A First Look at Multi-Access Connectivity for Mobile Networking

Philipp S. Schmidt
Telekom Innovation Laboratories / TU Berlin
Berlin, Germany
philipp@net.t-labs.tu-berlin.de

Ruben Merz
Swisscom
Strategy & Innovation
Bern, Switzerland
ruben.merz@ieee.org

Anja Feldmann
Telekom Innovation Laboratories / TU Berlin
Berlin, Germany
anja@net.t-labs.tu-berlin.de

ABSTRACT

Multi-access connectivity in wireless networks is becoming widely available. Most handsets on the market today support connectivity to multiple access technologies such as HSPA, LTE or WiFi and can make use of some of them in parallel. By offering multiple access options, the available bandwidth and robustness of the connectivity can be increased. There is a large design space to take advantage of multiple connectivity. Various solutions and protocols at different layers and locations in the network exist, originating from different viewpoints and problem statements. They range from static low-layer tunneling approaches like IWLAN or Mobile-IP to dynamic transport layer approaches like multipath TCP up to highly sophisticated application-layer specific protocols.

Focusing on the network layer and above, we review a variety of these protocols and solutions by categorizing them, exhibiting their properties, reasoning about their potential behavior in the presence of Internet traffic and take a glimpse on their impact on the network.

Categories and Subject Descriptors

C.2.2 [Computer-Communication Networks]: Network Protocols

Keywords

Multi-Access Connectivity, Capacity Sharing, Fixed Mobile Convergence, Offloading

1. INTRODUCTION

Wireless connectivity has become ubiquitous. According to a recent report by Ericsson [1], global worldwide mobile penetration is around 87% in the first quarter of 2012. In Western Europe, cellular data connectivity is widely available with the advent of HSPA. With flat-rates, mobile subscribers are taking full advantage of their smartphones and tablets. The amount of data-traffic transported by second and third generation cellular infrastructures is surpassing all expectations and several indicators are showing an extremely strong growth for mobile data-traffic over the coming years [2, 1]. For mobile operators, this growth is proving cumbersome to manage. In particular, in urban areas, cellular access networks are showing increasing signs of congestion, especially at peak hour.

While the ongoing deployment of LTE, with a higher radio access capacity, is expected to help, it is acknowledged that there is also a need to more efficiently use the available infrastructure [3]. A recent trend has been WiFi offloading [4, 5, 6, 7], where mobile operators take advantage of their WiFi infrastructure and of terminals with heterogeneous access interfaces. When in range from a WiFi access point (AP), the terminal will offload all IP traffic to the WiFi interface, freeing up resources in the cellular infrastructure. But behind WiFi offloading is the more general idea of taking advantage of all connectivity options available, possibly concurrently [8, 9]. Especially in urban environment, a terminal will have several access infrastructures in range, offering multiple connectivity options, typically WiFi, 3G or 4G. On the client side, with the proliferation of smartphones and tablets, there is an increasing number of terminals that have the ability to take advantage of these options.

There exist several technologies and protocols that allow a terminal to take advantage of multiple interfaces. They come with various tradeoffs, in terms of control path and signaling, support of mobility or compatibility with current network deployments. They operate at different layers and location in the network. Furthermore, while it can be beneficial to have multiple interfaces available, the optimal way of using them remains an issue largely open. For instance, [8] showed that simply binding interfaces together does not achieve the best throughput. Multipath TCP could be used [10], but it is restricted to TCP and its performance in a wireless environment remains to be assessed with a realistic mobile traffic distribution [11].

In this paper, our goal is first to present an overview of the different solutions, protocols and technologies available to support multiple connectivity in a cellular access scenario. We characterize them mainly in terms of location in the networking stack, support for mobility and flexibility in terms of traffic scheduling, i.e. whether and how traffic can be explicitly directed to a particular interface. We then present a preliminary evaluation of a subset of these solutions (3GPP IFOM [7], Multipath TCP and application layer) which are deployable in today’s mobile networks. In particular, we fo-
2. AN EXPLORATION OF MULTI-ACCESS CONNECTIVITY APPROACHES

To categorize multi-access connectivity approaches, we take a flow-based perspective. The most fine-grained definition of flow we want to take into account is a five-tuple (i.e., source/destination address, transport-layer protocol and source/destination port) identifying a connection between two Internet hosts. From this viewpoint, we consider five basic dimensions for our categorization.

Flow Granularity One of the most important questions is at which granularity flows are assigned to access interfaces. We see three types of granularity: (type 1) all flows are assigned to only one access interface at a time e.g. for IWLAN, (type 2) one or more flows can be assigned to any available access interface e.g. for IFOM, and (type 3) a flow is split among any available access interface e.g. for multipath TCP or multipath RTP.

Stack Location We identify multi-access solution functioning at the network layer (e.g. IP level like 3GPP IFOM), transport layer (e.g. multipath TCP, SCTP, multipath RTP) and application layer (RTP within SIP/IMS sessions).

Mobility Because of the inherent mobile behavior of wireless access networks, it is important to classify any multi-access connectivity solution in terms of flow mobility. While some solutions can re-negotiate the available access at any time (e.g. multipath TCP, 3GPP IFOM and all Mobile IP variants), some can only do this negotiation at connection establishment (e.g. pre RFC 5061 SCTP) or not at all as for many application layer approaches which will just establish a new flow in case of mobility.

Concurrency Having several access interfaces available does not necessarily imply that they can be used at the same time. For instance, 3GPP approaches such as IWLAN and UMA can only use one connection at a time. This classification criteria is highly related to the flow granularity one, but we still believe it is necessary in itself.

Flow Scheduling Some solutions allow fine-grained control over the assignment of a flows to particular interfaces. This is the case for all application layer solutions, but e.g. not for multipath TCP where this could interfere with the desired fairness properties.

In addition, from an architectural point of view, some solutions require a junction point located in the network, typically for Mobile IP.

We now discuss several protocols and solutions to take advantage of multi-access connectivity. Our list is definitely not exhaustive and is somewhat centered on solutions applicable to a cellular access context, but tries to be representative for the possible choices for multi-access connectivity throughout the network stack. Our classification is ordered by the position in the networking stack, going from the network layer up to the application layer. For a brief summary, see Table 1.

2.1 A Note on Lower Layers Support

The lower layers of 3G, 4G and other wireless networks have support for multi-access connectivity of different grains and complexity. While wireless LAN and 2G networks provide simple switch-overs between base stations, 3G networks use soft handovers where during a transition to a new base-station, a terminal remains connected to the older one. For LTE, there is ongoing work on cooperative multi-point (CoMP). With CoMP, multi-access connectivity is available within that single transport technology.

While these technologies are designed to be transparent to the user’s terminal, its operating system and applications, this also prevents them to choose whether to use those technologies. In heterogeneous access scenarios, solutions below the network layer can not support transitions between different access technologies. Looking forward to choosing and combining multiple access solutions, we decided to exclude all solutions below the network layer from our comparison due to the limitations mentioned above.

2.2 Network Layer

IWLAN & UMA.

Interworking with wireless LAN (I-WLAN) [12] allows for interworking and mobility between a WiFi and a 3G/4G infrastructure. It applies to the so-called packet domain in 3GPP lingo, i.e. I-WLAN is not usable for voice traffic in the so-called circuit switch domain. It re-uses the 3G/4G authentication infrastructure. Through the WiFi AP, a trusted tunnel is established from the terminal back to the 3GPP core network of the mobile operator via the WLAN access gateway (WAG).

I-WLAN is a network layer solution. It allows for a so-called 3GPP access point name (APN, which roughly corresponds to an IP address) to be assigned or moved to a particular access interface. In that sense it offers a type 2 flow granularity and, in theory, allows for concurrent access. The classification in terms of mobility is a bit cumbersome. While an APN (and all its associated flows) can be moved to another access interface, it is not possible to move flows across APN. In term of flow scheduling, it offers very little because a flow, once associated with a particular APN, remains bound to this APN.

Unlicensed mobile access (UMA) is the commercial name of 3GPP generic access network (GAN). From an architecture point of view, it is very similar to I-WLAN. However, it can transport all traffic, whether from the circuit or packet domain, over a WiFi connection.

UMA is a network layer solution. It offers type 1 flow granularity and hence, no concurrent access. In terms of
Table 1: Comparison of Multi-Access Connectivity Solutions

<table>
<thead>
<tr>
<th>Mobility</th>
<th>Concurrency</th>
<th>Flow Scheduling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow Granularity</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>UMA</td>
<td>type 1 (per terminal)</td>
<td>yes</td>
</tr>
<tr>
<td>I-WLAN</td>
<td>type 2 (per APN/IP)</td>
<td>yes</td>
</tr>
<tr>
<td>IFOM</td>
<td>type 2 (per 5-tuple)</td>
<td>yes</td>
</tr>
<tr>
<td>Multipath TCP</td>
<td>type 3 (per connection)</td>
<td>yes</td>
</tr>
<tr>
<td>SCTP</td>
<td>type 3 (per connection)</td>
<td>yes</td>
</tr>
<tr>
<td>IMS with regular RTP</td>
<td>type 2 (per RTP stream)</td>
<td>yes</td>
</tr>
<tr>
<td>IMS with multipath RTP</td>
<td>type 3 (per RTP stream)</td>
<td>yes</td>
</tr>
<tr>
<td>HTTP based solution</td>
<td>type 2 / type 3 using ranges</td>
<td>partially</td>
</tr>
</tbody>
</table>

mobility, the entire traffic can be switched from one access interface to another. It offers no possibility for flow scheduling.

**Mobile IPv6 / 3GPP IFOM.**

IP flow mobility and seamless offload (IFOM) is based on dual stack mobile IPv6 [7]. Note that, unlike I-WLAN and UMA, it is a LTE Release 10 and onward solution. Hence, it is essentially not currently deployable. With IFOM, the terminal can simultaneously connect to a 3GPP access and a WiFi access. It can exchange different IP flows belonging to the same APN through different accesses. IFOM also enables seamless IP flow mobility, with IP flows belonging to the same or different applications being moved seamlessly between access interfaces.

IFOM is a network layer solution. It offers type 2 flow granularity and concurrent access. In terms of mobility, flows can be switched from an interface to another. Flow scheduling can be done flexibly on a per-flow basis.

We ignore solutions like 3GPP LIPA and SIPTO, because they are not multiple access solution from the terminal perspective ([7] see for a comparison).

**2.3 Transport Layer**

**Multipath TCP.**

Multipath TCP [13] recently gained a lot of attention from the research community. It distributes the traffic of a TCP stream among multiple paths, performs congestion control for each TCP sub-stream and can provide fairness between multipath TCP and regular TCP streams sharing the same bottleneck [13]. With multipath TCP a sub-stream can be moved to another access interface and the IP address can be updated.

Multipath TCP is a transport layer solution. It offers type 3 flow granularity with concurrent access. It fully supports mobility since sub-streams can be seamlessly associated to another access interface, or have the IP address of any endpoint changed. But in terms of flow scheduling, it offers no support. The sub-flow assignment is automatically handled.

**SCTP.**

SCTP (see RFC 2960) is a reliable in-order datagram stream transport protocol originally intended for telecommunications signaling. It supports concurrent use of multiple access links and allows the application to influence the access interface selections on a datagram stream basis. It also supports mobility via access renegotiation (RFC 5061). Hence, SCTP is a transport layer solution. It offers type 3 flow granularity with concurrency, but with more flow scheduling control than multipath TCP. It also fully supports mobility, in a similar fashion than multipath TCP.

**2.4 Application Layer**

**Streaming with IMS and MPRTP.**

SIP (RFC 3261) and the IMS architecture built upon it are signaling protocols to negotiate media sessions (typically phone calls and streaming sessions). Interestingly, they can support the concurrent use of multiple contact IP addresses for the (concurrent) registration of one or more IMS user agents on terminals. This allows for the creation of multiple IMS signaling paths to terminals. On these signaling paths, signaling messages carry session description protocol (SDP) messages to negotiate media streams (i.e. calls or streaming). SDP allows for the (re-)negotiation of the streams of one media session over multiple paths. In turn, this enables application layer multipath transport.

From this point of view, IMS can offer application layer multipath support with a type 2 flow granularity and concurrent access. From a mobility and flow scheduling standpoint, it is also extremely flexible.

For multimedia streams, it is expected that RTP is used, which only supports a single path. A multipath extension to RTP is currently under discussion within the IETF [14].

Multipath RTP can offer type 3 flow granularity with concurrent access, can support mobility (via IMS, SDP signaling or the RTP control protocol)\(^1\). Multipath RTP then keeps statistics for each subflow, distributes the stream accordingly and reports an aggregate to the application to adopt the stream rate. Flow scheduling can be very flexible thanks to SDP.

**HTTP.**

Finally, generic multi access connectivity support can be provided at the application layer when HTTP is used. HTTP is by far the most popular application layer protocol in the Internet today. Because of the mostly stateless nature of its request/replay scheme, it is fairly straightforward to issue requests for a given application over multiple access inter-

\(^1\)But an IMS multimedia-session may already consist of multiple RTP streams (e.g. one for audio and one for video), note that then, multipath RTP could allow for the transparent distribution of any single stream over multiple links without the need for the application to initiate the streams to take care of.

\(^2\)For multipath RTP, there is currently no discussion on how to choose paths except the proposed use of Interactive Connectivity Establishment (ICE, RFC 5245) or using re-negotiation with SIP.
faces. In addition to that, using ranged requests it is possible to stripe the request of a large object over multiple available paths.

With HTTP (at the application layer), type 2 flow granularity with concurrency is possible. Flow scheduling can be also very flexible. But note that, as for IMS, HTTP has no information about the underlying state of the access interface. Mobility is, strictly speaking, not supported. But it can be emulated by stopping a flow connection on one access interface and using a range request to open a new flow connection on another access interface.

3. PRELIMINARY EVALUATION: COMPARING IFOM, MULTIPATH TCP AND APPLICATION LAYER SOLUTION VIA HTTP

From the set of protocols and architecture presented in the previous section, we decided to focus on 3GPP IFOM, Multipath TCP and a generic HTTP-based application layer support for this preliminary evaluation. We propose four dimensions in which multi-access connectivity approaches should be evaluated:

Latency Overhead This captures how much latency can be added, e.g. by tunneling to a junction point, or removed, e.g. by using an access interface with a lower latency per se. This has impact on user experience and throughput. To quantify the latency overhead, we take the difference between the latency with and without the multi-access connectivity solution. When a multi-access connectivity solution at a lower layer uses a junction point, we need to compare the latency overhead of that solution against using the available access interfaces natively at higher layers. Because we mainly focus on cellular carrier scenarios, this can have surprising effects.

Data Overhead This captures the additional bytes transmitted by in-band signaling like additional headers or additional encapsulations for tunneling. We do not take out of band signaling like tunnel setups in the cellular backhaul into this dimension but treat it as part of the network impact. While this is the most obvious dimension, it is probably not the most important one because adding a few bytes to packets has a lower network impact on wireless networks than sending additional packets.

Network Impact This is the most difficult property to capture for multi-access connectivity technologies. Indeed, it is very versatile and hard to put into a metric. The main aspects in this category are how any multi-access technology influences signaling overhead and how a solution behaves towards other flows in the presence of congestion on a common path. In this paper, we only approach this aspect from a high level perspective.

Flow-Size Effects Multi-access technologies might behave differently for different flow characteristics. We distinguish between the behavior towards short-lived flows (only few packets in each flow, e.g. DNS queries or HTTP dialogs never leaving initial TCP slow start), interactive flows which are long lived but are not limited by the bandwidth of the links and elephant flows which are long lived and are limited by the bandwidth of the links (see [15] for more applications of this classification).

For the evaluation we assume a scenario with a terminal connected through 3G and WiFi. Even though LTE/4G networks have a much lower delay than 3G networks, we stick to the 3G case because we believe 3G networks will remain a relevant case for the coming years. We assume the application and/or operating system has a choice to select the outgoing interface (3G or WiFi) and whether to use multiple access extensions for all layers above the network layer.

3.1 Network Layer: 3GPP IFOM with Dual Stack Mobile IPv6

All variants of Mobile IP³ introduce additional delay by going through a central junction point. The high delay of 2G/3G networks, compared to DSL or PON lines (behind a WiFi access), might negate this effect. With some unrepresentative measurements, we estimate using IFOM over DSL with a 3G provider as home network will add about 30 ms additional delay compared with a pure DSL access, but save about 50 ms compared to pure 3G access. Depending on the Mobile IPv6 variant used, the tunneling overhead is between 24 bytes (IPv6 Routing header type 2 only; see RFC 6275) and 68 bytes per packet (using RFC 6618 UDP encapsulation with AES as cypher). The network impact and flow size effects largely depend on the distribution of flows between the accesses, but are not standardized in any way.

While this allows for optimization that can benefit the network and has only a few bytes tunneling overhead, the use of a common junction point imposes an additional delay compared with higher layer approaches.

3.2 Transport Layer: Multipath TCP

Multipath TCP directly communicates to the endpoint over multiple paths, so it does not impose or save any latency. It adds data overhead at various stages of the protocol: 20 bytes for any SYN/ACK/SYN+ACK packet, an additional TCP handshake per path and 28 bytes each time part of the TCP connections sequence number space is mapped to some path, which is negligible for elephant flows and most interactive flows, but is unreasonable high for small flows. The developers of multipath TCP put great effort to make it TCP friendly. Multipath TCP doubles the TCP congestion control for the whole connection and each sub-flow, so multipath TCP flow should not harm normal TCP when sharing a common bottleneck. Depending on the interactions with lower layers, the overall network impact can be good or bad, e.g. by causing signaling traffic to re-establish cellular connections.

3.3 Application Layer: HTTP-based Solution

Providing multi-access connectivity at the application layer allows for a great granularity and does not impose additional latency and byte overhead. The drawback of application layer solutions is that they have to be implemented

³If not using a corresponding node, which is unlikely to happen because it reduces the control of the operator and is very hard to deploy without serious security impacts.
within the application, a protocol aware middle-box, or an application-level proxy. In all three cases it is hard to react to the conditions of the network. The network impact, if not externally controlled, is most likely negative. Two TCP connections to the same server take twice the bandwidth of a single one in the presence of congestion at a common point in the path to the server and are therefore unfair. Depending on the type of flows, different strategies might be useful: While elephant flows can be split over multiple ranged requests, small and interactive flows will likely profit from choosing the path with the lowest latency and packet loss.

3.4 Conclusion

Multi-access connectivity technologies exhibit different advantages depending on their location in the networking stack: While the protocols in the lower layers mostly support seamless mobility of coarse flow aggregates without negative network impact, application layer protocols allow for fine-grained distribution of flows, but suffer from uncertain network impact. Transport layer solutions like multipath TCP can provide fairness, but lack network awareness towards both the underlying network and the applications needs of different flow characteristics.

4. OUTLOOK & FUTURE WORK

In this paper, we have focused mainly on standardized solutions for multi-access connectivity. There are several related works that are of interest. In particular, intentional networking [16] is based on the idea that an application might label its needs on a per-socket basis. The system uses an external probing tool to discover networks and provides a modified socket API to applications which uses their own transport abstraction to connect to the server over the most appropriate way. The PERM framework in [17] uses the RTT and completion time of TCP connections as the basis for a heuristic to schedule TCP connections over multiple available WLAN networks to increase the bandwidth by using neighboring available networks. While this is potentially positive for the users experience, the network impact remains unclear because the access interfaces might share a common bottleneck. In [18], three scheduling schemes for distributing new TCP flows over multiple WLAN access are evaluated based on real network traces. While this mimics possible application layer approaches, it is implemented as a transparent addition to the socket library maintaining full compatibility to normal TCP servers. Another relevant question is how fairness in the presence of congestion can be realized with multi-access clients. A game-theoretical modeling can be found in [19].

Finally, there have been approaches from a peer-to-peer network point of view. For instance, ARIBA is a framework to build overlay systems that can exploit multi-access connectivity [20].

For future work, our current priority goes towards a more in-depth and thorough system evaluation using a combination of testbeds and emulation. Furthermore, we would like to explore the design space of application layer solutions when information about the underlying connections become available. The assignment of flows over a particular access interface is a problem that remains largely open and its dependence on traffic characteristics is challenging and exciting.

5. REFERENCES


