This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

DNS-based Solution for Operator Control of Selected IP Traffic Offload

Tarik Taleb, Konstantinos Samdanis, Stefan Schmid
NEC Laboratories Europe
{tarik.taleb, konstantinos.samdanis, stefan.schmid}@neclab.eu

***Abstract*—In this paper, we consider the Selected IP Traffic Offload (SIPTO) approach to handle increased data traffic of both local and macro-cellular networks. We devise different approaches based on operator defined offload policies on a per destination domain name basis, which offer operators fine-grained control of whether a new IP connection should be offloaded or provided via the core network. Two of our solutions are based on Network Address Translation (NAT) named simple-NATing and twice-NATing, while a third one employs simple tunneling. These solutions support all kinds of UEs including those that support a single Packet Data Protocol (PDP) context/Packet Data Network (PDN) connection. In a forth solution, we consider the case where a User Equipment (UE) supports multiple PDP contexts/PDN connections, with at least one dedicated for SIPTO traffic. A qualitative analysis and a simulation study are presented.**

I. Introduction

Along with the ever-growing community of mobile users and increasing number of mobile bandwidth-intensive applications, operators are looking for efficient networking solutions that ensure the scalability of their systems while minimizing investments in the infrastructure [1] [2]. In order to optimize usage of network resources, operators should be capable to selectively offload IP traffic as near to the edge of operator's network as possible.

Two concepts are currently highlighted in 3GPP's System Architecture working group (SA2), the Local IP Access (LIPA) and the Selected IP Traffic Offload (SIPTO) [3]. The first one is supported by residential and corporate femtocell deployments, enabling local network access. It allows a mobile terminal, connected via a femtocell (i.e. a 3G Home NodeB or LTE Home evolved NodeB), to directly connect to other IP-capable devices in the local network without detouring via operator's core network. The latter contributes to offload, selected IP traffic at both femtocell networks as well as at 3G or LTE macro-cellular access networks.

By breaking out selected traffic at entities close to the edge of the network, operators may avoid overloading their scarce resources, i.e. PDN (Packet Data Network) Gateways (P-GW) and Serving Gateways (S-GW).

The discussions and analysis in 3GPP currently focus on shaping the architecture in terms of positioning the local breakout/traffic offload point. Issues regarding security, charging, mobility, and traffic control/handling are yet to be investigated. This paper focuses on finding adequate solutions that jointly address the two latter issues, namely service continuity/mobility support and traffic control.

Effectively, operators are interested in controlling traffic, i.e. routing of particular traffic depending on the type of traffic and the destination IP network/address. With such control, they can flexibly regulate which IP traffic flows to be offloaded in order to allow lawful intercept, access optimization to specific Internet services, and value-added services provided by the operator (e.g. content filtering, child protection).

This paper introduces a DNS-based solution that meets such objectives and at the same time supports service continuity of SIPTO traffic with minimal complexity with respect to both the core network and User Equipment (UE). We consider SIPTO at H(e)NB and at macro (e)NBs focusing only on SIPTO traffic to the Internet, i.e., leaving out the aspects related to LIPA traffic towards a home network. Moving the traffic offload decisions to the network core, via the DNS-based solution significant resources savings at P/S-GWs in the core network may be achieved. These savings could be used to accommodate other types of services, e.g. VIP customers. In addition, the use of DNS may reduce the investment for accommodating increased traffic in a cost effective way by a simple software upgrade instead of purchasing new P/S-GWs.

The remainder of this paper is organized as follows. Section II presents the state of the art. Following a brief description of the Evolved Packet System (EPS) and its key components, Section III presents our DNS-based SIPTO traffic control methods. The implementation issues of the proposed solutions along with a qualitative and simulation based comparison in terms of their support of service continuity are discussed in Section IV. The paper concludes in Section V.

II. State of the Art

Data offload solutions and particularly LIPA/SIPTO are widely studied within the 3GPP SA2 group with the objective to identify architectural enhancements and functionalities for supporting LIPA for the H(e)NB subsystem and SIPTO for the H(e)NB subsystem as well as for the macro cellular network as required in [4][5]. In [3], different LIPA/SIPTO architectures are introduced offering an insight analysis of the main operations and the related requirements. Besides the 3GPP framework, a generic data offloading scheme focusing on local access networks is described in [9] concentrating on UEs with multiple and a single APNs.

A study that advances the architecture knowledge by analyzing the deployment scenarios of SIPTO with respect to QoS and quantifying the benefits of data offloading in the macro cellular network by introducing LIPA and SIPTO services in the home or local networks is introduced in [8]. Such study is centered on UMTS, considering the geographical position of micro and macro cells as well as the radio characteristics of the system and the losses due to indoor environment. The presented results demonstrate that the offloading benefits may

978-1-61284-233-2/11/$26.00 ©2011 IEEE

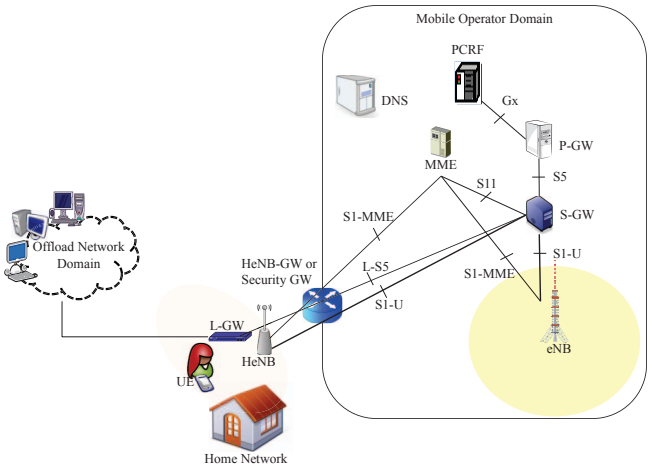

Fig. 1. Overall network architecture.

increase the operator's network capacity up to 100% but may vary depending on the interference among macro and micro cells.

Most of the current efforts for providing SIPTO services mainly focus on locating the functionality on the operator's network and on providing methods for distinguishing the SIPTO related traffic. This paper advances the existing methods by introducing a DNS-based approach for controlling SIPTO traffic handling. At the same time, the presented work is the first to consider service continuity for SIPTO traffic during handover. DNS is already employed for selecting geographically the nearest S-GWs/offload points for SIPTO macro-cellular network solutions. However, this is orthogonal to the objective of this work. The envisioned SIPTO service continuity is on per SIPTO flow basis, similar to IP flow mobility [10] and the distributed IP flow mobility considered in [7], though the network environment and the means of achieving mobility are different.

III. DNS-Based SIPTO Traffic Control

Fig. 1 depicts the major components of the envisioned architecture, namely a core DNS server, a Mobility Management Entity (MME), a S-GW, a PDN-GW, an eNB, a HeNB/Security-GW, a HeNB with a collocated local gateway (L-GW), and an offload network domain. The local gateway can be functioning as either a local small-scale P-GW in case of UEs supporting multiple APNs or as a simple local gateway only including some necessary P-GW functions [3]. In some solutions, a DNS proxy is assumed to be at the local GW. In this paper, two types of UEs are considered: UEs using one single PDN connection/IP address for both SIPTO and non-SIPTO traffic and UEs using multiple PDN connections/IP addresses with one dedicated for SIPTO traffic.

For SIPTO traffic control, a basic solution could be performed by enforcing IP flow-based routing policies at the eNB or HeNB. In this solution, the Policy and Charging Rules Function (PCRF) or HNB Management System (HMS) dynamically updates (H)eNBs with "IP flow filters" for policy-based routing in order to separate SIPTO and core network traffic. A major concern with this solution is the complexity of provisioning these IP flow filters dynamically to (H)eNBs or alternatively the complexity of pro-actively providing SIPTO IP flow filters to all (H)eNBs that are SIPTO enabled. This incurs high signalling cost. Additionally, with no additional intelligence, service continuity of SIPTO traffic cannot be supported with such IP flow routing policy.

A solution whereby the routing decision is taken centrally in the operator core network and offload decisions are explicitly provided to (H)eNBs during IP flow setup may be preferred. In the envisioned mechanisms, decision on which traffic is to be handled via the macro cellular network and which one to be offloaded is taken by the operator core network DNS. Four different options are explored in the following subsections that support SIPTO services requiring different UE capabilities with respect to multiple APNs support and different functionalities on the (H)eNB as well as on the core network.

A. Simple Source NATing Solution

In a simple "DNS-based SIPTO control" solution, referred to as Simple Source NATing, we consider UEs supporting one single APN for both SIPTO and non-SIPTO traffic. In this solution, the L-GW provides NAT services for SIPTO traffic by translating the IP address of the UE into a local IP address and adds it as an entry into its NAT table.

The SIPTO service is indicated by a flag in the DNS reply message that specifies the offload decision for the related IP traffic flow. In case of offload, the IP address of the peer provided in the DNS reply is stored at the HeNB and used for the traffic offload enforcement. It is worth noting that applications that do not work through NAT cannot be offloaded using this solution.

B. Twice-NATing Solution

In this solution, both source and destination addresses are translated as offload traffic crosses the L- GW. For non-SIPTO traffic a UE uses the IP of the peer, i.e. YouTube server, while for SIPTO traffic it uses an IP address of the local GW, referred to as destination NAT (DestNAT). The local GW performs twice-NATing on SIPTO traffic by translating the destination NAT address to the IP address of the peer and the UE global IP address into the local GW or external NAT address (Source NAT).

In this variant, the traffic offload is again triggered by a flag in the DNS reply message. The IP address of the peer is stored in the local GW and is associated with the local DestNAT. The address space for DestNAT is subject to certain limitations, since DestNAT should be routable in the operator network. IPv6 support or solutions based on the combination of source/destination port numbers in conjunction with the UE's IPv4 address, as described in [6], can be envisioned. It should be noted that the Twice-NATing solution also suffers from the same limitations as conventional NAT-based solutions.

C. Simple-Tunneling Solution

In the simple-tunneling approach, UEs forward SIPTO traffic by establishing a tunnel towards the L-GW, which performs simple source address translation. Non-SIPTO traffic is again routed through the operator core network using the UE's global IP address. To enable the simple tunneling approach, the UE is supplied, in the DNS reply message, with the L-GW's IP address, routable within the macro network, in addition to the IP address of the peer. Upon receiving the offload traffic, the L-GW "untunnels" packets and performs simple source address translation.

In this variant, the traffic offload is also triggered by a SIPTO flag in the DNS reply. Upon receiving the flag, the L-GW adds to the DNS response towards the UE its IP address. UEs realize that a particular connection is subject to SIPTO by the additional tunnel end-point addresses provided in the DNS reply and tunnels the uplink traffic to the local GW address. An alternative approach to tunneling could be source routing based on the IPv6 header options. UEs maintain per flow state information to decide how to route a flow. Such information is kept at the network-layer and is completely transparent to the application layer.

D. Multiple-APNs UEs-oriented Solution

In this solution we consider UEs that can support multiple PDN connections to different APNs, with at least one APN dedicated for SIPTO. In this variant, the operator DNS indicates to the UE which APN to use for a given IP flow. The DNS server must be aware of the configured APNs. UEs may inform the DNS server of their available APNs as part of the DNS request. The DNS server may also recommend a list of APNs in order of priority that is defined based on different parameters.

Alternatively, the operator DNS can employ a flag in the DNS reply to indicate that this IP flow should be offloaded. In this case, UE must be able to autonomously identify the adequate APN for SIPTO. In response to the DNS reply, the UE binds the new IP flow to the UE's IP address associated with the recommended PDN connection/APN. It is worth noting that UE requires only simple network-level functionality for the binding process, which is anyway supported by UEs supporting concurrent PDN connections.

IV. Evaluating Service Continuity

This section analyzes the performance of the proposed SIPTO approaches with respect to service continuity based on a qualitative analysis with respect to complexity and based on a simulation study that compares the main path selection options.

A. Qualitative Analysis

In this section, we evaluate the proposed solutions and assess their support of service continuity. Fig. 2 depicts all potential paths for both uplink (UL) and downlink (DL) traffic upon the handoff of a UE . There are three possible paths for downlink traffic, namely 1DL - 3DL, and five possible

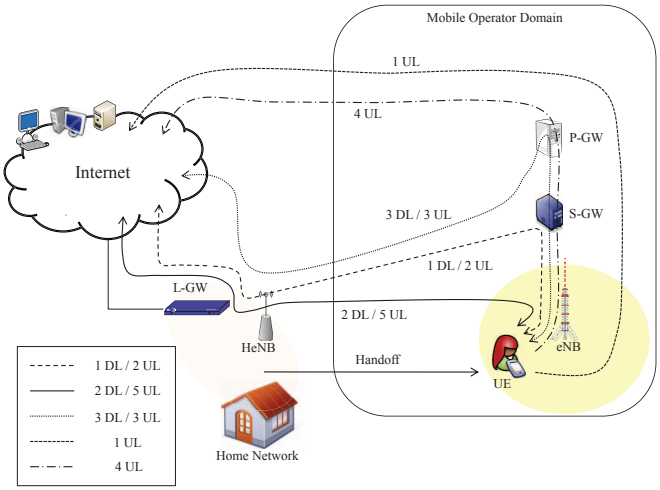

Fig. 2. DL/UL potential paths after handoff.

paths for uplink traffic namely, 1UL - 5UL. In case SIPTO is handled via IP flow filters, the uplink traffic breaks-out at the target (H)eNB using 1UL in Fig. 2 without providing service continuity support in case of handover. Similarly, in the DNS-based Simple Source NATing solution, service continuity cannot be supported since the corresponding peer will, upon a handover, receive the UE traffic with a different source IP address, namely the UE's global IP address.

Service continuity for ongoing SIPTO traffic can be supported only if the break-out point for ongoing connections remains the same, at the L-GW of the source (H)eNB. This implies a L-GW must remain as anchor point employing novel mechanisms to forward/route the uplink traffic from the UE to the L-GW. The traffic may traverse the core network before reaching the new location of the UE, a process that may need some further functionality on selected network elements. Paths that transverse the S-GW (e.g. 2UL in Fig. 2) require some extra functionality at S-GW to distinguish SIPTO from non-SIPTO traffic, break it out and route it to the L-GW at the source (H)eNB. Connecting directly the L-GWs at the source and target (H)eNBs as 2DL and 5UL and/or support of data forwarding over the X2 interface are optimal but may need (H)eNB/L-GW enhancement, which are outside the scope of this paper.

In case of the Twice-NATing solution, using the DestNAT, which is routable within the operator network and Source NAT, service continuity of the SIPTO traffic can be guaranteed by enforcing the downlink and uplink traffic to follow paths 1DL or 3DL and 2UL or 3UL as shown in Fig. 2, respectively. In the uplink, path 3UL can be easily established as this requires merely the Twice-NATing functionality in the L-GW, which needs to intercept packets sent to the DestNAT address. The use of the alternative path 2UL requires some extra functionality in the S-GW to detect traffic targeted to the L-GW based on the DestNAT address range. Considering the downlink, path 3DL follows the conventional standardized path, while the optimized 1DL requires functionality alternations at the S-GW.

TABLE I
COMPARISON AMONG THE SIPTO SOLUTIONS

	Single APN				Multiple APN DNS based
	IP flow filter	DNS based			
		Simple NATing	Twice NATing	IP-in-IP Tunneling	
System complexity & Cost	High	Low	Moderate	Moderate	Low
Transparency to UE	Yes	Yes	Yes	Transparent to application layer but network layer involved	Transparent to application layer but network layer involved
Service continuity support	No	No	Yes	Yes	Yes
Changes to DNS resolution	None	SIPTO flag	SIPTO flag & DestNAT	SIPTO flag & L-GW address	SIPTO flag or APN
Flushing of DNS caching at UE	No impact	Requires DNS cache flush upon (H)eNB change	Caching not possible (unless IPv6/IPv4 NAPT supported)	Requires DNS cache flush upon (H)eNB change	No impact
Packet inspection processing	Yes	Yes	Yes	Yes	No
Further issues	-	-	IP address space of LP-GW	-	APN priorization

In the Simple-Tunneling based solution, the IP address of the L-GW, used for the IP-in-IP tunnel, is routable within the operator network towards the source (H)eNB, and therefore service continuity of SIPTO traffic can be supported by enforcing the downlink and uplink traffic to follow paths 1DL or 3DL and 2UL or 3UL as in Fig. 2, respectively. In the uplink, path 3UL can be easily established as this requires merely the Simple Tunneling functionality in the L-GW, which needs to terminate the tunnel and route the traffic towards the destination. Path 2UL requires some extra functionality in the S-GW to detect traffic targeted to the L-GW based on the L-GW address range, while the downlink Path 3DL follows the conventional standardized path. Again the optimized 1DL is also supported but requires additionally S-GW functionality.

In the multiple APNs UE-oriented solution, service continuity for SIPTO traffic is supported as the standard mobility procedures ensure that the PDN connections are maintained during handover. The downlink and uplink traffic follow paths 1DL and 2UL as shown in Fig. 2, respectively. It should be noted that unlike the other schemes, the multiple APN approach avoids the cashing problems related with DNS results and peer addressing in service continuation process. A qualitative comparison among all introduced solutions is presented in Table IV-A.

B. Simulation Study

This section provides a comparative analysis using simulation among the different SIPTO service continuity supporting schemes, emphasizing their performance with respect to the various path selection options. The aim is to offer a quantitative insight view between schemes that employ paths traversing the P-GWs and S-GWs, respectively. The simulation study is performed using Matlab. The topology used is illustrated in Fig. 3. It consists of three eNBs with offloading capabilities being connected to the Internet via an offload network domain and to the operator's core network

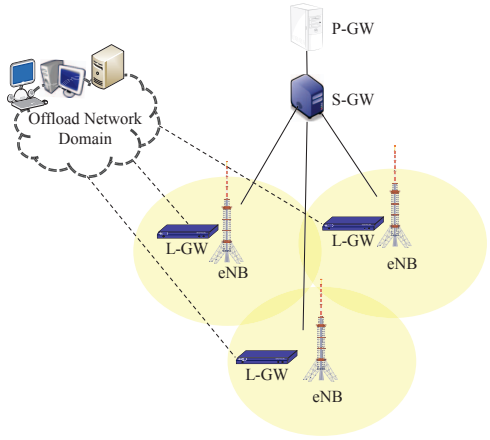

Fig. 3. Envisioned simulation topology.

via a S-GW and P-GW. The capacity of links connecting the eNBs with the S-GW is 25Mbps, while that of the ones connecting the S-GW with the P-GW is 50Mbps, introducing certain capacity limits in handling traffic for each particular network element.

Traffic is modeled as requests with specific source, destination and bandwidth requirements. Incoming users are assumed to arrive at each eNB with an equal probability and independently initiate a flow session following a Poisson distribution with $\lambda = 5$. Such session is routed either towards the core network or offloaded with equal probability. The session duration is assumed to be an exponential random variable with mean $\mu = 3$ min. The cell residence time is also assumed to be exponentially distributed with a mean $n = \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$ min. The underlying mobility model is random walk. The requested bandwidth is uniformly distributed within the interval [6.7 Kbps; 5 Mbps], representing services ranging from voice (3GPP AMR speech codec) to video applications.

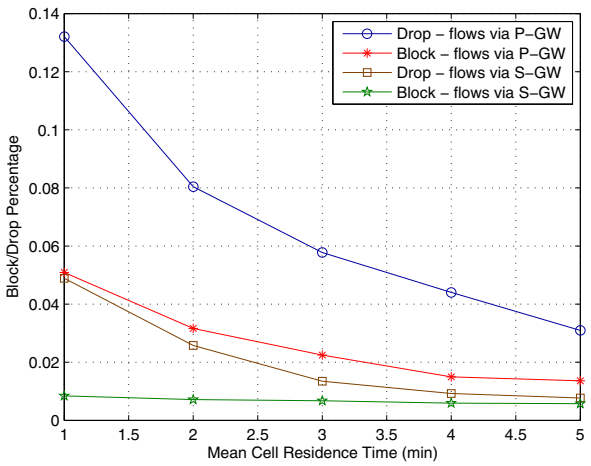

Fig. 4. Performance of SIPTO service continuity supporting schemes with respect to path options.

The evaluation is performed measuring the contribution of each method on blocking and dropping rates. The blocking/dropping rate is defined as the percentage of blocked/dropped sessions to the total amount of incoming sessions. The simulation running time is set to 100 min with the simulation process being repeated 100 times. Measures of dropping and blocking percentage are averaged in the interval [0; 1]. The blocking and dropping percentage for each method with respect to the selected path is illustrated in Fig. 4.

From the results it is obvious that sessions involving the P-GW consume more network resources and therefore experience a higher blocking and dropping percentage compared to the ones that transverse the S-GW only. The performance difference varies with the mean cell residence time since higher mobility results in increased resource consumption towards the P-GW. The performance difference is, however, significant, being more than three times higher in case of high mobility scenarios and twice higher in case of lower mobility scenarios.

Although the performance difference, in terms of dropping and blocking, is significant for flows that use the P-GW, with mean cell residence times that surpass the average session duration, i.e. $n > 4$, such difference is getting closer with the dropping difference being 2% and the blocking less than 1%.

These results suggest that in cases where mobility is relatively low, a SIPTO mobility solution that uses the P-GW instead of S-GW would be cheaper without the need for technological upgrades, degrading the performance slightly within acceptable limits, less than 3%. Therefore, solutions that use the P-GW may be a cost effective alternative for networks that particularly exhibit low mobility.

With respect to the SIPTO service continuity supporting schemes analyzed, the obtained results indicate that the multiple APNs UEs-oriented solution exhibits the most outstanding performance, in comparison to the simple tunneling and the twice-NATing solutions. The required additional system complexity is also minimal in case of the multiple APN UEs oriented solution. However, one obvious drawback of the solution is that it cannot support UEs that support only single PDN connection.

V. Concluding Remarks

In this paper, we proposed DNS-based solutions for providing flexible and fine-grained traffic offload control considering UEs supporting only a single PDN connection and UEs supporting multiple concurrent PDN connections. The service continuity support of SIPTO traffic during handover is achieved by enforcing both downlink and uplink traffic to traverse the L-GW at the (H)eNB, which anchors the IP flow/connection.

For UEs supporting only a single PDP Context/PDN connection, service continuity for SIPTO traffic can be supported either with no additional complexity in core network (when the traffic tunneled to the P-GW and then routed to the local GW based on normal IP routing) or with little additional complexity in the S-GW (when the traffic is directly routed to the L-GW by the S-GW).

For UEs supporting multiple PDN connection, an operator can simply control the traffic offload based on DNS replies that indicate to the UE which PDN connection to use for a particular IP flow/connection. For this, UEs require minimal extra functionality, but could also be involved in the decision process. Service continuity for SIPTO traffic is also supported in this solution based on the standard handover support.

The presented results provide an overview on the fundamental mechanisms, while a further assessment based on a test-bed, forms a future research direction. Finally, exploring the benefits of a load balancing based P/S-GW selection on the proposed solutions is another area of future research work.

Acknowledgment

The research leading to these results has been partially performed within the UniverSelf project (www.UniverSelf-project.eu) and received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement N 257513.

References

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2009-2014", White Paper, Feb. 2010.
[2] "MOBILE TRAFFIC GROWTH + COST PRESSURES = NEW SOLUTIONS?", New Mobile, Jan. 2010.
[3] 3GPP, "Local IP Access and Selected IP Traffic offload", TR 23.829, May 2010.
[4] 3GPP, "Service principles", TS 22.101, Jun. 2010.
[5] 3GPP, "Service requirements for Home NodeBs and Home eNodeBs", TS 22.220, Jun. 2010.
[6] K. Nishida, et.al, "An Unified Multiplex Communication Architecture for Simple Security Enhancements in IPv6 Communications", in Proc. EuroView, Aug. 2010.
[7] D. Liu, et.al, "Distributed Mobility Management", Network Working Group, IETF Draft, Work in Progress, Jan. 2011.
[8] D. Calin, et. al., "On Femto Deployment Architectures and Marcocell Offloading Benefits in Joint Marco-Femto Deployments", IEEE Communication Magazine, Vol.48, No. 1, Jan 2010.
[9] P. Tinnakornsrisuphap, et.al, "Local IP Access Scheme", US 2009/0268668 A1, Oct 2009.
[10] 3GPP, "IP Flow Mobility and Seamless Wireless Local Area Network (WLAN) Offload", TS 23.261, Jun 2010