Performance Analysis and Comparison of the MIPv6 and mSCTP Based Vertical Handoff^{*}

Shi Yan¹, Chen Shanzhi², Ai Ming¹, and Hu Bo¹

¹ State Key Laboratory of Networking and Switching, Beijing University of Posts and Telecommunications, 100876 ² China Academy of Telecommunications Technology, Beijing, 100083 catshiy@gmail.com, {chenshanzhi,aimingam}@yahoo.com.cn, hubo@bupt.edu.cn

Abstract. Vertical handoff is the key technology supporting session mobility in the future heterogeneous network environments. Based on the asymmetry feature of vertical handoff, this paper analyzes the handoff procedure of MIPv6 and mSCTP in forced and unforced handoff scenarios. Qualitative and quantitative analysis and comparison of the handoff performance including handoff delay, handoff packet loss and signaling overhead are given accordingly. Besides, the main factors influencing vertical handoff performance are pointed out and possible performance improvement schemes are discussed.

Keywords: Vertical handoff, MIPv6, mSCTP, Handoff performance.

1 Introduction

Handoff is one of the key technologies in mobility management. To support session mobility in the future ubiquitous and heterogeneous network environments, the research focus turns from horizontal handoff between homogeneous access technologies to vertical handoff between heterogeneous access technologies.

According to the handoff direction, vertical handoff can be classified into upward and downward handoff. Upward handoff is from the network with small coverage to the network with large coverage, e.g. the handoff from WLAN to GPRS. The reverse handoff is called downward. Different handoff directions bring notable asymmetry in handoff scenarios, procedures and performance. There are two typical scenarios – forced and unforced handoff. Forced handoffs are usually triggered by lower layer events incurring the change of interface availability. Since only one interface is available, the handoff is obligatory to avoid communication interruption. Unforced handoffs are triggered by users actively according to user policies, preferences or perceived QoS. And multiple interfaces are available simultaneously. Downward handoff must be unforced, while upward handoff may be forced or unforced.

In order to ensure the independency of access technologies, vertical handoff is implemented at network layer or above. MIP (Mobile IP)[1][2] and mSCTP (mobile

^{*} This work is supported by the National High-Technology (863) Program of China under Grant No. 2006AA01Z229, 2005AA121630 and 2003AA121530.

Stream Control Transport Protocol)[3][4] are two typical vertical handoff protocols. MIP operates at network layer and provides server-based handoff, while mSCTP operates at transport layer and provides end-to-end based handoff. [5-10] propose the vertical handoff schemes with simple performance analysis. Some handoff performance comparison of MIPv4 and mSCTP is made in [11]. But in all these works, the asymmetry feature, forced handoff scenario and error-prone feature of wireless links are not considered adequately. This paper analyzes the vertical handoff performance of MIPv6 and mSCTP. Qualitative and quantitative results are given. The asymmetry feature of vertical handoff and lossy feature of wireless links are considered. The rest of this paper is organized as follows: section 2 gives the qualitative analysis based on the unified analysis model. Section 3 gives quantitative results. Section 4 concludes this paper, the main performance influencing factors are pointed out and possible performance improvement schemes are discussed.

2 Vertical Handoff Performance Analysis

Generally, vertical handoff performance can be evaluated by handoff delay, handoff packet loss and signaling overhead. Handoff delay refers to the connection interrupted interval. Packet loss is often evaluated by the interval incurring loss. Signaling overhead is defined as the traffic caused by the signaling messages exchange during the handoff. In wireline networks, it is denoted by the product of signaling message packet size and transmission distance (hops). Considering the limited bandwidth on wireless links, the product should be multiplied by a factor α , where $\alpha > 1$.

2.1 Unified Analysis Model

The handoff performance analysis is based on different subprocedure combinations and sequences. The main subprocedures include: (1) L2 handoff: the link layer connection change procedure. (2) System discovery and movement detection: MN recognizes the change of network connectivity. (3) Access authentication: the necessary procedure for handoff across heterogeneous access networks belonging to different domains or operators. (4) Address configuration: the network assigns an IP address to MN. (5) Handoff protocol signaling subprocedure: the signaling exchange for implementing the handoff, e.g. the location registration procedure in MIPv6, the setting primary address procedure in mSCTP. (6) Assistant subprocedure: the handoff assistant operations, e.g. the unavailable IP address deletion in mSCTP.

2.2 Vertical Handoff Procedure and Qualitative Performance Analysis

A. Vertical Handoff Based on MIPv6

The forced vertical handoff procedure based on MIPv6 is illustrated in Fig.1, which includes several serial subprocedures: L2 handoff, movement detection, access authentication, address configuration and MIPv6 location registration. The introduced delay are denoted by T_{L2HO} , $T_{MoveDet}$, T_{Auth} , $T_{AddrCon}$ and T_{Regis} . We will analyze $T_{MoveDet}$ and T_{Regis} , which are protocol-related. The others are determined by specific access technologies.

The movement detection adopts L3 detection [2]. MN continuously listens to the RA (Router Advertisement) messages multicasted by the routers periodically with interval Δ_{RA} . If MN has not received any RA from the default router in Δ_{RA} , it can determine one RA miss. After several RA misses (here assumes three for a reasonable tradeoff between efficiency and robustness), MN can confirm an IP connectivity change. [12] defines the interval as 3-4s at least and 1350-1800s at most. In order to support mobility efficiently, [2] decreases it to 30-70ms. Hence, the movement detection duration can be given by $T_{MoveDet} = 3 \cdot \Delta_{RA}$.



Fig. 1. Forced Vertical Handoff Procedure Based on MIPv6

The location registration procedure includes HA registration and CN registration. [2] dictates that MN is permitted to send BU (Binding Update) to CN only after it receives BA (Binding Acknowledgement) from HA, that is, HA registration and CN registration are serial procedures.

In the unforced vertical handoff based on MIPv6, MN has multiple available networks simultaneously. As illustrated in Fig. 2, system discovery, authentication and address configuration can be completed when MN continues communication through the former network interface. Only after the handoff decision, the MIPv6 location registration procedure is triggered for link switch.



Fig. 2. Unforced Vertical Handoff Based on MIPv6

The detailed performance for MIPv6-based forced and unforced handoff is summarized in Table 1. The packet loss is denoted by the duration incurring loss.

B. Vertical Handoff Based on mSCTP

mSCTP supports vertical handoff because of its multi-homing feature and DAR (Dynamic Address Reconfiguration) extension [13]. With the movement of MN, mSCTPbased handoff includes three DAR operations: adding new acquired IP address, changing the primary IP address and deleting the unavailable IP address. All these DAR operations are implemented through a couple of ASCONF and ASCONF_ACK messages carrying different parameters. Fig. 3 illustrates the procedures of the downward unforced handoff and upward forced handoff.

The forced vertical handoff based on mSCTP includes three serial subprocedures: movement detection, setting primary IP address and deleting unavailable IP address.



Fig. 3. Unforced and Forced Handoff Based on mSCTP

Because of the multi-homing feature, the new primary IP address is the existing and authenticated address in the association. Hence, authentication and address configuration are not necessary.

mSCTP uses the path failure detection function [3] for movement detection. An error counter is maintained for the primary IP address. It will be incremented each time the retransmission timer T3-rtx expires. When the error counter is incremented successively and exceeds the threshold PMR (Path Maximum Retransmissions), this address is regarded as unreachable and the movement is detected. The initial value of T3-rtx is RIO_{nSCTP}^{T3} =1s. Hence, the movement detection duration is $T_{MoveDet} = \sum_{i=0}^{PMR} 2^i RTO_{mSCTP}^{T3}$.

The key signaling procedure for handoff based on mSCTP is changing the primary IP address. The old IP deletion belongs to the assistant procedure. It incurs signaling overhead but it should not be included in the handoff delay.

The unforced handoff includes several subprocedures: system discovery, authentication, address configuration, adding IP address, setting primary IP address. The former ones are completed when MN continues communication through the old interface. Only after handoff decision, the primary address is changed for path switch.

The detailed performance of mSCTP is summarized in Table 1. It is notable that the packet loss in unforced handoff is 0 because mSCTP supports soft handoff.

Scenarios		MIPv6	mSCTP
forced	Handoff Delay	$D_{MIPv6}^{forced} = T_{L2HO} + 3 * \Delta_{RA} + T_{Auth} + T_{AddrCon} + T_{HA} = Reals + T_{CN} = Reals$	$D_{mSCTP}^{forced} = \sum_{i=0}^{PMR} 2^{i} R T O_{mSCTP}^{T3}$ $+ T_{o} P_{o} + M$
	Packet Loss	$LD_{MIPv6}^{forced} = D_{MIPv6}^{forced}$	$LD_{mSCTP}^{forced} = D_{mSCTP}^{forced}$
	Signaling Overhead	$\begin{split} C^{forced}_{MIPv6} &= (L_{BU} + L_{BA}) \bullet (\alpha + d_{MN-HA} - 1) \\ &+ (L_{BU} + L_{BA}) \bullet (\alpha + d_{MN-CN} - 1) \end{split}$	$C_{mSCTP}^{forced} = 2 \cdot (L_{ASCONF} + L_{ASCONF-ACK}) \\ \cdot (\alpha + d_{MN-CN} - 1)$
unforced	Handoff Delay	$D_{MIPv6}^{unforced} = T_{HA_Regis} + T_{CN_Regis}$	$D_{mSCTP}^{unforced} = T_{SetPrimAddr}$
	Packet Loss	$LD_{MIPv6}^{unforced} = D_{MIPv6}^{unforced}$	$L_{mSCTP}^{unforced} = 0$
	Signaling Overhead	$C_{MIPv6}^{unforced} = (L_{BU} + L_{BA}) \cdot (\alpha + d_{MN-HA} - 1) $ + $(L_{MV} + L_{PA}) \cdot (\alpha + d_{MN-HA} - 1)$	$C_{mSCTP}^{forced} = 2 \cdot (L_{ASCONF} + L_{ASCONF-ACK})$ $\cdot (\alpha + d_{average} - 1)$

Table 1. Vertical Handoff Performance Comparison for MIPv6 and mSCTP

2.3 Analysis of the Handoff Signaling Delay

Handoff signaling procedure is unaviodable in any scenario. In this part, we will analyze the signaling delay (T_{Regis} in MIPv6 and $T_{SetPrimAddr}$ in mSCTP) in detail. Considering the error-prone feature of wireless links, the delay incurred by the error

recovery mechanism should be taken into account. Refering to and extending the analysis method in [14], the signaling delay analysis model is shown in Fig. 4. Here the delay consists of five parts: T_{MN} , T_{SAP} and $T_{CN}(T_{HA})$ denote the average processing delay at MN, SAP (Service Access Point, e.g. BS or AP), CN or HA, $T_{wirefine}$ denote the average transmission delay over wireless and wireline links.



Fig. 4. The Handoff Signaling Delay Analysis Model

Assuming an M/M/1 queuing model for MN, SAP, HA and CN, according to the queuing theory, the average processing delay at each entity is: $T_{Entity} = 1/(\mu_{Entity} - \lambda_{Entity})$, where λ and μ denote the signaling message arriving rate and processing rate. Assuming the distance between SAP and HA or CN is w (hops) and the average transmission delay per hop is \overline{t} , we have $\overline{T_{wireline}} = w \cdot \overline{t}$. The average transmission delay over wireless links will be discussed in the following parts.

A. Average Transmission Delay over Wireless Link in MIPv6

[2] defines timeout retransmission mechanism to ensure the reliable transmission of BU and BA messages. The initial value and the threshold of the retransmission timer are 1s and 32s [2], i.e., the max retransmission is $N_{MIPv6} = 5$.

Assuming the FER (Frame Error Rate) at link layer is p, the bandwidth of the wireless link is *B* bps and the inter-frame time is τ s, for the IP packet with size *L*, the frame numbers contained in it is $k = L/(B \cdot \tau/8)$.

For the HA registration procedure, the transmission error of either BU or BA will incur retransmission. Success registration implies both BU and BA are transmitted successfully, the probability of which is $p_s = (1-p)^{k_{BU}+k_{BA}}$. Therefore, with the maximum retransmission number $N_{MIP_{V6}}$, the success registration probability is given by: $P = p_s + (1-p_s) \cdot p_s + (1-p_s)^2 \cdot p_s + \dots + (1-p_s)^{N_{MIP_{V6}}-1} \cdot p_s$. Correspondingly, the average transmission delay over wireless link caused by HA registration is given by:

$$\overline{T}_{wireless_HA_Reg}^{MIPv6} = p_s \bullet T_{MIPv6}^{Iran} + (1 - p_s) p_s \bullet (T_{MIPv6}^{Iran} + RTO_{MIPv6}) + \cdots + (1 - p_s)^{N_{MIPv6} - 1} p_s \bullet (T_{MIPv6}^{Iran} + (2^0 + 2^1 + \cdots + 2^{N_{MIPv6} - 2}) RTO_{MIPv6})$$
(1)

where $T_{MIPv6}^{tran} = (D + (k_{BU} - 1)\tau) + (D + (k_{BA} - 1)\tau)$ is the transmission delay of BU and BA over the wireless link and D is the end-to-end frame propagation delay.

Obviously, for CN registration, we have the similar result.

B. Average Transmission Delay Over Wireless Link in mSCTP

mSCTP forms packets from the ASCONF and ASCONF-ACK control chunks and here we do not consider the bundling. mSCTP also uses retransmission mechanism for reliable transmission [13]. The timer is T4-rtx with initial value 1s. Considering

the multi-homing feature, the retransmission is implemented on the secondary path if it exists. In addition, an individual retransmission timer is maintained for each path.

For the forced handoff, the secondary path does not exist and the retransmission will be completed over the primary path. The maximum retransmission times should be PMR=5 [3]. Therefore, assuming the same FER for primary and secondary paths, we have similar conclusion to MIPv6.

For the unforced handoff, there exists the secondary path for retransmission and the maximum retransmission times should be AMR (Association Maximum Retransmission) =10 [3]. Considering the individual retransmission timer for each path, the average wireless transmission delay for unforced handoff can be given by:

$$\bar{T}_{wireless}^{mSCTP} = p_{s} \bullet T_{mSCTP}^{tran} + (1 - p_{s}) p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{2} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{tran} + RTO_{mSCTP}^{T4}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{ta}) + (1 - p_{s})^{3} p_{s} \bullet (T_{mSCTP}^{ta$$

3 Numerical Results

Based on the analysis in section 2 and referring to the typical parameters defined in the protocol recommendations [2][3][4][13] as well as related researches [6][15][17], the numerical results are given in the following part.



Fig. 5. Vertical Handoff Performance Comparison of MIPv6 and mSCTP in Upward Forced and Downward Unforced Handoff Scenarios

3.1 Vertical Handoff Performances Comparison

Fig. 5 shows the performance of MIPv6 and mSCTP in forced and unforced handoffs.

(1) For the same protocol, the handoff delay and packet loss in unforced handoff is better than forced handoff obviously. This is due to two reasons. Firstly, movement detection, authentication and address configuration contribute to the delay in forced handoff, but in unforced handoff they are completed before handoff is triggered and do not incur handoff delay. Secondly, signaling messages are transmitted over the higher performance link in downward unforced handoff



Fig. 6. Components of Vertical Handoff Delay

whereas over the lower performance link in upward forced handoff. This causes the difference in the signaling delay. Fig. 6 illustrates the components of the handoff delay in each scenario. We can see the effects of the above two reasons apparently. (2)Although authentication and address configuration are not necessary in mSCTP-based forced handoff, its handoff delay and packet loss is rather worse than MIPv6. This is due to the long duration for movement detection. mSCTP is designed for wireline networks initially and many parameter values are not suitable for wireless and mobile networks. E.g., the initial values of the retransmission timers are defined as 1s, which incurs large delay for movement detection. Reversely, MIPv6 decreases the interval between two successive RAs for better mobility supporting. This reduces the movement detection delay efficiently. (3) In the unforced handoff, MIPv6 has larger handoff delay. This is because MIPv6 requires the serial HA and CN registration. (4) As for handoff packet loss, it is proportional to the handoff delay in most scenarios. In mSCTP-based unforced handoff, the loss is 0 because mSCTP supports soft handoff.

3.2 Handoff Signaling Delay Comparison

Handoff signaling delay is the unavoidable component in any scenario. In unforced handoff, it is the major part of handoff delay. The signaling delay with the variation of FER is shown in Fig. 7, and Fig. 8 illustrates the detailed composition.



Fig. 7. Handoff Signaling Delay vs. FER

Fig. 8. Components of Signaling Delay

For both MIPv6 and mSCTP, the signaling delay of upward forced handoff is obviously lager than that of downward unforced handoff. We can see in Fig.8, the main difference exists in the wireless part transmission delay. This is because the packet loss rate at upper layer and the end-to-end frame propagation delay of lower bandwidth link are all larger than higher bandwidth link. In addition, the signaling delay of MIPv6 is lager than mSCTP due to the serial HA and CN registrations.

4 Conclusions

Performance Analysis and comparison of the MIPv6 and mSCTP based vertical handoff in different scenarios are given in this paper, considering the asymmetry in handoff directions, scenarios, procedures. Numerical results show that the handoff performances, especially the forced handoff performance based on standard protocols are not satisfying. For the more and more delay-sensitive realtime services with maximum acceptable interruption about 200ms, such a performance will be unacceptable. Therefore, it is stringent to improve the handoff performance. The main reasons include: (1) in forced handoff, the serial procedures such as movement detection, authentication and address configuration incur considerable delay; (2) the low bandwidth and error-prone feature of wireless link introduce high message transmission delay for error recovery; (3) some parameters defined in the standard protocols can not support handoff in wireless mobile networks efficiently.

The possible methods to improve the vertical handoff performance include: (1) In forced handoff, cross-layer interaction should be introduced to provide lower layers information to upper layer handoff protocols. This is helpful to avoid long movement detection delay and realize proactive handoff control rather than reactive handoff. The serial relationships can be changed, e.g. pre-authentication, pre-address-configuration, pre-registration etc. MIH (Media Independent Handoff) [16] defined by IEEE 802.21 WG provides an efficient cross-layer interaction mechanism. (2) Some parameters defined in the protocol standards should be modified for more efficient mobility supporting. (3) Special measures should be adopted to reduce the handoff packet loss. Soft handoff is another issue in handoff performance optimization.

References

- 1. Perkins, C.: IP Mobility Support for IPv4. RFC 3344 (2002)
- 2. Johnson, D.: Mobility Support in IPv6. RFC 3775 (2004)
- 3. Stewart, R., Xie, Q., et al.: Stream Control Transmission Protocol. RFC 2960 (2000)
- 4. Riegel, M., Tuexen, M.: mobile SCTP. draft-riegel-tuexen-mobile-sctp-06 (2006)
- Vogt, C., Zitterbart, M.: Efficient and Scalable, End-to-End Mobility Support for Reactive and Proactive Handoffs in IPv6. IEEE Communications Magazine, Vol. 44, No.6 (2006)
- Choi, H.-H., Song, O., Cho, D.-H.: A Seamless Handoff Scheme for UMTS–WLAN Interworking. Proceedings of IEEE GlobeCom (2004) 1559-1564
- Bernaschi, M., Cacace, F.: Vertical Handoff Performance in Heterogeneous Networks. Proceedings of ICPPW (2004) 100-107
- Lee, C.-W., Chen, L.-M., Chen, M.-C., Sun, Y.-S.: A Framework of Handoffs in Wireless Overlay Networks Based on Mobile IPv6. IEEE Journal on Selected Areas in Communications, Vol. 23, No. 11 (2005) 2118-2128
- Ma, L., Yu, F., Leung, V.C.M.: A New Method to Support UMTS/WLAN Vertical Handover Using SCTP. IEEE Wireless Communications, Vol.11, No.4 (2004) 44-51
- Ma, L., Yu, F., Leung, V.C.M.: SMART-FRX: A Novel Error-Recovery Scheme to Improve Performance of Mobile SCTP during WLAN to Cellular Forced Vertical Handoff. IEEE WCNC (2005) 1377-1382
- Song, J.-K.: Performance Evaluation of Handoff between UMTS/802.11b based on Mobile IP and Stream Control Transmission Protocol. Master Thesis (2005)
- 12. Narten, T., Nordmark, E., Simpson, W., Soliman, H.: Neighbor Discovery for IP Version 6. draft-ietf-ipv6-2461bis-06 (2006)
- 13. Stewart, R., Xie, Q., Tuexen, M., Maruyama, S., Kozuka, M.: SCTP Dynamic Address Reconfiguration. draft-ietf-tsvwg-addip-sctp-17 (2006)
- Banerjee, N., Basu, K., Das, S.K.: Hand-off Delay Analysis in SIP-based Mobility Management in Wireless Networks. Proceedings of IPDPS (2003)
- 15. Banerjee, N., Wu, W., Basu, K., Das, S.K.: Analysis of SIP-based mobility management in 4G wireless networks. Computer Communications, Vol. 27, No. 8 (2004) 697-707
- 16. Draft IEEE Std. P802.21/D00.05: Media Independent Handover Services. (2006)
- He, X.-Y., Liu, Q., Lei, Z.-M.: Location Pre-query Scheme to Internet Mobility Support. Journal of Software, Vol. 15, No. 2 (2004) 259-267