GEO-LOCALIZATION TO ENHANCE SCTP HANDOVER IN PUBLIC TRANSPORTS BETWEEN SATELLITE NETWORKS AND WLAN

Lionel Bertaux[1], Pascal Berthou[1], Thierry Gayraud[1], Paul D. Amer[2]
[1] CNRS ; LAAS ; 7 avenue du colonel Roche, F-31077 Toulouse, France
Université de Toulouse ; UPS, INSA, INP, ISAE; F-31077 Toulouse, France
[2] Computer and Information Sciences Department; Protocol Engineering Laboratory;
University of Delaware; Newark; DE 19716-2586
{lbertaux, berthou, gayraud}@laas.fr; amer@cis.udel.edu

Abstract

The rise of new ways of communication along with the spread of mobile Internet will encourage the evolution of public transports to provide uninterrupted Internet service to its customers. It is obvious that using multiple technologies such as Wi-Fi, 3G and satellite each with their own characteristics will enhance the connectivity. Current proposals for mobility are mainly based on the use of proxies and need the modification of the infrastructure. The multi-homing feature included in SCTP allows roaming without losses and without interrupting the session. This study focuses on the use of geo-localization to improve SCTP handover, and shows how it can be implemented and what improvements can be done.
KEYWORDS: geo-localization, handover, mobility, multi-homing, public transports, SCTP.

1. Introduction

Recent significant growth in mobile technologies along with the development of Internet services has increased the needs of travelling customers. In an “always connected” world, public transports (trains, busses, planes, etc.) have to offer more services to their users (emails, timetables, multimedia streaming, etc.) and also Internet access to company databases for data backup, real-time traffic information, etc. Having contiguous connectivity involves the use of several access technologies. This study considers the two main ones compatible with public transports constraints of cost, mobility, broadband, etc.: satellite networks and wireless terrestrial access points. To enhance the vertical handover between these two technologies, we propose to exploit known characteristics of the public transports: predetermined travel routes and geo-localization using on-board GPS and odometers.

Handovers (vertical or horizontal) impact communications at multiple layers. Network and below layers are concerned by handover since they are responsible of connecting to Access Points (APs) and reaching remote hosts. But the most impacted layer could be transport layer. Sessions are bound to transport connections or associations. If a connection goes down, the session fails and the application will need to establish a new session. To avoid such failures, network layer solutions provide uninterrupted connectivity to transport layer, but these solutions’ requirements can be fairly high in term of architecture modifications or mandatory mechanisms. In this paper, we choose to focus on SCTP, a transport protocol that can provide mobility through its multi-homing feature and its extensions.

The first part of this paper introduces the context of public transport with their characteristics and their mobility scheme. The second section presents existing mobile IP solutions, and explains why transport mobility solutions are a credible alternative. In part three, the proposed solutions are depicted with a time analysis of the handover and a brief description of the implementation. The fourth part includes the test scenario, the results and their analysis to demonstrate the enhancements that can be done to the handover delay with SCTP. The last part concludes with an evaluation of the proposed solution and a statement of future work.

2. Mobility in Public Transports

This section introduces public transports listing their key characteristics and describing their mobility schemes. A short description of conceivable access technologies is also provided.

2.1. Context

Public transports are going to evolve and provide more services to customers. One challenge is to have a continuous Internet connection onboard. To design a mobile network architecture fitted for public transports, it is essential to understand their characteristics:

- Vehicles have a fixed run, thus their localization is known,
- Network traffic is composed of flows with different requirements (low delay, large bandwidth, connectivity, etc),
- Vehicles are equipped with multiple communication technologies,
- There is a need for autonomous network configuration.

Knowing vehicle run and its localization can help to predict mobility. Our proposition is to use a "map" that lists all available networks along a vehicle's run to predict handovers based on current vehicle localization. An autonomous mobility solution based on this map will be detailed in section 4.

An Internet connection onboard a public transport would have several uses: backup data from the vehicle operator, critical information related to vehicle, web traffic and even VoIP. Such flows have different QoS (Quality of Service) requirements and priority levels. The access point choice should be done according to these requirements and restrictions should be applied if all requirements cannot be met. Communications can also be enhanced in case several networks are available: depending on their characteristics they can provide lower delay, higher bandwidth or larger coverage. Influence of network characteristics is detailed in section 2.2.

When changing access network, latency is introduced by a layer 2 handover and the network configuration. Therefore, performing a handover should be done only if necessary (i.e. if the handover enhances communications) and at the right time. Example: if a mobile node is connected to 3G and heads toward a Wi-Fi AP, performing a handover will be useful only if the node spends more time in range of the Wi-Fi than the latency introduced by handover. In addition, the handover method used needs to be taken into account: it defines the latency introduced by handover and it also impacts on communications efficiency. This matter is treated in section 3.

2.2. Communication Technologies

In public transport networks, several communication technologies are available. Since each technology has its assets and its downside, the physical and data Link layers will affect communications differently. Network efficiency depends on the following parameters: available bandwidth, end-to-end delay, QoS provided and service availability. Besides network characteristics, technology operating cost can be taken into account; providers can charge depending on data exchanges.

For information purposes, Table 1 presents parameter values for three communication technologies: Wi-Fi, 3G/UMTS and satellite communication network. Instead of considering such characteristics independently, it is important to keep a global view of available supports. An example is given by [SOR 2009] based on Diesel-Net experimentations. Even if Wi-Fi has a higher bandwidth than 3G, a larger global throughput is achieved on 3G due to its overall availability. But since users will be charged depending on data exchanges, it can be useless to use 3G permanently for non-critical applications.

	Bandwidth	Delay	Quality of Service	Availability	Cost
Wi-Fi	54 to 100Mbps	Low	802.11e	Range 100 to 200m	Low
3G/UMTS	Contract 2Mbps	Average Propagation	Depending Infrastructure	Range 50m to 3km	Amount Data
Satellite	Contract 1Mbps	Geo: 300ms Leo: 100ms	Depending Contract	Depending obstacles	High

Table 1 - Wi-Fi, 3G/UMTS and Satellite characteristics.

3. Solutions for mobile networks

As explained in section 2, public transports can keep communications alive by connecting to multiple networks but in most cases, changing the access network modifies the IP address of the node. Since transport protocols use the IP address and port to identify a node, communications can break during handovers. Many solutions have been proposed to improve mobility and avoid session failure.

3.1. Current Solutions

Since changing IP address is an issue for transport protocols, propositions have been made for TCP to support mobility. TCP-Migrate [SNO 00] uses an extra id negotiated between hosts during connection initiation. If mobile node changes its IP address, remote host will be able to identify TCP session thanks to the extra id. Another solution is MPTCP (MultiPath TCP [RFC 6182]); a transport protocol designed according Next Generation Transport recommendations. It consists in a layer above multiple simple TCP connections allowing using multiple networks simultaneously.

Most propositions to support mobility are located at network layer. [RFC 3344] - Now obsolete by [RFC 5944] - adds mobility support for IPv4 by the use of a Home Agent (HA) and a tunnel between Mobile Node and HA. Main issue of this solution is the triangular communication introduced between MN and Corresponding Node (CN) to route data from Home Address (HoA) to Care-of Address (CoA) attributed in the new network.

Solution proposed by Mobility Support in IPv6 [RFC 6275] uses the same mechanisms for IPv6 plus a "Binding Update" (BU) that allows the update of IP address used for communication and thus to communicate directly between Mobile Node (MN) and Corresponding Node (CN) after an authentication (in a mobile context, authenticate remote host is obviously mandatory for security reasons). Fast Handover for Mobile IPv6 (FMIPv6) defined in [RFC 5568] is one of the most interesting solution based on IPv6, it allows to pre-configure IP address before changing link by defining two entities : the Previous Access Router (PAR) and the New Access Router (NAR). This solution is oriented on inter-domain mobility. For intra domain mobility, a Hierarchical Mobile IPv6 (HMIPv6) defined in [RFC 5380] proposes the use of a Mobility Anchor Point (MAP) to avoid exchanges with the HA when changing access point in the same domain. However, for inter-domain mobility this solution introduced extra exchanges to configure the MAPs. Another solution called Proxy Mobile IPv6 (PMIPv6) proposes to handle mobility in the network with new entities: Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG). MN is no longer aware of mobility and has a fixed Home Network Prefix. All data is tunnelled between LMA and the MAG where MN is connected; propositions have also been made to avoid tunnelling when CN and MN are on the same link. IETF working group Network Mobility (nemo) is an extension of MIPv6 and allows sessions continuity for all nodes in the mobile network. It proposes to apply MIPv6 solutions to Mobile Router (MR) and thus defining necessary exchanges.

These solutions are interesting since they can provide transparent mobility for applications and ameliorations provided to MIPv6 allows reducing handover latency. However they all modify network architecture by adding new entities (HA, PAR, NAR, MAP ...) and some of them even add an overhead with extra headers or tunnelling. Moreover, as QoS support or multicast support, mobility support should be difficult to deploy in different admin domains. We have the conviction that in a near future these solutions will not be available in a large and heterogeneous context. The convergence between a satellite operator, a 3G provider and a private Wi-Fi network is a complex task that will take decades.

3.2. Stream Control Transmission Protocol (SCTP)

SCTP is an IP transport protocol that includes all TCP features: reliability, order delivery, flow control and congestion avoidance. Its first standard was written in 2000; the protocol is now defined by [RFC4960]. Main features introduced by SCTP are:

- Multi-streaming: allows creation of several streams in only one association (SCTP connection),
- Multi-homing: enables use of multiple network interfaces simultaneously.

By using several network interfaces, hosts define multiple paths between them and one is considered primary. If a timeout occurs, a packet is retransmitted on the secondary path. When several timeouts occur in a row, the primary path is considered unavailable and a secondary path becomes the primary. Extensions have been developed to dynamically add IP addresses, thus allowing mobility for SCTP (see section 3.3).

SCTP has been designed for terrestrial links. Studies have been made on SCTP's behaviour over satellite communication networks (showing it can perform as well as TCP [BER 10, SHA 03].

3.3. SCTP extensions for mobility

Thanks to multi-homing, SCTP can connect to multiple networks. An SCTP extension defined in [RFC 5061] allows dynamic IP addresses configuration, meaning that during an association's lifetime, addresses can be added or removed and primary path can be changed. If multi homed, a mobile node is thus able to establish connection with a new network, add the new IP address to association and choose the new network as the primary path. This way, a handover can be performed without breaking the transport connection and without triangular communications. For security reasons, this extension needs to be used with [RFC 4895] that allows chunk authentication so that no third party can generate requests to add new IP or change primary path. These extensions and SCTP together are called mSCTP (mobile SCTP). SCTP mobility has been compared with network layer mobility scheme (see [SIV 05] and [KIM 08]) and results show that it can reduce handover latency and even improve throughput while performing handover compared to mIPv6 mobility.

If mSCTP enables mobility support, it only provides API (Application Programming Interface) to application developers. There is no automatic handover sequence. [KIM 06] presents an algorithm mSCTP-DAC (mobile SCTP-Dynamic Address Configuration) to make handovers as fast as possible:

- When connecting to a new network, a node requests an IP address,
- New IP address is added to association and communicated to remote host along with a request to change the primary path,
- Heart Beat (HB) is sent on new path to check its availability,
- New path becomes primary once it is available for both hosts. Data is now sent on the new path.

To evaluate handover, a key delay is the time spent between network detection and its use as the primary path. The time needed to add an IP address and check path availability is fixed (bounded??) by the network RTT (Round Trip Time) since acknowledgments have to been taken into account. But delay introduced by first step (i.e. getting a new IP address) will depend on the method used and on hardware configuration. Dynamic Host Configuration Protocol (DHCP) is used in most networks to configure new nodes interfaces and introduces delay due to information exchanges. In the next section, a solution based on static IP address will be presented and compared with the above algorithm.

4. Using localization to avoid failover

4.1. Behaviour

Public transports can be easily localized via GPS and odometers. The solution introduced in this section puts together localization and a "map" of all available networks along vehicle run in order to improve multi homed SCTP handover.

When a public transport has to go inside the range of a known access point, it can be expedient to establish connection and use it as primary path during the time spent in its range. By knowing access points geographic position, a mobile node can anticipate handover by configuring a network interface in advance and sending its new IP address to remote hosts (via SCTP Add-IP requests). Besides localization, other factors have to be taken into account:

- In case several networks are available, choose the best fitted one based on parameters introduced in 2.2,
- Determine interest of handover: depends on time spent in range and associations states.

If at least one association is active and handover is necessary, the following sequence is performed:

- "Network map" advertises a new AP will soon be available and handover will be needed,

- Network interface is configured (start-up, network settling, IP address modification and route rules setting),
- Local host adds new IP address in association and sends Add-IP request to remote host,
- When the new network is detected, mobile node connects, sends a heart beat on new path and the request Change-Prim-Addr on old path.
- Upon acknowledgment receipt, new path can be used as primary.

Anticipating connection to a new access point allows configuring network interface and to adding the new IP address before handover and thus reducing the handover latency. Next sub-section analyzes this behaviour by describing exchanges needed between nodes for this approach.

4.2. Temporal Analysis

Main point of this analysis is to demonstrate in a logical way that the use of this solution is more efficient than a standard handover with mSCTP and therefore will be compared with mSCTP-DAC. Here are the abbreviations used in the analysis :

- Tc - Time needed to configure network interface (depends on method used and hardware configuration,
- Td_i - End-to-end delay between two hosts on path i,
- RTT_i - Round Trip Time on path i,
- t_0 - time when network is detected,
- t_1 - begin of configuration sequence,
- Tm - Time needed by mSCTP to fully perform handover,
- Ts_i - Time needed by our solution to fully perform handover from starting time t_i.
- A/Aa - Chunk ASCONF/ASCONF-Ack emission,
- HB/HBa - Chunk HB/HB-Ack emission.

Handover time that will be study is time spent between new network detection and data sent on new path.

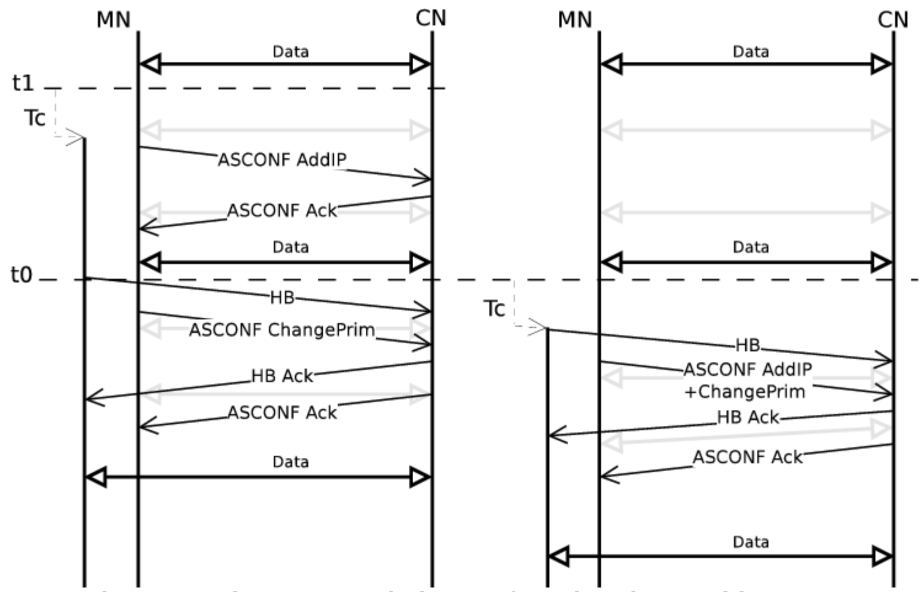

Fig. 1 – Exchanges needed to perform handover with SCTP.

Fig. 1 presents time necessary for handover with the two methods. Chunks ASCONF and HB can be sent at the same time, their position has been chosen here for reading purpose only. To consider handover effective, both hosts have to consider the new path as available: even if one node can emit on new path when receiving requests, we will take into account delivery time of acknowledgments. It has also to be noticed that in these graphs the CN is not multi homed: according to RFC it is not mandatory, only MN needs to be multi homed.

$$Tm = Tc + \max\left(Td_1(A) + Td_1(Aa),\ Td_2(HB) + Td_2(HBa)\right) \quad [1]$$

$$Tm = Tc + Td_1(A) + \max\left(Td_1(Aa),\ Td_2(HB) + Td_2(HBa)\right) \quad [2]$$

Equation [1] express handover time if heart beat is sent by mobile node along with ASCONF request, on contrary if CN is sending heart beat after receiving request a one-way delay is added [2]. Our solution can be expressed in two ways, by starting handover time at the beginning of configuration sequence (t_1) or by starting it when network is detected (t_0):

$$Ts_0 = \max(Td_1(A) + Td_1(Aa),\ Td_2(HB) + Td_2(HBa)) \quad [3]$$
$$Ts_1 = Tc + Td_1(A) + Td_1(Aa) + Ts_0 \quad [4]$$

For our solution to be effective and [3] to be true, the network interface has to be configure, the IP address sent to remote host and acknowledge before network detection. The minimal time needed for these actions can be deducted from [3] and [4]:

$$t_1 - t_0 \geq Tc + Td_1(A) + Td_1(Aa) \quad [5]$$

Expressions based on RTT:

$$Tm = Tc + \max(RTT_1, RTT_2) \quad [6]$$
$$Tm = Tc + Td_1(A) + \max(Td_1(Aa),\ RTT_2) \quad [7]$$
$$Ts_0 = \max(RTT_1,\ RTT_2) \quad [8]$$

Depending on which host sends the heart beat request and network characteristics, solution benefit can differ:

- In the worst case, heart beat is sent by corresponding node [7] :
  - Both networks suffer the same delay (Td1=Td2=Td) and solution benefit equals to configuration time plus end-to-end delay : Tc + Td,
  - If second network suffers a smaller delay (Td1>RTT2) then benefit is only configuration time Tc.
- Heart beat is sent by mobile node [6]: benefit is only configuration time Tc.

This solution is interesting since it allows decreasing handover time by at least Tc and at best 30% of total time. An important point of this analysis is the minimal time needed before network detection, proposed solution will only be efficient if [5] is verified. It is also noticeable that mSCTP is sending one more control message when network is detected, thus ASCONF chunks are grouped in the same SCTP packet and their size is small enough to ignore sending time compared to propagation time.

5. Experiments

5.1. Scenarios

The proposed solution has been checked on a test-bed composed of four nodes: two APs, one MN and one fixed CN. One access point is a Wi-Fi gateway and will help in emulating mobility by switching off/on the access network. The other access point is SATEM, a satellite link emulator that emulates connection to a communication satellite network. Both nodes are multi homed and connected on one path by satellite link and on the other path by Wi-Fi link.

To illustrate mobility, MN is connected to satellite link when association is initiated and Wi-Fi link is down. When MN detects Wi-Fi AP, connection is established and handover is performed. Application generates data at 50 kilo-Bytes per seconds with 500 Bytes packets size. Two handovers algorithms are compared: mSCTP-DAC and our solution.

5.2. Results and analysis

This section presents the measurements done on the platform. Both algorithms have been used with the above scenario several times to compare their behaviour and efficiency. Last results show the impact of multiple handovers on communications.

Fig. 2 (a) and (b) present application delay when performing a handover between satellite link and Wi-Fi link. In Fig. 2 (a), it can be seen that IP is added to association before connection to network while in Fig. 2 (b) can be added only after connection to network. When connected to network, Heart Beat is sent also faster with our solution and thus allows to perform handover faster. This is due to the time needed for route configuration and establishment: an additionnal latency is introduced between IP address configuration and HB emission (i.e. link avaibility).

With DAC algorithm, the IP is configured and added to association 20ms after connection to network. In the experiments we didnot use any DHCP server and IP addresses are configured in a static way. If a DHCP server is used, additionnal latencie is introduced as configuration time (see 4.2).

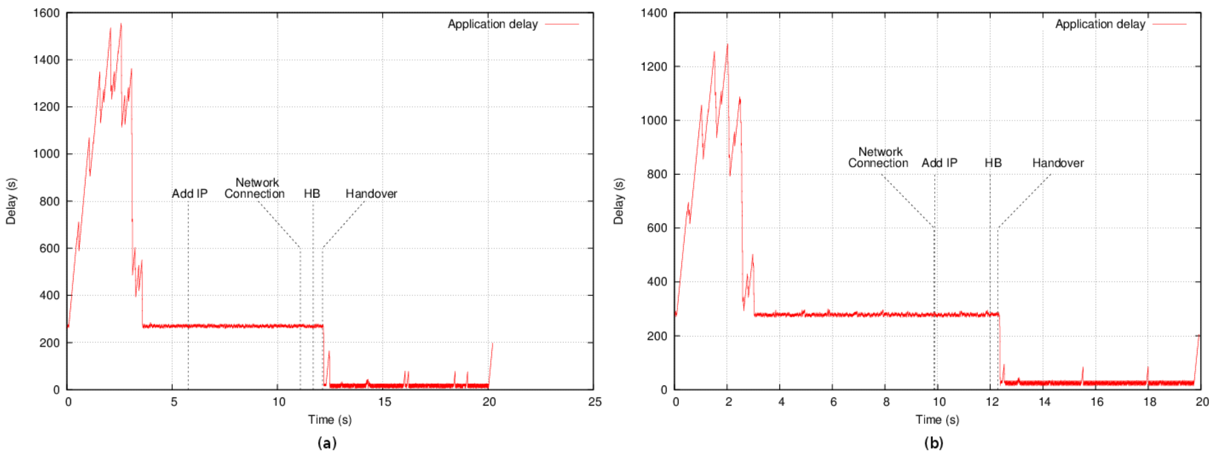

Fig. 2 - Application delay while performing handover from satellite to Wi-Fi with our solution (a) and with DAC algorithm (b).

In order to evaluate the differences between the two solutions, experiments have been run ten times. Table 2 presents average handovers latencies when performing handover. On the global delay between connection to the new network and emission of data chunks on the new path, our solution enhance handover by 900ms (corresponding to 13 meters at 50KMph). Table 2 also illustrates timings from network connection to HB emission and from network connection to route configuration. It can be seen that the most important latencie is introduced by route configuration and link establishment. Interresting point is that configuring network interface in advance can lowered that delay by half.

	Network Connection to Data Emission	Network Connection to HB Emission	Network Connection to Route Configuration
SOL	1,86s	0,890s	0,889s
DAC	2,79s	1,84s	1,84s

Table 2 – Average latencies when performing handover with DAC algorithm and our solution.

Fig. 3 (a) and (b) illustrates application behaviour while changing primary path several times. In this experiment, data was generated at 100 kilo-Bytes per seconds and there were no bandwidth limitation. Changes on global throughput when performing handover is limited by SCTP multi-homing : since congestion window is computed for each path, it can restart where it stops and no time is loss when going back on satellite network. In satellite networks, “slow-start” phase lasts several seconds and limits application throughput at start.

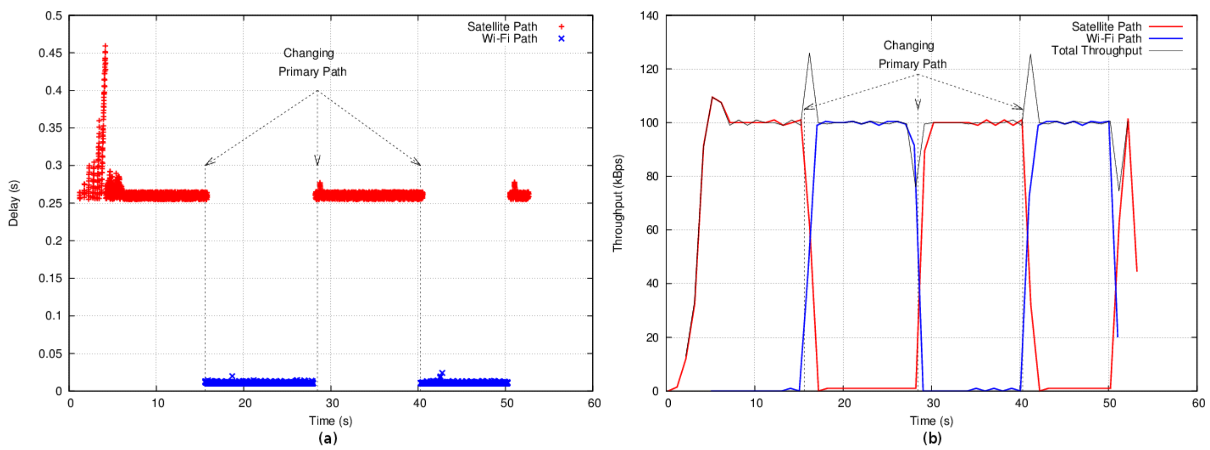

Fig. 3 – Impact on communication of multiple handovers: delay (a) and throughput (b).

Above results shows that SCTP handover can be improved with our solution by reducing the latency between network detection and link avaibility.

6. Conclusion and Perspectives

In order to guarantee network connectivity onboard, a public transport needs to resolve transport layer connexions breaks while performing handovers. Many solutions have been conceived at network layer as well as at transport layer, but few have been implemented. Indeed most of them use additional entities or functionalities that cannot be deployed in nowadays networks due to their architecture or private administration. Our solution focuses on transport layer and uses SCTP multi-homing features along with localization to predict handovers and perform them faster.

Results obtained from experiments verify expected enhancement from temporal analysis. It can save some time by pre-configuring network interfaces. By reducing handover time, more time spent in range of access point can be used for communications. In a public transport network, impact can be consequent since handovers can happen frequently with mid-range technologies (such as Wi-Fi).

Further studies will be dedicated to use this solution along with network layer solutions based on mobile IP: instead of having a static address, the mobile node could use our "network map" to request a new IP before network detection (as done in FMIPv6). Implementation also needs some improvements concerning "network map" to automate handover prediction, this map could be generating from a real bus lane by detecting available networks along its path.

These experiments were done using SATEM; a SATellite EMulator developed by ASTRIUM and real Wi-Fi access points. Implementations and tests with a real Operating System helped in defining configurable protocol parameters and accessible network information. This information can be used to configure a network simulator according to real-world possibilities. Making tests on a simulator will allow increasing the number of access points and simulating mobility between them.

References

[BER 10] Bertaux L., Gayraud T., Berthou P., *How is SCTP Able to Compete with TCP on a QoS-aware Satellite Network?*, Advances in Satellite and Space Communications (SPACOMM) 2010, Athens, 13-19 June 2010.

[KIM 06] Kim D.P. et al., *mSCTP-DAC: Dynamic Address Configuration for mSCTP Handover,* Embedded and Ubiquitous Computing EUC'06, August 2006.

[KIM 08] Kim D.P.; Koh S.J.; *Analysis of Handover Latency for Mobile IPv6 and mSCTP,* IEEE International Conference on Communications Workshops, pp.420-424, 19-23 May 2008

[SHA 03] Shaokian F., Atiquzzaman M., Ivancic W., *SCTP over satellite networks*, IEEE 18th Annual Workshop on *Computer Communications, CCW'2003*, pp. 112–116, October 2003.

[SNO 2000] Snoeren A.C., Balakriishnan H., *An End-to-End approach to Host Mobility*, 6th ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'00), Boston, August 2000.

[SOR 09] Soroush H., Banerjee N., Balasubramanian A., Corner M.D., Levine B.N., *DOME: A Diverse Outdoor Mobile Testbed*, ACM Intl. Workshop on Hot Topics of Planet-Scale Mobility Measurements (HotPlanet), June 2009.

[SIV 05] Sivagurunathan, S.K.; Jones, J.; Atiquzzaman, M.; Shaojian Fu; Yong-Jin Lee; *Experimental comparison of handoff performance of SIGMA and Mobile IP*", Workshop on High Performance Switching and Routing, HPSR, pp. 366- 370, 12-14 May 2005.

[RFC 4960] R. Stewa1rt, *Stream Control Transmission Protocol*, RFC 4960, September 2007.

[RFC 5061] R. Stewart et al., *SCTP Dynamic Address Configuration*, RFC 5061, September 2007.

[RFC 5568] R. Koodli, *Mobile IPv6 Fast Handovers*, RFC 5568, July 2009.

[RFC 5944] C. Perkins, *IP Mobility Support for IPv4 Revised*, RFC 5944, November 2010.

[RFC 6182] A. Ford, C. Raiciu, M. Handley, S. Barre, J. Iyengar, *Architectural Guidelines for Multipath TCP Development*, March 2011.

[RFC 6275] C. Perkins, D. Johnson, J. Arkko, *Mobility Support in IPv6*, July 2011.

[RFC 5380] H. Soliman, C. Castelluccia, K. ElMalki, L. Bellier, *Hierarchical Mobile IPv6 (HMIPv6) Mobility Management*, October 2008.