aware of the backhaul topology and user's dynamic location. Therefore when identifying that a specific user is experiencing poor QoE, modifying the bearer's quality parameters dynamically may not solve the problem efficiently.

A possible solution to overcome these limitations is to allow the backhaul network to report on congestion events and to be able to handle these events locally, following the PCRF's given policies per user/application.

We call it distributed PCEF located in the edges of the network, next to the base stations.

By taking into considerations the affects the transport network has on the end-to-end quality of service, the policy management in the mobile network becomes much more robust, dynamic and scalable.

9.2.1 Closing loop with the PCRF

Providing congestion information from the backhaul network towards the PCRF gives complete E2E awareness to customer's quality of service.

This information includes congestion events in the backhaul, affected users and relevant base stations.

Using this CSG-PCRF communication enables the following improvements that can be added to the mobile's policy control:

- Awareness to backhaul congestion
- Correlation between the transport topology and the users / base stations affected from backhaul congestion.
- Dynamic Qos policies and prioritization during congestion at the access link
- Offloading traffic from the mobile core when possible
- Traffic steering in order to solve the backhaul congestion, always choosing the best route based on the latency and bandwidth measurements done periodically. The steering can be DSCP aware, forwarding the high quality applications towards the low latency route.
- Security – intrusion prevention

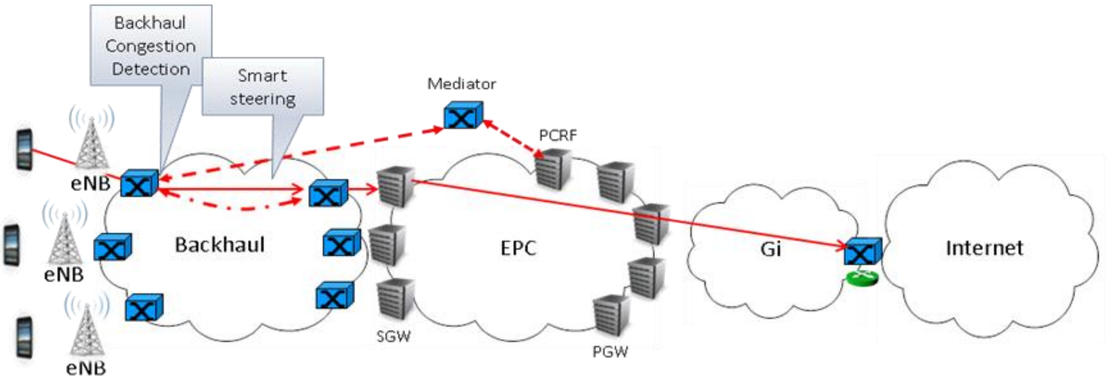


Figure 24. Dynamic Qos policies manages locally or received from the PCRF

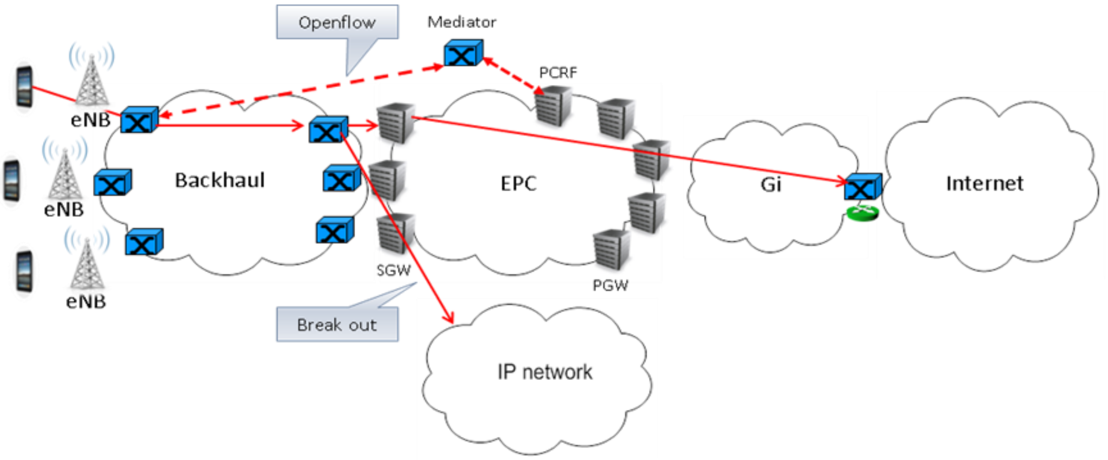


Figure 25. Offloading traffic when instructed to do so by the PCRF

CSGs-PCRF Mediator functionality

The mediator aggregates many CSGs and transfer reports from the CSGs (after analysis) to the PCRF. Since the PCRF is not aware of transport's topology nor enodeb or S-GW, the mediator should indicate to the PCRF who are the affected users, from specific backhaul congestion.

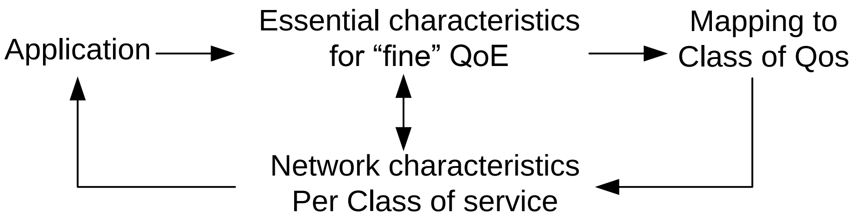
The suggested interface between the mediator and the PCRF is based on diameter protocol.

The protocol between the mediator and the CSGs can be based on openflow or SNMP, allowing changes in bandwidth profiles (open flow doesn't support traffic engineering for now), intrusion prevention and an ability to offload specific traffic from the mobile core.

9.2.2 Standards, Models and algorithms

9.2.2.1 Applications and their essential characteristics for "fine" QoE

In this section we would like to connect the different topics raised in previous sections according to the following diagram:


Application
Essential characteristics for "fine" QoE
Mapping to Class of Qos
Network characteristics Per Class of service

QoE is subjective and refers to personal feeling, yet each mobile operator will probably define what good QoE means for their users.

In order to guarantee end to end QoE, there should be a way to map the applications to classes of service and this way manage and monitor end to end the resources per class of service in the network.

Let's focus on the connection between Application and traffic characteristics:

Voice application For example, expects to experience essentially no packet loss and a minimal but fixed amount of packet delay. The best-effort IP network provides almost exactly the opposite performance required by the voice application, (varying amounts of packet loss and variable delay typically caused by network congested nodes).

Therefore, Qos management plays a critical role to ensure that diverse applications can be properly supported in a multiservice IP network.

The table below shows different application categories and their sensitivity to performance parameters:

Application	Bandwidth	Sensitive to		
		Loss	Delay	Delay variation
IP Telephony	Low	Med	High	High
Video Conferencing	High	Med	High	High
Streaming media	Low-High	Med	Med	Low
Client / Server Transactions	Low	High *	Med	Low
Email (store/forward)	Low	High *	Low	Low
Best Effort Traffic	Low-Med	Low	Low	Low

*Highly loss sensitive applications, runs over TCP that manages retransmissions.

When examining real time gaming and other interactive applications, there is a clear relationship between game session time and network QoS performance parameters.

A gamer will enjoy a longer game session when having a good quality of experience and available resources end to end.

The relations between game play time and Qos parameters are shown in the following diagrams:

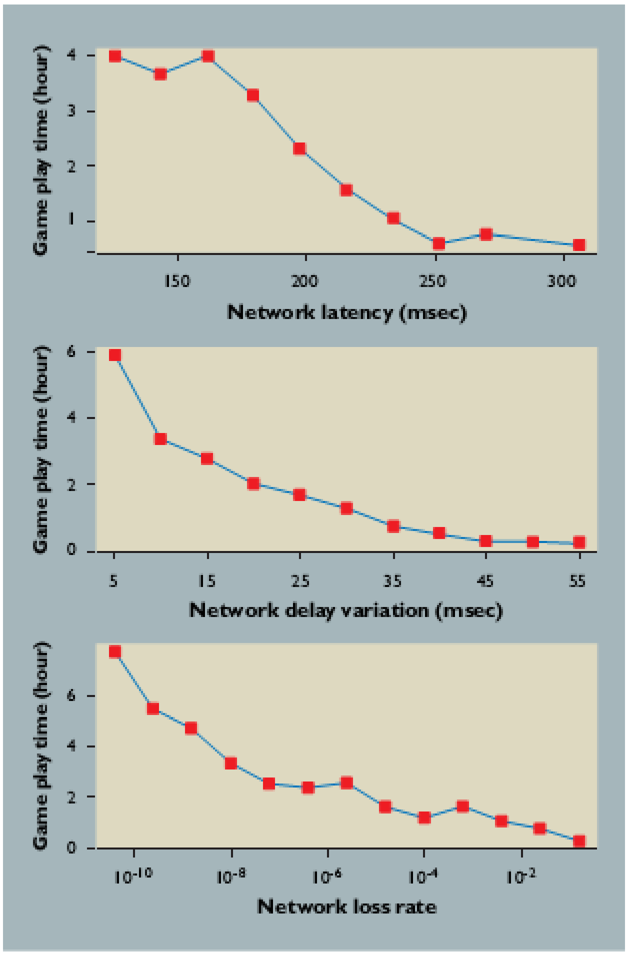


Figure 26. relations between game play time and Qos parameters

An operator can decide what QoE he would like to guarantee a user that is paying for a real time service.
Does two hours of on-line gaming will be sufficient?
How will it be reflected to network's resources?

A work has been done in the main standardization bodies to map applications to classes of service and to characterize the traffic of the different classes of service in order to assure QoE

Taken from the NGMN:

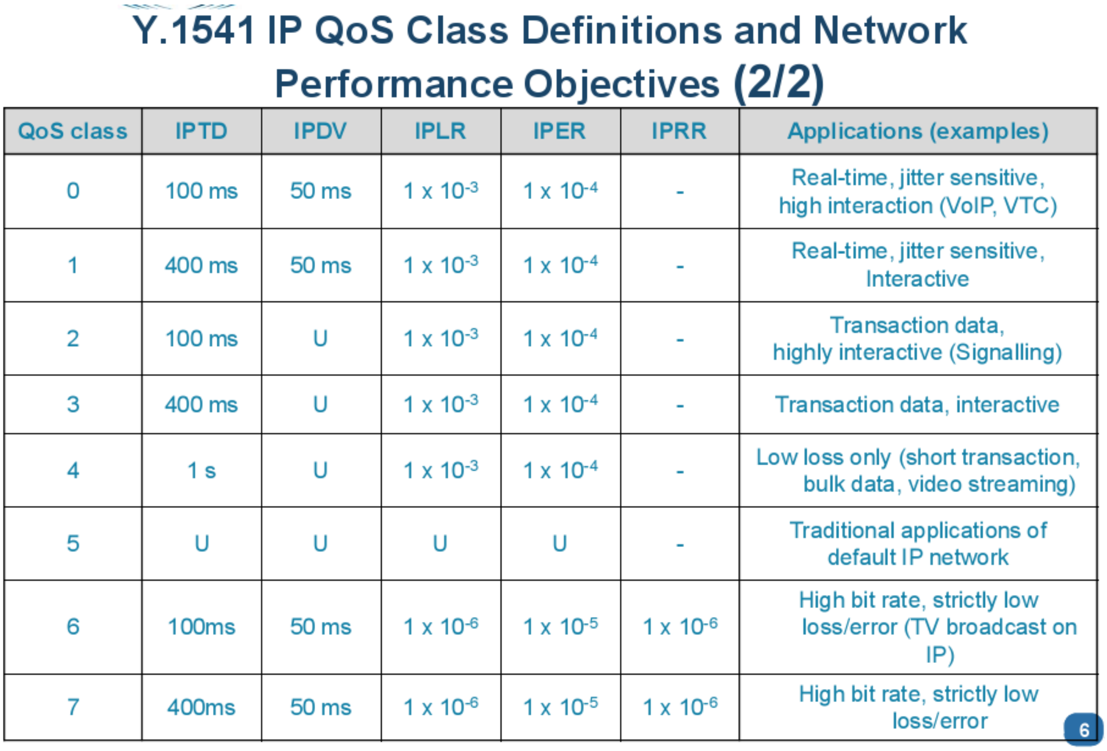
C1	Voice, Real-Time Gaming, Synchronization and Control Plane
C2	2G Data (EDGE) and Real-Time Video
C3	Premium Data (buffered Video, non-GBR Real-Time)

■	Everything Else

CLASS OF SERVICE	TOLERANCE		
	LOSS	DELAY	DELAY VARIANCE
	(10-x)	(ms)	(ms or L/M/H)
C1	10^{-6}	50	Very Low
C2	10^{-6}	100	Low
C3	10^{-6}	100	Moderate
■	-	-	-

Figure 27. NGMN – Target Performance Indicators for 4-CoS Classification Scheme

Taken from the ITU-T:



Y.1541 IP QoS Class Definitions and Network Performance Objectives (2/2)

QoS class	IPTD	IPDV	IPLR	IPER	IPRR	Applications (examples)
0	100 ms	50 ms	1×10^{-3}	1×10^{-4}	-	Real-time, jitter sensitive, high interaction (VoIP, VTC)
1	400 ms	50 ms	1×10^{-3}	1×10^{-4}	-	Real-time, jitter sensitive, Interactive
2	100 ms	U	1×10^{-3}	1×10^{-4}	-	Transaction data, highly interactive (Signalling)
3	400 ms	U	1×10^{-3}	1×10^{-4}	-	Transaction data, interactive
4	1 s	U	1×10^{-3}	1×10^{-4}	-	Low loss only (short transaction, bulk data, video streaming)
5	U	U	U	U	-	Traditional applications of default IP network
6	100ms	50 ms	1×10^{-6}	1×10^{-5}	1×10^{-6}	High bit rate, strictly low loss/error (TV broadcast on IP)
7	400ms	50 ms	1×10^{-6}	1×10^{-5}	1×10^{-6}	High bit rate, strictly low loss/error

9.2.2.2 3GPP's Applications and their essential characteristics for "fine" QoE

According to 3GPP , mobile operators should logically divide their network into ‘pipes’ called bearers, offering different service performance characteristics (QCI, ARP and GBR) to different applications. The different options are shown in the following table:

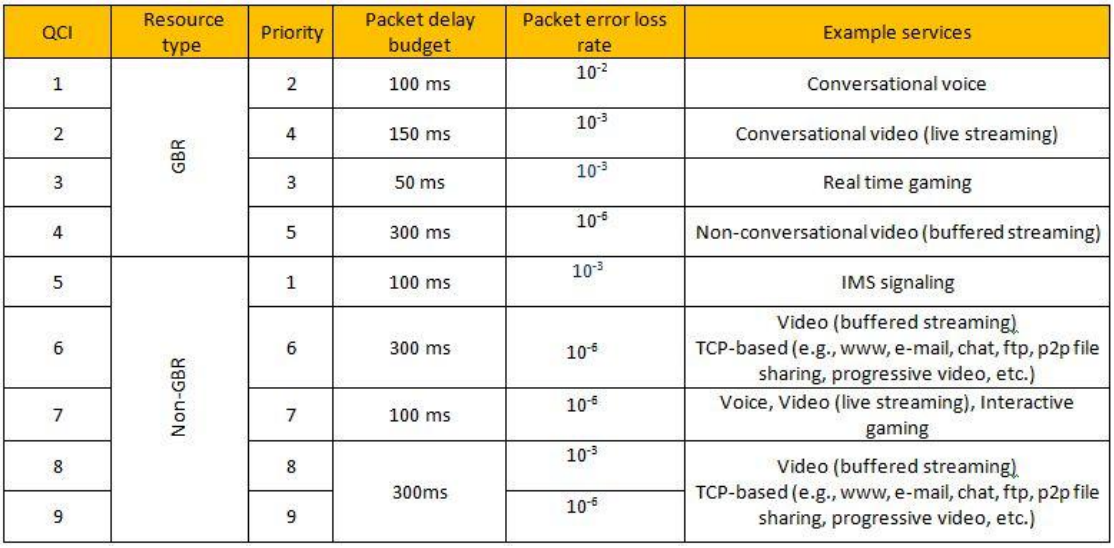
According to the 3GPP standards,

Each bearer is associated with the following bearer level QoS parameters:

– Qos Class Identifier (QCI);

– Allocation and Retention Priority (ARP).

In case a bearer is a GBR (guaranteed bit rate) bearer, it is additionally associated with the following bearer level QoS Parameters:



QCI	Resource type	Priority	Packet delay budget	Packet error loss rate	Example services
1	GBR	2	100 ms	10^{-2}	Conversational voice
2		4	150 ms	10^{-3}	Conversational video (live streaming)
3		3	50 ms	10^{-3}	Real time gaming
4		5	300 ms	10^{-6}	Non-conversational video (buffered streaming)
5	Non-GBR	1	100 ms	10^{-3}	IMS signaling
6		6	300 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	10^{-6}	Voice, Video (live streaming), Interactive gaming
8		8	300ms	10^{-3}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9		10^{-6}	

–Guaranteed Bit Rate (GBR);

–Maximum Bit Rate (MBR).

•Each APN is associated with an Aggregate Maximum Bit Rate (APN AMBR).

•Each UE is associated with UE Aggregate Maximum Bit Rate (UE AMBR).

Nine QCI values are defined in the Release 8 version of the 3GPP.

The ARP parameter is stored in the Subscriber profile (HSS) typically on a per APN basis

- Priority level: 1 – 15
- Pre emption capability: determines whether a bearer with a lower ARP priority level should be dropped to free up the required resources
- Pre emption vulnerability: determines whether a bearer is applicable for dropping by a preemption capable bearer with a higher ARP priority value

However, as mentioned before, the underlying mobile backhaul transport layer is out of scope for 3GPP in terms of Qos. This can result in service degradation because actually, in real aggregation networks, bottlenecks are an integral characteristic of the packet networks and packets will be dropped randomly disturbing running services and users.

9.2.2.3 Congestion detection model

What do we define as congestion event?

Congestion of a network is said to occur if packets are dropped or if delay increases significantly.

Network operators tend to define congestion in terms of the load on a network over a particular period of time: Aggregated volume of traffic or latency exceeding some thresholds over a period of time.

The growth in rich-media services makes the congestion management much more complex and Network should be managed according to the traffic it is carrying.

Traffic prioritization is done today by the operator using different techniques such as DPI. Link capacity provisioning is based on different policies for each service/application and user.

Our Model

We base our detection model on the following assumptions:

- Qos and QoE are mutually dependent and to achieve QoE, QoS is the basic building.

- Different services with different characteristics can be distinguished based on the DSCP marking, in order to allow end to end QoS.

- Our model's target is to identify congestion events and improve affected user's QoE in a way that will not harm most of the services running in parallel.

As stated before the offered measuring tool is Eth OAM Y.1731 performance monitoring (real time loss, delay and delay variation measurements).

Our goal is to determine if this mechanism is efficient in detection of backhaul congestion.

Remark: Time that passes till congestion detection should also be minimized, in order to detect congestion while it is happening.

Steps for achieving congestion detection:

1. Define performance parameters that are of interest (delay, loss ,delay variation)
2. Define time dependent thresholds per performance parameter per each class of service.
3. Perform real time analysis per class of service in order to detect congestion.

9.2.2.4 Available bandwidth algorithm - MoSeab

Taking into account available bandwidth measurement can improve the behavior of applications running on the best-effort network; this measurement indicates how much spare bandwidth can actually be used by an application and can help detect congestion.

In this research, we chose to be based on ABW (available bandwidth) measurement technique named MoSeab. The technique determines the ABW using active measurements i.e. injecting a sequence of probing packets at the sender and estimating ABW by analyzing them at the receiver.

The size of packet and the rate are set by the sender.

MoSeab uses "packet train", where the delay between packets is constant and set by the sender

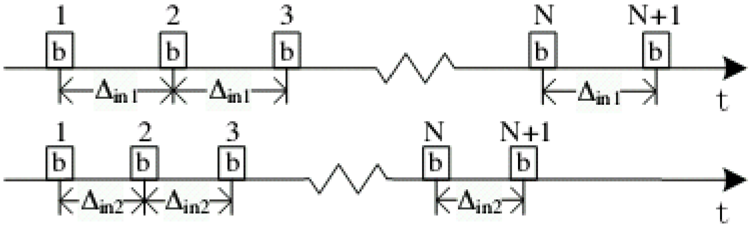


Figure 28. Packet train.

One way delay (OWD) measurements are done at the receiver.

By further analysis we can calculate the changing rate of the relative OWD.

MoSeab has proven valid even when there are multiple bottlenecks.

9.2.2.5 Eth OAM

Service providers get the ability to fully monitor a customer's end-to-end Ethernet service by using Y.1731/802.1ag Service OAM.

This standard supports an impressive array of OAM messages, including CC, LB, Link Trace (LT), AIS, RDI, Lock Signal (LCK), Test Signal (Test), Automatic Protection Switching (APS), Loss Measurement (LM) and Delay Measurement (DM) for performance monitoring.

Since Y.1731 has been finalized while 802.1ag is still in progress, we shall mainly use ITU terminology here. Y.1731 defines a maintenance entity (ME) that requires management.

In order to capture the multipoint-to-multipoint nature of Ethernet, MEs are grouped into ME groups (MEGs, referred to as Maintenance Associations or MAs in IEEE language).

Each MEG is given a unique ID, and OAM messages specify the MEG ID for which the message belongs to.

At the ends of managed entities we find MEG End Points (MEPs), which are the functions that generate and process OAM, frames to monitor and maintain the ME.

It is the responsibility of the MEP to prevent OAM messages from leaking out of the administrative domain to which they belong, or entering another domain. However, MEPs transparently pass OAM frames from other domains when they belong to a higher OAM level, thus enabling end-to-end management of customer connectivity.

For the purpose of measuring the QoS level per class of service the offered measuring tool is Eth OAM Y.1731 Performance monitoring (loss, delay and delay variation).

Eth OAM offers real time loss and synthetic frame loss measurements in case the network topology is not point to point. The transmission period in case of synthetic frame loss is 100msec (10 frames per second) and therefore we can assume that if real time loss measurements are not available, the eth OAM based congestion detection will be limited and be based mainly on delay and delay variation.

Obviously congestion events shorter than 200msec can be missed using this measurement tool, but it might be that these short congestion events could be ignored since these events are solved with no intervention and their effect on the user are probably minor but there is no public research we are aware of to document this statement.

For better congestion resolution, one way to go could be to sending more frequent OAM packets and increase the PM resolution. Another way is to add our own proprietary real time performance monitoring mechanism based on TCP.

9.2.3 The Proposed Functional Architecture

The CSG (cell site gateway) will include a policy enforcement function and a congestion detection function.

1. *The congestion detection function:*

   Basic configuration:

   Absolute thresholds configured, per class of service (DSCP value) for each performance parameter.

   The solution offered doesn’t change the prioritization of the different services; this is assumed to be taken care of by the Mobile operator’s network (PCRF and DPI functions).

   Time window configured per class of service (DSCP value) for each performance parameter.

   The time window is the time in which the performance parameters are analyzed and will be set by the operator according to his policies per service.

   Detection will be done based on real-time analysis of the following performance parameters:

   - Delay
   - Delay variation

- Packet loss ratio
- Available bandwidth on the available links (per backhaul class of service connection)
- Connectivity checks for both upstream and downstream

These measurements received from Y.1731/802.1ag service OAM and additional algorithms allow finding the current QoS level. By comparing the QoS level to the expected thresholds configured, congestion can be detected

Affected users will be the ones that experience violation of different threshold's over a configured time window per class of service.

Reports:

This function will generate a report which includes:

- Users (IP addresses) that are affected by backhaul congestion
- Congestion level (analysis of measured performance parameters)
- Base station ID

2. *The backhaul enforcement function:*

   The enforcement function, if instructed to, can act in the following ways:

   - Traffic steering: Transfer part of the traffic (can be DSCP based) through an additional route with better characteristics in terms of latency and bandwidth to ensure QoE consistency.
   - Offloading specific traffic from the mobile core (encapsulation)
   - Dropping specific flows (security issues)
   - Controlling the bandwidth of low priority flows by increasing / decreasing shaper's rate.
   - Increasing the traffic's priority in case there is available bandwidth end to end and we are allowed to change the traffic's priority.
   - Change color marking for prioritizing specific services without damaging the original prioritization of the mobile operator.

9.3 Validation system

First step in validating this approach is to simulate major building block such as the available bandwidth.

The simulations will be implemented in NS2.

The second step is to test the congestion detection based on Eth OAM (actions taken in case of detected congestion remains for further study).

We'll be using a CSG (ETX-203) and RAD's management system as a mediator to the PCRF, (based on SNMP).

Based on Eth-OAM measurements (Y.1731) done per class of service in the ETX, RAD's management system will be able to report on congestion events.

For further study:

RAD's management system will identify congestion events per class of service and will be able to react to these events, as if these events were reported to the PCRF and as a result, the PCRF updated policy.

Remark:

The solution should not change the mobile's policies without a specific instruction from PCRF so re-editing class of service is not an option here.

On following pages you'll find the test diagram description in which traffic with different classes of service will go through a network emulator, and based on Eth OAM the management system will detect congestion per class of service.

The Network emulator used here is paragon, which is able to filter different flows and add impairments only to the filtered flows.

In this case different classes of service will be differentiated using different vlans. Network impairments will be inserted based on G.1050 - Network model for evaluating multimedia transmission performance over IP.

Traffic will be generated using Ixia and impairments such as delay and packet loss inserted by the Paragon will be also monitored by Ixia using latency report.

Three different applications, represented by different QCI values were chosen for the test.

Thresholds per application for sufficient QoE were selected after correlating between the different standardization bodies mentioned in previous sections.

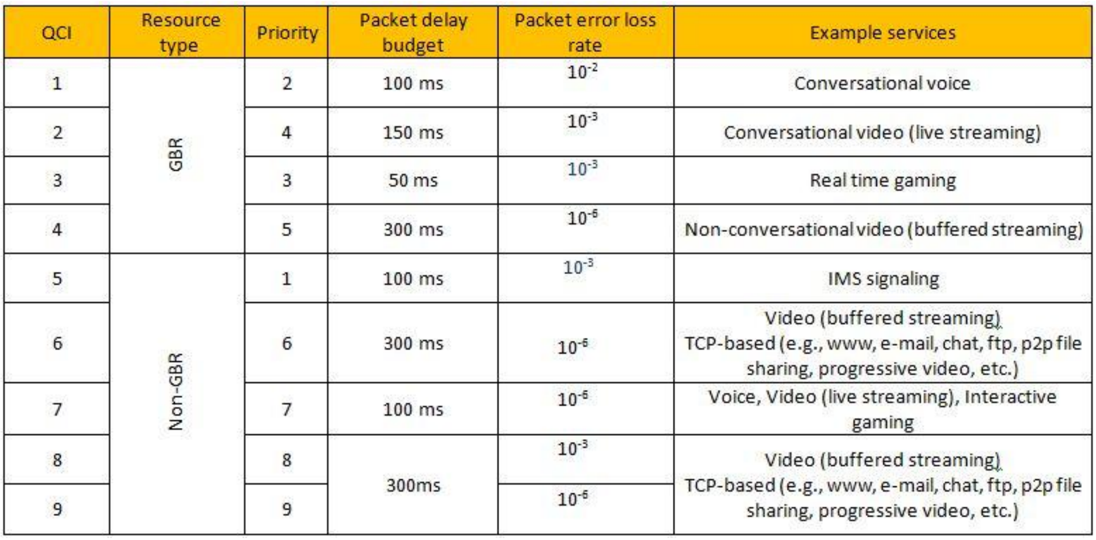
As mentioned before QoE is a subjective and that is why we believe each service provider will have his own QoE definitions.

The summary is shown below:

Selected thresholds for real time gaming based on all the above information:

QCI 3, real time gaming: gamer that plays up to 2 hours session

One way Packet delay 50 msec
One way Delay variation 20 msec
Loss rate 10^-6



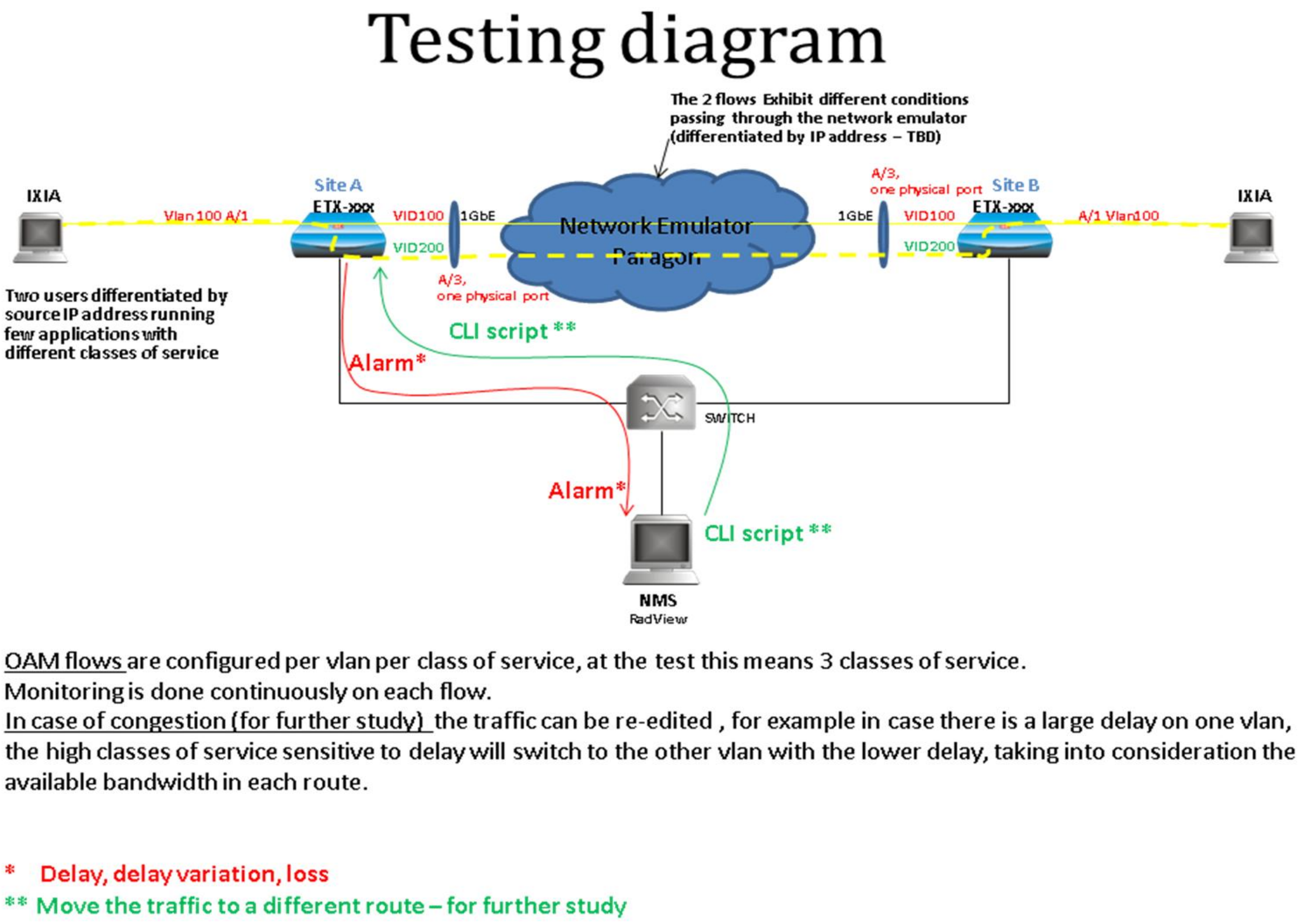
QCI	Resource type	Priority	Packet delay budget	Packet error loss rate	Example services
1	GBR	2	100 ms	10^{-2}	Conversational voice
2		4	150 ms	10^{-3}	Conversational video (live streaming)
3		3	50 ms	10^{-3}	Real time gaming
4		5	300 ms	10^{-6}	Non-conversational video (buffered streaming)
5	Non-GBR	1	100 ms	10^{-3}	IMS signaling
6		6	300 ms	10^{-6}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	10^{-6}	Voice, Video (live streaming), Interactive gaming
8		8	300ms	10^{-3}	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9		10^{-6}	

Selected thresholds for VoIP services based on all the above information:

QCI 1, conversational voice
The ITU-T p.800.1 showed that a satisfactory QoE for VoIP can be obtained when the network operates within QoS limits of:
One way Packet delay 100 msec
One way Delay variation 50 msec
Packet loss 10^-2

Selected thresholds for default bearer based on all the above information:

QCI 9(default bearer), best effort class:
One way Packet delay 300 msec and loss rate 10^-2



Testing diagram

The 2 flows Exhibit different conditions passing through the network emulator (differentiated by IP address – TBD)

IXIA

Vlan 100 A/1

Site A

ETX-xxx

VID100

VID200

1GbE

A/3, one physical port

Network Emulator Paragon

1GbE

A/3, one physical port

Site B

ETX-xxx

VID100

VID200

A/1 Vlan100

IXIA

Two users differentiated by source IP address running few applications with different classes of service

CLI script **

Alarm*

SWITCH

Alarm*

CLI script **

NMS

RadView

OAM flows are configured per vlan per class of service, at the test this means 3 classes of service.

Monitoring is done continuously on each flow.

In case of congestion (for further study) the traffic can be re-edited , for example in case there is a large delay on one vlan, the high classes of service sensitive to delay will switch to the other vlan with the lower delay, taking into consideration the available bandwidth in each route.

* Delay, delay variation, loss

** Move the traffic to a different route – for further study

Figure 29. Eth OAM testing diagram.

9.3.1 Network Impairments

Few scenarios were examined based on different networks models from G.1050 and different classes of service and their congestion thresholds taken from 3GPP.

9.3.1.1 Network models G.1050

Table 2/G.1050 – Impairment Ranges for Well-Managed Network (Profile A)

Impairment Type	Units	Range (min to max)
One Way latency	ms	20 to 100 (regional) 90 to 300 (intercontinental)
Jitter (peak to peak)	ms	0 to 50
Sequential Packet Loss	ms	Random loss only (except when link failure occurs)
Rate of Sequential Loss	sec^{-1}	Random loss only (except when link failure occurs)
Random Packet Loss	%	0 to 0.05
Reordered Packets	%	0 to 0.001

Table 3/G.1050 – Impairment Ranges for Partially-Managed Network (Profile B)

Impairment Type	Units	Range (min to max)
One Way latency	ms	20 to 100 (regional) 90 to 400 (intercontinental)
Jitter (peak to peak)	ms	0 to 150
Sequential Packet Loss	ms	40 to 200
Rate of Sequential Loss	sec^{-1}	$\leq 10^{-3}$*
Random Packet Loss	%	0 to 2
Reordered Packets	%	0 to 0.01

* Sequential Packet Loss occurs 1 every 1000 seconds

Table 4/G.1050 – Impairment Ranges for Un-managed Network (Profile C)*

Impairment Type	Units	Range (min to max)
One Way latency	ms	20 to 500
Jitter (peak to peak)	ms	0 to 500
Sequential Packet Loss	ms	40 to 10,000
Rate of Sequential Loss	sec^{-1}	$\leq 10^{-1}$**
Random Packet Loss	%	0 to 20
Reordered Packets	%	0 to 0.1

9.4 Further work

- Load management in case of congestion
- Insert improvements into the available bandwidth algorithm
- Check Openflow (future versions) as an optional interface between the CSGs and the mediator.
- Be part of SDN solution that allows dynamic load balancing and elasticity of the network
- Check the closed loop concept with PCRF vendors

9.5 Conclusions

Available bandwidth calculations over the transport and congestion detection using Eth OAM tools can be used to help the PCRF troubleshoot the backhaul congestion dynamically and effectively.

The current Eth OAM standard is limited in identifying short congestion events because of its resolution but these short congestion events (<200msec) should probably be solved without intervention of the policy mechanism in the mobile network.

In case of detected congestion, allowing the network to heal itself locally by local policy enforcer can help offload the PCRF and lower the recovery time from congestion in the backhaul network.

10. MACSEC for security

10.1 Introduction to proposed solution

LTE mobile networks need security capabilities for user authentication, data integrity and optionally encryption at a carrier grade level. The 3GPP standardized security approach for LTE is based on IPsec tunneling between the eNBs and security gateways (SEGs) located at the core networks.

This approach suffers from several disadvantages:

- Forcing all connections to traverse the SEG is reasonable for hub-and-spoke backhaul, but contrary to the LTE approach of a flat IP network.
- In particular, X2 connections between neighboring eNBs, which need to have very low latency for advanced features such as network MIMO, are instead routed via SEGs, significantly increasing their latency.
- Due to their complex functionality and configuration requirements, SEGs are expensive from both CAPEX and OPEX points of view.

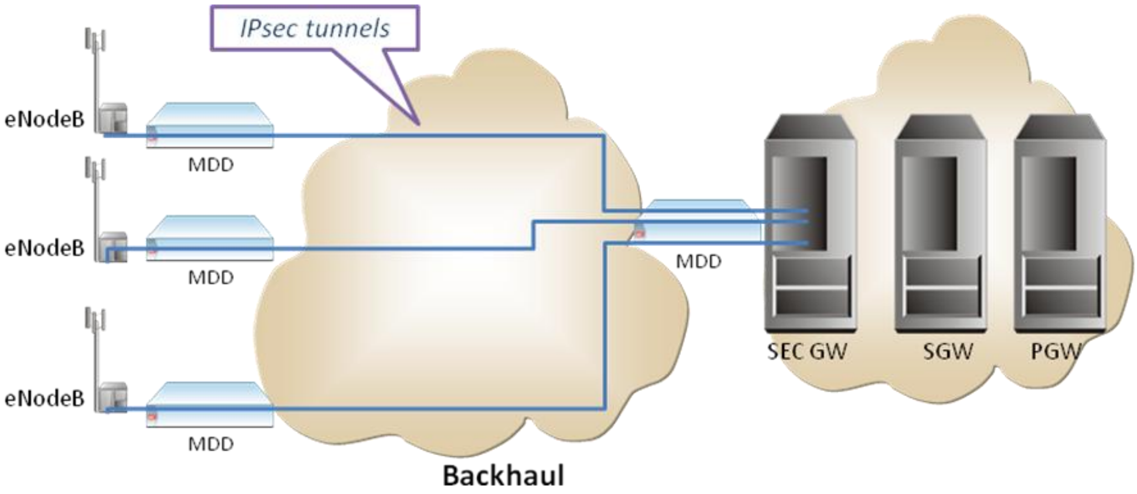


Figure 30. 3GPP security architecture

10.2 Detailed description of the proposed solution

The proposed solution reduces latency and expenditure while maintaining a high level of security, based on three principles:

- The backhaul network is based on Ethernet with MACsec (802.1AE).
- An extension to MACsec enables end-to-end security for multi-hop connections.
- Wirespeed security functionality is implemented in access NIDs in a cost-effective manner.

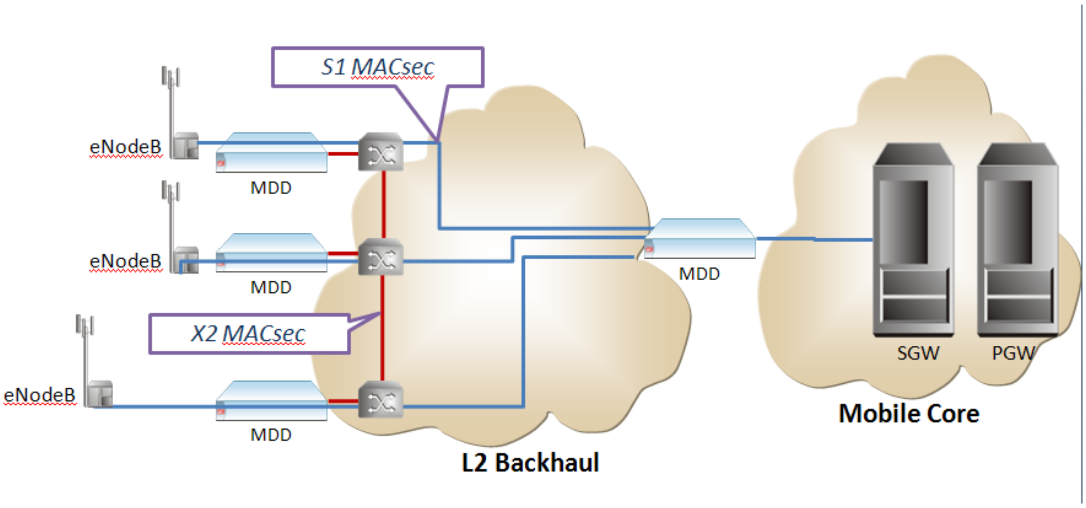


Figure 31. MACsec architecture.

Notes:

1. The X2 connections may be true L2 interconnections rather than logical overlays, thus avoiding additional latency of the latter topology. Only S1 traffic needs to be forwarded towards the core network.
2. The SEG is replaced by a centralized Mobile Demarcation Device (MDD) that functions not only as a demarcation but also as a security component for terminating the S1 MACsec.
3. The proposed MACsec is an extension to standard single-hop MACSec and provides an end-to-end L2 tunnel from eNB to the mobile core or from eNB to eNB.

10.2.1 Secured L2VPN

The proposed secure L2VPN solution provides a point-to-point secure channel for eLine topologies, a point-to-multipoint secure channel for eTree, and a multipoint-to-multipoint secure channel for eLAN. The security is provided at L2, both for single hop (single physical LAN connection), and at end-to-end (between EVC peers) cases. The solution is based on standard protocols: IEEE 802.1AE (2006) and IEEE 802.1X-2010, and it provides frame source authentication, frame integrity protection, replay protection, and optional confidentiality. For end-to-end security an extension to 802.1AE may be required.

One unique advantage of the solution is the ability to create a multisite secure L2 VPN over a switched network, using standard protocols. Secure L2 VPN solutions have several advantages over secure L3 VPNs – among them improved cost performance, lower latency, faster protection switching, and tight integration with port access control.

10.2.2 Security services

The proposed solution can provide the following security services:

1. Source authentication: source authentication guarantees that the frame was indeed sent by the entity identified by the MAC address / EVC peer claimed in the frame header.
2. Integrity protection: integrity protection guarantees that the frame was not modified on route from source to destination.
3. Replay protection: replay protection guarantees that each original frame is delivered only once.
4. Confidentiality: confidentiality guarantees that the information in the frame can not be read by unauthorized entities.

10.2.3 Operation modes

MDDs support two secure operation modes: single-hop (standard and interoperable) and end-to-end (proprietary extension).

10.2.3.1 Single-hop operation

When a port is configured for single-hop operation, a SecTag is inserted into the frame as the first tag followed by a 12 byte Initialization Vector (IV), per 802.1AE. Integrity protection is applied to the entire frame, starting from the Source Address and encryption may be applied to the payload. The original FCS is discarded, a 16 byte Integrity Check Value is appended, followed by a new FCS.

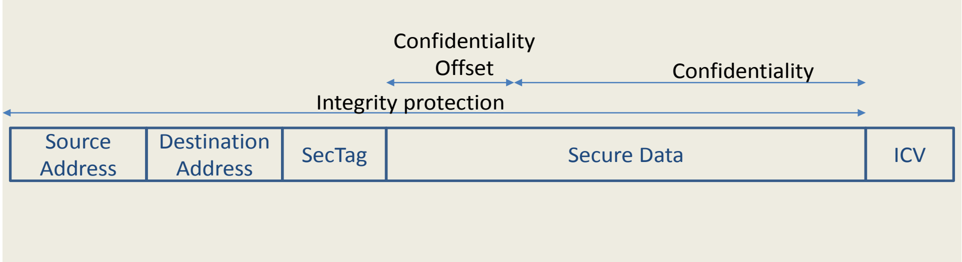


Figure 32. Single hop operation mode

As this mode conforms to standard 802.1AE, it can be used between any two devices supporting standard MACSEC.

10.2.3.2 End-to-end operation

When a port is configured to work in end-to-end mode, then the SecTag and its IV are inserted after the S-tag (if one exists) before the C-tag. The integrity protection is applied to a pseudo frame created by removing the S-tag, encryption is optionally applied to the payload after a configured offset. Thus, pushing, popping, or swapping of S-tags do not invalidate the frame's security.

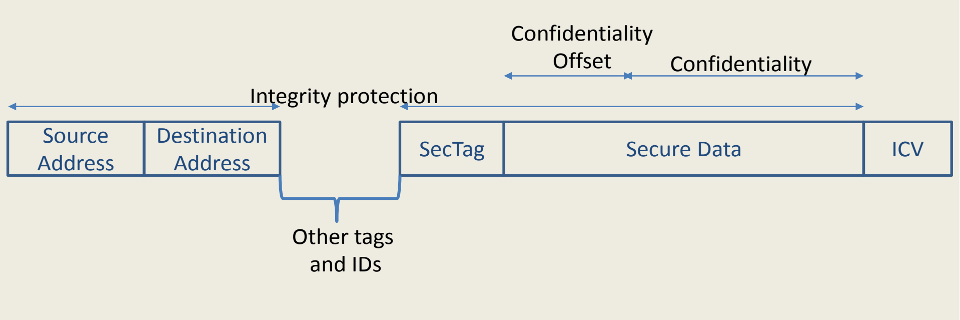


Figure 33.End–to-end operation mode

This end-to-end mode is a proprietary extension, not presently supported by all vendors. It is proposed here due to several limitations of standard MACsec:

1. In standard MACsec the S-tag, which conventionally indicates the EVC, may be encrypted, and thus unavailable for IVL switch forwarding.
2. Even when the S-tag is unencrypted, conventional S-tag processing would invalidate the frame.
3. Thus standard MACsec requires termination of the security association at every switch along the path.

10.2.4 NB and MDD authentication

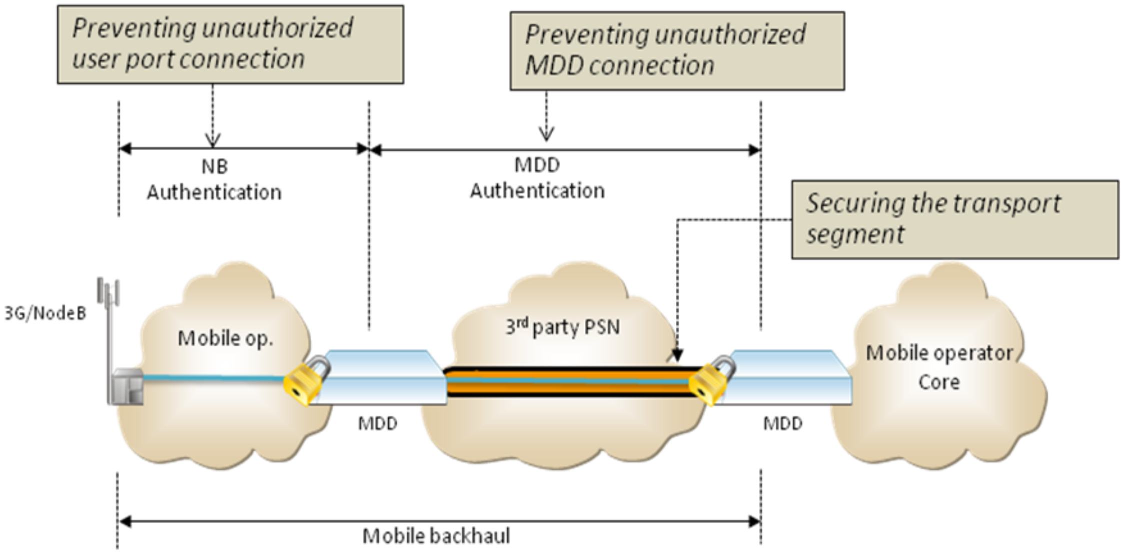

Preventing unauthorized user port connection
Preventing unauthorized MDD connection
NB Authentication
MDD Authentication
Securing the transport segment
3G/NodeB
Mobile op.
3rd party PSN
Mobile operator Core
MDD
MDD
Mobile backhaul

Figure 34. NB and MDD authentication.

10.2.4.1 NB authentication

A NB desiring to connect to the MDD port must initiate authentication before forwarding traffic to the port. The MDD must authenticate the NB before forwarding traffic from/to this NB.

10.2.4.2 MDD authentication

An MDD attempting to join an EVC must be authenticated in order to receive the secure EVC secret key which enables it to send and receive MACSec frames from the other EVC members.

10.2.4.3 Authentication Process

NB or MDD authentication is initiated by an 802.1X supplicant that sends (EAPOL) start messages to the appropriate authenticator(s). The authentication is initiated by the NB or by the MDD attempting to the remote MDD or to the centralized MDD. The participants in each 802.1X authentication exchange are the supplicant, the authenticator and the authentication server.

10.2.5 Out of Franchise (OOF) use case

Another important use case for the secured L2VPN is protecting an OOF network segment, i.e., a network segment that is leased by the mobile operator from a wholesaler. In this case the mobile operator who desires to provide a truly secured L2VPN cannot guarantee end-to-end security without protecting the OOF segment. This can be accomplished using (the mobile operator's or the wholesaler's) demarcation devices to perform the required security functions. In fact, the demarcation device is the ideal placement for the security functionality, as this assists the operator in locating the attacked segment and localizing the threat.

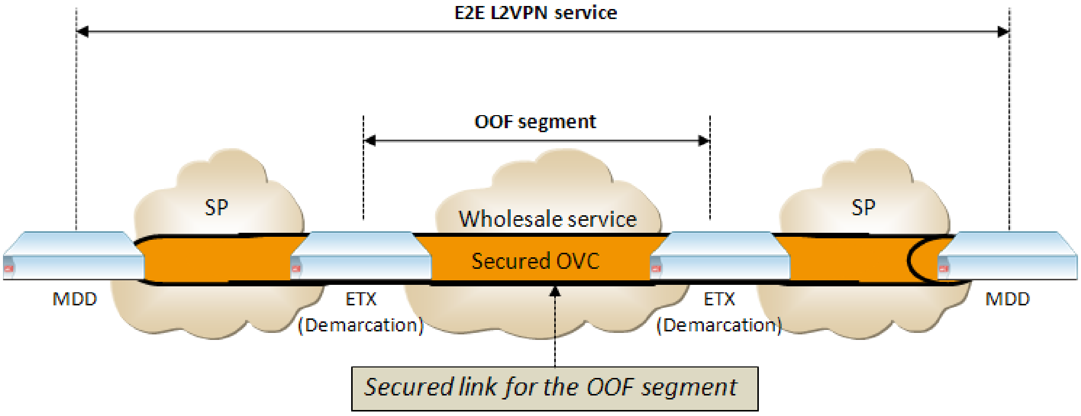


Figure 35. E2E L2VPN service.

10.3 Main results

10.3.1 Throughput and Performance

- For equivalent computational resources, the throughput of L2 encryption can be higher than that of L3 encryption, due to lower per packet overhead.

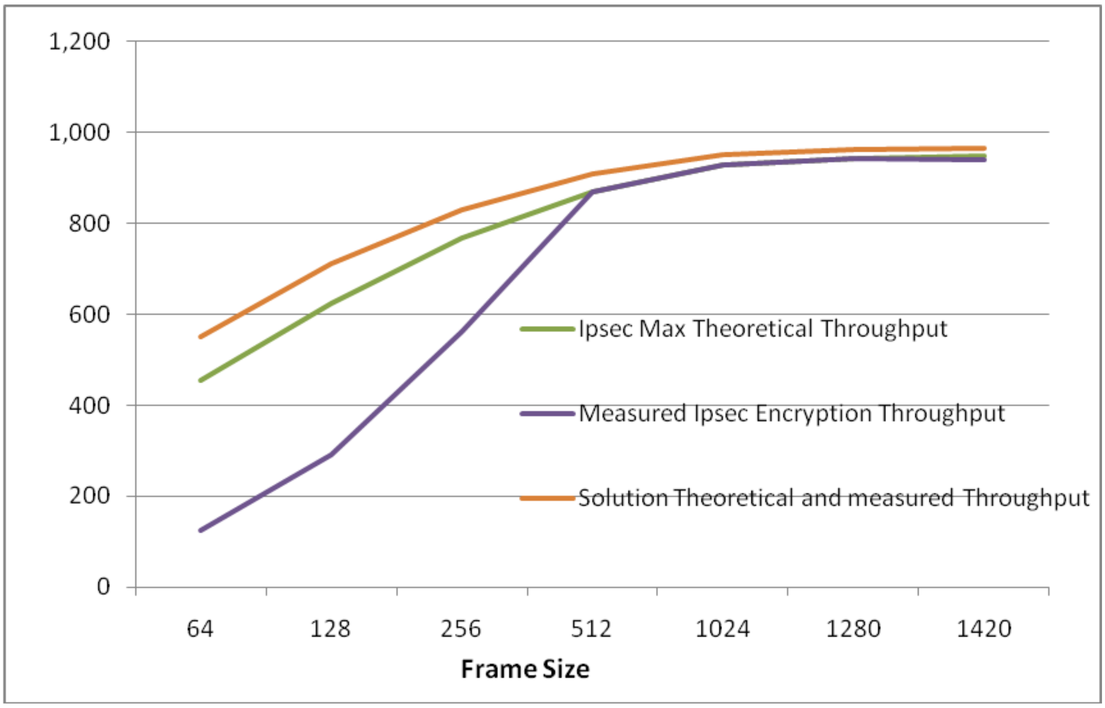


Figure 36.L3 encryption performance.

10.3.2 Latency

- Wirespeed implementation of the proposed solution adds negligible per-packet delay (approximately 0.1 □sec).
- The latency of IPsec is about 300□sec for a 1420 byte packet.

  Note: This measurement was done for a software based device implementing IPsec. Most of IPsec implementations are software based and this is due to the many modes required to be supported in order to become IPsec compliant.

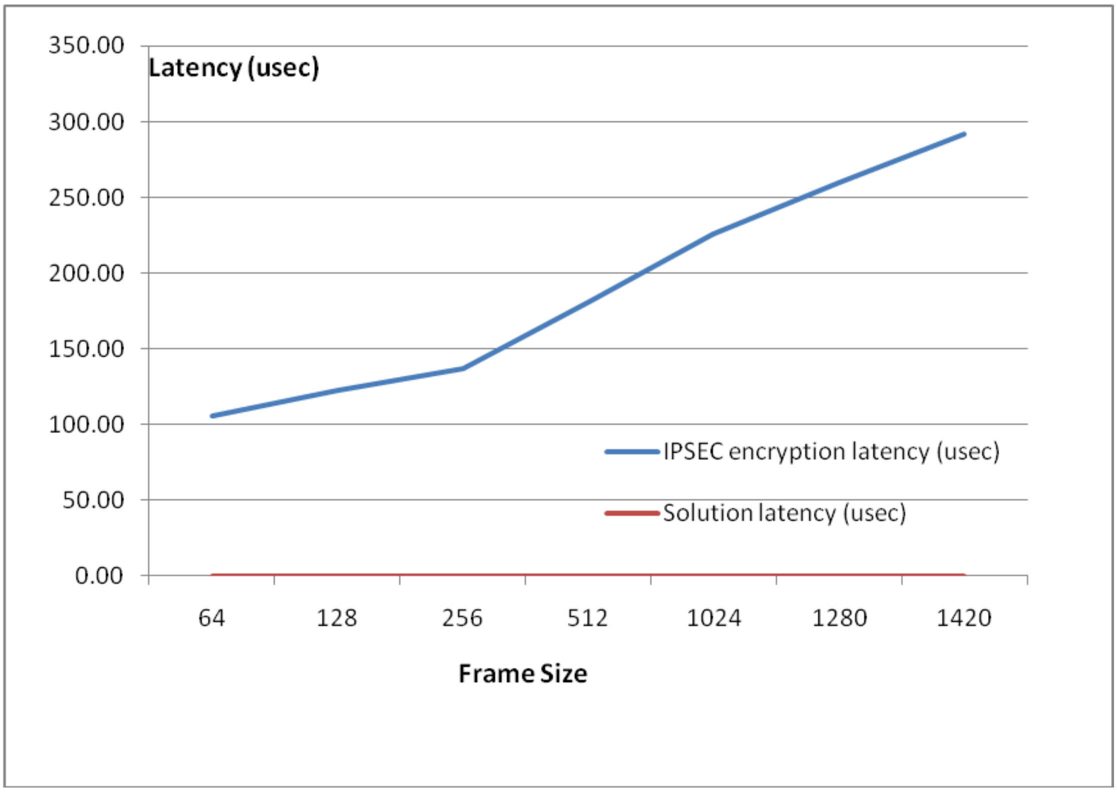


Figure 37. L3 encryption latency

10.4 Future extensions

- For wide scale deployment it will be necessary to standardize the end-to-end solution and/or to perform interoperability testing with other vendors.
- In order to reduce operational load, it will be necessary to develop a simple security key distribution model.

10.5 Conclusions

- The proposed solution provides an affordable security solution based on Ethernet and MACSec. No security gateway is required.
- The proposed solution secures S1 connections between the (e)NB and a central MDD.
- The proposed solution enables X2 connection latency to remain low, as these connections need not be routed via a security gateway.
- The proposed solution adds negligible processing delay, as compared to IPsec implementations that may add significant delay.

11. Summary

The proposed technologies provide performance improvements mainly in the area of mobility, security and end to end delay. The results show that proposed technologies will improve the overall performance in mobile networks. These technologies overcome some of the limitations identified in MEVICO project such as i) throughput gain in 3GPP access and backhaul, ii)reducing the backhaul and RAN influence on E-E delay, iii) reducing the recovery time from link failures or congestions and/or OPEX reduction in the in the backhaul or core transport network layer, iv) efficient load distribution in the backhaul and in the core and Offload gain due to the usage of multi-access capabilities and Capacity aggregation and E2E QoE sustainment, v) reducing the service interruption delay due to handover, vi) reducing the handover related signaling load on the network and vii) reducing the E-E delay between UE and content.

12. References

[MBMS2011] Universal Mobile Telecommunications System (UMTS); LTE; Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description, ETSI TS 123 246 V10.0.0, March 2011.

[Saada2012] H. Asaeda et al., “Multicast Routing Optimization by PIM-SM with PMIPv6”, Internet Draft, March 2012, (Work in Progress).

[RFC 4861] T. Narten et al. “Neighbor Discovery for IP version 6 (IPv6)”, RFC 4861, September 2007.

[RFC 5213] S. Gundavelli et al., "Proxy Mobile IPv6", RFC 5213, August 2008.

[Hui2012] M. Hui et al, “Fast Handover for Multicast in Proxy Mobile IPv6”, Internet Draft, March 2012, (Work in Progress).

[RFC 6224] T. Schmidt, “Base Deployment for Multicast Listener Support”, April 2011, RFC 6224.

[RFC 3261] J. Rosenberg et al.,“ SIP: Session Initiation Protocol", RFC 3261, June 2002.