Simplified Wireless Connectivity for 5G Machine Type Communication

Peyman TalebiFard†‡, Stere Preda†, Catherine Truchan†, Suresh Krishnan†,
Denis Monette†, and Victor C.M. Leung‡
†Ericsson Research, Montreal, QC, Canada.
{stere.preda, catherine.truchan, suresh.krishnan, denis.monette}@ericsson.com
‡University of British Columbia, Vancouver, BC, Canada. {PeymanT, VLeung}@ece.ubc.ca

***Abstract*—Ubiquity of applications supporting the networked society demands an efficient communication system to support Machine Type Communication (MTC) devices. The current architecture is not optimized for MTC traffic patterns and therefore, may lead to excessive signaling. In this paper, we propose a simplified approach to wireless connectivity to cope with the requirements of Massive MTC devices. The essence of our solution is based on an optimized and simplified IPv6 connectivity at the existing base stations for a large number of MTC devices while supporting the coexistence of legacy mobile devices. This paper focuses on the core network aspect of the end-to-end connectivity. By analyzing the signaling load we illustrate the feasibility of the proposed solution and demonstrate that it outperforms the existing architecture through simplification of connectivity and elimination of unnecessary functions.**

Keywords: 5G, Core Network, IPv6, Massive MTC, Simplified Connectivity.

I. INTRODUCTION

Various reports forecast that a large number of mobile devices will be connected by 2020 and beyond [1]. The market expansion of Internet of Things (IoT) and Machine Type Communications (MTC) [16], [1] calls for rapid evolution of radio access and core networks. Design requirements are already in place to ensure that future mobile applications and devices are certified and are able to fulfill the operator's requirements [17].

A substantial number of MTC devices are also expected to generate lower Average Revenue Per User (ARPU). As a result, the operators may establish a set of policies and requirements for resources that are dimensioned for low revenue use cases [2].

Therefore, sustaining the traffic growth from large sets of diverse mobile devices and communication patterns remains a challenge which will be referred to, in this paper, as the 26 Billion challenge (26B) [1].

The 3rd Generation Partnership Project (3GPP) has listed the network requirements for a generic MTC deployment and proposes an enhanced Evolved Packet Core (EPC) solution to meet them, i.e., with MTC-dedicated EPC functionalities [4], [5]. It has already been shown in [15] that the 3GPP proposed architecture could be further enhanced. However, our proposed solution differs significantly from the 3GPP based approach because our analysis indicates that the packet core signaling load and the user plane overhead remain high. The existing proposals have the following disadvantages:

- Data plane management overhead: excessive number of M-MTC devices and frequent establishment and modifications of Data Radio Bearers (DRB)s, impose a significant load on the control plane signaling. A significant portion of the signaling overhead on the Mobility Management Entity (MME) is due to the establishment and reconfiguration of GTP tunnels between SGW and PGW.
- Management and control of network are agnostic from type of traffic as envisioned by different M-MTC device profiles.

In this paper, we propose a Simplified Wireless Connectivity (SWC) approach to address the 26B challenge for communication of MTC devices particularly for the case of Massive MTC (M-MTC) for future 5G networks. The emphasis of this work is to improve the core network functionality with the following objectives:

- Efficient and scalable: the traffic should benefit from efficient IP connectivity that is optimized for MTC communication patterns, which could include infrequent and small volumes of data.
- Optimal and simple: the utilization of network should be improved and payload data traffic should be maximized. A simplified network provides better manageability.

This is a work in progress and the evaluation results presented are preliminary. The cornerstone of our proposal is simplified IP based connectivity through the base station with a *dual personality* to support the coexistence of legacy mobile devices as well as M-MTC devices. As a consequence, some legacy core functionalities are still required for SWC, such as those for security mechanisms that have proven to be successful in 3GPP Evolved Packet Systems (EPS). Furthermore, SWC should be easily extensible by software to provide enhanced network functions and meet new requirements such as those for M-MTC. To incrementally support such basic connectivity, SWC should be equipped with functions that improve network utilization. Given the large number of devices that need to be connected, IPv6 is the protocol of choice for such deployments.

The contribution of this paper is twofold: (1) we propose a flexible architecture that enables an efficient core connectivity to meet M-MTC requirements in the 26B context. Given its modular approach, SWC allows easy integration of legacy operator services. (2) Through leveraging IPv6 based mechanisms, re-using the principle of separating the user and control planes, our proposal leads to a simplification of the control plane and the data plane for the different MTC scenarios. We analytically demonstrate that SWC reduces the signaling load.

The remainder of the paper is structured as follows. Section II summarizes research related to MTC signaling overhead and proposed architectures. Section III presents the details of the proposed architecture and its functional entities. Section IV provides an analysis and evaluation of the proposed architecture with preliminary numerical results.

II. RELATED WORK

A considerable amount of the research work related to the MTC 26B context falls into the category of Radio Access Network (RAN) overload and congestion control. A set of RAN enhancements are proposed in 3GPP TR 37.869 [7] pertaining to Small Data Transmission (SDT) that principally addressed low bandwidth use cases. Some of the proposals are specifically targeted to reduce the signaling overhead over the air interface with direct impact on the UE technology (e.g., small data transfer using Radio Resource Control (RRC) messages). These RAN procedure simplifications have led to research efforts which optimize the packet core and related procedures. As an example, a series of EPC congestion scenarios with potential solutions and recommendations are proposed in 3GPP TR 23.843 [8].

To address some of the challenges observed in the proposed architecture of 3GPP TS 23.682 [5] related to the overload of signaling, the authors of [11] propose a congestion mechanism to be implemented by the MME. The authors have identified the challenges caused by high rate of MTC triggering via the Service Capability Server (SCS) and the MTC Interworking Function (MTC-IWF) that can lead to MME overload. Therefore, a congestion mechanism implementing a direct notification from the MME to MTC-IWF is proposed in [11]. This should further be used by the MTC-IWF to reject or aggregate MTC triggering events. Another approach to increase the MME capacity as suggested in [15] is profiling UEs and MTC devices into classes and processing their signaling through aggregation schemes. The authors of [15] confirm that many of the Information Elements (IEs) are common to all MME messages and propose a mechanism to aggregate many IEs into a "profile IE" associated to a group of devices. Although the main drawback of this scheme is identified as being the increased processing time by the message aggregator, the MME has less messages to process. Furthermore, it is proposed in [15] that the MTC-IWF be directly involved in paging low mobility MTC devices. The MTC-IWF sends paging requests to the enhanced NodeB (eNB) after receiving "low-mobility UE" information from the Home Subscriber Server (HSS).

Network programmability paradigms have also been proposed to alleviate the signaling of the EPC control plane. Not directly related to the MTC applications, the work in [13] uses OpenFlow protocol to replace the MME-eNB and the MME-Serving Gateway (SGW) control protocols. In [14], the authors explore the problem of packet core virtualization and show how the user plane capacity is impacted by the SGW especially when it has to handle high rates of control plane packets. Other solutions are based on optimized routing. For instance, a new bearer model is proposed in [12] which reduces bearer management signaling by setting up multiple "virtual" bearers at once. This model associates up to three IPv6 addresses per default bearer. Each IP address identifies a different breakout point (i.e., eNB, S-GW or P-GW). The Packet Data Convergence Protocol (PDCP) is modified to support new Packet Data Unit (PDU) types which are mapped to each IPv6 address. The eNB uses these PDU types to route the traffic without doing an IP lookup.

III. PROPOSED SOLUTION

In this section, we delve into the SWC general architecture model and show how the SWC design satisfies our objective of providing optimized connectivity for M-MTC devices. We define 3 communication pattern categories, namely,

1) Low frequency and low bandwidth
2) Low frequency and high bandwidth
3) High frequency and low bandwidth

Note that high frequency and high bandwidth devices are standard UEs and therefore, are handled as such. In our proposed SWC solution, we ensure co-existence with legacy RAN and EPC. This architecture also leverages IPv6 mechanisms and intrinsically enables local breakout. For the SWC architecture, we primarily address the issues prevalent to M-MTC which includes devices such as:

- Smart Wearables: Measure pressure, temperature, heart rate, blood pressure, body temperature, breathing rate and volume, skin moisture, etc.
- Sensor Networks: Metering (e.g., gas, energy, and water), city or building lights management, environment (e.g., pollution, temperature, humidity, noise) monitoring, and vehicle traffic control represent prominent examples of services in a smart city.
- Mobile Video Surveillance: may evolve to be available on aircrafts, drones, cars, and safety and security personnel for monitoring.
- Asset tracking devices and telemetry.

The objective of the proposed solution is to provide efficient, scalable and simplified core connectivity for a better network utilization through alleviation of user plane signaling overhead. In addition, SWC provides backward compatibility to support legacy devices. In order to meet these goals, SWC assumes the following:

- 5G/LTE and IPv6 are supported by the M-MTC devices.
- An M-MTC device hosts only MTC applications which benefit from our connectivity proposal. IPv6 multi-addressing could be used, for instance, to bind different

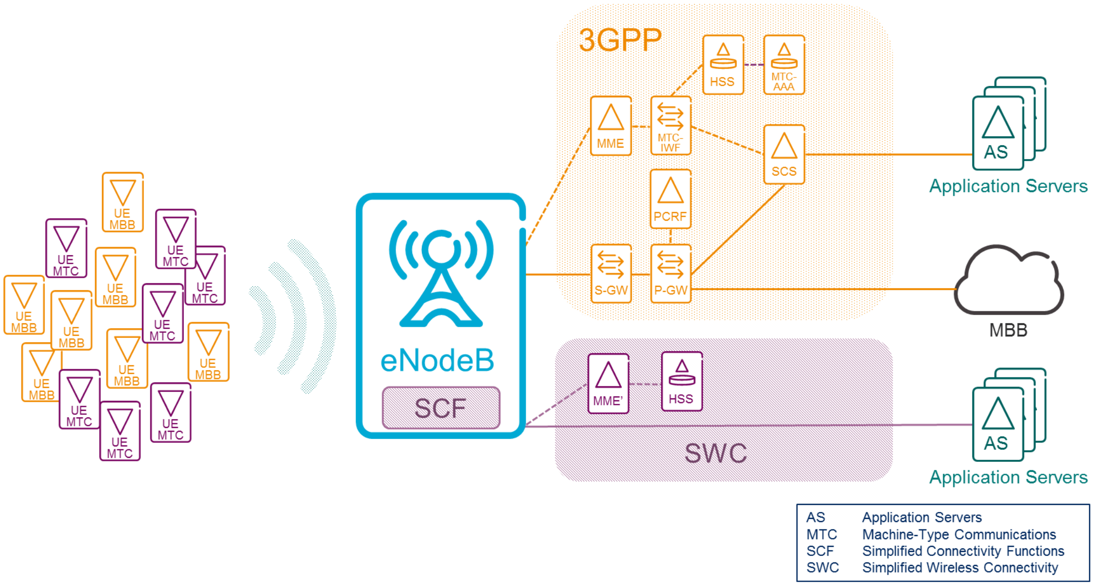

Fig. 1. Proposed SWC architecture with main functional entities. The dual personality base station is capable of handling different types of devices.

MTC applications to separate IPv6 Global Unicast Addresses (GUA).

- A simplified approach for MTC device triggering is envisioned. Although MTC devices may be mobile, session continuity may not be necessary and therefore X2 or S1 handover is not necessary. In addition, nomadicity does not require that the packet data protocol (PDP) context be preserved. Upon attaching to a new eNB, a new IPv6 session will be established and follows the enhanced re-attachment procedure through Simplified Connectivity Function (SCF) at the eNB.

A. Architecture Model

Figure 1 provides an overview of the SWC architecture and the main functional entities.

1) Simplified Connectivity Function (SCF): is the main functional entity in SWC. Every SWC-enabled eNB runs an SCF module that contains basic IPv6 node and router functionalities. It provides simple connectivity through a transparent SGi/Gi like reference point to the Internet. SCF advertises itself to all M-MTC devices served by the current eNB as the first-hop IP router (e.g., send solicited and unsolicited Router Advertisements). The M-MTC devices use IPv6 stateless address auto-configuration (SLAAC) [3]. SCF should also be able to relay or respond to DHCPv6 requests from M-MTC devices for other configuration options. Additionally, SCF provides the following services:

- SCF maintains a mapping of M-MTC devices to IPv6 address(es). Every M-MTC device will have a DRB established for all the traffic generated by the MTC applications. Therefore, SCF will map a DRB to the IP sessions of the device. SCF maintains a mapping relation UE_ID ↔ IPv6 address, where UE_ID could include UE identifiers usually available on the eNB, such as Cell Radio Network Temporary Identifier (C-RNTI) that identifies a device in the cell, System Architecture Evolution (SAE)-Temporary Mobile Subscriber Identity (S-TMSI) or International Mobile Subscriber Identity (IMSI) (used in RRC connection requests), E-UTRAN Cell Global Identifier (ECGI), etc.
- SCF intrinsically implements local breakout because of the SCF routing functionality.

In uplink, SCF ensures fast traffic steering for the DRB data associated to the SCF context based on the IP header. In downlink, SCF relays the data to the right active DRB of the connected UE. The IP connectivity perceived by the UE is depicted in Figure 2. This illustrates an end-to-end path from the UE to a Peer Entity via the eNB/SCF. The SCF forwards the traffic between the Radio Bearers and IP connections.

To support charging, monitoring, authentication and authorization we envision a simplified auxiliary function module capable of supporting multiple SCFs. This module can be a standalone entity or co-located with MME' appropriately dimensioned for the required bandwidth.

2) MME': is a MME with reduced functionality. It provides only the functions which are essential for SWC including the following functionalities:

- NAS signaling without the EPS Session Management (ESM) procedures. This is because SWC no longer requires the EPS bearer management procedures as the interface to SGW is no longer present.
- The interface with HSS for authentication and authorization purposes is maintained; the NAS security procedures are still required in SWC.
- Paging procedure is required for UE reachability.

MME' triggers the setup of the UE SCF Context at the base station based on either: (1) an explicit indication from the UE via the initial Attach Request including a newly defined field for the SWC support within the UE capabilities IE; or (2) the subscription profile from HSS indicating that the UE is SWC capable.

In the proceeding sections, we further describe the SWC enhancements for M-MTC device deployment.

B. User Plane Setup

Before the UE can send or receive data from the AS, it must first acquire an IPv6 address. Upon the initial attach, the UE - MME' NAS signaling will indicate that the UE is to be served by the SWC and not the legacy core. The MME' does not trigger ESM procedures leading to EPS GTP-based bearer establishment. Once the DRB is established, the UE auto-configures (SLAAC) its IPv6 global address with the prefix acquired from SCF. SCF manages a pool of IPv6 prefixes as defined by the operator's policies and maintains the UE SCF context. This context includes the mapping between UE identifiers available on the eNB and the UE IPv6 address. The SCF at the eNB steers the traffic based on the configured IPv6 address.

The UE is now ready to register its IP address with the AS. Depending on the UE technology, it can learn the IPv6 address of the AS by either hardcoding this information in the UE, using a Fully Qualified Domain Name (FQDN), or via stateless DHCPv6 exchange with the SCF.

The UE context at the SCF is maintained while the UE is in idle state. The AS uses the global IPv6 address of a device to initiate a trigger. An idle UE can also be paged directly by the eNB that has maintained the last valid UE context at the SCF. Downlink data or the trigger requests are buffered at the SCF/eNB which pages the target UE (e.g., using its S-TMSI). In case of failure to reach the UE, the MME' is notified to proceed with paging the tracking area.

The UE SCF context and its IPv6 address (and prefix) change when a UE attaches to a new eNB due to link failures, retransmission or nomadicity (device on the move). Upon a re-attach to the mobile network, the M-MTC device must re-register its new IP address with the AS. RAN events such as link failures and retransmission indications may cause the UE to attach to a new eNB and trigger a new IPv6 auto-configuration [6].

IV. ANALYSIS AND EVALUATION

In this section, an analytical evaluation of the signaling load performance for mobile originating and mobile terminating traffic is presented. Since this is a work in progress, presented results are preliminary. For the case of M-MTC, signaling floods are inevitable in scenarios such as power failures and shutdowns at the eNBs, flood of notifications or attach requests by a large number of devices, or simultaneous attempts by multiple MTC AS to trigger an event for a large group of MTC devices. Device triggering is required when the IP address of a MTC device is not known or the device is not reachable by the Service Capability Server (SCS) or AS. The proposed solution by 3GPP as stated in TS 23.682 [5] suggests that device triggering can be performed over Tsp and T5 interfaces that may eventually lead to an MME overload and control plane congestion. A M-MTC device will transition from Idle to Connected state under the following circumstances:

- MTC Mobile Originated Traffic. For example, a periodic MTC timer that triggers Uplink (UL) data to send.
- MTC Mobile Terminating Traffic originated by the network. For example, trigger which causes UL or Downlink (DL) data transmission such as on-demand sensors (e.g., hydro meter UL data) or actuators (e.g., DL data to a light switch).

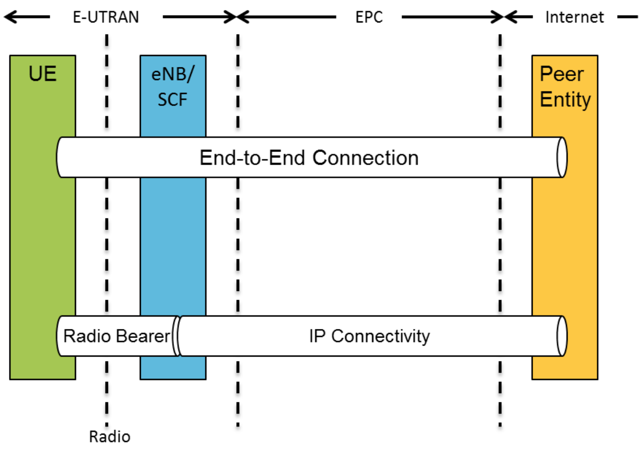

Fig. 2. IP Connectivity as perceived by the UE through SCF.

In the case for SDT, triggering may be over NAS and the MTC device response is transmitted either over NAS or user plane. Numerous optimizations are suggested in the technical recommendation document 3GPP TR 23.887 [9].

A. Model Formulation

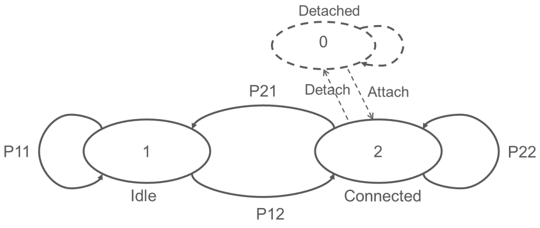

Fig. 3. States of a UE as seen by the core network signaling.

In order to quantify the signaling cost, we consider two EPS Connection Management (ECM) states for the UE after the UE is attached, i.e., the UE can be in idle or connected states. Under the assumption of modeling as depicted in Figure 3, the cost of signaling can be expressed as follows.

$$N_T = N_s \sum_{i,j} C_{ij}.P_{ij}$$

where N_s is the number of subscribers (devices), P_{ij} is the probability of transition from state i to state j and C_{ij} is the cost of signaling for state (i, j) and calculated as $C_{ij} = N_{ij}.\lambda$, where N_{ij} is the number of messages and λ is the aggregate of all Poisson [10] data arrivals (mobile originating or mobile terminating) where $\lambda = \lambda_1 + \lambda_2 + ...\lambda_k$ for k independent application types. The theorem [1] stated in [10] forms the basis of this assumption.

To evaluate the performance of signaling in the proposed architecture, we consider the 3GPP architecture as in TS 23.682 [5] and signaling cost as a baseline for the performance benchmark. The signaling performance analysis is done based on the number of messages as well as the number of bytes. Three modes of operation for the MTC are envisioned by 3GPP namely, *direct mode*, *indirect mode* and *hybrid mode* [5]. We consider the indirect mode of operation throughout the analysis in the paper.

There are several factors that cause the release of bearers such as UE transition to idle mode as a result of inactivity timer expiry, radio link failure, RRC connection re-establishment rejection and Tracking Area Update (TAU). We take into account only the first cause that is the idle mode transition with inactivity timer expiry.

The UE inactivity timer is set at eNB as one of the configuration parameters. If there is no activity within a time frame T_i the inactivity timer triggers the idle transition and the DRB resources will be released. To determine the expected cost of signaling caused by the transitions from connected to idle, it is necessary to determine the probability that such a transition will happen.

We introduce a new event as shown in Figure 4 to represent the transition from “connected” to “idle”. The idle transition event will start T_i units of time after the last signaling event is processed. The idle trigger event follows the same trend as the arrival rate of other events. As mentioned previously, events are represented as Poisson processes with aggregated arrival rate of λ and session duration of exponential distribution with parameter μ and average $\frac{1}{\mu}$. We denote the inactivity timer expiry event as a Poisson process of rate λ_i that is consistent with the arrival of signaling traffic. Successful trigger of inactivity timer T_i expiry requires that $Z > \frac{1}{\mu} + T_i$.

We make the following assumptions throughout the analysis. We model eNB as a system of M/M/∞ [10] characteristic that describes a system with Poisson process on arrivals and exponentially distributed service time, assuming there are infinitely many servers, so signaling arrivals immediately follow the processing without waiting in any queue.

Considering the signaling arrivals as a Poisson process of rate λ and $x = \frac{1}{\mu}$, the length of the interval from x until the next arrival will be independent of the event occupying $[t_1, x] = \frac{1}{\mu}$. Therefore we can derive the idle transition probability as follows.

$$Pr(Idle \quad Transition) = Pr(Z > \frac{1}{\mu} + T_i)$$

$$Pr(Idle \quad Transition) = e^{-\lambda(\frac{1}{\mu} + T_i)}.$$

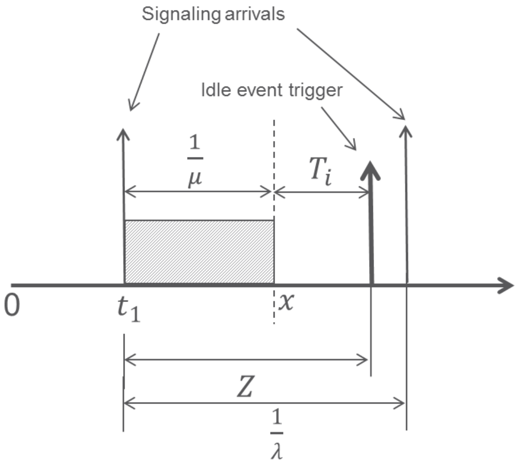

Fig. 4. Modeling the idle transition event.

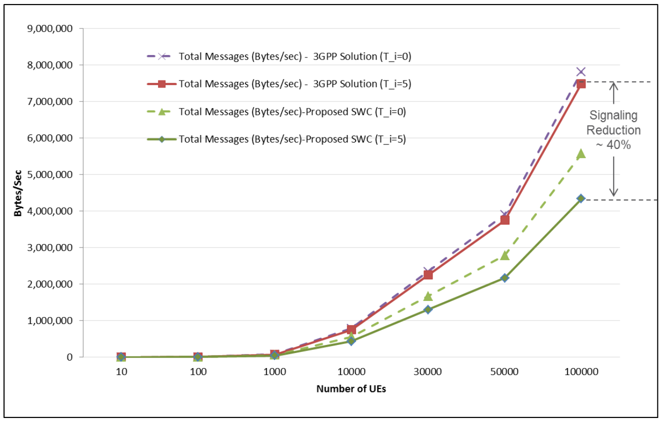

Fig. 5. Signaling load in bytes/sec as the number of UEs increases ($\lambda = 0.05$ and $\mu = 0.1$).

Expanding this by a power series we have

$$Pr(Idle \quad Transition) = \frac{\mu}{\mu + \lambda + \lambda\mu T_i} + o(T_i),$$

where $\lim_{T_i \to 0} \frac{o(T_i)}{T_i} = 0$.

B. Signaling Cost Evaluation Results

We demonstrate the performance of the proposed architecture based on data collected from real network traces and using the indirect mode as stated in the 3GPP TS 23.682 [5]. The numerical results are based on the parameters as shown in Table I.

Parameter	Value
Pr(Mobile Terminating Traffic)	0.8
Pr(Mobile Originating Traffic)	0.2
λ	0.05
μ	0.1
T_i (sec)	0 and 5

TABLE I
NUMERICAL RESULT PARAMETERS.

---

[1] If $\{N_1(t); t > 0\}$ and $\{N_2(t); t > 0\}$ are independent Poisson processes of rate λ_1 and λ_2, and $N(t) = N_1(t) + N_2(t)$, $\forall t > 0$, then $\{N(t); t > 0\}$ is a Poisson process of rate $\lambda = \lambda_1 + \lambda_2$.

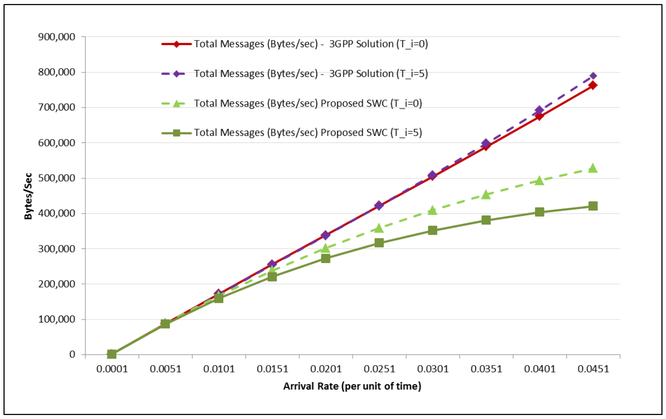

Fig. 6. Signaling load in bytes/sec w.r.t signaling arrival rates (Number of UEs=10,000).

Event (SCTP Payload)	**Size (bytes)**
S1AP: DL NAS Transport	142
S1AP: UL NAS Transport	158
Paging over S1AP	43
Initial UE Message S1AP	60
DL NAS S1AP	142
UL NAS S1AP	158
UE Context Release Request	29
UE Context Release Command	25
UE Context Release Complete	23
Authentication Information Request Diameter	356
Authentication Information Response Diameter	372
Authentication Request S1AP	64
Authentication Response S1AP	61
Security Mode Command	42
Security Mode Command Complete	69

TABLE II

MESSAGE SIZES CAPTURED BY REAL NETWORK TRACES.

The graphs of Figure 5 and Figure 6 show the comparison of the proposed architecture signaling load in bytes/sec with the baseline scenario. In Figure 5, we show the comparison of the proposed SWC and the 3GPP as the number of MTC devices increases. This result is based on $\lambda = 0.05$ and $\mu = 0.1$. Graph of Figure 6 shows the comparison of signaling load with respect to signaling arrival rate as the number of UEs is 10,000. The choice of the inactivity timer (T_i) parameter is implementation specific. While a zero value reduces the power consumption of the device, it causes more signaling load on the network. We ran our analysis with different values of T_i and observed similar improvements. For illustration purposes, we show the result for both $T_i = 0$s and $T_i = 5$s as shown in Table I. Examples of message sizes for some events are shown in Table II.

The performance gains may be attributed to:

- Elimination of SGW and PGW nodes and the associated signaling load due to bearer management. The GTP tunnels are replaced with IP connectivity between the SCF at eNB and the end point.
- Reduction of signaling required for session continuity (i.e., no support required for handovers).

V. CONCLUSION

This paper identified the challenges caused by the rapid growth of M-MTC traffic for the 26B challenge. To support the networked society and cope with the proliferation of applications supporting massive MTC devices, a simple, efficient, and scalable communication system is required.

In this paper, a Simplified Wireless Connectivity (SWC) approach based on IPv6 is proposed. The objective of the proposed architecture is to simplify and alleviate the signaling load through elimination of unnecessary functions and an optimized IPv6 connectivity at the base station. Furthermore, the essence of the proposed solution is to support the co-existence and interworking with the legacy mobile devices by a dual personality base station. The architecture and connectivity model were presented with a description of the main functional entities. The feasibility and advantages of the proposed solution are also illustrated by analytical reasoning. The numerical results demonstrate a significant reduction in network signaling load.

REFERENCES

[1] Ericsson Mobility Report, June 2015.
[2] NGMN Alliance, Next Generation Mobile Networks, White paper (2015).
[3] Narten, T., Thomson, S., and Jinmei, T., IPv6 stateless address autoconfiguration, IETF RFC 4862, 2007.
[4] 3GPP TS 22.368, Service requirements for Machine-Type Communications (MTC); Stage 1 (Rel. 13). Dec-2014.
[5] 3GPP TS 23.682, Architecture enhancements to facilitate communications with packet data networks and applications (Rel. 13). March 2015.
[6] Krishnan, S., and G. Daley, Simple Procedures for Detecting Network Attachment in IPv6. No. RFC 6059, 2010.
[7] 3GPP TR 37.869, Study on enhancements to machine-type-communication (MTC) and other mobile data applications enhancements, 2013.
[8] 3GPP TR 23.843, Study on Core Network Overload Solutions, 2013.
[9] 3GPP TR 23.887, Study on Machine-Type Communications (MTC) and other mobile data applications communications enhancements V12.0.0,2013.
[10] Gallager, R. G., Stochastic processes: theory for applications, Cambridge University Press, 2013.
[11] Ksentini, A., Taleb, T., Ge, X., and Honglin, H., Congestion-aware MTC device triggering, In *IEEE International Conference on Communications (ICC)*, 2014 (pp. 294-298).
[12] Korhonen, J., Savolainen, T., Wolfner, G., and Laganier, J., Evolving the 3GPP bearer model towards multiple IPv6 prefixes and next-hop routers, Telecommunication Systems, 59(2), 193-209.
[13] Sama, M. R., Ben Hadj Said, S., Guillouard, K., and Suciu, L., Enabling network programmability in LTE/EPC architecture using OpenFlow, *12th IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2014 (pp. 389-396).
[14] Rajan, A. S., Gobriel, S., Maciocco, C., Ramia, K. B., Kapury, S., Singhy, A.,and Janaz, R., Understanding the bottlenecks in virtualizing cellular core network functions, *IEEE International Workshop on Local and Metropolitan Area Networks (LANMAN)*, 2015 (pp. 1-6).
[15] Taleb, T., and Ksentini, A., On alleviating MTC overload in EPS, In Ad hoc Networks Journal (2014).
[16] OECD, Machine-to-Machine Communications: Connecting Billions of Devices OECD Digital Economy Papers 192, OECD Publishing, January 2012.
[17] AT&T Network-ready Devices. [Online]. Available from http://www.att.com/modules (last accessed: Jan 8, 2016).